Viral and host factors related to the clinical outcome of COVID-19

https://doi.org/10.1038/s41586-020-2355-0

Received: 14 March 2020

Accepted: 13 May 2020

Published online: 20 May 2020

Check for updates

Xiaonan Zhang[1,7], Yun Tan[2,7], Yun Ling[1,7], Gang Lu[2,7], Feng Liu[2,7], Zhigang Yi[1,3,7], Xiaofang Jia[1], Min Wu[1], Bisheng Shi[1], Shuibao Xu[1], Jun Chen[1], Wei Wang[1], Bing Chen[2], Lu Jiang[2], Shuting Yu[2], Jing Lu[2], Jinzeng Wang[2], Mingzhu Xu[1], Zhenghong Yuan[3], Qin Zhang[4], Xinxin Zhang[5], Guoping Zhao[6], Shengyue Wang[2]✉, Saijuan Chen[2]✉ & Hongzhou Lu[1]✉

In December 2019, coronavirus disease 2019 (COVID-19), which is caused by the new coronavirus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was identified in Wuhan (Hubei province, China)[1]; it soon spread across the world. In this ongoing pandemic, public health concerns and the urgent need for effective therapeutic measures require a deep understanding of the epidemiology, transmissibility and pathogenesis of COVID-19. Here we analysed clinical, molecular and immunological data from 326 patients with confirmed SARS-CoV-2 infection in Shanghai. The genomic sequences of SARS-CoV-2, assembled from 112 high-quality samples together with sequences in the Global Initiative on Sharing All Influenza Data (GISAID) dataset, showed a stable evolution and suggested that there were two major lineages with differential exposure history during the early phase of the outbreak in Wuhan. Nevertheless, they exhibited similar virulence and clinical outcomes. Lymphocytopenia, especially reduced CD4$^+$ and CD8$^+$ T cell counts upon hospital admission, was predictive of disease progression. High levels of interleukin (IL)-6 and IL-8 during treatment were observed in patients with severe or critical disease and correlated with decreased lymphocyte count. The determinants of disease severity seemed to stem mostly from host factors such as age and lymphocytopenia (and its associated cytokine storm), whereas viral genetic variation did not significantly affect outcomes.

The COVID-19 outbreak was first identified in Wuhan and appeared to be linked to Huanan Seafood Wholesale Market (HSWM). The causal agent, SARS-CoV-2[1,2], is closely related to a bat coronavirus (RaTG13)[2], although its receptor binding domain is more similar to that of pangolin coronaviruses[3]. Currently, several questions remain regarding the origin, evolution and host interactions of SARS-CoV-2. First, although HSWM has been widely proposed to be the original outbreak site of SARS-CoV-2, a significant number of the initial cases did not have contact with this market[4]. This casts doubt on the idea of a singular event of zoonotic spillover to humans in the initial outbreak. Second, additional data are required to discern whether the virulence of SARS-CoV-2 has altered as a result of genomic sequence evolution during the spread of the disease. Third, although SARS-CoV-2 infection can cause life-threatening respiratory disease, most cases manifest only mild pneumonia[5]. The factors associated with disease outcome have yet to be fully characterized. We have systematically analysed key immunological parameters spanning the course of infection in patients, obtained viral genomes directly from clinical samples, and delineated factors associated with prognosis and epidemiological features.

Overview of enrolment

The basic clinical and epidemiological features of the cohort (326 patients in Shanghai between 20 January and 25 February 2020) are summarized in Extended Data Table 1. Four categories of infected case were defined. Five individuals were asymptomatic; that is, they had no obvious fever, respiratory symptoms or radiological manifestations. Most patients (293) had mild disease with fever and radiological manifestations of pneumonia. Twelve patients who had symptoms of dyspnoea and signs of expanding ground-glass opacity in the lung within 24–48 h of admission were defined as severe cases. The remaining 16 patients deteriorated into acute respiratory distress syndrome and required mechanical ventilation or extracorporeal membrane oxygenation; these patients were categorized as critical (Extended

[1]Shanghai Public Health Clinical Center, Fudan University, Shanghai, China. [2]National Research Center for Translational Medicine, Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, Ruijin Hospital Affiliated to Shanghai Jiao Tong University (SJTU) School of Medicine, Shanghai, China. [3]Key Laboratory of Medical Molecular Virology, Shanghai Medical College, Fudan University, Shanghai, China. [4]Tong Ren Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [5]Research Laboratory of Clinical Virology, Ruijin Hospital Affiliated to Shanghai Jiao Tong University (SJTU) School of Medicine, Shanghai, China. [6]Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. [7]These authors contributed equally: Xiaonan Zhang, Yun Tan, Yun Ling, Gang Lu, Feng Liu, Zhigang Yi. ✉e-mail: wsy12115@rjh.com.cn; sjchen@stn.sh.cn; luhongzhou@shphc.org.cn

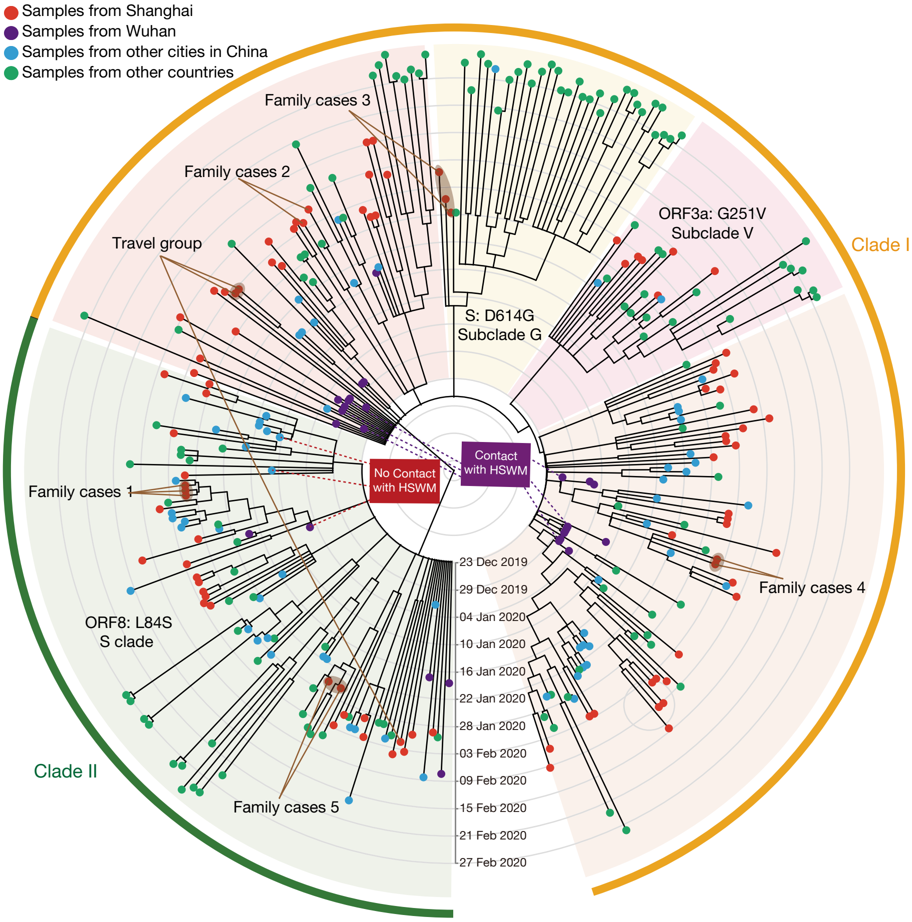


Fig. 1 | Phylogenetic analysis of the assembled SARS-CoV-2 genomes. We used 94 SARS-CoV-2 genome sequences and 221 published sequences to construct a time-resolved phylogeny tree. Clades I and II are marked and variations that distinguish branches of the tree are indicated. Concentric circles represent sampling dates. Each tip circle represents a single sample; colours indicate case locations (key). Cases with a history of contact with HSWM are highlighted.

Data Table 1). As of 1 April 2020, 315 (96.63%) of the patients had been discharged, and 6 (1.84%) had died.

Nucleotide variation in viral genomes

Sequencing data from 112 samples (sputum or oropharyngeal swab) passed quality control and were used for nucleotide variation calling (Extended Data Fig. 1). Compared to the first-released genome (Wuhan-Hu-1), we identified 66 synonymous and 103 nonsynonymous variants in 9 protein-coding regions (Extended Data Fig. 2a, b). Substitution rates in most genes (*ORF1ab*, *S*, *ORF3a*, *E*, *M* and *ORF7a*) were similar (around 3.5×10^{-4} per site per year), whereas variation rates in *ORF8* (9.51×10^{-4} per site per year) and *N* (1.05×10^{-3} per site per year) were higher (Extended Data Fig. 2a, b). The recurrence of variations in the viral genome is similar between samples from Shanghai and the GISAID dataset (Extended Data Fig. 2c).

Genomic phylogeny analysis

We next used the viral genomes from 94 patients (which were more than 90% complete) together with 221 sequences of SARS-CoV-2 from the GISAID database for phylogeny analysis. Two major clades were identified (Fig. 1, Extended Data Fig. 3a, b), both of which included cases diagnosed in early December 2019[1,2]. Clade I included several subclades, such as those characterized by ORF3a: p.251G>V (subclade V), or S: p.614D>G (subclade G). Clade II is distinguished from clade I by two linked variations—ORF8: p.84L>S (28144T>C) and ORF1ab: p.2839S (8782C>T) (Fig. 1, Extended Data Fig. 3a). The both major clades and their subclades were found in the Shanghai cohort, suggesting that there were multiple origins of transmission into Shanghai. We did not observe significant expansion of clades or subclades in Shanghai.

Additionally, the viral genomes from six patients with a clear history of contact with HSWM[1,2], the suspected initial outbreak site, were all clustered into clade I, whereas those from three patients diagnosed at the same time without a history of contact with HSWM[6,7] were clustered into clade II (Fig. 1). We analysed the sequences around nucleotides 8,782 and 28,144 of SARS-CoV-2 in samples from patients with or without a history of contact with HSWM and in the bat coronavirus Bat-SARS-CoV-RaTG13. Virus genomes found in patients without contact with HSWM were identical to Bat-SARS-CoV-RaTG13 at these two sites (Extended Data Fig. 3c).

We compared the clinical manifestations of patients infected with viruses of either clade I or clade II. We found no statistical difference in disease severity ($P = 0.88$), lymphocyte count ($P = 0.79$), CD3 T cell count ($P = 0.21$), C-reactive protein level ($P = 0.83$) or D-dimer level ($P = 0.19$), or in the duration of virus shedding after onset ($P = 0.79$) (Extended Data Table 2). Thus, these two clades of virus exhibited similar pathogenic effects despite their genome sequence variations. Likewise, we found no significant association between disease severity and the 13 most frequent genetic variations (synonymous and non-synonymous) (Extended Data Fig. 4).

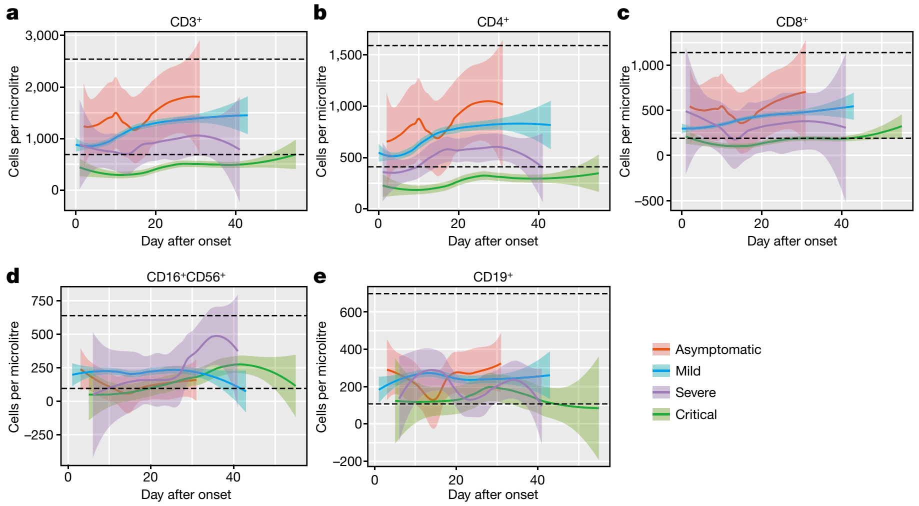


Fig. 2 | Lymphocyte numbers in patients during hospitalization. a–**e**, Temporal changes in CD3$^+$ (**a**), CD4$^+$ (**b**), CD8$^+$ (**c**), CD16$^+$CD56$^+$ (**d**) and CD19$^+$ (**e**) cell counts in each group. Data are shown as median ± 95% confidence interval and the normal range for each cell type is indicated with dashed lines. **a**–**c**, $n = 325$; **d**, **e**, $n = 220$.

Host factors associated with disease severity

A notable feature of our cohort was that some infected individuals (five cases; 1.53%) did not develop obvious symptoms even though substantial virus shedding could be detected. As shown in Extended Data Fig. 5a, an asymptomatic patient showed no obvious lesions in the lungs upon admission or five days later. By contrast, unilateral and bilateral opacity lesions were observed in patients with mild (Extended Data Fig. 5b) or critical COVID-19, and the latter deteriorated quickly over just two days (Extended Data Fig. 5c).

We further analysed the immunological and biochemical parameters of the patients (Extended Data Table 3). A prominent feature of COVID-19 was progressive lymphocytopenia, particularly in patients categorized as severe or critical (initial test result after admission, $P = 6 \times 10^{-6}$). Detailed analysis of lymphocyte subtypes revealed that CD3$^+$ T cells were most significantly affected ($P < 10^{-6}$), with CD4$^+$ and CD8$^+$ T cells sharing similar trends (CD4$^+$ T cell, $P < 10^{-6}$; CD8$^+$ T cell, $P = 1 \times 10^{-5}$). Notably, the changes in T lymphocytes were statistically significant not only in critical cases but also in the other three categories (asymptomatic, mild and severe; CD3$^+$ T cells, $P = 0.013$; CD8$^+$ T cells, $P = 0.004$). By contrast, for CD19$^+$ B cells, although a significant decline was found in critical cases ($P = 1 \times 10^{-5}$), patients in the other categories showed no obvious changes ($P = 0.47$). We further examined the longitudinal cell counting data for each group. It was clear that the CD3$^+$ T lymphocytes exhibited a gradual decline ($P < 0.05$ on dayd 7, 8, 11, 14–18, 22–25, 28 and 29 after onset, Kruskal–Wallis test) as the disease deteriorated (Fig. 2a), a trend that was also seen in CD4$^+$ and CD8$^+$ T cells (Fig. 2b, c). However, it was not found for natural killer (NK) (CD16$^+$CD56$^+$) or B (CD19$^+$) cells (Fig. 2d, e).

We next compared the clinical parameters grouped by comorbidity and found a significantly higher risk for disease progression when the disease was complicated by co-existing conditions ($P = 0.01$) (Extended Data Table 4), although the median age of the comorbidity group was also higher ($P = 0.02$). Indeed, univariate logistic regression analysis indicated that age ($P < 0.0001$), lymphocyte counts upon admission ($P < 0.0001$), comorbidities ($P = 0.01$) and gender ($P = 0.014$) (higher risk for male) were the main factors associated with disease severity (Extended Data Table 5). Multivariate analysis showed that age ($P = 0.002$) and lymphocytopenia ($P = 0.002$) were two major independent factors, whereas comorbidities did not reach statistical significance.

The levels of eleven cytokines (IFN-α, IFN-γ, IL-1β, IL-2, IL-4, IL-5, IL-6, IL-8, IL-10, IL-12 and IL-17) in serum were measured upon admission and during treatment. Among them, IL-6 ($P < 10^{-6}$) and IL-8 ($P = 1 \times 10^{-5}$) (Extended Data Table 3) showed the most significant changes. Notably, the levels of these two cytokines were inversely correlated with lymphocyte count (Fig. 3a, b, Extended Data Table 5). Furthermore, we combined the longitudinal cytokine data of each group and plotted their fluctuation patterns against the time point from onset. We aggregated the highest IL-6 data from each patient from day 6 to day 10 after onset and compared patients classed as critical with those classed as non-critical. Patients categorized as critical showed significantly higher levels of IL-6 ($P = 0.001$, two-sided Mann–Whitney U test) (Fig. 3c). There was a similarly significant difference in IL-8 level when data were aggregated from day 16 to day 20 after onset ($P = 0.006$) (Fig. 3d). These data suggest that there is a strong link between inflammatory cytokines and the pathogenesis of SARS-CoV-2 infection.

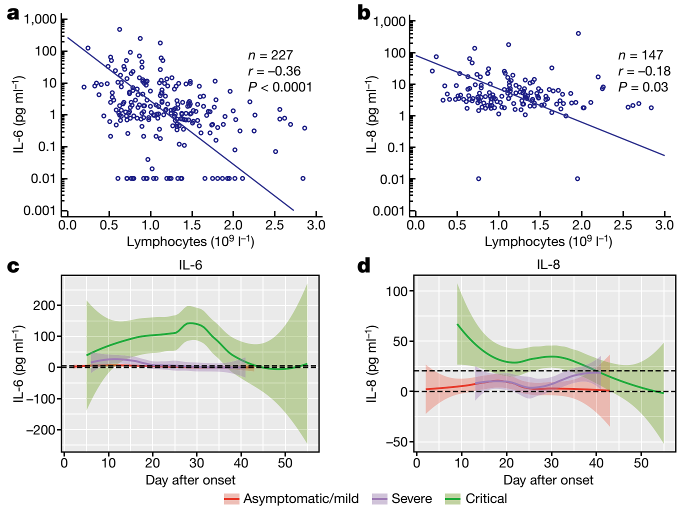


Fig. 3 | Correlation between inflammatory cytokines and lymphocyte counts. a, **b**, Levels of serum IL-6 (**a**) and IL-8 (**b**) upon admission plotted against lymphocyte count. Two-sided Spearman's correlation analysis with no adjustment of multiple comparisons. **c**, **d**, Temporal changes in IL-6 (**c**, $n = 230$) and IL-8 (**d**, $n = 149$) in each group during hospitalization. Data are shown as median ± 95% confidence interval and the normal range for each cytokine is indicated with dashed lines.

Discussion

Our analysis of some recently treated patients provides further evidence that the viral genome of SARS-CoV-2 is largely stable. Consistent with recently published results[8], we found that the observed small sequence variations divided the viral genomes into two major clades. We noted that six sequences recovered from patients with a history of contact with HSWM all fell into clade I, whereas three genomes from patients diagnosed in the same period but without exposure to HSWM were clustered into clade II. Thus, these two major haplotypes are likely to represent two lineages derived from a common ancestor that evolved independently in early December 2019 in Wuhan, only one of which (clade I) was spawned within the HSWM, where a high density of stalls, vendors and customers might have facilitated human-to-human transmission. Consistent with this idea, epidemiological investigations of the earliest cases found in Wuhan before 18 December 2019 identified two patients that were linked to HSWM and five that were not[4]. Our time-resolved phylogeny analysis suggests that the earliest zoonotic spillover event might have occurred in late November 2019, which is in agreement with a previous analysis[8].

Nevertheless, we found no significant differences in clinical features, mutation rate or transmissibility between patients infected with clade I or II virus. Our data are in agreement with a lack of selection against either clade, as suggested[9], but is at odds with a previous conclusion, the L/S-type classification of which was based on the same two linked polymorphisms[10]. The presumed difference in transmissibility might be due to sampling bias, as the early uploaded sequences in the GISAID database were recovered from a limited number of critically ill patients and duplicate assemblies from the same patients were not uncommon[1,2,11].

A recent analysis of 1,099 cases of COVID-19 in China found lymphocytopenia to be one of the most common features in laboratory tests[5]. Here, we have confirmed this observation and further shown that $CD3^+$ T cells were the major cell type that was suppressed in infected patients, whereas $CD19^+$ B cells and $CD16^+CD56^+$ NK cells exhibited less suppression. Indeed, lymphopenia and, in particular, reduced $CD4^+/CD8^+$ cell counts, are also a major manifestation of SARS-CoV infection[12]. Furthermore, our longitudinal monitoring of major cytokines indicated that IL-6 and IL-8 were negatively correlated with lymphocyte count and that IL-6 kinetics was highly related to disease severity. At present, the relationships between virological activity, cytokine release and lymphocytopenia remain unclear. We hypothesize that the immunopathological response against SARS-CoV-2, involving a cytokine storm and loss of $CD3^+$ T lymphocytes, could constitute—at least in part—an underlying mechanism for disease progression and fatality. The macrophages in the lung could serve as the first driver of the cytokine storm in the early phase of COVID-19 pneumonia[13], and subsequent lymphocyte infiltration mobilized by the cytokines, as observed in infected patients[14,15] and Rhesus macaques[16], may explain the lymphocytopenia, although probable cytokine-induced T cell depletion cannot be ruled out.

In conclusion, by closely monitoring the molecular and immunological data in 326 patients with COVID-19, we find evidence that adverse outcome is associated with depletion of $CD3^+$ T lymphocytes, which is tightly linked to bursts of cytokines such as IL-6 and IL-8. Targeted sequencing of 94 individuals who were infected during late January to February indicated limited variation in the viral genome, which suggests stable evolution. Two major lineages of the virus derived from one common ancestor may have originated independently from Wuhan in December 2019 and contributed to the current pandemic, although we find no major difference in clinical manifestation or transmissibility between them. Our data provide further evidence for the respective roles played by viral and host factors in disease mechanism and underscore the importance of early intervention in therapy.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2355-0.

1. Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med*. **382**, 727–733 (2020).
2. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
3. Lam, T. T. et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* (2020).
4. Li, Q. et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med*. **382**, 1199–1207 (2020).
5. Guan, W. J. et al. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med*. **382**, 1708–1720 (2020).
6. Chan, J. F. et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* **395**, 514–523 (2020).
7. Lu, R. et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020).
8. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med*. **26**, 450–452 (2020).
9. MacLean, O. A., Orton, R., Singer, J. B. & Robertson, D. L. Response to "On the origin and continuing evolution of SARS-CoV-2", http://virological.org/t/response-to-on-the-origin-and-continuing-evolution-of-sars-cov-2/418 (2020).
10. Tang, X. et al. On the origin and continuing evolution of SARS-CoV-2. *Natl Sci. Rev*. https://doi.org/10.1093/nsr/nwaa036 (2020).
11. Ren, L. L. et al. Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. *Chin. Med. J. (Engl.)* **133**, 1015–1024 (2020).
12. Wong, R. S. et al. Haematological manifestations in patients with severe acute respiratory syndrome: retrospective analysis. *Br. Med. J*. **326**, 1358–1362 (2003).
13. Tian, S. et al. Pulmonary pathology of early-phase 2019 novel coronavirus (COVID-19) pneumonia in two patients with lung cancer. *J. Thorac. Oncol*. **15**, 700–704 (2020).
14. Xu, Z. et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir. Med*. **8**, 420–422 (2020).
15. Wang, C. et al. Alveolar macrophage activation and cytokine storm in the pathogenesis of severe COVID-19. Preprint at https://doi.org/10.21203/rs.3.rs-19346/v1 (2020).
16. Shan, C. et al. Infection with novel coronavirus (SARS-CoV-2) causes pneumonia in the Rhesus macaques. Preprint at https://doi.org/10.21203/rs.2.25200/v1 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Ethics statement

This study was approved by the Shanghai Public Health Clinical Center Ethics Committee (no. YJ-2020-S015-01). Informed consent was obtained from all enrolled patients.

Patients

This study involved 326 patients, who had tested positive for SARS-CoV-2 RNA and were admitted to the Shanghai Public Health Clinical Center (the designated hospital receiving all COVID-19 cases in Shanghai) between 20 January and 25 February 2020. In addition to routine clinical tests, measurement of serum cytokines was performed on 228 patients. Their basic demographic, epidemiological and clinical characteristics are shown in Extended Data Table 1. The median age of the patients was 51 years (range 15–88) with a male:female sex ratio of 1.10:1. Among these 326 patients, 125 (38.34%) had at least one comorbidity; the most common were hypertension (76 patients), diabetes (24), coronary heart disease (13), chronic hepatitis B (10), chronic obstructive pulmonary disease (2), chronic renal disease (2) and cancer (3). Disease severity was categorized into four stages—asymptomatic, mild, severe and critical—according to the guidelines on the Diagnosis and Treatment of COVID-19 issued by the National Health Commission, China[17]. In brief, asymptomatic disease was defined as normal body temperature, lack of respiratory symptoms and no pulmonary radiological manifestation; mild disease as having fever, respiratory symptoms and radiological evidence of pneumonia; severe disease as meeting one of the following manifestations: respiratory rate >30/min, oxygen saturation levels (S_pO_2) <93%, arterial partial pressure of oxygen (P_aO_2)/fraction of inspired oxygen (F_iO_2)(P_aO_2/F_iO_2 ratio) $\leq$ 300 mm Hg or pulmonary imaging with multi-lobular lesions or lesion progression exceeding 50% within 48 h; and critical disease as one of the following: acute respiratory distress syndrome requiring mechanical ventilation, shock, or complications with other organ failure.

Nucleic acid extraction, molecular screening and genome sequencing

Swabs and sputum samples were collected for nucleic acid extraction using an automatic magnetic extraction device and accompanying kit (Shanghai Bio-Germ) and screened using a semiquantitative RT–PCR kit (Shanghai Bio-Germ) with amplification targeting the *ORF1a/b* and *N* genes. Deep sequencing was done using the nucleic acid extracted from patients confirmed as having COVID-19 by RT–PCR in Shanghai Public Health Clinical Center. We used a multiplexed amplicon strategy as described[18] and the primers were synthesized as described (https://github.com/artic-network/artic-ncov2019/blob/master/primer_schemes/nCoV-2019/V1/nCoV-2019.tsv). The primers were split into 10 subpools each containing 9–10 pairs for specific amplification of 400-bp viral sequence using the remaining cDNA from the diagnostic test. The PCR amplicons were purified using AMPure DNA cleanup steps. The amplicon libraries were generated using a NanoPrep for Illumina kit (IDT) according to the manufacturer's instructions. In brief, the procedures included end-repair, 3′ end adenylation, adaptor ligation and PCR amplification, followed by assessing DNA library quality. Amplicon sequencing was performed with established Illumina protocols on MiSeq platform (Illumina) according to a 2 × 300-bp protocol in the National Research Center for Translational Medicine (Shanghai).

Viral genomic sequence variation calling

All clean reads were mapped to the SARS-CoV-2 genome (Wuhan-Hu-1, GenBank accession number MN908947) using BWA (version 0.7.17)[19]. Variations were called with mpileup tools in samtools[20]. Low-quality variations with depth lower than 10 and Qual score lower than 50 were filtered using bcftools (version 1.9).

Phylogenetic analysis

Sequencing reads were trimmed using Trimmomatic (version 0.39)[21] to remove low-quality regions, adaptor sequences and sequencing primers. Clean reads were used to build virus genome assemblies with VirGenA (version 1.4)[22]. A post-assembly procedure was manually performed to remove low-quality content and potential sequencing artefacts. Ninety-four assemblies with coverage above 90% qualified for phylogeny analysis. MAFFT (version 7.453)[23] made the multi-sequence alignment after trimming off Ns on both ends of the genome sequences. The computation and visualization platform used for the phylogeny analysis was Nextstrain (version 1.15.0)[24]. The module we selected for phylogenetic tree building was IQ-TREE (version 1.6.12)[25]. Automatic substitution model selection was performed and the TIM+F+I model was selected to build the maximum likelihood phylogeny tree based on Bayesian information criteria (BIC) score. TreeTime (version 0.7.3)[26] was used for time-resolved phylogeny analysis. The resulting phylogeny tree was visualized using auspice from the Nextstrain package. All bioinformatics analyses were performed using the ASTRA supercomputing platform (Sugon) with Optane memory technology in the National Research Center for Translational Medicine (Shanghai).

Cytokine quantification and lymphocyte subset counting

A Becton Dickinson (BD) cytometric bead array (human Th1/Th2/Th17 cytokine kit and Human Inflammatory Cytokine Kit) was used quantify serum cytokines (IFNα, IFNγ, IL-1β, IL-2, IL-4, IL-5, IL-6, IL-8, IL-10, IL-12 and IL-17). $CD3^+$ T, $CD4^+$ T, $CD8^+$ T, $CD19^+$ B, and $CD16^+CD56^+$ NK cells were stained using BD Multitest 6-colour TBNK reagent in Trucount tubes and analysed using the BD FACSCanto II flow cytometer. The longitudinal plots of cytokines and cell count data were visualized using the geom_smooth tool in the ggplot2 R package.

Statistical analysis

Two sided Mann–Whitney *U* tests and Kruskal–Wallis tests were used to compare two and more than two groups of variables, respectively. χ^2 and Fisher's exact test were used for analysing contingency tables. Spearman's rank correlation test was used to evaluate correlations. No statistical methods were used to predetermine sample size. Investigators were not blinded to patient group during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The 94 genome sequences with over 90% coverage were deposited in GISAID (https://www.gisaid.org/) (accessions EPI_ISL_416316–EPI_ISL_416409) and the phylogeny result is accessible at http://ncov.linc.org.cn. The amplicon sequencing reads for variant calling have been deposited with NCBI Bioproject (PRJNA627662) and NODE (http://www.biosino.org/node/project/detail/OEP000877).

17. National Health Commission of the People's Republic of China *Diagnosis and Treatment Protocol for COVID-19 (Trial Version 7)* http://en.nhc.gov.cn/2020-03/29/c_78469.htm (2020).
18. Quick, J. et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protocols* **12**, 1261–1276 (2017).
19. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
20. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
21. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
22. Fedonin, G. G., Fantin, Y. S., Favorov, A. V., Shipulin, G. A. & Neverov, A. D. VirGenA: a reference-based assembler for variable viral genomes. *Brief. Bioinform.* **20**, 15–25 (2019).

23. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol*. **30**, 772–780 (2013).
24. Hadfield, J. et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
25. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol*. **32**, 268–274 (2015).
26. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: maximum-likelihood phylodynamic analysis. *Virus Evol*. **4**, vex042 (2018).

Acknowledgements This work was supported by National Megaprojects of China for Infectious Diseases (2017ZX10103009-001, 2018ZX10305409-001-005); Double First-Class Project of Fudan University (no. IDF162005); Scientific Research for Special Subjects on 2019 Novel Coronavirus of Shanghai Public Health Clinical Center (no. 2020YJKY01); Double First-Class Project (WF510162602) from the Ministry of Education, State Key Laboratory of Medical Genomics; Overseas Expertise Introduction Project for Discipline Innovation (111 Project, B17029); National Key R&D Program of China (2019YFA0905902); and the Shanghai Guangci Translational Medical Research Development Foundation. We acknowledge all healthcare personnel involved in the diagnosis and treatment of patients in Shanghai Public Health Clinical Center and we thank Z. Chen for guidance regarding study design, interpretation of results and critical revision of the manuscript. We also thank all researchers who shared SARS-CoV-2 genome sequences in GISAID. We thank the late Y. Hu of Shanghai Public Health Clinical Center for her lifelong commitment to infectious disease research.

Author contributions S.C., H.L., Xiaonan Zhang, S.W. and Z. Yuan conceived the study. Xiaonan Zhang, Y.L., Z. Yi, X.J., M.W., B.S., S.X., J.C., Q.Z. and W.W. collected patient samples and epidemiological and clinical data. Xiaonan Zhang, X.J. and M.W. performed viral RNA isolation and PCR. S.Y., J.L., L.J., G.L. and J.W. performed sequencing and sequence assembly. Xiaonan Zhang, Y.T., F.L., G.L., B.S., S.X., J.C., B.C., M.X., S.W. and S.C. carried out data acquisition, analysis and interpretation. Xiaonan Zhang, Y.T., F.L. and G.L. drafted the manuscript. S.C., S.W., Xinxin Zhang and G.Z. revised the final manuscript.

Competing interests The authors declare no competing interests.

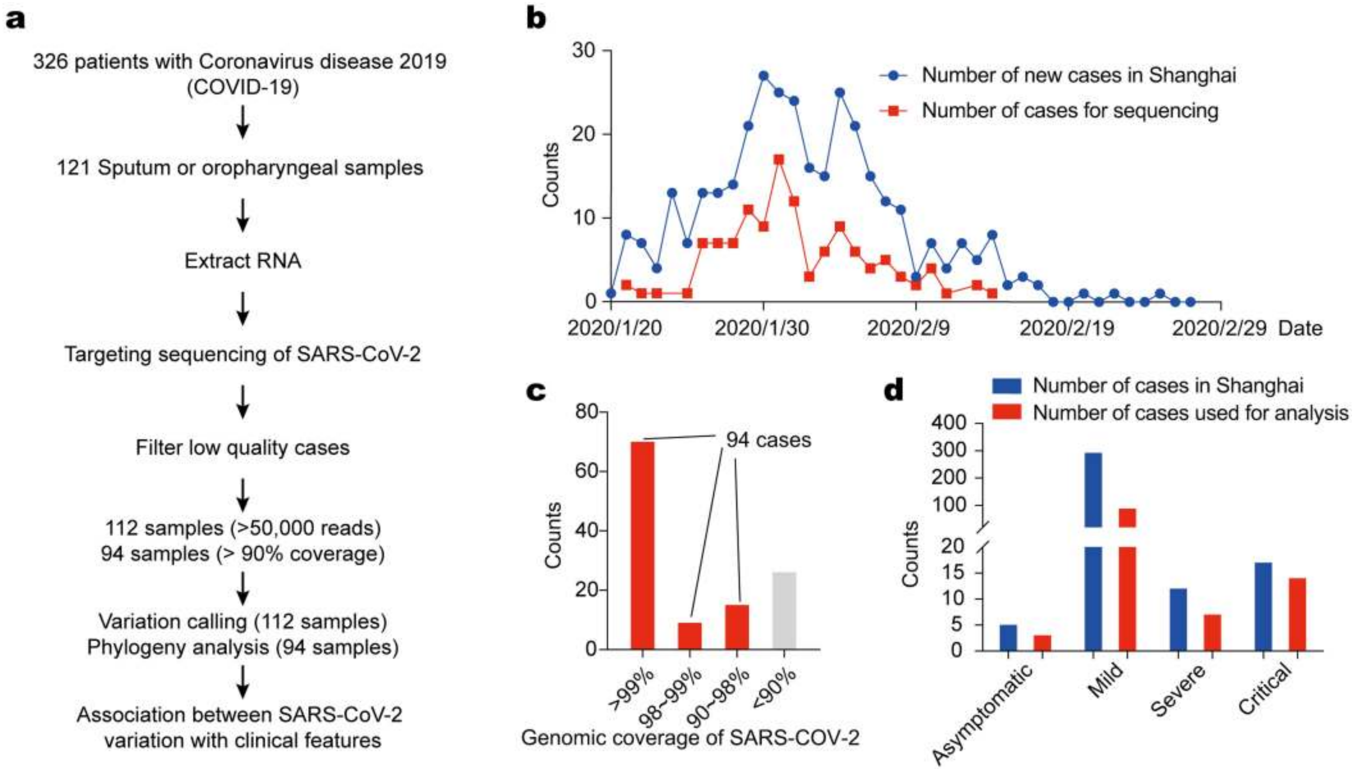
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2355-0.

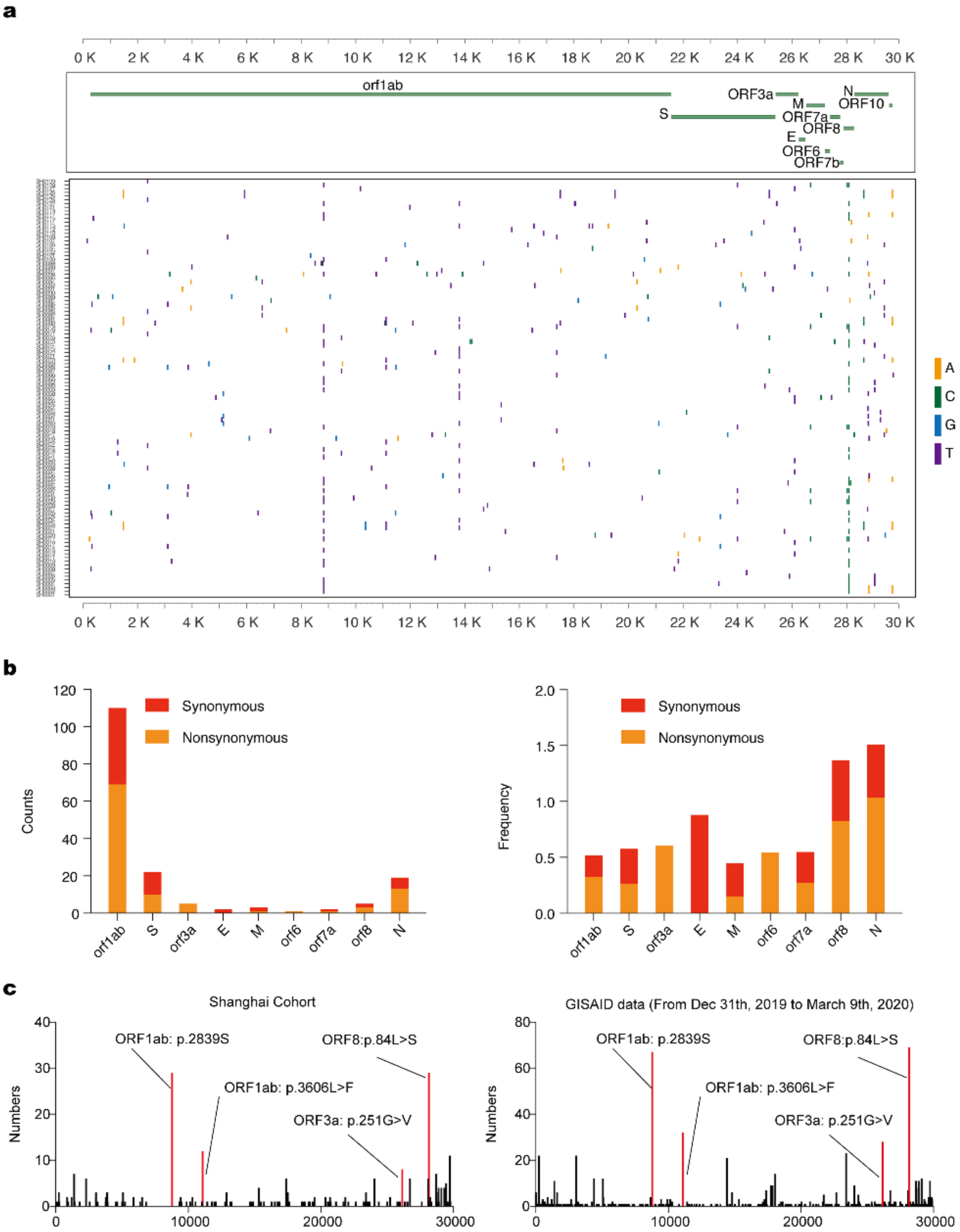
Correspondence and requests for materials should be addressed to S.W., S.C. or H.L.

Peer review information *Nature* thanks Luke O'Neill and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

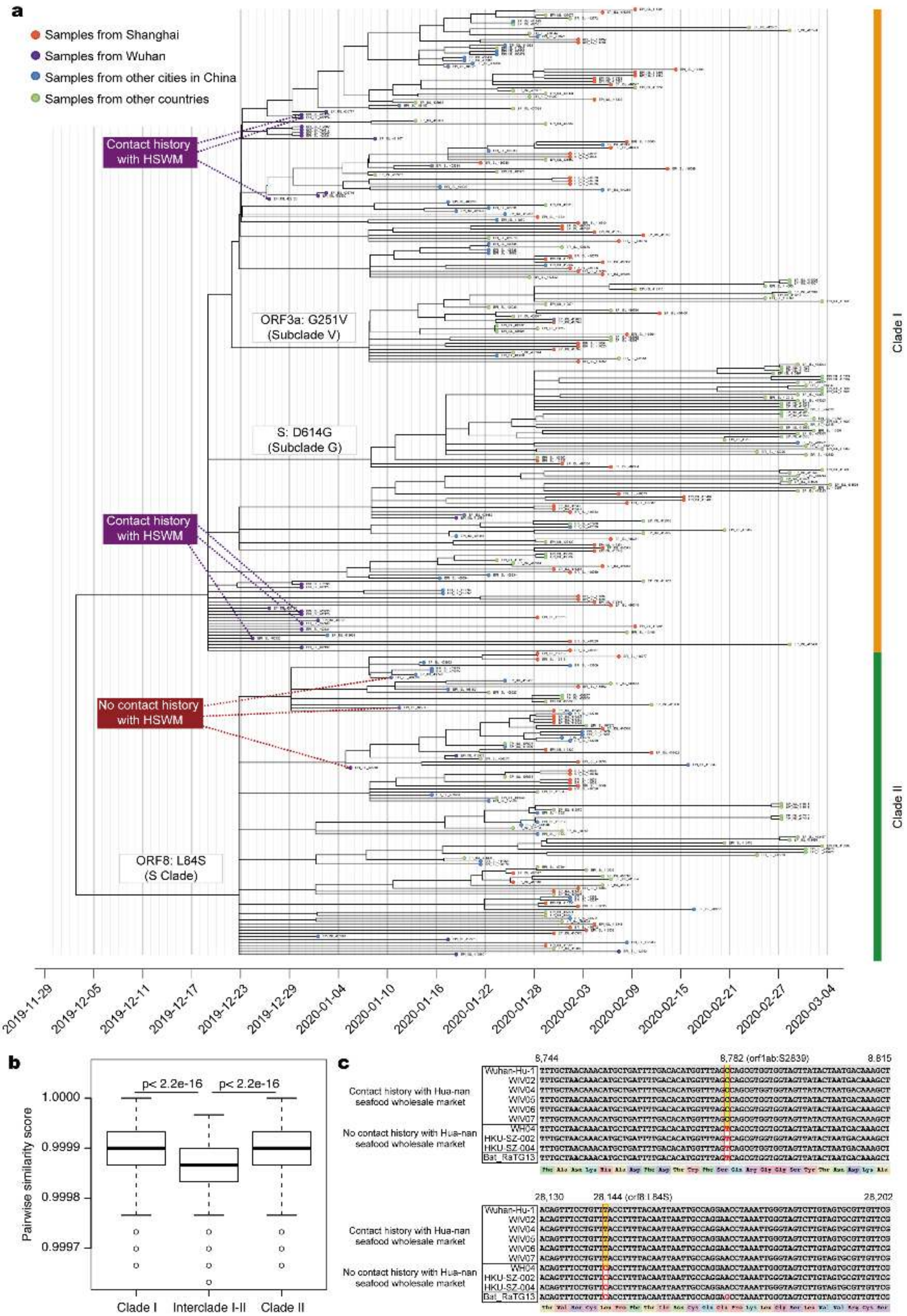
Reprints and permissions information is available at http://www.nature.com/reprints.



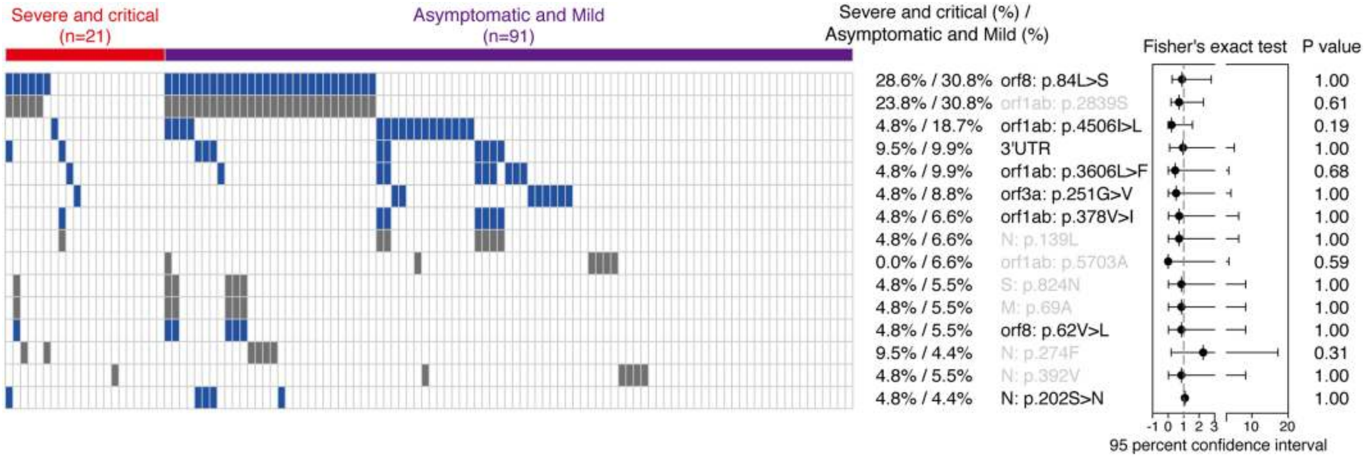
Extended Data Fig. 1 | High-throughput targeted sequencing process. **a**, Design of experiment and analysis pipeline. A total of 121 samples, including sputum and oropharyngeal swab samples, from patients with COVID-19 were used for viral RNA extraction and sequencing. **b**, Comparison of the number of patients with COVID-19 used for sequencing with total cases in the Shanghai cohort between 20 January and 25 February 2020. **c**, Coverage of the SARS-CoV-2 genome per sample in four bins (>99%, 98–99%, 90–98%, and <90%). Ninety-four samples have coverage of over 90%. **d**, Numbers of individuals with severe or critical and mild or asymptomatic COVID-19. Blue bar, cases included in this study; red bar, cases used for variant calling or phylogeny analysis.



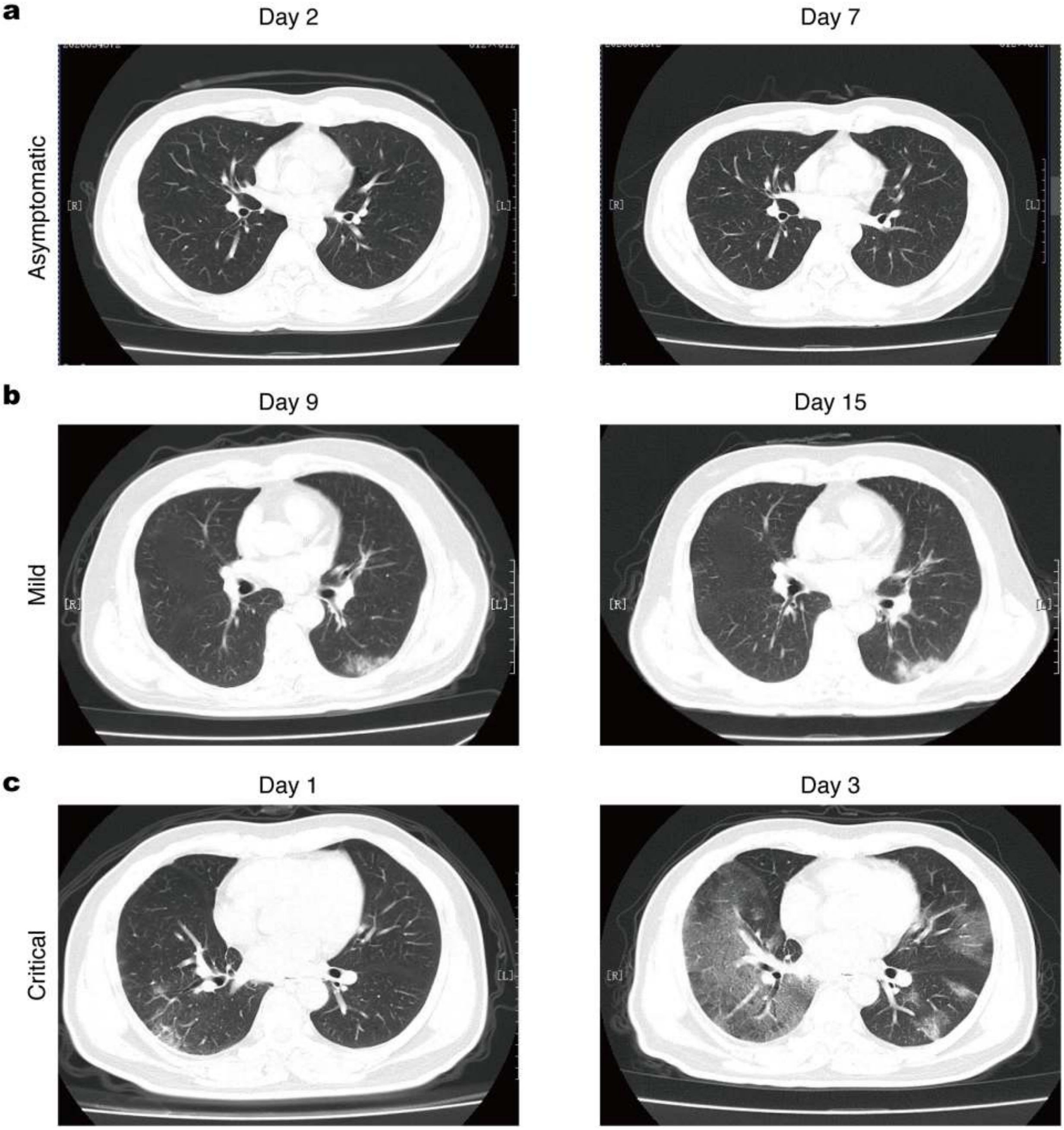
Extended Data Fig. 2 | Characteristics of single nucleotide variations in 112 Shanghai SARS-CoV-2 samples. a, Single nucleotide variations between the SARS-CoV-2 reference genome (Wuhan-Hu-1) and genome sequences in the Shanghai cohort, shown by vertical colour bars. Orange, A; blue, G; green, C; purple, T. The top panel shows the open reading frame of each gene. **b**, Summary of 169 variations in nine open reading frames. Variation counts and the ratio of synonymous to nonsynonymous mutations are plotted. Red represents synonymous; orange represents nonsynonymous. **c**, Frequencies of variations in Shanghai cohort and published GISAID dataset.



Extended Data Fig. 3 | Phylogenetic analysis of the assembled SARS-CoV-2 genomes. a, A total of 94 SARS-CoV-2 genome sequences and 221 published sequences (as in Fig. 1a) were used for construction of a time-resolved rectangular phylogeny tree. Clade I and clade II are marked, and variations that distinguish branches of the phylogeny tree are indicated. Each tip circle represents a single sample. The position of each sample along the *x*-axis corresponds to the sample collection date. Case locations are marked by colours (see key). GISAID accession identifiers are displayed alongside each tip circle. Cases with or without a history of contact with HSWM are highlighted. **b**, Box plot (centre, median; box, interquartile range (IQR); whiskers, 1.5× IQR) of pairwise similarity scores within clade I ($n = 22{,}791$), clade II ($n = 5{,}050$) and between clades I and II (interclade I–II, $n = 21{,}614$). Two-sided unpaired *t*-test. **c**, Alignment of sequences around nucleotides 8,782 and 28,144 using Bat-SARS-CoV-RaTG13 sequence and SARS-CoV-2 sequences recovered from patients with or without known history of exposure to HSWM.



Extended Data Fig. 4 | SARS-CoV-2 variation in severe or critical and mild cases of COVID-19. Variations of SARS-CoV-2 from patients with differing severity of COVID-19 are plotted. Blue, nonsynonymous mutations; grey, synonymous mutations. Red bar, severe or critical COVID-19; purple bar, mild COVID-19. Two-sided Fisher's exact test; 95% confidence intervals and odds ratios are shown. *P* values are shown on the right.



Extended Data Fig. 5 | Computed tomographic scans of three typical patients. a, Asymptomatic; **b**, mild; **c**, critical. Days after admission to hospital are shown above scans. Computed tomography scans were performed once for each patient on a specific date during treatment. Images are representative of 5 asymptomatic, 7 mild and 13 critical cases.

Extended Data Table 1 | Clinical features of enrolled patients and subgroups

	Entire cohort (n=326)	Phylogeny (n=94)	Cytokine analysis (n=228)
Age			
Median (range)-year	51 (15-88)	53 (24-85)	53.5 (21-85)
<=39 --no. (%)	109 (33.44%)	32 (34.04%)	75 (32.89%)
40-49 --no. (%)	49 (15.03%)	10 (10.64%)	25 (10.96%)
50-59 --no. (%)	55 (16.87%)	18 (19.15%)	40 (17.54%)
60-69 --no. (%)	80 (24.54%)	22 (23.40%)	63 (27.63%)
>=70 --no. (%)	33 (10.12%)	12(12.76%)	25 (10.96%)
Gender			
Female (%)	155(47.54%)	44 (46.81%)	110 (48.24%)
Male (%)	171 (52.46 %)	50 (53.19%)	118 (51.75%)
Exposure to source of transmission within 14 days			
Local residents of Hubei --no. (%)	90 (27.60%)	26 (27.66%)	56 (24.56%)
Recently been to Hubei --no. (%)	80 (24.54%)	25 (26.60%)	47 (20.61%)
Contacted with people from Hubei --no. (%)	52 (15.96%)	14 (14.89%)	45 (19.74%)
Other or unknown --no. (%)	104 (31.90%)	29 (30.85%)	80 (35.09%)
Highest Temperature (°C)	38.0 (36.6 -40.0)	38.2 (36.6-40.0)	38.0 (36.0-40.0)
Disease Severity			
Asymptomatic --no. (%)	5 (1.53%)	0 (0)	3 (1.32%)
Mild --no. (%)	293 (89.88%)	78 (82.98%)	200 (87.72%)
Severe --no. (%)	12 (3.68%)	5 (5.32%)	10 (4.39%)
Critical --no. (%)	16 (4.91%)	11 (11.70%)	15 (6.58%)
Death --no. (%)	6 (1.84%)	5 (5.32%)	5 (2.19%)
Any Co-morbidities	125 (38.34)	38 (40.42%)	90 (39.47%)
Hypertension --no. (%)	76 (23.31%)	24 (25.53%)	60 (26.32%)
Diabetes --no. (%)	24 (9.13%)	9 (9.57%)	18 (7.89%)
Coronary heart disease --no. (%)	13 (3.99%)	5 (5.32%)	7 (3.07%)
Chronic hepatitis B --no. (%)	10 (3.07%)	5 (5.32%)	4 (1.75%)
Chronic obstructive pulmonary disease --no. (%)	2 (0.61%)	1 (1.06%)	1 (0.44%)
Chronic renal diseases --no. (%)	2 (0.61%)	1 (1.06%)	1 (0.44%)
Cancer --no. (%)	3 (0.92%)	2 (2.13%)	2 (0.88%)

Extended Data Table 2 | Clinical features of patients infected with different clades of SARS-CoV-2

	Clade I (n=78)	Clade II (n=34)	P value
Age-year	57.5 (41-65)	46.5 (32-62)	0.02 [a]
Disease status			
Critical	10	4	
Non-critical	68	30	0.88 [b]
Leukocytes counts (×10^9/L, normal range 3.5-9.5)	4.46 (3.80-5.66)	4.80 (4.11-5.65)	0.58 [a]
Lymphocytes (× 10^9/L, normal range 1·1–3·2)	1.07 (0.76-1.40)	1.01 (0.71-1.37)	0.79 [a]
CD3+T cell counts (/µL, normal range 690-2540)	720 (503-943)	606 (411-854)	0.21 [a]
CD8+T cell counts (/µL, normal range 190-1140)	222 (164-371)	188 (135-283)	0.14 [a]
CD4+T cell counts (/µL, normal range 410-1590)	404 (279-608)	367 (219-578)	0.38 [a]
Platelets (×10^9/L, normal range 125-350)	158.5 (130-208)	178 (151-207)	0.18 [a]
Haemoglobin (g/L, 115-150)	137 (124-149)	141 (130-150)	0.22 [a]
C-reactive protein (mg/L, normal range 0.9-1.8)	13.35 (5.4-41.6)	18.35(5.67-32.4)	0.83 [a]
Lactose dehydrogenase (U/L, normal range 120-250)	221 (186-280)	240 (195-299)	0.23 [a]
Complement C3 (mg/L, normal range <3)	1.13 (1.01-1.29)	1.15 (1.01-1.29)	0.81 [a]
D-dimer (µg/L, normal range 0-0.5)	0.415 (0.30-0.71)	0.35 (0.25-0.57)	0.19 [a]
IL-6 (pg/ml, normal range <5.4 pg/ml)	1.46 (0.69-8.17)	2.12 (0.75-6.29)	0.92 [a]
IL-8 (pg/ml, normal range <20.6 pg/ml)	4.72 (3.41-6.03)	3.40 (2.61-7.03)	0.32 [a]
Duration of virus shedding after onset (days)	16 (11-24)	18 (12-24)	0.79 [a]

Data are presented as median (IQR) [a]Two-sided Mann–Whitney *U* test. [b]χ^2 test.

Extended Data Table 3 | Immunological and biochemical parameters associated with disease severity

	Asymptomatic (n=5)	Mild (n=293)	Severe (n=12)	Critical (n=16)	P value*
Leukocytes counts ($\times10^9$/L, normal range 3.5-9.5)	7.13 (4.03-8.72)	4.83 (4.03-5.89)	4.30 (3.53-7.76)	5.52 (4.12-7.74)	0.30
Lymphocytes ($\times 10^9$/L, normal range 1·1–3·2)	1.59 (1.23-2.08)	1.25 (0.85-1.49)	0.91 (0.66-1.40)	0.64 (0.44-0.88)	6×10^{-6}
CD3$^+$T cell count (/µL,normal range 690-2540)	1208(1012-1591)	778 (553-1041)	500 (379-705)	234 (152-474)	$<1\times10^{-6}$
CD8$^+$T cell count (/µL, normal range 190-1140)	495 (405-615)	265 (171-393)	133 (102-199)	96 (52-211)	1×10^{-5}
CD4$^+$T cell count (/µL, normal range 410-1590)	634 (529-909)	455 (314-650)	332 (226-541)	130 (96-254)	$<1\times10^{-6}$
Platelets ($\times10^9$/L, normal range 125-350)	187 (165-219)	182 (145-225)	162 (130-198)	156 (119-202)	0.23
CD19$^+$ B cell counts (/µL, normal range 107-698)	289 (223-320)	228 (160-309)	221 (108-269)	111 (58-135)	1×10^{-5}
CD16$^+$CD56$^+$ NK cells (/µL, normal range 95-640)	160 (102-219)	188 (130-267)	170 (113-279)	82 (41-174)	0.04
Haemoglobin (g/L, 115-150)	139 (132-147)	135 (126-148)	142 (126-146)	145 (126-150)	0.71
C-reactive protein (mg/L, normal range 0.9-1.8)	<3	11.4 (3.9-25.2)	46.1 (17.4-92.8)	71.2 (37.0-120.0)	$<1\times10^{-6}$
Alanine aminotransferase (U/L, normal range 7-40)	17 (9-24.3)	21 (15-32)	29 (20-37)	27 (17.5-33.0)	0.16
Lactose dehydrogenase (U/L, normal range 120-250)	174 (162-180)	225 (193-269)	355 (309-401)	371 (293-460)	$<1\times10^{-6}$
Complement C3 (mg/L, normal range <3)	1.02 (0.92-1.06)	1.17 (1.04-1.30)	1.19 (0.99-1.32)	1.00 (0.86-1.28)	0.03
D-dimer (µg/L, normal range 0-0.5)	0.3 (0.2-0.4)	0.4 (0.3-0.7)	0.7 (0.5-1.6)	0.8 (0.5-1.3)	2×10^{-5}
IL-6 (pg/ml, normal range <5.4 pg/ml)	0.3 (0.1-5.7)	1.1 (0.5-3.2)	6.0 (1.1-11.1)	33.0 (9.0-110.6)	$<1\times10^{-6}$
IL-8 (pg/ml, normal range <20.6 pg/ml)	7.2 (3.1-7.3)	3.5 (2.3-6.0)	8.1 (3.6-16.8)	21.4 (7.2-58.7)	1×10^{-5}

*Kruskal–Wallis test. Data are presented as median (IQR).

Extended Data Table 4 | Clinical features of patients with and without comorbidities

Co-morbidities	No (n=201)	Yes (n=125)	P value
Age-year	48 (35-63)	54 (39-66)	0.02 [a]
Disease status			
Critical	5	11	
Non-critical	196	114	0.01 [b]
Leukocytes counts ($\times 10^9$/L, normal range 3.5-9.5)	4.70 (3.91-5.93)	5.01 (4.10-6.02)	0.34 [a]
Lymphocytes ($\times 10^9$/L, normal range 1·1–3·2)	1.13 (0.81-1.49)	1.12 (0.77-1.48)	0.65 [a]
CD3$^+$T cell counts (/µL, normal range 690-2540)	752 (504-1023)	716 (497-1042)	0.58 [a]
CD8$^+$T cell counts (/µL, normal range 190-1140)	259 (171-396)	257 (143-386)	0.40 [a]
CD4$^+$T cell counts (/µL, normal range 410-1590)	448 (301-633)	415 (306-651)	0.67 [a]
Platelets ($\times 10^9$/L, normal range 125-350)	179 (145-224)	178 (142-221)	0.65 [a]
Haemoglobin (g/L, 115-150)	136 (127-148)	135 (125-149)	0.87 [a]
C-reactive protein (mg/L, normal range 0.9-1.8)	11.05 (3.0-22.9)	16.2 (4.98-49.8)	0.005 [a]
Lactose dehydrogenase (U/L, normal range 120-250)	225 (192-277)	239 (199-303)	0.11 [a]
Complement C3 (mg/L, normal range <3)	1.13 (1.01-1.29)	1.23 (1.08-1.30)	0.04 [a]
D-dimer (µg/L, normal range 0-0.5)	0.41 (0.29-0.70)	0.45 (0.29-0.84)	0.21 [a]
IL-6 (pg/ml, normal range <5.4 pg/ml)	1.09 (0.45-4.28)	1.51 (0.74-8.50)	0.06 [a]
IL-8 (pg/ml, normal range <20.6 pg/ml)	3.71 (2.58-6.72)	5.25 (2.30-9.17)	0.36 [a]
Onset to hospitalization (day)	4 (2-7)	4 (2-7.5)	0.62
Duration of virus shedding after onset (days)	15 (10-20)	15 (10-21.25)	0.33 [a]

Data are presented as median (IQR). [a]Two-sided Mann–Whitney *U* test. [b]χ^2 test.

Extended Data Table 5 | Univariate and multivariate logistic regression analyses of factors associated with disease severity

Clinical Variables *	Odds Ratio	95% CI		P value
		Lower	Upper	
UNIVIARIATE ANALYAIS				
Age -year	1.128	1.07	1.19	<0.0001
Gender	0.221	0.067	0.858	0.014
Lymphocyte count	0.0099	0.0011	0.0899	<0.0001
CD3$^+$ T cell counts	0.9932	0.9901	0.9962	<0.0001
CD4$^+$ T cell counts	0.9878	0.9823	0.9932	<0.0001
CD8$^+$ T cell counts	0.9900	0.9840	0.9959	0.0003
CD19$^+$ B cell counts	0.9797	0.9703	0.9892	<0.0001
CD16$^+$CD56$^+$ NK cell counts	0.9949	0.9891	1.0007	0.045
IL-6	1.0623	1.0329	1.0926	<0.0001
IL-8	1.0133	0.9992	1.0276	0.0164
Comorbidity	3.40	1.29	11.20	0.01
MULTIVARIATE ANALYSIS				
Age -year	1.09	1.03	1.16	0.002
Lymphocyte count	0.03	0.003	0.273	0.002

*Critical cases (*n* = 16) versus asymptomatic, mild or severe cases of COVID-19 (*n* = 310).

nature research

Corresponding author(s):	Hongzhou Lu, Saijuan Chen, Shengyue Wang
Last updated by author(s):	Apr 23, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
□	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
□	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
□	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	□	A description of all covariates tested
□	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
□	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
□	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	□	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	□	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	□	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Miseq control software (version 2.6.2.1)
Data analysis	Trimmomatic (version 0.39), BWA (version 0.7.17), Samtools (version 1.10), VirGenA (version 1.4), MAFFT (version 7.453), IQ-TREE (version 1.6.12), TreeTime (version 0.7.3), Nextstrain (version 1.15.0), R (version 3.6.2), ggplot2 (version 3.3.0), bcftools (version 1.9), Graphpad Prism (version 6), Medcalc (version 15).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The 94 genome sequences with over 90% coverage were deposited to GISAID (EPI_ISL_416316-- EPI_ISL_416409). The phylogeny result is accessible via web address http://ncov.linc.org.cn. The amplicon sequencing reads for variant calling were deposited to NCBI Bioproject (PRJNA627662).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	A total of 326 patients, who were tested positive for SARS-CoV-2 RNA and were admitted into Shanghai Public Health Clinical Center from Jan 20th to Feb 25th were included. In addition to routine clinical tests, measurement of serum cytokine was performed on 228 patients. Phylogenetic analysis was performed on genome sequences (>90% complete) recovered from 94 patients. The sample size was determined by the maximum available clinical and laboratory information in our center. No prior sample size calculation was performed.
Data exclusions	No data excluded.
Replication	All the clinical and immunological data was generated in laboratories within Shanghai Public Health Clinical Center which undertook regular quality controls and inter-laboratory consistency evaluations. All the genetic sequence data were generated using the standard practice of molecular biology and next generation sequencing standards.No experimental replication was performed for sequencing experiments and clinical measurements.
Randomization	Participants were chosen randomly.
Blinding	The measurements were performed without prior knowledge of the participant groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	A total of 326 patients, who were tested positive for SARS-CoV-2 RNA and were admitted into Shanghai Public Health Clinical Center from Jan 20th to Feb 25th. The median age of the patients was 51 years (range 15-88) with a male: female sex ratio of 1.10. 125 cases (38.34% had at least one co-morbidiy, the most common were hypertension (76 cases), diabetes (24 cases), coronary heart disease (13 cases), chronic hepatitis B (10 cases), chronic obstructive pulmonary disease (2 cases), chronic renal disease (2 cases) and cancer (3 cases).
Recruitment	All the available COVID-19 patients who were willing to participate in this study were recruited in this study. No self-selection biase existed to the best of our knowledge.
Ethics oversight	The study was approved by the ethics committee of the Shanghai Public Health Clinical Center (Approval No. YJ-2020-S015-01)

Note that full information on the approval of the study protocol must also be provided in the manuscript.