Japan COVID-19 Task Force: a nation-wide consortium to elucidate host genetics of COVID-19 pandemic in Japan

Ho Namkoong and Ryuya Edahiro et al.

(Full lists of the authors and affiliations are at the end of the manuscript file)
Abstract
To elucidate the host genetic loci affecting severity of SARS-CoV-2 infection, or Coronavirus disease 2019 (COVID-19), is an emerging issue in the face of the current devastating pandemic. Here, we report a genome-wide association study (GWAS) of COVID-19 in a Japanese population led by the Japan COVID-19 Task Force, as one of the initial discovery GWAS studies performed on a non-European population. Enrolling a total of 2,393 cases and 3,289 controls, we not only replicated previously reported COVID-19 risk variants (e.g., LZTFL1, FOXP4, ABO, and IFNAR2), but also found a variant on 5q35 (rs60200309-A at DOCK2) that was associated with severe COVID-19 in younger (<65 years of age) patients with a genome-wide significant p-value of $1.2 \times 10^{-8}$ (odds ratio = 2.01, 95% confidence interval = 1.58-2.55). This risk allele was prevalent in East Asians, including Japanese (minor allele frequency [MAF] = 0.097), but rarely found in Europeans. Cross-population Mendelian randomization analysis made a causal inference of a number of complex human traits on COVID-19. In particular, obesity had a significant impact on severe COVID-19. The presence of the population-specific risk allele underscores the need of non-European studies of COVID-19 host genetics.
Introduction

Coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been a serious global public health issue. Although promising vaccines have recently become available, the emergence of SARS-CoV-2 variants may delay the end of this pandemic. One of the clinical characteristics of COVID-19 is its diverse clinical presentation, ranging from asymptomatic infection to fatal respiratory/multi-organ failure. For example, elderly people are known to be at high risk for severe diseases, particularly those who have multiple complications, such as obesity, hypertension, diabetes, and chronic renal failure. The risk for severe disease and mortality may differ depending on populations; Asians, for instance, may have less severe clinical presentation than Europeans. However, practical explanations on this paradox have been elusive.

Human genetic backgrounds influence susceptibility to and/or severity of infectious diseases. Genome-wide association studies (GWAS) have identified a number of genetic loci associated with infectious diseases, including class I human leukocyte antigen (HLA) and CCR5 implicated in HIV infection, 18q11 in tuberculosis, HBB in malaria infection, and 16p21 in Mycobacterium avium complex (MAC). Thus, human host genetics is expected to define the susceptibility to COVID-19 and explain the populational differences in disease mortality and morbidity. The Severe Covid-19 GWAS Group reported that a variant at LZTFL1 on 3p21 confers a risk for severe COVID-19 in Europeans with an odds ratio (OR) of as high as 2.0. The risk variant at 3p21, known to be inherited from the archaic hominin-derived genome sequences of Neanderthals, is also associated with various clinical manifestations of COVID-19 such as severe respiratory failure and venous thromboembolism. Of interests, this variant demonstrated globally heterogeneous allele frequency spectra: highest in South Asians (>40%), moderate in Europeans (~10%), but rarely present among East Asians. Despite its dominant impact in Europeans, this region does not explain COVID-19 severities among East Asians.

Further GWAS and whole genome sequencing (WGS) have identified host susceptibility genes. In addition to the ABO locus, OAS1/3/2, TYK2, DPP9, and IFNAR2 have been identified. WGS revealed rare loss-of-function (LoF) mutations in TLR7 and type I IFN-related genes in severe COVID-19 patients. In particular, the international consortium COVID-19 Human Genome Initiatives (HGI) has been conducting a large-scale meta-GWAS. However, the vast majority of the existing genomic studies were performed on European populations. Considering the global diversity of COVID-19 severity, COVID-19 host genetics analysis in non-European populations should provide novel insights.
The Japan COVID-19 Task Force was established in early 2020 as a nation-wide multicenter consortium to overcome the COVID-19 pandemic in Japan (https://www.covid19-taskforce.jp/en/home; Figure 1). This collaborative research networks are composed of >100 hospitals throughout Japan led by core academic institutes (Supplementary Table 1). Since the beginning of the pandemic, the Japan COVID-19 Task Force has longitudinally collected DNA, RNA, and plasma from >3,400 COVID-19 cases along with detailed clinical information (as of April 2021). The collected bioresources have been invested for multi-omics analysis towards open science for the community. In this study, we report the result of an initial large-scale discovery GWAS of COVID-19 in a Japanese population with systemic comparisons to that in Europeans to elucidate the global host genetics landscape of COVID-19.
Figure 1. Japan COVID-19 Task Force.

Japan COVID-19 Task Force is a nation-wide consortium to overcome COVID-19 pandemic in Japan, which was established in early 2020. Japan COVID-19 Task Force consists of >100 hospitals (red dots) led by core academic institutes (blue labels), and collected DNA, RNA, and plasma from >3,400 COVID-19 cases along with detailed clinical information.
Results
Overview of the study participants
In this study, we enrolled unrelated 2,393 patients with COVID-19 who required hospitalization from April 2020 to January 2021, from >100 hospitals participating in the Japan COVID-19 Task Force. COVID-19 diagnoses of all cases were confirmed by physicians of each affiliated hospital based on clinical manifestations and a positive PCR test result. As for the control, we enrolled unrelated 3,289 subjects ahead of the COVID-19 pandemic who represent a general Japanese population. All the participants were confirmed to be of Japanese origin on the basis of a principal component analysis (Supplementary Figure 1). Detailed characteristics of the participants are described in Supplementary Table 2.

Of the 2,393 COVID-19 cases, 990 ultimately had severe infection as defined by oxygen support, artificial respiration, and/or intensive-care unit hospitalization), while 1,391 cases had non-severe diseases. Severity information was not available for the remaining 12. As reported previously\(^3,18\), the severe COVID-19 cases were relatively more aged (65.3 ± 13.9 years [mean ± SD]) and a higher proportion of males (73.9%), compared with non-severe cases (49.3 ± 19.2 years and 57.2 of males, respectively).

GWAS of COVID-19 in Japanese identified a population-specific risk variant at DOCK2
We conducted a GWAS of COVID-19 in a Japanese population. After applying stringent quality control (QC) filters and genome-wide genotype imputation using a population-specific reference panel of Japanese\(^{19–21}\), we obtained 13,485,123 variants with minor allele frequency (MAF) ≥ 0.001 and imputation score (Rsq) ≥ 0.5 (13,116,003, 368,566 and 554 variants for autosomal, X-chromosomal, and mitochondrial variants, respectively). As illustrated in the follow-up analysis stratified by deep clinical information at the LZTFL1 locus, several COVID-19 risk variants are expected to confer relatively larger effects in severe and younger cases than in mild (and self-reported) cases or elder cases\(^{14}\). This suggests that COVID-19 GWAS likely to have a higher statistical power when focusing on severe and younger cases\(^{12,16,22}\). We thus separately conducted stratified GWAS of severe COVID-19 cases (\(n\)Case = 990), younger cases (age < 65, \(n\)Case = 1,484), and their combinations (\(n\)Case = 440), as well as all the cases (\(n\)Case = 2,393), in comparisons with the controls. We selected the age of 65 as a threshold, since ages ≥65 years is defined as a aggravation risk factor in the clinical management guide of patients with COVID-19 in Japan\(^4\). We did not observe
inflation of GWAS test statistics ($\lambda_{GC} < 1.007$; Supplementary Figure 2), suggesting no evidence of population stratification, as well as potential biases, in our GWAS.

GWAS between all COVID-19 cases vs. controls yielded no positive signals satisfying a genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ (Supplementary Figure 2)$^{29}$. By contrast, when the comparison was made between younger cases with severe COVID-19 and respective controls, where the highest prior probability of discovery of a positive association was expected, we identified a genetic locus on 5q35 that satisfied genome-wide significance ($P = 1.2 \times 10^{-8}$ at rs60200309; Figure 2a). The A allele of the lead SNP (rs60200309) that was located at an intergenic region downstream of the DOCK2 gene was associated with an inflated risk for severe COVID-19 infection with an OR of 2.01 (95% confidence interval [95%CI] = 1.58-2.55, $P = 1.2 \times 10^{-8}$; Table 1 and Figure 2b). The risk rs60200309-A allele was also associated with a significantly increased risk of COVID-19 in other comparisons including all COVID-19 cases and controls regardless of severity (OR = 1.24, 95%CI = 1.09-1.41, $P = 0.0011$; Supplementary Table 3), and within-case severity analysis (i.e., severe vs non-severe cases; OR = 1.27, 95%CI = 1.03-1.57, $P = 0.028$ for all ages; OR = 1.90, 95%CI = 1.43-2.52, $P = 1.1 \times 10^{-5}$ for ages <65). To evaluate the effect of the time of enrollment, we conducted a stratified analysis according to the time of enrollment for younger cases with severe COVID-19. The rs60200309-A allele consistently showed an inflated risk throughout the recruitment period (OR = 2.38, 1.79, 1.92 in the stratified analysis according to the trisected recruitment periods of April 2020 – July 2020, August 2020 – October 2020, and November 2020 – January 2021, respectively; OR = 2.00 and $P = 2.0 \times 10^{-8}$ in the all-period meta-analysis; Supplementary Table 3). This allele, however, did not seem to confer any significant COVID-19 risk in elder cases ($P > 0.069$). These results suggest a susceptibility of patients with the rs60200309-A allele to severe COVID-19 in the Japanese population, particularly in younger cases with severe COVID-19.

We then looked up COVID-19 risk of the DOCK2 variant in different ancestries (3,138 hospitalized COVID-19 cases vs 891,375 controls from the pan-ancestry meta-analysis available at https://rgc-covid19.regeneron.com/)$^{24,25}$. We observed the same directional effect with a marginal association signal (OR = 1.73, 95%CI = 0.95-3.15, $P = 0.072$, MAFCase = 0.0025, MAFControl = 0.0008; Supplementary Table 4). Meta-analysis of the Japanese discovery GWAS and the replication study from the pan-ancestry study yielded a genome-wide significant association showing an OR = 1.97 (95%CI = 1.57-2.46, $P = 1.2 \times 10^{-9}$; Supplementary Table 5). We note that rs60200309 does not exist in the public summary statistics provided by the COVID-19 Host Genetics Initiative (release 5)$^{16}$. Nevertheless,
given the low allele frequencies of the relevant \textit{DOCK2} allele in non-East Asian populations, further population-specific and cross-population replication studies are warranted.

\textbf{Population-specific features of the \textit{DOCK2} risk allele associated with COVID-19}

Of interest, the risk allele at \textit{DOCK2} (rs60200309-A) identified in this study was common in East Asians (= 0.097) with the highest frequency in Japanese (= 0.125), less frequent in native Americans (= 0.049), but very rare in Europeans, African, and south Asians (< 0.005; from 1000 Genomes Project Phase3v5 database; Figure 2c). When we referred to the results of WGS-based natural selection screening in Japanese\textsuperscript{19}, the rs60200309-A allele marginally positively selected in Japanese ($P_{\text{SDS}} = 0.051$). It is suggested that the rs60200309-A rapidly increased its frequency among the Japanese population during the past several thousand years. These population-specific features of the \textit{DOCK2} variant partly explain the reason why it was not identified in the previous European COVID-19 GWAS studies despite their larger sample sizes, and should provide a rationale for further accelerating COVID-19 host genetics researches on non-European populations.
Figure 2. Severe and younger COVID-19 GWAS in the Japanese population

(a) A Manhattan plot of the severe and younger COVID-19 GWAS (440 cases and 2,377 controls). A dotted line represents the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$. Manhattan and quantile-quantile plots of all GWAS results are in Supplementary Figure 2.

(b) A regional association plot at the DOCK2 locus. Dots represent SNPs with colors according to linkage disequilibrium ($r^2$) with the lead SNP of rs60200309.

(c) Allele frequency spectra of the rs60200309-A allele in the 1000 Genomes Project Phase3v5 database.
Table 1. Association of the DOCK2 variant with COVID-19 risk in the Japanese population

<table>
<thead>
<tr>
<th>rsID</th>
<th>Chr:position (b37)</th>
<th>Allele</th>
<th>Gene</th>
<th>Age</th>
<th>Phenotype</th>
<th>No. subjects</th>
<th>Risk allele freq. (A)</th>
<th>OR (95%CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs60200309</td>
<td>5:169519612</td>
<td>G/A</td>
<td>DOCK2</td>
<td>All age</td>
<td>COVID-19 vs control</td>
<td>2,393</td>
<td>0.119</td>
<td>0.101</td>
<td>1.24 (1.09-1.41)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Severe COVID-19 vs control</td>
<td>990</td>
<td>0.129</td>
<td>0.101</td>
<td>1.39 (1.16-1.66)</td>
</tr>
<tr>
<td>rs60200309</td>
<td>5:169519612</td>
<td>G/A</td>
<td>DOCK2</td>
<td>Age &lt; 65</td>
<td>COVID-19 vs control</td>
<td>1,484</td>
<td>0.123</td>
<td>0.099</td>
<td>1.32 (1.13-1.55)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Severe COVID-19 vs control</td>
<td>440</td>
<td>0.159</td>
<td>0.099</td>
<td>2.01 (1.58-2.55)</td>
</tr>
</tbody>
</table>
Cross-population comparisons of the COVID-19-associated variants

We then conducted cross-population comparisons of allele frequency spectra and genetic risk of the previously reported COVID-19-associated variants\textsuperscript{12,16,22}. Of the 11 associated variants evaluated in our Japanese GWAS, we replicated the associations with 7 variants ($P < 0.05$ in any 4 phenotypes of the case-control GWAS; \textit{LZTFL1}, \textit{FOXP4}, \textit{TMEM65}, \textit{ABO}, \textit{TAC4}, \textit{DPP9}, and \textit{IFNAR2}; Figure 3 and Supplementary Table 6). For all nominally associated signals ($P < 0.05$), we observed same directional effects of the alleles as in Europeans. ORs for severe and younger COVID-19 cases were highest among the phenotype patterns in six of the 7 loci, confirming our strategy that focusing on such cases should efficiently highlight the host genetic risk of COVID-19.

The most significant replication was observed at the \textit{FOXP4} locus, where the risk allele was known to be more prevalent in East Asians than in Europeans, and expected to have a higher power to be detected in East Asians\textsuperscript{16} (OR = 1.29, 95%CI = 1.13-1.46, $P = 9.1 \times 10^{-5}$ for the severe COVID-19 cases). Regarding the strongest risk variant in Europeans at \textit{LZTFL1} (rs35081325), we replicated the associations despite its rare frequency in Japanese (= 0.0013 in controls) with the highest risk in the severe and younger COVID-19 cases (OR = 11.8, 95%CI = 1.64-85.5, $P = 0.014$). These observations propose shared host genetics backgrounds of COVID-19 across populations.
Odds ratios of the COVID-19-associated variants in the Japanese population are indicated.

Figure 3. Forest plots of the risk of COVID-19-associated variants in Japanese
Odds ratios of the COVID-19-associated variants in the Japanese population are indicated.
Relatively less dominant impact of the HLA gene variants on COVID-19 risk

Given their critical impact on immune responses and contribution to host genetics of various infectious diseases, HLA gene variants have been investigated for their possible role in the response to COVID-19 infection with controversial discussions. To address this issue, we applied in silico imputation of both classical and non-classical HLA variants using the HLA reference panel of Japanese ($n = 1,118$). After imputing the HLA variants, we did not observe association signals satisfying neither of the genome-wide significance ($P < 5.0 \times 10^{-8}$) or HLA-wide significance thresholds ($P < 0.05/2,482 \text{ variants } = 2.0 \times 10^{-5}$; Supplementary Figure 3 and Supplementary Table 7). Most significant HLA variant associations in each of the case-control phenotypes were followings: all cases (HLA-DQa1 Glu175, OR = 0.84, 95%CI = 0.77-0.92, $P = 2.9 \times 10^{-4}$), severe cases (HLA-DRB1 amino acid position 38, $P = 6.7 \times 10^{-4}$), younger cases (HLA-DRB1 amino acid positions 96 and 104, $P = 8.3 \times 10^{-5}$), and severe and younger cases (HLA-C*03:04, OR = 1.48, 95%CI = 1.16-1.90, $P = 0.0020$). While further accumulation of sample sizes is warranted, the current results may not support a major impact of the HLA variants on host genetics in Japanese despite previous associations with other infectious diseases.

Link between ABO blood type and COVID-19 risk

ABO blood types are defined by the variants on the coding region of the ABO gene on 9q34, which are pleiotropic on various complex human traits including infectious diseases (e.g., malaria resistance of blood type O). Motivated by replicated COVID-19 risk of the ABO locus in Japanese, we conducted ABO blood type-based risk analysis.

Among the four major ABO blood types (A, B, AB, and O with 39.0%, 21.8%, 9.5%, and 29.7% in our Japanese GWAS, respectively), the O blood type was consistently associated with a protective effect on COVID-19 in case-control phenotypes ($P < 0.05$), most evidently in severe and younger cases (OR = 0.73, 95%CI = 0.56-0.93, $P = 0.014$; Figure 4 and Supplementary Table 8), as reported previously. We found increased risk of the AB blood type, especially in severe cases (OR = 1.41, 95%CI = 1.10-1.81, $P = 0.0065$ for all ages, and OR = 1.40, 95%CI = 1.00-1.94, $P = 0.048$ for age < 65, in comparison with the other blood types). Increased severity risk of the AB blood type was also significant when compared with the A or B blood types (OR > 1.34, $P < 0.041$ for all ages). To our knowledge, this is the initial study to report severe COVID-19 risk of the AB blood type. The ABO blood type distributions are heterogeneous among worldwide populations, and Japanese is the one with the highest
AB blood type frequency\textsuperscript{35}, which might have provided statistical power to detect its risk on severe COVID-19 in our study.
Figure 4. Forest plots of the ABO blood type associations with COVID-19 in Japanese
Odds ratios of the ABO blood types in the Japanese population are indicated.
Causal inference on COVID-19 by cross-population Mendelian randomization

COVID-19 pandemic has exposed global populations to an emergent health risk. To predict individuals’ risk on COVID-19-related outcomes, elucidation of the medical conditions that can affect COVID-19 susceptibility is warranted. While medical record-based epidemiological studies assessing comorbidity have identified multiple risk factors, there remains various clinical status where causal inference on COVID-19 is controversial.

To make a causal inference, we applied cross-population two-sample Mendelian randomization (MR) analysis. Two-sample MR utilizes GWAS summary statistics to infer causality between correlated phenotypes. Further, cross-population MR analysis could provide robust evidence of causality. As exposure phenotypes, we selected a series of clinical states and diseases where increased or decreased comorbidity with COVID-19 have been discussed: metabolic traits (obesity [body mass index; BMI] and type 2 diabetes [T2D]), chronic respiratory diseases and related phenotypes (cigarettes per day [CPD] and asthma), blood pressure, renal function (estimated glomerular filtration rate [eGFR]), serum uric acids (UA) and gout, and rheumatic diseases (rheumatoid arthritis [RA] and systemic lupus erythematosus [SLE]). Lists of the GWAS studies of the exposure phenotypes are in Supplementary Table 9.

In the Japanese population, MR results were contrastive between the severe COVID-19 cases and all COVID-19 cases (Figure 5 and Supplementary Table 10). As for the severe COVID-19 cases, a causal effect was demonstrated only for obesity ($P = 0.0067$ and $0.0074$ for all age and age < 65, respectively). By contrast, we observed causal effects of asthma ($P = 0.0061$ and $0.018$ for all age and age < 65, respectively), UA ($P = 0.019$ for age < 65), and gout ($P = 0.0048$ and $0.0027$ for all age and age < 65, respectively), while SLE ($P = 0.0014$ for all age) showed a protective effect.

We then looked up the MR results in Europeans by using publicly released GWAS summary statistics of COVID-19 Host Genetics Initiative (release 5). We observed significant causal inferences of obesity consistent with those in Japanese. The causal effect of obesity was observed for self-reported, hospitalized, and severe COVID-19 in Europeans ($P = 8.5 \times 10^{-9}$, $3.2 \times 10^{-11}$, and $6.2 \times 10^{-6}$, respectively) as previously reported, while effect sizes were twice as high in hospitalized and severe COVID-19 ($\beta > 0.398$) when compared with self-reported COVID-19 ($\beta = 0.175$). Obesity is one of the major risk factors for COVID-19 severity and critical outcomes, and our cross-population MR analysis provided evidence of causality on this link. Causal inference of decreased renal function ($P = 0.043$ for severe COVID-19) and
T2D \( (P = 0.019 \text{ for self-reported COVID-19 and } P = 0.0078 \text{ for hospitalized COVID-19}) \) was also observed in Europeans.

Our cross-population MR analysis provided several phenotypes with significant MR results observed only in Japanese (i.e., risk of asthma, UA, and gout, and protective role of SLE). This suggests existence of populational heterogeneity in the impacts of causal links from the baseline clinical manifestations to COVID-19 susceptibility. Hyperuricemia is reported as one of the major risk factors of severe COVID-19 in Japan\(^{18}\), which is consistent with a Japanese-specific causal inference of UA and gout in the Japanese MR analysis. There exist controversial discussions on the risk of SLE patients on COVID-19 infection\(^{38,39}\). Our results suggest a possibility that genetically-determined susceptibility to SLE, and its underlying immunophenotypes, could make patients protective against COVID-19 infection.
Figure 5. Cross-population Mendelian randomization analysis of the COVID-19 GWAS

Forest plots of the Mendelian randomization (MR) analysis results of causal inference on the COVID-19 GWAS in Japanese (left panel) and Europeans (right panel). Since effect sizes (= beta) of MR are not scalable among the phenotypes and populations, normalized beta is indicated. For each phenotype and population, the standard error for the COVID-19 GWAS with the largest sample size (i.e., “COVID-19 vs control” for Japanese and “Self-reported COVID-19 vs control (C2)” for Europeans) was set to be 0.1. Abbreviations of the exposure phenotypes and the detailed MR results are in Supplementary Table 10.
Discussion

In this study, we reported GWAS of COVID-19 in a Japanese population led by the Japan COVID-19 Task Force, a nation-wide consortium to battle against the COVID-19 pandemic. This is one of the initial and largest COVID-19 host genetics studies in non-European populations to date. Our study highlighted multiple genetic variants associated with the COVID-19 risk shared across populations such as LZTFL1, ABO, and FOXP4, as well as the identification of a population-specific risk variant at the DOCK2 locus. Stratified analysis of these susceptibility loci supported the expectation that host genetics of COVID-19 should be enhanced when focusing the analysis on younger cases with severe COVID-19. Rather unexpectedly, contribution of HLA variants to COVID-19 host susceptibility, if ever present, was not remarkable, compared with previous findings on other infectious diseases. As for the ABO blood type classification, we newly identified the risk of the AB blood type to severe COVID-19. Finally, cross-population MR analysis disclosed a causal inference of a number of complex human traits on COVID-19, such as an elevated risk of obesity on severe COVID-19. Our results highlight population-specific risk allele host genetic backgrounds, which underscores the need of non-European studies for COVID-19 host genetics.

DOCK2 (dedicator of cytokinesis 2) is a Rac activator involved in chemokine signaling, type I interferon production, and lymphocyte migration\cite{41,42}, where the pathophysiology of COVID-19 have been implicated\cite{43,44}. Multiple COVID-19 risk alleles elucidated by host genetics studies are in this axis (e.g., LZTFL1-SLC6A20-CCR9, TYK2, IFNAR2, and IRF7)\cite{12,16,22}. Of note, autosomal recessive DOCK2 deficiency is a Mendelian disorder with combined immunodeficiency and severe invasive infection (OMIM #616433)\cite{45}. DOCK2 has recently been reported to be suppressed in bronchoalveolar lavage fluid (BALF) cells of COVID-19 patients\cite{46}. Given that LoF caused by an inborn error is a key to fine-map host susceptible genes for infectious diseases\cite{26}, DOCK2 could be considered as one of the key genes to determine the risk for, as well as potential targets, of COVID-19 therapy and drug discovery. Further functional studies linking the DOCK2 variant to molecular and clinical phenotypes should be required to elucidate the mechanism by which this DOCK2 allele confer the risk of severe COVID-19.

While our nation-wide longitudinal efforts have enabled the current COVID-19 GWAS in Japanese, a number of replication studies involving independent populations are required to confirm the current result. In the near future, a growing number of GWAS studies should be conducted regarding COVID-19 host genetics, which through public sharing contribute to guide a global health strategy against the COVID-19 pandemic.
Methods

Study participants
All the cases affected with COVID-19 were recruited through Japan COVID-19 Task Force. We enrolled the hospitalized cases diagnosed as COVID-19 by physicians using the clinical manifestation and PCR test results, who were recruited from April 2020 to January 2021 at any of the >100 the affiliated hospitals (Supplementary Table 1 and 2). Control subjects were collected as general Japanese populations at Osaka University Graduate School of Medicine and affiliated institutes. Individuals determined to be of non-Japanese origin either of self-reporting or by principal component analysis were excluded as described elsewhere (Supplementary Figure 1). All the participants provided written informed consent as approved by the ethical committees of Keio University School of Medicine, Osaka University Graduate School of Medicine, and affiliated institutes.

GWAS genotyping and quality control
We performed GWAS genotyping of the 2,520 COVID-19 cases and 3,341 controls using Infinium Asian Screening Array (Illumina, CA, USA). We applied stringent QC filters to the samples (sample call rate < 0.97, excess heterozygosity of genotypes > mean + 3SD, related samples with PI_HAT > 0.175, or outlier samples from East Asian clusters in principal component analysis with 1000 Genomes Project samples), and variants (variant call rate < 0.99, significant call rate differences between cases and controls with $P < 5.0 \times 10^{-8}$, deviation from Hardy-Weinberg equilibrium with $P < 1.0 \times 10^{-6}$, or minor allele count < 5), as described elsewhere. Details of the QC for the mitochondrial variants are described elsewhere. After QC, we obtained genotype data of 489,539, 15,161, and 217 autosomal, X-chromosomal, and mitochondrial variants, respectively, for 2,393 COVID-19 cases and 3,289 controls.

Genome-wide and HLA genotype imputation
We used SHAPEIT4 software (version 4.1.2) for haplotype phasing of autosomal genotype data, and SHAPEIT2 software (v2.r904) for X-chromosomal genotype data. After phasing, we used Minimac4 software (version 1.0.1) for genome-wide genotype imputation. We used the population-specific imputation reference panel of Japanese ($n = 1,037$) combined with 1000 Genomes Project Phase3v5 samples ($n = 2,504$). Imputations of the mitochondrial variants were conducted as described elsewhere, using the population-specific reference panel ($n = 1,037$). We applied post-imputation QC filters of MAF ≥ 0.1% and imputation score (Rsq) > 0.5. We note that the genotypes of the lead variant in the GWAS (rs60200309) were
obtained by imputation (Rsq = 0.88). We assessed accuracy by comparing the imputed dosages with WGS data for the part of the controls (n = 236), and confirmed high concordance rate of 97.5%.

HLA genotype imputation was performed using DEEP*HLA software (version 1.0), a multitask convolutional deep learning method31. We used the population-specific imputation reference panel of Japanese (n = 1,118), which included both classical and non-classical HLA gene variants for imputation30. Before imputation, we removed the overlapping samples between the GWAS controls and the reference panel (n = 649), from the GWAS data side. We imputed HLA alleles (2-digit and 4-digit) and the corresponding HLA amino acid polymorphisms, and applied post-imputation QC filters of MAF ≥ 0.5% and imputation score (r² in cross-validation) > 0.7.

Case-control association test
We conducted GWAS of COVID-19 by using logistic regression of the imputed dosages of each of the variants on case-control status, using PLINK2 software (v2.00a3LM AVX2 Intel [6 Jul 2020]). We included sex, age, and the top five principal components as covariates in the regression model. We set the genome-wide association significance threshold of P < 5.0 × 10⁻⁸.23 We obtained the association of the DOCK2 variant (rs60200309) from the pan-ancestry meta-analysis available at https://rgc-covid19.regeneron.com24,25. We obtained the meta-analysis results of the phenotype of "hospitalized COVID-19 vs COVID-19 negative or COVID-19 status unknown" with the largest case sample size. Meta-analysis of the Japanese discovery GWAS and the pan-ancestry analysis was conducted using an inverse-variance method assuming a fixed-effects model.

As for the imputed HLA variants, we conducted (i) association test of binary HLA markers (2-digit and 4-digit HLA alleles, respectively amino acid residues) and (ii) an omnibus test of each of the HLA amino acid positions, as described elsewhere30. Binary maker test was conducted using the same logistic regression model and covariates as in the GWAS. Omnibus test was conducted by a log likelihood ratio test between the null model and the fitted model, followed by a χ² distribution with m-1 degree(s) of freedom, where m is the number of the residues. R statistical software (version 3.6.0) was used for the HLA association test. In addition to the genome-wide significance threshold, we set the HLA-wide significance threshold based on Bonferroni’s correction for the number of the HLA tests (α = 0.05).
Estimation of the ABO blood types and analysis

We estimated the ABO blood types of the GWAS subjects based on the five coding variants at the ABO gene (rs8176747, rs8176746, rs8176743, rs7853989, and rs8176719)\(^{32,33}\). We phased the haplotypes of these five variants based on the best-guess genotypes obtained by genome-wide imputation, and estimated the ABO blood type as described elsewhere\(^{34}\). We could unambiguously determine the ABO blood type of 99.1% of the subjects.

Blood group-specific ORs were estimated based on comparisons of A vs AB/B/O, B vs A/AB/O, AB vs A/B/O, and O vs A/AB/B. We conducted a logistic regression analysis including age, sex and the top 5 principal components as covariates. R statistical software (version 3.6.3) was used for the ABO blood type analysis.

Cross-population MR analysis

We conducted two-sample MR analysis as described elsewhere\(^{36,37}\). As an outcome phenotype, we utilized the GWAS summary statistics of Japanese (current study) and Europeans (release 5 from COVID-19 Host Genetics Initiative\(^{16}\)). Lists of the Japanese and European GWAS studies used as the exposure phenotypes are in Supplementary Table 9.

We extracted the independent lead variants with genome-wide significance (or the proxy variants in linkage disequilibrium \(r^2 \geq 0.8\) in the EAS or EUR subjects of the 1000 Genomes Project Phase3v5 databases) from the GWAS results of the exposure phenotypes. We applied the inverse variance weighted (IVW) method using the TwoSampleMR package (version 0.5.5) in R statistical software (version 4.0.2).
References


Acknowledgements
We would like to sincerely thank all the participants involved in this study, and all the members of Japan COVID-19 Task Force for their supports. We thank Mr. Johji Kitano, e-Parcel Corporation, and Ascend Corporation for voluntarily supporting Japan COVID-19 Task Force.
We thank COVID-19 Host Genetics Initiative for publicly sharing the GWAS summary statistics of COVID-19. This study was supported by AMED (JP20nk0101612, JP20fk0108415, JP21jk0210034, JP21km0405211, and JP21km0405217), JST CREST (JPMJCR20H2), MHLW (20CA2054), Takeda Science Foundation, the Mitsubishi Foundation, and Bioinformatics Initiative of Osaka University Graduate School of Medicine, Osaka University. The super-computing resource was provided by Human Genome Center (the Univ. of Tokyo).

Author contributions

Competing interests
The authors declare no conflicts of interests.
Authors and Affiliations

Kamimaki\textsuperscript{139}, Yoshiteru Tominaga\textsuperscript{139}, Tomoo Ishii\textsuperscript{140}, Mitsuyoshi Utsugi\textsuperscript{141}, Akihiro Ono\textsuperscript{141}, Toru Tanaka\textsuperscript{142}, Takeru Kashiwada\textsuperscript{142}, Kazue Fujita\textsuperscript{142}, Yoshinobu Saito\textsuperscript{142}, Masahiro Seike\textsuperscript{142}, Yosuke Omae\textsuperscript{143}, Yasuhiro Nannya\textsuperscript{8}, Takafumi Ueno\textsuperscript{144}, Tomomi Takano\textsuperscript{145}, Kazuhiko Katayama\textsuperscript{146}, Masumi Ai\textsuperscript{147}, Atsushi Kumanogoh\textsuperscript{3,148,5,149}, Toshiro Sato\textsuperscript{150}, Naoki Hasegawa\textsuperscript{1}, Katsushi Tokunaga\textsuperscript{143}, Makoto Ishii\textsuperscript{4}, Ryuji Koike\textsuperscript{151}, Yuko Kitagawa\textsuperscript{152}, Akinori Kimura\textsuperscript{153}, Seiya Imoto\textsuperscript{9}, Satoru Miyano\textsuperscript{6,158}, Seishi Ogawa\textsuperscript{8,154,155,158}, Takanori Kanai\textsuperscript{20,158}, Yukinori Okada\textsuperscript{2,5,156,149,158}.

1. Department of Infectious Diseases, Keio University School of Medicine, Tokyo, Japan.
2. Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan.
3. Department of Respiratory Medicine and Clinical Immunology, Osaka University Graduate School of Medicine, Suita, Japan.
4. Division of Pulmonary Medicine, Department of Medicine, Keio University School of Medicine, Tokyo, Japan
5. Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita, Japan.
6. M\&D Data Science Center, Tokyo Medical and Dental University, Tokyo, Japan.
7. Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
8. Department of Pathology and Tumor Biology, Kyoto University, Kyoto, Japan.
9. Division of Health Medical Intelligence, Human Genome Center, the Institute of Medical Science, the University of Tokyo, Tokyo, Japan.
10. Department of Respiratory Medicine, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan.
11. Department of General Medicine, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan.
12. Department of Emergency and Disaster Medicine, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan.
13. Department of Cardiovascular Biology and Medicine, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan.
14. Department of Internal Medicine and Rheumatology, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan.
15. Department of Nephrology, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan.
16. Atopy (Allergy) Research Center, Juntendo University Graduate School of Medicine, Tokyo, Japan.
17. Department of Emergency and Critical Care Medicine, Keio University School of Medicine, Tokyo, Japan
18. Department of Anesthesiology, Keio University School of Medicine, Tokyo, Japan.
19. Department of Laboratory Medicine, Keio University School of Medicine, Tokyo, Japan
20. Division of Gastroenterology and Hepatology, Department of Medicine, Keio University School of Medicine, Tokyo, Japan.
21. Keio University Health Center, Tokyo, Japan.
22. Department of Respiratory Medicine, Saitama Cardiovascular and Respiratory Center, Kumagaya, Japan.
23. JCHO (Japan Community Health care Organization) Saitama Medical Center, Internal Medicine, Saitama, Japan.
24. Department of Respiratory Medicine, Tokyo Women's Medical University, Tokyo, Japan.
25. Department of General Medicine, Tokyo Women's Medical University, Tokyo, Japan.
26. Clinical Research Center, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan.
27. Department of Medical Informatics, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan.
28. Respiratory Medicine, Tokyo Medical and Dental University, Tokyo, Japan.
29. Clinical Laboratory, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan.
30. Kawasaki Municipal Ida Hospital, Department of Internal Medicine, Kawasaki, Japan.
31. Department of Respiratory Medicine, Osaka Saiseikai Nakatsu Hospital, Osaka, Japan.
32. Department of Infection Control, Osaka Saiseikai Nakatsu Hospital, Osaka, Japan.
33. Department of Infectious Diseases, Tosei General Hospital, Seto, Japan.
34. Department of Respiratory, Allergic Diseases Internal Medicine, Tosei General Hospital, Seto, Japan.
35. Department of Emergency and Critical Care Medicine, Kansai Medical University General Medical Center, Moriguchi, Japan.
36. Fukujuji hospital, Kiyose, Japan.
37. Department of Pulmonary Medicine, Saitama City Hospital, Saitama, Japan.
38. Department of Infectious Diseases, Saitama City Hospital, Saitama, Japan.
39. Department of General Thoracic Surgery, Saitama City Hospital, Saitama, Japan.
40. Department of Pulmonary Medicine, Eiju General Hospital, Tokyo, Japan.
41. Division of Infection Control, Eiju General Hospital, Tokyo, Japan.
42. Department of Hematology, Eiju General Hospital, Tokyo, Japan.
43. Saiseikai Utsunomiya Hospital, Utsunomiya, Japan.
44. Department of Respiratory Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan.
45. Department of Infectious Diseases, Tohoku University Graduate School of Medicine, Sendai, Japan.
46. Department of Respiratory Medicine, Kitasato University Kitasato Institute Hospital, Tokyo, Japan.
47. Department of Pulmonary Medicine, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan.
48. Department of Neurosurgery, Faculty of Medicine, the University of Tokyo, Tokyo, Japan.
49. Department of Integrative Physiology and Bio-Nano Medicine, National Defense Medical College, Tokorozawa, Japan.
50. Department of Otorhinolaryngology-Head and Neck Surgery, Osaka University Graduate School of Medicine, Suita, Japan.
51. Department of Head and Neck Surgery, Aichi Cancer Center Hospital, Nagoya, Japan.
52. Department of Neurosurgery, Osaka University Graduate School of Medicine, Suita, Japan.
53. Department of Otolaryngology and Head and Neck Surgery, Kansai Rosai Hospital, Hyogo, Japan.
54. Division of Infection Control and Prevention, Osaka University Hospital, Suita, Japan.
55. Department of Biomedical Ethics and Public Policy, Osaka University Graduate School of Medicine, Suita, Japan.
56. Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan.
57. Tachikawa Hospital, Tachikawa, Japan.
58. Department of Emergency and Critical Care Medicine, Tokyo Women's Medical University Medical Center East, Tokyo, Japan.
59. Department of Medicine, Tokyo Women's Medical University Medical Center East, Tokyo, Japan.
60. Department of Pediatrics, Tokyo Women's Medical University Medical Center East, Tokyo, Japan.
61. Internal Medicine, Sano Kosei General Hospital, Sano, Japan.
62. Japan Community Health care Organization Kanazawa Hospital, Kanazawa, Japan.
63. Department of Respiratory Medicine, Saiseikai Yokohamashi Nanbu Hospital, Yokohama, Japan.
64. Department of Clinical Laboratory, Saiseikai Yokohamashi Nanbu Hospital, Yokohama, Japan.
65. Internal Medicine, Internal Medicine Center, Showa University Koto Toyosu Hospital, Tokyo, Japan.
66. Department of Respiratory Medicine, Japan Organization of Occupational Health and Safety, Kanto Rosai Hospital, Kawasaki, Japan.
67. Department of General Internal Medicine, Japan Organization of Occupational Health and Safety, Kanto Rosai Hospital, Kawasaki, Japan.
68. Ishikawa Prefectural Central Hospital, Kanazawa, Japan.
69. Kanagawa Cardiovascular and Respiratory Center, Yokohama, Japan.
70. Department of Respiratory Medicine, National Hospital Organization Tokyo Medical Center, Tokyo, Japan.
71. Department of Allergy, National Hospital Organization Tokyo Medical Center, Tokyo, Japan.
72. Department of General Internal Medicine and Infectious Diseases, National Hospital Organization Tokyo Medical Center, Tokyo, Japan.
73. Department of Respiratory Medicine, Toyohashi Municipal Hospital, Toyohashi, Japan.
74. Keiyu Hospital, Yokohama, Japan.
75. KKR Sapporo Medical Center, Department of respiratory medicine, Sapporo, Japan.
76. Division of General Internal Medicine, Department of Internal Medicine, St. Marianna University School of Medicine, Kawasaki, Japan.
77. Department of Emergency and Critical Care Medicine, St. Marianna University School of Medicine, Kawasaki, Japan.
78. Japanese Red Cross Medical Center, Tokyo, Japan.
79. Matsumoto City Hospital, Matsumoto, Japan.
80. Department of Emergency and Critical Care Medicine, Faculty of Medicine, Fukuoka University, Fukuoka, Japan.
81. Department of Infection Control, Fukuoka University Hospital, Fukuoka, Japan.
82. Department of Rheumatology, National Hospital Organization Hokkaido Medical Center, Sapporo, Japan.
83. Department of Respiratory Medicine, National Hospital Organization Hokkaido Medical Center, Sapporo, Japan.
84. Department of Emergency and Critical Care Medicine, National Hospital Organization Hokkaido Medical Center, Sapporo, Japan.
85. NHO Kanazawa Medical Center, Kanazawa, Japan.
86. Nihon University School of Medicine, Department of Internal Medicine, Division of Respiratory Medicine, Tokyo, Japan.
87. Musashino Red Cross Hospital, Musashino, Japan.
88. Division of Respiratory Medicine, Social Welfare Organization Saiseikai Imperial Gift Foundation, Inc., Saiseikai Kumamoto Hospital, Kumamoto, Japan.
89. Department of Respiratory Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan.
90. Fukuoka Tokushukai Hospital, Department of Internal Medicine, Kasuga, Japan.
91. Fukuoka Tokushukai Hospital, Respiratory Medicine, Kasuga, Japan.
92. Department of Endocrinology, Hematology and Gerontology, Chiba University Graduate School of Medicine, Chiba, Japan.
93. Department of Emergency and Critical Care Medicine, Chiba University Graduate School of Medicine, Chiba, Japan.
94. National Hospital Organization Kumamoto Medical Center, Kumamoto, Japan.
95. Division of Infectious Diseases and Respiratory Medicine, Department of Internal Medicine, National Defense Medical College, Tokorozawa, Japan.
96. Sapporo City General Hospital, Sapporo, Japan.
97. Department of Pulmonary Medicine, Tokyo Saiseikai Central Hospital, Tokyo, Japan.
98. Department of Respiratory Medicine, Fujisawa City Hospital, Fujisawa, Japan.
99. Department of Medicine and Biosystemic Science, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan.
100. Ome Municipal General Hospital, Ome, Japan.
101. Research Institute for Diseases of the Chest, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan.
102. Department of Infectious Disease and Clinical Research Institute, National Hospital Organization Kyushu Medical Center, Fukuoka Japan.
103. Division of Respirology, Rheumatology, and Neurology, Department of Internal Medicine, Kurume University School of Medicine, Kurume, Japan.
104. Department of Infectious Disease, National Hospital Organization Kyushu Medical Center, Fukuoka Japan.
105. Daini Osaka Police Hospital, Osaka, Japan.
106. Department of Emergency and Critical Care Medicine, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan.
107. Department of Hematology, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan.
108. Department of Nephrology, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan.
112. Department of Cardiovascular Surgery, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan.
113. Division of Pulmonary Medicine, Department of Medicine, Tokai University School of Medicine, Isehara, Japan.
114. Department of Anesthesiology and Intensive Care Medicine, Kyoto Prefectural University of Medicine, Kyoto, Japan.
115. Department of Infection Control and Laboratory Medicine, Kyoto Prefectural University of Medicine, Kyoto, Japan.
116. Department of Respiratory Internal Medicine, St Marianna University School of Medicine, Yokohama-City Seibu Hospital, Yokohama, Japan.
117. KINSHUKAI Hanwa The Second Hospital, Osaka, Japan.
118. Gifu University School of Medicine Graduate School of Medicine, Emergency and Disaster Medicine, Gifu, Japan.
119. Department of Respiratory Medicine, Tokyo Medical University Hospital, Tokyo, Japan.
120. JA Toride medical hospital, Toride, Japan.
121. Okayama Rosai Hospital, Okayama, Japan.
122. Himeji St. Mary's Hospital, Himeji, Japan.
123. Emergency & Critical Care, Niigata University, Niigata, Japan.
124. Emergency & Critical Care Center, National Hospital Organization Kyoto Medical Center, Kyoto, Japan.
125. National Hospital Organization Tokyo Hospital Hospital, Kiyose, Japan.
126. Fujioka General Hospital, Fujioka, Japan.
127. Department of General Medicine, School of Medicine, International University of Health and Welfare Shioya Hospital, Ohtawara, Japan.
128. Department of Pharmacology, School of Pharmacy, International University of Health and Welfare Shioya Hospital, Ohtawara, Japan.
129. Department of Respiratory Medicine, International University of Health and Welfare Shioya Hospital, Ohtawara, Japan.
130. Department of Clinical Laboratory, International University of Health and Welfare Shioya Hospital, Ohtawara, Japan.
131. Department of Cardiology, Pulmonology, and Nephrology, Yamagata University Faculty of Medicine, Yamagata, Japan.
132. Division of Respiratory Medicine and Allergology, Department of Medicine, School of Medicine, Showa University, Tokyo, Japan.
133. Department of Pulmonary Medicine, Fukushima Medical University, Fukushima, Japan.
134. Kansai Electric Power Hospital, Osaka, Japan.
135. Division of Infectious Diseases, Kumamoto City Hospital, Kumamoto, Japan.
136. Department of Respiratory Medicine, Kumamoto City Hospital, Kumamoto, Japan.
137. Department of Emergency and Critical Care Medicine, Tokyo Metropolitan Police Hospital, Tokyo, Japan.
138. Department of Respiratory Medicine, Gunma University Graduate School of Medicine, Maebashi, Japan.
139. National hospital organization Saitama Hospital, Wako, Japan.
140. Tokyo Medical University Ibaraki Medical Center, Inashiki, Japan.
141. Department of Internal Medicine, Kiryu Kosei General Hospital, Kiryu, Japan.
142. Department of Pulmonary Medicine and Oncology, Graduate School of Medicine, Nippon Medical School, Tokyo, Japan.
143. Genome Medical Science Project (Toyama), National Center for Global Health and Medicine, Tokyo, Japan.
144. Department of Biomolecular Engineering, Graduate School of Tokyo Institute of Technology, Tokyo, Japan.
145. School of Veterinary Medicine, Kitasato University, Towada, Japan.
146. Laboratory of Viral Infection I, Department of Infection Control and Immunology, Ōmura Satoshi Memorial Institute & Graduate School of Infection Control Sciences, Kitasato University, Tokyo, Japan.
147. Department of Insured Medical Care Management, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan.
148. Department of Immunopathology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan.
149. Center for Infectious Disease Education and Research (CiDER), Osaka University, Suita, Japan.
150. Department of Organoid Medicine, Keio University School of Medicine, Tokyo, Japan.
151. Medical Innovation Promotion Center, Tokyo Medical and Dental University, Tokyo, Japan.
152. Department of Surgery, Keio University School of Medicine, Tokyo, Japan.
153. Institute of Research, Tokyo Medical and Dental University, Tokyo, Japan.
154. Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto, Japan.
155. Department of Medicine, Center for Hematology and Regenerative Medicine, Karolinska Institute, Stockholm, Sweden.
156. Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan.
157. These authors contributed equally.
158. These authors jointly supervised the study.