

# 1 **Shotgun Transcriptome and Isothermal Profiling of SARS-CoV-2 Infection Reveals Unique** 2 **Host Responses, Viral Diversification, and Drug Interactions**

## 3 4 **Authors:**

5 Daniel J. Butler<sup>1#</sup>, Christopher Mozsary<sup>1#</sup>, Cem Meydan<sup>1,2,3#</sup>, David Danko<sup>1,4#</sup>, Jonathan Foox<sup>1,2#</sup>,  
6 Joel Rosiene<sup>5,6</sup>, Alon Shaiber<sup>5,6,7</sup>, Ebrahim Afshinnekoo<sup>1,2,3</sup>, Matthew MacKay<sup>1</sup>, Fritz J.  
7 Sedlazeck<sup>8</sup>, Nikolay A. Ivanov<sup>1,2,9</sup>, Maria Sierra<sup>1</sup>, Diana Pohle<sup>10</sup>, Michael Zietz<sup>11</sup>, Undina  
8 Gisladottir<sup>11</sup>, Vijendra Ramlall<sup>11,12</sup>, Craig D. Westover<sup>1</sup>, Krista Ryon<sup>1</sup>, Benjamin Young<sup>1</sup>,  
9 Chandrima Bhattacharya<sup>1</sup>, Phyllis Ruggiero<sup>6</sup>, Bradley W. Langhorst<sup>13</sup>, Nathan Tanner<sup>13</sup>, Justyna  
10 Gawrys<sup>6</sup>, Dmitry Meleshko<sup>1,4</sup>, Dong Xu<sup>14</sup>, Peter A. D. Steel<sup>15</sup>, Amos J. Shemesh<sup>15</sup>, Jenny  
11 Xiang<sup>14,16</sup>, Jean Thierry-Mieg<sup>17</sup>, Danielle Thierry-Mieg<sup>17</sup>, Robert E. Schwartz<sup>18</sup>, Angelika  
12 Iftner<sup>10</sup>, Daniela Bezdán<sup>10</sup>, John Siple<sup>6</sup>, Lin Cong<sup>6</sup>, Arryn Craney<sup>6</sup>, Priya Velu<sup>6</sup>, Ari M.  
13 Melnick<sup>18</sup>, Iman Hajirasouliha<sup>1,2,7</sup>, Stacy M. Horner<sup>19,20</sup>, Thomas Iftner<sup>10</sup>, Mirella Salvatore<sup>16</sup>,  
14 Massimo Loda<sup>6</sup>, Lars F. Westblade<sup>6,16</sup>, Melissa Cushing<sup>6</sup>, Shawn Levy<sup>21</sup>, Shixiu Wu<sup>22,23</sup>,  
15 Nicholas Tatonetti<sup>11</sup>, Marcin Imielinski<sup>5,6,7\*</sup>, Hanna Rennert<sup>6\*</sup>, Christopher E. Mason<sup>1,2,3,24\*</sup>

## 16 17 **Affiliations:**

18 <sup>1</sup>Department of Physiology and Biophysics, Weill Cornell Medicine, NY, USA.  
19 <sup>2</sup>The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational  
20 Biomedicine, Weill Cornell Medicine, NY, USA  
21 <sup>3</sup>WorldQuant Initiative for Quantitative Prediction, Weill Cornell Medicine, NY, USA  
22 <sup>4</sup>Tri-Institutional Computational Biol. & Medicine Program, Weill Cornell Medicine, NY, USA  
23 <sup>5</sup>New York Genome Center, NY, USA  
24 <sup>6</sup>Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, NY, USA  
25 <sup>7</sup>Englander Institute for Precision Medicine and the Meyer Cancer Center, Weill Cornell  
26 Medicine, NY, USA  
27 <sup>8</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA  
28 <sup>9</sup>Clinical & Translational Science Center, Weill Cornell Medicine, NY, USA  
29 <sup>10</sup>Institute of Medical Virology and Epidemiology of Viral Diseases, University Hospital  
30 Tuebingen, Germany  
31 <sup>11</sup>Department of Biomedical Informatics, Department of Systems Biology, Department of  
32 Medicine, Institute for Genomic Medicine, Columbia University, NY, USA  
33 <sup>12</sup>Department of Cellular, Molecular Physiology & Biophysics, Columbia University, NY, USA  
34 <sup>13</sup>New England Biolabs, MA, USA  
35 <sup>14</sup>Genomics Resources Core Facility, Weill Cornell Medicine, NY, USA  
36 <sup>15</sup>Department of Emergency Medicine, Weill Cornell Medicine, NY, USA  
37 <sup>16</sup>Division of Infectious Diseases, Department of Medicine, Weill Cornell Medicine, NY, USA  
38 <sup>17</sup>National Center for Biotechnology Information, National Library of Medicine, National  
39 Institute of Health, MD, USA  
40 <sup>18</sup>Department of Medicine, Weill Cornell Medicine, NY, USA  
41 <sup>19</sup>Department of Molecular Genetics and Microbiology, Duke University Medical Center, NC,  
42 USA  
43 <sup>20</sup>Department of Medicine, Duke University Medical Center, NC, USA  
44 <sup>21</sup>HudsonAlpha Discovery Institute, Huntsville, AL, USA  
45 <sup>22</sup>Hangzhou Cancer Institute, Hangzhou Cancer Hospital, Hangzhou, China  
46 <sup>23</sup>Department of Radiation Oncology, Hangzhou Cancer Hospital, Hangzhou, China

47 <sup>24</sup>The Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, NY, USA

48 #These authors contributed equally to this work

49 \*To whom correspondence should be addressed

50

### 51 **Corresponding Authors:**

52 Marcin Imielinski (mai9037@med.cornell.edu); Hanna Rennert (har2006@med.cornell.edu);

53 Christopher E. Mason (chm2042@med.cornell.edu)

54

### 55 **Abstract**

56 The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has caused thousands of  
57 deaths worldwide, including >18,000 in New York City (NYC) alone. The sudden emergence of  
58 this pandemic has highlighted a pressing clinical need for rapid, scalable diagnostics that can  
59 detect infection, interrogate strain evolution, and identify novel patient biomarkers. To address  
60 these challenges, we designed a fast (30-minute) colorimetric test (LAMP) for SARS-CoV-2  
61 infection from naso/oropharyngeal swabs, plus a large-scale shotgun metatranscriptomics  
62 platform (total-RNA-seq) for host, bacterial, and viral profiling. We applied both technologies  
63 across 857 SARS-CoV-2 clinical specimens and 86 NYC subway samples, providing a broad  
64 molecular portrait of the COVID-19 NYC outbreak. Our results define new features of SARS-  
65 CoV-2 evolution, nominate a novel, NYC-enriched viral subclade, reveal specific host responses  
66 in interferon, ACE, hematological, and olfaction pathways, and examine risks associated with  
67 use of ACE inhibitors and angiotensin receptor blockers. Together, these findings have  
68 immediate applications to SARS-CoV-2 diagnostics, public health, and new therapeutic targets.

69

70 **Keywords:** severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), coronavirus  
71 disease 2019 (COVID-19), loop-mediated isothermal amplification (LAMP), quantitative reverse  
72 transcription polymerase chain reaction (qRT-PCR), next-generation sequencing (NGS), RNA-  
73 seq, global health.

74

### 75 **Introduction**

76 In March 2020, the World Health Organization (WHO) declared a pandemic of the coronavirus  
77 disease 2019 (COVID-19), an infection caused by severe acute respiratory syndrome coronavirus  
78 2 (SARS-CoV-2) (He *et al.*, 2020). Since its start, more than three million of cases and more  
79 than two hundred of thousand deaths have been reported (<https://coronavirus.jhu.edu>), with an  
80 especially high burden in New York City (NYC). Since the presenting symptoms of COVID-19  
81 can resemble those of common viral respiratory infections, a molecular diagnosis is required to  
82 reliably distinguish a SARS-CoV-2 infection from influenza and other respiratory illnesses  
83 (Guan *et al.*, 2020, Zhou *et al.*, 2020).

84

85 The current gold standard (qRT-PCR-based) approaches to SARS-CoV-2 molecular testing are  
86 largely limited to hospital or academic laboratories and reserved for the most severe cases, with  
87 limited accessibility to the general public. As a result, the true prevalence of SARS-CoV-2 in the  
88 population is mostly unknown, particularly among pre-symptomatic, mild symptomatic, or  
89 asymptomatic cases. Though several novel, scalable biotechnological approaches for viral testing  
90 have recently emerged (e.g., CRISPR-Cas12a (Broughton *et al.*, 2020) or CRISPR-Cas13  
91 (Metsky *et al.*, 2020) on paper-based detection systems, or loop-mediated isothermal  
92 amplification (LAMP) (Tanner *et al.*, 2015, Zhang *et al.*, 2020, Yu *et al.*, 2020, Schmid-Burgk

93 *et al.*, 2020), these have not been validated against gold-standard clinical assays or next-  
94 generation sequencing (NGS).

95  
96 The lack of widely available, rapid SARS-CoV-2 diagnostics has fundamentally limited the  
97 public health approach to COVID-19, including hampering ability to implement contact tracing  
98 and accurate estimation of infection fatality rates. In addition, the persistence of SARS-CoV-2  
99 across a range of surfaces (van Doremalen *et al.*, 2020) and hospital areas (Ong *et al.*, 2020)  
100 raises the specter of COVID-19 spread via fomite transmission. A key question is whether the  
101 environmental surface distribution of SARS-CoV-2 in high-touch, high-traffic areas (e.g.  
102 subways) may have driven its rapid emergence in certain regions (e.g., NYC). As yet, little  
103 information exists about RNA viruses in public areas.

104  
105 Nonetheless, genomic viral surveillance resources, such as Global Initiative on Sharing All  
106 Influenza Data (GISAID), the Johns Hopkins University (JHU) Dashboard, and Nextstrain  
107 (<https://nextstrain.org/ncov>) (Gardy *et al.*, 2015, Dong *et al.*, 2020, Meyers *et al.*, 2020, Hadfield  
108 *et al.*, 2018) have enabled dynamic tracking of patient-derived viral evolution during the  
109 COVID-19 pandemic. These efforts still only cover a fraction (<0.5%) of documented cases,  
110 motivating ongoing viral profiling efforts. Unlike the results of simple binary diagnostic tests,  
111 full-length viral genome sequences can reveal patterns of strain evolution, temporal and  
112 geographic trajectories of viral spread, and correlations between specific genotypes and clinical  
113 features (e.g. disease severity, comorbidities, and viral load). These genomic epidemiology  
114 efforts have already played a crucial public health role in confirming community spread of  
115 SARS-CoV-2 in the USA (Zhao *et al.*, 2020, Gonzalez-Reiche *et al.*, 2020, Fauver *et al.*, 2020)  
116 as well as in China (Lu *et al.*, 2020). However, standard approaches to viral profiling (i.e.  
117 targeted sequencing) fail to provide information on either the host immune response or microbial  
118 co-infections, both which might impact the clinical presentation of COVID-19 and provide  
119 directions for therapeutic intervention and management.

120  
121 To address these gaps in knowledge of disease and technology, we designed and validated a  
122 rapid reverse transcriptase LAMP (RT-LAMP) assay to detect SARS-CoV-2 infection from  
123 nasopharyngeal swab specimens and oropharyngeal swab lysates. We then integrated this assay  
124 into a large-scale host and viral profiling platform that combines targeted diagnostics with  
125 shotgun metatranscriptomics (**Figure 1a**). We applied this total RNA-seq platform to 857  
126 samples from 735 confirmed or suspected COVID-19 cases at New York-Presbyterian Hospital-  
127 Weill Cornell Medical Center (NYPH-WCMC) and 86 environmental samples collected from  
128 high-transit areas in the NYC subway in early March 2020. Our results define genetic features of  
129 the NYC outbreak and implicate specific host responses in SARS-CoV-2 infection, including  
130 perturbations of the angiotensin converting enzyme (ACE) pathway.

131  
132 **Results**

133  
134 ***Rapid, single tube detection of SARS-CoV-2***

135 We developed a colorimetric assay to quickly detect and quantify SARS-CoV-2 viral load in  
136 patient and environmental samples, utilizing a set of six LAMP primers and simple single tube  
137 protocol (**Figure 1a-b**). Primers were designed to create two nested loops and amplify within the  
138 SARS-CoV-2 nucleocapsid gene (N gene), which enabled a 30-minute reaction workflow.

139 Related pathogens, high prevalence disease agents, and normal or pathogenic microbiota that are  
140 reasonably likely to be encountered in the clinical samples were evaluated to identify the percent  
141 homology between the primer sequences and these other organisms, and the probes were also  
142 designed to avoid known polymorphisms (see Methods).

143  
144 To validate the assay, we first evaluated two synthetic RNAs (see Methods) whose sequences  
145 matched the viral sequences of patients from Wuhan, China and Melbourne, Australia (**Supp.**  
146 **Fig. 1**). The first control (MT007544.1) was used to test the analytical sensitivity via the limit of  
147 detection (LoD), titrated from 1 million molecules of virus ( $10^6$ ) down to a single copy, using  
148 serial  $\log_{10}$  dilutions. The reaction output was measured at 0-, 20-, and 30-minute intervals  
149 (**Figure 1c**, **Supp. Figure 2a**) before the samples were heated to 95°C for inactivation. LAMP  
150 fluorescence correlated closely with SARS-CoV-2 RNA viral copies (**Figure 1d**), with an  
151 overlap of the median signal from negative controls at lower levels (0-10 total copies per  
152 reaction) of viral RNA (n=10). The LoD was found to be between 5–25 viral total copies for the  
153 dual primer, single-tube reaction (N gene and E gene), and this was then replicated to show a  
154 similar LoD on the second synthetic control from a patient from Wuhan (**Supp. Figure 2b, 3a**).

155  
156 To optimize the LAMP assay for clinical samples, we first tested the reaction on a range of  
157 sample types, dilutions, and reaction volumes. We used a set of 201 samples from known or  
158 suspected COVID-19 cases that were tested for SARS-CoV-2 by a commercial qRT-PCR assay  
159 that has been the primary clinical test at NYPH-WCMC since early March. These comprised 69  
160 nasopharyngeal (NP) swab samples that tested positive (qRT-PCR positives, Ct<40) and 132  
161 samples that tested negative (qRT-PCR negative, Ct  $\geq$  40) (see Methods). qRT-PCR positive  
162 samples showed a much higher LAMP fluorescence than qRT-PCR negative samples, with  
163 superior performance observed at higher (10 $\mu$ L vs 5 $\mu$ L) input volumes (**Supp. Figure 3b**). qRT-  
164 PCR positive samples that failed to generate LAMP fluorescence were associated with lower  
165 viral genome abundance (high cycle threshold [Ct] value qRT-PCR). We obtained similar  
166 performance on bulk oropharyngeal swab lysate (**Supp. Figure 3c-e**), including increasing  
167 reaction sensitivity as a function of viral copy number. This lysate protocol required only a 30-  
168 minute lysis time for the oral collections and 30-minute LAMP reaction time. These results  
169 indicate robust performance of the LAMP assay across a broad range of purified nucleic acids as  
170 well as crude cellular lysates (see Methods).

171  
172 Analysis of Receiver Operator Characteristic (ROC) curves revealed an optimal threshold of  
173 11,140 QuantiFluor relative fluorescence units (RFUs), which yielded an overall sensitivity of  
174 95.6% and specificity of 99.2% (**Figure 1e**). We observed higher LAMP sensitivity at higher  
175 viral loads, as determined by qRT-PCR Ct values (**Supp. Fig. 4**). The highest viral load (Ct <20)  
176 showed 100.0% sensitivity and 97.4% specificity, compared with the sensitivity at the lowest  
177 viral load at 80.0% (Ct >28, RFU below 7010). These same LAMP assay thresholds yielded  
178 consistent test positivity for synthetic RNA positive controls (Twist Biosciences) as well as  
179 clinical spike-in carrier RNAs (20/20) (**Supp. Fig. 4**), and consistent signal from clinical viral  
180 positives in Vero E6 cells (100.0%, 12/12) and blank clinical buffer negatives (100.0% 8/8).

### 181 ***Shotgun metatranscriptomics platform for viral, microbiome, and host genomics***

182 To provide orthogonal validation of our LAMP assay and further investigate the biological  
183 characteristics of qRT-PCR positive and negative specimens, we developed a shotgun  
184

185 metatranscriptomics platform utilizing total RNA-seq (RNA-sequencing with ribosomal RNA  
186 depletion) to profile all RNAs from patient and environmental swab specimens (**Figure 1a**). We  
187 sequenced 857 RNA-seq libraries (**Figure 2**) across 735 COVID-19 cases and controls  
188 (including the 201 tested with LAMP) to an average of 63.2M read pairs per sample. This  
189 spanned 243 qRT-PCR positive and 546 qRT-PCR negative samples, 17 positive controls (Vero  
190 E6 cells), and 33 negative controls (buffer). Replicate RNA-seq libraries from clinical samples  
191 were merged into a final set of 216 qRT-PCR positive samples and 519 qRT-PCR-negatives.  
192

193 We aligned total RNA-seq reads to the human genome and NCBI reference databases (see  
194 methods), Kraken2 classification of non-human sequences revealed abundant bacterial or SARS-  
195 CoV-2 RNA alignments, with a lower proportion of sequences mapping to fungi, archaea, or  
196 other viruses (**Figure 2a**). Positive controls and clinical samples demonstrating qRT-PCR Ct  
197 values consistent with medium and high viral loads were significantly ( $p < 1 \times 10^{-16}$ ) enriched in  
198 SARS-CoV-2 genome alignments (median reads per kilobase per million mapped reads, RPKM)  
199 relative to negative controls and qRT-PCR negative clinical samples (**Figure 2b**). A comparison  
200 of LAMP fluorescence, qRT-PCR Ct values, and total-RNA Seq RPKM across 201 clinical  
201 specimens demonstrated consistent estimates of SARS-CoV-2 viral abundance across these three  
202 technologies ( $R_{\text{seq\_vs\_Ct}} = -0.84$ ,  $R_{\text{seq\_vs\_lamp}} = 0.82$ ,  $R_{\text{lamp\_vs\_Ct}} = -0.80$ ) (**Figure 2c, Supp. Figure**  
203 **5a**), as well as RNA-seq extraction replicates (**Supp. Figure 5b**). These results provide  
204 additional validation for the LAMP SARS-CoV-2 assay on clinical samples.  
205

206 Analysis of total RNA-seq sequences also enabled mapping of likely co-infections and  
207 colonization with commensal species across both clinical positives and negatives in our sample  
208 set. We identified additional RNA viruses and organisms that distinguished high, medium, and  
209 lower viral load patients (**Supp Table 1, Supp. Figure 6**), with overall similarity observed  
210 across patients in the bacterial RNA fractions (**Supp. Figure 7a**). However, other known  
211 respiratory viruses were observed in some of the COVID-19-positive patients (3.2%), including  
212 human coronaviruses 229E, coronaviruses NL63, influenza A, and human mastadenovirus and  
213 metapneumovirus (**Supp. Figure 7a**). Most patients (209/216, 96.8%) that presented with the  
214 SARS-CoV-2 virus showed no evidence of other respiratory viruses, although bacterial mapping  
215 showed 102 patients with reads mapping to *Mycoplasma pneumonia* (relative abundance  $>$   
216 0.01%) (**Supp. Table 1**).  
217

### 218 ***Environmental sampling of SARS-CoV-2 in the NYC subway***

219 Having validated a rapid LAMP assay and shotgun metatranscriptomics platform, we next  
220 investigated the environmental distribution of SARS-CoV-2 in high-transit areas of the NYC  
221 subway at the beginning of the NYC pandemic. We collected 62 samples from handrails, kiosks,  
222 and floors in Grand Central and Times Square subway stations between March 6 and 13th, 2020  
223 and generated 86 RNA-seq libraries. Each sample was collected using a sterile and DNA/RNA  
224 free swab, following the MetaSUB sampling protocols for nucleic acid stabilization (Danko *et*  
225 *al.*, 2020).  
226

227 To further investigate these 86 environmental samples for the presence of SARS-CoV-2, we  
228 generated total RNA-seq libraries and investigated the distribution of non-human sequences.  
229 These samples demonstrated a mix of fungal and archaeal species that was consistent with  
230 underground subway origin (Danko *et al.*, 2020) (**Supp. Table 1**). However, we did not observe

231 significant counts or proportions (**Figure 3a**) of SARS-CoV-2 reads, which was particularly  
232 clear with dual-index library preparations (**Supp. Figure 8**). However, a broad range of other  
233 bacterial and viral species was found, including a large set of phages (e.g. *Streptomyces* phage  
234 VWB), desiccation-tolerant bacteria (e.g. *Deinococcus radiodurans*), and more abundant  
235 bacterial and archaeal RNA than the clinical samples (**Supp. Table 1**). Of note, these were  
236 checked against a database of putative false positives (**Supp. Table 2**), which was created from  
237 *in silico* fragmentation of the SARS-CoV-2 genome and mapping against the same database.  
238 Taken together, these results indicate that high transit surfaces were not likely to harbor  
239 significant levels of SARS-CoV-2 in the early phases of the NYC epidemic.

240

### 241 ***SARS-CoV-2 assemblies from shotgun metatranscriptomes***

242 The abundance of SARS-CoV-2 alignments from total RNA-seq was sufficient to provide >10x  
243 coverage of the viral genomes and yield high quality, full-length assemblies for 155 samples,  
244 including 146 (67.6%) of qRT-PCR positive patient samples (**Figure 3a, Supp. Figure 6**). The  
245 97 qRT-PCR positive samples that yielded partial or no assembly demonstrated low or uneven  
246 genome-wide read depth, though these were increased on average relative to qRT-PCR  
247 negatives. We also generated full length SARS-CoV-2 assemblies for 9 samples that were found  
248 negative by qRT-PCR (**Figure 3a**). Each of these demonstrated a high RPKM viral load, an  
249 abundance of reads evenly covering the SARS-CoV-2 genome, and high (>0.5) variant allele  
250 fractions (VAFs) of SNV's commonly seen in positive samples, as well as a recent build of  
251 GISAID (**Figure 3b**). Apart from these common variant sites, we did not find additional  
252 recurrent mutations that might explain why these cases were not detected by qRT-PCR (e.g.  
253 primer site mutation).

254

255 When examining the viral genomes, we identified a total of 1,147 instances of 165 unique  
256 variants across 155 assemblies, including 1,143 single nucleotide variants (SNVs) and 4  
257 deletions. Variant calling of assemblies and RNA-seq reference alignments showed most  
258 (87.5%) variants with VAFs >0.95 and high (>100) numbers of variant-supporting reads.  
259 Analysis of VAF posterior probability distributions identified a subset of alleles whose VAFs  
260 were confidently (probability > 0.95) above 0.05 but below 0.95. We labeled these alleles as  
261 "het" (heterogenous) variants and refer to the remaining (high VAF) consensus assembly variants  
262 as "homogeneous." Many of these het variants were associated with robust read support despite  
263 frequently having VAFs below 0.5 and not being incorporated into the consensus assembly  
264 sequence. A subset of samples (11/155) samples harbored at least 5 het variants, with a  
265 significant enrichment ( $P = 0.0054$ ) in qRT-PCR negative cases (**Figure 3c**). Analyzing the  
266 population allele frequencies (PAF) of these het variants across GISAID revealed the vast  
267 majority (91%) to be rare (PAF<0.1) or even private to that sample, indicating that these were  
268 unlikely to be due to cross contamination (**Figure 3d**).

269

270 We then asked whether het variants might indicate the presence of a viral subclone, arising either  
271 through co-infection or intra-host adaptation. Analyzing het-rich (>5 het variants) samples,  
272 revealed that set of het variants observed in a single sample showed similar VAFs, yielding clear  
273 peaks when summing their individual posterior probability densities (**Figure 3e**). These results  
274 were consistent with the presence of a viral subclone at some fraction  $f$  corresponding to the  
275 modes of each of these VAF distributions (0.1-0.2). Interestingly, each of the three samples  
276 shown in **Figure 3e** also harbored a clear peak of homogenous variants with VAFs near 1. These

277 results suggest that the viral subclone is not the result of a coinfection by two viruses, which  
278 would yield a second peak at 1-*f* at every major clone variant that has the reference allele in the  
279 minor subclone. Het variants were found to occur at common sites of consensus variants across  
280 samples, however, rare private hets were also present (**Figure 3f**).

281

### 282 ***Global phylogenetic relationship of NYC strains***

283 We next integrated 141 full-length NYPH-WCMC sequences with 7,806 SARS-CoV-2  
284 sequences obtained from a recent GISAID build (downloaded on 4/20/2020) into a maximum  
285 likelihood phylogenetic tree using the Nextstrain algorithm (Hadfield *et al.*, 2018) (**Figure 4a**,  
286 see Methods). This integration resulted in 141/155 NYPH-WCMC sequences integrated into the  
287 tree. Analyzing worldwide groupings annotated by the Nextstrain database, we found a high  
288 proportion (>87%, 123/141) of incorporated samples were associated with A2, a large Western  
289 European-derived clade that comprises nearly half of the sequences in GISAID, which agrees  
290 with recently published findings (Gonzalez-Reiche *et al.*, 2020). In addition, we found a minority  
291 of samples from clades with mixed Asian and European origins (A1a, 6% of cases; B1, 4% of  
292 cases; B, 3% of cases). Among these, we identified a clear predominance (94.1%, 146/155) of  
293 the "L" strain, which is defined by a reference base at position 28143 (ORF8) and previously  
294 associated with severe cases in Wuhan (Tang *et al.*, 2020) (**Figure 4b**).

295

296 Among the A2 cases in our cohort, we found a recurrent 9 bp in-frame deletion (g.686\_694del)  
297 in the gene encoding non-structural protein 1 (NSP1). NSP1 is a putative SARS-CoV-2 virulence  
298 factor that is highly divergent across betacoronaviruses (Narayanan *et al.*, 2015). This deletion  
299 was assembled in 3 of 93 NYPH-WCMC samples, in which it showed robust read support with  
300 near >0.95 VAF across thousands of high-quality alignments to the locus (**Figure 4c**). An  
301 identical deletion was observed in 15 other samples in the GISAID database, spanning samples  
302 from England, Iceland, and Canada, as well as three additional cases from New York. Among  
303 these, 12 of 15 were labeled as part of clade A2 by Nextstrain, while two (Washington state)  
304 cases were associated with a distant clade B1 (**Figure 4d**). This deletion removes three amino  
305 acids (p.141\_143KSD), including a residue that is variant (143Y>F) relative to the SARS-CoV-1  
306 genome (**Figure 4c**). Interestingly, a 16th case (Washington, clade A2, **Figure 4c**) showed a  
307 related but not identical 9 bp in-frame deletion (g.686\_694del) that removes the same three  
308 residues. These residues occur in a conserved portion of the C-terminal region of NSP1, which  
309 has been linked to host chemokine dysregulation and translational inhibition in SARS-CoV  
310 (Narayanan *et al.*, 2015).

311

### 312 ***A novel SARS-CoV-2 subclade enriched in NYC cases***

313 Given the predominance of clade A2 among NYC cases, we next asked whether specific variants  
314 were enriched in NYC A2 samples relative to other regions' A2 samples. Comparing variant  
315 frequencies between A2 cases in our data and non-NYC cases in GISAID, we identified six  
316 enriched loci with FDR<0.1 (Fisher's exact test, one-tailed) that were over-represented in NYC  
317 A2 samples (**Figure 5a**). All six of these variants were significantly enriched (Bonferroni-  
318 adjusted p-value <0.05) in a second analysis comparing all other cases from NYC in GISAID vs.  
319 the rest of the world (**Figure 5a**). Visualizing the most significant of these variants across the  
320 Nextstrain derived phylogenetic tree (**Figure 5b**) demonstrated that the highest frequency of  
321 these variants (25563G>T), defined a novel subclade of A2, which we call A2-25563. This  
322 subclade comprised 84.4% (572/678) of (NYPH-WCMC and GISAID) NYC samples, and

323 92.7% (572/617) of the A2 portion of NYC cases (Sinai and NYU). Interestingly, while all three  
324 of the NYPH-WCMC A2 cases harboring the aforementioned NSP1 9-bp deletion were in A2-  
325 25563, many of the remaining cases were distributed across other subclades. Unlike the SNVs  
326 used to define A2-25563, this deletion appeared to be highly polyphyletic with 7 unique  
327 phylogenetic clusters associated with the deletion across a variety of clades (**Figure 4c, Figure**  
328 **5b**).

329  
330 Investigating the geographical and temporal distribution of non-NYC samples in A2-25563, we  
331 found a global distribution with likely origin in Western Europe (**Figure 5c, 5d**). The first noted  
332 case of A2-25563 in GISAID was in Belgium on 2/21/2020, but since then, this putative  
333 subclade has represented a minority (20%) of Western European cases. While the prevalence of  
334 A2-25563 in NYC appeared to be consistently high across sampling dates, it constituted only a  
335 moderate proportion of cases from the Western and (non-New York portion of the) Eastern USA  
336 (<50%) in early March (**Figure 5d**). Strikingly, we found that of the prevalence of the A2-25563  
337 clade across Eastern and Western United States steadily and significantly increased through late  
338 March and early April, while remaining stable in the Midwest. Among the early locations  
339 showing high fractions of A2-25563 was West Africa, which has steadily decreased in  
340 prevalence since early March (**Figure 5d**). The SNVs defining A2-25563 subclade were non-  
341 synonymous variants targeting genes encoding the non-structural protein 2 (NSP2), the viral  
342 replicase, and ORF3a, a poorly characterized SARS-CoV-2 protein with putative roles in  
343 inflammation (Siu *et al.*, 2019). This subclade defining ORF3a site (p.Gln57His) is distinct from  
344 that which defines the previously characterized L strain (p.Leu84Ser) (Tang *et al.*, 2020).

#### 345 346 ***Defining the SARS-CoV-2 host transcriptome***

347 We leveraged the comprehensive nature of the total RNA-seq profiles to better understand the  
348 host transcriptome during SARS-CoV-2 infection. Cell proportion analyses using the MUSIC  
349 deconvolution algorithm showed enriched proportions of cell types spanning goblet, ciliated  
350 airway, and epithelial cells across all samples (**Supp. Figure 9**). Differentially expressed genes  
351 (DEGs) associated with SARS-CoV-2 infection were calculated using limma voom and DESeq2  
352 (see Methods). Overall, 757 significant DEGs ( $q < 0.01$ , >1.5-fold change) were found the qRT-  
353 PCR positive vs. negative samples (**Supp. Table 3**), spanning 350 up-regulated DEGs and 407  
354 down-regulated DEGs, and a total of 8851 unique genes ( $q < 0.01$ , >1.5-fold change) across  
355 different subgroups including a subset that were distinct from other detected viral respiratory  
356 infections (**Figure 6a, Supp. Figure 12**). These spanned viral response pathways, innate immune  
357 response, and interferon signaling (**Figure 6b,d, Supp. Figure 10a**), but also included iron,  
358 olfaction, calcium, and ribosome dysregulation.

359  
360 Differentially expressed host genes indicated a wide range of antiviral responses, including a  
361 common interferon response across all ranges of viral levels, which was significantly higher  
362 when compared to SARS-CoV-2 negative samples that harbored other respiratory viruses  
363 (**Figure 6a,b**). Notably, host cells showed an increase in angiotensin converting enzyme 2  
364 (*ACE2*) expression (**Figure 6b**) ( $p\text{-value} = 1.4 \times 10^{-9}$ ), which is the SARS-CoV-2 cellular receptor  
365 (Hoffmann *et al.*, 2020). This critical gene for viral entry (Sungnak *et al.*, 2020) exhibited a  
366 dose-dependent expression concomitant with the higher levels of SARS-CoV-2 virus, along with  
367 IFI27 (Interferon Alpha Inducible Protein 27,  $p < 2.2 \times 10^{-16}$ ) and IFI6 (Interferon Alpha Inducible  
368 Protein 6,  $p < 2.2 \times 10^{-16}$ ). The DEGs also included HERC6 (HECT Containing E3 Ubiquitin



369 Protein Ligase Family Member 6), which aids Class I MHC-mediated antigen processing and  
370 Interferon-Stimulated Genes (ISGs) (**Figure 6b,c,d**), underscoring the impact of the virus on  
371 these cells' immune response (Oudshoorn *et al.*, 2012). Also, a subset of cytokines (CXCL10,  
372 CXCL11, CCL8) showed the highest spike of expression in the higher viral load sub-group  
373 (**Supp. Figure 10**), matching previous results from animal models and infected cells (Blanco-  
374 Melo *et al.*, 2020). The host transcriptome that exhibited the greatest amount of DEGs were  
375 those with the highest viral titer (**Figure 6a**).

376  
377 Down-regulated genes and those with a negative enrichment score (NES) were functionally  
378 distinct (**Figure 6d**). This included a significant decrease in gene expression for the olfactory  
379 receptor pathway genes (q-value =0.0005, **Supp. Table 4**), which is consistent with a COVID-19  
380 phenotype wherein patients lose their sense of smell. Other down-regulated genes included the  
381 transmembrane serine protease *TMPRSS-11B*, which regulates lung cell growth (Updegraff *et al.*,  
382 2018) and *ALAS2*, a gene which makes erythroid ALA-synthase (Ajioka *et al.*, 2006) that is  
383 found in developing erythroblasts (red blood cells). ALA-synthase plays an important role in the  
384 production of heme TRIM2 E3 ubiquitin ligase induced during late erythropoiesis, which  
385 indicated a connection to hematological and iron (heme) regulation during infection (**Figure 6d**).  
386 Accordingly, genes in a related biological network were significantly enriched based on Gene  
387 Ontology (GO) pathways for iron regulation (q-value 0.04, **Supp. Table 4**). Both the up-  
388 regulated and down-regulated gene expression differences in were distinct from those of house-  
389 keeping genes (**Supp. Figure 11**), which stayed mostly stable during infection.

#### 390 ***ACE inhibitor/angiotensin receptor blocker usage correlates with COVID-19***

391 Finally, given our observation of increased *ACE2* gene expression in patients with high SARS-  
392 CoV-2 viral load, we then investigated the interplay of receiving pharmacologic angiotensin  
393 converting enzyme inhibition (ACEI) or angiotensin receptor blockers (ARBs) for hypertension  
394 and clinical features of COVID-19 disease. Since *ACE2* expression can be increased in patients  
395 taking ACEIs and ARBs (Agata *et al.*, 2006), the observed correlation of viral titer with *ACE2*  
396 expression may be attributed to the pre-infection use of such inhibitors, which is common in  
397 older patients and those with comorbidities (Fang *et al.*, 2020). To address this, we analyzed an  
398 observational cohort of 8,856 patients with suspected SARS-CoV-2 infection from New York  
399 Presbyterian Hospital-Columbia University Irving Medical Center (NYPH-CUIMC) for their  
400 usage of ACEIs and ARBs (4,829 who tested positive). We found that use of ACEIs/ARBs was  
401 strongly associated with testing positive in patients suspected of SARS-CoV-2 infection (Odds  
402 Ratio, OR=2.7, and 95% Confidence Interval, CI=2.2-3.94, p=7.43E-28, **Figure 6e**). This result  
403 was consistent when corrected for age, sex, and IL-6 (OR=1.5 CI:1.2-1.8), when corrected for  
404 comorbidities (**Table 1**) (OR=1.754 CI:1.4-2.0) and when only considering patients with  
405 hypertension (OR=1.8, CI: 1.4-2.3). For comparison, we repeated this analysis with calcium  
406 channel blockers (CCBs), another class of anti-hypertensives, and found no association between  
407 CCB exposure and infection. We found that patients with hypertension who take ACEIs and  
408 ARBs have higher infection rates than patients with hypertension on CCBs (OR=1.5 CI: 1.2-2.0,  
409 p=0.0011) (**Table 2**).

410  
411  
412 We found a range of effects when investigating the relationship between ACEI/ARB usage and  
413 morbidity and mortality. In an uncorrected univariate analysis, ACEI/ARB usage conferred an  
414 increased risk of intubation and mortality for SARS-CoV-2 positive patients (intubation: HR=2.8

415 CI:2.3-3.5, 6.99E-24) (**Figure 6e** and **Figure 6f**) and mortality: HR=2.2 CI:1.8-2.7,  $p=3.18E-14$   
416 (**Figure 6e** and **Figure 6g**). These results hold when correct for age, sex, and IL6 (**Figure 6e**).  
417 However, when correcting for comorbidities, only intubation was significant (HR=1.7 CI:1.4-  
418 2.2,  $p=5.91E-07$ ). Further, in the analysis where only patients with hypertension were included,  
419 we found modest evidence for both (i) a risk of intubation in the univariate model (HR=1.4  
420 CI:1.1-1.8,  $p=0.02$ ) and (ii) a protective effect of ACEI/ARBs for mortality in the covariate  
421 models (HR=0.8 CI:0.6-1.0,  $p=0.03$ ).  
422

423 Additionally, we confirmed previously reported risk factors for both mechanical respiration and  
424 mortality. For requirement of mechanical respiration, we found significant risk effects from male  
425 sex ( $p=2.76E-03$ ), diabetes ( $p=2.64E-02$ ), and IL-6 ( $p=1.07E-121$ ). In addition, for mortality we  
426 found significant effects from age ( $p=1.10E-86$ ), male sex ( $p=1.98E-04$ ), diabetes ( $p=6.04E-04$ ),  
427 being overweight ( $p=3.86E-03$ ), and IL-6 ( $p=8.68E-23$ ). Finally, in a post-hoc comparative  
428 analysis between specific ACEIs, we found no significant differences in infection rates,  
429 intubation, or mortality with any single ACEI drug over another (**Supp. Figure 13**). In a similar  
430 analysis on ARBs, we found that olmesartan was associated with an increased risk of mortality  
431 (HR=2.3 CI: 1.1-4.9,  $p=0.03$ ,  $N=28$ ) in the hypertension-only analysis. We found no risk in the  
432 larger cohort nor for any other single ARB (**Supp. Figure 13**, with complete modeling results  
433 are provided in **Supplemental Tables 6-9**).  
434

### 435 **Discussion**

436 During a large-scale pandemic with exponential spread, such as COVID-19, scalable methods for  
437 diagnosis and screening are crucial for both mitigation and containment (Lan *et al.*, 2020, Liu *et*  
438 *al.*, 2020). While hospital-grade, core lab devices can achieve massive throughput (thousands of  
439 samples per day), a key limitation of these assays is accessibility of testing facilities to patients,  
440 the logistics of sample transport, and timely test reporting. These limitations become even more  
441 stark in the context of widespread quarantines and nationwide lockdowns, where requiring  
442 patients to travel (even for viral testing) incurs significant personal and public health risks.  
443

444 The most urgent diagnostic need in this situation is for scalable rapid point-of-care tests that can  
445 be potentially implemented in the home. Our validation of a rapid one-tube, dual-primer  
446 colorimetric SARS-CoV-2 assay with both qRT-PCR and total RNA-seq provides a potential  
447 solution to this problem. Further work will be needed to assess whether this LAMP assay can  
448 detect the presence of SARS-CoV-2 at even lower (but clinically relevant) viral concentrations in  
449 specimen types that are less cumbersome to collect than naso/oropharyngeal swabs (e.g. saliva,  
450 stool). As we demonstrate, this LAMP SARS-CoV-2 assay can be also applied for environmental  
451 sampling, which may be crucial in the containment and recovery phases of this pandemic.  
452 Specifically, LAMP positivity may quickly indicate if an area is infectious and a negative result  
453 (with appropriate confirmation) will possibly represent a lower risk. Indeed, these tools and  
454 methods can help create a viral “weather report” if broadly used and partnered with continual  
455 validation.  
456

457 Total RNA-sequencing data enabled a complete genetic map of the viruses in a significant subset  
458 of our samples. Remarkably, we found 9 qRT-PCR negative cases among the 155 cases that had  
459 sufficient reads to assemble the SARS-CoV-2 genome *de novo*. Though these likely qRT-PCR  
460 false negatives could not be attributed to specific sequence changes (e.g. primer site mutations),

461 their high frequency underscores the limitation of "gold standard" qRT-PCR approaches for  
462 SARS-CoV-2 detection. These results further highlight the need for open-source primer design  
463 (the RT-PCR methods for this FDA-approved test are proprietary), so these assays can be  
464 updated as a more granular picture of strain diversity and evolution is obtained through  
465 worldwide sequence efforts.

466  
467 Notably, our phylogenetic analysis nominates an A2 subclade (A2-25563, defined by  
468 25563G>T) which comprises the majority of known NYC samples, including those sequenced  
469 outside of this study (Gonzalez-Reiche *et al.*, 2020). Though remaining NYC cases show a wide  
470 distribution across all identified clades, the predominance (>80% in NYC) of such a narrowly  
471 defined set of sequences within NYC from a rare ( $\leq 20\%$ ) Western European subclade is striking.  
472 The steady increase of A2-25563 prevalence in other regions in the Eastern and Western USA  
473 (but not Midwest) through late March and early April may be consistent with emigration patterns  
474 of NYC residents who left the city during the lockdown. The dominance of the proposed A2-  
475 25563 subclade in NYC suggests either (1) a very early introduction by a single patient  
476 harboring A2-25563; or (2) multiple A2-25563 founder events; or (3) disproportionate  
477 community transmission of strains within this subclade. The first possibility may be potentially  
478 linked to expansion among cases in the New Rochelle cluster, though additional genotyping  
479 and/or metadata aggregation from those sequenced cases would be required to assess this  
480 hypothesis. The third possibility could be consistent with A2-25563 harboring differential  
481 fitness with respect to transmissibility or virulence relative to other A2 viruses.

482  
483 Future studies correlating viral genotypes with patient outcomes in larger cohorts will be  
484 necessary to determine whether any of these A2-25563 associated variants functionally influence  
485 viral transmission or disease severity. The polyphyletic pattern and convergent evolution pattern  
486 observed with NSP1 p.141\_143KSD, including clades outside of A2, raises the possibility that  
487 this variant arose multiple times and may be under positive selection. Given the small number of  
488 these variants observed in our analysis (14), larger and more statistically powered datasets will  
489 be required to evaluate this hypothesis. The VAF distribution in patients harboring het variants  
490 suggests that intrahost diversification may lead to the development of SARS-CoV-2 subclones in  
491 a subset of COVID-19 cases. The relatively large numbers of these het variants (>10)  
492 presumably arising during the course of a single infection rising to a reasonably high oligoclonal  
493 VAFs (0.1-0.2) suggests that this intrahost diversification may be rapid and also associated with  
494 positive selection.

495  
496 Our results demonstrate that distinct host transcriptional programs are activated during viral  
497 infection of the naso-/oropharynx with SARS-CoV-2. This includes upregulation of specific  
498 interferon pathway genes (*SHFL*, *IFI6*, *IFI27*, and *IFIT1*) that have been previously associated  
499 with the innate antiviral host immune response against other positive-strand RNA viruses (e.g.  
500 hepatitis C, Dengue virus). These results provide clinical relevance for recently published results  
501 from animal and cellular models of SARS-CoV-2 (Blanco-Melo *et al.*, 2020). Our analyses also  
502 implicate expression perturbations of the ACE pathway in SARS-CoV-2 host response, including  
503 *ACE2*. ACE2 is the cognate cellular receptor for SARS-CoV and SARS-CoV-2 coronaviruses,  
504 and a recently proposed drug target for SARS-CoV-2 (Monteil *et al.*, 2020). Patients presenting  
505 with COVID-19 frequently harbor comorbidities such as hypertension, diabetes mellitus, and  
506 coronary heart diseases, all which have been associated with increased disease severity (Fang *et*

507 *al.*, 2020, Ferrario *et al.*, 2005). Since these comorbidities are frequently treated with ACE  
508 pathway perturbing medications (ACEIs and ARBs), one possibility is that these medications  
509 may make patients more susceptible to SARS-CoV-2 infection.

510  
511 Though previous epidemiological studies have reported increased mortality and morbidity in  
512 COVID-19 patients with hypertension, we sought to examine this risk in the context of ACEI or  
513 ARB use (Patel and Verma, 2020). Our retrospective clinical analysis shows that ACEI/ARB  
514 exposure increased COVID-19 risk, but showed equivocal effects for morbidity and mortality  
515 when conditioning on hypertension status. A recent study reported a protective effect of  
516 ACEI/ARB exposure on mortality (Zhang *et al.*, 2020). We were able to replicate their result by  
517 conditioning on hypertension. However, this same analysis also revealed no effect, or possibly  
518 even a risk, from ACEI/ARB exposure on mechanical respiration and a strong risk of infection.  
519 Further, the association with infection was consistent regardless of conditioning on hypertension  
520 and in a direct comparison to CCB exposure, mitigating some of the concerns about confounding  
521 risk. Taken together, the results from these studies suggest caution should be taken before any  
522 clinical guidelines are put in place. Prospective clinical trials or multi-centered blinded studies  
523 may be necessary before more concrete conclusions can be drawn. For example, if some patients  
524 are more susceptible because they are already expressing high levels of *ACE2*, this could help  
525 with targeting *ACE2* in these patients as a prophylactic method. However, if the cells respond to  
526 infection with *ACE2* expression, and this leads to the cytokine storm seen in patients, then this  
527 could be used as a downstream treatment (post-infection), for when *ACE2* interacts with  
528 TMPRSS2, such as the ongoing trials with camostat mesylate (Hoffman *et al.*, 2020).

529  
530 Finally, it is notable that the majority of the testing for SARS-CoV-2 so far has relied on  
531 nasopharyngeal specimen collection, yet preliminary results here and elsewhere demonstrate  
532 SARS-CoV-2 detection from oral specimens is feasible and likely optimal (Woelfel *et al.*, 2020,  
533 Wylie *et al.*, 2020). Further studies comparing nasopharyngeal, oropharyngeal, buccal, and saliva  
534 collection approaches, as well as a comparison of different swab types, are needed. Depending  
535 on the availability of reagents and resources, as well as automation, a LAMP-based approach on  
536 such sample types could allow facilities to increase testing capabilities by orders of magnitude.  
537 Since viral pandemics can have significant, long-lasting detrimental impacts for affected  
538 countries, it is crucial to deploy methods that can track and profile cases (e.g. RNA-seq, LAMP,  
539 qRT-PCR) and provide a comprehensive view of host and viral biology. These methods can help  
540 mitigate the medical and socioeconomic harm from viral outbreaks, as well as establish  
541 protective surveillance networks that can help defend against future pandemics.

#### 542 543 **Acknowledgements**

544 We thank the Core Facilities at Weill Cornell Medicine, the Clinical Laboratories at New York  
545 Presbyterian Hospital, the Scientific Computing Unit (SCU), OneCodex, the XSEDE  
546 Supercomputing Resources and the GISAID Initiative curators and submitters (**Supp Table 10**).  
547 We also thank New England Biolabs for providing the reagents for preliminary testing of the  
548 LAMP protocols, as well as Eileen Dimalanta and Ted Davis for technical discussions. The  
549 authors wish to thank the following members of the HudsonAlpha Discovery team who  
550 supported the RNASeq experiments described in the manuscript: Colleen Cowan, John Mote,  
551 Arianna Pionzio, Melanie Robinson, and Madison Robison.

552

553 **Funding**

554 We are grateful for support from the STARR Foundation (I13-0052) the Vallee Foundation, the  
555 WorldQuant Foundation, The Pershing Square Sohn Cancer Research Alliance, Citadel, the  
556 National Institutes of Health (R01MH117406, R25EB020393, R01AI151059), the Bill and  
557 Melinda Gates Foundation (OPP1151054), the NSF (1840275), the National Center for  
558 Advancing Translational Sciences of the National Institutes of Health (UL1TR000457, CTSC),  
559 and the Alfred P. Sloan Foundation (G-2015-13964). FJS supported by the National Institute of  
560 Allergy and Infectious Diseases (1U19AI144297-01). MZ supported by T15LM007079. NPT  
561 and UOG supported by R35GM131905. NAI was supported by the National Center for  
562 Advancing Translational Sciences of the National Institutes of Health under Award Number  
563 TL1TR002386.

564

565 **Author Contributions**

566 CEM led the study design and coordination and HR led the clinical collection and validation  
567 work in the NYP CLIA laboratory, as well as with MI. Overall supervision and protocol  
568 development and implementation for the Zymo RNAClean and NEB assays (SL). DB and CMz  
569 performed the LAMP experiments to validate the assay, established a method to quantify LAMP  
570 output, and developed a protocol for clinical use of the assay. CW, DX, PR, JG, and JX assisted  
571 with sample preparation and sequencing. CeM, DD, JF, AS, JR, MM, EA, IH, DM, MI, BWL,  
572 MZ, UG, NPT, NAI, CEM performed analyses. DD, CMz, NAI, MS, BY, KR, CB coordinated  
573 and collected environmental samples. EA submitted the IRB application and helped with clinical  
574 coordination. Help and insights for analysis from EA, SL, MI, MS, LFW, ML, MC, HR, KR all  
575 led to the figures and analyses. All authors reviewed, edited, and approved the manuscript.

576

577 **Conflicts of Interest**

578 Nathan Tanner and Bradley W. Langhorst are employees at New England Biolabs.

579

580 **IRB**

581 Samples were collected and processed through the Weill Cornell Medicine Institutional Review  
582 Board (IRB) Protocol 19-11021069. Observational cohort analysis (ACEI/ARB) was done  
583 through the Columbia University IRB Protocol AAAL0601.

584

585 **Data Accessibility**

586 Viral sequences were uploaded into GISAID (Global Initiative on Sharing All Influenza Data)  
587 site (<https://www.gisaid.org>), and patient data is being deposited into dbGAP (accession  
588 pending).

589

590 **Materials and Methods**

591 *Sample Collection and Processing*

592 Patient specimens were collected with patients' consent at New York Presbyterian Hospital-  
593 Weill Cornell Medical Center (NYPH-WCMC) and then processed for qRT-PCR.

594 Nasopharyngeal (NP) swab specimens were collected using the BD Universal Viral Transport  
595 Media system (Becton, Dickinson and Company, Franklin Lakes, NJ) from symptomatic  
596 patients.

597

598 *Extraction of Viral RNA and qRT-PCR detection*

599 Total viral RNA was extracted from deactivated samples using automated nucleic acid extraction  
600 on the QIA Symphony and the DSP Virus/Pathogen Mini Kit (QIAGEN). One step reverse  
601 transcription to cDNA and real-time PCR (RT-PCR) amplification of viral targets, E (envelope)  
602 and S (spike) genes and internal control, was performed using the Rotor-Gene Q thermocycler  
603 (QIAGEN).

604

#### 605 *Twist Synthetic RNAs*

606 We used two fully synthetic RNAs made by in vitro transcription (IVT) from Twist Biosciences,  
607 which were synthesized in 5kb pieces with full viral genome coverage of SARS-CoV-2. They  
608 were sequence verified to ensure >99.9% viral genome coverage, and come as 1,000,000 copies  
609 per  $\mu\text{L}$ , 100 $\mu\text{L}$  per tube. The two controls are from Wuhan, China (MN908947.3) and  
610 Melbourne, Australia (MT007544.1). Reference sequence designs came from NCBI:  
611 <https://www.ncbi.nlm.nih.gov/nuccore/MT007544> and  
612 <https://www.ncbi.nlm.nih.gov/nuccore/MN908947.3>.

613

#### 614 *Reverse Transcriptase, quantitative real-time PCR (RT-PCR)*

615 Clinical samples were extracted as described above and then tested with qRT-PCR using primers  
616 for the E (envelope) gene, which detects all members of the lineage B of the beta-CoVs,  
617 including all SARS, SARS-like, and SARS-related viruses, and a second primer set for the S  
618 (spike) gene, which specifically detects the SARS-CoV-2 virus. The reaction also contains an  
619 internal control that served as an extraction control and a control for PCR inhibition.

620

621 Samples were annotated using qRT-PCR cycle threshold (Ct) value for SARS-CoV-2 primers.  
622 Subjects with Ct less than or equal to 18 were assigned "high viral load" label, Ct between 18  
623 and 24 were assigned "medium viral load" and Ct between 24 and 40 were assigned "low viral  
624 load" classes, with anything above Ct of 40 was classified as negative. We also predicted a  
625 combined viral load score using Ct, GloMax QuantiFluor readout from LAMP experiments and  
626 fraction of SARS-CoV-2 matching NGS reads in a sample. For this score (40-Ct), (LAMP  
627 readout) and ( $\log_{10}(\text{SARS-CoV-2 fraction} + 1\text{e-}6)$ ) were all normalized between zero and one  
628 individually, and summed together using a combination weight of 5 for Ct, 3 for LAMP and 2  
629 for NGS.

630

#### 631 *LAMP Primer Sequences*

632 Primers were designed using PrimerExplorer (v4.0), as per guidelines in Zhang *et al.*, 2020. This  
633 specifically utilized the LAMP-compatible primers for the on the COVID-19 reference genome  
634 (NCBI). LAMP's inherent specificity (using 4-6 primers vs. 2 for PCR amplification) in  
635 combination with this *in-silico* analysis revealed there is limited opportunity for cross-reactivity  
636 to allow for false-positive reporting or affect performance of the N-gene primers for SARS-CoV-  
637 2 detection (**Supp Table 5**). Overall, the primers had less than 80% homology with the vast  
638 majority of tested pathogen sequences. For any organisms where a primer hit >80% homology,  
639 only one of the primers (forward or reverse) had significant homology making an amplified  
640 product extremely unlikely. Overall, the results of this analysis predict no significant cross-  
641 reactivity or microbial interference. We also assessed the potential impact of sequence variation  
642 in circulating strains that might lead to poor amplification. In the thousands of sequences  
643 deposited in GISAID (Shu and McCauley, 2017), only one site in the priming region was  
644 observed to be polymorphic. The polymorphism (T30359C) was only observed in 106 of 6753

645 (<2%) sequences with coverage of this region. This variant overlaps the priming site of the LB  
646 primer but is not near a 3-prime end and is not anticipated to cause amplification failure. The  
647 data from Figure 1 show the use of a single-tube, dual-primer protocol, wherein both the N2 gene  
648 and E gene primers are present.

649

650 **Primer Sequences: rActin (5'-3')**

651 ACTB-F3 AGTACCCCATCGAGCACG  
652 ACTB-B3 AGCCTGGATAGCAACGTACA  
653 ACTB-FIP GAGCCACACGCAGCTCATTGTATCACCAACTGGGACGACA  
654 ACTB-BIP CTGAACCCCAAGGCCAACCGGCTGGGGTGTGAAAGGTC  
655 ACTB-LoopF TGTGGTGCCAGATTTTCTCCA  
656 ACTB-LoopB CGAGAAGATGACCCAGATCATGT

657

658 **Gene E Primer Set (5'-3')**

659 E1-F3 TGAGTACGAACTTATGTACTCAT  
660 E1-B3 TTCAGATTTTAAACACGAGAGT  
661 E1-FIP ACCACGAAAGCAAGAAAAAGAAGTTCGTTTCGGAAGAGACAG  
662 E1-BIP TTGCTAGTTACACTAGCCATCCTTAGGTTTACAAGACTCACGT  
663 E1-LB GCGCTTCGATTGTGTGCGT  
664 E1-LF CGCTATTA ACTATTAACG

665

666 **Gene N2 Primer Set (5'-3')**

667 N2-F3 ACCAGGAACTAATCAGACAAG  
668 N2-B3 GACTTGATCTTTGAAATTTGGATCT  
669 N2-FIP TTCCGAAGAACGCTGAAGCGGAACTGATTACAAACATTGGCC  
670 N2-BIP CGCATTGGCATGGAAGTCACAATTTGATGGCACCTGTGTA  
671 N2-LF GGGGGCAAATTGTGCAATTTG  
672 N2-LB CTTCGGGAACGTGGTTGACC

673

674 *The LAMP Reaction Setup*

675 For each well or Eppendorf tube, we utilized a set of six primers (above) for Gene N, the M1800  
676 LAMP Master Mix (NEB), water, and 11.5µL of the sample. The protocol is as follows:

677

- 678 1. Reagents added:
  - 679 a. 12.5 µL M1800 LAMP mix (NEB)
  - 680 b. 1-5 µL LAMP primers (Gene N or N2/E mix)
  - 681 c. 1-11.5 µL of sample
  - 682 d. Remaining volume (to 25 µl) H<sub>2</sub>O
- 683 2. Vortex, spin down;
- 684 3. Place on Thermocycler at 65°C for 30 minutes with lid at 105 °C;
- 685 4. Remove tubes, place on ice for 5 seconds;
- 686 5. Visualize over lab bench/ice/paper.

687

688 *Oropharyngeal Lysate LAMP Run*

689 Nasopharyngeal and/or oropharyngeal swab samples from 201 patients were collected using a  
690 dry cotton swab (cliniswab DS, Aptaca Spa (Italy), #2170/SG/CS). Crude extraction was

691 performed according to pending unpublished European patent [No. 20168 593.0]. In summary,  
692 the dry swab was transferred to a 15 ml falcon (Greiner Bio-one, #188.271), filled with 0.5 ml  
693 saline solution, and shaken vigorously for 30 min. Afterwards the liquid is transferred to a screw-  
694 cap (Sarstedt, #72.692.005). 10 $\mu$ l of the crude RNA extract was added to 12.5  $\mu$ l 2x NEB LAMP  
695 master mix (#M1800L), 2.5  $\mu$ l of water, 1  $\mu$ l 25x primer master mix Gen N, and was incubated  
696 at 65 °C for 30-40 min. A sample of a patient whom tested positive using an approved qRT-PCR  
697 test (sample #1123) was used as an internal control. The read-out was performed visually by  
698 color change from pink to yellow or orange.

699  
700 The RNA isolation was performed with 50  $\mu$ l of the crude extract on the QIA Symphony with the  
701 DSP Virus/Pathogen Kit (Qiagen, #937055). The RT-PCR was performed using 5  $\mu$ l of 85  $\mu$ l  
702 eluate with TIB MolBiol Lightmix® MODULAR SARS AND WUHAN CoV E-Gene Kit.  
703 Analysis was done with the LightCycler(R) 480 II software and calculated CP values were used  
704 for statics and graphical analysis. The Standard curve with a synthetic RNA control (Twist  
705 Bioscience, #MT007544.1) was generated using the lamp assay and in parallel with the qRT-  
706 PCR. The control RNA was diluted serially 10-fold with water and the absolute copy number,  
707 ranging from 10<sup>5</sup> to 10<sup>-5</sup> was analyzed. Gel electrophoresis after visual read-out of the LAMP  
708 assay was done by loading 5  $\mu$ l of the lamp reaction with 5  $\mu$ l 2x loading dye on a 1.5 % agarose  
709 (Seakem LE Agarose, Lonza #50004) together with 5  $\mu$ l of a 1kB DNA Ladder (Roche). The  
710 electrophoresis was performed at 90 V constant.

711  
712 The photometric read-out of the standard curve was performed in a plate reader. To this end the  
713 lamp reaction was transferred into a 96-well V-shaped cell-culture plate (Greiner Bio-One,  
714 #651180). After measuring the absorbances at 432nm and 560nm, the ratio abs(432)/abs(560)  
715 was calculated. The relative absorbance abs(432)/abs(560) was calculated by subtraction of the  
716 negative (water) control from all samples (including the negative control). All values above a  
717 threshold of 0.1 are considered as a positive assay read-out and are marked "+", all other  
718 values are negative and are marked "-". Statistical and graphical analysis were performed with  
719 GraphPad Prism 8.0.4.

720  
721 *Light Intensity and Data Processing*

722 Completed reactions were analyzed with the Promega GloMax Explorer (Promega GM3500)  
723 fluorometer using the QuantiFluor ONE dsDNA system (Promega E4871). This system recorded  
724 fluorometric readout from each well using an emission filter of 500-550nm, an excitation filter  
725 set at blue 475 nm, and a high sensitivity setting on the Glomax software. Values were then  
726 tabulated and compared with controls (positive and negative). The intensity threshold of 2.5x  
727 negative control was used as the threshold for positive detection.

728  
729 *DNase treatment, rRNA depletion, and RNaseq library construction*

730 For library preparation, all samples' total nucleic acid (TNA) were treated with DNase 1 (Zymo  
731 Research, Catalog # E1010), which cuts both double-stranded and single-stranded DNA. Post-  
732 DNase digested samples were then put into the NEBNext rRNA depletion v2  
733 (Human/Mouse/Rat), Ultra II Directional RNA (10ng), and Unique Dual Index Primer Pairs  
734 were used following the vendor protocols from New England Biolabs (except for the first  
735 flowcell, see supplemental figures). Kits were supplied from a single manufacturer lot.  
736 Completed libraries were quantified by Qubit or equivalent and run on a Bioanalyzer or



737 equivalent for size determination. Libraries were pooled and sent to the WCM Genomics Core or  
738 HudsonAlpha for final quantification by Qubit fluorometer (ThermoFisher Scientific),  
739 TapeStation 2200 (Agilent), and QRT-PCR using the Kapa Biosystems Illumina library  
740 quantification kit.

#### 741 *Taxonomic Classification of Sequence Data*

742 All complete genome or chromosome level assemblies from RefSeq database for archaea,  
743 bacteria, protozoa, fungi, human and viruses including SARS-CoV and SARS-CoV-2 genomes  
744 were downloaded and used for building a classification database for Kraken2 (k=35,  $\ell=31$ )  
745 (O'Leary *et al.*, 2016; Wood *et al.*, 2019).

746  
747  
748 To get an approximation for the positive and negative classification rate, the BMAP random-  
749 reads script was used to simulate 10 million 150bp paired-end Illumina reads from the database  
750 sequences (Segata *et al.*, 2016). For the negative test all sequences in the database excluding  
751 SARS-CoV and SARS-CoV-2 genome were removed from the sequences and the simulated  
752 reads were mapped with the Kraken2 database (Supp Table 1).

753  
754 For the positive test, the same process was repeated using only SARS-CoV-2 genome (**Supp.**  
755 **Table 2**). Positive results show >99% of SARS-CoV-2 reads uniquely map to either SARS-CoV  
756 or SARS-CoV-2, with the remaining 1% are ambiguous, potentially matching multiple taxa  
757 (**Supp. Table 2**). All sequences were classified using the Kraken2 database. To remove the  
758 potential contamination of reads that are homologous across multiple species we used Kraken2  
759 outputs to filter sequences to either human (uniquely matching *Homo sapiens* and no other taxon  
760 in our database), SARS-CoV-2 (either matching SARS-CoV or SARS-CoV-2 due to homology  
761 between these two viruses), and remaining reads that may be coming from unclassified, archaeal,  
762 bacterial, viral, fungal, protozoan or ambiguously mapping reads to human or SARS-CoV (Li,  
763 2015).

764  
765 Using kraken2 classifications common respiratory pathogens were identified in clinical samples.  
766 Any SARS-CoV-2 negative sample with >0.01% relative abundance (normalized after the  
767 exclusion of any human, SARS-CoV-2 and uncharacterized reads) for presence of viral  
768 pathogens were classified as potential unrelated viral infection (**Supp. Figure 7**). These samples  
769 were used as controls during specific differential expression comparisons to compare the  
770 common effects of viral infections on host cells from SARS-CoV-2 infection.

#### 771 *Viral Genome Assembly*

772 Reads unambiguously mapping to SARS-CoV or SARS-CoV-2 were aligned to the Wuhan-Hu-1  
773 (Genbank accession MN908947.3) reference using bwa mem (Li, 2013). Variants were called  
774 using iVar, and pileups and consensus sequences were generated using samtools (Li *et al.*, 2009;  
775 Grubaugh *et al.*, 2019; Greenfield *et al.*, 2020). Any sample with >99% coverage above 10x  
776 depth for SARS-CoV-2 genome were taken as reliable samples, which resulted in 155 samples  
777 (146 positive, 9 negative). 155 clinical samples were compared to 7,797 SARS-CoV-2 sequences  
778 from GISAID (as of April 23, 2020) (9, 10). All sequence filtering, alignments, phylogenetic  
779 inference, temporal ordering of sequences and geographic reconstruction of likely transmission  
780 events were done using Nextstrain (Kato and Standley, 2013; Sagulenko *et al.*, 2018; Hadfield  
781 *et al.*, 2018). Sequence identity and coverage metrics were calculated using Mview (Brown *et*

783 *al.*, 1998). Phylogenetic trees were created using Nextstrain's augur as described above, and  
784 visualized using the ggtree package in R (Yu, 2020).

785

#### 786 *Viral variant calling and allelic fraction estimation*

787 Full-length viral consensus sequences were aligned to the Wuhan-Hu-1 reference using bwa  
788 mem (Li, 2013) with default settings. Variants were identified by enumerating the coordinates  
789 and query / reference subsequences associated with mismatches (SNV) and gaps in the query  
790 (deletion) and reference (insertions) using R/Bioconductor (GenomicRanges, Rsamtools,  
791 Biostrings packages) and Imielinski lab gChain packages (<https://github.com/mskilab/gChain>).  
792 Exhaustive variant calling on read alignments was additionally performed using bcftools mpileup  
793 and call, with variant read support (VAF, alternate allele count) enumerated with the  
794 R/Bioconductor Rsamtools package.

795

796 For each variant, a posterior distribution of variant allele fraction (VAF) was computed using a  
797 beta distribution with shape parameters  $\theta$  comprising reference and alternate allele counts and  
798 pseudo count of 0.5. Variants were classified as het (heterogenous) if  $P(VAF > 0.05 \wedge VAF <$   
799  $0.95 | \theta) > 0.95$ . For a given specimen, posterior VAF distributions of  $k$  heterogenous variants  
800 were then integrated using a histogram density estimator by summing the posterior VAF density  
801 across all variants at each point of a grid of 100 points evenly spaced in the (0,1) line. This  
802 (unnormalized) mixture density was visualized alongside the individual VAF densities as an  
803 estimate of the probability density of a putative viral subclone.

804

#### 805 *SARS-CoV-2 genome wide association study*

806 Population reference and alternate allele counts were computed for all SNV and indel variant  
807 calls detected in at least three cases in the A2 clade subset of our NYPH-WCMC sequences (123  
808 cases, see above). Similar counts were computed across 537 New York and 7269 non-New York  
809 sequences that were obtained from GISAID and incorporated into the Nextstrain tree (see  
810 above). A one-tailed Fisher's Exact test was used to compare allele counts in NYPH-WCMC vs.  
811 non-New York GISAID cases. All variants with FDR < 0.1 were reported as significant.  
812 Significant variants from the initial analysis were replicated in a similar comparison of New  
813 York vs non-New York GISAID sequences.

814

#### 815 *Cell Deconvolution Analysis*

816 Bulk RNAseq count data was deconvoluted into cell composition matrices using the MUSIC  
817 algorithm (Wang *et al.*, 2019) on a reference single cell RNAseq data from upper respiratory  
818 epithelium obtained from nasal brushes and upper airway and lung cells ((Vieira *et al.*, 2019;  
819 Lukassen *et al.* 2020).

820

#### 821 *Human Transcriptome Analysis*

822 The reads that mapped unambiguously to the human reference genome via Kraken2 were used to  
823 detect the host transcriptional response to the virus. Reads matching *Homo sapiens* were trimmed  
824 with TrimGalore, aligned with STAR (v2.6.1d) to the human reference build GRCh38 and the  
825 GENCODE v33 transcriptome reference, gene expression was quantified using featureCounts,  
826 stringTie and salmon using the nf-core RNAseq pipeline (Pertea *et al.*, 2015; Malinen *et al.*,  
827 2005; Johnson *et al.*, 2007; Robinson *et al.*, 2010; Naccache *et al.*, 2014; Zamani *et al.*, 2017;  
828 Ewels *et al.*, 2019). Sample QC was reported using fastqc, RSeQC, qualimap, dupradar, Preseq

829 and MultiQC (Okonechnikov *et al.*, 2016; Andrews, 2015; Ewesl *et al.*, 2016; Sayols *et al.*,  
830 2016; Wang *et al.*, 2012). Samples that had more than 10 million human mapped reads were  
831 used for differential expression analysis. Reads, as reported by featureCounts, were normalized  
832 using variance-stabilizing transform (vst) in DESeq2 package in R for visualization purposes in  
833 log-scale (Love *et al.*, 2014). Limma voom and DESeq2 were used to call differential expression  
834 with either Positive cases vs Negative, or viral load (High/Medium/Low/None excluding any  
835 samples with evidence of other viral infections) as reported by either qRT-PCR cycle threshold  
836 (Ct) values, or using the inverted normalized Ct value as continuous response for viral levels,  
837 where Ct of 15 is 1.0 and Ct of >40 is taken as 0 (Law *et al.*, 2014). Genes with BH-adjusted p-  
838 value < 0.01 and absolute log<sub>2</sub> fold-change greater than 0.58 (at least 50% change in either  
839 direction) were taken as significantly differentially regulated (Benjamini and Hochberg, 1995).  
840 The same approaches were repeated correcting for potential confounders in our data in two ways.  
841 In the first correction ciliated cell fraction (as predicted by MUSIC) was added as another  
842 covariate to our model. For the second correction SVA was run on the data and the resulting two  
843 surrogate variables were included in a multivariate model (Leek *et al.*, 2012). The complete gene  
844 list for all comparisons are given in Supp Table 3. Resulting gene sets were ranked using log<sub>2</sub>  
845 fold-change values within each comparison and put into GSEA to calculate gene set enrichment  
846 for molecular signatures database (MSigDB), MGI Mammalian Phenotypes database and  
847 ENCODE transcription factor binding sets (Liberzon *et al.*, 2011; Subramanian *et al.*, 2005;  
848 Kuleshov *et al.* 2016; Sergushichev, 2016; Smith *et al.*, 2018). Any signature with adjusted p-  
849 value < 0.01 and absolute normalized enrichment score (NES) >= 1.5 were reported (Supp Table  
850 3).

#### 851 *Cross-reactivity Analysis*

852 Primers were compared with a list of sequences from organism from the same genetic family as  
853 SARS-CoV-2 and other high-priority organisms listed in the United States Food and Drug  
854 Administration's Emergency Use Authorization Template  
855 (<https://www.fda.gov/media/135900/download>). Using the sequence names in the EUA template,  
856 the NCBI taxonomy database was queried to find the highest quality representative sequences for  
857 more detailed analysis. Primers were compared to this database using Blast 2.8.1 and the  
858 following parameters (word size: 7, match score: 2, mismatch score: -3, gap open cost: 5, gap  
859 extend cost: 2). Up to 1000 hits with e-value > 10 were reported.

#### 861 *Inclusivity Analysis*

862 Unique, full-length, human-sample sequences were downloaded from the GISAID web interface.  
863 These sequences were aligned to NC\_045512.2 (Wuhan SARS-CoV-2) using minimap2 -x asm5  
864 and visually inspected using IGV 2.8.0 with allele frequency threshold set to 0.01.

#### 866 *ACE Inhibitor/Angiotensin Receptor Blocker Cohort Analysis*

867 We compared usage of ACE inhibitors (ACEI) and angiotensin receptor blockers (ARB) in an  
868 observational cohort analysis of 8,856 patients with suspected SARS-CoV-2 infection (4,829 of  
869 which tested positive). ACEIs and ARBs are commonly taken continuously for several years  
870 (Bonarjee *et al.*, 2001). We defined a cohort of ACEI/ARB exposed patients as those that have  
871 an ACEI/ARB prescription order sometime after January 1st, 2019. We compared the frequency  
872 of ACEI/ARB exposure in three cohort comparisons:

874 i. SARS-CoV-2 tested positive patients versus SARS-CoV-2 tested negative patients,

- 875           ii. SARS-CoV-2 positive patients who require mechanical ventilation versus those who  
876 did not, and  
877           iii. SARS-CoV-2 positive patient survival versus death.  
878 In addition, we perform one post-hoc comparison to evaluate the individual effects of particular  
879 ACEIs or ARBs among SARS-CoV-2 positive patients who are exposed.

880

#### 881 *Cohort and Data Source*

882 Our cohort data for SARS-CoV-2 suspected patients is extracted from the electronic health  
883 records at NYPH-CUIMC. We used data collected starting on March 10th, 2020 through April  
884 20th, 2020. In addition, we used data from 90,989 patients, who were not tested for SARS-CoV-  
885 2 infection and were 18 years or older, with available electronic health records from January 1st,  
886 2019 through September 24th, 2019 to represent a comparison population of patients. In both  
887 cases, data extracted included disease diagnoses, laboratory measurements, medication and  
888 pharmacy orders, and patient demographics. We derived mortality from a death note filed by a  
889 resident or primary provider that records the date and time of death. Intubation was used as an  
890 intermediary endpoint and is a proxy for a patient requiring mechanical respiration. We used  
891 note types that were developed for patients with SARS-CoV-2 infection to record that this  
892 procedure was completed. We validated outcome data derived from notes against the patient's  
893 medical record using manual review.

894

#### 895 *Experimental Statistical Methods*

896 We conducted univariate analysis of the frequency differences of ACEI/ARB exposure and  
897 multivariate regression analysis to account for established risk factors for SARS-CoV-2  
898 outcomes (i.e. age, sex, baseline IL-6 upon admission, and comorbidities) (Zhou et al., 2020).  
899 We use logistic regression for analysis for comparing cohort (i) and performed survival analysis  
900 using a Cox proportional hazards model for cohort comparisons (ii) and (iii). We cannot  
901 determine from our data the date of infection. For the study start date for the patient, we use the  
902 date of testing positive or the first date a diagnosed patient showed symptoms, whichever came  
903 first, minus seven days. Quantitative variables (i.e., age and the first IL-6 measurement) are  
904 linearly scaled to [0,1] to facilitate comparison of model coefficients to those of dichotomous  
905 variables. Prior to conducting our multivariate analysis, we evaluated and removed correlated  
906 covariates. Finally, we repeated the previously described analysis for a cohort restricted to  
907 patients with a documented history of clinically diagnosed hypertension. We include all model  
908 results in the supplemental tables.

909

#### 910 *Comorbidity Definitions*

911 Risk factors were assigned using OMOP CDM concept IDs 316866 for “Hypertensive Disorder”,  
912 317576 for “Coronary Arteriosclerosis”, 201820 for “Diabetes mellitus”, and 437525 for  
913 “Overweight”. In each case, the concept ID and all descendant concepts were used to define the  
914 risk factor phenotype, and individuals were assigned the phenotype if they were assigned any of  
915 the codes.

916

#### 917 *Statistical and visualization software*

918 All electronic health data analyses were performed in Python 3.7 and all models were fit using R  
919 3.6.3. Survival analyses (Cox regressions and survival curves) were performed with the survival  
920 package for R, version 3.1-12. Additional statistical analyses, processing, transformation, and

921 visualization of genomic data were completed in R / Bioconductor ('Rsamtools',  
922 'GenomicRanges', 'Biostrings') and additional Imielinski Lab R packages ('gTrack', 'gChain',  
923 'gUtils', 'RSeqLib') available at <https://github.com/mskilab>. Visualization of phylogenies was  
924 completed using Auspice and the 'ggtree' and 'ape' libraries for R.

925  
926

## 927 **References**

928 Agata J, Ura N, Yoshida H, Shinshi Y, Sasaki H, Hyakkoku M, Taniguchi S, Shimamoto K.  
929 Olmesartan Is an Angiotensin II Receptor Blocker with an Inhibitory Effect on Angiotensin-  
930 Converting Enzyme. *Hypertension Research* 29, 865-874 (2006).

931

932 Ajioka RS, Phillips JD, Kushner JP. Biosynthesis of heme in mammals. *Biochim Biophys Acta*.  
933 2006 Jul;1763(7):723-36

934

935 Amanat F, Nguyen T, Chromikova V, Strohmeier S, Stadlbauer D, Javier A, Jiang K, Asthagiri-  
936 Arunkumar G, Polanco J, Bermudez-Gonzalez M, Caplivski D, Cheng A, Kedzierska K,  
937 Vapalahti O, Hepojoki J, Simon V, Krammer F. A serological assay to detect SARS-CoV-2  
938 seroconversion in humans. *MedRxiv*. 2020 Mar 18.

939 <https://www.medrxiv.org/content/10.1101/2020.03.17.20037713v1>

940

941 Andrews S. FASTQC A Quality Control tool for High Throughput Sequence Data. *Babraham*  
942 *Inst.* (2015).

943

944 Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful  
945 Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* (1995)

946

947 Broughton JP, Deng X, Yu G, Fasching CL, Singh J, Streithorst J, Granados A, Sotomayor-  
948 Gonzalez A, Zorn K, Gopez A, Hsu E, Gu W, Miller S, Pan C, Guevara H, Wadford D, Chen J,  
949 Chiu CY. Rapid Detection of 2019 Novel Coronavirus SARS-CoV-2 Using a CRISPR-based  
950 DETECTR Lateral Flow Assay. 2020 Mar; 10: 1-27.

951 <https://medrxiv.org/cgi/content/short/2020.03.06.20032334v1>

952

953 Brown, N.P., Leroy C., Sander C. (1998). MView: A Web compatible database search or  
954 multiple alignment viewer. *Bioinformatics*. 14 (4):380-381.

955

956 Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, Cox AJ, Kruglyak S,  
957 Saunders CT. Manta: rapid detection of structural variants and indels for germline and cancer  
958 sequencing applications, *Bioinformatics*, Volume 32, Issue 8, 15 April 2016, Pages 1220–1222

959

960 Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, Marth GT, Quinlan AR,  
961 Hall IM. SpeedSeq: Ultra-fast personal genome analysis and interpretation. *Nature Methods*  
962 12(2015): 966–968.

963

964 Danko D, Bezdan D, Afshinnkoo E, et al...Mason CE. Global Genetic Cartography of Urban  
965 Metagenomes and Anti-Microbial Resistance

966 <https://www.biorxiv.org/content/10.1101/724526v1>

967  
968 Deng X, Achari A, Federman S, Yu G, Somasekar S, et al...Chiu CY. Metagenomic sequencing  
969 with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nature*  
970 *Microbiology*. 2020 Mar; 5(3): 443–454.  
971  
972 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time.  
973 *Lancet Infect Dis*. 2020 Feb 4.  
974  
975 Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID’s innovative contribution to  
976 global health. *Glob. Challenges* (2017)  
977  
978 Ewels P, Magnusson M, Lundin S, Källner M. MultiQC: Summarize analysis results for multiple  
979 tools and samples in a single report. *Bioinformatics* (2016)  
980  
981 Ewels PA et al., nf-core: Community curated bioinformatics pipelines. *Nature* 2020 Feb 13.  
982  
983 Farek J, Hughes D, Mansfield A, Krasheninina O, Nasser W, Sedlazeck FJ, Khan Z, Venner E,  
984 Metcalf G, Boerwinkle E, Muzny DM, Gibbs RA, Salerno W. xAtlas: Scalable small variant  
985 calling across heterogeneous next-generation sequencing experiments. bioRxiv 295071; doi:  
986 <https://doi.org/10.1101/295071>  
987  
988 Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at  
989 increased risk for COVID-19 infection? *Lancet Respir Med*. 2020 Apr;8(4):e21.  
990  
991 Fauver JR, et al., Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United  
992 States. *Cell* 2020.  
993  
994 Gardy J, Loman NJ, Rambaut A. Real-time digital pathogen surveillance - the time is now.  
995 *Genome Biology*. 2015 Jul 30;16(1):155.  
996  
997 Gonzalez-Reiche A, Hernandez MH, Sullivan M, Ciferri B, Alshammary H, Obla A, Fabre S,  
998 Kleiner G, Polanco J, Khan Z, Albuquerque B, Guchte A, Dutta J, Francoeur N, Salom B,  
999 Oussenko I, Deikus G, Soto J, Sridhar SH, Wang Y, Twyman K, Kasarskis A, Altman DR,  
1000 Smith M, Sebra R, Aberg J, Krammer F, Garcia-Sarstre A, Luksza M, Patel G, Paniz-Mondolfi  
1001 A, Gitman M, Sordillo EM, Simon V, Bakel H. Introductions and early spread of SARS-CoV-2  
1002 in the New York City area. MedAxiv. 2020. Apr 11.  
1003 <https://www.medrxiv.org/content/10.1101/2020.04.08.20056929v1>  
1004  
1005 Guan WJ, Ni Zy, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li  
1006 LJ, Zeng G, Yuen KY, Chen RC, Tang CL, Wang T, Chen PY, Xiang J, Li SY, Wang JL, Liang  
1007 ZJ, Peng YX, Wei L, Liu Y, Hu YH, Peng P, Wang JM, Liu JY, Chen Z, Li G, Zheng ZJ, Qiu  
1008 SQ, Luo J, Ye CJ, Zhu SY, Zhong NS, China Medical Treatment Expert Group for Covid-19.  
1009 Clinical Characteristics of Coronavirus Disease 2019 in China. *NEJM*. 2020 Feb 28.  
1010  
1011 Greenfield N, SARS-CoV-2 variant calling and consensus assembly pipeline. GitHub (2020),  
1012 (available at <https://github.com/onecodex/sars-cov-2/>).

1013  
1014 Grubaugh ND et al., An amplicon-based sequencing framework for accurately measuring  
1015 intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* (2019)  
1016  
1017 Hadfield J, Megill C, Bell S, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, and  
1018 Neher RA. 2018. Nextstrain: Real-Time Tracking of Pathogen Evolution. *Bioinformatics* 34  
1019 (23): 4121–23.  
1020  
1021 He F, Deng Y, Li W. Coronavirus Disease 2019 (COVID-19): What we know? *J Med Virol.*  
1022 2020 Mar 14.  
1023  
1024 Hoffmann M, Weber HK, Schroeder Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G,  
1025 Wu N, Nitsche A, Müller MA, Drosten C, Pöhlmann S. SARS-CoV-2 Cell Entry Depends on  
1026 ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell.* 2020. Mar  
1027 5.  
1028  
1029 Jeffares DC, Jolly C, Hoti M, et al. Transient structural variations have strong effects on  
1030 quantitative traits and reproductive isolation in fission yeast. *Nature Communications.* 2017  
1031 Jan;8:14061.  
1032  
1033 Jenkins C and Orsburn B. In silico approach to accelerate the development of mass  
1034 spectrometry-based proteomics methods for detection of viral proteins: Application to COVID-  
1035 19. 2020 Mar 10. <https://www.biorxiv.org/content/10.1101/2020.03.08.980383v1>  
1036  
1037 Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using  
1038 empirical Bayes methods. *Biostatistics* (2007).  
1039  
1040 Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements  
1041 in performance and usability. *Mol. Biol. Evol.* (2013).  
1042  
1043  
1044 Konrad R, Eberle U, Dangel A, Treis B, Berger A, Bengs K, Fingerle V, Liebl B, Ackermann N,  
1045 Sing A. Rapid Establishment of Laboratory Diagnostics for the Novel Coronavirus SARS-CoV-2  
1046 in Bavaria, Germany, February 2020. *Eurosurveillance.* 2020 Mar 5.  
1047  
1048 Kim D, Lee J, Yang J, Kim JW, Kim N, Chang H. The architecture of SARS-CoV-2  
1049 transcriptome. *BioRxiv* <https://www.biorxiv.org/content/10.1101/2020.03.12.988865v2>  
1050  
1051 Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016  
1052 update. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkw377  
1053  
1054 Lan L, Xu D, Ye G, Xia C, Wang S, Li Y, Xu H. Positive RT-PCR Test Results in Patients  
1055 Recovered from COVID-19. *JAMA* 2020 Feb 27.  
1056  
1057 Langmead B, and Salzberg S. 2012. Fast Gapped-Read Alignment with Bowtie 2. *Nature*  
1058 *Methods* 9 (4): 357–59.

1059  
1060 Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J.  
1061 The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported  
1062 Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*. 2020 Mar 10.  
1063 M20-0504.  
1064  
1065 Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: Precision weights unlock linear model  
1066 analysis tools for RNA-seq read counts. *Genome Biol.* (2014). doi:10.1186/gb-2014-15-2-r29  
1067  
1068 Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural  
1069 variant discovery. *Genome Biol.* 2014;15(6):R84.  
1070  
1071 Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The SVA package for  
1072 removing batch effects and other unwanted variation in high-throughput experiments.  
1073 *Bioinformatics* (2012). doi:10.1093/bioinformatics/bts034  
1074  
1075 Li H. A statistical framework for SNP calling, mutation discovery, association mapping and  
1076 population genetical parameter estimation from sequencing data. *Bioinformatics*.  
1077 2011;27(21):2987–2993.  
1078  
1079 Li H, Seqtk: Toolkit for processing sequences in FASTA/Q formats. GitHub (2015).  
1080  
1081 Li H, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv  
1082 Prepr. arXiv (2013), doi:arXiv:1303.3997 [q-bio.GN].  
1083  
1084 Li H et al., The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (2009).  
1085  
1086 Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. Substantial undocumented infection  
1087 facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*. 2020 Mar 16.  
1088  
1089 Liberzon A et al., Molecular signatures database (MSigDB) 3.0. *Bioinformatics* (2011).  
1090  
1091 Liu R, Han H, Liu F, Lv Z, Wu K, Liu Y, Feng Y, Zhu C. Positive Rate of RT-PCR Detection of  
1092 SARS-COV-2 Infection of 4880 Cases from One Hospital in Wuhan, China, from Jan to Feb  
1093 2020. *Clinica Chimica Acta* 505 (2020) 172-175.  
1094  
1095 Lokugamag KG, Hage A, Schindewolf C, Rajsbaum R, Menachery VD. SARS-CoV-2 is  
1096 sensitive to type 1 interferon pretreatment. *BioRxiv*  
1097 <https://www.biorxiv.org/content/10.1101/2020.03.07.982264v3>  
1098  
1099 Love MI et al., Moderated estimation of fold change and dispersion for RNA-seq data with  
1100 DESeq2. *Genome Biol.* (2014), doi:10.1186/s13059-014-0550-8.  
1101  
1102 Lu J et al., Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* 2020  
1103



- 1104 Lukassen, S. et al. SARS-CoV-2 receptor ACE 2 and TMPRSS 2 are primarily expressed in  
1105 bronchial transient secretory cells . *EMBO J.* (2020). doi:10.15252/embj.20105114  
1106
- 1107 Malinen E et al., Analysis of the fecal microbiota of irritable bowel syndrome patients and  
1108 healthy controls with real-time PCR. *Am. J. Gastroenterol.* 100, 373–382 (2005).  
1109
- 1110 Metsky HC, Freije CA, Kosoko-Thoroddsen TF, Sabeti PC, Myhrvold C. CRISPR-based  
1111 surveillance for COVID-19 using genomically-comprehensive machine learning design. 2020  
1112 Mar 2. *BioRxiv*. <https://www.biorxiv.org/content/10.1101/2020.02.26.967026v2>.  
1113
- 1114 Meyers L, Dien Bard J, Galvin B, Nawrocki J, Niesters HGM, Stellrecht KA, St George K, Daly  
1115 JA, Blaschke AJ, Robinson C, Wang H, Cook CV, Hassan F, Dominguez SR, Pretty K,  
1116 Naccache S, Olin KE, Althouse BM, Jones JD, Ginocchio CC, Poritz MA, Leber A, Selvarangan  
1117 R. Enterovirus D68 Outbreak Detection Through A Syndromic Disease Epidemiology Network.  
1118 *J Clin Virol*, 124, 104262, 2020 Mar.  
1119
- 1120 Monteil V, Kwon H, Prado P, Hagelkrüys A, Wimmer RA, Stahl M, Leopoldi A,  
1121 Garreta E, Hurtado Del Pozo C, Prosper F, Romero JP, Wirnsberger G, Zhang H,  
1122 Slutsky AS, Conder R, Montserrat N, Mirazimi A, Penninger JM. Inhibition of  
1123 SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble  
1124 Human ACE2. *Cell*. 2020 Apr 17.  
1125
- 1126 Naccache SN et al., A cloud-compatible bioinformatics pipeline for ultrarapid pathogen  
1127 identification from next-generation sequencing of clinical samples. *Genome Res.* (2014),  
1128
- 1129 Narayanan K, Ramirez SI, Lokugamage KG, Makino S. Coronavirus nonstructural  
1130 protein 1: Common and distinct functions in the regulation of host and viral gene  
1131 expression. *Virus Res.* 2015 Apr 16;202:89-100.  
1132
- 1133 Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: Advanced multi-sample quality  
1134 control for high-throughput sequencing data. *Bioinformatics* (2016),  
1135
- 1136 Ong SWX, Tan YK, Chia PY, Lee TH, Ng OT, Wong MSY, Marimuthu K. Air, Surface  
1137 Environmental, and Personal Protective Equipment Contamination by Severe Acute Respiratory  
1138 Syndrome Coronavirus 2 (SARS-CoV-2) From a Symptomatic Patient. *JAMA*. 2020 Mar 4.  
1139
- 1140 O’Leary NA et al., Reference sequence (RefSeq) database at NCBI: Current status, taxonomic  
1141 expansion, and functional annotation. *Nucleic Acids Res.* (2016).  
1142
- 1143 Oudshoorn D, van Boheemen S, Sánchez-Aparicio MT, Rajsbaum R, García-Sastre A, Versteeg  
1144 GA. HERC6 is the main E3 ligase for global ISG15 conjugation in mouse cells. *PLoS One*.  
1145 2012;7(1):e29870.  
1146
- 1147 Patel AB, Verma A. COVID-19 and Angiotensin-Converting Enzyme Inhibitors and Angiotensin  
1148 Receptor Blockers. What is the evidence? *JAMA* 2020 Mar 24.  
1149

- 1150 Peci A, Winter AL, King EC, Blair J, Gubbay JB. Performance of rapid influenza diagnostic  
1151 testing in outbreak settings. *J Clin Microbiol.* 2014 Dec;52(12):4309-17.  
1152
- 1153 Perteua M et al., StringTie enables improved reconstruction of a transcriptome from RNA-seq  
1154 reads. *Nat. Biotechnol.* (2015)  
1155
- 1156 Quilty B, Clifford S, Flasche S, Eggo RM. Effectiveness of airport screening at detecting  
1157 travellers infected with 2019-nCoV. 2020. Feb 02. *MedRxiv.*  
1158 <https://www.medrxiv.org/content/10.1101/2020.01.31.20019265v1>  
1159
- 1160 Robinson MD, Oshlack A, A scaling normalization method for differential expression analysis of  
1161 RNA-seq data. *Genome Biol.* (2010).  
1162
- 1163 Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant  
1164 discovery by integrated paired-end and split-read analysis. *Bioinformatics*, Volume 28, Issue 18,  
1165 15 September 2012, Pages i333–i339.  
1166
- 1167 Robinson, James T., Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander,  
1168 Gad Getz, and Jill P. Mesirov. 2011. Integrative Genomics Viewer. *Nature Biotechnology* 29 (1):  
1169 24–26.  
1170
- 1171 Rodriguez-Santana J, Burchard EG, Seibold MA. Type 2 and interferon inflammation strongly  
1172 regulate SARS-CoV-2 related gene expression in the airway epithelium. *Biorxiv.* 2020. Apr 10.  
1173 <https://www.biorxiv.org/content/10.1101/2020.04.09.034454v1>  
1174
- 1175 Sagulenko P, Puller V, Neher RA, TreeTime: Maximum-likelihood phylodynamic analysis.  
1176 *Virus Evol.* (2018), doi:10.1093/ve/vex042.  
1177
- 1178 Sayols S, Scherzinger D, Klein H. dupRadar: A Bioconductor package for the assessment of  
1179 PCR artifacts in RNA-Seq data. *BMC Bioinformatics* (2016), doi:10.1186/s12859-016-1276-2.  
1180
- 1181 Sergushichev A, An algorithm for fast preranked gene set enrichment analysis using cumulative  
1182 statistic calculation. *bioRxiv* (2016), doi:10.1101/060012.  
1183
- 1184 Santarpia JL, Rivera DN, Herrera V, Morwitzer MJ, Creager H, Santarpia GW,  
1185 Crown KK, Brett-Major DM, Schnaubelt E, Broadhurst MJ, Lawler JV, Reid SP, Lowe JJ.  
1186 Transmission Potential of SARS-CoV-2 in Viral Shedding Observed at the University of  
1187 Nebraska Medical Center. 2020. Mar 26. *MedRxiv.*  
1188 <https://www.medrxiv.org/content/10.1101/2020.03.23.20039446v2>  
1189
- 1190 Schmid-Burgk JL Li D, Feldman D, Ślabicki M, Borrajo J, Strecker J, Cleary B, Regev A,  
1191 Zhang F. LAMP-Seq: Population-Scale COVID-19 Diagnostics Using a Compressed  
1192 Barcode Space. *BioRxiv.* <https://www.biorxiv.org/content/10.1101/2020.04.06.025635v1>  
1193
- 1194 Segata N et al., BBMap short-read aligner, and other bioinformatics tools. *Bioinformatics* (2016).  
1195

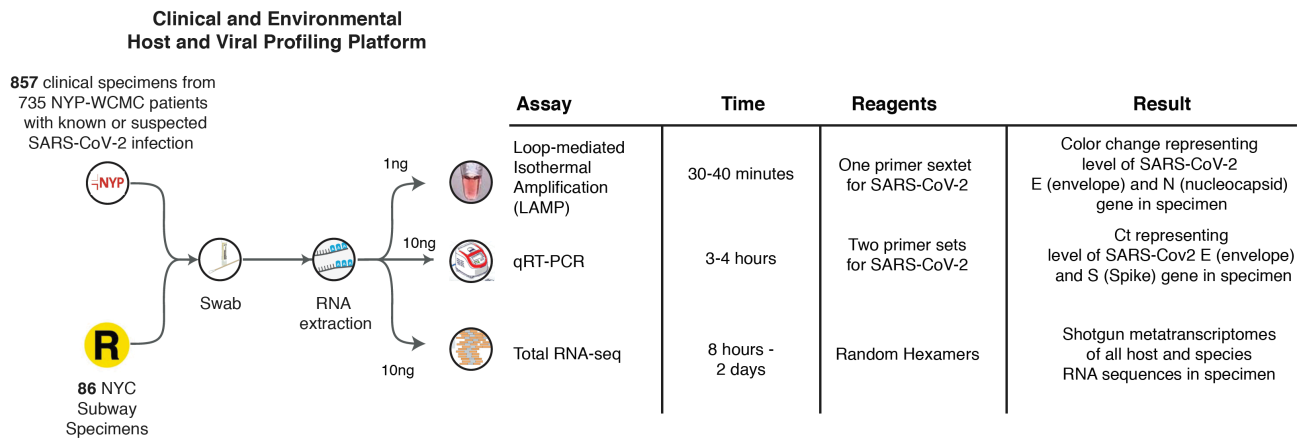
1196 Sajuthi SP, DeFord P, Jackson ND, Montgomery MT, Everman JL, Rios CL, Pruesse E, Nolin  
1197 JD, Plender EG, Wechsler ME, Mak A, Eng C, Salazar S, Medina V, Wohlford EM, Huntsman  
1198 S, Nickerson D, Germer S, Zody MC, Abecasis G, Kang H, Rice KM, Kumar R, Oh S,  
1199  
1200 Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to  
1201 reality. *Eurosurveillance* (2017).  
1202  
1203 Siu KL, Yuen KS, Castaño-Rodriguez C, Ye ZW, Yeung ML, Fung SY, Yuan S, Chan CP, Yuen  
1204 KY, Enjuanes L, Jin DY. Severe acute respiratory syndrome coronavirus ORF3a protein  
1205 activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC.  
1206 The FASEB Journal (2019), doi: 10.1096/fj.201802418R.  
1207  
1208 Smith CL, Blake JA, Kadin JA, Richardson JE, Bult CJ, Mouse Genome Database (MGD)-2018:  
1209 Knowledgebase for the laboratory mouse. *Nucleic Acids Res.* (2018),  
1210  
1211 Subramanian A et al., Gene set enrichment analysis: A knowledge-based approach for  
1212 interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* (2005),  
1213  
1214 Sungnak W, Huang N, Bécavin C, Berg M, Queen R, et al., SARS-CoV-2 entry factors are highly  
1215 expressed in nasal epithelial cells together with innate immune genes. *Nature Medicine* (2020).  
1216  
1217 Taiaroa G, Rawlinson D, Featherstone L, Pitt M, Caly L, Druce J, Purcell D, Harty L, Tran T,  
1218 Roberts J, Catton M, Williamson D, Coin L, Duchene S. Direct RNA sequencing and early  
1219 evolution 2 of SARS-CoV-2. 2020 Mar 7. *BioRxiv*.  
1220 <https://www.biorxiv.org/content/10.1101/2020.03.05.976167v1>  
1221  
1222 Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z, Cui J, Lu J.  
1223 On the origin and continuing evolution of SARS-CoV-2, *National Science Review*. 2020. Mar 3.  
1224  
1225 Tanner NA, Zhang Y, Evans TC. Visual Detection of Isothermal Nucleic Acid Amplification  
1226 Using pH-Sensitive Dyes. *Biotechniques*. 58 (2), 59-68. 2015 Feb 1.  
1227  
1228 Updegraff BL, Zhou X, Guo Y, Padanad MS, Chen PH, Yang C, Sudderth J, Rodriguez-Tirado  
1229 C, Girard L, Minna JD, Mishra P, DeBerardinis RJ, O'Donnell KA. Transmembrane Protease  
1230 TMPRSS11B Promotes Lung Cancer Growth by Enhancing Lactate Export and Glycolytic  
1231 Metabolism. *Cell Rep*. 2018 Nov 20;25(8):2223-2233.e6.  
1232  
1233 van Doremalen N, Bushmaker T, Morris D, Holbrook M, Gamble A, Williamson B, Tamin A,  
1234 Harcourt J, Thornburg N, Gerber S, Lloyd-Smith J, de Wit E, Munster V. Aerosol and surface  
1235 stability of HCoV-19 (SARS-CoV-2) compared to SARS-CoV-1. 2020. Mar 17. *New England*  
1236 *Journal of Medicine*.  
1237  
1238 Vieira Braga, F. A. et al. A cellular census of human lungs identifies novel cell states in health  
1239 and in asthma. *Nat. Med.* (2019). doi:10.1038/s41591-019-0468-5  
1240

- 1241 Wang L, Wang S, Li W, RSeQC: Quality control of RNA-seq experiments. *Bioinformatics*  
1242 (2012)  
1243
- 1244 Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with  
1245 multi-subject single-cell expression reference. *Nat. Commun.* (2019). doi:10.1038/s41467-018-  
1246 08023-x  
1247
- 1248 Woelfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, Niemeyer D, Kelly  
1249 TCJ, Vollmar P, Rothe C, Hoelscher M, Bleicker T, Brünink S, Schneider J, Ehmann R,  
1250 Zwirgmaier K, Drosten C, Wendtner C. Virological assessment of hospitalized cases of  
1251 coronavirus disease 2019 . 2020. Mar 8.  
1252 <https://www.medrxiv.org/content/10.1101/2020.03.05.20030502v1>  
1253
- 1254 Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.*  
1255 (2019).  
1256
- 1257 Wylie AL, Fournier J, Casanovas-Massana A, Campbell M, Tokuyama M, Vijayakumar P, et al.,  
1258 Saliva is more sensitive for SARS-CoV-2 detection in COVID-19 patients than nasopharyngeal  
1259 swabs. *MedRxiv*. <https://www.medrxiv.org/content/10.1101/2020.04.16.20067835v1>  
1260
- 1261 Yu G. Using ggtree to visualize data on tree-like structures. *Current Protocols in Bioinformatics*,  
1262 2020, 69:e96. doi: 10.1002/cpbi.96.  
1263
- 1264 Yu L, Wu S, Hao X, Li X, Liu X, Ye S, Han H, Dong X, Li X, Li J, Liu J, Zhang W, Pelechano  
1265 V, Chen W, Yin X. Rapid Colorimetric Detection of COVID-19 Coronavirus Using a Reverse  
1266 Transcriptional Loop-Mediated Isothermal Amplification (RT-LAMP) Diagnostic Platform:  
1267 iLACO. 2020 Feb 24.  
1268 <https://www.medrxiv.org/content/10.1101/2020.02.20.20025874v1>  
1269
- 1270 Zamani S et al., *Mycobacterium avium* subsp. *paratuberculosis* and associated risk factors for  
1271 inflammatory bowel disease in Iranian patients. *Gut Pathog.* 9, 1 (2017).  
1272
- 1273 Zhao Z, Sokhansanj BA, Rosen G. Characterizing geographical and temporal dynamics of novel  
1274 coronavirus SARS-CoV-2 using informative subtype markers. *BioRxiv*  
1275 <https://www.biorxiv.org/content/10.1101/2020.04.07.030759v3>  
1276
- 1277 Zhang P, Zhu L, Cai J, Lei F, Qin J, et al. Association of Inpatient Use of Angiotensin  
1278 Converting Enzyme Inhibitors and Angiotensin II Receptor Blockers with Mortality Among  
1279 Patients With Hypertension Hospitalized With COVID-19. *Circulation Research*. 2020 Apr 17.  
1280
- 1281 Zhang Y, Odiwuor N, Xiong J, Sun L, Nyaruaba RO, Wei H, Tanner NA. Rapid Molecular  
1282 Detection of SARS-CoV-2 (COVID-19) Virus RNA Using Colorimetric LAMP. 2020. Feb 29.  
1283 *MedRxiv* <https://www.medrxiv.org/content/10.1101/2020.02.26.20028373v1>  
1284
- 1285 Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li  
1286 H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B. Clinical course and risk factors for mortality of

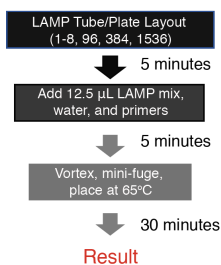
1287 adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020  
1288 Mar 11.  
1289

## Figures and Legends for “Shotgun Transcriptome and Isothermal Profiling of SARS-CoV-2 Infection Reveals Unique Host Responses, Viral Diversification, and Drug Interactions” (Butler *et al.* 2020)

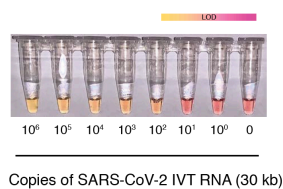
a



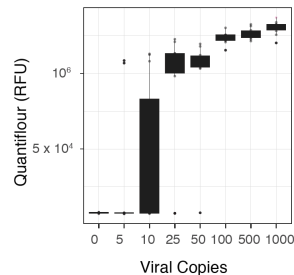
b



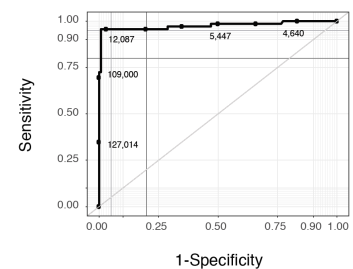
c



d

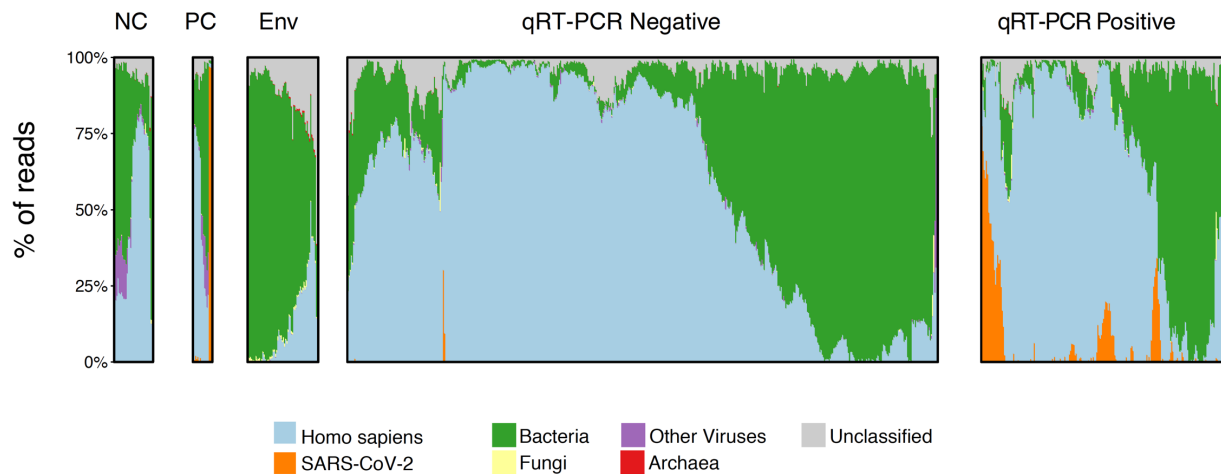


e

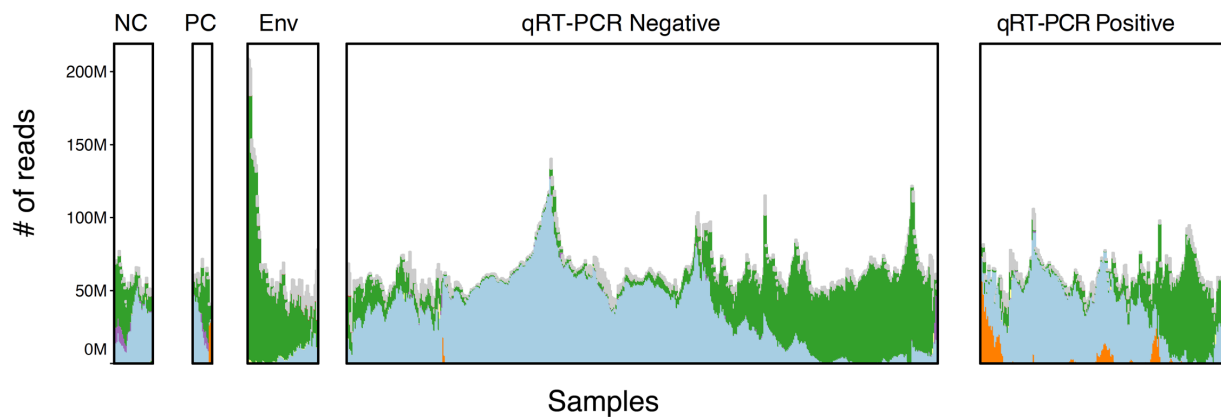


**Figure 1. Sample Processing, the Loop-Mediated Isothermal (LAMP) Reaction and Synthetic RNA Validation.** (a) Clinical and environmental samples collected with nasopharyngeal (NP) and isohelix swabs respectively, were tested with RNA-sequencing, qRT-PCR, and LAMP. (b) The test samples were prepared using an optimized LAMP protocol from NEB, with a reaction time of 30 minutes. (c) Reaction progress was measured for the Twist SARS-CoV-2 synthetic RNA (*MT007544.1*) from 1 million molecules of virus (10<sup>6</sup>), then titrated down by log<sub>10</sub> dilutions. The colorimetric findings of the LAMP assay are based on a yellow to pink gradient with higher copies of SARS-CoV-2 RNA corresponding to a yellow color. The limit of detection (LoD) range is shown with a gradient after 30 minutes between 10 and 100 viral copies (lower right). (d) Replicates of the titrated viral copies using LAMP, as measured by Quantifluor fluorescence. (e) The sensitivity and specificity of the LAMP assay from 201 patients (132 negative and 69 positive for SARS-CoV-2, as measured by qRT-PCR). Thresholds are DNA quantified by the Quantifluor.

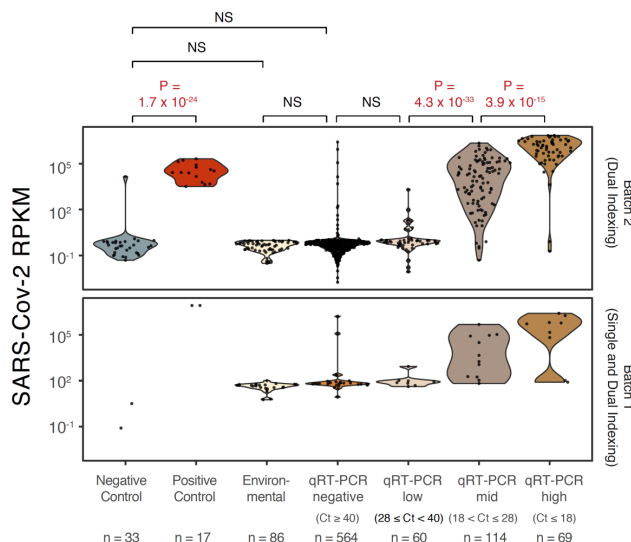
a



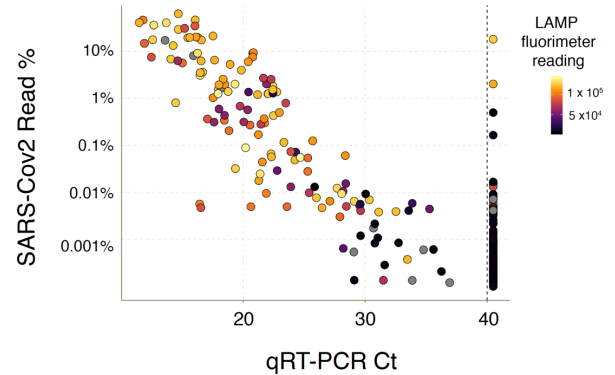
b



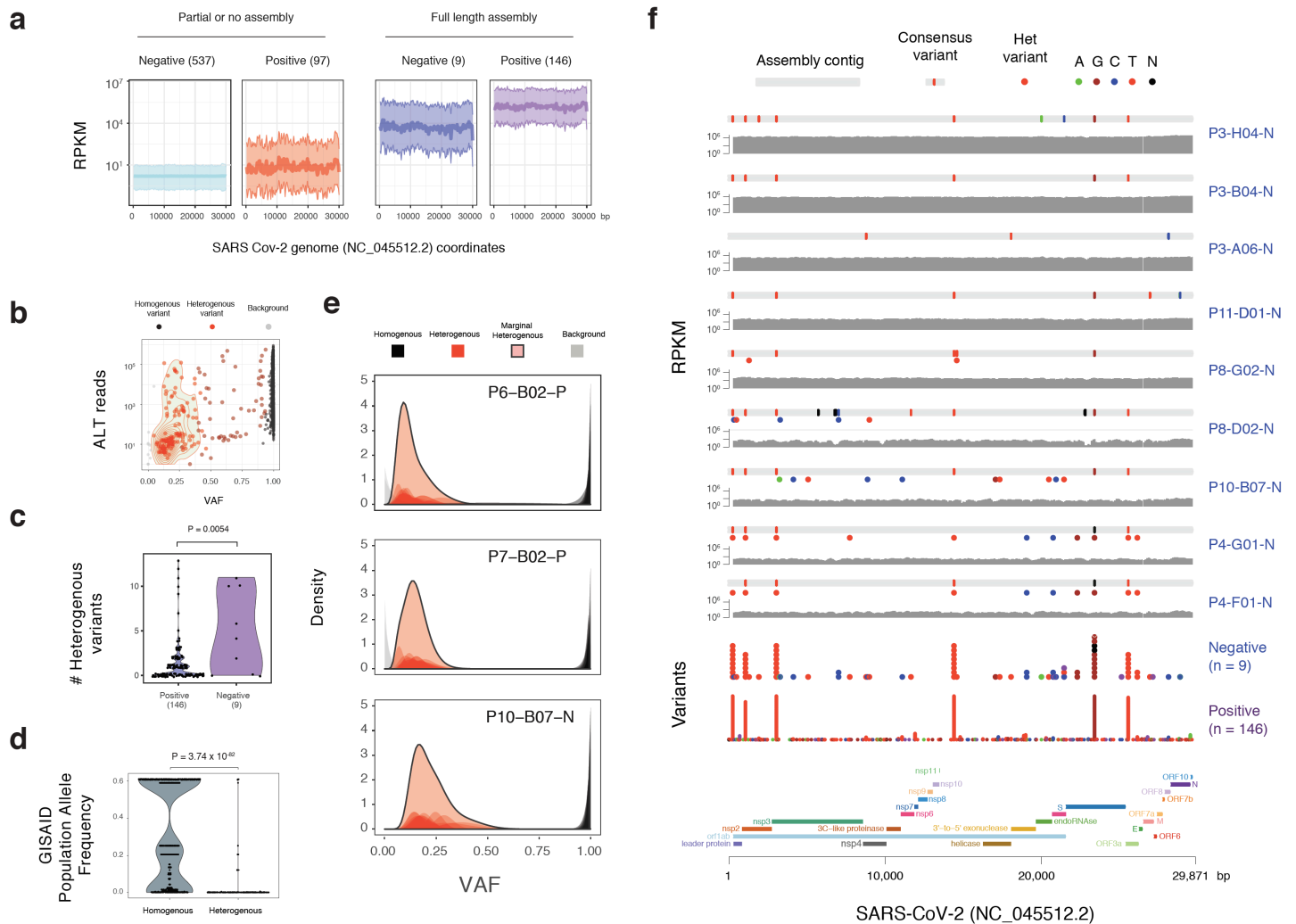
c



d

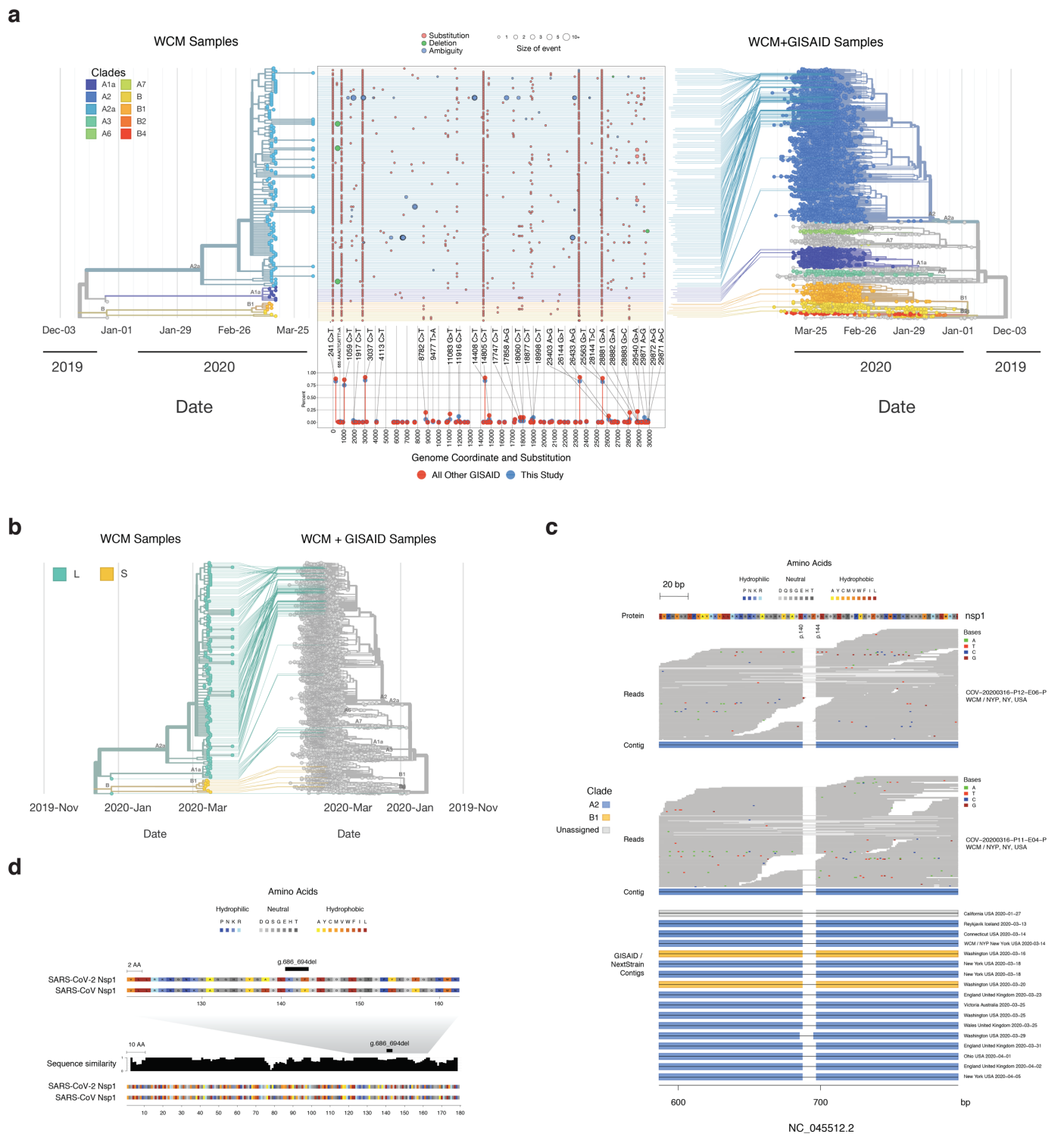


**Figure 2. Full transcriptome profiles of SARS-CoV-2 Patients with NGS, qRT-PCR, and LAMP.** (a,b) Read mapping to archaea (red), bacteria (green), fungi (yellow), human (blue), and SARS-CoV-2 (orange), and other viruses (grey), across the clinical controls (CN, CP), environmental samples, qRT-PCR negative, and qRT-PCR positive samples. (c) Clinical samples tested by qRT-PCR (Positive, n=255, or Negative, n=564) were sequenced and run through the LAMP assay. These results were compared to the buffer blanks (Negative Control, CN, n=33), Vero E6 cell extracts with SARS-CoV-2 infection (Positive Controls, CP, n=17), and Subway Samples (Environmental, Env, n=86). Read proportions are shown on the y-axis. (d) SARS-CoV-2 abundance, as measured with NGS and percentage of reads (y-axis) is compared to the Ct Threshold for qRT-PCR (x-axis), with lower Ct values representing higher viral abundance, and the LAMP reaction output (Fluorimeter values, black to yellow scale).



**Figure 3. Viral genome assemblies and variants** (a) Clinical samples tested by qRT-PCR (Negative, N) were compared to RNA-seq coverage (Read per Kilobase per Million Reads, RPKM, y-axis) for the set of fully-assembled positive samples (n=146, 9) vs. those with partial or no assemblies (left, n=537, 97). (b) Variant allele frequencies (VAF, x-axis) for alternative alleles (y-axis) were calculated for all variants across viral strains, with heterogenous (het, 5%<x<95%, red) variants shown as well as homogenous (black) variants with VAF>0.95. (c) Proportion of het variants (y-axis) is shown for those tested by qRT-PCR and shown as negative or positive (x-axis). (d) The frequency of variants present in the GISAID global database of virus sequences (y-axis) is shown for variant types (x-axis). (e) The distribution and density of the VAF for three exemplar samples are shown relative to their variant type (top) (f) The assembled contigs (grey) the consensus variants (red dots) relative to the reference, and coverage (RPKM) are shown for a set of clinical samples, relative to the reference SARS-CoV-2 genome annotation (bottom), gene segments (colored bars), and histogram of variants from qRT-PCR positive and negative samples (variants y-axis).

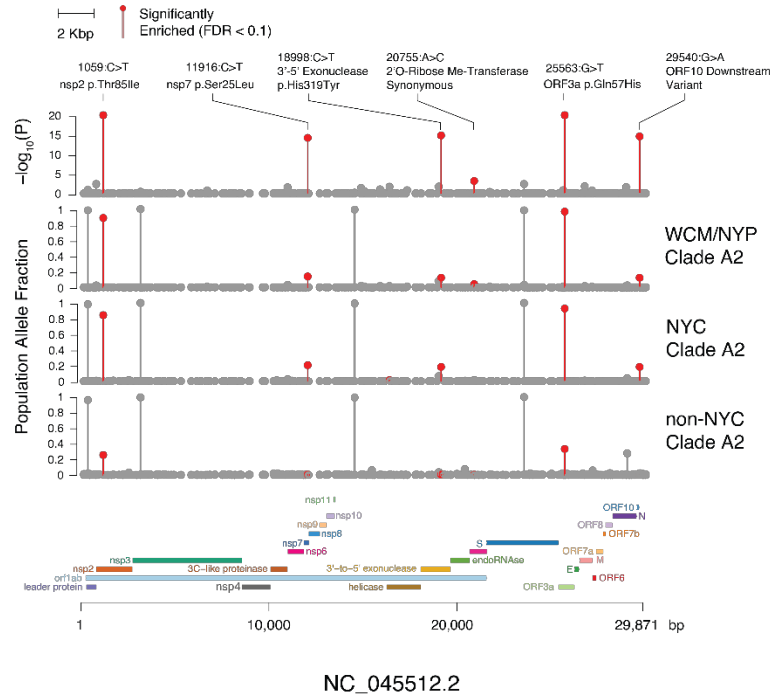




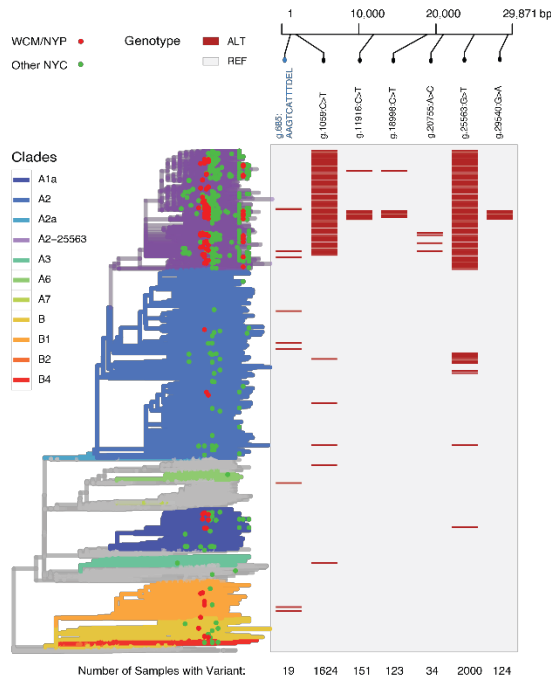
**Figure 4 – The mutational landscape SARS-CoV-2.** (a) The phylogenetic placement of these SARS-CoV-2 samples is shown on the tree (left) and the global map of known SARS-CoV-2 genomes (right). Genetic variants called from the RNA-seq data (middle) show a range of variants that are distinct from the Wuhan reference strain, and the samples from this study, highlighted in blue, show enrichment for European and Asian alleles. The annotation track on the bottom shows variants called in these samples alongside all other GISAID samples. Samples that diverge from GISAID by more than 5% in either direction are annotated, including their coordinate and substitution event. (b) Proportion of the L (green) and S strain (yellow) are shown for the NYC viruses. Phylogeny of samples from this study on the left and total GISAID samples on the right, with a map of variants in this study's samples in the middle, colored by event type and sized by number of nucleotides impacted. Annotation track on the bottom shows frequency of alternate alleles in this study and in the GISAID database. (c) The 9-bp deletion in ORF1b (NSP1 protein) that was detected in samples from three NYPH-WCMC patients was confirmed in the GISAID database (bottom tracks). Read-level support is shown in the top tracks for

two of the variants, with aligned contigs visualized below. Assembly alignments for 17 additional cases harboring this deletion, including 1 additional case from this study are shown in the bottom track. (d) Alignment of SARS-CoV-2 to SARS-CoV NSP1 protein sequence is shown, with an enlarged view showing a sequence similarity track (normalized protein alignment score in 5 amino acid sliding window) and the 9-bp deletion region delineated.

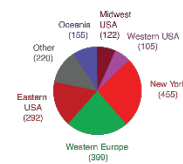
a



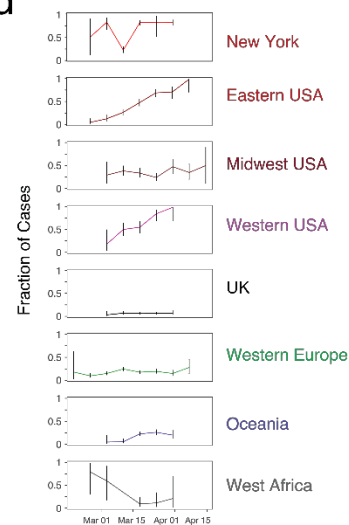
b



c



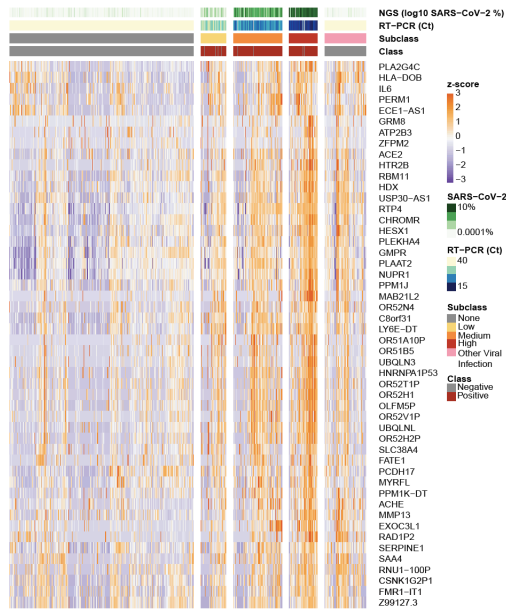
d



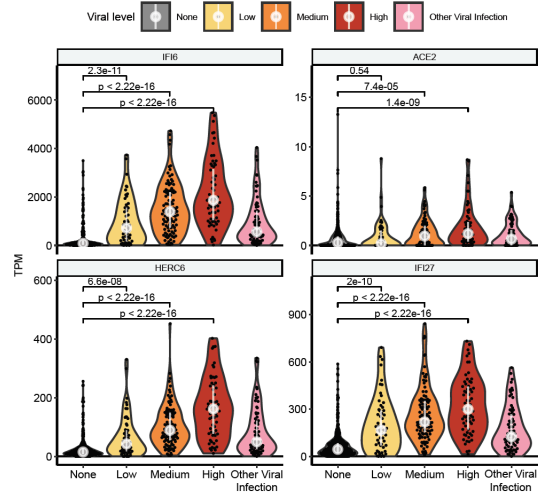
**Figure 5. Variant and Subclade Analysis.** (a) Six variant alleles were found to be significantly enriched by population allelic fraction within this set of 93 WCM/NYP cases as compared with non-NYC strains of Nextstrain clade A2 (1059:C>T  $P = 2.13 \times 10^{-47}$ , 11916:C>T  $P = 6.26 \times 10^{-15}$ , 18998:C>T  $P = 1.53 \times 10^{-15}$ , 20755:A>C  $P = 6.6 \times 10^{-4}$ , 25563:G>T  $P = 7.92 \times 10^{-50}$ , 29540:G>A  $P = 2.68 \times 10^{-15}$ ). These variants demonstrated similar PAF enrichment within all other NYC strains as compared to non-NYC clade A2 (1059:C>T  $P = 3.36 \times 10^{-146}$ , 11916:C>T  $P = 3.53 \times 10^{-74}$ ,

18998:C>T  $P = 3.94 \times 10^{-73}$ , 20755:A>C  $P = 7.71 \times 10^{-3}$ , 25563:G>T  $P = 1.45 \times 10^{-154}$ , 29540:G>A  $P = 2.07 \times 10^{-72}$ ). (b) (left) Phylogenetic tree produced by the Nextstrain analysis with clade affiliations and nodes corresponding to WCM/NYP in red and other NYC cases in green. (right) occurrence of the six NYC-enriched alleles and the 9 nucleotide deletion across genomes. (c) Raw counts of cases present within this A2-25563 subclade demonstrated a predominance of European and North American cases, with Western Europe and New York together comprising the majority of strains. (d) Fraction of A2-25563 cases from each region of the world.

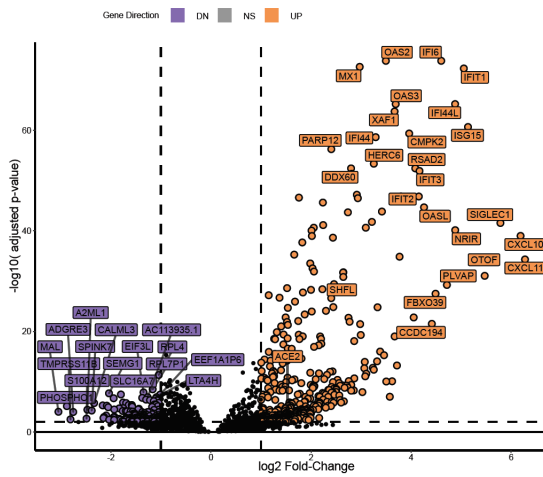
a



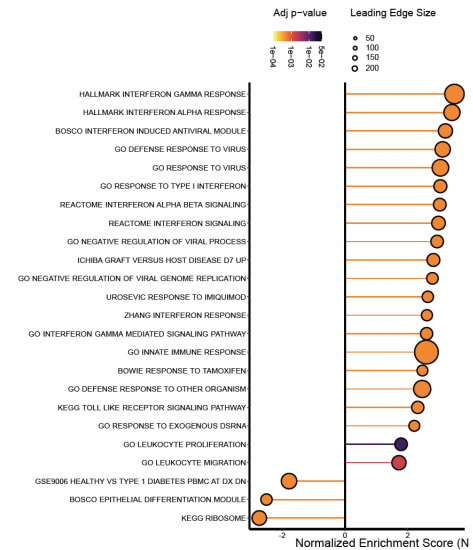
b



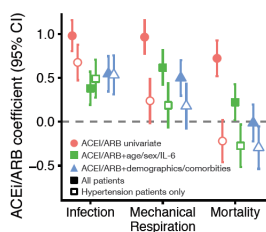
c



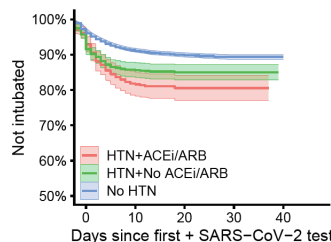
d



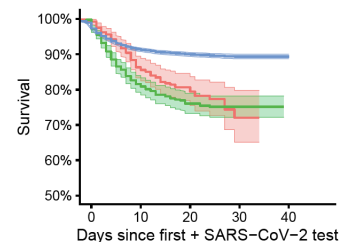
e



f



g



**Figure 6: Host transcriptome responses and risk to SARS-CoV-2.** (top row) Samples were quantified by a range of viral load, including RNA-seq ( $\log_{10}$  SARS-CoV-2 % of reads), and qRT-PCR (Ct values) to create a three-tier range of viral load for the positive samples (right) compared to the clinically-annotated negative samples (class, red or grey) and those samples with other viral infections that were SARS-CoV-2 negative by qRT-PCR. (b) The differentially expressed genes of qRT-PCR positive patients compared to qRT-PCR patients showed up-regulated (orange) genes as well as down-regulated (purple) genes. (b) Up-regulated genes, with boxplots across all samples, include *IFI6*, *ACE2*, *SHFL*, *HERC6*, *IFI27*, and *IFIT1*, based on data from (c), which is the total set of DEGs. The full set is shown in an intersecting heat map, with a core set of up-regulated genes (orange) distinct from the set of down-regulated genes (purple), compared to genes that are not significantly differently expressed (grey) in any comparison (DESeq2, q-value <0.01,  $|\log_{2}FC| > 0.58$ ). (d) GSEA enrichment of significant pathways, with color indicating statistical significance and circle size the number of genes on the leading edge (e) Regression coefficients for variables indicating exposure/history of exposure to ACEI/ARBs inhibitors for each of the three cohort comparisons: (left) test outcome in a cohort of patients suspected of SARS-CoV-2 infection, (middle) requirement of mechanical respiration in patients who tested positive, (right) mortality in patients who tested positive. Univariate analyses are shown as red circles. The green triangles coefficients are when correcting for age, sex, and baseline IL-6 levels taken upon admission. The blue squares are from a model that includes age, sex, and IL-6 as well as comorbidities including CAD/CHD, diabetes, being overweight, or obesity, and self-reported race and ethnicity. Additionally, open markers indicate the same analyses but restricted to only patients with clinically diagnosed hypertension. (f) Curves for patients requiring mechanical respiration (identified by intubation procedure notes) as a function of drug class (g) and survival, conditioned on hypertension status and ACEI/ARB exposure status.

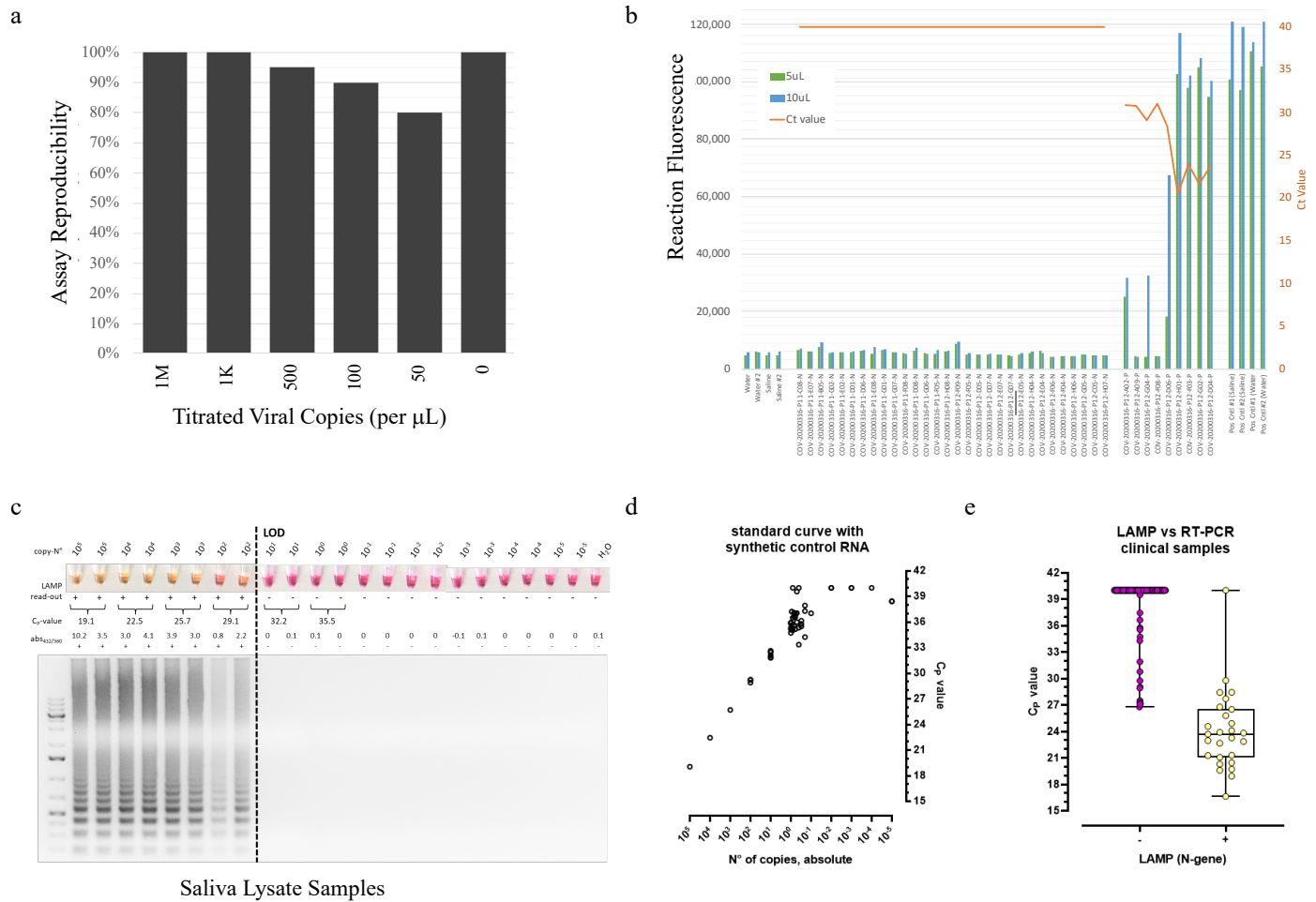
**Table 1. Baseline characteristics of SARS-CoV-2 suspected cohort.** General population refers to a comparison cohort of individuals administered drugs at NYPH-CUIMC between January 1, 2019 – September 24, 2019 who were not later tested for SARS-COV-2 infection.

Variable	General	Tested	COV+	COV+/Intubated	COV+/Died
N (% of tested)	90989	8856 (100%)	4829 (54.5%)	534 (6%)	528 (6%)
ACEI	2201 (2.4%)	408 (4.6%)	310 (6.4%)	78 (14.6%)	57 (10.8%)
ARB	2113 (2.3%)	400 (4.5%)	295 (6.1%)	70 (13.1%)	65 (12.3%)
ACEI and ARB	163 (<0.1%)	28 (0.3%)	22 (0.5%)	7 (1.3%)	4 (0.8%)
Median age (50%)	34.6 (27.5-42.2)	57.2 (38.3-71.8)	62.4 (47.8-75)	65.4 (54.4-74.8)	78.2 (69.9-87)
Male	31228 (34.3%)	4001 (45.2%)	2546 (52.7%)	336 (62.9%)	309 (58.5%)
Black/African-American	1739 (1.9%)	2047 (23.1%)	1229 (25.5%)	125 (23.4%)	95 (18%)
Caucasian	108 (0.1%)	2587 (29.2%)	1249 (25.9%)	127 (23.8%)	141 (26.7%)
Asian or Pacific Islander	1564 (1.7%)	237 (2.7%)	106 (2.2%)	4 (0.7%)	4 (0.8%)
Other race	79 (0.1%)	2134 (24.1%)	1250 (25.9%)	156 (29.2%)	165 (31.2%)
Missing race	140 (0.2%)	1851 (20.9%)	995 (20.6%)	122 (22.8%)	123 (23.3%)
Hispanic/Latino	13875 (15.2%)	2811 (31.7%)	1704 (35.3%)	260 (48.7%)	251 (47.5%)
Non-Hispanic/Latino	23537 (25.9%)	3576 (40.4%)	1811 (37.5%)	151 (28.3%)	156 (29.5%)
Other ethnicity		95 (1.1%)	61 (1.3%)	7 (1.3%)	13 (2.5%)
Missing ethnicity	53577 (58.9%)	2374 (26.8%)	1253 (25.9%)	116 (21.7%)	108 (20.5%)
Hypertension	15298 (16.8%)	2200 (24.8%)	1392 (28.8%)	243 (45.5%)	310 (58.7%)
CAD/CHD	1978 (2.2%)	831 (9.4%)	511 (10.6%)	92 (17.2%)	137 (25.9%)
Diabetes	7239 (8.0%)	1340 (15.1%)	867 (18%)	163 (30.5%)	203 (38.4%)
Overweight	12878 (14.2%)	923 (10.4%)	560 (11.6%)	95 (17.8%)	101 (19.1%)
No RF	64695 (71.1%)	6223 (70.3%)	3212 (66.5%)	264 (49.4%)	198 (37.5%)
Median IL-6 (50%)	-	67.8 (29.3-152.9)	70.2 (30.8-154.4)	152.8 (72.9-278.1)	146.1 (81.2-284.8)

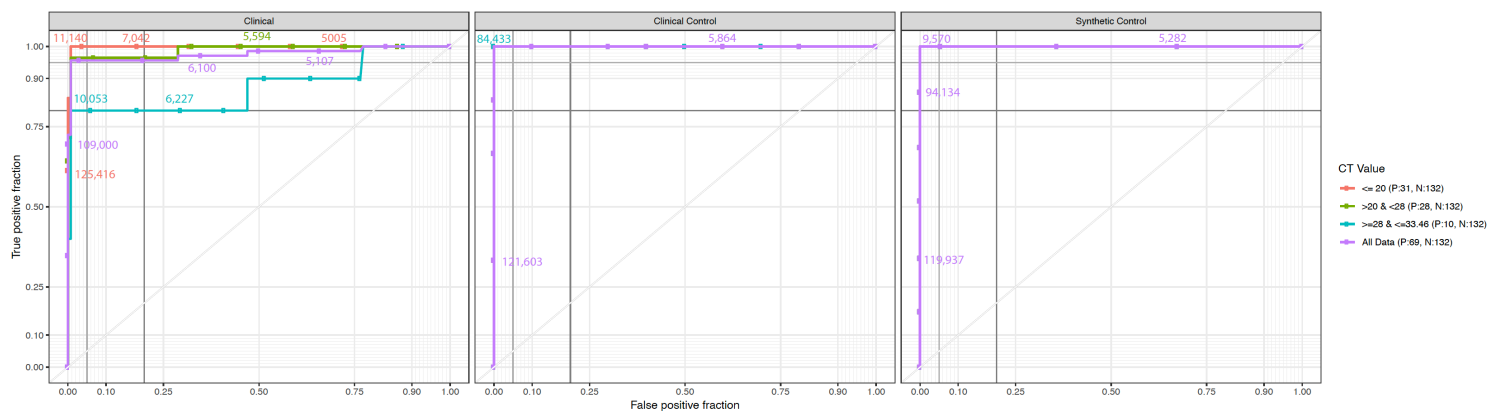
**Table 2. Comparison between ACE/ARB exposure and CCB exposure.** To evaluate if evidence of hypertension treatment is confounding association between ACE/ARB exposure and infection. These analyses were completed using only patients with hypertension. In each test the ACE/ARB groups showed significantly greater rates of infection. CCB exposure was not found to be associated with infection.

<b>Comparison</b>	<b>OR</b>	<b>95% CI</b>	<b>p-value</b>
ACEi vs none	1.481	1.121–1.965	4.7e–03
ARB vs none	1.776	1.298–2.449	2.2e–04
ACEi/ARB vs none	1.603	1.28–2.012	2.6e–05
CCB vs none	1.052	0.846–1.308	0.67
ACEi vs CCB	1.408	1.034–1.922	0.02
ARB vs CCB	1.688	1.201–2.388	2.0e–03
ACEi/ARB vs CCB	1.524	1.175–1.98	1.1e–03
ACEi vs ¬ACEi	1.253	0.963–1.637	0.09
ARB vs ¬ARB	1.511	1.128–2.043	4.6e–03
ACEi/ARB vs ¬ACEi/ARB	1.374	1.135–1.668	9.0e–04
CCB vs ¬CCB	0.891	0.744–1.068	0.2

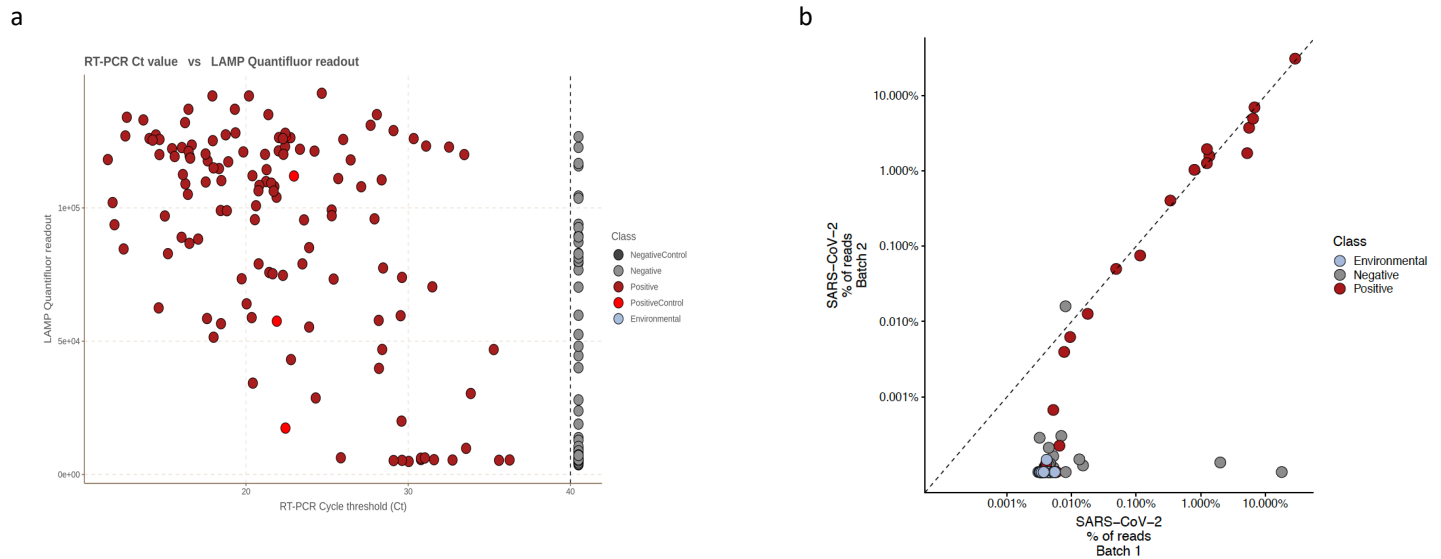




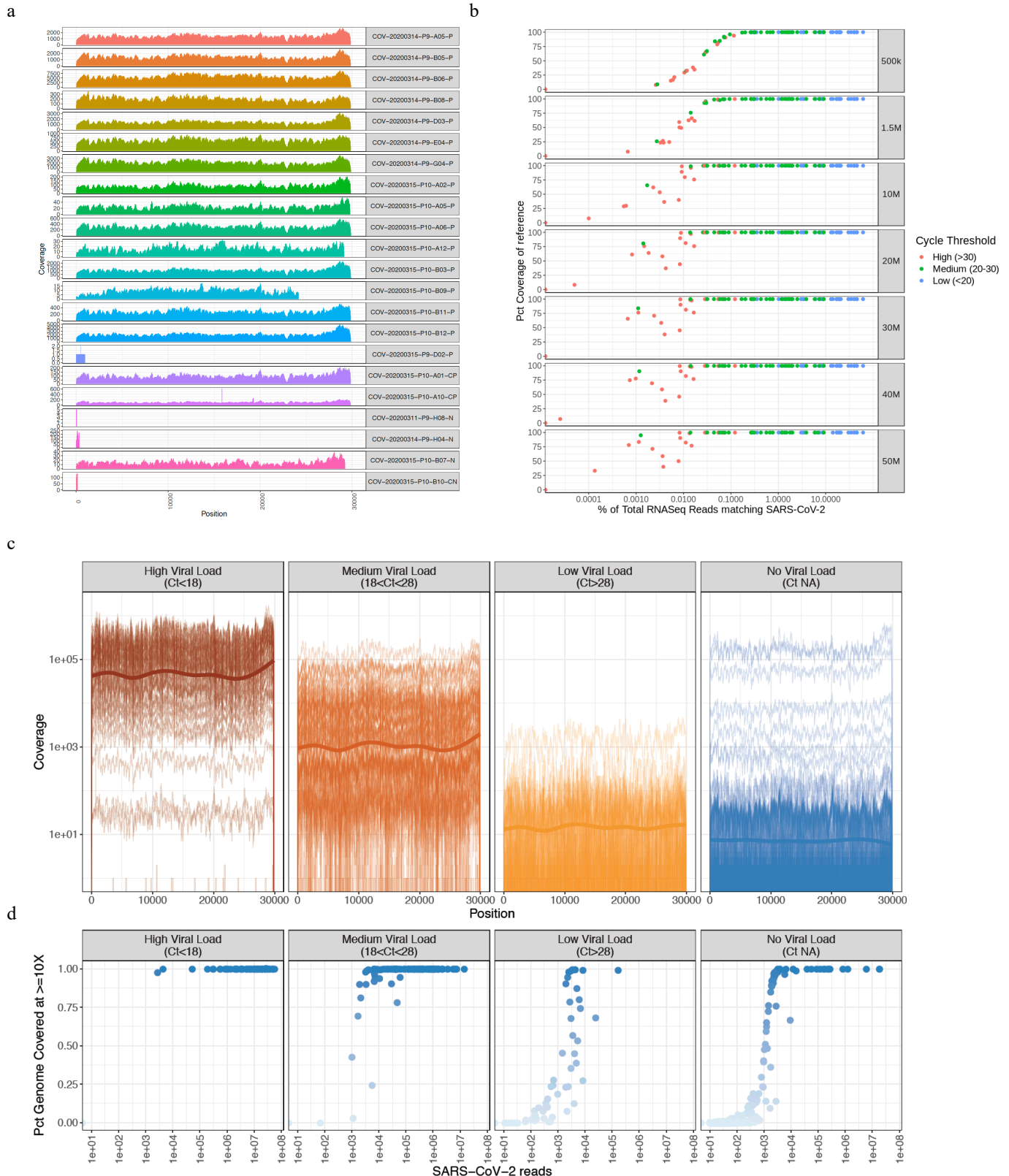
**Supplemental Figure 3. Reproducibility, sensitivity, and specificity for the LAMP Assay.** (a) Testing with a synthetic SARS-CoV-2 RNA that was serially titrated and measured in replicates ( $n=10$ ) showed 100% and 95% reproducibility at 1,000 and 500 copies, respectively, with lower rates at lower viral titers. (b) Replicates of a clinically positive (by qRT-PCR) sample at 10 uL (blue) compared to 5uL (green) showed high concordance, with greater sensitivity with increased reaction volume. (c) Whole oropharyngeal swab lysates from clinical positive ( $C_p$ -value  $>0$ ) and negative samples ( $C_p = NA$ ) were used to test the LAMP reaction. (d) The standard curve with synthetic RNA was also tested relative to absolute number of copies (x-axis) and the  $C_p$  value (y-axis). (e) The  $C_p$  value for the LAMP positive (+, right) and negative (-, left) were compared to the  $C_p$  value from qRT-PCR (y-axis).



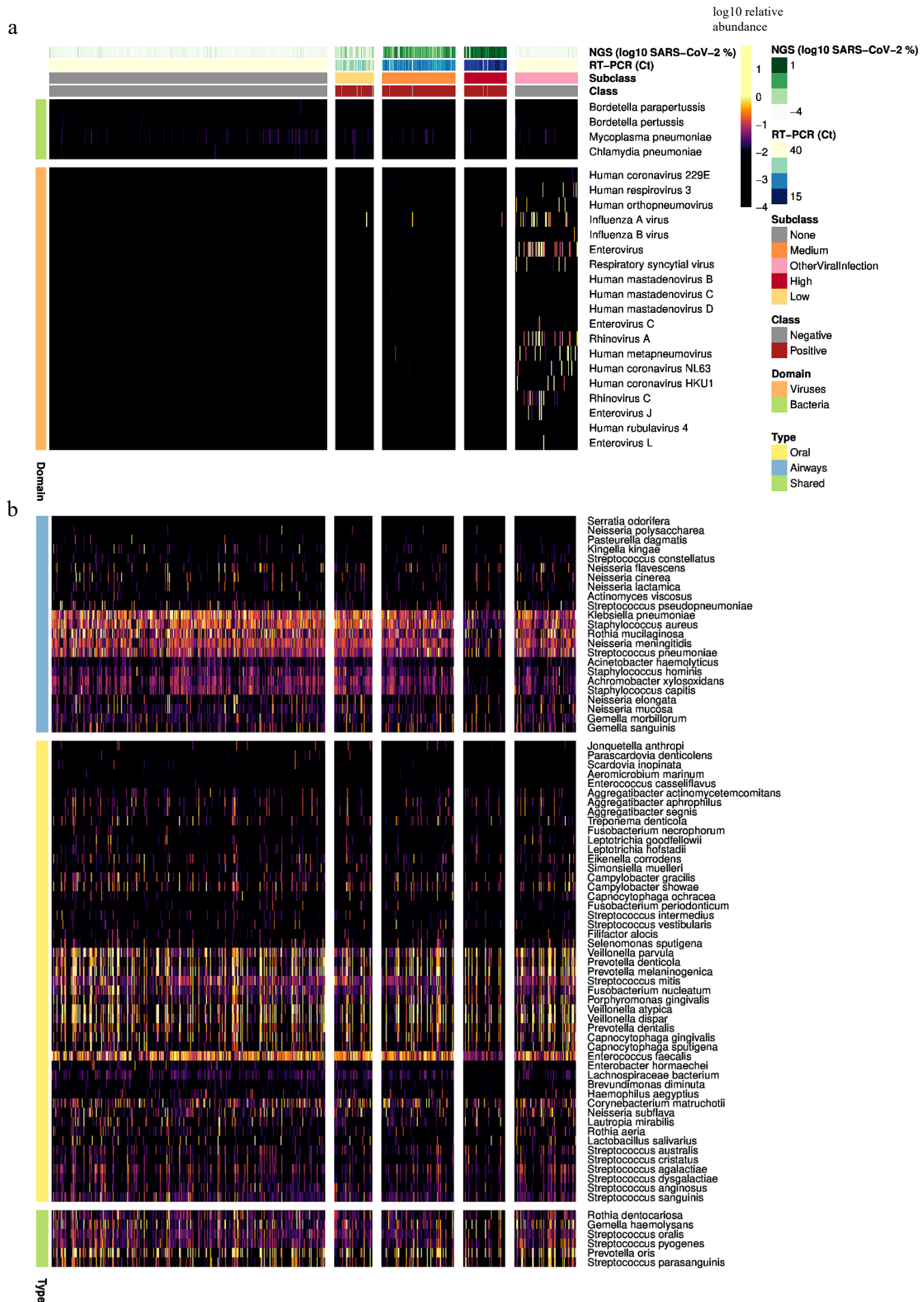




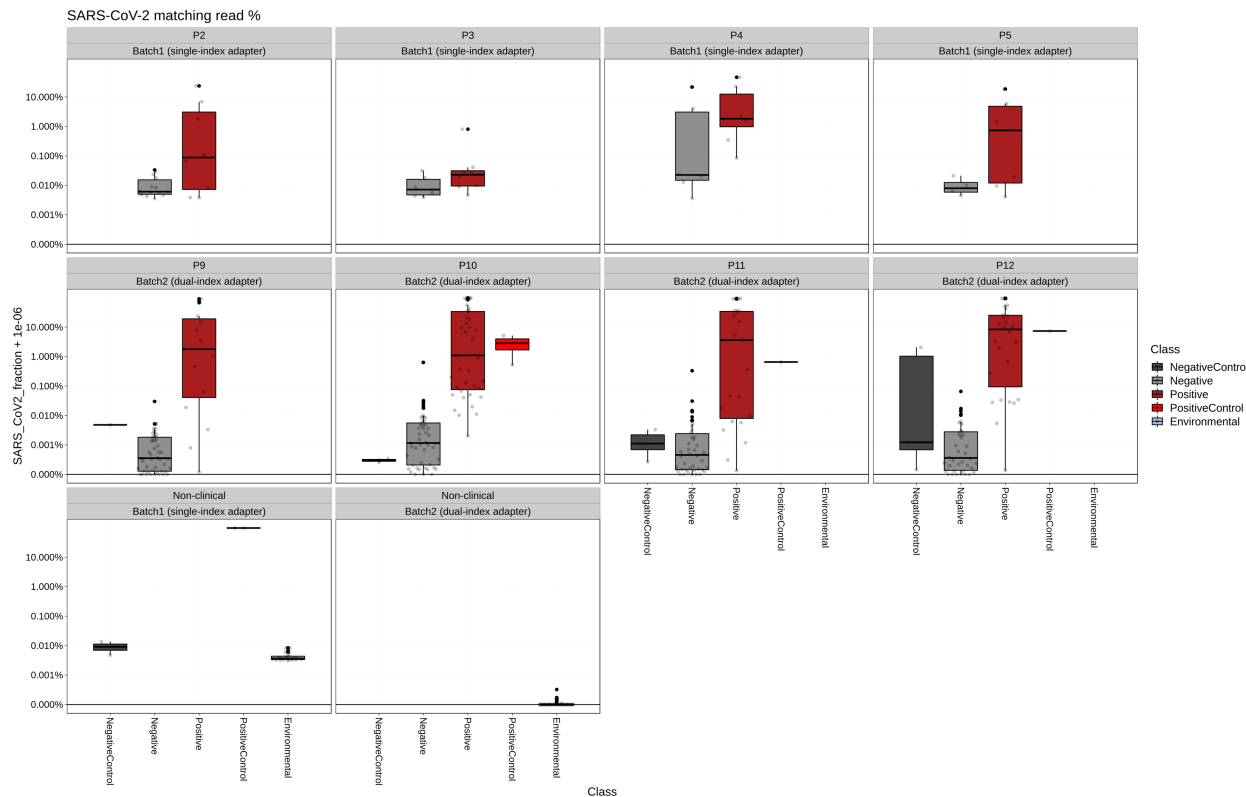
**Supplementary Figure 5. Correlation between LAMP Reaction Output and qRT-PCR as well as RNA-seq replicates.** (a) Clinical samples tested by qRT-PCR (Positive, dark red or Negative, light grey) were run with the LAMP assay and compared to the buffer blanks (Negative Control, dark grey), Synthetic RNAs or Vero 6 cell extracts with SARS-CoV-2 infection (Positive Controls, light red), and Subway Samples (Environmental, blue, lower right). The DNA abundance, as measured with the GloMax QuantiFluor (y-axis) is compared to the Ct Threshold for qRT-PCR (x-axis), with lower Ct values representing higher viral abundance. (b) RNA-seq replicates from batch 1 (single index) vs. batch 2 (dual index adapters), showing the broader dynamic range of dual-index libraries ( $R=0.85$ ).



**Supplementary Figure 6. Viral genomes from RNA-seq data and titration of coverage** (a) The coverage plot across the SARS-CoV-2 genomes (viral coordinate on bottom, colored by sample) from a representative set of clinical positive samples. Sample names with the suffixes CN and N are clinical negative (buffer), P are qRT-PCR positive, and CP are Vero E6 cells with virus. (b) Downsampling (right annotation) of the samples and mapping to the SARS-CoV-2 genome to gauge the percent coverage (y-axis) as a function of the viral quantification by qRT-PCR (Ct thresholds, low <20, medium 20-30, and high >30). (c) Average coverage statistics for the low, medium, and high Ct samples, as well as the mean coverage for each of these samples. (d) The cycle threshold (x-axis) vs. the coverage of the genome (y-axis and color depth) for the total RNA-seq.

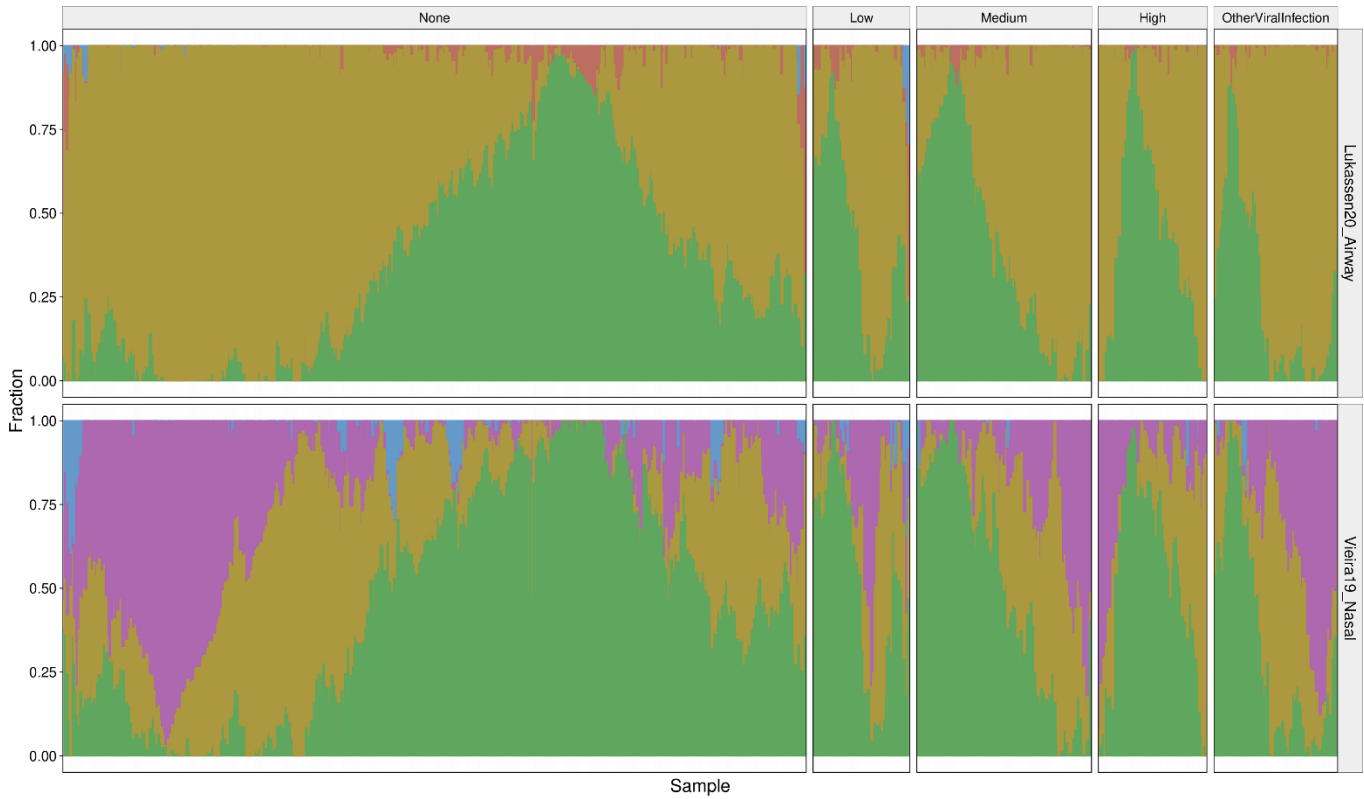


**Supplemental Figure 7. Metatranscriptome profiles of the patient cohorts.** (top row, all panels) Samples were quantified by a range of viral detection methods, including RNA-seq (log<sub>10</sub> SARS-CoV-2 % of reads) and qRT-PCR (Ct values) to create a three-tier range of viral load for the positive samples (right) compared to the clinically-annotated negative samples (class, red or grey). (bottom) The detected microbial species (horizontal lines) are plotted as log<sub>10</sub> abundance. (a) Common respiratory pathogens plotted as a log<sub>10</sub> abundance of mapped reads, with each organism as a line and each vertical column as a patient. (b) The same layout of samples as (a), but with bacteria annotated from the Human Microbiome Project (HMP) as normal airway (blue, top portion), oral (yellow, middle set), or both oral and airway flora (green, bottom).

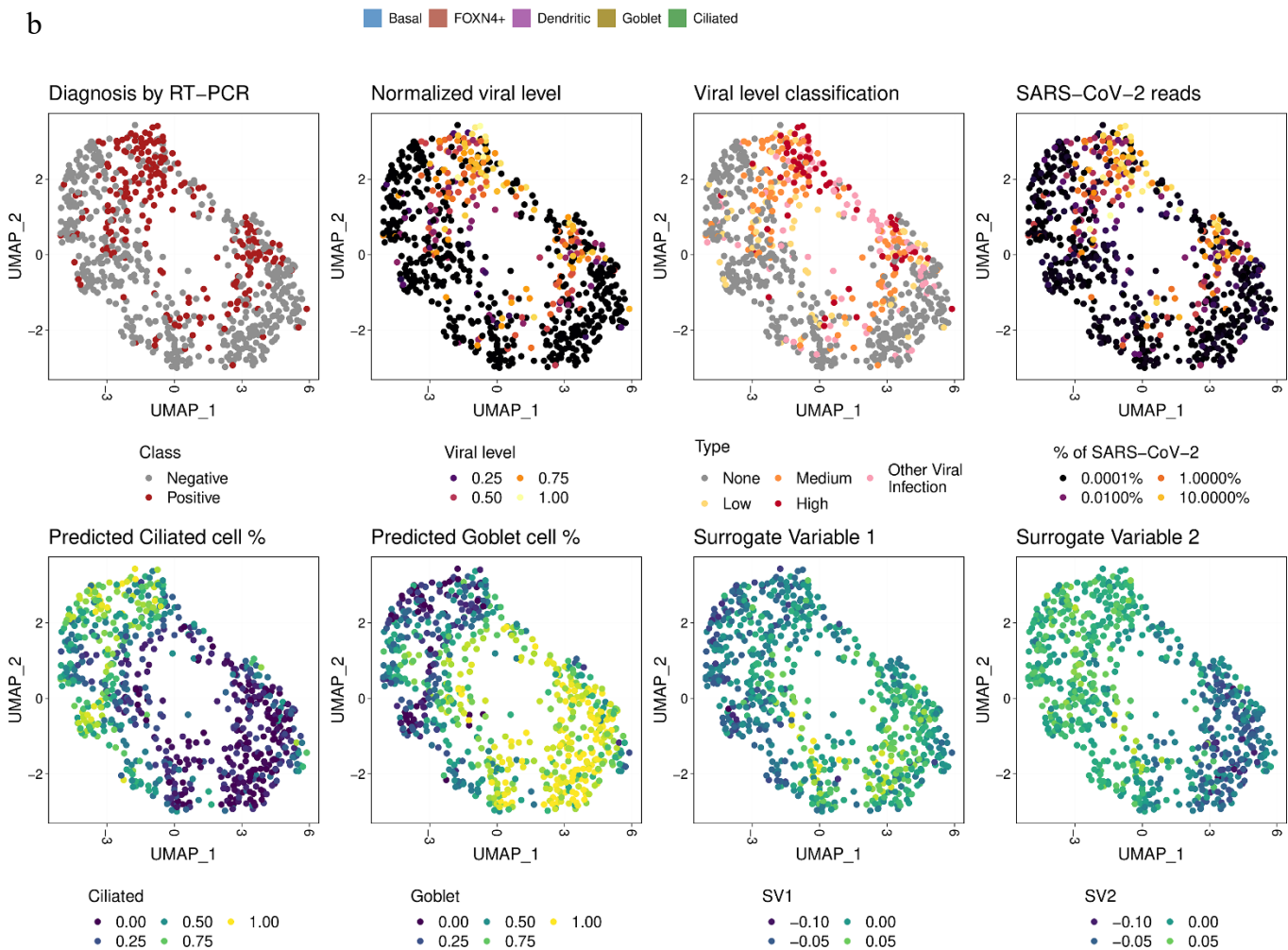


**Supplemental Figure 8. Proportion of RNA-seq reads mapping SARS-CoV-2.** Clinical samples tested by qRT-PCR (Positive, dark red or Negative, light grey) were sequenced and compared to the buffer blanks (Negative Control), dark grey, Synthetic RNAs or Vero 6 cell extracts with SARS-CoV-2 infection (Positive Controls, light red), and Subway Samples (Environmental, blue). Read totals are shown on the y-axis. Differences between single index barcodes are plotted across each of the plates of samples that were processed.

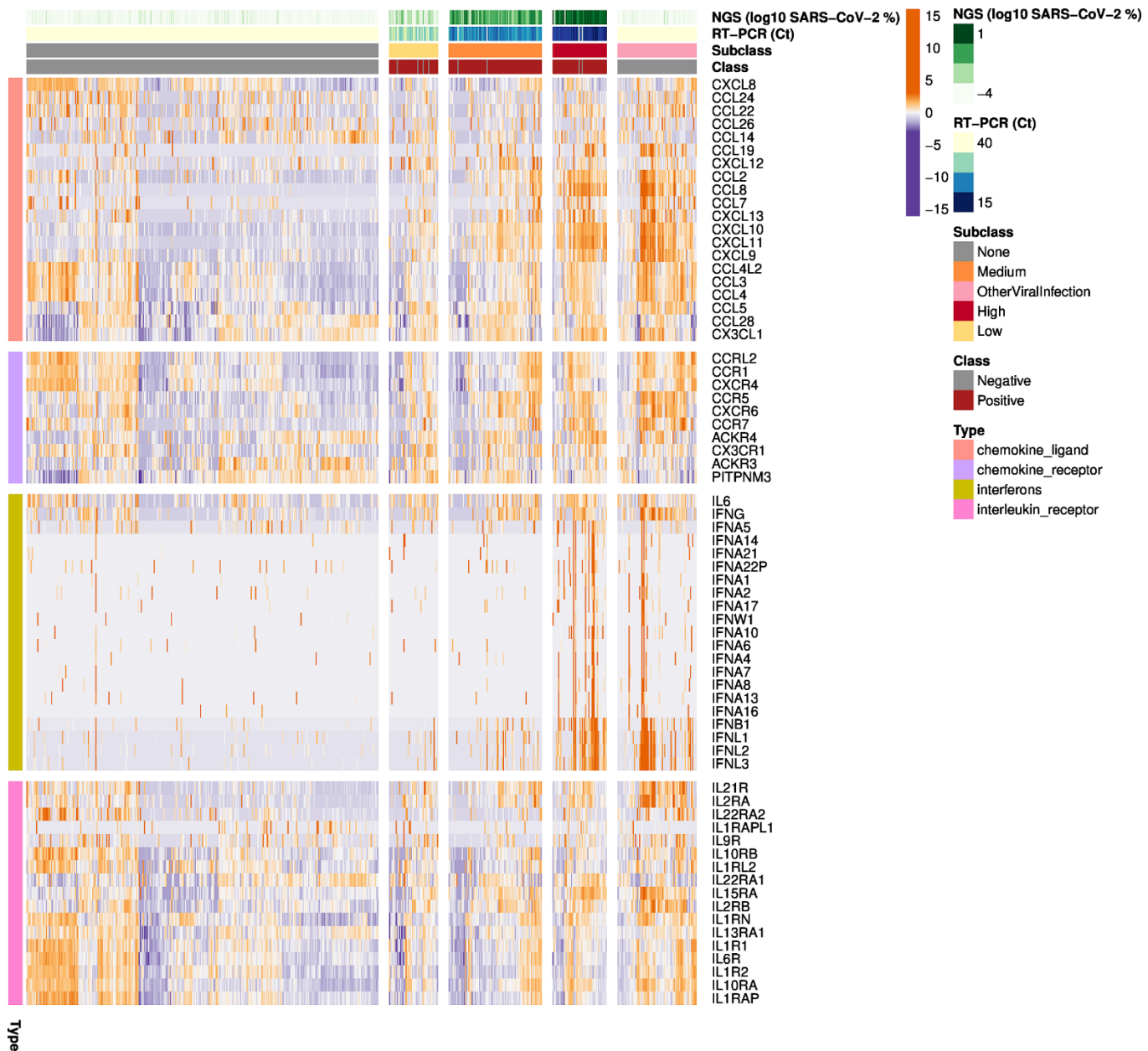
a



b

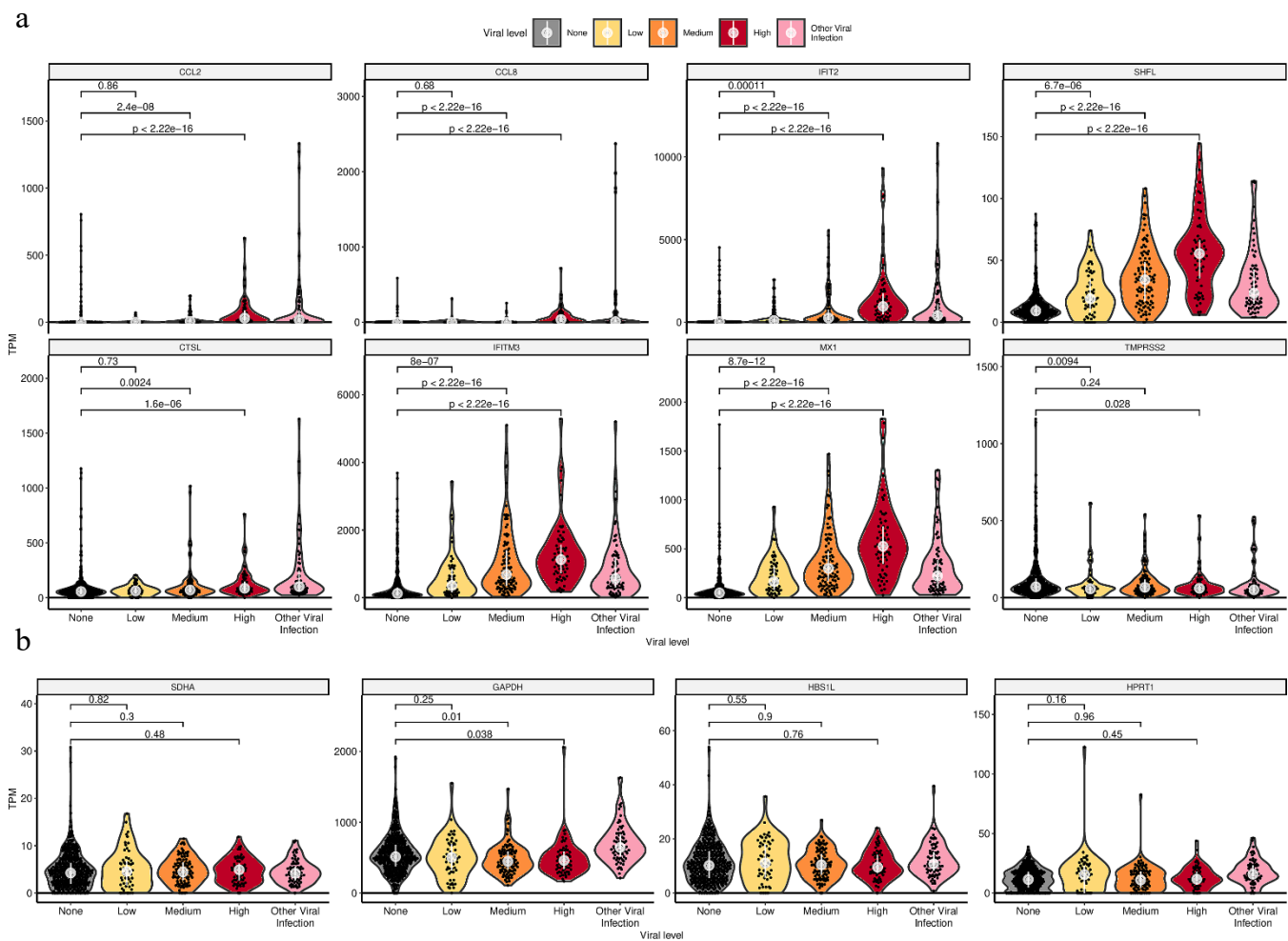


**Supplemental Figure 9. Cellular sub-type deconvolution from nasopharyngeal (NP) swabs.** The MUSIC algorithm was used to separate the cellular sub-type gene expression signatures present in the total RNA-seq data from the NP swabs. (a) The fraction of cells estimated for each cell type (y-axis) was calculated for the clinical samples that tested positive by RT-qPCR in different viral levels, as well as negative by RT-qPCR, with proportions shown for goblet cells (yellow), ciliated cells (green), basal cells (blue), dendritic cells (purple) and FOXN4+ cells (red). First row shows deconvolution using the single cell transcription reference set of upper airway cells by Lukassen *et al* 2020. and the second row shows a similar approach using nasal epithelial cells by Vieria *et al*. (b) Clinical samples are embedded in two-dimensions using dimensionality reduction by UMAP and separate features of samples are projected onto the points including RT-PCR based diagnosis, viral levels, fraction of SARS-CoV-2 reads as captured by NGS, predicted ciliated and goblet cell fractions from cell deconvolution using Lukassen airway reference, and the two predicted surrogate variables.

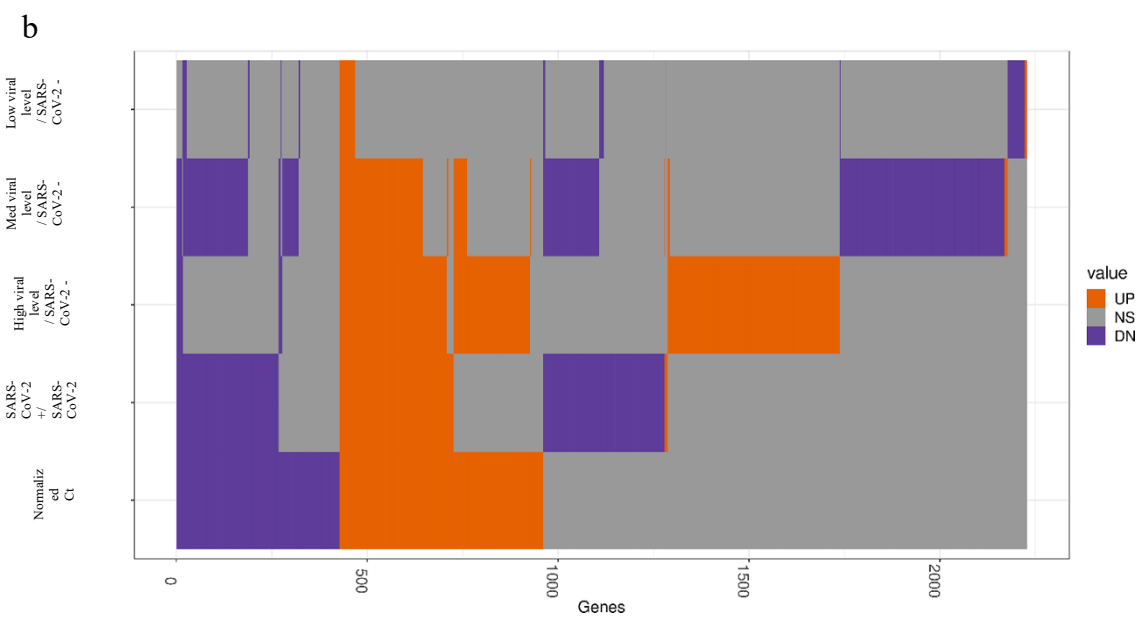
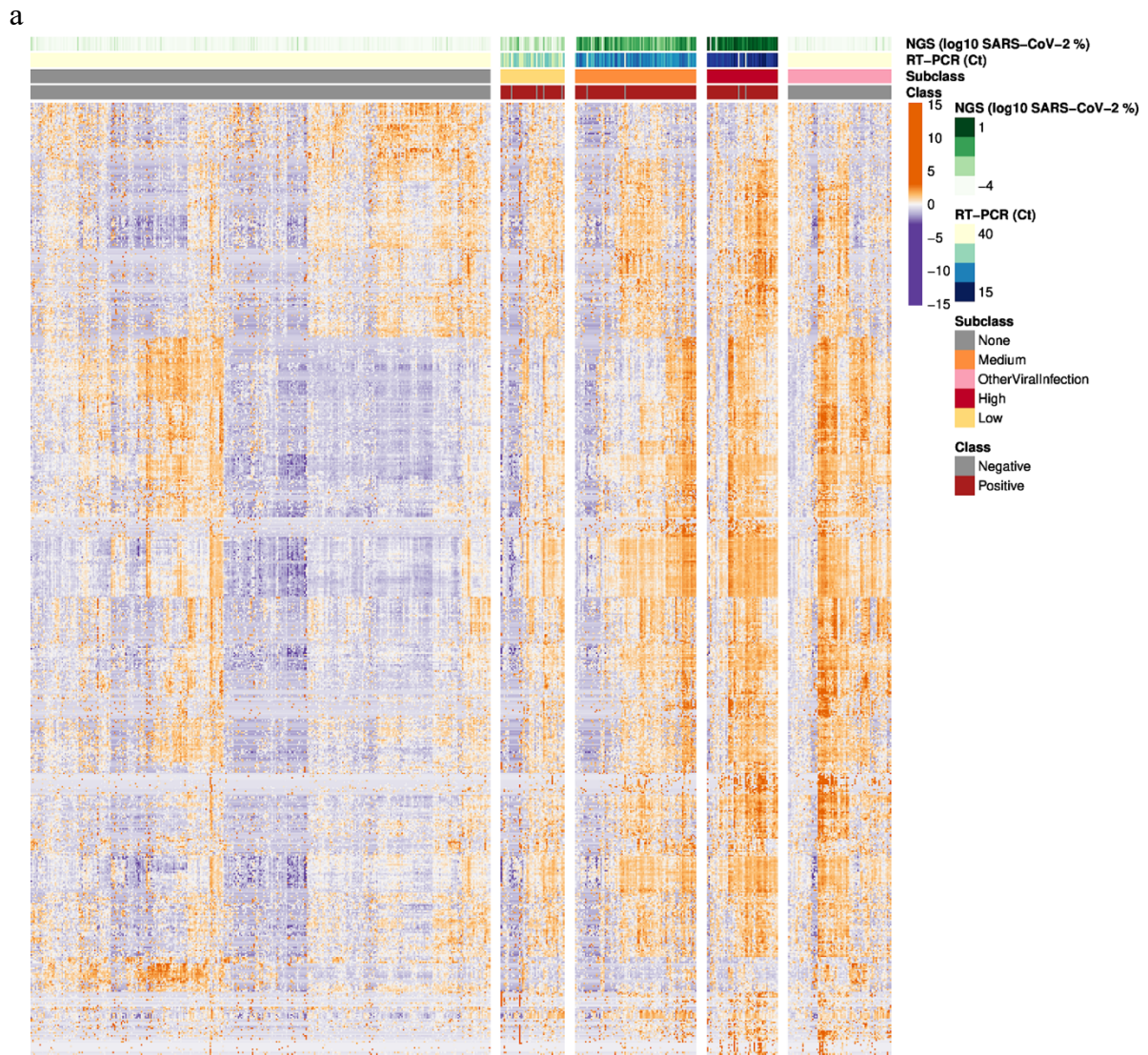


**Supplemental Figure 10. Cytokine and interferon profiles of the host transcriptome.** Top rows: samples were quantified by a range of viral detection methods, including LAMP (QuantiFluor), RNA-seq (log10 SARS-CoV-2 % of reads), and qRT-PCR (Ct values) to create a three-tier range of viral load for the positive samples (right) compared to the clinically-annotated negative samples (class, red or grey). (bottom) The differentially expressed genes of SARS-CoV-2

positive patients compared to SARS-CoV-2 negative patients showed up-regulated (orange) genes as well as down-regulated (purple) genes. Heatmap is separated into Chemokine ligand, chemokine receptor, interferon and interleukin receptor profiles for the samples (x-axis) is plotted for each related gene (y-axis).

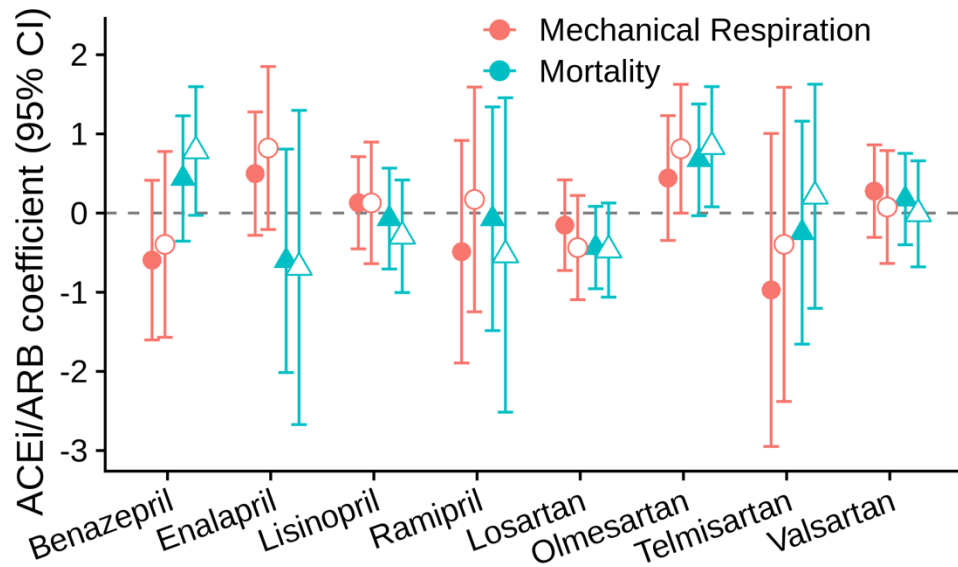


**Supplemental Figure 11. Host gene expression relative to SARS-CoV-2 infection.** Clinical samples were sequenced with RNA-seq and quantified to a set of genes for their expression levels. Samples with no virus (grey) were compared to those with low (yellow), medium (orange), and high (red) expression levels, based on qRT-PCR. p-values are calculated by Wilcoxon rank sum test and are not adjusted for multiple testing for a given gene. (a) Additional genes that were differentially expressed, or reported to be important for SARS-CoV-2 entry. (b) Expression of housekeeping genes in different groups.





**Supplemental Figure 12. Differentially expressed genes.** (a) Top rows: samples were quantified by a range of viral detection methods, including LAMP (QuantiFluor), RNA-seq (log<sub>10</sub> SARS-CoV-2 % of reads), and qRT-PCR (Ct values) to create a three-tier range of viral load for the positive samples (right) compared to the clinically-annotated negative samples (class, red or grey). (bottom) The differentially expressed genes of SARS-CoV-2 positive patients compared to SARS-CoV-2 negative patients showed up-regulated (orange) genes as well as down-regulated (purple) genes. All differentially expressed genes with adjusted p-value < 0.001 and |log<sub>2</sub> fold-change| > 1.5 (> 2.82 fold) are shown across different patient groups by viral level and presence of other viral pathogens. (b) Intersection heatmap of differentially expressed genes across different comparisons with genes in x-axis and comparisons in y-axis, with a core set of up-regulated genes (orange) distinct from the set of down-regulated genes (purple), compared to genes that are not significantly differently expressed (grey) in any comparison (Limma voom, q-value < 0.01, |log<sub>2</sub>FC| > 0.58).



**Supplemental Figure 13. ACEi/ARBs Multivariate and Comparative Analyses.** Comparison of the effects of different ACEIs (benazepril, enalapril, lisinopril, ramipril) and ARBs (losartan, olmesartan, telmisartan, valsartan) when compared to patients exposed to other ACEIs or ARBs, respectively. There were no significant findings amongst the different ACEIs. However, among the ARBs we found that olmesartan was associated with a significant increase in mortality (HR=2.3 CI: 1.1-4.9, p=3.05E-02).

**Supplementary Table 1. Metatranscriptome Profiles of All Samples.**

Appended.

**Supplementary Table 2. Taxonomic Mis-assignment Filter.**

Appended.

**Supplementary Table 3. Differentially Expressed Genes in SARS-CoV-2 +/- patients.** Appended.

**Supplemental Table 4. Gene Ontology Pathways.**

Appended.

**Supplementary Table 5. SARS-CoV-2 Rapid Colorimetric LAMP Detection Test: N Primers Specificity**

Appended.

**Supplementary Table 6. Logistic Regression Coefficient**

Appended.

**Supplementary Table 7. Hypertension Logistic Regression Coefficient**

Appended.

**Supplementary Table 8. COX Regression Coefficients**

Appended.

**Supplementary Table 9. Hypertension COX Regression Coefficients**

Appended.

**Supplementary Table 10: GISAID COVID-19 Acknowledgments**

Appended.