

1 **Title**

2 SARS-CoV-2 human challenge reveals single-gene blood transcriptional biomarkers that discriminate early
3 and late phases of acute respiratory viral infections.

4 **Authors**

5 Joshua Rosenheim¹, Rishi K Gupta², Clare Thakker¹, Tiffeney Mann¹, Lucy CK Bell¹, Claire M Broderick³,
6 Kieran Madon⁴, Loukas Papargyris³, Pete Dayananda³, Andrew J Kwok^{5,6}, James Greenan-Barrett⁷, Helen R
7 Wagstaffe³, Emily Conibear⁴, Joe Fenn⁴, Seran Hakki⁴, Rik GH Lindeboom⁸, Lisa M Dratva⁸, Briac Lemetais¹,
8 Caroline M Weight¹, Cristina Venturini⁹, Myrsini Kaforou³, Michael Levin³, Mariya Kalinova¹⁰, Alex Mann¹⁰,
9 Andrew Catchpole¹⁰, Julian C Knight⁵, Marko Z. Nikolić⁷, Sarah A. Teichmann⁸, Ben Killingley¹¹, Wendy
10 Barclay³, Benjamin M Chain¹, Ajit Lalvani⁴, Robert S Heyderman¹, Christopher Chiu³, Mahdad Noursadeghi¹.

11 **Affiliations**

- 12 1. Division of Infection and Immunity, University College London, London, UK
13 2. Institute of Health Informatics, University College London, London, UK
14 3. Department of Infectious Disease, Imperial College London, London, UK
15 4. NIHR Health Protection Research Unit in Respiratory Infections, National Heart and Lung Institute, Imperial
16 College London, London, UK
17 5. Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK.
18 6. Department of Medicine and Therapeutics, Faculty of Medicine, The Chinese University of Hong Kong,
19 Shatin, Hong Kong.
20 7. Department of Respiratory Medicine, University College London Hospitals NHS Foundation Trust, London,
21 UK
22 8. Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge CB10 1SA, UK
23 9. Infection, Immunity and Inflammation Department, Great Ormond Street Institute of Child Health,
24 University College London, London, UK
25 10. hVIVO Services Ltd., London, UK
26 11. Department of Infectious Diseases, University College London Hospital NHS Foundation Trust, London,
27 UK

28 **Correspondence**

29 Professor Mahdad Noursadeghi, Division of Infection and Immunity, University College London, London, UK.
30 Email: m.noursadeghi@ucl.ac.uk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

31 **Summary**

32 Evaluation of host-response blood transcriptional signatures of viral infection have so far failed to test whether
33 these biomarkers reflect different biological processes that may be leveraged for distinct translational
34 applications. We addressed this question in the SARS-CoV-2 human challenge model. We found differential
35 time profiles for interferon (IFN) stimulated blood transcriptional responses represented by measurement of
36 single genes. MX1 transcripts correlated with a rapid and transient wave of type 1 IFN stimulated genes (ISG)
37 across all cell types, which may precede PCR detection of replicative infection. Another ISG, IFI27, showed a
38 delayed but sustained response restricted to myeloid peripheral blood mononuclear cells, attributable to gene
39 and cell-specific epigenetic regulation. These findings were reproducible in diverse respiratory virus
40 challenges, and in natural infection with SARS-CoV-2 or unselected respiratory viruses. The MX1 response
41 achieved superior diagnostic accuracy in early infection, correlation with viral load and identification of virus
42 culture positivity, with potential to stratify patients for time sensitive antiviral treatment. IFI27 achieved superior
43 diagnostic accuracy across the time course of symptomatic infection. Compared to blood, measurement of
44 these responses in nasal mucosal samples was less sensitive and did not discriminate between early and late
45 phases of infection.

46 **Background**

47 Host response biomarkers of viral infection have multiple potential clinical applications. These include
48 diagnostic triage tests to direct prioritisation of confirmatory laboratory investigations, and to guide clinical
49 management decisions with the aim of reducing unnecessary antibacterial prescribing, or directing infection
50 control measures and antiviral treatment. Attention has mostly focussed on biomarker discovery in whole blood
51 samples that enable easy and technically consistent access. Genome-wide transcriptional profiling has
52 emerged as the most common unbiased data-driven approach due to the maturity of technical and analytical
53 workflows¹.

54 Numerous blood transcriptional signatures for host responses to viral infections have been identified in this
55 way using case-control studies of natural infection or experimental viral challenge in humans, designed to
56 discover the most parsimonious measurements that discriminate viral infections from healthy controls or other
57 diseases. We previously tested the accuracy of such blood transcriptional signatures of viral infection, identified
58 by systematic review, to detect incident SARS-CoV-2 infection². We showed that the majority were highly
59 correlated, and collectively driven by type 1 interferon (IFN) responses. Many, including single gene transcripts
60 (such as that of IFI27) provided near perfect discrimination of PCR positive individuals compared to uninfected
61 controls. In some, the transcriptional biomarkers identified infections before the first positive viral PCR in
62 nasopharyngeal samples. The sensitivity of IFI27 measurements was further leveraged to provide evidence
63 for abortive infections associated with virus specific T cell responses without detection of the virus by PCR³.

64 In observational studies of natural infection, it is not possible to synchronise the time course of exposure and
65 replicative infection. This has precluded identification of temporally distinct host response biomarkers that may
66 offer optimal solutions for different translational applications such as diagnostic triage or patient stratification
67 for antiviral therapies. To address this limitation, we leveraged the first controlled human challenge model of
68 SARS-CoV-2 infection, complemented with high frequency sampling to measure viral replication and host
69 responses spanning the full time course of viral replication⁴. We updated our previous systematic review to
70 undertake comprehensive head-to-head evaluation of all reported host transcriptional signatures of viral

71 infection to date. We compared their ability to discriminate between groups of participants with and without
72 evidence of replicative infection using whole blood samples stratified by time since experimental inoculation.
73 For selected biomarkers, representative of differential host-responses over the time course, we evaluated
74 associations with symptoms and viral load. We investigated their cellular source in single cell transcriptomic
75 data, and the potential epigenetic mechanisms that may underpin their differential expression. We compared
76 their measurement in blood and nasal swabs, and explored the extent to which our findings were generalisable
77 to other respiratory viruses, in both experimental challenge and natural infection studies.

78 **Methods**

79 ***Research ethics***

80 Regulatory approvals for the human studies presented herein were provided by the UK Health Research
81 Authority under the following reference numbers: 20/UK/2001 and 20/UK/0002 for the SARS-CoV-2 challenge
82 study; 20/NW/0231 for the INSTINCT study; 19/LO/1441 for the H3N2 influenza challenge study.

83 ***Identification of blood transcriptional signatures of viral infection***

84 We updated our previous systematic review of blood transcriptional biomarkers for viral infection². In the
85 current analysis, we amended our previous eligibility criteria to identify concise blood transcriptional signatures
86 discovered or applied with a primary objective of diagnosis of viral infection from human whole-blood or
87 peripheral blood mononuclear cell samples, excluding those exclusively intended to stratify severity of
88 infection. Other eligibility criteria remained the same as our previous review. In our update, we searched
89 MEDLINE for articles published up to 31 December 2022, using comprehensive MeSH and keyword terms for
90 “viral infection”, “transcriptome”, “biomarker”, and “blood”, as previously². Additional studies were identified in
91 reference lists. Title and abstract screening was independently performed by two reviewers (CT and JGB);
92 shortlisted articles were reviewed in full, with input from a third reviewer (RKG) to resolve conflicts. For eligible
93 signatures, constituent genes, modelling approaches and gene weightings were extracted, with verification by
94 a second reviewer. Multi-gene signatures are referred to using a prefix of the first-author’s name from the
95 corresponding publication, and a suffix of the number of component genes. Single-gene signatures are referred
96 to by the gene symbol.

97 ***Human challenge and patient cohorts***

98 The SARS-CoV-2 human challenge model in healthy seronegative volunteers has been described previously⁴.
99 Briefly, 36 SARS-CoV-2 volunteers were inoculated intranasally with a standardized dose of D614G-containing
100 pre-alpha wild-type SARS-CoV-2 under quarantine conditions. From 24 hours after inoculation, virus was
101 quantified by PCR and culture in samples obtained at 12 hourly intervals from nose (mid-turbinate) and throat
102 swabs for at least 14 days of quarantine, or longer if they remained in quarantine beyond 14 days because
103 they still had detectable virus. A final sample was obtained at 28 days after challenge. Blood samples for RNA
104 sequencing were collected into PAXgene tubes (Qiagen) before virus challenge, 6 hours after challenge, daily
105 thereafter for 14 days and on day 28. Mid-turbinate MW013, X nose swabs (MW013, MedWire) for RNA
106 sequencing were collected before virus challenge, and on days 1, 3, 5, 7, 10 and 14 after challenge, preserved
107 in RNAprotect (Qiagen). Quantitation of viral load and symptoms has been described previously⁴. Two
108 individuals who seroconverted in the interval between screening and inoculation were excluded from the
109 present analysis, on the basis that they experienced a recent infection that may affect the biomarker expression
110 that is the focus of this study.

111 The SARS-CoV-2 household contact study (INSTINCT) has been described previously⁵. Briefly, 52 household
112 contacts of SARS-CoV-2 infected index cases recruited within 5 days of index case symptom onset provided
113 nasopharyngeal swabs and blood RNA samples collected in PAXgene tubes on day of enrolment (day 0), day
114 7, day 14 and day 28. Nasopharyngeal swabs were used to measure viral copy number using PCR against
115 the E-gene.

116 The Influenza H3N2 human challenge model has been described previously⁶. Briefly, 20 healthy volunteers
117 were inoculated intranasally with a standardized dose of Influenza A/Belgium/4217/2015 (H3N2) under
118 quarantine conditions. From 24 hours after inoculation, virus was quantified by PCR in nasal lavage samples
119 obtained at 12 hourly intervals. Participants were ascertained to have replicative viral infection if found to have
120 consecutive positive PCR tests at least 24 hours after challenge. Blood samples for RNA sequencing were
121 collected into PAXgene tubes before virus challenge and days 1, 2, 3, 7, 10, 14, and 28 after challenge. Nasal
122 curettage samples were collected on days -14 (baseline), 1, 2, 3, 7, 10 and 14 and preserved in TRIzol
123 (ThermoFisher Scientific) as previously described⁷.

124 All RNA samples were stored at -80°C until processing.

125 ***Transcriptional profiling***

126 Total RNA was extracted from SARS-CoV-2 challenge PAXgene tubes using the PAXgene Blood RNA kit
127 (Qiagen), including on-column DNase treatment and depleted of globin mRNA using the GLOBINclear Human
128 Kit (Thermo Fisher Scientific). Total RNA was extracted from the INSTINCT SARS-CoV-2 household contact
129 study and the H3N2 influenza challenge PAXgene tubes using the Qiasymphony PAXgene blood RNA kit, with
130 subsequent DNase I treatment (Zymo) and clean-up using the RNA Clean and Concentrator-96 kit (Zymo),
131 followed by globin mRNA and rRNA depletion using NEBNext® Globin & rRNA Depletion kits (New England
132 BioLabs). Total RNA from SARS-CoV-2 challenge nasopharyngeal swabs and curettage samples was
133 extracted using the RNeasy mini kit (Qiagen), including on-column DNase treatment. RNA concentrations were
134 quantified using Qubit 2.0 Fluorometer (ThermoFisher Scientific). RNA integrity scores were determined using
135 the Bioanalyser (RNA Nano 6000 Chip, Agilent) or 4200 Tape Station (Agilent).

136 Blood RNA samples from SARS-CoV-2 challenge underwent total RNA sequencing. DNA libraries were
137 constructed using the KAPA RNA HyperPrep Kit with RiboErase (Roche) and sequenced on the Illumina
138 NovaSeq 6000 platform using the NovaSeq 6000 S4 Reagent Kit (200 cycles) (Illumina), giving a median of
139 69.3 million (range 29.3-152.8) 100 base pair (bp) paired-end reads per sample. Nose swab RNA samples
140 underwent mRNA sequencing. DNA libraries were constructed using the Kappa mRNA HyperPrep kit (Roche)
141 and sequenced on the Illumina NextSeq platform the using the NextSeq 500/550 High Output Kit (75 cycles)
142 (Illumina), giving a median of 32.4 million (range 3.2-176.2) 41bp paired-end reads per sample. Blood RNA
143 samples from the INSTINCT SARS-CoV-2 household contact study underwent mRNA sequencing. DNA
144 libraries were constructed using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (New
145 England Biolabs) and sequenced on the Illumina HiSeq 4000 using the HiSeq 3000/4000 PE Cluster and SBS
146 kits (Illumina), giving a median of 26.1 million (range 18.34-56.04) 75bp paired-end reads per sample. Nose
147 curettage RNA samples from the H3N2 human challenge underwent mRNA sequencing. DNA libraries were
148 constructed using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs)
149 and sequenced on the Illumina HiSeq 4000 using the HiSeq 3000/4000 SBS kit (Illumina), giving a median of
150 74.9 million (range 44.2-122) 75bp paired-end reads per sample. Whole blood RNA samples from the H3N2

151 influenza challenge underwent mRNA sequencing, DNA libraries were constructed with the NEBNext® Ultra
152 II Directional RNA Library Prep Kit for Illumina (New England BioLabs) and sequenced on the Illumina
153 NovaSeq 6000 platform using the NovaSeq 6000 S2 200 cycles Flowcell (Illumina), with a target of 40 million
154 paired-end reads per sample.

155 SARS-CoV-2 challenge sequencing reads were mapped to the reference transcriptome (Ensembl Human
156 GRCh38 release 108) using Kallisto (version 0.46.1)⁸. 371 blood RNA samples from 34 seronegative
157 individuals gave a median of 28.7 million (range 12.2-73.5) mapped reads per sample. 99 nose swab RNA
158 samples gave a median of 23.9 million (range 2.5-138.6) mapped reads per sample. Transcript-level output
159 Deseq2 normalised counts and transcripts per million values were summed on gene level and annotated with
160 Ensembl gene ID, gene name, and gene biotype using the tximport (version 1.20.0) and biomaRt (version
161 2.48.0) Bioconductor packages in R⁹⁻¹³.

162 Sequencing reads from the INSTINCT SARS-CoV-2 household contact study were mapped to the reference
163 transcriptome (NCBI Human GRCh38.p13) using STAR aligner (version 2.7.1a)¹⁴. 134 blood RNA samples
164 from 52 individuals gave a median of 14.37 million (range 7.30-37.67) mapped reads per sample. Read count
165 matrices were generated using featureCounts from the Rsubread package¹⁵ and normalised using the variance
166 stabilised transformation from the DESeq2 package.

167 For the sequencing reads of whole blood RNA from the H3N2 influenza challenge, quality control was
168 performed using with FastQC (v 0.11.7; <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and
169 adapter sequences were removed using Trimmomatic (v 0.36)¹⁶. The reads were mapped against hg38
170 reference genome using STAR aligner (v 2.7.1a). The featureCounts tool from Subread package (v 1.5.2) was
171 used for transcript quantification. Computed gene counts were used for downstream analyses. Whole blood
172 samples from 19 donors gave a median of 10.1 million (range 8.5-12.7) mapped reads per sample (read length
173 = 100bp). Transcript-level output Deseq2 normalised counts were annotated with Ensembl gene ID and gene
174 name using biomaRt (version 2.46.3) Bioconductor packages in R. Sequencing reads from the H3N2 challenge
175 were mapped to the reference transcriptome (Ensembl Human GRCh38.p13) using STAR (version 2.7.10a).
176 Nasal RNA samples from 17 donors gave a median of 21.4 million (range 3.32-34.7) mapped reads per
177 sample. Transcript-level output Deseq2 normalised counts were annotated with Ensembl gene ID and gene
178 name using biomaRt (version 2.52.0) Bioconductor packages in R.

179 For RNAseq datasets that were generated in more than one batch, processing batch effects were excluded by
180 principal component analysis of all RNAseq data (Supplementary Figure 2). Additional genome-wide
181 transcriptomic microarray data were derived from previously published experimental challenge datasets of
182 other respiratory viruses (GEO accession: GSE73072)¹⁷ and from a natural infection study of respiratory
183 viruses (GEO accession: GSE68310)¹⁸. In each case, we used log-2 transformed and normalised data
184 matrices to quantify biomarker scores, standardized to baseline samples.

185 **Signature scores**

186 Analyses were performed in R (version 4.0.2). Scores for candidate transcriptional signatures were calculated
187 as per the original author's descriptions using transcripts per million values, as previously. Where component
188 genes could not be identified in the RNA sequencing dataset (for example due to genes being withdrawn from
189 the reference transcriptome), these genes were excluded from calculations. Scores were standardised to Z
190 scores by subtracting the mean and dividing by the standard deviation of pre-inoculation samples, and were

191 multiplied by -1 for scores intended to decrease in the presence of viral infection. Discrimination of each
192 signature for the outcome of replicative infection was calculated as the area under the receiver operating
193 characteristic curve (AUROC), with 95% confidence intervals, and stratified by day since inoculation, using the
194 pROC package in R¹⁹. Correlation between signatures and with viral loads was quantified as Spearman rank
195 correlation coefficients.

196 ***Analysis of ATACseq data***

197 Publicly available ATAC (Assay for Transposase Accessible Chromatin) sequencing fastq datasets derived
198 from unstimulated human monocytes, B-cells and CD4 T-effector cells (GEO accession: GSE118189,
199 European Nucleotide Archive accession: PRJNA484801)²⁰ were analysed with the nf-core ATAC-seq analysis
200 pipeline (v2.0) curated in Nextflow^{21,22}, using default parameters. Adaptors were trimmed using trimgalore
201 (v0.6.7) and reads were aligned to the reference genome (NCBI GRCh38) using BWA (v 0.7.17)²³. Duplicate
202 reads were identified using picard (v2.27.4)²⁴. Reads were filtered using SAMtools (v1.16.1)²⁵. BEDtools
203 (v.2.30.0)²⁶ was used to remove duplicates, reads mapping to blacklisted regions and mitochondrial DNA,
204 multimappers, unmapped reads or those not marked as primary alignments. Replicate datasets were merged
205 using picard for some downstream analyses. Normalised scaled bigWig files were created using BEDtools and
206 tracks were visualised using Integrative Genomics Viewer (v2.16.0)²⁷. Peak calling was performed using
207 MACS2 (v2.2.7.1)²⁸ in broadpeak mode. Peaks were annotated to gene features using HOMER (v4.11)²⁹ and
208 a consensus peak-set was generated using BEDtools. Matrices of reads falling within consensus peaks were
209 generated using featureCounts from the subread package (v2.0.1)¹⁵ for quantitation.

210 Publicly available single-cell ATACseq data from the COMBAT consortium³⁰ (EGAD00001007963; Zenodo:
211 <https://doi.org/10.5281/zenodo.6120249>) were reanalysed for read counts per cell type in established COVID
212 infection from hospitalised COVID patients. Data were processed as described in the original publication using
213 the ArchR software package (v0.9.3)³¹. The sequencing reads at the *IFI27* locus were plotted per cell type with
214 the plotBrowserTrack function.

215 **Results**

216 ***Blood transcriptional signatures of viral infection***

217 We updated our previous systematic review of the literature, to identify 26 blood transcriptional signatures
218 associated with viral infection (Supplementary Figure 1A, Supplementary Table 1)³²⁻⁵⁶. These included six
219 single gene biomarkers. The remaining multigene signatures were made up of 2-47 constituent genes. The
220 composition of these signatures was generally distinct, reflected by low Jaccard indices in a matrix of pairwise
221 comparisons (Supplementary Figure 1B).

222 ***Viral infection outcomes in the SARS-CoV-2 controlled human challenge model***

223 34 SARS-CoV-2 seronegative healthy volunteers subjected to nasal inoculation of a standardized dose of
224 SARS-CoV-2 divided into two groups with (N=18) and without (N=16) evidence of sustained replicative
225 infection from 2 days after challenge (Figure 1). Although the individual viral load profiles were different in nose
226 and throat swabs, both measurements segregated the same participants into two groups with and without
227 replicative infection.

228 **Blood transcriptional biomarker discrimination of participants with and without sustained replicative**
229 **SARS-CoV-2 infection**

230 Blood transcriptional biomarker scores were calculated for each of the 26 signatures identified by systematic
231 review, from RNA sequencing of whole blood samples at selected time points before and after viral inoculation
232 (Figure 2A). Across this time course, all the biomarkers showed a transient increase in expression
233 (Supplementary Figure 3) associated with replicative SARS-CoV-2 infection. We first ranked all biomarkers by
234 their ability to discriminate between participants with and without replicative infection by area under the receiver
235 operating characteristic curve (AUROC) across 14 days. We limited calculations to data from days 3, 7, 10
236 and 14, in order to achieve equal sampling frequency distribution across the time course of infection
237 (Supplementary Figure 3). Point estimates of the AUROCs ranged between 0.6-0.99. 22 of the 26 biomarkers
238 with point estimates ranging 0.92-0.99 were statistically comparable with overlapping 95% confidence
239 intervals, suggesting most biomarkers were able to accurately discriminate participants with and without
240 replicative infection.

241 **Identification of blood transcriptional biomarkers of early and late phases of SARS-CoV-2 infection**

242 Next, we compared the AUROC of each signature stratified by time point. Most achieved near perfect
243 discrimination of participants with and without replicative infection in days 4-10 (Supplementary Figure 4A).
244 We found greater variation in performance of each signature before and after this time interval, suggesting
245 differential ability to identify early and late phases of viral infection. To investigate this hypothesis further, we
246 focused on the single gene transcripts with highest AUROC on day 3 (MX1) and on day 14 (IFI27). On day 3,
247 MX1 achieved an AUROC of 0.97 (0.93-1) which reduced to 0.8 (0.64-0.96) by day 14. In contrast, IFI27
248 achieved an AUROC of 0.73 (0.56-0.91) on day 3, increasing to 1 by day 14 (Figure 2B). These findings
249 reflected an early but transient increase in MX1 expression and a comparatively delayed but sustained
250 increase in IFI27 expression (Figure 2C). A number of other single gene biomarkers (IFI44L, IFIT3 and RSAD2)
251 were highly correlated to MX1 and distinct from IFI27 (Supplementary Figure 4B).

252 **Relationship of MX1 and IFI27 expression in blood to symptoms and SARS-CoV-2 viral load**

253 Most biomarker discovery and validation has focused on naturally acquired symptomatic viral infection. We
254 and others have shown that host response biomarkers are able to detect asymptomatic infection^{2,39}. Consistent
255 with this, we found no correlation between blood transcriptional scores and prospective quantitation of daily
256 symptom scores among individuals who developed replicative infection (Figure 3A). We found elevated
257 biomarker scores ($>Z2$ threshold) at time points in which participants who experienced replicative infection
258 were completely asymptomatic. This was more evident with MX1 measurements at early time points and with
259 IFI27 measurements at late time points.

260 In addition, we investigated the relationship between blood transcriptional signature scores and viral load
261 stratified by time from inoculation in samples from individuals who developed replicative infection. Examples
262 of elevated MX1 and IFI27 scores ($Z>2$) were evident at time points with negative virus PCR in contemporary
263 nose or throat swabs (Figure 3B). Elevated MX1 scores associated with negative virus PCR tests were more
264 evident at early time points, and elevated IFI27 scores associated with negative virus PCR tests were more
265 evident at late time points. Importantly, MX1 and IFI27 scores in the normal range ($<Z2$) were also evident at
266 time points with positive virus PCR tests in contemporary samples. False-negative biomarker results were
267 more evident for IFI27 at early time points, and for MX1 at late time points. To underscore the differential

268 temporal relationship of each biomarker with viral load, we examined longitudinal biomarker measurements
269 per participant who developed replicative infection, indexed by time from first PCR detection of virus (>4 Log₁₀
270 copies/mL) in nasal swabs, which we have recently reported to correlate best with viral emissions⁵⁷. The rise
271 in MX1 scores was generally co-incident with PCR detection of the virus, and in some individuals evident
272 before detection of virus by PCR. However, the MX1 response generally peaked before the peak in viral load,
273 suggesting that clearance of MX1 transcript enrichment was faster than clearance of the virus. In contrast,
274 IFI27 scores increased after detection of the virus and remained elevated after viral load started to fall (Figure
275 3C).

276 Both blood transcriptional biomarkers showed statistically significant correlation with viral load when including
277 PCR negative time points (Supplementary Figure 5A) consistent with the fact that they provided good
278 discrimination of groups of participants with and without replicative infection. However, when restricting the
279 analysis to time points with positive virus PCR tests, we found a significant correlation only to MX1, suggesting
280 this biomarker provided better prediction of viral load than IFI27 (Supplementary Figure 5B). Consistent with
281 this observation, we also found that MX1 provided a better biomarker of infectiousness than IFI27, by predicting
282 positive viral culture in contemporary samples. Among individuals who developed replicative infection, blood
283 MX1 transcript levels discriminated virus culture positivity in nose or throat samples with AUROC 0.85 (0.79-
284 0.92), significantly better than IFI27 which achieved AUROC of 0.66 (0.57-0.75). In this analysis false positive
285 MX1 levels were limited to early time points, consistent with the observation that the rise in MX1 levels can
286 precede PCR detection of the virus (Figure 4).

287 ***Differential regulation of MX1 and IFI27 expression in blood***

288 Both MX1 and IFI27 are widely recognised as interferon stimulated genes (ISG)^{58,59}. To explore this
289 relationship among participants in the replicative infection group, we compared MX1 and IFI27 levels with the
290 average expression of a multigene signature (“STAT1 regulated module”) that we had previously derived and
291 validated as a measure of type 1 IFN bioactivity⁶⁰. Both biomarkers showed a statistically significant correlation
292 with the STAT1 module, but the relationship with MX1 was stronger with near perfect correlation and
293 covariance, suggesting that IFI27 expression was subject to additional levels of transcriptional regulation
294 (Figure 5A-B). To obtain a deeper insight into the mechanisms of differential regulation of MX1 and IFI27, we
295 investigated their expression in our previously reported single cell RNA sequencing analysis of PBMC from a
296 subset of participants with replicative infection in the present SARS-CoV-2 challenge study⁶¹. We found a clear
297 increase of MX1 expression in all major PBMC subsets in pooled day 3 data, and subsequent reduction by
298 day 7. In contrast, increased expression of IFI27 was almost exclusively restricted to myeloid cells (monocytes
299 and conventional dendritic cells). Modest upregulation was evident at day 3, but then increased further at day
300 7 and day 10 before reducing again by day 14, although expression levels remained higher than baseline
301 through to day 28 (Figure 5C).

302 In published ATAC sequencing data²⁰, we tested the hypothesis that differential time and cellular distribution
303 of MX1 and IFI27 expression reflected differential epigenetic regulation (chromatin accessibility) of *MX1* and
304 *IFI27* loci in circulating immune cells. In datasets from unstimulated monocytes, CD4 T effector cells and B-
305 cells from healthy individuals, we found evidence that the *MX1* locus contained areas of open chromatin
306 (enrichment of sequencing peaks) close to the transcription start site and exon-1 (Supplementary Figure 6A-
307 B), which would enable rapid transcriptional upregulation of this gene across multiple cell types. In contrast,

308 the *IFI27* locus contained little evidence of open chromatin (Supplementary Figure 6A-B) in any of these cell
309 types, and therefore inaccessible for rapid transcriptional upregulation. To evaluate subsequent epigenetic
310 modifications following infection, we leveraged single cell ATACseq data from patients admitted to hospital
311 with COVID-19³⁰. Despite the sparsity in single cell data and relatively low coverage of the *IFI27* locus, in
312 samples from patients with acute COVID-19, we found a higher number of *IFI27* sequencing reads in
313 monocytes compared to all lymphocyte populations. This difference was less evident in data from convalescent
314 patients (Supplementary Figure 6C), and consistent with transient cell-type specific opening of the *IFI27* locus
315 in established infection, providing a mechanistic basis for the temporal delay and cellular restriction of *IFI27*
316 responses compared to *MX1*.

317 ***Generalisable differential utility of blood MX1 and IFI27 transcriptional biomarkers in acute respiratory***
318 ***virus infection.***

319 In order to investigate whether the differential host responses represented by *MX1* and *IFI27* were
320 generalisable to other acute respiratory viral infections, we investigated their expression profiles in collated
321 data from previously reported influenza, respiratory syncytial virus, and rhinovirus human challenges among
322 participants with evidence of infection following inoculation as per original study definitions¹⁷. In every case
323 *MX1* upregulation in whole blood transcriptional profiles preceded that of *IFI27* (Figure 6A). The data from
324 these experiments were limited to approximately 6 days post-challenge and did not allow us to fully compare
325 the temporal profiles of these biomarker measurements to the present SARS-CoV-2 challenge. Therefore, we
326 undertook transcriptional profiling of blood samples from another recent H3N2 influenza human challenge
327 model that included sampling beyond day 7⁶. This analysis also reproduced our findings in the SARS-CoV-2
328 challenge (Figure 6B).

329 We further sought to extend the generalisability of our findings to natural infections. In a household contact
330 study of index cases with COVID-19⁵, blood transcript levels of *MX1* and *IFI27* achieved equivalently good
331 discrimination of contacts with and without prevalent SARS-CoV-2 infection at recruitment (day 0, AUROC
332 0.97, 0.92-1). This level of discrimination was maintained for *IFI27* in follow up samples 7 days later, but
333 significantly reduced for *MX1*, consistent with earlier resolution of this biomarker (Figure 7A). In a further data
334 set from patients with unselected community acquired respiratory virus infections, we evaluated *MX1* and *IFI27*
335 expression in whole blood transcriptional profiles of individuals with PCR confirmed respiratory virus infections
336 within 48 hours of symptom onset, in four sequential samples on alternate days¹⁸. Compared to baseline (pre-
337 infection) samples from the same individuals, increased levels of *MX1* expression ($Z > 2$) were largely confined
338 to early time points day 0-2 after presentation within 4 days of symptom onset. Increased levels of *IFI27*
339 expression ($Z > 2$) were evident over a longer time course including day 4-6 after presentation, up to 8 days
340 after symptom onset (Figure 7B, Supplementary Figure 7A). Across all time points, *IFI27* measurements
341 achieved statistically better AUROC than *MX1* measurements for discrimination of infection from baseline
342 uninfected samples (Supplementary Figure 7B). However, when the analysis was stratified by sample time
343 point, *MX1* achieved the highest AUROC for discrimination of infected samples on the day of presentation
344 (Supplementary Figure 7C). The AUROC for *MX1* reduced significantly at each subsequent time point. The
345 time point stratified analysis of *IFI27*, showed stable AUROC discrimination of infection. A combined biomarker
346 signature, comprising the average expression of *MX1* and *IFI27* improved the AUROC discrimination at early
347 time points compared to *IFI27* alone, and at late time points compared to *MX1* alone (Supplementary Figure
348 7C).

349 ***Comparison of host response biomarkers of acute respiratory virus infection in blood and nose***
350 ***samples***

351 The potential to measure host response transcriptional signatures in samples from upper respiratory tract
352 swabs has recently been reported^{62,63}. We compared MX1 and IFI27 transcript measurements in samples from
353 blood and nose swabs in the present SARS-CoV-2 challenge. Surface nose swabs only yielded adequate RNA
354 for sequencing in 103 of 238 samples (43%), reflecting an inherent technical limitation in this approach.
355 Nonetheless, for nose samples which did yield RNA sequencing data, we found clear evidence of MX1 and
356 IFI27 responses in participants who developed a replicative infection. In comparison to blood measurements
357 of these biomarkers, the signal strength in nose swab samples was weaker than in blood, the response in the
358 nose was delayed in comparison to blood, and the differential time course for each biomarker evident in blood
359 samples was lost in nose swab samples (Figure 7C). These findings were replicated in blood and nasal
360 mucosal curettage samples from the H3N2 influenza human challenge and indicate that in general, blood
361 biomarker measurements are likely to provide better diagnostic discrimination for prevalent infection as well
362 as better differentiation of early and late phases of infection, compared to nasal swabs (Figure 7C).

363 **Discussion**

364 We present a comprehensive evaluation of previously reported transcriptional signatures as host response
365 biomarkers of viral infection in high frequency longitudinal blood and nose swab samples from the first SARS-
366 CoV-2 human challenge experiment. We provide compelling evidence showing that single gene transcripts for
367 MX1 and IFI27 in blood, discriminate temporally distinct phases of infection, and we show that these findings
368 are generalisable across a range of clinically important respiratory viruses in both experimental and naturally
369 acquired infections. The earliest phase of replicative SARS-CoV-2 infection was associated with rapid
370 upregulation of MX1 transcripts in blood, which may precede PCR detection of the virus and correlated with
371 PCR positive viral load measurements. In contrast, blood transcriptional upregulation of IFI27 occurred after
372 PCR detection of the virus. IFI27 expression did not correlate with PCR positive viral load measurements and
373 was sustained above baseline levels after viral clearance. Of note, transcriptional upregulation of both
374 biomarkers was independent of symptoms.

375 Both MX1 and IFI27 are widely recognised as ISGs^{58,59}. The MX1 response closely reflected generalised type
376 1 ISG expression across all major cell types. We focused on MX1 because it achieved the highest single gene
377 point estimate AUROC for discriminating groups of individuals with and without replicative viral infection at the
378 first time point at which any biomarker achieved significant discrimination. Alternative interferon inducible single
379 gene biomarkers such as IFI44L, IFIT3 and RSAD2 provided statistically comparable discrimination at this
380 time point, and are highly correlated to MX1. These biomarkers are likely to share the same mechanisms for
381 transcriptional regulation, and offer the same utility as MX1. Delayed transcriptional upregulation of IFI27
382 compared to other canonical ISGs has also been reported following in vitro stimulation of cells with IFN^{64,65}. In
383 vivo, IFI27 expression in blood samples was restricted to myeloid PBMC. We found evidence of differential
384 epigenetic silencing of the IFI27 locus compared to the MX1 locus in resting PBMC, and cell type specific
385 epigenetic modulation of this locus in monocytes during established COVID-19 infection. These data provide
386 a mechanistic explanation for the differential temporal and cellular expression of the two biomarkers, namely
387 that IFI27 is epigenetically silenced in resting cells but becomes accessible for transcription in specific myeloid
388 lineages during the evolving immune response to infection.

389 The differential temporal expression of MX1 and IFI27 in the SARS-CoV-2 challenge model was replicated in
390 challenge experiments with multiple influenza strains, respiratory syncytial virus, or rhinovirus, and in data from
391 household contacts with naturally acquired SARS-CoV-2 infection. Likewise, in unselected community
392 acquired symptomatic respiratory virus infections, in which MX1 measurements achieved high diagnostic
393 accuracy for infections within 4 days of symptom onset, and IFI27 measurements achieved higher diagnostic
394 accuracy at later time points. The combination of both measurements provided highest diagnostic accuracy
395 across all time points.

396 In SARS-CoV-2 and H3N2 Influenza challenge experiments, we found no evidence that measuring these
397 biomarkers in nose swab samples offered any advantage to blood samples. Lower upregulation of gene
398 expression in nose samples compared to blood, and the loss of temporal differentiation between the
399 biomarkers may reflect the biology of superficial cell populations compared to circulating leukocytes.
400 Unexpectedly, upregulation of these biomarkers in nose samples of individuals with replicative infection was
401 also temporally delayed compared to the blood. This finding is also evident in our analysis of comparative
402 single cell sequencing data from blood and nose swab samples⁶¹. Whether it reflects faster transmission of
403 IFN signalling to circulating blood cells, later onset of viral replication in the nose compared to the throat, or
404 local suppression of IFN signalling by the virus in the nasal mucosa require future mechanistic investigation.

405 We propose that these blood transcriptional biomarkers of early and late phase of viral infection offer a range
406 of research and clinical applications. Upregulation of MX1 expression may be used to detect pre-symptomatic
407 infection in contacts of index cases. Its correlation with viral load and culture results suggests it may also
408 provide clinical utility to infer infectiousness and thus trigger infection control interventions. Finally, since
409 specific antiviral treatment efficacy diminishes with duration of infection^{66,67}, MX1 measurements may be used
410 to stratify patients most likely to benefit from treatment. Conversely, IFI27 expression may offer a more time-
411 stable diagnostic triage test for viral infection due to sustained upregulation beyond the initial acute phase of
412 infection. To account for differential temporal profiles, a combined approach including both MX1 and IFI27 (as
413 averaged expression, or where a positive test for either gene triggers further confirmatory testing) may be the
414 optimal approach to diagnostic triage.

415 Comprehensive identification of reported blood transcriptional biomarkers of viral infection by systematic
416 review, and their application in standardised human SARS-CoV-2 and influenza human challenges with high
417 frequency sampling are major strengths of this study, thus enabling identification of differential temporal
418 profiles of MX1 and IFI27 responses. Single cell data from the SARS-CoV-2 challenge model, and analyses
419 of publicly available data also allowed investigation of the mechanism for differential temporal profiles of MX1
420 and IFI27 responses, and to confirm reproducibility of our findings across a range of respiratory virus infections.

421 Our conclusions are currently limited to data derived from individuals with non-severe infection. Therefore,
422 future validation in hospitalised cohorts for whom the time of exposure can be estimated will be required to
423 assess whether severe disease alters the temporal profiles of these biomarkers. Finally, we do not address
424 the specificity of our findings for respiratory virus infections. Therefore, we have limited our discussion of
425 potential translational applications to diagnostic triage tests to trigger confirmatory virological investigations,
426 stratification of patients with confirmed viral infections for antiviral treatment, and pre-symptomatic screening
427 of contacts of index cases of confirmed viral infections. Notably, for each of these applications, the
428 generalisability of blood transcriptional biomarkers across respiratory viruses may be considered a strength.

429 Translation of these viral biomarkers to near-patient platforms is now required to enable further evaluation of
430 clinical utility and impact in prospective observational and interventional studies.

431 **Contributions**

432 JR, RKG and MN conceived this study. RH, CC and MN obtained funding for this study. MK, AM, AC, BK, WB
433 and CC conceived, obtained funding and supervised conduct of the SARS-CoV-2 challenge study. CT, LKB,
434 RKG, JGB and MN undertook the systematic review of blood transcriptional biomarkers of viral infection. JR,
435 RKG, LKB and MN undertook analysis of publicly available bulk RNAseq and ATACseq data. JR, CT, TM,
436 LKB, JGB, HW, BL, CW, CV, BMC and RH undertook sample processing and data analysis for the SARS-
437 CoV-2 challenge study. RL, LD, MZN and ST contributed single cell RNAseq data from the SARS-CoV-2
438 challenge study. CMB, LP, PD, MK, ML and CC contributed bulk data from the H3N2 influenza challenge
439 study. KM, EC, JF, SH and AL contributed data from the INSTINCT SARS-CoV-2 household contact study.
440 AJK and JCK contributed data single cell ATACseq data from the COMBAT study. JR, RKG and MN wrote the
441 manuscript with input from all the authors.

442 **Declaration of interests**

443 S.A.T. has received remuneration for Scientific Advisory Board Membership from Sanofi, GlaxoSmithKline,
444 Foresite Labs and Qiagen. S.A.T. is a co-founder and holds equity in Transition Bio. A.M., A.C., M.K., M.M.
445 and A.B. are full time employees at hVIVO Services Ltd.

446 **Funding acknowledgements**

447 This research was supported by the Wellcome Trust (224530/Z/21/Z). AL acknowledges funding by the NIHR
448 Health Protection Research Units (HPRU) in Respiratory Infections (NIHR200927). BMC acknowledges
449 funding by the Rosetrees Foundation. CMB, MK and ML acknowledge funding by NIHR Biomedical Research
450 to Imperial College London. CMW acknowledges funding from the Medical Research Council (MR/T016329/1).
451 CT acknowledges funding from the Wellcome Trust (102186/B/13/Z). LCKB acknowledges funding from the
452 NIHR (Academic Clinical Fellowship Programme). LMD acknowledges funding from the European Union's
453 Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No
454 955321. M.Z.N. acknowledges funding from a MRC Clinician Scientist Fellowship (MR/W00111X/1), Action
455 Medical Research (GN2911) and funding from the Rutherford Fund Fellowship allocated by the MRC UK
456 Regenerative Medicine Platform 2 (MR/5005579/1). MN acknowledges funding from the Wellcome Trust
457 (207511/Z/17/Z) and by NIHR Biomedical Research Funding to UCL and UCLH. R.H. is a NIHR Senior
458 Investigator. RKG acknowledges funding from the National Institute for Health Research (NIHR302829).

459 **Data sharing statement**

460 Processed RNAseq data are available from ArrayExpress from the SARS-CoV-2 challenge study (accession
461 number: E-MTAB-12993), H3N2 influenza challenge study (accession numbers: E-MTAB-13038 and E-MTAB-
462 13041). Raw sequencing data from these studies and the INSTINCT SARS-CoV-2 household contact study
463 will be made available via the European Genome-Phenome Archive under managed data access.

464 **References**

465 1 Gupta RK, Noursadeghi M. Toward a more generalizable blood RNA signature for bacterial and viral
466 infections. *Cell Rep Med* 2022; **3**: 100866.

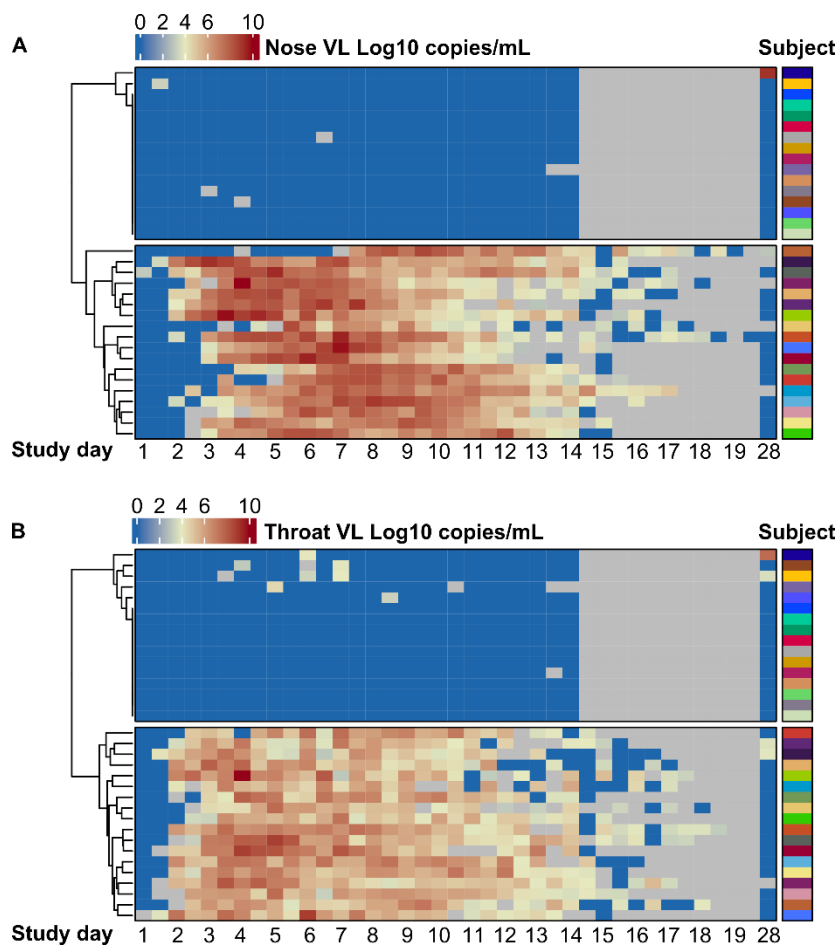
- 467 2 Gupta RK, Rosenheim J, Bell LC, Chandran A, Guerra-Assuncao JA, Pollara G *et al.* Blood transcriptional
468 biomarkers of acute viral infection for detection of pre-symptomatic SARS-CoV-2 infection: a nested, case-
469 control diagnostic accuracy study. *Lancet Microbe* 2021; **2**: e508–e517.
- 470 3 Swadling L, Diniz MO, Schmidt NM, Amin OE, Chandran A, Shaw E *et al.* Pre-existing polymerase-specific
471 T cells expand in abortive seronegative SARS-CoV-2. *Nature* 2022; **601**: 110–117.
- 472 4 Killingley B, Mann AJ, Kalinova M, Boyers A, Goonawardane N, Zhou J *et al.* Safety, tolerability and viral
473 kinetics during SARS-CoV-2 human challenge in young adults. *Nat Med* 2022; **28**: 1031–1041.
- 474 5 Derqui N, Koycheva A, Zhou J, Pillay TD, Crone MA, Hakki S *et al.* Risk factors and vectors for SARS-
475 CoV-2 household transmission: a prospective, longitudinal cohort study. *The Lancet Microbe* 2023; **0**.
476 doi:10.1016/S2666-5247(23)00069-1.
- 477 6 Temple DS, Hegarty-Craver M, Furberg RD, Preble EA, Bergstrom E, Gardener Z *et al.* Wearable Sensor-
478 Based Detection of Influenza in Presymptomatic and Asymptomatic Individuals. *The Journal of Infectious*
479 *Diseases* 2023; **227**: 864–872.
- 480 7 Dhariwal J, Cameron A, Trujillo-Torralbo M-B, del Rosario A, Bakhsoliani E, Paulsen M *et al.* Mucosal
481 Type 2 Innate Lymphoid Cells Are a Key Component of the Allergic Response to Aeroallergens. *Am J*
482 *Respir Crit Care Med* 2017; **195**: 1586–1596.
- 483 8 Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat*
484 *Biotechnol* 2016; **34**: 525–527.
- 485 9 Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve
486 gene-level inferences. *F1000Res* 2015; **4**: 1521.
- 487 10 Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A *et al.* BioMart and Bioconductor: a
488 powerful link between biological databases and microarray data analysis. *Bioinformatics* 2005; **21**: 3439–
489 3440.
- 490 11 Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biology* 2010; **11**:
491 R106.
- 492 12 Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets
493 with the R/Bioconductor package biomaRt. *Nat Protoc* 2009; **4**: 1184–1191.
- 494 13 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with
495 DESeq2. *Genome Biology* 2014; **15**: 550.
- 496 14 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S *et al.* STAR: ultrafast universal RNA-seq
497 aligner. *Bioinformatics* 2013; **29**: 15–21.
- 498 15 Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment
499 and quantification of RNA sequencing reads. *Nucleic Acids Research* 2019; **47**: e47.
- 500 16 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*
501 2014; **30**: 2114–2120.
- 502 17 Liu T-Y, Burke T, Park LP, Woods CW, Zaas AK, Ginsburg GS *et al.* An individualized predictor of health
503 and disease using paired reference and target samples. *BMC Bioinformatics* 2016; **17**: 47.
- 504 18 Zhai Y, Franco LM, Atmar RL, Quarles JM, Arden N, Bucacas KL *et al.* Host Transcriptional Response to
505 Influenza and Other Acute Respiratory Viral Infections--A Prospective Cohort Study. *PLoS Pathog* 2015;
506 **11**: e1004869.
- 507 19 Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C *et al.* pROC: an open-source package for
508 R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; **12**: 77.
- 509 20 Calderon D, Nguyen MLT, Mezger A, Kathiria A, Müller F, Nguyen V *et al.* Landscape of stimulation-
510 responsive chromatin across diverse human immune cells. *Nat Genet* 2019; **51**: 1494–1505.

- 511 21 Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables
512 reproducible computational workflows. *Nat Biotechnol* 2017; **35**: 316–319.
- 513 22 Patel H, Espinosa-Carrasco J, Ewels P, bot nf-core, Langer B, Syme R *et al.* nf-core/atacseq: nf-
514 core/atacseq v2.0 - Iron Iguana. 2022. doi:10.5281/zenodo.7384115.
- 515 23 Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*
516 2009; **25**: 1754–1760.
- 517 24 Picard Tools - By Broad Institute. <https://broadinstitute.github.io/picard/> (accessed 11 May2023).
- 518 25 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO *et al.* Twelve years of SAMtools and
519 BCFtools. *GigaScience* 2021; **10**: giab008.
- 520 26 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*
521 2010; **26**: 841–842.
- 522 27 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G *et al.* Integrative genomics
523 viewer. *Nat Biotechnol* 2011; **29**: 24–26.
- 524 28 Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE *et al.* Model-based Analysis of CHIP-
525 Seq (MACS). *Genome Biology* 2008; **9**: R137.
- 526 29 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P *et al.* Simple combinations of lineage-
527 determining transcription factors prime cis-regulatory elements required for macrophage and B cell
528 identities. *Mol Cell* 2010; **38**: 576–589.
- 529 30 Ahern DJ, Ai Z, Ainsworth M, Allan C, Allcock A, Angus B *et al.* A blood atlas of COVID-19 defines
530 hallmarks of disease severity and specificity. *Cell* 2022; **185**: 916-938.e58.
- 531 31 Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY *et al.* ArchR is a scalable software
532 package for integrative single-cell chromatin accessibility analysis. *Nat Genet* 2021; **53**: 403–411.
- 533 32 Andres-Terre M, McGuire HM, Pouliot Y, Bongen E, Sweeney TE, Tato CM *et al.* Integrated, Multi-cohort
534 Analysis Identifies Conserved Transcriptional Signatures across Multiple Respiratory Viruses. *Immunity*
535 2015; **43**: 1199–1211.
- 536 33 Cappuccio A, Chawla DG, Chen X, Rubenstein AB, Cheng WS, Mao W *et al.* Multi-objective optimization
537 identifies a specific and interpretable COVID-19 host response signature. *Cell Syst* 2022; **13**: 989-1001.e8.
- 538 34 Gómez-Carballa A, Barral-Arca R, Cebey-López M, Bello X, Pardo-Seco J, Martín-Torres F *et al.*
539 Identification of a Minimal 3-Transcript Signature to Differentiate Viral from Bacterial Infection from Best
540 Genome-Wide Host RNA Biomarkers: A Multi-Cohort Analysis. *Int J Mol Sci* 2021; **22**: 3148.
- 541 35 Henrickson SE, Manne S, Dolfi DV, Mansfield KD, Parkhouse K, Mistry RD *et al.* Genomic Circuitry
542 Underlying Immunological Response to Pediatric Acute Respiratory Infection. *Cell Reports* 2018; **22**: 411–
543 426.
- 544 36 Herberg JA, Kaforou M, Wright VJ, Shailes H, Eleftherohorinou H, Hoggart CJ *et al.* Diagnostic Test
545 Accuracy of a 2-Transcript Host RNA Signature for Discriminating Bacterial vs Viral Infection in Febrile
546 Children. *JAMA* 2016; **316**: 835–845.
- 547 37 Tang BM, Shojaei M, Parnell GP, Huang S, Nalos M, Teoh S *et al.* A novel immune biomarker IFI27
548 discriminates between influenza and bacteria in patients with suspected respiratory infection. *European*
549 *Respiratory Journal* 2017; **49**. doi:10.1183/13993003.02098-2016.
- 550 38 Gómez-Carballa A, Cebey-López M, Pardo-Seco J, Barral-Arca R, Rivero-Calle I, Pischedda S *et al.* A
551 qPCR expression assay of IFI44L gene differentiates viral from bacterial infections in febrile children. *Sci*
552 *Rep* 2019; **9**: 11780.

- 553 39 McClain MT, Constantine FJ, Nicholson BP, Nichols M, Burke TW, Henao R *et al.* A blood-based host
554 gene expression assay for early detection of respiratory viral infection: an index-cluster prospective cohort
555 study. *Lancet Infect Dis* 2021; **21**: 396–404.
- 556 40 Li HK, Kaforou M, Rodriguez-Manzano J, Channon-Wells S, Moniri A, Habgood-Coote D *et al.* Discovery
557 and validation of a three-gene signature to distinguish COVID-19 and other viral infections in emergency
558 infectious disease presentations: a case-control and observational cohort study. *Lancet Microbe* 2021; **2**:
559 e594–e603.
- 560 41 Lopez R, Wang R, Seelig G. A molecular multi-gene classifier for disease diagnostics. *Nature Chem* 2018;
561 **10**: 746–754.
- 562 42 Lydon EC, Henao R, Burke TW, Aydin M, Nicholson BP, Glickman SW *et al.* Validation of a host response
563 test to distinguish bacterial and viral respiratory infection. *EBioMedicine* 2019; **48**: 453–461.
- 564 43 Pennisi I, Rodriguez-Manzano J, Moniri A, Kaforou M, Herberg JA, Levin M *et al.* Translation of a host
565 blood RNA Signature distinguishing bacterial from viral infection into a platform suitable for development
566 as a point-of-care test. *JAMA Pediatrics* 2021. doi:doi:10.1001/jamapediatrics.2020.5227.
- 567 44 Roers A, Hochkeppel HK, Horisberger MA, Hovanessian A, Haller O. MxA Gene Expression after Live
568 Virus Vaccination: A Sensitive Marker for Endogenous Type I Interferon. *J Infect Dis* 1994; **169**: 807–813.
- 569 45 Rao AM. A robust host-response-based signature distinguishes bacterial and viral infections across
570 diverse global populations. *Cell Reports Medicine* 2022.
- 571 46 Ravichandran S, Banerjee U, Dr GD, Kandukuru R, Thakur C, Chakravorty D *et al.* VB10, a new blood
572 biomarker for differential diagnosis and recovery monitoring of acute viral and bacterial infections.
573 *eBioMedicine* 2021; **67**. doi:10.1016/j.ebiom.2021.103352.
- 574 47 Samy A, Maher MA, Abdelsalam NA, Badr E. SARS-CoV-2 potential drugs, drug targets, and biomarkers:
575 a viral-host interaction network-based analysis. *Sci Rep* 2022; **12**: 11934.
- 576 48 Sampson DL, Fox BA, Yager TD, Bhide S, Cermelli S, McHugh LC *et al.* A Four-Biomarker Blood
577 Signature Discriminates Systemic Inflammation Due to Viral Infection Versus Other Etiologies. *Sci Rep*
578 2017; **7**: 2914.
- 579 49 Sampson D, Yager TD, Fox B, Shallcross L, McHugh L, Seldon T *et al.* Blood transcriptomic discrimination
580 of bacterial and viral infections in the emergency department: a multi-cohort observational validation study.
581 *BMC Med* 2020; **18**: 185.
- 582 50 Steinbrink JM, Myers RA, Hua K, Johnson MD, Seidelman JL, Tsalik EL *et al.* The host transcriptional
583 response to Candidemia is dominated by neutrophil activation and heme biosynthesis and supports novel
584 diagnostic approaches. *Genome Med* 2021; **13**: 108.
- 585 51 Sweeney TE, Wong HR, Khatri P. Robust classification of bacterial and viral infections via integrated host
586 gene expression diagnostics. *Sci Transl Med* 2016; **8**: 346ra91.
- 587 52 Trouillet-Assant S, Viel S, Ouziel A, Boisselier L, Rebaud P, Basmaci R *et al.* Type I Interferon in Children
588 with Viral or Bacterial Infections. *Clin Chem* 2020; **66**: 802–808.
- 589 53 Tsalik EL, Henao R, Nichols M, Burke T, Ko ER, McClain MT *et al.* Host gene expression classifiers
590 diagnose acute respiratory illness etiology. *Science Translational Medicine* 2016; **8**: 322ra11-322ra11.
- 591 54 Xu N, Hao F, Dong X, Yao Y, Guan Y, Yang L *et al.* A two-transcript biomarker of host classifier genes for
592 discrimination of bacterial from viral infection in acute febrile illness: a multicentre discovery and validation
593 study. *The Lancet Digital Health* 2021; **3**: e507–e516.
- 594 55 Yu J, Peterson DR, Baran AM, Bhattacharya S, Wylie TN, Falsey AR *et al.* Host Gene Expression in Nose
595 and Blood for the Diagnosis of Viral Respiratory Infection. *J Infect Dis* 2019; **219**: 1151–1161.

- 596 56 Zaas AK, Burke T, Chen M, McClain M, Nicholson B, Veldman T *et al.* A Host-Based RT-PCR Gene
597 Expression Signature to Identify Acute Respiratory Viral Infection. *Science Translational Medicine* 2013;
598 **5**: 203ra126-203ra126.
- 599 57 Zhou J, Singanayagam A, Goonawardane N, Moshe M, Sweeney F, Sukhova K *et al.* Viral Emissions into
600 the Air and Environment after SARS-CoV-2 Human Challenge: A Phase 1, Open Label, First-in-Human
601 Study. 2022. doi:10.2139/ssrn.4301808.
- 602 58 Haller O, Kochs G. Mx genes: host determinants controlling influenza virus infection and trans-species
603 transmission. *Hum Genet* 2020; **139**: 695–705.
- 604 59 Parker N, Porter AC. Identification of a novel gene family that includes the interferon-inducible human
605 genes 6–16 and ISG12. *BMC Genomics* 2004; **5**: 8.
- 606 60 Chandran A, Rosenheim J, Nageswaran G, Swadling L, Pollara G, Gupta RK *et al.* Rapid synchronous
607 type 1 IFN and virus-specific T cell responses characterize first wave non-severe SARS-CoV-2 infections.
608 *Cell Rep Med* 2022; **3**: 100557.
- 609 61 Rik G H Lindeboom, Kaylee B Worlock, Lisa M Dratva, Masahiro Yoshida, David Scobie, Helen R
610 Wagstaffe *et al.* Human SARS-CoV-2 challenge resolves local and systemic response dynamics. *medRxiv*
611 2023; : 2023.04.13.23288227.
- 612 62 Shojaei M, Shamshirian A, Monkman J, Grice L, Tran M, Tan CW *et al.* IFI27 transcription is an early
613 predictor for COVID-19 outcomes, a multi-cohort observational study. *Frontiers in Immunology* 2023;
614 **13**.<https://www.frontiersin.org/articles/10.3389/fimmu.2022.1060438> (accessed 13 Apr2023).
- 615 63 Cheemarla NR, Hanron A, Fauver JR, Bishai J, Watkins TA, Brito AF *et al.* Nasal host response-based
616 screening for undiagnosed respiratory viruses: a pathogen surveillance and detection study. *The Lancet*
617 *Microbe* 2023; **4**: e38–e46.
- 618 64 Leaman DW, Chawla-Sarkar M, Jacobs B, Vyas K, Sun Y, Ozdemir A *et al.* Novel Growth and Death
619 Related Interferon-Stimulated Genes (ISGs) in Melanoma: Greater Potency of IFN- β Compared with IFN-
620 α 2. *Journal of Interferon & Cytokine Research* 2003; **23**: 745–756.
- 621 65 Rosebeck S, Leaman DW. Mitochondrial localization and pro-apoptotic effects of the interferon-inducible
622 protein ISG12a. *Apoptosis* 2008; **13**: 562–572.
- 623 66 Muthuri SG, Venkatesan S, Myles PR, Leonardi-Bee J, Khuwaitir TSAA, Mamun AA *et al.* Effectiveness
624 of neuraminidase inhibitors in reducing mortality in patients admitted to hospital with influenza A
625 H1N1pdm09 virus infection: a meta-analysis of individual participant data. *The Lancet Respiratory*
626 *Medicine* 2014; **2**: 395–404.
- 627 67 Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC *et al.* Remdesivir for the Treatment
628 of Covid-19 — Final Report. *New England Journal of Medicine* 2020; **383**: 1813–1826.

629 **Figure 1.**

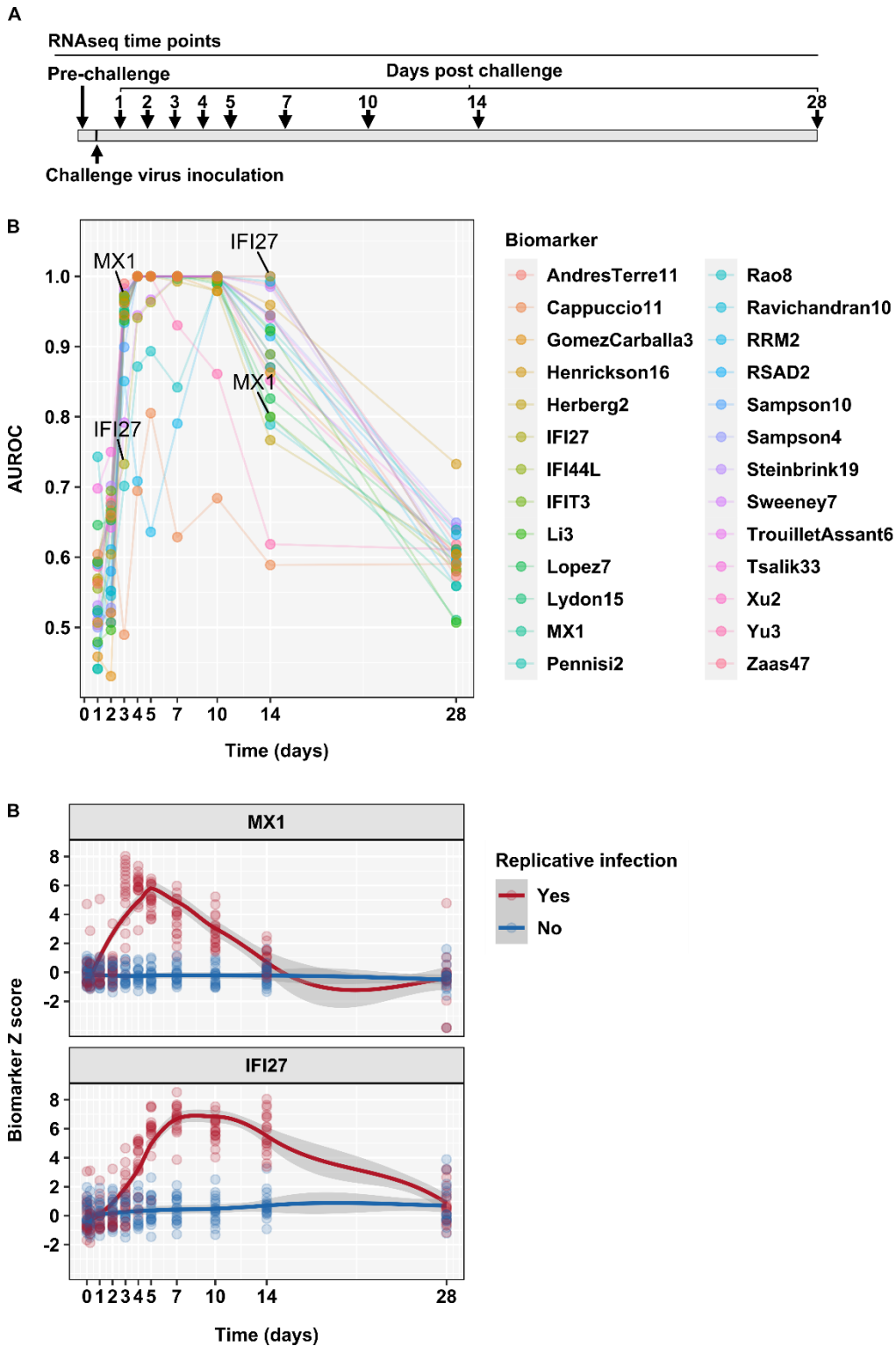


630

631 ***SARS-CoV-2 PCR viral load in nose and throat swabs following virus challenge.***

632 Quantitative viral load measurements by PCR from **(A)** nose and **(B)** throat swabs per participant (rows)
633 stratified by time point (columns) after virus challenge, and clustered into two groups of participants with (N=18)
634 and without (N=16) evidence of replicative virus replication. Grey colour denotes unavailable data points.

635 **Figure 2.**



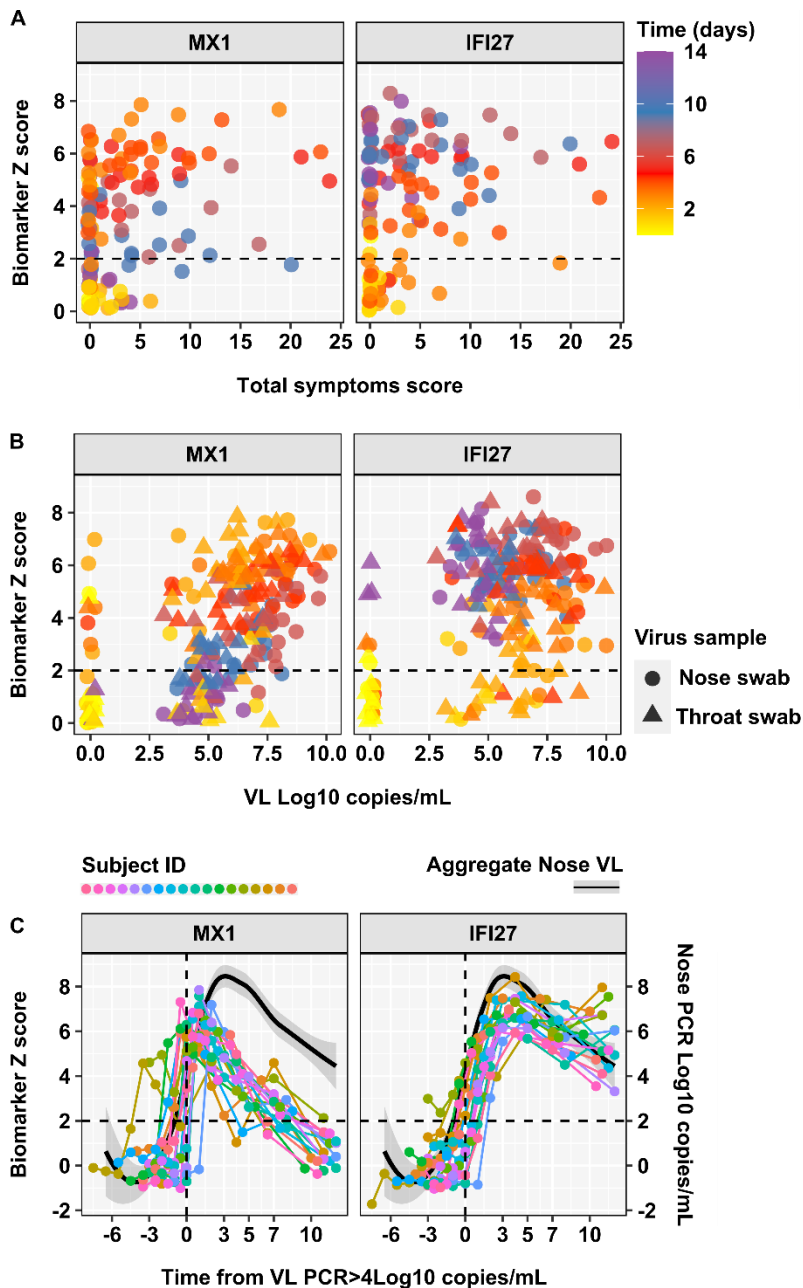
636

637 **Blood transcriptional discrimination of participants with and without replicative infection by time from**
 638 **SARS-CoV-2 challenge.**

639 **(A)** Time points for blood RNA sampling in relation to virus challenge. **(B)** Point estimates for area under the
 640 receiver operating characteristic curve (AUROC) stratified by blood transcriptional signature and time after
 641 virus challenge. **(C)** Individual (data points) and less smoothed summary (line $\pm 95\%$ CI) for standardised

642 blood transcript levels of MX1 and IFI27 in sequential time points after challenge, for participants with (N=18)
643 and without (N=16) replicative viral infection.

644 **Figure 3.**

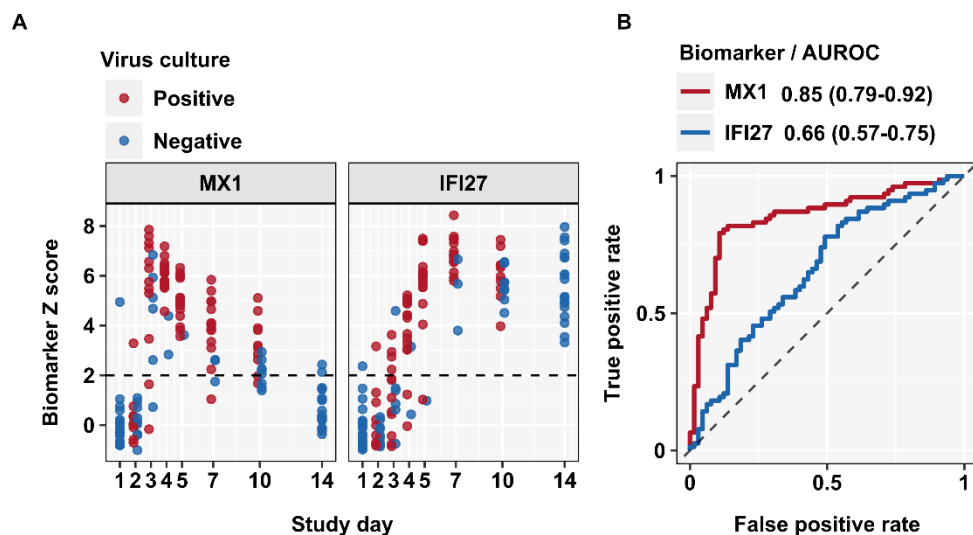


645

646 **Relationship between blood transcript levels of MX1 and IFI27 with symptoms and viral load by time**
647 **from SARS-CoV-2 challenge.**

648 Individual standardised blood transcript levels of MX1 and IFI27 against (A) symptom scores and (B) nose and
649 throat viral loads in participants who developed replicative virus infection (N=18), stratified by time after virus
650 challenge, with dashed line to represent the threshold (Z>2) for elevated transcript levels. (C) Individual
651 standardised blood transcript levels of MX1 and IFI27 (connected data points, left axis) and loess smoothed
652 summary for nose viral load (black line ±95% CI, right axis) in participants who developed replicative virus
653 infection (N=18), by time from virus detection in nose swabs >4 Log₁₀ copies/mL.

654 **Figure 4.**

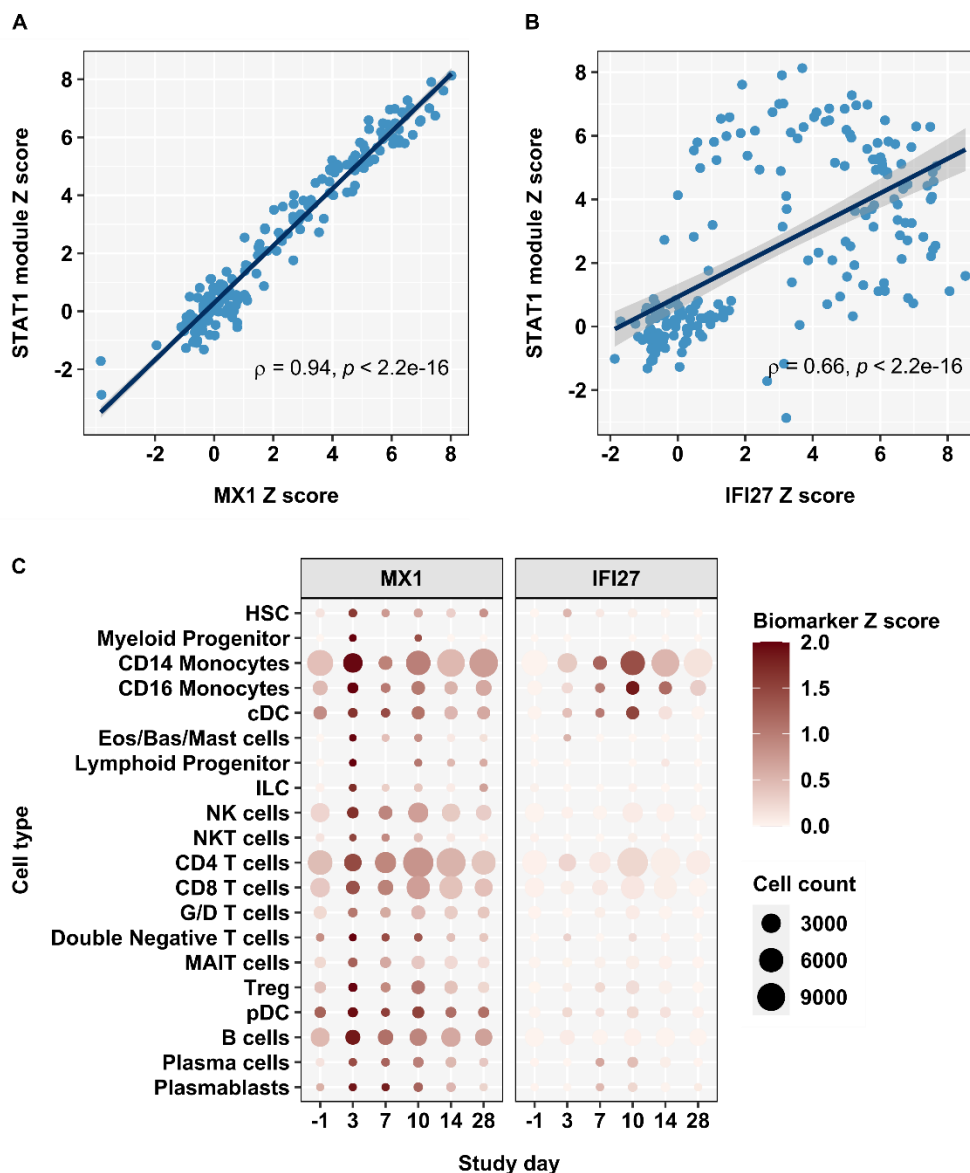


655

656 ***Discrimination of virus culture positivity by blood transcript levels of MX1 and IFI27.***

657 **(A)** Individual standardised blood transcript levels of MX1 and IFI27 at each time point for all individuals who
658 develop replicative infection (N=18), stratified by contemporary virus culture positivity in either nose or throat
659 swabs, with dashed line to represent the threshold ($Z > 2$) for elevated transcript levels. **(B)** Area under the
660 receiver operating characteristic curve (AUROC) discrimination of virus culture positivity by blood transcript
661 levels of MX1 and IFI27 across all time points, showing AUROC point estimates and 95% confidence intervals.

662 **Figure 5.**

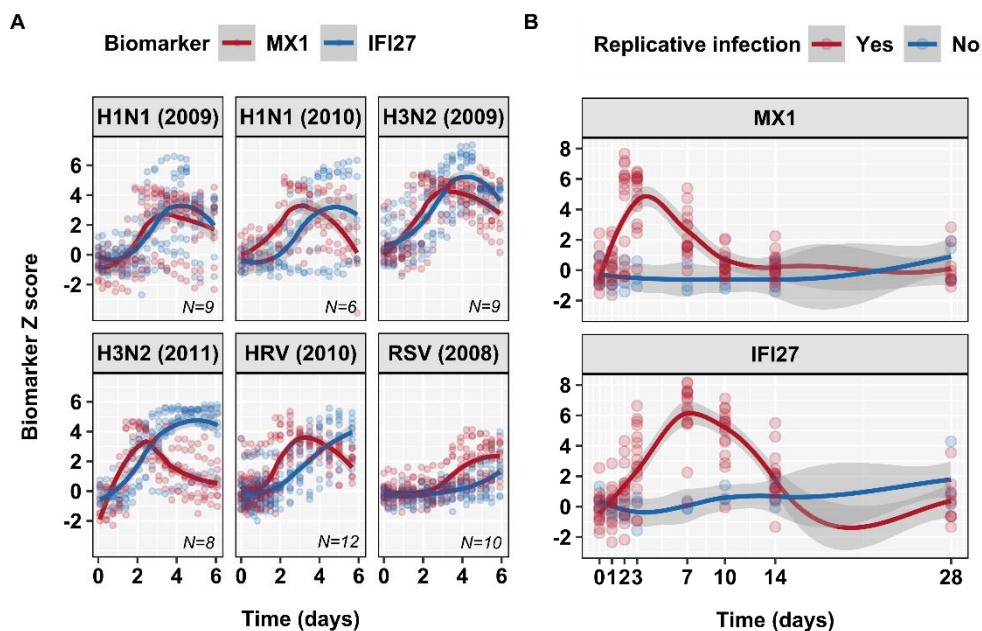


663

664 **Differential regulation of MX1 and IFI27 expression in blood.**

665 **(A)** Individual standardised blood transcriptional expression of a type 1 IFN stimulated gene signature (STAT1
666 module) by standardised blood transcriptional expression of MX1 and IFI27 at all time points in participants
667 who developed replicative virus infection (N=18), showing Spearman correlation coefficients and p value. **(B)**
668 Standardised blood transcriptional expression of MX1 and IFI27 stratified by cell type and time after virus
669 challenge, in pooled data from participants who developed replicative virus infection (N=6).

670 **Figure 6.**

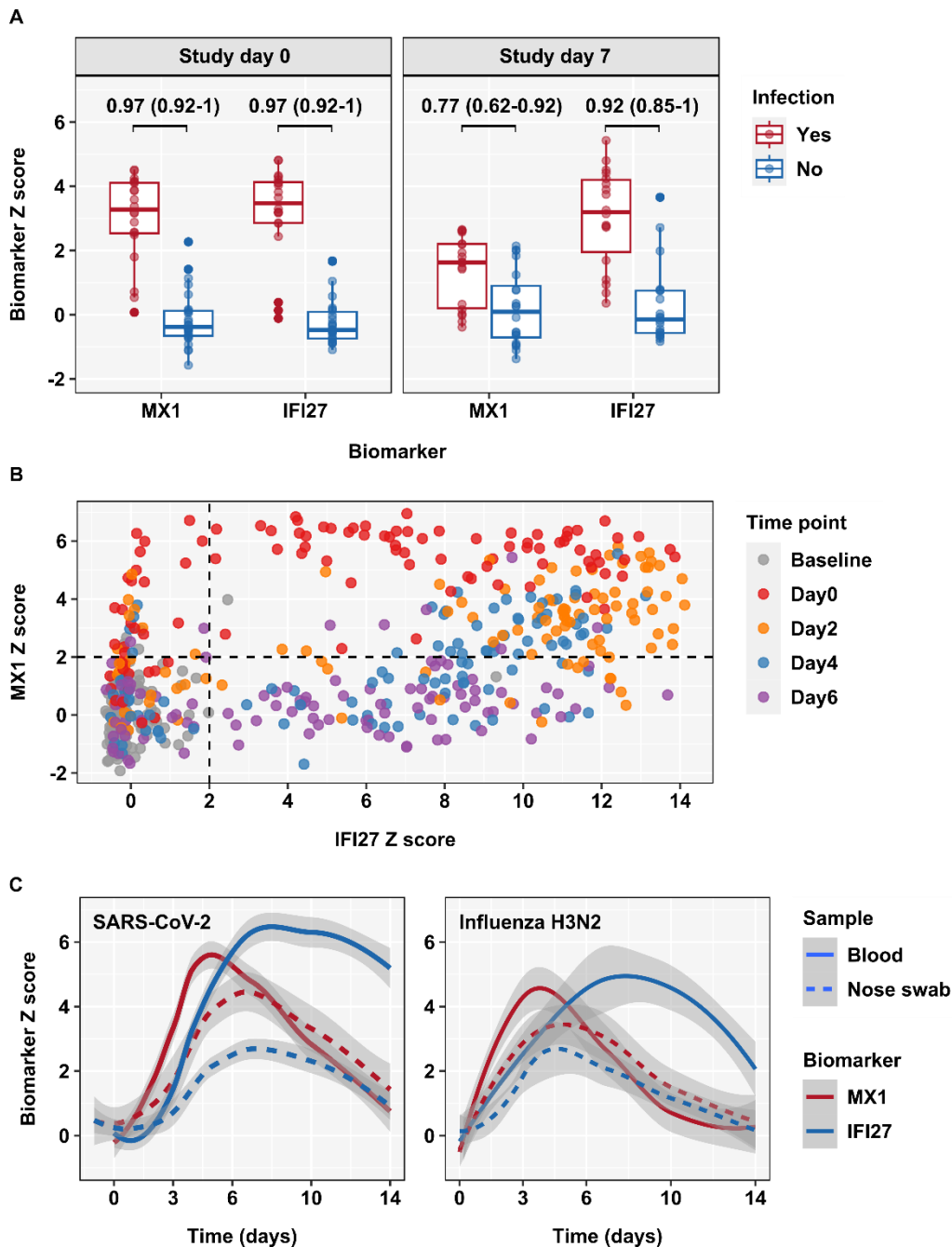


671

672 **Generalisable differences in temporal profiles of blood MX1 and IFI27 expression in diverse respiratory**
673 **virus challenges.**

674 **(A)** Individual (data points) and loess smoothed summary (line $\pm 95\%$ CI) for standardised blood transcript
675 levels of MX1 and IFI27 over the first 6 days after challenge in selected human respiratory virus challenge
676 models (GSE73072) among participants who develop replicative infection, and **(B)** over 14 days in an H3N2
677 influenza human challenge model, among participants with (N=16) and without (N=3) replicative infection.

678 **Figure 7.**



679

680 **Differences in temporal profiles of blood MX1 and IFI27 expression in naturally acquired respiratory**
 681 **virus infections, and delayed responses in the nose to virus challenge.**

682 **(A)** Discrimination between SARS-CoV-2 infected (N=20) and uninfected (N=26) household contacts of index
 683 cases with COVID-19 by blood transcript levels of MX1 and IFI27, at participant recruitment (study day 0) and
 684 7 days later. Data points represent individual study participants, summarised by box and whisker plots.
 685 Discrimination accuracy is shown as AUROC point estimate and 95% confidence intervals. **(B)** Individual
 686 standardised blood transcript levels of MX1 against IFI27 for sequential samples before infection (baseline,
 687 N=128) and at time points indicated (day0, N=103; day 2, N=106, day 4, N=100; day 6, N=102) after
 688 presentation within 48 hours of symptoms onset among prospectively recruited participants with unselected
 689 respiratory virus infections. **(C)** Loess smoothed summary (line \pm 95% CI) for standardised transcript levels of
 690 MX1 and IFI27 in blood (N=18) and nose samples (N=5-13) from participants who developed replicative virus

691 infection by time after SARS-CoV-2 challenge, and in blood (N=16) and nose samples (N=12-13) from
692 participants who developed replicative virus infection by time after H3N2 influenza challenge.