Serum KL-6 levels predict clinical outcomes and are associated with MUC1 polymorphism in Japanese patients with COVID-19

Shuhei Azekawa,1 Shotaro Chubachi,1 Takanori Asakura,1,2,3 Ho Namkoong,4 Yasunori Sato,5 Ryuya Edahiro,6,7 Ho Lee,1 Hiromu Tanaka,1 Shiro Otake,1 Kensuke Nakagawara,1 Takahiro Fukushima,1 Mayuko Watase,1 Kaori Sakurai,1 Tatsuya Kusumoto,1 Katsunori Masaki,1 Hirofumi Kamata,1 Makoto Ishii,8 Naoki Hasegawa,4 Yukinori Okada,6,9,10,11,12,13 Ryuji Koike,14 Yuko Kitagawa,15 Akinori Kimura,16 Seiya Imoto,17 Satoru Miyano,18 Seishi Ogawa,19 Takanori Kanai,20 Koichi Fukunaga,1 The Japan COVID-19 Task Force

ABSTRACT

Background Krebs von den Lungen-6 (KL-6) is a known biomarker for diagnosis and monitoring of interstitial lung diseases. However, the role of serum KL-6 and the mucin 1 (MUC1) variant (rs4072037) in COVID-19 outcomes remains to be elucidated. We aimed to evaluate the relationships among serum KL-6 levels, critical outcomes and the MUC1 variant in Japanese patients with COVID-19.

Methods This is a secondary analysis of a multicentre retrospective study using data from the Japan COVID-19 Task Force collected from February 2020 to November 2021, including 2226 patients with COVID-19 whose serum KL-6 levels were measured. An optimal serum KL-6 level cut-off to predict critical outcomes was determined and used for multivariable logistic regression analysis. Furthermore, the relationship among the allele dosage of the MUC1 variant, calculated from single nucleotide polymorphism typing data of genome-wide association studies using the imputation method, serum KL-6 levels and COVID-19 critical outcomes was evaluated.

Results Serum KL-6 levels were significantly higher in patients with COVID-19 with critical outcomes (511±442 U/mL) than those without (279±204 U/mL) (p<0.001). Serum KL-6 levels ≥304 U/mL independently predicted critical outcomes (adjusted OR (aOR) 3.47, 95% CI 2.44 to 4.95). Moreover, multivariable logistic regression analysis with age and sex indicated that the MUC1 variant was independently associated with increased serum KL-6 levels (aOR 0.24, 95% CI 0.28 to 0.32) but not significantly associated with critical outcomes (aOR 1.11, 95% CI 0.80 to 1.54).

Conclusion Serum KL-6 levels predicted critical outcomes in Japanese patients with COVID-19 and were associated with the MUC1 variant. Therefore, serum KL-6 level is a potentially useful biomarker of critical COVID-19 outcomes.

INTRODUCTION

COVID-19 was first discovered in China in December 2019 and has since spread globally,1 infecting more than 600 million people and killing 6.5 million.2 Despite great advances in the development of vaccines and treatments, the disease continues to affect numerous individuals worldwide. Therefore, developing biomarkers that can predict COVID-19 severity is essential for effectively using healthcare resources and identifying targets for new therapeutic agents.3

Krebs von den Lungen-6 (KL-6) is a mucin-like, high-molecular-weight glycoprotein classified as a mucin 1 (MUC1) antigen.4,5 Serum KL-6 levels are a biomarker for diagnosis and monitoring of disease activity in patients with various interstitial lung diseases (ILDs).6 Its
levels are also increased in various respiratory infections, including pneumocystis pneumonia,7 mycobacterial infections8 9 and viral infection.10 Since the beginning of the COVID-19 pandemic, elevated serum KL-6 levels have been reported to be associated with COVID-19 severity.11-15 Moreover, the extent of lung lesions in COVID-19-associated pneumonia is related to increased serum KL-6 levels and poor prognosis.16 17 However, some studies have found no association between KL-6 levels and disease severity.18-20 Therefore, the clinical role of KL-6 in COVID-19 is still controversial, which might be a consequence of the varied and limited sample sizes of the aforementioned studies.

KL-6 levels vary following MUC1 gene polymorphisms and demonstrate racial differences.21 22 In ILDs, MUC1 polymorphisms have been reported to be associated with disease susceptibility.21-25 However, no studies conducted so far have focused on the relationships among serum KL-6 levels, KL-6-associated genetic polymorphisms and COVID-19 severity. The Japan COVID-19 Task Force was founded in early 2020 as a national multicentre consortium to combat COVID-19. Since February 2020, more than 100 institutions across Japan have contributed clinical data and specimens from patients with COVID-19 to improve knowledge of this disease.26-31 Using this large pool of available data, the present study aimed to (1) determine whether serum KL-6 is associated with important clinical outcomes, such as mortality, extracorporeal membrane oxygenation (ECMO) and invasive positive pressure ventilation (IPPV), that more directly reflect COVID-19 prognosis than its severity; and (2) clarify the association among serum KL-6 levels, important clinical outcomes and KL-6-associated genetic polymorphisms.

METHODS
Study design and setting
Data for all COVID-19 cases used for the present secondary analysis of a multicentre retrospective study were obtained by the Japan COVID-19 Task Force from February 2020 to September 2021.26 Patients were at least 18 years old and diagnosed with COVID-19 based on detection of SARS-CoV-2 via PCR or antigen analysis. The data of patients who agreed to participate in the present study were registered in an electronic case record form. All patients who participated in the present study provided written or oral informed consent.

Figure 1 shows the enrolment process. Of the 3424 patients enrolled, 1198 were excluded because they lacked clinical information on KL-6 levels (956 patients), were non-Japanese (69 patients), presented malignancy (165 patients) or presented an ILD as a comorbidity (8 patients).6 32 Finally, 2226 patients were analysed. For assessing KL-6-associated genetic polymorphisms, 1119 patients available for genome-wide association study (GWAS) were included.

Data collection and definition
The following clinical information was extracted from the electronic case record forms: age, gender, height, weight, smoking history, comorbidities, symptoms and signs at initial presentation, blood test results, and radiological test results. Symptoms and signs included those detected at the time of referral and on admission and those detected during hospitalisation. Blood results and radiographic imaging using chest X-ray (CXR) and CT were collected within 48 hours of admission. The attending physician at each facility determined when to perform the follow-up CXR. Rapid deterioration of CXR was defined as a deterioration of lung infiltrates in more than 50% of the lung fields within 48 hours compared with the CXR on admission, based on the criteria for severe disease in COVID-19-positive patients.33 Serum KL-6 levels were measured using the Nanopia KL-6 Reagent Kit (Sekisui Medical, Tokyo, Japan). KL-6 levels were measured in the serum within 48 hours of admission, but longitudinal KL-6 levels were not collected. COVID-19 severity was defined as follows: most severe, patients requiring support using high-flow oxygen devices, IPPV or ECMO, or cases leading to death; severe, patients requiring support using low-flow oxygen devices; and asymptomatic/mild, asymptomatic/symptomatic patients not requiring oxygen support.26 To investigate the association between serum KL-6 levels and clinically important outcomes, critical outcomes were defined as follows: use of IPPV, ECMO or high-flow oxygen supply or patient death during hospitalisation.27

Genotype imputation
GWAS genotyping was performed on samples from 2520 COVID-19 cases using the Illumina Asian Screening Array (Illumina, San Diego, California, USA). We applied stringent quality control (QC) filters to the samples and variants. The 2393 COVID-19 cases passed the sample QC. Details of the QC are described elsewhere.26 After QC, we performed genome-wide genotype imputation. Details of
genotype imputation are described elsewhere. Out of 2393 patients with COVID-19, 1119 cases had information on rs4072037 mutation. The mutation of rs4072037 in \( MUC1 \) is known to be associated with serum KL-6 levels. The dosage effects of rs4072037 on serum KL-6 levels were evaluated using a linear regression model with age and sex as covariates.

### Statistical analysis

For baseline characteristics, summary statistics were performed using frequencies and proportions for categorical data, and means and SD for continuous variables. For comparisons between the groups with high and low KL-6 levels, unpaired t-test or Wilcoxon rank-sum test was used for continuous variables, while \( \chi^2 \) test or Fisher's exact test was used for categorical variables. The Dunnett's multiple comparisons test was used to compare serum KL-6 levels among the three groups classified following the type of chest imaging findings. Receiver operating characteristic (ROC) curve analysis was performed to determine appropriate serum KL-6 cut-off values for critical outcomes using the Youden index. To investigate the association between KL-6 levels and critical outcomes, a multivariable logistic regression analysis was performed, adjusting for the following clinical characteristics: age, sex, hypertension, haemoglobin A1c (HbA1c) \( \geq 6.5\% \), cardiovascular disease, chronic obstructive pulmonary disease (COPD), asthma, chronic obstructive pulmonary disease.
Liver disease and chronic kidney disease. To elucidate the clinical relevance of serum KL-6 levels, we calculated the Pearson correlation coefficients between KL-6 and inflammatory biomarkers and performed multivariable logistic regression analyses for critical outcomes using these inflammatory biomarkers as covariates, in addition to serum KL-6 and previously reported risk factors.

RESULTS
Baseline characteristics of patients
Patients with COVID-19 whose serum KL-6 levels were measured on admission had more severe disease and more frequently had critical outcomes than those whose serum KL-6 levels were not measured (online supplemental table 1). The baseline clinical characteristics of the 2226 patients included in the analysis are shown in table 1. The mean age of the patients was 56.3 years (SD, 17.0), and 1529 were male (68.7%). The 434 patients (19.5%) experiencing a critical outcome during hospitalisation were significantly older than the patients without such characteristics. Regarding CT images, patients with bilateral ground glass opacity (GGO) and consolidation were significantly more likely to show comorbidities, such as hypertension (50.9% vs 29.0%), diabetes mellitus (33.0% vs 14.1%), cardiovascular disease (14.7% vs 8.4%), COPD (7.1% vs 3.0%), hyperuricaemia (15.9% vs 9.0%) and chronic kidney diseases (15.7% vs 5.0%), than the group not experiencing such outcomes.

Figure 2 Differences in serum KL-6 levels based on the chest imaging findings on admission. Group comparisons were performed using Dunnett’s multiple comparisons test. *P<0.05, **P<0.01, ***P<0.001. CXR, chest X-ray; GGO, ground glass opacity; KL-6, Krebs von den Lungen-6.

Figure 3 Comparison of serum KL-6 levels with and without post-hospitalisation complications (bacterial infection, thrombosis, heart failure and fungal infection). Unpaired t-test was performed. ***P<0.001. KL-6, Krebs von den Lungen-6.

KL-6 levels predict in-hospital complications
Figure 3 depicts the serum KL-6 levels following the presence or absence of complications during hospitalisation. Serum KL-6 levels were significantly higher in the group with bacterial and/or fungal infection, thrombosis and heart failure than in the group without these complications.

KL-6 levels predict critical outcomes
Regarding disease severity on admission, patients with severe or most severe disease had significantly higher serum KL-6 levels than those with asymptomatic/mild disease (online supplemental figure 1). Serum KL-6 levels were also higher in patients with a rapid deterioration of chest imaging findings within 48 hours of admission compared with those without (online supplemental figure 1). Figure 4A depicts the serum KL-6 levels following COVID-19 severity. Serum KL-6 levels were significantly higher in patients with severe and most severe COVID-19 than in those with mild COVID-19 severity (most severe 510±443 U/mL vs mild 248±162 U/mL; severe 333±163 U/mL vs mild 248±162 U/mL). Serum KL-6 levels were significantly higher in patients with critical outcomes than in those without them (figure 4B). Figure 4C shows the oxygen demand and mortality on admission and at the time of the most severe illness for patients with high and low serum KL-6 on admission. It was shown that nearly half of the patients with elevated serum KL-6 levels on admission have a higher oxygen demand than on admission and that the oxygen demand tended to increase over time. In addition, ROC curve analysis was performed to evaluate the diagnostic value of KL-6 levels for the critical outcomes. The optimal cut-off value of KL-6 levels for the critical outcomes was 304 U/mL, with an area under the curve of 0.73 (95% CI 0.71 to 0.77) (figure 4D). Each critical outcome was consistent with this result (online supplemental figure 2). Even after excluding patients with critical outcomes on admission, serum KL-6 levels were associated with subsequent poor outcomes (online supplemental table 2). Clinical characteristics were more frequently male than female (74.7% vs 67.2%). The group experiencing critical outcomes was also significantly more likely to show comorbidities, such as hypertension (50.9% vs 29.0%), diabetes mellitus (33.0% vs 14.1%), cardiovascular disease (14.7% vs 8.4%), COPD (7.1% vs 3.0%), hyperuricaemia (15.9% vs 9.0%) and chronic kidney diseases (15.7% vs 5.0%), than the group not experiencing such outcomes.
Figure 4 (A) Comparison of serum KL-6 levels following the severity of COVID-19. Group comparisons were performed using Dunnett’s multiple comparisons tests. (B) Serum KL-6 levels of patients with or without critical outcomes. Unpaired t-tests were performed. (C) Alluvial plot showing trends on admission, worst oxygen demand and mortality by serum KL-6 level on admission. ‘KL-6 High’ means serum KL-6 ≥304 U/mL and ‘KL-6 Low’ means serum KL-6 <304 U/mL. ‘High Flow’ means that high-flow nasal cannula or non-invasive positive pressure ventilation was used for treatment. ‘Low Flow’ means some oxygen demand without IPPV or high-flow nasal cannula or non-invasive positive pressure high-flow nasal cannula or non-invasive positive pressure was used for treatment. (D) Receiver operating characteristic curve for outcomes of COVID-19 based on the serum levels of KL-6. ***P<0.001. AUC, area under the curve; IPPV, invasive positive pressure ventilation; KL-6, Krebs von den Lungen-6.

Figure 5 Multivariable logistic regression analysis for ascertaining the relationship between critical outcomes and serum KL-6 and already known risk factors for COVID-19. Forest plot showing the adjusted OR, 95% CI and p value of the parameters of the analysis. BMI, body mass index; COPD, chronic obstructive pulmonary disease; HbA1c, haemoglobin A1c; KL-6, Krebs von den Lungen-6.

Clinical relevance of MUC1 polymorphisms
Table 2 shows that the allele dosage of the MUC1 variant (rs4072037) adjusted for age and sex was significantly associated with serum KL-6 levels (aOR 0.24, 95% CI 0.28 to 0.32). This suggested that the presence of the C allele, a reference allele of rs4072037, was associated with increased KL-6 levels.

DISCUSSION
The present study provided two novel findings with clinical relevance. First, serum KL-6 levels could predict the clinical outcomes of Japanese patients with COVID-19 in this large multicentre study; additionally, we successfully determined the optimal cut-off value of serum KL-6 levels to predict such critical outcomes. It has already been reported that serum KL-6 effectively predicts the severity of COVID-19 or fibrotic change in patients with severe COVID-19. However, in a large patient population including patients with mild disease, for the first time in this study, KL-6 was stratified according to disease severity and was shown to be associated with hard endpoints. The present study also allowed us to perform a multivariate analysis where we showed an association between KL-6 levels and clinical outcomes using confounding factors, including biomarkers other than serum KL-6 and known risk factors. Moreover, the results showed that serum KL-6 levels were related to complications during hospitalisation. Therefore, the serum KL-6 levels of patients showed that KL-6 levels are independently associated with IPPV (aOR 3.81, 95% CI 2.59 to 5.61) and ECMO (aOR 6.41, 95% CI 2.56 to 16.05), but not mortality (aOR 1.52, 95% CI 0.77 to 3.03) (online supplemental figure 3). Serum CRP, ferritin and D-dimer levels were weakly correlated with serum KL-6 levels (online supplemental table 4). Multivariable logistic regression analysis also showed that KL-6 was independently associated with CRP, ferritin and D-dimer for critical outcomes (online supplemental figure 4).

compared by dividing the patients into two groups: the high KL-6 group (KL-6 ≥304 U/mL) and the low KL-6 group (KL-6 <304 U/mL) (online supplemental table 3). Concerning clinical signs, the high KL-6 group had significantly higher rates of unconsciousness and shortness of breath, both suggesting severe COVID-19, than the low KL-6 group. Regarding laboratory data, white cell count, lactate dehydrogenase, uric acid, ferritin, D-dimer and C reactive protein (CRP) levels were significantly higher in the high KL-6 group than in the low KL-6 group. These serological biomarkers are known to reflect inflammation in patients with COVID-19.

The multivariable logistic regression analysis revealed that high KL-6 levels (adjusted OR (aOR) 3.47, 95% CI 2.44 to 4.95) as well as older age (aOR 1.02, 95% CI 1.01 to 1.04), male status (aOR 1.63, 95% CI 1.08 to 2.46), increased body mass index (aOR 1.05, 95% CI 1.01 to 1.09), chronic kidney disease (aOR 2.71, 95% CI 1.64 to 4.40) and HbA1c levels ≥6.5% (aOR 1.95, 95% CI 2.44 to 4.95) were independently associated with critical outcomes (figure 5). Using the same covariates as for critical outcomes, multivariable logistic regression analyses
with COVID-19 may be a clinical indicator for providing more aggressive treatment. Second, the MUC1 variant (rs4072037) was related to serum KL-6 levels but not to the critical outcomes of patients with COVID-19. To the best of our knowledge, this is the first study to examine the association of MUC1 polymorphism with serum KL-6 levels and COVID-19 clinical outcomes.

The present study provides strong evidence that KL-6 levels can predict the critical outcomes of COVID-19 through significant multicentre validation. However, the efficacy of several serological biomarkers, such as those related to the inflammatory response and coagulation predisposition, in predicting COVID-19 critical outcomes has been previously reported. Notably, our results also showed that KL-6 was independently associated with critical outcomes, even when adding CRP, ferritin and D-dimer, which are known serological biomarkers of COVID-19, to the multivariable analysis. This suggests that serum KL-6 levels can be a complementary measurement of these biomarkers (online supplemental figure 4).

Serum KL-6 is a known biomarker of lung injury and fibrosis, reflecting damage or regeneration of type 2 pneumocytes, based on its previously reported association with the severity and prognosis of ILDs and acute respiratory distress syndrome (ARDS). In the present study, KL-6 levels were also associated with the degree of COVID-19-derived pneumonia on imaging evaluation. A possible mechanism underlying the increased serum KL-6 levels in severe COVID-19 cases is the increased production of type 2 pneumocytes due to direct viral infection and leakage into the blood due to disruption of the alveolar–capillary barrier. Patients with severe COVID-19 often end up having ARDS, and deaths due to COVID-19 are associated with a higher incidence of ARDS than that in COVID-19 survivors. It has been reported that autopsy samples of the lungs of patients with COVID-19 show a pattern of diffuse alveolar damage (DAD), a pathology of ARDS. Therefore, it can be inferred that the primary mechanism underlying the increase in serum KL-6 levels is the leakage of KL-6 from the lungs into the blood due to virus proliferation and destruction of the alveolar epithelium and basement membrane with DAD. In fact, not only blood but also epithelial lining fluid KL-6 levels are increased in patients with ARDS, and high levels of this biomarker have also been reported in the bronchoalveolar lavage fluid of patients with severe COVID-19.

In the present study, KL-6 levels were associated with bilateral shadows on imaging findings, consistent with the results of previous studies. The KL-6 cut-off value for predicting critical outcomes was 304 U/mL. This value was lower than the cut-off values found for diagnosing ILDs and poor prognostic factors for ARDS and ILDs. Furthermore, the KL-6 cut-off value was lower than that previously reported for predicting the severity of COVID-19. We hypothesise that this may be because previous studies had fewer participants but included more patients with relatively severe disease compared with that in the present study. The alveolar–capillary permeability in COVID-19 is elevated in the acute phase and may further increase with lung fibrosis. A previous report on COVID-19 suggested that peak serum KL-6 levels during hospitalisation are more helpful than those at diagnosis in predicting clinical outcomes. Therefore, further evaluation of serum KL-6 levels over time should be explored in future studies.

The present study also revealed the relationships between the MUC1 variant and serum KL-6 levels in Japanese patients with COVID-19. The results obtained here are consistent with those of previous reports demonstrating that the MUC1 variant affects KL-6 levels in ILDs and sarcoidosis. MUC1 is an extracellular protein anchored to the epithelial surface and involved in morphogenetic signalling. The presence of the rs4072037 SNP (single nucleotide polymorphism) causes alternative splicing of the exon regions under its regulation, leading to abnormal transcription. MUC1 has been demonstrated to be a critical innate immunity modulator, acting as an important and necessary anti-inflammatory agent during airway infection. It is controversial whether the MUC1 variant affects the severity of ILDs. In support of this hypothesis, it has been reported that the MUC1 variant affects the severity of pulmonary alveolar proteinosis. However, some studies showed that rs4072037 may be involved in disease susceptibility to ILDs but is not related to the severity of the disease. For example, high frequency of the rs4072037 C allele and serum KL-6 levels have been reported in patients with antisyntethase syndrome compared with those in healthy control subjects. It has also been reported that in patients with systemic sclerosis-ILD, the rs4072037 mutation is significantly associated with increased serum KL-6 levels but is not a predictor of a
lower diffusing capacity for carbon monoxide. In the present study, the MUC1 variant did not correlate with COVID-19 severity. Moreover, it has been reported that the optimal cut-off value of serum KL-6 levels discriminating patients with ILD from healthy control subjects differs following the rs4072037 genotype. The results of the present study suggest that high serum KL-6 levels may be due not only to severe COVID-19, but also to the MUC1 variant, which may result in false-positive serum KL-6 in patients with severe COVID-19. Although it is not possible to measure MUC1 variants in actual clinical practice at this time, caution should be exercised in interpreting serum KL-6 in patients with COVID-19.

Recently, it has been recognised that COVID-19 symptoms can persist after the disease, which is referred to as long COVID or postacute COVID-19 syndrome. It is estimated that 22.9% of patients with COVID-19 have residual shortness of breath. It has also been reported that within 1–6 months of discharge from the hospital, 55.7% of patients have residual abnormal findings on CT scans and 44.3% have residual abnormal pulmonary function tests. Furthermore, 56% of patients treated with mechanical ventilation have been reported to have reduced pulmonary diffusion capacity 6 months after discharge. Serum KL-6 levels may also help predict and diagnose long COVID. An association between residual CT shadows and KL-6 levels at 3 months has been reported. Future studies should examine the relationship among serum KL-6 levels, changes on imaging and long COVID on admission, during hospitalisation and after discharge.

There are several limitations to the present study. First, image scoring was not available. Previous studies have reported a correlation between CT quantitative pneumonia range and serum KL-6 levels. In this large, multicentre study, CT could only be evaluated qualitatively; future comparisons with quantitative evaluations are desirable. Second, serum KL-6 levels were not measured over time. Serum KL-6 levels continue to rise during hospitalisation and peak later in severe cases than in mild cases; moreover, peak serum KL-6 levels have been reported to be more useful in predicting COVID-19 severity and death than serum KL-6 levels on admission. Although measuring peak levels may improve sensitivity and specificity, the fact that serum KL-6 levels on admission were predictive of outcome is both convenient and clinically useful. Third, clinical outcomes are significantly higher in patients whose serum KL-6 levels were measured on admission than in those who did not. It is speculated that this result might be because KL-6 was aggressively measured in patients with more severe COVID-19 on admission to assess the extent of lung disease. This selection bias may affect the prediction values of serum KL-6 values for critical outcomes. Fourth, in this study, we were unable to show the prevalence and severity of ARDS with serum KL-6 levels because we could not assess PaO₂ (partial pressure of arterial oxygen) levels via arterial blood gas analysis on admission. Severe COVID-19 is associated with ARDS, but the effect of ARDS on serum KL-6 levels of patients with COVID-19 cannot be determined from this study. However, we revealed that serum KL-6 levels were higher in patients with more severe disease on admission or those with rapid deterioration. These findings support that serum KL-6 is a predictive marker in patients with ARDS-like severe COVID-19 or rapidly deteriorating COVID-19 on early admission.

In conclusion, the levels of serum KL-6 within 48 hours of admission were associated with critical outcomes in a large Japanese multicentre study involving patients with COVID-19. Further, the MUC1 variant was associated with serum KL-6 levels within 48 hours of admission but not with critical outcomes. Thus, serum KL-6 levels on admission may be a useful biomarker of critical COVID-19 outcomes.

Author affiliations
1Division of Pulmonary Medicine, Department of Medicine, Keio University School of Medicine Graduate School of Medicine, Tokyo, Japan
2Department of Clinical Medicine (Laboratory of Bioregulatory Medicine), Kitasato University School of Pharmacy, Tokyo, Japan
3Department of Respiratory Medicine, Kitasato University, Kitasato Institute Hospital, Tokyo, Japan
4Department of Infectious Diseases, Keio University School of Medicine, Tokyo, Japan
5Department of Preventive Medicine and Public Health, Keio University School of Medicine, Tokyo, Japan
6Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan
7Department of Respiratory Medicine and Clinical Immunology, Osaka University Graduate School of Medicine, Suita, Japan
8Department of Respiratory Medicine, Nagoya University Graduate School of Medicine Faculty of Medicine, Nagoya, Japan
9Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita, Japan
10Center for Infectious Disease Education and Research (CiDER), Osaka University, Suita, Japan
11Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
12Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan
13Department of Genome Informatics, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan
14Medical Innovation Promotion Center, Tokyo Medical and Dental University, Tokyo, Japan
15Department of Surgery, Keio University School of Medicine, Tokyo, Japan
16Institute of Research, Tokyo Medical and Dental University, Tokyo, Japan
17Division of Health Medical Intelligence, Human Genome Center, the Institute of Medical Science, The University of Tokyo, Tokyo, Japan
18M&D Data Science Center, Tokyo Medical and Dental University, Tokyo, Japan
19Department of Pathology and Tumor Biology, Kyoto University Graduate School of Medicine Faculty of Medicine, Kyoto, Japan
20Division of Gastroenterology and Hepatology, Department of Internal Medicine, Keio University School of Medicine Graduate School of Medicine, Tokyo, Japan

Acknowledgements We would like to thank all participants involved in this study and all members of the Japan COVID-19 Task Force who regularly engaged in clinical and research work on COVID-19.

Collaborators The list of all members who contributed to this study is shown in the online supplemental information.
Contributors SA, SC, TA, HH, NH, KM, HK, MI, and KF conceptualised the present study. SA, SC, TA, HH, NH, KM, HK, MI, and KF drafted the paper. HH, NS, RE, RL, HT, STI, KN, TF, MW, KS, TK, KM, HK, MI, NH, YO, RK, YK, AK, SI, SM, SOg, TKa, and KF reviewed and edited the paper. All authors approved the final draft of the manuscript for publication. SC accepts full responsibility for the work and the conduct of the study, accesses to the data, and controls the decision to publish.

Funding This study was supported by the AMED (JP20kk0101612; JP20fk0108415, JP19k02100034, JP21km0405211), JP21km0405217, and JP21wm0325031 and JP21fk0108573), JST CREST (JPMJCR20H2), JST PRESTO (JPMJPR21R7) and MHLW (20CA2054).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval This study involves human participants and was approved by the Ethics Committee of Keio University School of Medicine and related research institutions (#20200061). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been properly peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/. ORCID iD Katsunori Masaki http://orcid.org/0000-0003-0909-9409

REFERENCES
33 Docherty AB, Harrison EM, Green CA, et al. Features of 20133 UK patients in hospital with COVID-19 using the ISARIC WHO clinical