

## Longitudinal virological changes and underlying pathogenesis in hospitalized COVID-19 patients in Guangzhou, China

Zhengtu Li<sup>1†</sup>, Yinhu Li<sup>4†</sup>, Ruilin Sun<sup>3†</sup>, Shaoqiang Li<sup>1†</sup>, Lingdan Chen<sup>1†</sup>, Yangqing Zhan<sup>1†</sup>, Mingzhou Xie<sup>2</sup>, Jiasheng Yang<sup>3</sup>, Yanqun Wang<sup>1</sup>, Airu Zhu<sup>1</sup>, Guoping Gu<sup>1</sup>, Le Yu<sup>2</sup>, Shuaicheng Li<sup>4</sup>, Tingting Liu<sup>2</sup>, Zhaoming Chen<sup>1</sup>, Wenhua Jian<sup>1</sup>, Qian Jiang<sup>1</sup>, Xiaofen Su<sup>1</sup>, Weili Gu<sup>1</sup>, Liyan Chen<sup>1</sup>, Jing Cheng<sup>1</sup>, Jincun Zhao<sup>1</sup>, Wenju Lu<sup>1</sup>, Jinping Zheng<sup>1</sup>, Shiyue Li<sup>1</sup>, Nanshan Zhong<sup>1</sup> & Feng Ye<sup>1\*</sup>

<sup>1</sup>State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, China;

<sup>2</sup>Beijing YuanShengKangTai (ProtoDNA) Genetech Co., Ltd., Beijing 100190, China;

<sup>3</sup>Department of Pulmonary and Critical Care Medicine, Guangdong Second Provincial General Hospital, Guangzhou 510317, China;

<sup>4</sup>Department of Computer Science, City University of Hong Kong, Hong Kong 999077, China

Received January 29, 2021; accepted April 6, 2021; published online April 28, 2021

Prolonged viral RNA shedding and recurrence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in coronavirus disease 2019 (COVID-19) patients have been reported. However, the clinical outcome and pathogenesis remain unclear. In this study, we recruited 43 laboratory-confirmed COVID-19 patients. We found that prolonged viral RNA shedding or recurrence mainly occurred in severe/critical patients ( $P < 0.05$ ). The average viral shedding time in severe/critical patients was more than 50 days, and up to 100 days in some patients, after symptom onset. However, chest computed tomography gradually improved and complete absorption occurred when SARS-CoV-2 RT-PCR was still positive, but specific antibodies appeared. Furthermore, the viral shedding time significantly decreased when the A1,430G or C12,473T mutation occurred ( $P < 0.01$  and  $FDR < 0.01$ ) and increased when G227A occurred ( $P < 0.05$  and  $FDR < 0.05$ ). High IL1R1, IL1R2, and TNFRSF21 expression in the host positively correlated with viral shedding time ( $P < 0.05$  and false discovery rate  $< 0.05$ ). Prolonged viral RNA shedding often occurs but may not increase disease damage. Prolonged viral RNA shedding is associated with viral mutations and host factors.

### SRAS-CoV-2, longitudinal virological changes, viral genome, host gene expression

**Citation:** Li, Z., Li, Y., Sun, R., Li, S., Chen, L., Zhan, Y., Xie, M., Yang, J., Wang, Y., Zhu, A., et al. (2021). Longitudinal virological changes and underlying pathogenesis in hospitalized COVID-19 patients in Guangzhou, China. *Sci China Life Sci* 64, 2129–2143. <https://doi.org/10.1007/s11427-020-1921-5>

## INTRODUCTION

Both local and imported cases of coronavirus disease 2019 (COVID-19) have been identified in China (Chinese Center for Disease Control and Prevention, 2020a). The number of confirmed COVID-19 cases reported all over the world has

risen exponentially. As of 23 November 2020, 58,892,589 confirmed COVID-19 cases and 1,388,240 deaths had been reported and the World Health Organization (WHO) had indicated a very high global risk level for this disease (World Health Organization, 2020c). Some asymptomatic carriers of COVID-19 have no respiratory tract symptoms but can still potentially transmit the virus (Gandhi et al., 2020; Li et al., 2020b; Rothe et al., 2020); China's current epidemic pressure comes from asymptomatic carriers. Therefore, COVID-

<sup>†</sup>Contributed equally to this work

\*Corresponding author (email: [tu276025@gird.cn](mailto:tu276025@gird.cn))

19 is still actively investigated in China; effectively controlling and preventing further development of the global epidemic is a priority.

Understanding the evolution and transmission patterns of a virus after it enters a new population includes understanding the kinetics of viral shedding, the contagiousness of a virus, and the effect of viral mutation on clinical outcomes. Understanding the viral evolution and transmission patterns is crucial for the adoption of effective strategies for disease control and prevention and for enhancing clinical treatment (Faria et al., 2018; Grubaugh et al., 2017; Ladner et al., 2019; Liu et al., 2020). Several studies have attempted to elucidate the genomic epidemiology of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Kasibhatla et al., 2020; Lu et al., 2020; Phan, 2020). Furthermore, some reports also focus on prolonged viral shedding and seroconversion (Liu et al., 2020). However, analyses of the correlation between virus evolution, prolonged viral shedding, and the occurrence and development of disease are still lacking. Initial outbreaks in China included approximately 18.6% of cases with severe/critical courses (Guan et al., 2020). This severe/critical situation may result from the viral recognition receptor, which is expressed predominantly in the lung (Li et al., 2020a). However, the correlation between viral mutations and disease severity still needs to be explored.

Therefore, we conducted a cohort study including 43 hospitalized confirmed COVID-19 patients in Guangzhou to longitudinally investigate viral shedding, viral mutations, and seroconversion of specific IgM and IgG antibodies against SARS-CoV-2. Moreover, we present an optimized nanopore target sequencing-based clinical framework (Charalampous et al., 2019; Faria et al., 2017; Rhee and Burns, 2007) for SARS-CoV-2 genome and mutation detection, which was verified by Sanger sequencing. We combined the phylogenetic relationships of viruses, host gene expression profiling, and longitudinal clinical disease data to investigate the genetic evolution and epidemiology of SARS-CoV-2, the function of the mutant genes, and the influence of virus evolution on the occurrence and development of disease. Our results provide valuable information for the prevention and control of COVID-19 in China and globally.

## RESULTS

### Demographics and baseline clinical characteristics

Between January and March 2020, 43 patients with a laboratory-confirmed diagnosis of COVID-19 were enrolled in the study; 19 patients were categorized as mild/moderate, 5 as severe, and 19 as critical on admission. The mean age of all patients was 43 years and 44.2% were male. Critical patients were significantly older than mild/moderate patients

(63 vs. 43 years,  $P < 0.001$ ). Severe or critical patients had at least one underlying comorbidity; diabetes and hypertension were the most common comorbidities. Viral shedding time was significantly prolonged in critical patients compared with severe or mild/moderate patients (88 vs. 68 and 19 days,  $P < 0.05$ ). Leukocytes, neutrophils, and neutrophil-to-lymphocyte ratio were significantly elevated in critical patients compared with mild/moderate patients, but absolute lymphocyte counts and total counts of CD3<sup>+</sup> T, CD4<sup>+</sup> T, and CD8<sup>+</sup> T cells were significantly decreased in critical patients (Table 1).

### Virological and sero-antibody analysis indicates active replication of SARS-CoV-2 in a new population, based on a longitudinal investigation

Although the positive detection of SARS-CoV-2 in clinical specimens from the upper respiratory tract has previously been described (Young et al., 2020; Zou et al., 2020), the detection sensitivity was higher in the lower respiratory tract and fecal specimens. Therefore, to investigate the active replication of SARS-CoV-2 in a new population, we included upper respiratory tract, lower respiratory tract, and fecal specimens at different times of hospitalization in our analysis.

The viral load was significantly higher and the viral shedding time was significantly longer in lower respiratory tract specimens compared with upper respiratory tract specimens. Viral shedding persistently tested positive in sputum specimens after nasopharyngeal testing was negative (Figure 1A–C). The average viral load in sputum was  $10.84 \times 10^7$  copies per mL, with a maximum of  $8.7 \times 10^8$  copies per mL. The average viral RNA load was  $10.2 \times 10^6$  copies per swab, and the maximum load was  $1.1 \times 10^8$  copies per swab. Interestingly, persistent fecal viral shedding occurred in some cases and one case showed persistently positive fecal specimens (the maximum load was  $1.1 \times 10^8$  copies per mL), even after nasopharyngeal and sputum tests were negative (Figure 1D–F). More interestingly, the viral RNA of SARS-CoV-2 can be detected in other clinical specimens, including serum, plasma, eye conjunctival swab, urine, saliva, throat washing, and even gastric juice during the early days of symptom onset (data not shown).

Viral shedding time was significantly prolonged in severe/critical patients compared with mild/moderate patients ( $P < 0.05$ ) and long viral shedding was mainly found in severe/critical patients. Among the severe and critical patients, the viral shedding time was lengthened more than 50 days after the onset of symptoms. Even though patients persistently tested positive up to 100 days after the onset of symptoms, the clinical symptoms recovered at this point, which was not reported in the previous reports (Figure 1D–F).

**Table 1** Demographics and clinical characteristics of COVID-19 patients

Variable	Mild/moderate	Severe	Critical	P value <sup>a)</sup>
	(N=19)	(N=5)	(N=19)	
Age (years)	43 (31–57)	50 (35–78)	63 (50–70)	0.094
Sex, Male, n (%)	8 (47.1)	4 (80.0)	12 (70.6)	0.296
Duration of hospital admission, days	19 (17–28)	68 (49–82)	88 (59–122)	<0.05
Viral shedding duration, days	18 (10–30)	54 (40–64)	59 (50–81)	<0.05
Comorbidity				<0.05
Current smoker	0	1 (20)	5 (29.4)	<0.05
Hypertension	2 (11.8)	0	5 (29.4)	0.398
Diabetes	0	1 (20)	5 (29.4)	<0.05
Heart disease	1 (5.9)	0	3 (17.6)	0.601
Cerebrovascular disease	1 (5.9)	0	0	1.000
Chronic lung disease	1 (5.9)	1 (20)	1 (5.9)	1.000
Chronic kidney disease	0	0	1 (5.9)	1.000
Chronic liver disease	0	0	2 (11.8)	0.485
Laboratory examination				
White blood cells ( $\times 10^9$ cells L <sup>-1</sup> )	4.32 (3.51–4.60)	4.70 (3.65, 5.09)	9.60 (7.25, 11.75)	<0.05
Neutrophil absolute count ( $\times 10^9$ cells L <sup>-1</sup> )	2.26 (1.72, 3.15)	3.43 (2.80, 4.00)	8.70 (6.45, 11.10)	<0.05
Neutrophil ratio (neut %)	61.25 (46.43, 68.13)	75.20 (37.78, 82.05)	92.60 (84.15, 93.80)	<0.05
Lymphocyte absolute count ( $\times 10^9$ cells L <sup>-1</sup> )	1.36 (1.07, 1.94)	0.50 (0.99, 0.40)	0.40 (0.25, 0.50)	<0.05
T cell count	562 (389, 800)	528 (175, 740)	310 (174, 355)	<0.05
Th cell count	318 (251, 440)	240 (75, 481)	220 (108, 308)	0.067
Ts cell count	242 (153, 290)	219 (80, 271)	63 (29, 124)	<0.05

a) P value means a comparison of mild/moderate and critical cases. Because the number of severe cases was small, no statistical comparison was made.

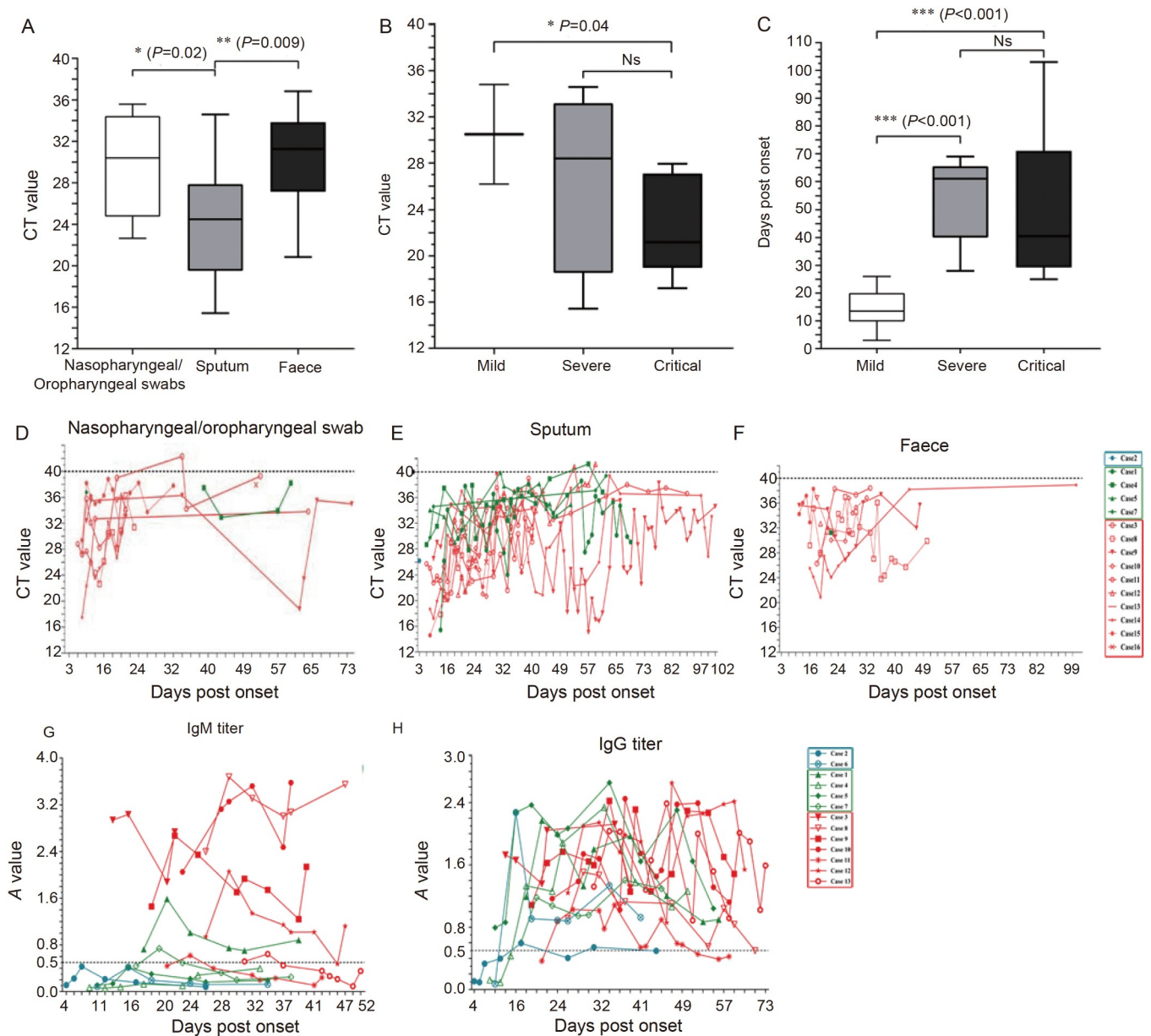
Serum IgM and IgG antibodies were detected by enzyme-linked immunosorbent assay (ELISA) (Figure 1G and H). In general, the IgM and IgG antibodies appeared simultaneously on day 11 after the onset of symptoms. The IgM antibody test turned negative first, but some severe/critical patients persistently tested positive for IgM and IgG antibodies even more than 90 days after the onset of symptoms. Prolonged viral shedding was detected in severe/critical COVID-19 patients after clinical symptoms had resolved and specific antibodies emerged, especially IgM antibodies. Second, the IgM antibody turned negative when the viral RNA turned negative. However, when the viral RNA turned positive again in “recurrent” cases, the patients also tested positive for IgM antibodies.

#### Dynamic virological and neutralizing antibody analysis of COVID-19 patients with prolonged viral RNA shedding or recurrence provides proof of replication sites and infectivity of SARS-CoV-2

Among the 43 patients, six patients (13.95%) with SARS-CoV-2 exhibited prolonged viral RNA shedding or recur-

rence, and the viral shedding time was more than 50 days after symptom onset in one patient with prolonged viral shedding. These were all severe/critical COVID-19 patients (Figure 2). Therefore, to better understand the replication sites and infectivity of SARS-CoV-2, these six patients were analyzed separately.

One question we were concerned about is whether there is still infection in those patients in the later period of prolonged viral shedding or when recurrence occurs. Our findings showed that although the real-time polymerase chain reaction (RT-PCR) of SARS-CoV-2 was positive, live virus isolations from sputum samples on multiple occasions were negative (Figure 2). The neutralizing antibody tested for SARS-CoV-2 was positive, which is conducive to clearing the live virus. These results suggest that the virus detected with RT-PCR may be a virus fragment, not a live virus, during prolonged shedding. To confirm that these viruses were fragmented, the coverage of SARS-CoV-2 was tested using optimized nanopore target sequencing. The results showed that the coverage of the virus gradually decreased as the disease improved, except for two cases (Cases 5 and 11) (Figure S1 in Supporting Information). The coverage de-



**Figure 1** Detailed virological and sero-antibody analysis of COVID-19 patients. A, Comparison of viral loads among different types of specimens. B, Comparison of viral loads among different clinical disease types. C, Comparison of the shedding time among different clinical disease types. D–F, Dynamic changes in longitudinal virology observed in nasopharyngeal/oropharyngeal swabs, sputum, and fecal specimens. G and H, The dynamic changes in specific IgM and IgG antibodies against SARS-CoV-2. \*\*\*,  $P < 0.001$ ; \*\*,  $P < 0.01$ ; \*,  $P < 0.05$ .

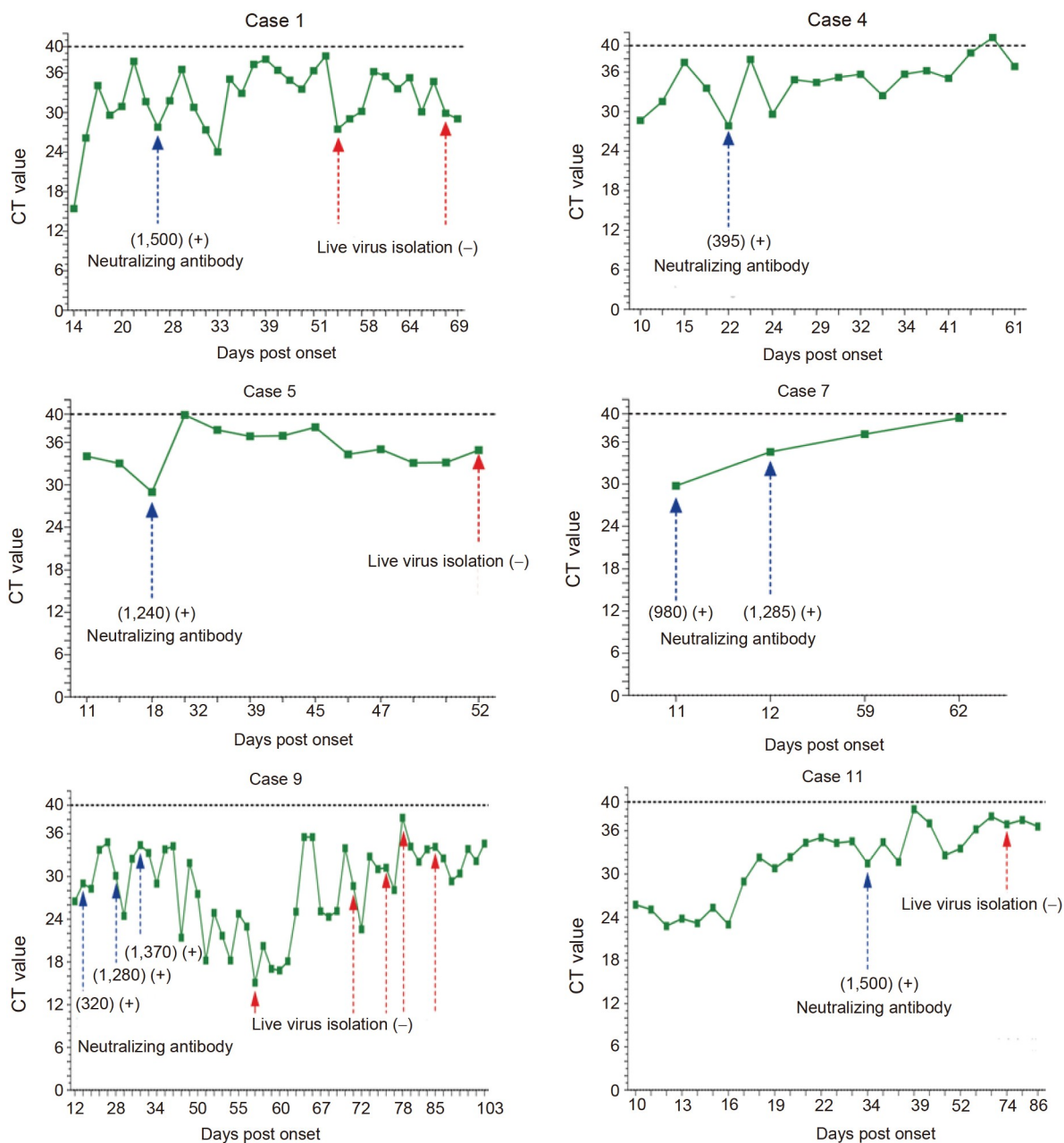
crease indicates that the virus was not intact and was likely fragmented. Interestingly, the two recurrent cases (Cases 4 and 7) tested negative at the time of discharge from the hospital. These patients had close contact with family members after recurrence before being hospitalized again; however, no family members were infected.

### Clinical effects of prolonged viral RNA shedding or recurrence on COVID-19 patients

Prolonged viral RNA shedding or recurrence has been ob-

served in some COVID-19 patients. Furthermore, the viruses with RT-PCR positivity may be virus fragments, not live viruses. However, the question remains, does viral RNA shedding or recurrence have adverse effects on COVID-19 patients? Therefore, we compared patients with or without prolonged viral RNA shedding or recurrence. Lengths of hospital stays were increased in patients with prolonged viral shedding or recurrence due to the management of epidemic prevention and control. In contrast, other factors, including the ratio of mechanical ventilation, the proportion of severe or critical cases, and mortality, were not different between



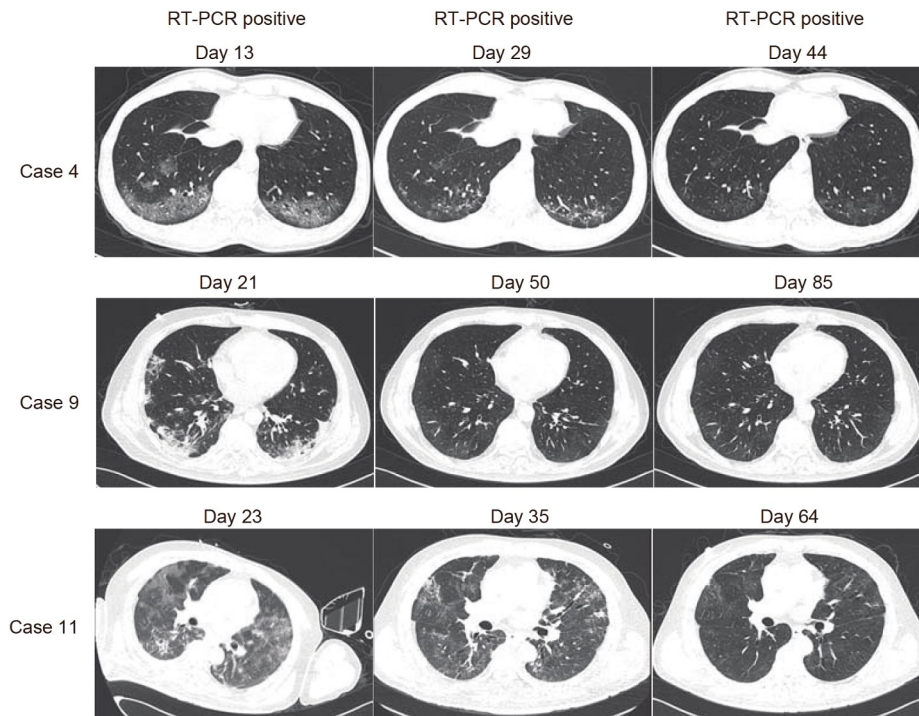


**Figure 2** Dynamic virological and neutralizing antibody analysis of COVID-19 patients. Dynamic changes in the virus, live viral cultures, and neutralizing antibody tests are shown in the six prolonged viral RNA shedding or recurrence patients, including Cases 1, 4, 5, 7, 9, and 11. (+) means neutralizing antibody testing positive when the number >100.

the two groups (Table S1 in Supporting Information). Moreover, chest computed tomography (CT) showed that the ground glass shadow in the lung caused by SARS-CoV-2 infection, gradually improved as the disease progressed, even exhibiting complete absorption when the RT-PCR of SARS-CoV-2 was still positive in some COVID-19 patients with prolonged viral RNA shedding or recurrence (Figure 3). Furthermore, there were no significant differences in clinical features between patients with or without prolonged viral RNA shedding or recurrence (Table S1 in Supporting Information).

### Analysis of SARS-CoV-2 mutations and their relationships with viral shedding times explains the cause of prolonged viral RNA shedding or recurrence based on the virus itself

Aligning the sequencing data from patients to the SARS-CoV-2 genome revealed 128 single-nucleotide polymorphisms (SNPs), 122 of which were located in gene regions (Figure 4A). We then focused on the 11 high-frequency SNPs ( $\geq 6$  samples, Figure 4B): eight SNPs were located in the ORF1ab gene (loci: 865, 1,430, 7,392, 8,782, 12,473,



**Figure 3** Imaging of chest computed tomography (CT) scan. The chest CT of COVID-19 patients with prolonged viral RNA shedding or recurrence (Cases 4, 9, and 11) showed different courses of illness, including early, middle, and convalescent periods. Furthermore, the chest CT showed positive SARS-CoV-2 when the RT-PCR test was also positive.

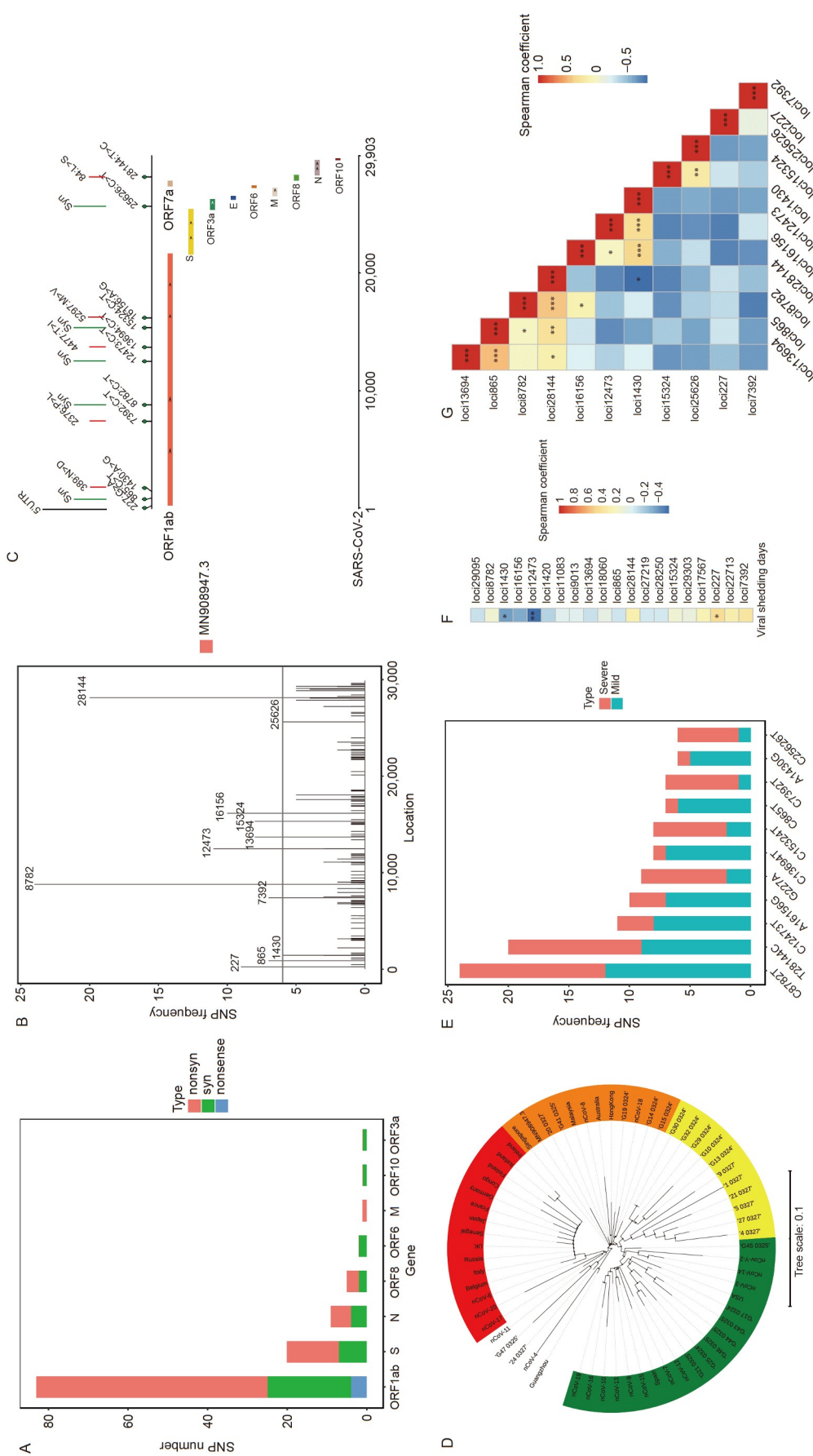
13,694, 15,324, and 16,156), one SNP was located in the ORF3a gene (loci: 25,626), one SNP was located in the ORF8 gene (loci: 28,144), and one SNP was located in the 5' UTR (loci: 227, [Figure 4C](#)). Furthermore, four high-frequency SNPs in ORF1ab were nonsynonymous mutations and the mutation that occurred on loci 13,694 changed a threonine to isoleucine ([Figure 4C](#)). After detecting the corresponding amino acid for the SNP C13,694T, we hypothesized that the protein structure and the function of the RNA-directed RNA polymerase (RdRP) may be affected by this mutation (Figure S2 in Supporting Information). A phylogenetic tree was constructed based on SARS-CoV-2 strains from our COVID-19 patients and SARS-CoV-2 strains from other countries. Four major clades were discovered ([Figure 4D](#)). One clade included the SARS-CoV-2 strains from Europe and Africa (marked with red); one clade included the SARS-CoV-2 strains from Asia (except Japan), Australia, and some of our patients; and the two other clades contained the SARS-CoV-2 virus strains from most of our patients (marked with orange and green).

After detecting the distribution of high-frequency SNPs in COVID-19 patients, we discovered that G227A, C7,392T, C15,324T, and C25,626T occurred more frequently in severe patients. In contrast, C865T, A1,430G, and C13,694T occurred more frequently in mild patients ([Figure 4E](#)). To further explore the relationships between the high-frequency SNPs and viral shedding time, we used a Spearman coefficient analysis.

The viral shedding time in COVID-19 patients significantly decreased when the A1,430G ( $r=-0.382$ ,  $P<0.05$  and false discovery rate (FDR) $<0.05$ ) or C12,473T ( $r=-0.555$ ,  $P<0.01$  and FDR $<0.01$ ) mutation occurred. In contrast, the viral shedding time in COVID-19 patients increased when the G227A mutation occurred ( $r=0.437$ ,  $P<0.05$  and FDR $<0.05$ , [Figure 4F](#)). In addition, the potential co-mutations were also analyzed: C865T and C13,694T ( $r=0.612$ ,  $P<0.001$  and FDR $<0.001$ ), C8,782T and T28,144C ( $r=0.455$ ,  $P<0.001$  and FDR $<0.001$ ), A1,430G and C12,473T ( $r=0.629$ ,  $P<0.001$  and FDR $<0.001$ ), and A16,156G ( $r=0.809$ ,  $P<0.001$  and FDR $<0.001$ ) were probably co-mutations in SARS-CoV-2 ([Figure 4G](#)). These potential co-mutations provide a deeper understanding of the SARS-CoV-2 evolutionary routes. To verify the precision of the discovered SNPs, we used Sanger sequencing to verify the SNPs (G227A, C865T, C7,392T, C13,694T, C15,324T, and C25,626T) in some samples. C865T, C13,694T, and C15,324T were confirmed, but G227A, C7,392T, and C13,694T were not detected in the selected samples (Figure S3 in Supporting Information).

#### Correlation analysis of blood gene expression profiling and viral shedding explains the cause of prolonged viral RNA shedding or recurrence based on host factors

To elucidate the host factors leading to prolonged viral RNA



**Figure 4** Analysis of single-nucleotide polymorphisms (SNPs), the phylogenetic tree, and correlations between SNPs and viral shedding times. **A**, The distributions of all SNPs in the genetic regions of SARS-CoV-2. The synonymous, nonsense, and nonsynonymous mutations are represented by blue, gray, and red, respectively. **B**, The frequency of SNPs in SARS-CoV-2 strains from 43 COVID-19 patients. The loci of high-frequency mutations are marked. **C**, The genes in which high-frequency SNPs are located. The corresponding amino acids are linked to the high-frequency SNPs by lines. The black, green, and red lines indicate that the SNPs are inter-genetic, synonymous, and nonsynonymous mutations, respectively. **D**, The phylogenetic tree is composed of four major clades. The viral strains in the same clade are marked with the same color. **E**, The distribution of SNPs in severe and mild patients. The red and blue bars represent the frequency of the SNPs in severe and mild patients, respectively. **F**, The correlations between SNPs and viral shedding time. The red and blue colors represent positive and negative correlations, respectively. **G**, The correlations among SNPs. The red and blue colors represent positive and negative correlations, respectively. \*,  $P < 0.05$  and  $FDR < 0.05$ ; \*\*,  $P < 0.01$  and  $FDR < 0.01$ ; \*\*\*,  $P < 0.001$  and  $FDR < 0.001$ .



shedding or recurrence, blood gene expression profiling was performed from samples collected from patients with and without prolonged viral shedding. The intergroup analysis factors were defined as the presence (P group) or absence (NP group) of prolonged viral RNA shedding or recurrence. Based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) gene annotation, 226 KEGG functional pathways from 34 KEGG functional categories at the second level exhibited differential enrichment in the two groups (Figure 5A). Among these differentially enriched KEGG functional categories, the functional item “immune system” contained 22 pathways, among which “PI3K-Akt signaling pathway” and “Cytokine-cytokine receptor interaction” were significantly enriched in the P group ( $P < 0.01$ ,  $FDR < 0.01$ , Figure 5B). Compared with the NP group, the expression of the IL1R1 (interleukin 1 receptor type 1), IL1R2, and TNFRSF21 (tumor necrosis factor receptor superfamily member 21) genes increased significantly in the P group. In contrast, the expression of teneurin transmembrane protein 4 (TENM4) and alpha hemoglobin stabilizing protein (AHSP) decreased ( $\log_2$  fold change,  $P < 0.01$  and  $FDR < 0.01$ , Figure 5C). Applying the Gene Ontology (GO) database, we discovered that genes related to the function “immune system process” exhibited higher expression in the P group (Figure 5D; Table S2 in Supporting Information). Furthermore, 22 GO functional items related to host immune regulation exhibited significantly different enrichment between the P and NP groups (Figure 5E).

According to the KEGG analysis, genes related to immune response-related cell surface receptors (such as GO:0002433 and GO:0002429) were highly expressed in patients with long viral shedding times (Figure 5E). Subsequently, the relationships between genes with differential expression and viral shedding time in COVID-19 patients were explored. The high expression of the IL1R1, IL1R2, and TNFRSF21 genes positively correlated with viral shedding time in patients ( $P < 0.05$  and  $FDR < 0.05$ , Figure 5F). In contrast, the expression of the CCL5 gene negatively correlated with viral shedding time ( $P < 0.01$  and  $FDR < 0.01$ , Figure 5F).

## DISCUSSION

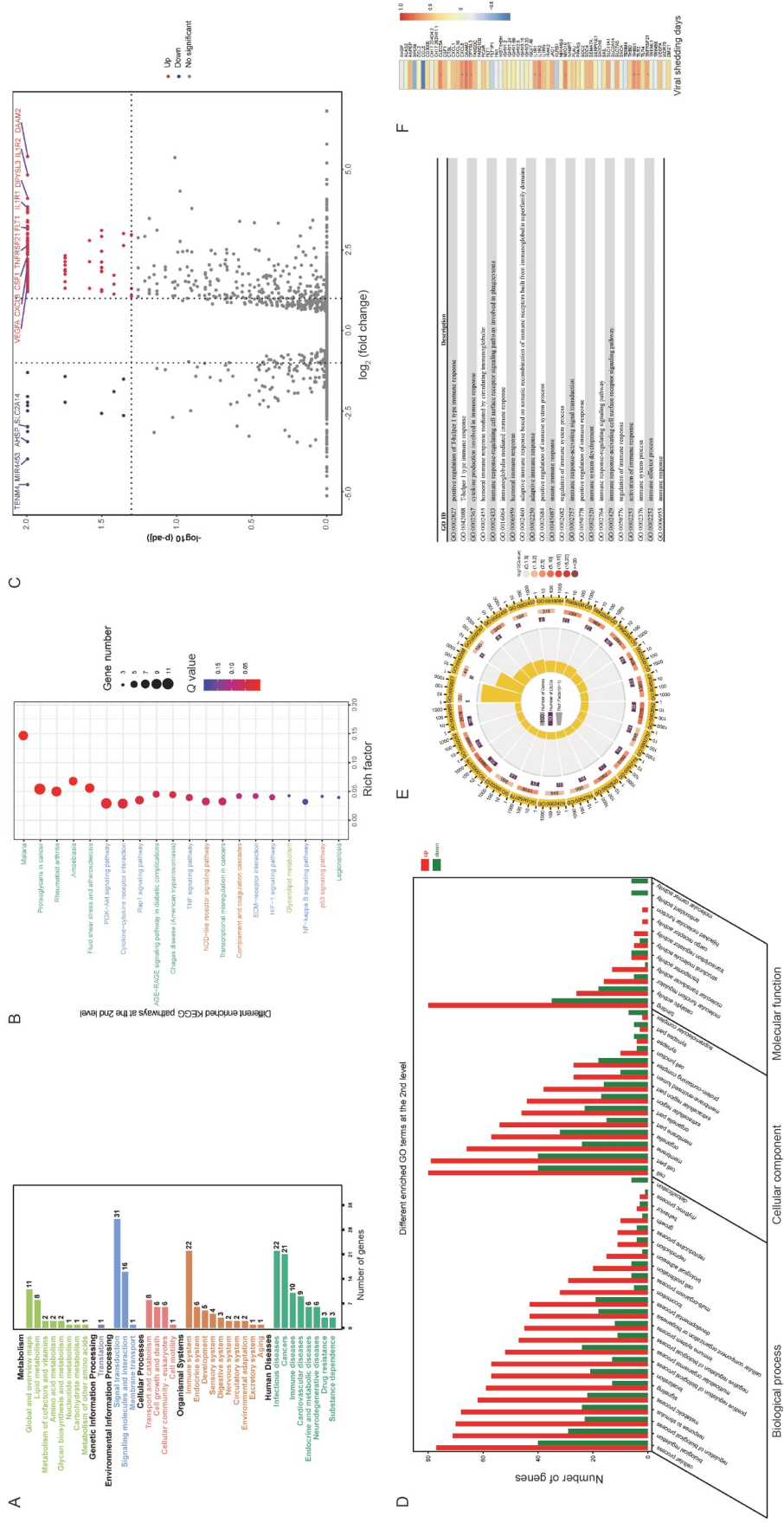
We established a study cohort of hospitalized COVID-19 patients, including mild/moderate, severe, and critical cases, to investigate the active replication, viral mutations, phylogenetics, and specific antibody response to SARS-CoV-2. This study explored the complete cycle of disease, from the onset of symptoms to day 28 after discharge from the hospital. There are several important highlights of this study. (i) The viral load was significantly higher and the viral shedding time was significantly lengthened in lower respiratory tract specimens more than 100 days after symptom onset. The

viral load was prolonged even in stool specimens. (ii) Viral RNA shedding was significantly prolonged in severe and critical patients, especially in recurrent patients. (iii) Prolonged viral RNA shedding was accompanied by the production of specific and neutralizing antibodies; however, live virus isolation was negative. (iv) Viral mutations, including A1,430G, C12,473T, and G227A, and the gene expression of IL1R1, IL1R2, and TNFRSF21 in the host were positively related to prolonged viral RNA shedding or recurrence.

Results from this longitudinal virological investigation indicated that the viral load was significantly higher and the viral shedding time was significantly lengthened in sputum (Figure 1). This finding suggests that lower respiratory tract specimens are more suitable for monitoring and evaluation of COVID-19 infection. This conclusion is consistent with reports concerning avian influenza viruses, such as human H7N9 and H5N6 (Yang et al., 2015). Of note, viral RNA shedding was also prolonged in the stool of some COVID-19 patients (Figure 1), as previously reported (Xiao et al., 2020b). This was not reported for avian influenza virus infection. Furthermore, the stool of COVID-19 patients could be used to isolate the live virus (Xiao et al., 2020b). In addition, we found that virus RNA can also be detected in other specimens, such as eye conjunctival swabs, urine, and even gastric juice. The live virus can also be isolated from tears (Xiao et al., 2020a). Therefore, although the respiratory tract is considered the main method of transmission, other transmission routes should also be avoided, and relevant protective measures are needed. Of course, more research is needed to prove this hypothesis.

One important question is whether prolonged viral shedding or recurrence is infectious. We could find no proof of infectivity in our study. One of the most interesting hypotheses to explain this conclusion is that the virus was inactive at that moment, although the nucleic acid RT-PCR was positive for SARS-CoV-2. Patients can persistently test positive, even more than 100 days after symptom onset in some cases, based on our longitudinal virological investigation (Figures 1 and 2). Previously, the longest shedding time reported was 45–63 days after symptom onset (Liu et al., 2020; Xiao et al., 2020a). During hospitalization, live virus isolations were negative at the late stage of viral shedding or recurrence in our patients, and specific neutralizing antibodies for SARS-CoV-2 existed before the live virus isolation (Figure 2). The antibodies appeared early on day 11 and seroconversion in 100% of patients occurred by day 16 after symptom onset (Figures 1 and 2). These findings are in agreement with other studies; seroconversion in 50% of patients occurred by day 7 and in all patients by day 14 (Wölfel et al., 2020). Simultaneously, no live viruses were isolated after day 7 (Wölfel et al., 2020). Another study also suggested that viruses could only be isolated from respiratory specimens collected within the first 18 days (Liu et





**Figure 5** Gene expression in P and NP patients. A, The KEGG annotation results of genes with different expression levels between the P and NP groups. The bold items on the Y coordinate represent the first level functional categories in the KEGG database and the second level functional categories in the KEGG database are marked with different colors. B, The top 20 enriched functional items at the KEGG second level. The diameters of the bubbles indicated their gene numbers. C, A volcano plot of genes with different expression levels between the P and NP groups. The upregulated and downregulated genes in the P group are marked by red and green, respectively. D, The GO annotation results of genes with different expression levels between the P and NP groups. The upregulated and downregulated genes in the P group are marked by red and green, respectively. E, The enriched immunological functions in the P group after GO annotation. The GO items are marked in the outer circle, the q-values of the differentially enriched GO items are marked in the second circle, and the included gene numbers are indicated by the length of the bars in the third circle. The detailed functions of the GO items are described in the table on the right. F, Spearman correlations between the viral shedding time and the genes with different expression levels in patients with prolonged viral shedding times. Red and blue indicate positive and negative correlations, respectively, between the gene expression and viral shedding time. \*,  $P < 0.05$ , \*\*,  $P < 0.01$  and  $FDR < 0.01$ .

al., 2020). These reports and our results suggest that the virus becomes inactive and may even become a viral fragment after a specific antibody against SARS-CoV-2 is produced in COVID-19 patients, although viral RT-PCR remains positive. Furthermore, we tested whether viruses are fragmented using nanopore target sequencing for the first time. These results showed that the integrity of the virus was progressively compromised and the virus was fragmented during viral shedding (Figure S3 in Supporting Information). More rigorous experiments are needed to support this conclusion. We also found that prolonged viral RNA shedding or recurrence did not have serious effects on the patient, including symptoms and outcomes (Figure 3; Table S1 in Supporting Information), which further confirms that the virus is inactive. Perhaps because of this, the WHO's clinical management of COVID-19 also states that a patient can be removed from isolation when symptoms have been resolved for more than 3 days, rather than based on RT-PCR results (World Health Organization, 2020a).

We also sought to explain why prolonged or recurrent viral shedding occurs in severe and critical patients. We believe this can be explained by both viral and host factors. On the one hand, the virus itself mutates. In this study, 11 high-frequency SNPs were discovered in 43 COVID-19 patients from Guangdong Province, and 3 of them may have significant impacts on the virulence of SARS-CoV-2. First, the mutation at loci 15,324 (C15,324T) (Figure 4) is located in the RdRP of the ORF1ab gene. C15,324T was recognized as a common SNP in COVID-19 patients (Banerjee et al., 2020) and we found that the hosts who were infected by SARS-CoV-2 with this mutation had longer viral shedding times. Thus, we speculate that the mutation might be related to the viral reproductive capacity of SARS-CoV-2. Second, the mutation at loci 13,694 is also located in the RdRP region of ORF1ab. Based on the predicted protein structure of RdRP (Waterhouse et al., 2018), we found that the C13,694T mutation should lead to changes in the protein structure and activity, which might reduce the virulence of SARS-CoV-2. Last, the mutation at loci 865 is located in the ORF1ab gene. Although it was a synonymous mutation, we found that C865T normally occurred in SARS-CoV-2 strains from mild patients, and might co-mutate with C13,694T. In addition, we also discovered mutation hotspots that were reported in previous studies, including C8,782T, C12,473T, and T28,144C (Lu et al., 2020; Pachetti et al., 2020). However, we found that some SNPs could not be verified with Sanger sequencing. We speculate that this discrepancy was caused by the limited tested samples or the biased sequencing error due to target amplification.

The phylogenetic analysis indicated that the SARS-CoV-2 viruses from countries in Europe and Africa are more closely related, and the strains from countries in Asia are more closely related, except for the virus from Japan, which might

be related to infections on the “Diamond Princess” cruise ship. Based on these results, we hypothesize that several SARS-CoV-2 subtypes might exist simultaneously during this epidemic (corresponding to the findings of a previous report (Forster et al., 2020)), but a deeper analysis of these relationships is urgently needed to help limit this epidemic.

On the other hand, host factors may also contribute to the prolonged shedding of the virus. In a previous study, male sex, delayed admission to the hospital after illness onset, and invasive mechanical ventilation were associated with prolonged SARS-CoV-2 RNA shedding (Xu et al., 2020). However, few studies reported the correlation between gene expression in the host and prolonged viral RNA shedding or recurrence. Therefore, we examined host transcriptomics (Figure 5). Based on previous research, we know that the excessive secretion of cytokines in the peripheral blood of COVID-19 patients leads to a hyperimmune response (Chen et al., 2020; Lee et al., 2020). Furthermore, the level of IL-6 in COVID-19 patients positively correlates with disease severity (Chen et al., 2020). In this study, COVID-19 patients with a prolonged viral shedding time exhibited higher expression of the IL1R1 and IL1R2 genes compared to patients with a non-prolonged viral shedding time. These gene expression changes explain the elevated activation of the IL-1 signaling pathway in these patients. In addition, the AHSP gene plays an important role in maintaining the stability and solubility of alpha hemoglobin (Viprakasit et al., 2004). Low expression of AHSP may reduce the stability of hemoglobin in patients with a prolonged viral shedding time, which decreases the oxygen-carrying capacity of red blood cells. Moreover, the relationship between viral shedding time and gene expression in patients confirms the phenomenon of cytokine storms in these COVID-19 patients. The high expression of the IL1R1 and IL1R2 genes aggravate the disease severity. Lee et al. (2020) highlighted the role of type I interferons in the overreaction of the immune system. We also discovered the high expression of the TNFRSF21 gene in patients with prolonged viral shedding times, which may explain the high expression of TNF in these patients. In general, these results support the over-activated type I interferon expression hypothesis in patients with prolonged viral shedding times, which provides a basis for the treatment of immunological suppression in COVID-19 patients.

This study had some limitations. First, the case size was small, especially for the cases with complete shedding times, and the number of follow-up cases needs to be supplemented. Second, since severe and critical cases are referred to our hospital halfway through the disease, early specimens are lacking. Thus, we were not able to accurately judge how the specific antibody of SARS-CoV-2 affects the activity of the virus. Third, this study successfully utilized nanopore target sequencing for clinical detection and genomic analysis of SARS-CoV-2. However, there are some defects to this sys-

tem. First, during the process of target amplification, biased amplification errors might occur due to the primer sequences, which would affect the accuracy of SNPs. Second, although Sanger sequencing was used to confirm the SNPs in this study, the impact of these SNPs on viral virulence needs to be further verified by functional experiments. Last, the sample numbers/ratio for the mild/moderate, severe, and critical groups only reflect the number of enrolled COVID-19 patients in this study and do not represent the clinical epidemiological investigation results in reality.

In summary, prolonged viral RNA shedding or recurrence usually occurs in severe or critical COVID-19 patients. Specimens from the lower respiratory tract or fecal specimens are more suitable for monitoring and evaluating viral shedding. The virus was not infectious at the late stage of viral shedding or recurrence because it was fragmented. Both viral and host factors contribute to the prolonged viral RNA shedding, including viral mutations and changes in host gene expression. Longitudinal virological investigations are very important, especially for new viruses. This information is useful in understanding virological dynamics during the process of illness, as well as for the clinical management and treatment of patients with COVID-19.

## MATERIALS AND METHODS

### Patient recruitment

Between January and March 2020, COVID-19-confirmed patients, who agreed to participate in this investigation for scientific purposes, were recruited from the First Affiliated Hospital of Guangzhou Medical University and Guangdong Second Provincial General Hospital, Guangzhou, China. These patients were admitted to the hospitals designated for COVID-19 patients by the health authorities. The following definitions were used in this study ([Chinese Center for Disease Control and Prevention, 2020b](#); [World Health Organization, 2020b](#)). Confirmed COVID-19 patients were defined as patients who had clinical symptoms and a positive RT-PCR test from respiratory, digestive, or blood samples for SARS-CoV-2. Mild/moderate COVID-19 cases were defined as patients with fever and respiratory symptoms and imaging features of pneumonia. Severe COVID-19 cases were defined as patients with the following conditions: shortness of breath and/or respiratory rate  $\geq 30$  breathes per min, peripheral oxygen saturation  $\leq 93\%$  at rest, or partial pressure of oxygen/fraction of inspired oxygen  $\leq 300$  mmHg. Critical COVID-19 cases were defined as patients with the following conditions: respiratory failure requiring mechanical ventilation, shock, or other organ failure requiring intensive care unit care. Recurrence was defined as two consecutive negative PCR tests with an interval of 24 h.

### Specimen collection

Clinical specimens, including oropharyngeal swabs and sputum samples, were obtained in accordance with WHO and Chinese guidelines ([Chinese Center for Disease Control and Prevention, 2020b](#); [World Health Organization, 2020b](#)). Oropharyngeal/nasopharyngeal swab specimens were collected with synthetic fiber swabs. Each swab was then placed into a separate sterile tube containing 3 mL of viral transport medium. Sputum specimens were collected in sterile specimen containers. Anal swabs and/or fecal specimens were collected in sterile specimen containers. The specimens were stored between 2°C and 8°C until ready for shipment to the standardized and accredited laboratory of the Guangdong Centers for Disease Control and Prevention (CDC). Specimens from COVID-19 patients were collected at different times after hospitalization and up until 28 days after negative RT-PCR (Figure S4 in Supporting Information).

### Nucleic acid isolation and RT-PCR detection of SARS-CoV-2

Viral lysis buffer (AVL, 560  $\mu\text{L}$ ) was added to each 140- $\mu\text{L}$  sputum or oropharyngeal swab specimen in a biosafety level 3 laboratory. Total RNA was extracted using a QIAamp<sup>®</sup> Viral RNA Mini kit (Qiagen, Germany) according to the manufacturer's instructions. Samples were eluted with 60  $\mu\text{L}$  elution buffer. A 5  $\mu\text{L}$  aliquot of RNA was used for real-time RT-PCR, which targeted the NP gene using a HiScript<sup>®</sup> II U+ One Step qRT-PCR probe kit (Vazyme, Nanjing, China). The final reaction mix contained 200  $\text{nmol L}^{-1}$  primer and 100  $\text{nmol L}^{-1}$  probe. Real-time RT-PCR was performed using the following conditions: 55°C for 15 min and 95°C for 30 s, followed by 45 cycles of amplification at 95°C for 10 s and 60°C for 30 s.

### Isolation and culture of SARS-CoV-2

Vero E6 cells were used for virus isolation. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum (FBS). An RT-PCR-positive sputum specimen was saved in viral transport media (DMEM containing 1% bovine serum albumin, 15  $\mu\text{g mL}^{-1}$  amphotericin, 100 units per mL penicillin G, and 100  $\mu\text{g mL}^{-1}$  streptomycin). Before virus isolation, the sample was filtered with a 0.45- $\mu\text{m}$  strainer and diluted 1:10 with DMEM containing 2% FBS and antimicrobial drugs. The cells were infected at 37°C for 1 h. The inoculum was removed and replaced with a fresh culture medium. The cells were incubated at 37°C and observed daily for a cytopathic effect. If there was no obvious cytopathic effect until day 6 post-infection, the cells and supernatant were scraped up, freeze-thawed once, and overlaid on new cells for a second-round passage.



## Design and detection of nanopore target sequencing for SARS-CoV-2

A new-generation pathogenic gene sequencing platform (fourth generation), called nanopore target sequencing, and the analysis platform for SARS-CoV-2 were developed. This platform was applied to overcome the methodological deficiencies of previous related studies. The detailed SARS-CoV-2 nanopore target sequencing procedure is presented in Figure S5 in Supporting Information.

### Bioinformatic methods

#### *Data filtering*

Data in FAST5 format were generated by nanopore sequencing and converted into FASTQ format with Guppy basecaller (version 3.0.3). NanoFilt (version 1.7.0) was used to filter the raw FASTQ data according to the following criteria: read lengths longer than 100 bp after the removal of the adapter sequences and an overall quality higher than 10. The clean data were retained for subsequent analysis.

#### *Detection of genomic mutations*

Using Minimap2 (version 2.17-r941) (Li, 2018), we aligned the clean data to the SARS-CoV-2 reference genome (NCBI MN908947.3). Based on the alignment results, we adopted BCFtools (version 1.8) to detect the genetic mutations in each sample. The candidate mutations were retained if their average quality was greater than 15 and the coverage depth was greater than 20. To avoid errors brought by PCR amplification, mutations were excluded if they were located in primer regions by using a previously developed method (Yu et al., 2020). The locations and functions of the final mutations were annotated based on the SARS-CoV-2 reference genome.

#### *Phylogenetic tree construction*

In addition to the Wuhan SARS-CoV-2 virus strain (MN908947.3), we also gathered published information on other SARS-CoV-2 strains from the GISAID website (<https://www.gisaid.org/>), including strains from Guangzhou, China (EPI\_ISL\_412967), Japan (EPI\_ISL\_410531), Malaysia (EPI\_ISL\_416866), Iceland (EPI\_ISL\_417852), Finland (EPI\_ISL\_418385), Ireland (EPI\_ISL\_418516), Belgium (EPI\_ISL\_418656), the United Kingdom (EPI\_ISL\_418667), Hong Kong, China (EPI\_ISL\_419214), Italy (EPI\_ISL\_419254), Japan (EPI\_ISL\_419296), Germany (EPI\_ISL\_419551), Australia (EPI\_ISL\_420013), Democratic Republic of the Congo (EPI\_ISL\_420032), France (EPI\_ISL\_420061), Senegal (EPI\_ISL\_420074), Russia (EPI\_ISL\_420080), Singapore (EPI\_ISL\_420105), and Spain (EPI\_ISL\_420119). Taking MN908947.3 as the reference sequence, the SNPs in the published SARS-CoV-2

strains were detected using Mummer (version: 3.23) (Kurtz et al., 2004). The SNP information was used to establish phylogenetic relationships using Mega X (version 10.0.4) by applying the neighbor-joining method (bootstrap: 1,000), and these relationships were visualized using iTOL (version 5.5.1) (Letunic and Bork, 2007).

#### *Detection of mutation-related protein structures*

After SNP annotation, the location of amino acids was specified for the nonsynonymous mutations on the SARS-CoV-2 RdRP gene. Based on the primary, secondary, and tertiary structures of RdRP on the Research Collaboratory for Structural Bioinformatics protein database (<http://www.rcsb.org/>), the functional domain of the altered amino acids was analyzed. Then, the secondary structure was analyzed and labeled in the tertiary structure.

#### *Coverage of the SARS-CoV-2 genome*

The depths of the bases in the SARS-CoV-2 genome were calculated after aligning the nanopore sequencing reads to the SARS-CoV-2 genome using Minimap2 (version 2.17-r941) (Li, 2018). Viral genomic bases with depths greater than 1 were kept and counted for all the samples and the coverage of the SARS-CoV-2 genome was calculated for the samples.

### Sanger sequencing

To identify the genetic type of loci 13,694 (ORF1ab gene) and loci 227 (5'-untranslated region (5'UTR)), PCR amplification for these loci was performed: forward primer 5'-GCTGATGCACAATCGTTTTT-3' and reverse primer 5'-TCATCACAACAATTGTATGTGACA-3' for loci 13,694; and forward primer 5'-AACAAACCAACCAACTTTCGA-3' and reverse primer 5'-TCCGAACGTTTGATGAACAC-3' for loci 227. The PCR product was sequenced by GBI (Shenzhen, China).

### Antibody detection for SARS-CoV-2

#### *ELISA*

Anti-human IgM ( $\mu$ -chain specific) antibody or the N protein of SARS-CoV-2 (IgG) was used as the coating. The plasma of patients was diluted at 1:100 for testing. HRP-labeled N protein of SARS-CoV-2 (IgM) or anti-human IgG (H+L) antibody labeled with HRP was used as the secondary antibody. The color was developed using tetramethylbenzidine and the reaction was terminated with  $H_2SO_4$ . Then, the absorbance at  $A_{450}$  was measured. Positive and negative controls were analyzed at the same time.

#### *Microneutralization antibody assay*

Vero E6 cells were seeded in 96-well plates. After 18 h, 100



PFU (plaque forming units) of the virus was incubated with an equal amount of diluted serum for 1 h. The mixture was transferred to the 96-well plates and incubated for 1 h. The liquid was discarded. The plates were sealed with 0.8% CMC and incubated for 24 h. On the second day, polyoxymethylene was used for inactivation. Then, the plates were sealed, fixed, and perforated. The plates were incubated with SARS-CoV-2 antibody as the primary antibody for 1 h and then anti-human IgG (H+L) antibody as the secondary antibody for 1 h. Finally, the TrueBlue substrate was added to develop the color of the solution. An ELISpot reader was used to read the spots and compute the inhibition rate of the virus and determine the serum neutralization capacity. If it was necessary to determine the specific neutralization titer of the serum, the serum was diluted by a gradient. In addition, a neutralization experiment was carried out to determine the maximum serum dilution with neutralization activity. This experiment was performed in the P3 laboratory.

### RNA purification and gene expression profiling from venous whole blood

Venous whole blood samples (5 mL) were collected into Tempus tubes (Applied Biosystems, USA) and RNA was purified according to the manufacturer's instructions. Gene expression from whole blood samples was determined using the Affymetrix Human Gene 2.0 ST Array following the manufacturer's instructions. The arrays were scanned by Affymetrix Gene Chip Scanner 3000 (Affymetrix, USA). Command Console Software (Affymetrix) was used to control the scanner and summarize the probe cell intensity data (CEL file generation) with default settings. The raw data were then normalized by Expression Console. All raw and normalized microarray data were deposited into the NCBI GEO database (Accession number: GSE114466).

### Data analysis of gene expression profiling

For each transcription region, a fragment per kilobase of transcript per million mapped reads (FPKM) value was calculated to quantify its expression abundance and variations. The FPKM formula is as follows:

$$FRKM = \frac{10^6 C}{NL / 10^3}$$

$C$  means the number of reads that mapped to the fragment.  $N$  means the number of reads that mapped to the whole genome.  $L$  stands for the length of the fragment.

The mRNA differential expression analysis between two different groups was performed using Cuffdiff software. Genes with FDRs below 0.05 and an absolute fold change >2 were considered differentially expressed genes (DEGs).

Genes usually interact with each other to perform specific

biological functions. Pathway-based analyses help to understand the biological functions of a gene. KEGG (Kanehisa and Goto, 2000) is the major public pathway-related database. Pathway enrichment analysis identified significantly enriched metabolic pathways or signal transduction pathways in the DEGs compared with the whole genome background. The formula for calculating the  $P$ -value is

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$

Here,  $N$  is the number of all genes with a KEGG annotation,  $n$  is the number of DEGs in  $N$ ,  $M$  is the number of all genes annotated to specific pathways, and  $m$  is the number of DEGs in  $M$ . The calculated  $P$ -value was FDR-corrected, taking  $FDR \leq 0.05$  as a threshold. Pathways meeting this condition were defined as significantly enriched pathways in the DEGs.

GO is an international standardized gene functional classification system that offers a dynamically updated controlled vocabulary and a strictly defined concept to comprehensively describe the properties of genes and their products in any organism. GO has three categories: molecular function, cellular component, and biological process. The calculation formula is the same as for the KEGG analysis.

### Statistical analysis

The means and standard deviations were calculated to represent the general features of COVID-19 mild or severe/critical patients. Fisher's exact tests were used to compare the categorical variables. The Kruskal-Wallis test was used for the continuous numerical variables. Two-sided comparisons with a  $P$ -value less than 0.05 were considered significant. The Spearman coefficients between the viral shedding time and the occurrence of SNPs, as well as the relationships among the SNPs, were calculated by the "psych" package in R. The statistical results from the Wilcoxon rank-sum test and Spearman analyses were adjusted with the Benjamini and Hochberg method ( $FDR < 0.05$ ) (Grant et al., 2005) using "p.adjust" in R. The bar plots in this study were drawn using the "ggplot2" package and the heatmaps were drawn with the "pheatmap" package in R.

### Data availability

The raw sequence data for SARS-CoV-2 in this paper have been deposited in the Genome Sequence Archive in the BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under BioProject PRJCA002503 with

accession ID CRA002522, which is publicly accessible at <https://bigd.big.ac.cn/gsa>. All other data are available from the authors upon reasonable request.

**Compliance and ethics** *The author(s) declare that they have no conflict of interest. The study conformed to the provisions of the Helsinki Declaration of 1975 (as revised in 2008) concerning Human and Animal Rights and was approved by the ethics committee at the First Affiliated Hospital of Guangzhou Medical University, China (Ethical number: 2020-36). All the patients provided written informed consent and volunteered to participate in the scientific research.*

**Acknowledgements** *This work was supported by the Guangzhou Institute of Respiratory Health Open Project (Funds provided by China Evergrande Group) (2020GIRHMS14), Zhongnanshan Medical Foundation of Guangdong Province (2020B1111340004 and ZNSA-2020003), the Traditional Chinese Medicine Bureau of Guangdong Province (2020ZYYJ05), and the funding from Guangzhou Institute of Respiratory Health. Furthermore, we would like to thank the patients, along with the nurses and clinical staff who provided patient care; the staff at the hospital respiratory medicine departments; the staff at the hospital clinical labs; the technical staff of the State Key Laboratory of Respiratory Disease for their excellent assistance; and the staffs at Guangdong Centers for Disease Control for the diagnosis of SARS-CoV-2 infection. Furthermore, we would also like to thank the AJE team for polishing the English language of this manuscript.*

## References

- Banerjee, S., Dhar, S., Bhattacharjee, S., and Bhattacharjee, P. (2020). Decoding the lethal effect of SARS-CoV-2 (novel coronavirus) strains from global perspective: molecular pathogenesis and evolutionary divergence. *bioRxiv*.
- Charalampous, T., Kay, G.L., Richardson, H., Aydin, A., Baldan, R., Jeanes, C., Rae, D., Grundy, S., Turner, D.J., Wain, J., et al. (2019). Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat Biotechnol* 37, 783–792.
- Chen, R., Sang, L., Jiang, M., Yang, Z., Jia, N., Fu, W., Xie, J., Guan, W., Liang, W., Ni, Z., et al. (2020). Longitudinal hematologic and immunologic variations associated with the progression of COVID-19 patients in China. *J Allergy Clin Immunol* 146, 89–100.
- Chinese Center for Disease Control and Prevention. (2020a). COVID-19 Outbreak Reported (in Chinese)..
- Chinese Center for Disease Control and Prevention. (2020b). The Guideline of Diagnosis and Treatment of COVID-19 (the Seventh Edition) (in Chinese).
- Faria, N.R., Quick, J., Claro, I.M., Thézé, J., de Jesus, J.G., Giovanetti, M., Kraemer, M.U.G., Hill, S.C., Black, A., da Costa, A.C., et al. (2017). Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* 546, 406–410.
- Faria, N.R., Kraemer, M.U.G., Hill, S.C., Goes de Jesus, J., Aguiar, R.S., Iani, F.C.M., Xavier, J., Quick, J., du Plessis, L., Dellicour, S., et al. (2018). Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science* 361, 894–899.
- Forster, P., Forster, L., Renfrew, C., and Forster, M. (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci USA* 117, 9241–9243.
- Gandhi, M., Yokoe, D.S., and Havlir, D.V. (2020). Asymptomatic transmission, the Achilles' Heel of current strategies to control COVID-19. *N Engl J Med* 382, 2158–2160.
- Grant, G.R., Liu, J., and Stoekert, C.J. (2005). A practical false discovery rate approach to identifying patterns of differential expression in microarray data. *Bioinformatics* 21, 2684–2690.
- Grubaugh, N.D., Ladner, J.T., Kraemer, M.U.G., Dudas, G., Tan, A.L., Gangavarapu, K., Wiley, M.R., White, S., Thézé, J., Magnani, D.M., et al. (2017). Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* 546, 401–405.
- Guan, W.J., Ni, Z.Y., Hu, Y., Liang, W.H., Ou, C.Q., He, J.X., Liu, L., Shan, H., Lei, C.L., Hui, D.S.C., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 382, 1708–1720.
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28, 27–30.
- Kasibhatla, S.M., Kinikar, M., Limaye, S., Kale, M.M., and Kulkarni-Kale, U. (2020). Understanding evolution of SARS-CoV-2: A perspective from analysis of genetic diversity of RdRp gene. *J Med Virol* 92, 1932–1937.
- Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. *Genome Biol* 5, R12.
- Ladner, J.T., Grubaugh, N.D., Pybus, O.G., and Andersen, K.G. (2019). Precision epidemiology for infectious disease control. *Nat Med* 25, 206–211.
- Lee, J.S., Park, S., Jeong, H.W., Ahn, J.Y., Choi, S.J., Lee, H., Choi, B., Nam, S.K., Sa, M., Kwon, J.S., et al. (2020). Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Sci Immunol* 5, eabd1554.
- Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128.
- Li, G., He, X., Zhang, L., Ran, Q., Wang, J., Xiong, A., Wu, D., Chen, F., Sun, J., and Chang, C. (2020a). Assessing ACE2 expression patterns in lung tissues in the pathogenesis of COVID-19. *J Autoimmun* 112, 102463.
- Li, G., Li, W., He, X., and Cao, Y. (2020b). Asymptomatic and presymptomatic infectors: hidden sources of coronavirus disease 2019 (COVID-19). *Clin Infect Dis* 71, 2018.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100.
- Liu, W.D., Chang, S.Y., Wang, J.T., Tsai, M.J., Hung, C.C., Hsu, C.L., and Chang, S.C. (2020). Prolonged virus shedding even after seroconversion in a patient with COVID-19. *J Infect* 81, 318–356.
- Lu, J., du Plessis, L., Liu, Z., Hill, V., Kang, M., Lin, H., Sun, J., François, S., Kraemer, M.U.G., Faria, N.R., et al. (2020). Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* 181, 997–1003. e9.
- Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., Masciovecchio, C., Angeletti, S., Ciccozzi, M., Gallo, R.C., et al. (2020). Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med* 18, 179.
- Phan, T. (2020). Genetic diversity and evolution of SARS-CoV-2. *Infect Genets Evol* 81, 104260.
- Rhee, M., and Burns, M.A. (2007). Nanopore sequencing technology: nanopore preparations. *Trends Biotech* 25, 174–181.
- Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W., et al. (2020). Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *N Engl J Med* 382, 970–971.
- Viprakasit, V., Tanphaichitr, V.S., Chinchang, W., Sangkla, P., Weiss, M.J., and Higgs, D.R. (2004). Evaluation of alpha hemoglobin stabilizing protein (AHSP) as a genetic modifier in patients with  $\beta$  thalassemia. *Blood* 103, 3296–3299.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46, W296–W303.
- Wölfel, R., Corman, V.M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M.A., Niemeyer, D., Jones, T.C., Vollmar, P., Rothe, C., et al. (2020). Virological assessment of hospitalized patients with COVID-2019. *Nature* 581, 465–469.
- World Health Organization. (2020a). Clinical management of COVID-19. WHO reference number: WHO/2019-nCoV/clinical/2020.5.
- World Health Organization. (2020b). Clinical management of severe acute

- respiratory infection when novel coronavirus (nCoV) infection is suspected.
- World Health Organization. (2020c). Coronavirus disease 2019 (COVID-19) Situation Report–157.
- Xiao, A.T., Tong, Y.X., and Zhang, S. (2020a). False negative of RT-PCR and prolonged nucleic acid conversion in COVID-19: Rather than recurrence. *J Med Virol* 92, 1755–1756.
- Xiao, F., Sun, J., Xu, Y., Li, F., Huang, X., Li, H., Zhao, J., Huang, J., and Zhao, J. (2020b). Infectious SARS-CoV-2 in feces of patient with severe COVID-19. *Emerg Infect Dis* 26, 1920–1922.
- Xu, K., Chen, Y., Yuan, J., Yi, P., Ding, C., Wu, W., Li, Y., Ni, Q., Zou, R., Li, X., et al. (2020). Factors associated with prolonged viral RNA shedding in patients with coronavirus disease 2019 (COVID-19). *Clin Infect Dis* 71, 799–806.
- Yang, Z.F., Mok, C.K.P., Peiris, J.S.M., and Zhong, N.S. (2015). Human infection with a novel avian influenza A (H5N6) virus. *N Engl J Med* 373, 487–489.
- Young, B.E., Ong, S.W.X., Kalimuddin, S., Low, J.G., Tan, S.Y., Loh, J., Ng, O.T., Marimuthu, K., Ang, L.W., Mak, T.M., et al. (2020). Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. *JAMA* 323, 1488–1494.
- Yu, Y., Li, Z., Li, Y., Yu, L., Jia, W., Ye, F., and Li, S.C. (2020). CovProfile: profiling the viral genome and gene expressions of SARS-COV2. bioRxiv.
- Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., Yu, J., Kang, M., Song, Y., Xia, J., et al. (2020). SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med* 382, 1177–1179.

## SUPPORTING INFORMATION

The supporting information is available online at <https://doi.org/10.1007/s11427-020-1921-5>. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.