

# Viral and host factors related to the clinic outcome of the SARS-CoV-2 infection

## **Xiaonan Zhang**

Shanghai Public Health Clinical Center, Fudan University

## **Yun Tan**

National Research Center for Translational Medicine (Shanghai), Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

## **Yun Ling**

Shanghai Public Health Clinical Center, Fudan University

## **Gang Lu**

National Research Center for Translational Medicine (Shanghai), Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

## **Feng Liu**

National Research Center for Translational Medicine (Shanghai) and State Key Laboratory of Medical Genomics, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

## **Zhigang Yi**

Shanghai Public Health Clinical Center and Key Laboratory of Medical Molecular Virology, Shanghai Medical College, Fudan University

## **Xiaofang Jia**

Shanghai Public Health Clinical Center, Fudan University

## **Min Wu**

Shanghai Public Health Clinical Center, Fudan University

## **Bisheng Shi**

Shanghai Public Health Clinical Center, Fudan University

## **Shuibao Xu**

Shanghai Public Health Clinical Center, Fudan University

## **Jun Chen**

Shanghai Public Health Clinical Center, Fudan University

## **Wei Wang**

Shanghai Public Health Clinical Center, Fudan University

## **Bing Chen**

Shanghai Institute of Hematology, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

## **Lu Jiang**

National Research Center for Translational Medicine (Shanghai), Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

**Shuting Yu**

National Research Center for Translational Medicine (Shanghai), Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

**Jing Lu**

Shanghai Institute of Hematology, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

**Jingzheng Wang**

National Research Center for Translational Medicine (Shanghai), Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

**Mingzhu Xu**

Shanghai Public Health Clinical Center, Fudan University

**Zhenghong Yuan**

Key Laboratory of Medical Molecular Virology, Shanghai Medical College, Fudan University

**Qin Zhang**

Tong Ren Hospital, Shanghai Jiao Tong University School of Medicine

**Xinxin Zhang**

Research Laboratory of Clinical Virology, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

**Guoping Zhao**

Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences

**Shengyue Wang (✉ [wsy12115@rjh.com.cn](mailto:wsy12115@rjh.com.cn))**

National Research Center for Translational Medicine (Shanghai) and State Key Laboratory of Medical Genomics, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

**Saijuan Chen (✉ [sjchen@stn.sh.cn](mailto:sjchen@stn.sh.cn))**

National Research Center for Translational Medicine (Shanghai) and State Key Laboratory of Medical Genomics and Shanghai Institute of Hematology, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

**Hongzhou Lu (✉ [luhongzhou@shphc.org.cn](mailto:luhongzhou@shphc.org.cn))**

Shanghai Public Health Clinical Center, Fudan University

---

**Research Article**

**Keywords:** COVID-19, Genomic sequences, clinical outcomes, virulence

**Posted Date:** March 25th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-19344/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on May 20th, 2020. See the published version at <https://doi.org/10.1038/s41586-020-2355-0>.

# Abstract

At least three months have been passed since the outbreak of the severe acute respiratory disease, COVID-19 in Wuhan city, China in December 2019, caused by the infection of a novel coronavirus, SARS-CoV-2.<sup>1,2</sup> Due to its rapid spread throughout China and abroad, knowledge sharing for both its epidemiology and clinic manifestations is urgently need. Here we analyzed the clinical, molecular and immunological data from 326 confirmed cases of SARS-CoV-2 infection in Shanghai. Genomic sequences assembled from 112 quality samples together with uploaded sequences in Global Initiative on Sharing All Influenza Data (GISAID) showed a stable evolution and suggested two major lineages with differential exposure history during the earliest outbreak in Wuhan. Nevertheless, they exhibited similar virulence and clinical outcomes. Lymphocytopenia, especially the reduced CD4+ and CD8+ T cell counts upon admission, was predictive of disease progression. High level of IL-6 and IL-8 during treatment was observed in severe and critical patients and correlated with decreased lymphocyte count. The determinants of disease severity seemed to stem mostly from host factors such age, lymphocytopenia and its associated cytokine storm whereas viral genetic variation did not significantly affect the outcomes. This comprehensive analysis on the molecular, immunological and clinical data provides a panorama of the key determinants related to the disease outcomes which should be helpful for improving the current combat against this extremely aggressive pandemic.

Authors Xiaonan Zhang, Yun Tan, Yun Ling, Gang Lu, Feng Liu, and Zhigang Yi contributed equally to this work.

## Introduction

In December 2019, an outbreak of infectious pneumonia emerged in Wuhan with a strong link to contacts with a seafood market (Huanan Seafood Wholesale Market, HSWM). The causal agent was soon identified to be a novel betacoronavirus sharing around 85% identity with SARS-CoV (severe acute respiratory syndrome coronavirus)<sup>1,2</sup>, hence named SARS-CoV-2. The genome sequence of SARS-CoV-2 is closest to a bat coronavirus RaTG13<sup>1</sup>. Additional studies revealed high sequence similarity with two pangolin coronaviruses especially in the receptor binding domain<sup>3-5</sup>. The SARS-CoV-2 caused disease was later officially named coronavirus disease 19 (COVID-19). On March 11, the World Health Organization has formally declared COVID-19 a world pandemic underscoring the gravity of the current situation<sup>6</sup>. As of March 13, there have been over 133,000 laboratory confirmed cases of COVID-19 around the globe. In this ongoing epidemic, public health concerns and the urgent need for effective therapeutic and intervention measures require a deep understanding of its epidemiology, transmissibility and pathogenesis.

In the last two decades, zoonotic coronaviruses have caused two large-scale outbreaks, i.e. SARS<sup>7,8</sup> and MERS<sup>9</sup> (Middle East respiratory syndrome) by crossing inter-species barrier. However, the scale of the current SARS-CoV-2 outbreak has overwhelmingly surpassed the previous two. During the SARS and MERS epidemics, continued monitoring of viral genomes from various geographical regions and different

phases of the outbreak was instrumental for understanding the source, evolution and spread of the diseases. In contrast, although HSWM is widely believed to be the birthplace of SARS-CoV-2, in the initial phase of the outbreak (December 2019), there were significant number of cases (before Jan 1, 2020) who did not have contact with this market<sup>10</sup>. This raises doubt on its singular origin. Second, the question that whether SARS-CoV-2 has evolved leading to altered virulence during the spread of the disease remains to be answered with additional molecular epidemiological data. Thirdly, although SARS-CoV-2 infection can cause life-threatening respiratory disease, a majority of infected cases manifested mild pneumonia<sup>11</sup>. The underlying viral and host factors associated with disease outcome are yet to be fully characterized.

In this study, we systematically analyzed key immunological parameters spanning the course of infection and obtained viral genomes directly from clinical samples in order to delineate key factors associated with prognosis and to monitor epidemiological feature of the epidemic.

## Results

### Overview of the enrollment

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, which was the designated hospital receiving all the COVID-19 cases in Shanghai, were included in this study. Their basic clinical and epidemiological information were 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. Among them, 90 were residents of Hubei province coming to Shanghai. 80 were Shanghai residents who had recently been to Hubei province, 52 contacted people from Hubei and 104 had unidentifiable exposures. Four categories of infected cases were defined: Five individuals were *asymptomatic* as having no obvious fever, respiratory symptoms or radiological manifestations (asymptomatic cases). A majority of the patients, 292 patients had mild disease with fever and radiological manifestations of pneumonia (mild cases). There were 12 patients with *severe* symptoms of dyspnea (respiratory rate >30/min) and signs of expanding ground-glass opacity in the lung within 24–48 hours and were defined as severe cases (severe case). Another 17 patients deteriorated into acute respiratory distress syndrome (ARDS) and required mechanical ventilation or extracorporeal membrane oxygenation (ECMO) and thus, were categorized as *critical* (Extended Data Table 1). As of March 13, 308 patients had been discharged, three had deceased (fatality rate 1.14%). The most common comorbidities 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).

### Nucleotide variation in viral genomes

We sequenced the viral genome using sputum and oropharyngeal swab samples from enrolled patients. Sequencing data from a total of 112 samples passed quality control and were used for nucleotide variation calling (Extended Data Fig. 1). As compared to the first genome of isolated virus (Wuhan-Hu-1), a total of 66 synonymous and 103 nonsynonymous variants were identified in nine protein coding regions (Extended Data Fig. 2a-b). Mutation rates of ORF1ab, S, ORF3a, E, M and ORF7a were similar (~0.5%), while variation rates of the ORF8 (1.37%) and N (1.51%) were much higher (Extended Data Fig. 2a-b). The recurrence of variations in the viral genome is similar between Shanghai samples and the GISAID datasets (Extended Data Fig. 2c).

## Genomic phylogeny analysis

We next used 94 samples with over 90% genome coverage together with 221 sequences of SARS-CoV-2 (GISAID) for the phylogenetic analysis. Two major clades were identified (Fig. 1a), both of which included some earliest reported cases<sup>1,2</sup>. Several subclades in the clade I, such as those characterized by ORF3a: p.251G>V (subclade V), or S: p.614D>G (subclade G) were also observed (Fig. 1a). Clade II is distinguished from that of Clade I by two linked variations ORF8: p.84L>S (28144T>C) and ORF1ab: p.2839S (8782C>T). The sequences of the Shanghai cohort were found throughout the two major clades and all of their subclades, suggesting multiple origins of transmission into Shanghai. No significant expansion of clades/subclades in Shanghai were observed. Although it is consistent with the stable status of this viral genome observed since the first isolate, it may also be attributable to the limited transmission locally in Shanghai due to the effective control with early detection, reporting and isolation.

Additionally, we found that six cases with clear contact history to the HSWM<sup>1,2</sup>, the suspected early outbreak site, were all clustered into clade I, while three cases without contact history to HSWM<sup>12,13</sup> were all clustered into clade II (Fig. 1a), implying that the virus might not exclusively originate from HSWM. The sequences around nt8,782 and nt28,144 of SARS-CoV-2 were analyzed in HSWM/non-HSWM-related samples and bat coronavirus Bat-SARS-CoV-RaTG13 (Fig. 1b). The non-HSWM sequences were identical to Bat-SARS-CoV-RaTG13 at these two sites, suggesting that clade II might be an evolutionarily ancestral form.

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 ( $p = 1.00$ , Fisher's exact test), lymphocytes count ( $p = 0.79$ ), CD3 T cell counts ( $p = 0.21$ ), C-reactive protein ( $p = 0.83$ ), or D-dimer ( $p = 0.19$ ) and duration of virus shedding after onset ( $p = 0.79$ , Mann-Whitney U test) (Extended Data Table 2). Thus, these two clades of virus exhibited similar pathogenic effects despite their genome sequence diversity.

## Host factors associated with disease severity

A notable feature of this SARS-CoV-2 infection cohort was that some infected individuals (5 cases, 1.53% in our cohort) did not develop obvious symptom although significant virus shedding could be

detected. As shown in Fig. 2A, no obvious lesions in the lungs were found in an asymptomatic patient upon admission till five days afterward. In comparison, unilateral and bilateral opacity lesions were observed in a mild case (Fig. 2B) and in a critical case, the latter quickly deteriorated in just two days (Fig. 2C).

We further analyzed the immunological and biochemical parameters of the four categories of patients (Extended Data Table 3). A prominent feature of SARS-CoV-2 infection was the progressive lymphocytopenia, particularly in severe and critical categories ( $p = 8.3E-5$ , Kruskal-Wallis test). Detailed analysis of the subsets in the lymphocytes revealed that CD3+ T cells were most significantly affected ( $p = 3E-6$ , Kruskal-Wallis test), with CD4+ and CD8+ T cells sharing similar trends (CD4+ T cell,  $p = 4E-6$ , CD8+ T cell,  $p = 4E-4$ , Kruskal-Wallis test). For CD19+ B cells, although significant decline was found in critical cases ( $p = 6E-5$ , Kruskal-Wallis test), no obvious changes were observed among asymptomatic, mild and severe cases ( $p = 0.47$ ). This situation was in contrast to that of T lymphocytes where the changes of cell number was statistically significant not only in critical cases as above mentioned but also in the other three categories (asymptomatic, mild and severe) in terms of both CD3+ T cells ( $p = 0.013$ ) and CD8+ T cells ( $p = 0.004$ ). We subsequently combined the longitudinal cell counting data in each group and plotted their patterns according to the time point post onset. It was clear that the CD3+ T lymphocytes (including both CD4+ and CD8+ cells) exhibited graded decline as the disease became more severe (Fig. 3 A-C) but no such trend was found for B cell and NK cell (Fig. 3D-E). Indeed, univariate logistic regression analysis indicated that age ( $p < 0.0001$ ), lymphocyte counts upon admission ( $p < 0.0001$ ), comorbidities ( $p = 0.006$ ) and gender ( $p = 0.006$ ) were the major factors associated with disease severity (Extended Data Table 4). Multivariate analysis showed that age ( $p = 0.004$ ) and lymphocytopenia ( $p = 0.02$ ) were the two major independent factors whereas comorbidities did not reach to statistical significance.

The levels of eleven cytokines (IFN- $\alpha$ , IFN- $\gamma$ , IL-1 $\beta$ , 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 and IL-8 showed most significant changes. Remarkably, these two cytokines exhibited an inverse correlation with lymphocyte count (Fig. 4A-B, Extended Data Table 5) suggesting that lymphocytopenia could be mechanistically linked to inflammatory cytokine release. Furthermore, we combined the longitudinal cytokine data in each group and plotted their fluctuation patterns according to the time point post onset. We observed that the level of IL-6 in critical group fluctuated during treatment but was well above the other groups (Fig. 4C). A similar trend was also found in the kinetic of IL-8 (Fig. 4D). These data indicated a critical role of inflammatory cytokines in the pathogenesis of SARS-CoV-2 infection.

## Discussion

At the time of writing of this manuscript, the COVID-19 epidemic is still ongoing, with more than 133,000 laboratory-confirmed cases in more than 100 countries (more than 80,000 in China alone). Since RNA virus genomes are known to mutate quickly due to error-prone replication, uncertainties remain regarding how the virus genome diversity might influence its relationship with the host environment and cellular

context. Here we performed the largest genomic sequence analysis of SARS-CoV-2 to date, utilizing 94 assembled sequences (>90% complete) from patients' samples in our cohort in combination with 221 genomes deposited in the GISAID database. Our analysis of some very recently treated patients provides further evidence that the viral genome is largely stable during this epidemic<sup>14</sup>.

Close inspection of the major variations in the genomes revealed that there were two major haplotypes due to variations at nt 8,782 and 28,144. It is intriguing to notice that sequences recovered from patients exposed to HSWM all belong to Clade I whereas three genomes from patients without exposure to HSWM are within Clade II. Thus, these two major haplotypes are likely to represent two lineages derived from a common ancestor that independently evolved in the early December 2019 in Wuhan city, one of which (Clade I) was spawned within the HSWM. Indeed, epidemiological investigations on the earliest cases in Wuhan before Dec. 18th identified two patients linked but five not linked to HSWM<sup>10</sup>, which was consistent with this idea.

However, patients infected with clade I or II virus did not exhibit significant difference in a suite of clinical features, mutation rate or transmissibility which was at odds with the conclusion made by Tang et al, whose L/S type was based on the same two linked polymorphisms<sup>14</sup>. The presumed difference in transmissibility might be due to sampling bias since the initially uploaded sequences were recovered from a limited number of critically ill patients and duplicate assemblies from the same patients' samples were not uncommon<sup>1,2,15</sup>.

Since viral genetic variations seemed to play a minor role in clinical outcome, we further evaluated host factors affecting disease severity. Guan et al has analyzed 1099 cases of COVID-19 in China and found lymphocytopenia to be one of the most common features in laboratory tests<sup>11</sup>. In this study, we confirmed this observation and further delineated the cell types that are suppressed. We found that CD3+ T cells are the major cell type for inhibition whereas CD19+ B cell and CD16+ CD56+ NK cells exhibited less suppression. Furthermore, our longitudinal monitoring of major cytokines indicated that IL-6 and IL-8 were negatively correlated with lymphocyte count and IL-6 kinetics was highly related to the disease severity. It is therefore reasonable to postulate that the immunopathological response against SARS-CoV-2 involving "cytokine storm" and loss of CD3+ T lymphocytes could constitute, at least in part, an underlying mechanism for disease progression and ultimately fatality. The macrophages in the lung could serve the first driver of the "cytokine storm" which has been described in early phase of COVID-19 pneumonia<sup>16</sup>. The subsequent mass lymphocyte infiltration mobilized by the cytokines as observed in infected patients<sup>17</sup> and Rhesus Macaques<sup>18</sup> should explain the lymphocytopenia. Due to the unavailability of potent antiviral therapies targeting SARS-CoV-2 at present, early intervention strategies such as blockade of "cytokine storm" by using IL-6 and/or IL-6 receptor inhibitors in patients with severe lymphocytopenia could eventually bring therapeutic benefits. Indeed, the use of Tocilizumab, a monoclonal antibody targeting IL-6 receptor has been recommended for the treatment of severe COVID-19 patients by the National Health Commission of China.<sup>19</sup> While a phase II randomized clinical trial is

ongoing (Dr. XU Xiaolin, personal communication), more than 400 cases have benefited this treatment in a real world trial in China since mid-February 2020.

In conclusion, by close monitoring the molecular and immunological data in our 326 cases of COVID-19 patients, we suggest that adverse outcome is associated with depleted CD3+ T lymphocytes which is tightly linked to burst of cytokines such as IL-6 and IL-8. Targeted sequencing of additional cases infected during late-January to mid-February indicated limited variations in the viral genome suggestive of a stable evolution. Two major lineages of virus, which spread out independently from Wuhan in Dec 2019, contributed to the current pandemic although no major difference in clinical manifestation or transmissibility was found. Our data provide further evidence for the respective roles played by the viral and host factors in disease mechanism and underscore the importance of early intervention in therapeutic strategies.

## Declarations

## Author Contribution

S. C., H. Z., X. Z., S. W. Z. Y. conceived the study. X. Z., Y. L., Z. Y., X. J., M. W., B. S., S. X., J. C., W. W. collected the patient samples, epidemiological and clinical data, X. Z., X. J., M. W. processed viral RNA isolation and PCR, S. Y., J. L., L. J., G. L., J. W. performed sequencing and sequence assembly. X. Z., Y. T., F. L., G. L., B. S., S. X., J. C., S. W., S. C. M. X. accessed data acquisition, analysis, and interpretation. X. Z., Y. T., F. L., G. L. drafted the manuscript. S. C., S. W., X. Z., G. Z. revised the final manuscript.

## Competing interests

The authors declare no competing interests.

## Acknowledgement

This work is supported by National Megaprojects of China for Infectious Diseases (2017ZX10103009-001 2018ZX10305409-001-005); Double First-class project of Fudan University (No.IDF162005) and the 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) and National Key R&D Program of China (2019YFA0905902). We acknowledge all healthcare personnel involved in the diagnosis and treatment of patients in Shanghai Public Health Clinical Center, we thank Prof. Zhu Chen for guidance in study design, interpretation of results and critical revision of the manuscript. This manuscript is dedicated to Prof. Yunwen Hu of Shanghai Public Health Clinical Center, whose life-long commitment to infectious disease research would never be forgotten.

# Ethics

Here we state that this study was approved by the Shanghai Public Health Clinical Center Ethics Committee (No. YJ-2020-S015-01). Informed consents had been obtained from the enrolled patients.

## References

1. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* (2020) doi:10.1038/s41586-020-2012-7.
2. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* **382**, 727–733 (2020).
3. Lam, T. T.-Y. *et al.* Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. <http://biorxiv.org/lookup/doi/10.1101/2020.02.13.945485> (2020) doi:10.1101/2020.02.13.945485.
4. Xiao, K. *et al.* Isolation and Characterization of 2019-nCoV-like Coronavirus from Malayan Pangolins. <http://biorxiv.org/lookup/doi/10.1101/2020.02.17.951335> (2020) doi:10.1101/2020.02.17.951335.
5. Wong, M. C., Javornik Cregeen, S. J., Ajami, N. J. & Petrosino, J. F. Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019. <http://biorxiv.org/lookup/doi/10.1101/2020.02.07.939207> (2020) doi:10.1101/2020.02.07.939207.
6. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020>.
7. The Chinese SARS Molecular Epidemiology Consortium. Molecular Evolution of the SARS Coronavirus During the Course of the SARS Epidemic in China. *Science* **303**, 1666–1669 (2004).
8. Guan, Y. *et al.* Molecular epidemiology of the novel coronavirus that causes severe acute respiratory syndrome. *The Lancet* **363**, 99–104 (2004).
9. Cotten, M. *et al.* Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study. *The Lancet* **382**, 1993–2002 (2013).
10. Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* NEJMoa2001316 (2020) doi:10.1056/NEJMoa2001316.
11. Guan, W.-J. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMoa2002032.
12. Chan, J. F.-W. *et al.* A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* **395**, 514–523 (2020).
13. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* **395**, 565–574 (2020).

14. Tang, X. *et al.* On the origin and continuing evolution of SARS-CoV-2. *National Science Review* nwaa036 (2020) doi:10.1093/nsr/nwaa036.
15. Ren, L.-L. *et al.* Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. *Chinese Medical Journal* 1 (2020) doi:10.1097/CM9.0000000000000722.
16. Tian, S. *et al.* Pulmonary pathology of early phase 2019 novel coronavirus (COVID-19) pneumonia in two patients with lung cancer. *J Thorac Oncol* (2020) doi:10.1016/j.jtho.2020.02.010.
17. Xu, Z. *et al.* Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir Med* (2020) doi:10.1016/S2213-2600(20)30076-X.
18. Infection with Novel Coronavirus (SARS-CoV-2) Causes Pneumonia in the Rhesus Macaques. (2020) doi:10.21203/rs.2.25200/v1.
19. National Health Commission of China. Guidelines for the Diagnosis and Treatment of COVID-19 (7th edition).

## Methods

### Patients

A total of 326 SARS-CoV-2 positive patients hospitalized in Shanghai Public Health Clinical Center were analyzed. Their basic demographic, epidemiological and clinical characteristics were shown in Extended Data Table 1. The disease severity was categorized into four stages, i.e., asymptomatic, mild, severe and critical, according to the Guidelines on the Diagnosis and Treatment of novel coronavirus issued by the National Health Commission, China (7th edition). Briefly, asymptomatic disease was defined as normal body temperature, lack of respiratory symptoms and no pulmonary radiological manifestation, mild disease was defined as having fever, respiratory symptoms and radiological evidence of pneumonia. Severe cases were defined by meeting one of the following manifestations: respiratory rate > 30/min, oxygen saturation levels (SpO<sub>2</sub>) < 93%, arterial partial pressure of oxygen (PaO<sub>2</sub>)/fraction of inspired oxygen (FiO<sub>2</sub>)(PaO<sub>2</sub>/FiO<sub>2</sub> ratio) < = 300mmHg or pulmonary imaging shows multi-lobular lesions or lesion progression exceeding 50% within 48 hours. Critical condition was defined as one of the following: acute respiratory distress syndrome requiring mechanical ventilation, shock, complications with other organ failure.

### Nucleic acid extraction, molecular screening and genome sequencing

Swabs and sputum samples were collected for nucleic acid extraction using automatic magnetic extraction device and accompanying kit (Shanghai Bio-Germ) and screened with a semi-quantitative RT-PCR kit (Shanghai Bio-Germ) with amplification targeting the ORF1a/b and N gene. Deep sequencing was done using the nucleic acid extracted from the RT-PCR confirmed cases hospitalized in Shanghai Public Health Clinical Center. We used a multiplexed amplicon strategy as described<sup>1</sup> the primers were

synthesized as described ([https://github.com/artic-network/artic-ncov2019/blob/master/primer\\_schemes/nCoV-2019/V1/nCoV-2019.tsv](https://github.com/artic-network/artic-ncov2019/blob/master/primer_schemes/nCoV-2019/V1/nCoV-2019.tsv)). The primers was split into 10 subpools each containing 9–10 pairs for specific amplification of 400bp 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. Briefly, the procedures included end-repair, 3' ends adenylation, adapter 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.

## Viral genomic sequence variation calling

All clean reads were mapped to the SARS-CoV-2 genome (Wuhan-Hu-1) using BWA (version 0.7.17)<sup>2</sup>. Variations were called with mpileup tools in samtools<sup>3</sup>. Low-quality variations that depth lower than 10 and Qual score lower than 50 were filtered by bcftools.

## Phylogenetic analysis

Sequencing reads were trimmed with Trimmomatic (version 0.39)<sup>4</sup> to remove low quality region, adapter sequences and sequencing primers<sup>5</sup>. Clean reads were used to build virus genome assemblies with VirGenA (version 1.4)<sup>5</sup>. A post-assembly procedure was manually performed to remove low quality content and potential sequencing artifacts. 94 assemblies with coverage above 90% were qualified for phylogeny analysis. MAFFT (version 7.453)<sup>6</sup> made the multi-sequence alignment after trimming off N's on both ends of the genome sequences. IQ-TREE (version 1.6.12)<sup>7</sup> package was used in substitution model selection and phylogeny tree construction. TIM+F substitution model was chosen to build the Maximum Likelihood phylogeny tree with 1000 bootstraps. TreeTime (version 0.7.3)<sup>8</sup> was used for time resolved phylogeny analysis. The resulting phylogeny tree was visualized by auspice from Nextstrain package (version 1.15.0)<sup>9</sup>. Genome sequences used in this analysis were deposited to the GISAID database and the phylogeny result is accessible via web address <http://ncov.linc.org.cn>.

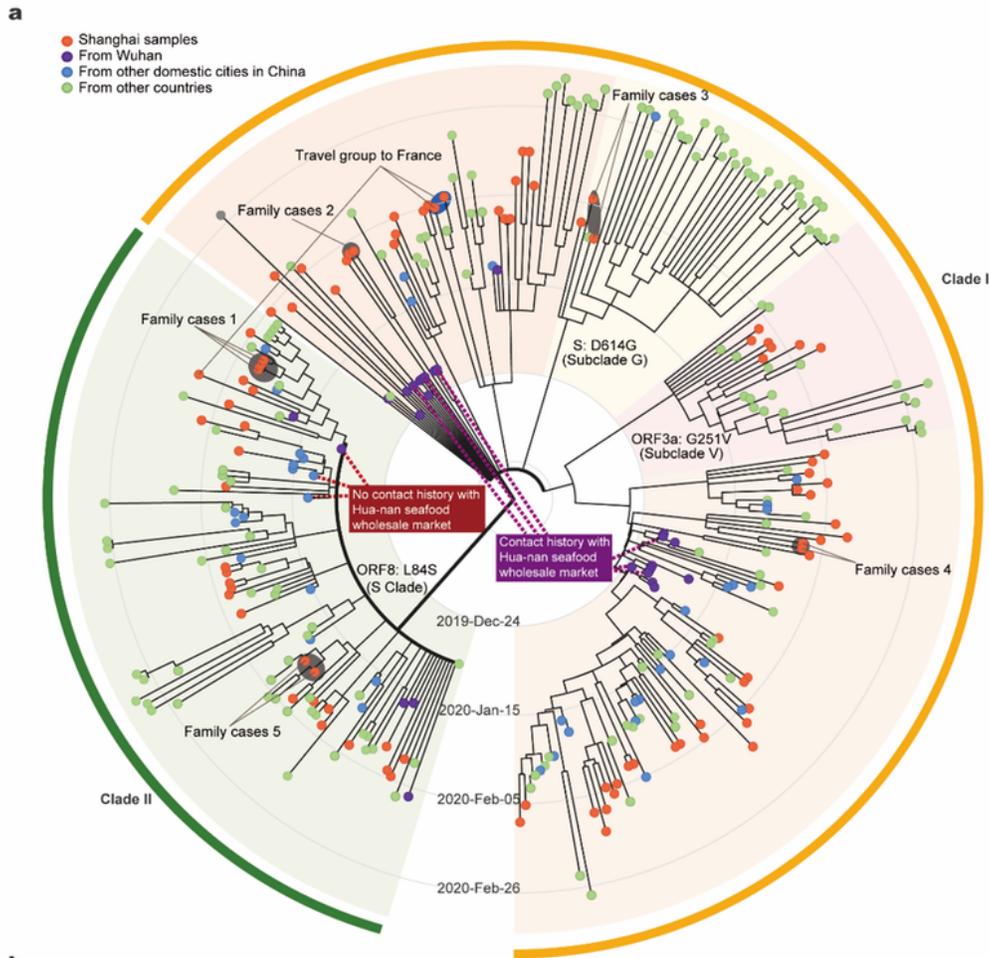
## Cytokine quantification and lymphocyte subset counting

Becton-Dickinson (BD) cytometric bead array (human Th1/Th2/Th17 cytokine kit and Human Inflammatory Cytokine Kit) was used quantify serum cytokines (IFN- $\alpha$ , IFN- $\gamma$ , IL-1 $\beta$ , IL-2, IL-4, IL-5, IL-6, IL-8, IL-10, IL-12 and IL-17). The absolute counts of CD3+ T as well as the CD4 and CD8 subpopulations, CD19+ B, and CD16+CD56+ natural killer (NK) cells was stained using BD Multitest™ 6-color TBNK reagent with Trucount tubes and analyzed using the BD FACSCanto™II flow cytometer. The longitudinal cytokine and cell count data were visualized using the geom\_smooth tool in the ggplot2 package.

## References for Methods

1. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* **12**, 1261–1276 (2017).
2. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
3. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
4. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
5. Fedonin, G. G., Fantin, Y. S., Favorov, A. V., Shipulin, G. A. & Neverov, A. D. VirGenA: a reference-based assembler for variable viral genomes. *Briefings in Bioinformatics* **20**, 15–25 (2019).
6. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
7. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. - PubMed - NCBI. <https://www.ncbi.nlm.nih.gov/pubmed/?term=IQ-TREE%3A+a+fast+and+effective+stochastic+algorithm+for+estimating+maximum-likelihood+phylogenies>.
8. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol* **4**, vex042 (2018).
9. Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).

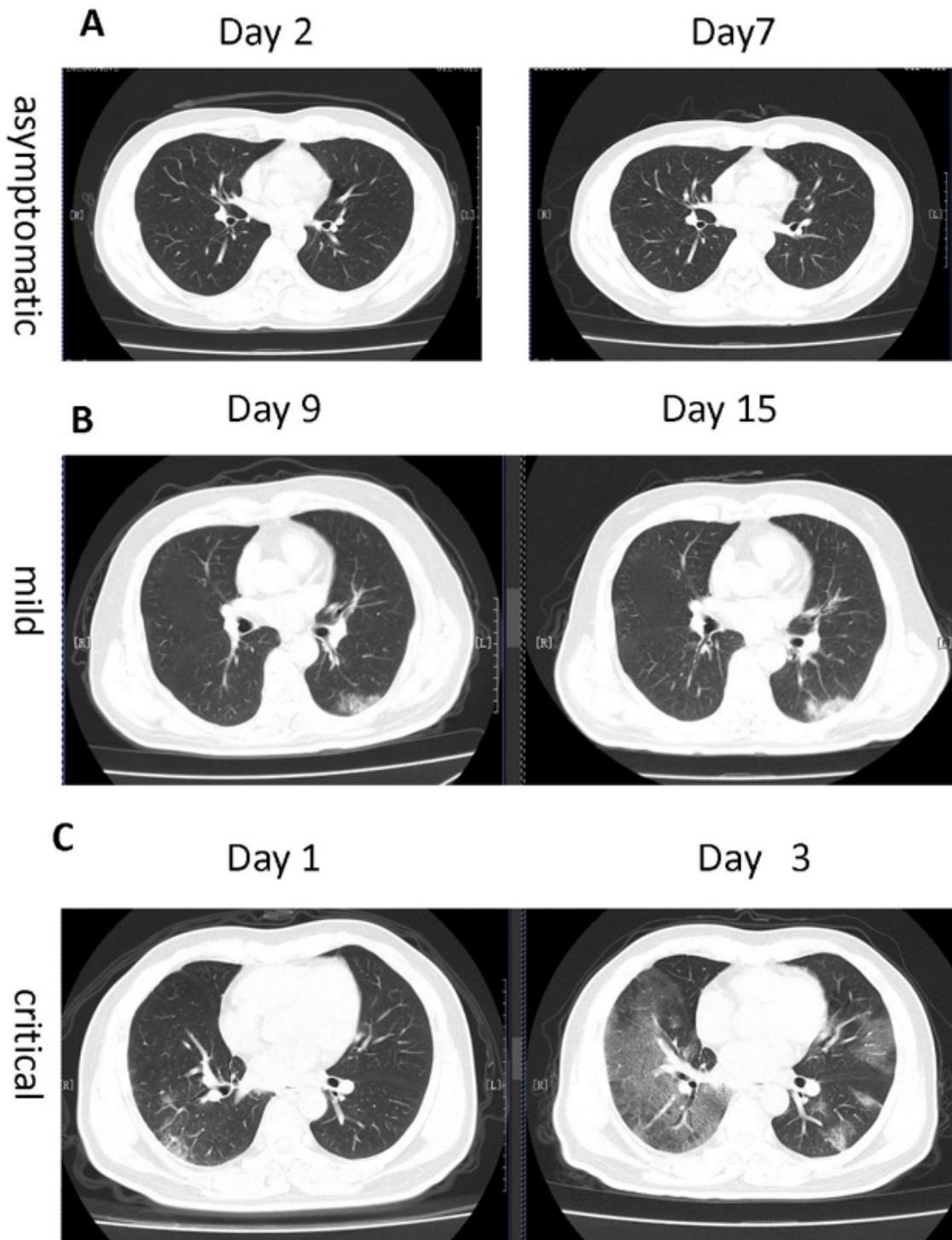
## Figures



**Figure 1**

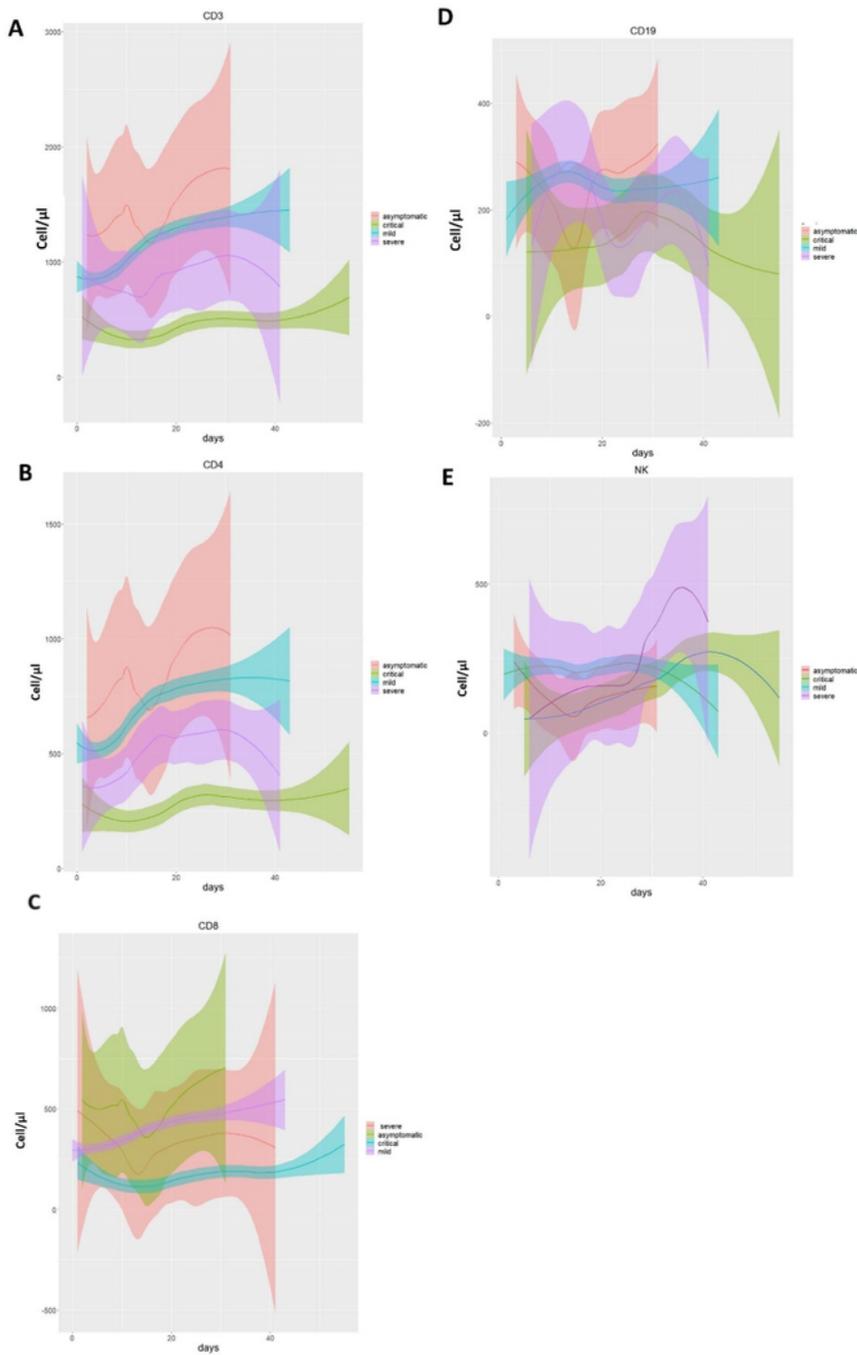
a, A total of 94 SARS-CoV-2 genome sequences and 221 published sequences were used for construction of time-resolved phylogeny tree. Clade I and clade II are marked and variations that determines the branches of the phylogeny tree were indicated. Concentric circles at the graph representing time scale of sampling date. Each tip circle stands for a single sample, while the color marks the case location (red represent Shanghai, purple represents Wuhan, blue represents other domestic cities in China, and green

represents other countries and districts). Case related to HSWM were highlighted. b, Alignment of sequences around nucleotide 8782 and 28144 using public sequences recovered from patients with known history of exposure to HSWM and Bat-SARS-CoV-RaTG13.



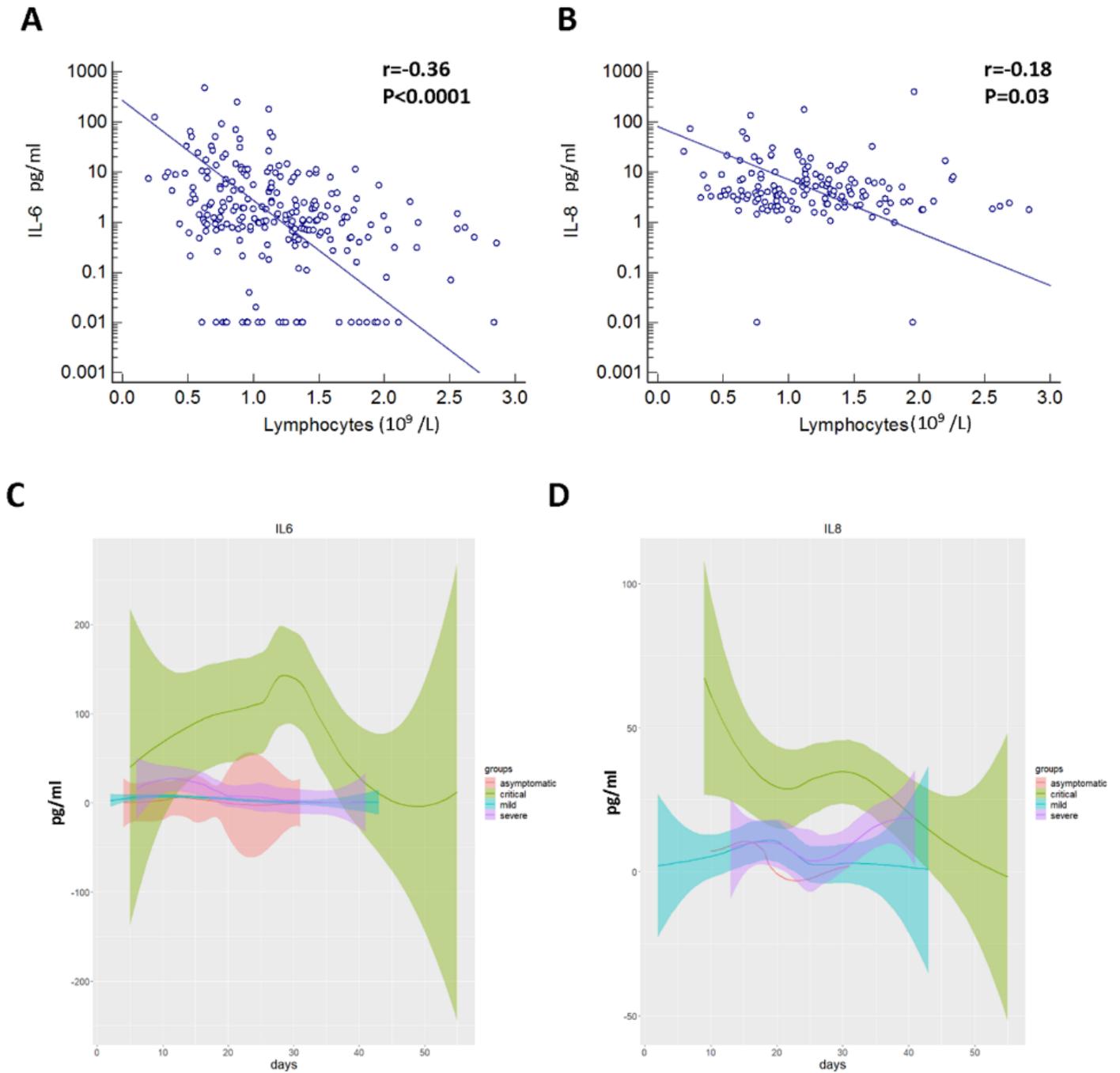
**Figure 2**

Computed tomographic scans of three typical patients (A, asymptomatic, B, mild and C, critical ) during hospitalization.



**Figure 3**

Temporal changes in CD3+ (A), CD4+ (B), CD8+ (C), CD19+ (E) and CD16+ CD56+ (F) cell counts in each group during hospitalization. The median value of each time-point (day from onset) was connected using geom smooth method and the 95% interval was plotted as colored shadow.



**Figure 4**

Correlation between inflammatory cytokines and lymphocyte counts. The level of serum IL-6 (A) and IL-8 (B) upon admission was plotted against lymphocyte count and Spearman's correlation analysis was performed. Temporal changes of IL-6 (C) and IL-8 (D) in each group during hospitalization was shown. The median value of each time-point (day from onset) was connected and the 95% interval was plotted as colored shadow.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [LuExtendedData.pdf](#)