frontiers | Research Topics

# Genome-wide association studies of COVID-19 among diverse human populations

**Edited by**
Zhongshan Cheng, Yi Cai, Jingxuan Zhang, Ke Zhang and Hongsheng Gui

**Published in**
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-83251-816-8
DOI 10.3389/978-2-83251-816-8

## About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

# Genome-wide association studies of COVID-19 among diverse human populations

**Topic editors**

Zhongshan Cheng — St. Jude Children's Research Hospital, United States
Yi Cai — Shenzhen University, China
Jingxuan Zhang — Hubei University of Medicine, China
Ke Zhang — Guizhou Medical University, China
Hongsheng Gui — Henry Ford Health System, United States

**Citation**

Cheng, Z., Cai, Y., Zhang, J., Zhang, K., Gui, H., eds. (2023). *Genome-wide association studies of COVID-19 among diverse human populations*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-816-8

# Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 01 December 2022
DOI 10.3389/fgene.2022.1088026



OPEN ACCESS

EDITED AND REVIEWED BY
Maxim B. Freidin,
Queen Mary University of London,
United Kingdom

*CORRESPONDENCE
Zhong-Shan Cheng,
cheng.zhong.shan@gmail.com

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 02 November 2022
ACCEPTED 14 November 2022
PUBLISHED 01 December 2022

CITATION
Cheng Z-S (2022), Editorial: Genome-wide association studies of COVID-19 among diverse human populations. *Front. Genet.* 13:1088026. doi: 10.3389/fgene.2022.1088026



# Editorial: Genome-wide association studies of COVID-19 among diverse human populations

Zhong-Shan Cheng*

St. Jude Children's Research Hospital, Memphis, TN, United States

KEYWORDS

**COVID-19, long COVID, GWAS, SNP, genetic association, sex-biased, African**

Editorial on the Research Topic
Genome-wide association studies of COVID-19 among diverse human populations

It is my pleasure to share with research communities this Research Topic on genome-wide association studies (GWAS) relating to coronavirus disease 2019 (COVID-19), titled "*Genome-wide association studies of COVID-19 among diverse human populations*." Our effort to identify COVID-19 risk factors that are relevant to individuals with a variety of different ancestries is a fundamentally important endeavor, and such findings can be generated by mining open-source data consisting of GWAS summary statistics shared by the GRASP COVID-19 portal (Thibord et al., 2022) and the COVID-19 Host Genetic Initiative (HGI) (Velavan et al., 2021).

The COVID-19 pandemic was caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Hu et al., 2021), which has devastated global populations during the past 3 years. Up to this point, the global death toll caused by COVID-19 is ~6.6 million according to the COVID-19 dashboard (Dong et al., 2020). Our knowledge of SARS-CoV-2 infection has increased at the price of a heavy fatality rate over the past 3 years. We now know that SARS-CoV-2 exploits human host factors to infect target cells by utilizing its viral spike (S) protein to bind human ACE2 protein on the surface of cells, and subsequent cleavage of the S protein by the human serine protease TMPRSS2 primes the infection by allowing fusion of the viral and lysosomal membranes (Hoffmann et al., 2020). COVID-19 induced by SARS-CoV-2 infection varies in symptoms, ranging from asymptomatic or mild COVID-19 to COVID-19 that may even be life-threatening; this manifests in the form of acute respiratory distress or excessive inflammation, which are hallmarks of severe COVID-19 (Wong et al., 2004; Zhang et al., 2020). An increasing number of studies also suggest that 5%–50% of COVID-19 patients later experience long COVID (Ledford, 2022), this being a more complex condition with over 200 symptoms, including "brain fog," fever, fatigue, exhaustion, heart damage,

and other chronic conditions such as depression and cognitive impairments that last for months or even years after SARS-CoV-2 infection. The exact cause of long COVID is still under investigation, with one leading theory pointing to the persistence of fragments of SARS-CoV-2 virus in the tissue of long COVID patients, causing prolonged symptoms (Yang et al., 2021), and other hypotheses suggesting that SARS-CoV-2 infection triggers a shift in the balance of the renin–angiotensin system (RAS) from the Mas (angiotensin-converting enzyme 2 [ACE2]/angiotensin [Ang] 1–7/Mas) axis to the RAS (Ang-converting enzyme [ACE]/Ang II/Ang II type I receptor [AT1R]) axis (Uysal et al., 2022). A new study has further revealed that the development of autoantibodies against Ang II in severe COVID-19 is correlated with blood pressure dysregulation and COVID-19 severity (Briquez et al., 2022).

Genetic factors have been found to contribute to variation in COVID -19 symptoms; although rapidly evolving and mutating SASR-CoV-2 strains may also lead to different COVID-19 symptoms, the effect of this variation on COVID-19 severity may not be comparable to the major effect contributed by host genetic factors. For example, in the Research Topic, Gjorgjievska et al. report on two cases in which two SARS-CoV-2 variants, namely Omicron BA.1 and BA.2, were detected in a single individual, and the Omicron BA.2 variant subsequently became the dominant one in both these two patients. However, their symptoms were mild, with no symptoms after 4 days. In further support of the above view, a promoter SNP of the SARS-CoV-2 receptor ACE2 has been demonstrated by Luo et al. to be associated with both higher levels of *ACE2* expression in brain tissues and lower risk of COVID-19 hospitalization; variation in *ACE2* may impact both severe COVID-19 and long COVID. Aside from genetic elements of the SARS-CoV-2 receptor *ACE2*, other genetic factors associated with predisposition to severe COVID-19 have been identified (Anastassopoulou et al., 2020; Gandhi et al., 2020; Severe Covid-19 GWAS Group, 2020; Pairo-Castineira et al., 2021), such as single nucleotide polymorphisms (SNPs) or indels located in *OAS* genes, *IFNAR2*, *CCR9*, and *CXCR6*, as well as other genes (Pairo-Castineira et al., 2021; Zhou et al., 2021). In the Research Topic, Zecevic et al. detected a SNP close to *KLHL1* that is associated with COVID-19 in the Serbian population. Glessner et al. also identified several genes (*SEMA6D*, *FMN1*, *ACTN1*, *PDS5B*, *NFIA*, *ADGRL3*, *MMP27*, *TENM3*, *SPRY4*, *MNS1*, and *RSU1*) potentially associated with COVID-19 susceptibility in children. Li et al. reported the nominal associations of potential regulatory SNPs mapped to *TNF*, *IFNAR2*, *APOE*, *FOXP4-AS1*, *ABO*, and *IFITM3*, with COVID-19 in Chinese population.

Sex differences in responses to severe COVID-19 and long COVID are increasingly recognized by researchers. During the early pandemic, with no COVID-19 vaccine available to the global population, there was a tendency toward higher rates of COVID-19 hospitalization and death among men. This may be attributed to the differences between sexes in immune responses to SARS-CoV-2 infection, as women produce more robust immune responses to the infection (Takahashi et al., 2020; Shattuck-Heidorn et al., 2021; Takahashi et al., 2021; Danielsen et al., 2022). Nevertheless, women have been reported to have a higher chance of experiencing long COVID (Sylvester et al., 2022). In the Research Topic, Luo et al. performed sex-biased GWAS analysis for COVID-19 hospitalization and identified three immunity genes (*TRIM21*, *TRIM29*, and *PVLR1*) and two brain-related genes (*KNDC1* and *STK32C*) showing sex-biased effects on COVID-19 hospitalization. The expression of the two aforementioned genes related to brain function is interrupted during SARS-CoV-2 infection, strongly suggesting their potential involvement in COVID-19 or long COVID.

Genetic ancestry may also affect COVID-19 severity. Raza and Abbasi report that the COVID-19 risk SNP rs2236757 residing in the *IFNAR2* gene has undergone recent positive selection in the African population. However, most recently published findings on COVID-19 risk SNPs or genes are mainly found in populations of European ancestry; few genetic risk factors for COVID-19 have been identified in other populations, such as African populations. In the Research Topic, Petersen et al. discuss the limited reporting of COVID-19 risk markers among African populations and point out that this limitation could result in health disparities between African populations and other groups, as it is difficult to extrapolate published COVID-19 genetic risk factors from European populations to African or other populations.

Since multiple comorbidities, such as hypertension, diabetes, and lung cancer, have been observed along with severe COVID-19, genetic factors associated with predisposition to both COVID-19 and these comorbidities are reasonable targets for investigation. To investigate the connection between hypertension and severe COVID-19, Cai et al. performed meta-analyses of 68 observational studies related to COVID-19 severity and hypertension; they concluded that hypertension patients with COVID-19 have a 1.8-fold chance of experiencing severe COVID-19. Further genome-wide cross-trait meta-analysis revealed that genetic variants of genes expressed in the lung, such as *CCR1*/*CCR5* and *IL10RB*, may confer liability to both hypertension and severe COVID-19. Additionally, Luo et al. (under peer review) searched for genetic factors affecting the outcomes of both hypertension and COVID-19 and identified one regulatory SNP located in the gene *SPEG* that is potentially associated with severe COVID-19 in women but not men. Furthermore, Faridzadeh et al. report an association between an indel of *ACE1* and severe COVID-19 in the Iranian population.

In summary, the open-source data consisting of GWAS summary statistics shared by GRASP and HGI have dramatically propelled global research communities toward discoveries of genetic risk markers for COVID-19. Long COVID is now emerging among COVID-19 patients and may devastate more patients worldwide. Given the greater complexity of long COVID and the substantial number of patients who may experience this condition, global research communities need to work together to reduce the medical burden imposed by it and ultimately to eliminate it.

## Author contributions

Z-SC conceived and wrote the manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

Anastassopoulou, C., Gkizarioti, Z., Patrinos, G. P., and Tsakris, A. (2020). Human genetic factors associated with susceptibility to SARS-CoV-2 infection and COVID-19 disease severity. *Hum. Genomics* 14, 40. doi:10.1186/s40246-020-00290-4

Briquez, P. S., Rouhani, S. J., Yu, J., Pyzer, A. R., Trujillo, J., Dugan, H. L., et al. (2022). Severe COVID-19 induces autoantibodies against angiotensin II that correlate with blood pressure dysregulation and disease severity. *Sci. Adv.* 8, eabn3777. doi:10.1126/sciadv.abn3777

Danielsen, A. C., Lee, K. M., Boulicault, M., Rushovich, T., Gompers, A., Tarrant, A., et al. (2022). Sex disparities in COVID-19 outcomes in the United States: Quantifying and contextualizing variation. *Soc. Sci. Med.* 294, 114716. doi:10.1016/j.socscimed.2022.114716

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *Lancet. Infect. Dis.* 20, 533–534. doi:10.1016/S1473-3099(20)30120-1

Gandhi, R. T., Lynch, J. B., and Del Rio, C. (2020). Mild or moderate covid-19. *N. Engl. J. Med.* 383, 1757–1766. doi:10.1056/NEJMcp2009249

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280. e278. doi:10.1016/j.cell.2020.02.052

Hu, B., Guo, H., Zhou, P., and Shi, Z. L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* 19, 141–154. doi:10.1038/s41579-020-00459-7

Ledford, H. (2022). How common is long COVID? Why studies give different answers. *Nature* 606, 852–853. doi:10.1038/d41586-022-01702-2

Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., Pasko, D., et al. (2021). Genetic mechanisms of critical illness in COVID-19. *Nature* 591, 92–98. doi:10.1038/s41586-020-03065-y

Severe Covid-19, GWAS Group, Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., et al. (2020). Genomewide association study of severe covid-19 with respiratory failure. *N. Engl. J. Med.* 383, 1522–1534. doi:10.1056/NEJMoa2020283

Shattuck-Heidorn, H., Danielsen, A. C., Gompers, A., Bruch, J. D., Zhao, H., Boulicault, M., et al. (2021). A finding of sex similarities rather than differences in COVID-19 outcomes. *Nature* 597, E7–E9. doi:10.1038/s41586-021-03644-7

Sylvester, S. V., Rusu, R., Chan, B., Bellows, M., O'Keefe, C., and Nicholson, S. (2022). Sex differences in sequelae from COVID-19 infection and in long COVID syndrome: A review. *Curr. Med. Res. Opin.* 38, 1391–1399. doi:10.1080/03007995.2022.2081454

Takahashi, T., Ellingson, M. K., Wong, P., Israelow, B., Lucas, C., Klein, J., et al. (2020). Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature* 588, 315–320. doi:10.1038/s41586-020-2700-3

Takahashi, T., Ellingson, M. K., Wong, P., Israelow, B., Lucas, C., Klein, J., et al. (2021). Reply to: A finding of sex similarities rather than differences in COVID-19 outcomes. *Nature* 597, E10–E11. doi:10.1038/s41586-021-03645-6

Thibord, F., Chan, M. V., Chen, M. H., and Johnson, A. D. (2022). A year of Covid-19 GWAS results from the GRASP portal reveals potential genetic risk factors. *HGG Adv.* 3, 100095. doi:10.1016/j.xhgg.2022.100095

Uysal, B., Akça, T., Akacı, O., and Uysal, F. (2022). The prevalence of Post-COVID-19 hypertension in children. *Clin. Pediatr.* 61, 453–460. doi:10.1177/00099228221085346

Velavan, T. P., Pallerla, S. R., Ruter, J., Augustin, Y., Kremsner, P. G., Krishna, S., et al. (2021). Host genetic factors determining COVID-19 susceptibility and severity. *EBioMedicine* 72, 103629. doi:10.1016/j.ebiom.2021.103629

Wong, C. K., Lam, C. W., Wu, A. K., Ip, W. K., Lee, N. L., Chan, I. H., et al. (2004). Plasma inflammatory cytokines and chemokines in severe acute respiratory syndrome. *Clin. Exp. Immunol.* 136, 95–103. doi:10.1111/j.1365-2249.2004.02415.x

Yang, A. C., Kern, F., Losada, P. M., Agam, M. R., Maat, C. A., Schmartz, G. P., et al. (2021). Dysregulation of brain and choroid plexus cell types in severe COVID-19. *Nature* 595, 565–571. doi:10.1038/s41586-021-03710-0

Zhang, B., Zhou, X., Qiu, Y., Song, Y., Feng, F., Feng, J., et al. (2020). Clinical characteristics of 82 cases of death from COVID-19. *PLoS One* 15, e0235458. doi:10.1371/journal.pone.0235458

Zhou, S., Butler-Laporte, G., Nakanishi, T., Morrison, D. R., Afilalo, J., Afilalo, M., et al. (2021). A Neanderthal OAS1 isoform protects individuals of European ancestry against COVID-19 susceptibility and severity. *Nat. Med.* 27, 659–667. doi:10.1038/s41591-021-01281-1

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 March 2022
doi: 10.3389/fgene.2022.859508

# An Evolutionary Insight Into the Heterogeneous Severity Pattern of the SARS-CoV-2 Infection

*Rabail Zehra Raza* and Sumra Wajid Abbasi*

*NUMS Department of Biological Sciences, Faculty of Multidisciplinary Studies, National University of Medical Sciences, Rawalpindi, Pakistan*

The ongoing pandemic of COVID-19 has elaborated an idiosyncratic pattern of SARS-CoV-2-induced symptoms in the human host. Some populations have succumbed to the SARS-CoV-2 infection in large numbers during this pandemic, whereas others have shown a resilient side by manifesting only milder or no symptoms at all. This observation has relayed the onus of the heterogeneous pattern of SARS-CoV-2-induced critical illness among different populations to the host genetic factors. Here, the evolutionary route was explored and three genetic loci, i.e., rs10735079, rs2109069, and rs2236757, associated with COVID-19 were analyzed. Among the three, the risk allele A at genetic locus rs2236757 residing in the *IFNAR2* gene was observed to have undergone recent positive selection in the African population.

**Keywords: COVID-19, host genetic factors, natural selection, population genomics, SNP variants**

OPEN ACCESS

***Edited by:***
*Ramcés Falfán-Valencia, Instituto Nacional de Enfermedades Respiratorias-México (INER), Mexico*

***Reviewed by:***
*Arutha Kulasinghe, The University of Queensland, Australia*
*Leiliang Zhang, Shandong First Medical University, China*
*Prashanth N. Suravajhala, Amrita Vishwa Vidyapeetham University, India*

****Correspondence:***
*Rabail Zehra Raza*
*rabail.zehra@numspak.edu.pk*

***Specialty section:***
*This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics*

***Received:*** *21 January 2022*
***Accepted:*** *28 February 2022*
***Published:*** *22 March 2022*

***Citation:***
*Raza RZ and Abbasi SW (2022) An Evolutionary Insight Into the Heterogeneous Severity Pattern of the SARS-CoV-2 Infection. Front. Genet. 13:859508. doi: 10.3389/fgene.2022.859508*

## INTRODUCTION

Coronaviruses have been around for the last 2 decades and were declared pathogenic to humans in the early 21st century after the first severe acute respiratory syndrome (SARS) outbreak (Cui et al., 2019). The recent worldwide surge in the novel SARS-CoV-2 infection during 2020 has made it a global pandemic. SARS-CoV-2 has a single-stranded RNA in its genome which depends on RNA-dependent RNA polymerase for its replication (Siqueira et al., 2021). RNA viruses are prone to mutations. The more the RNA virus replicates, the more changes it accumulates in the genome because of a lack of proofreading polymerase activity (Shen et al., 2020). Because of this rapid intra-host replication, highly related viral entities of RNA viruses (quasi species) arise in the infected host (Siqueira et al., 2021). Within-host evolution of viruses has previously been reported for many RNA viruses such as MERS, SARS-CoV-1 and influenza (Xue et al., 2018; Al Khatib et al., 2020). In the case of COVID-19, Shen et al. (2020) identified 0 to 51 viral entities per hospitalized COVID-19 patient from the Chinese District, Wuhan, in December 2019. The SARS-CoV-2 quasi species has also been analyzed in relation to disease severity in COVID-19 patients. One such study reported significant diversity in SARS-CoV-2 genomes at the sub-consensus sequence level between mild and severe patients and observed a considerable increase in the number of coding and non-coding variants in severe cases as compared to the mild ones (Al Khatib et al., 2020). However, scarcity of significant variation in SARS-CoV-2 genomes at the consensus level (where similarity of all viral sequences is greater than 99.8%) has led scientists to believe that the host genetic factors, for instance, age, gender,

**Abbreviations:** EHH, extended haplotype homozygosity; GWAS, genome wide association study; LD, linkage disequilibrium; SNV, single nucleotide variant.

and other underlying comorbidities, along with environmental and social factors, play a vital role in determining COVID-19 severity among patients (Guan et al., 2020).

The World Health Organization (WHO) has reported more than 100 million confirmed cases of COVID-19 across 223 countries since the start of the pandemic. The phenotypic results of the SARS-CoV-2 infection are in stark contrast, with some patients showing mild to no visible symptoms and others undergoing fatal respiratory distress (Siqueira et al., 2021). In multiple studies, people with male gender, older age, smoking history, cancer, and other underlying comorbidities such as obesity, hypertension, and autoimmune disorders have been identified as vulnerable groups to getting severely infected with SARS-CoV-2 (Atkins et al., 2020). Although a broader risk group for COVID-19 mortality with pre-existing comorbidities has been identified, the dilemma of idiosyncratic symptomatic responses to SARS-CoV-2 infection in otherwise healthy patients is still under discussion (Hu et al., 2020; Williamson et al., 2020). It also remains a conundrum as to why certain populations have shown a much greater mortality rate associated with COVID-19 than others. For instance, in Africa, the number of deaths reported from SARS-CoV-2 infection was predicted to be much higher given the continent's higher population density, weaker healthcare systems, lower finances, and lack of preparedness in the wake of a global pandemic (Mbow et al., 2020; Maeda and Nkengasong, 2021). However, on the contrary, the number of COVID-19 deaths reported in Africa has been much lower than expected. According to the Africa CDC, the number of COVD-19 deaths till November 2020, made up 3.6% of the total worldwide cases (https://africacdc.org/covid-19) (Maeda and Nkengasong, 2021). In the recent upsurge of the OMICRON crisis in Africa, the casualty rate has surpassed 0.2 million by early 2022, as reported by the Africa CDC (https://africacdc.org/covid-19), which is not equal to even half of the casualties (0.86 million) reported from the US alone because of the SARS-CoV-2 pandemic. Although myriad reasons could be called upon for populations who seemingly did not get affected by COVID-19 as much as others, such as poor reporting, testing, and having a younger population, to name a few, the fickle nature of the symptoms among the same human host at different geographical distributions needs a robust investigation (Chitungo et al., 2020). Various aspects of the COVID-19 host-specific severity have been explored, of which rapid mutations in the SARS-CoV-2 RNA genome have also been taken into account between the severe and milder cases. However, the results do not suffice the answer as to why some populations showed a greater casualty rate.

To gauge the disparity in the number of COVID-19 deaths among different populations or even between the individuals of the same population, several studies have put forth the significance of within-host diversity of SARS-CoV-2 genomes between mild and severe cases of COVID-19 (Al Khatib et al., 2020; Shen et al., 2020). Within-host diversity of SARS-CoV-2 genomes has been determined at the consensus and sub-consensus levels in mild and severe cases of COVID-19. Although the within-host diversity of SARS-CoV-2 genomes has been identified at the sub-consensus level, indicating more variants in the SARS-CoV-2 genomes in severe cases, the importance of host genetic factors in creating erratic immune responses to the SARS-CoV-2 infection in some individuals cannot be ignored. Therefore, host genetic factors are deemed crucially important in the case of the COVID-19 severity conundrum among the populations. In order to analyze the heterogeneous trend of COVID-19 severity, evolution of the host genome with regard to COVID-19-associated genetic loci in different populations could show promising results. In this study, population-wise haplotype-based analysis was conducted by employing 1000 Genomes phase III data on three genetic loci associated with COVID-19 and signatures of selection on them were analyzed (Nature, 2015).

# MATERIALS AND METHODS

## Data Collection

In this study, two GWAS studies conducted for COVID-19 associations meeting the genome-wide significance threshold (P-value $< 5 \times 10^{-8}$) were referred to (Group, 2020; Pairo-Castineira et al., 2020). Among the two studies, the older investigation published in June 2020 identified the association of two SNPs, rs11385942 (INDEL: INsertion-DELetion) and rs657152 (SNV: single nucleotide variant) with COVID-19 in a European cohort (Italian and Spanish). The former SNP rs11385942 with a genome-wide association P-value = $1.15 \times 10^{-10}$ was located in a chromosomal location harboring many immunity-related genes such as *CXCR6*, *CCR1* and *CCR2* in close proximity (Group, 2020). The latter SNP rs657152 (A > C) is situated in the ABO blood group locus with a P-value = $4.95 \times 10^{-8}$ in the meta-analysis (Group, 2020). The second GWAS study was published in December 2020 after investigating the critical care patients of the UK and identified associations of three SNPs, rs10735079 (SNV: A > G, P-value = $1.65 \times 10^{-8}$), rs2109069 (SNV: A > G, P-value = $3.98 \times 10^{-12}$), and rs2236757 (SNV: G > A, P-value = $4.99 \times 10^{-8}$) with critical COVID-19-induced illness (Pairo-Castineira et al., 2020). Among the three SNPs, the neighboring genes such as *IFNAR2* and *OAS* genes are the immunity-related genes involved in the innate anti-viral defense response by the host (Pairo-Castineira et al., 2020).

## 1000 Genomes Phase III SNP Data

In this study, the 1000 Genomes Phase III SNP data for the analysis was referred to. There were shortlisted three single nucleotide variations (SNVs) among the aforementioned COVID-19-associated SNPs with neighboring/flanking genes because of their immunity-related function, i.e., rs10735079 (A > G), rs2109069 (A > G) and rs2236757 (G > A) residing in the *OAS* gene cluster, within *DPP9* and within *IFNAR2*, respectively (Pairo-Castineira et al., 2020). Because of the limitation that only SNVs can be used for haplotype-based tests in this study, it was not shortlisted for analysis even though the genes lying within the vicinity have an immunity-related function (Group, 2020). In order to collect the SNP data for a regional analysis of length as long as 1 Mb, VCF files pertaining to a 0.5 Mb region were collected on either side of the three aforementioned SNPs

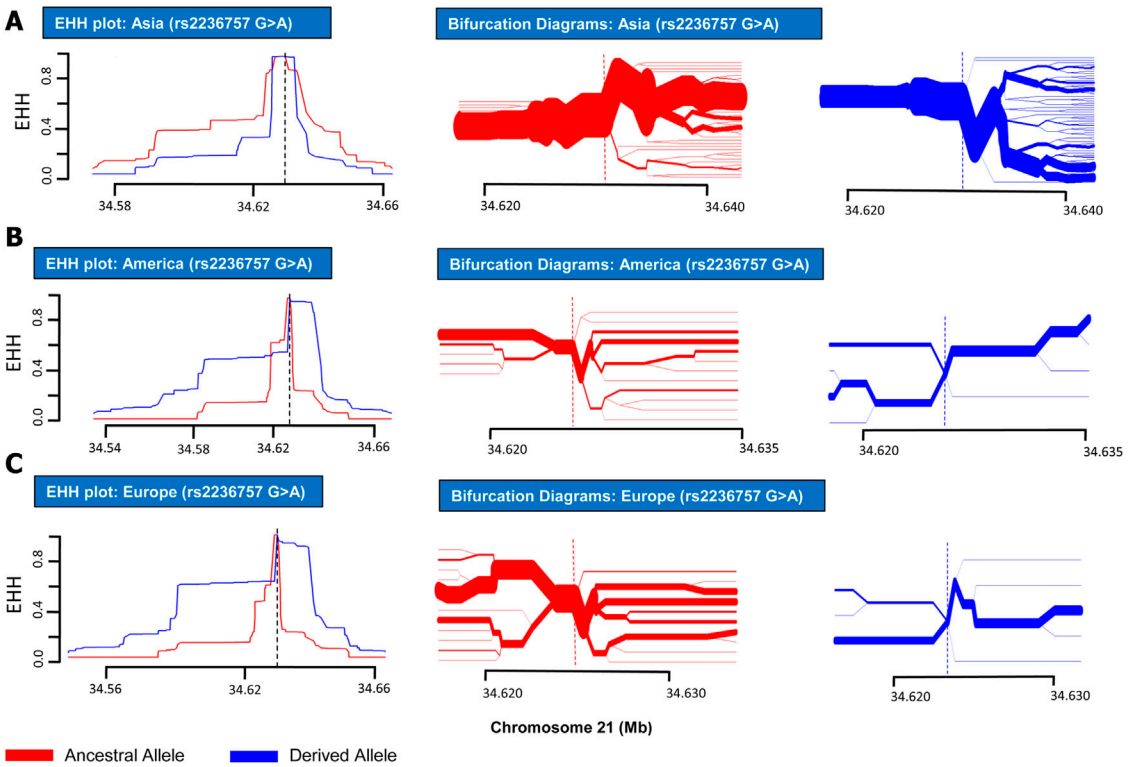



**FIGURE 1 |** EHH plots and Bifurcation Diagrams for SNP rs2236757 (in non-African Populations). EHH plots and Bifurcation Diagrams for SNP rs2236757 in Asian **(A)**, American **(B)** and European **(C)** populations. EHH = I on Y-axis indicates all haplotypes carrying either ancestral or derived state of the allele are matching upto this point. X-axis contains coordinates for human chromosome 21. Ancestral allele is shown before the derived allele, separated by a ">" symbol. In the EHH plots, smaller area under the curve for both ancestral and derived alleles (G > A) shows no signs of recent positive selection in any of the aforementioned populations.

from the 1000 Genomes Phase III SNP data (Nature, 2015; Zehra et al., 2018). All three SNPs had a minor allele frequency ≥0.05 and were used to assess signals of positive selection by the subsequent haplotype-based tests in 2504 individuals of the 1000 Genomes Phase III data belonging to African, European, Asian, and American samples.

## Haplotype Based Selection Tests

To build a selection regime in a population, the two haplotypes of an individual, acquired from each parental chromosome, are necessary. This explains haplotype inference or phasing, a critical stage in population genetics research to separate the genotype information inherited from both parents (Salem et al., 2005). As phased haplotypes are needed to calculate the Extended Haplotype Homozygosity (EHH) test and haplotype bifurcation diagrams, the VCF files were first phased using fastPHASE to reconstruct haplotypes (Sabeti et al., 2002; Scheet and Stephens, 2006). EHH plots and haplotype bifurcation diagrams were made using the rehh package in R (Gautier and Vitalis, 2012). Furthermore, in order to gauge the genetic differentiation between the aforementioned subpopulations, Weir and Cockerham fixation index ($F_{st}$) values were also evaluated using the VCFtools (Danecek et al., 2011). The $F_{st}$ values ≥0.1 were considered significant. Moreover, Haploreg (version 4.1) and linkage disequilibrium (LD) calculator at the Ensembl genome browser were also used for corroborating the haplotype blocks of adjacent SNPs with LD ($r^2$) ≥ 0.8 that confirmed the long, unbroken haplotypes resulted by applying EHH test and the haplotype bifurcation diagrams (Ward and Kellis, 2012; Cunningham et al., 2015).

# RESULTS AND DISCUSSION

Polymorphisms in the host genes such as *ACE2, TMPRSS2*, and *ADAM17* have been associated with their expression levels and ultimately influence the mechanism of SARS-CoV-2 infectivity and severity (Brest et al., 2020). In the human genome, mutations or genetic variants (alleles) on a locus can contribute to fitness and, because of the advantageousness they impart on the phenotypic fitness of the species, can undergo positive selection. Positive selection on beneficial alleles increases their frequency in a population, whereas negative selection discards the deleterious alleles (Karlsson et al., 2014). In a phenomenon known as linkage disequilibrium (LD), the signals of positive selection on a genomic position increase the frequency of the beneficial allele along with the neighboring alleles in a non-random manner, which in turn reduces genetic diversity in the entire locus (Cadzow et al., 2014). Therefore, in light of the non-

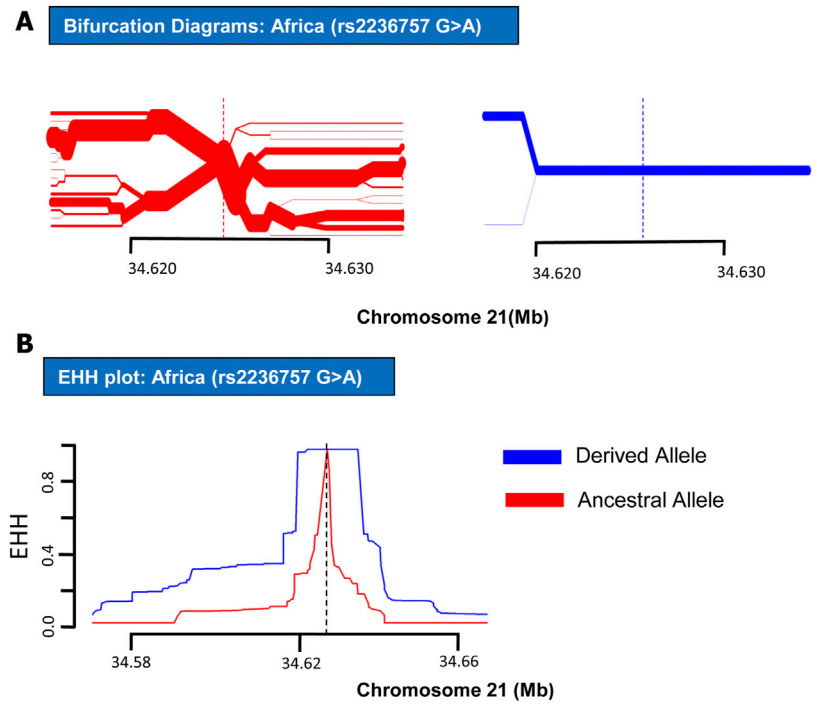



**FIGURE 2 |** EHH plot and Bifurcation Diagrams for SNP rs2236757 (in African Populations). EHH plots and Bifurcation Diagrams of SNP rs2236757in African populations. **(A)** Bifurcation Diagram of the derived variant of the SNP rs2236757 (shown in blue) shows long haplotype and absolutely no branching at the nodes upto 14.6 kb region. **(B)** EHH plot for SNP rs2236757 shows derived allele A (shown in blue) is under positive selection. EHH = 1 indicates all haplotypes carrying either ancestral or derived state of the allele are matching upto this point.

**TABLE 1 |** Weir and Cockerham $F_{st}$ values evaluated for SNPs rs10735079, rs2109069 and rs2236757.

| S.No. | SNPs | Genomic coordinates (hg19) | Genes | A > D[a] | Weir and cockerham $F_{st}$ | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Africa | Asia | America | Europe |
| 1 | rs10735079 | chr12: 113380008 | OAS1, OAS2, OAS3 | A > G | 0.025 | 0 | 0 | 0.039 |
| 2 | rs2109069 | chr19: 4719443 | DPP9 | A > G | 0.0007 | 0.008 | 0 | 0.054 |
| 3 | rs2236757 | chr21: 34624917 | IFNAR2 | G > A | **0.120** | 0.075 | 0.007 | 0.020 |

[a]*The bold value indicate a higher Weir and Cockerham $F_{st}$ value for SNP rs2236757 in the African population*

random association of the alleles associated with COVID-19 with their neighboring alleles, we can provide you with useful contextual information on seeing the pattern of positive selection in different human populations and the selective advantage it might be imparting on a certain population.

In the wake of a pandemic, two significant GWAS studies have been put forth that have successfully associated five genetic loci with COVID-19 severity. In this work, three out of five SNPs (also SNVs) associated with COVID-19 severity lie in or within the close proximity of immunity-related genes were focused on from an evolutionary perspective (see methods). The shortlisted three SNPs in this study are a result of a GWAS conducted on 2244 critical care patients with COVID-19 in the UK (Pairo-Castineira et al., 2020). The three novel COVID-19-associated SNPS are 1) rs10735079 in gene cluster of *OAS1, OAS2* and *OAS3*, 2) rs2109069 within *DPP9* near gene encoding tyrosine kinase 2 (*TYK2*) and 3) and rs2236757 in the interferon receptor gene *IFNAR2* (Pairo-Castineira et al., 2020).

In order to analyze positive selection on the aforementioned three SNPs, statistical approaches such as EHH tests and haplotype bifurcation diagrams were applied to the SNP data collected from the 1000 Genomes Phase III (Sabeti et al., 2002). By applying EHH tests and haplotype bifurcation diagrams, it was found that the derived minor allele "A" of SNP rs2236757 residing in the *IFNAR2* gene has undergone recent positive selection in the African population alone out of the four population categories (African, European, Asian, and American), whereas no positive selection signals were identified in any of the population categories for the ancestral major allele "G" of SNP rs2236757 (**Figure 1**). In 1322 haplotypes of samples of African individuals from 1000 Genomes Phase III, unbroken haplotypes, indicative of stronger linkage

disequilibrium, were observed to be up to 15 kb in length at an EHH value of 1 for derived minor allele "A" of SNP rs2236757 (**Figure 2**). In LD analysis carried out at Ensembl, it was observed that the SNP rs2236757 is co-inherited with the neighboring SNP rs2073361 in CLM and MXL (America), with LD ($r^2$) of 0.8486 and 0.9394, respectively (Cunningham et al., 2015). The higher LD ($r^2$) values indicate that the two SNPs are in strong LD and one of them is the causal SNP for such a behavior. Moreover, in Haploreg, LD ($r^2$) was also observed to be 1 for the SNP rs2236757 inclusive of the neighboring SNPs up to the said ~15 kb region in the African population, hence, indicating non-random association between the neighboring alleles and the SNP rs2236757 (Sabeti et al., 2002; Ward and Kellis, 2012). Furthermore, the $F_{st}$ value of rs2236757 in the African population was calculated via VCFtools and observed to be 0.12. The $F_{st}$ value higher than 0.1 is generally indicative of a significant high level of genetic differentiation between one population and the rest of the populations (**Table 1**) (Danecek et al., 2011). It is also interesting to note that the major allele "G" of SNP rs2236757 was found to be conserved in all of the 37 Eutherian mammalian species at Ensembl (Cunningham et al., 2015). On similar lines, EHH plots and haplotype bifurcation diagrams when applied to the remaining SNPs/SNVs rs10735079 and rs2109069 did not indicate longer, unbroken haplotypes of considerable length. Therefore, no positive selection signals were observed in any of the sub-populations on the respective derived and ancestral alleles of the SNPs rs10735079 and rs2109069 (**Supplementary Figures S1, S2**). A schematic flow of the results obtained can be viewed in **Supplementary Figure S3**.

The evolutionarily selected interferon (IFN)-mediated innate immune response is inbred in genomes and provides a powerful initial line of defense against invading pathogens (Schneider et al., 2014). Type 1 IFNs comprise the largest class that exhibit varied binding affinity with the IFNAR1/2 receptor complex and as a result diversified anti-viral responses are induced and amplified in the host (Moraga et al., 2009). In a recent cohort-based study, pulmonary tissue samples from the severely affected patients of COVID-19 and pH1N1 influenza showed differential expression of two genes, *IFI27* and *IFI6,* both belonging to type 1 IFNs (Kulasinghe et al., 2021). The findings for differential expression of the *IFN* genes controlling the immunoregulatory responses have also been corroborated in transcriptomic profiling of the hospitalized COVID-19 patients (Ahern et al., 2021). In most cases of COVID-19 patients, genetic aberrations in antiviral innate immune interferon (IFN) loci and dysregulation of IFNs have also been correlated with the severity of the SARS-CoV-2 infection (Lopez et al., 2020).

IFNAR2 is a subunit of the type 1 IFN receptor complex. Upon binding of type 1 IFNs with the surface receptor complex, JAK kinases are induced along with the activation of STAT transcription factors, which in turn initiate the transcription of the immune response genes (Saleh et al., 2004). In recent GWAS studies, polymorphisms in the *IFNAR2* gene have shown a direct association with COVID-19 hospitalizations (Smieszek et al., 2021). IFNAR2 protein has also been nominated along with ACE2 as drug targets for expedited clinical trials (Gaziano et al., 2021). In summary, our results have indicated recent positive selection on derived risk allele "A" of SNP rs2236757 within the *IFNAR2* gene in the African population in the shape of a long, unbroken ~15 kb haplotype (**Figure 2**). However, it has been established that some risk alleles may be positively selected individually or as part of an underlying biological function because of a currently unknown advantage they may have imparted on the host genome (Corona et al., 2010). In spite of the presented data, because of the dubious nature of COVID-19 spread among different populations in the face of the emerging new variants, it is not yet conclusively possible to point out a population which could be at a selective advantage and therefore with a lower mortality rate due to COVID-19. Nonetheless, the identified positive selection on a risk allele of SNP rs2236757 in the intronic region of the *IFNAR2* gene holds importance. This study confers the idea that natural selection within immunity-related can be used as a tool in addressing the symptomatic idiosyncrasy of the current COVID-19 pandemic. Moreover, the results also highlight the need for more GWAS studies inclusive of diverse population data and subsequently extensive assessment of the genetic aberrations that can be done under the light of evolution to understand the heterogeneous severity pattern of COVID-19 among different human populations.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

RR conceived the project, analyzed the data, and wrote the manuscript. SA analyzed the data and wrote the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.859508/full#supplementary-material

## REFERENCES

Ahern, D. J., Ai, Z., Ainsworth, M., Allan, C., Allcock, A., Ansari, A., et al. (2021). A Blood Atlas of COVID-19 Defines Hallmarks of Disease Severity and Specificity. *Cell* S0092-8674 (22), 00070–00078. doi:10.1016/j.cell.2022.01.012

Al Khatib, H. A., Benslimane, F. M., Elbashir, I. E., Coyle, P. V., Al Maslamani, M. A., Al-Khal, A., et al. (2020). Within-Host Diversity of SARS-CoV-2 in COVID-19 Patients with Variable Disease Severities. *Front Cel Infect Microbiol* 10, 575613. doi:10.3389/fcimb.2020.575613

Atkins, J. L., Masoli, J. A. H., Delgado, J., Pilling, L. C., Kuo, C.-L., Kuchel, G. A., et al. (2020). Preexisting Comorbidities Predicting COVID-19 and Mortality in

the UK Biobank Community Cohort. *The Journals Gerontol. Ser. A* 75, 2224–2230. doi:10.1093/gerona/glaa183

Brest, P., Refae, S., Mograbi, B., Hofman, P., and Milano, G. (2020). Host Polymorphisms May Impact SARS-CoV-2 Infectivity. *Trends Genet.* 36, 813–815. doi:10.1016/j.tig.2020.08.003

Cadzow, M., Boocock, J., Nguyen, H. T., Wilcox, P., Merriman, T. R., and Black, M. A. (2014). A Bioinformatics Workflow for Detecting Signatures of Selection in Genomic Data. *Front. Genet.* 5, 293. doi:10.3389/fgene.2014.00293

Chitungo, I., Dzobo, M., Hlongwa, M., and Dzinamarira, T. (2020). COVID-19: Unpacking the Low Number of Cases in Africa. *Public Health Pract.* 1, 100038. doi:10.1016/j.puhip.2020.100038

Corona, E., Dudley, J. T., and Butte, A. J. (2010). Extreme Evolutionary Disparities Seen in Positive Selection across Seven Complex Diseases. *PLoS ONE* 5, e12236. doi:10.1371/journal.pone.0012236

Cui, J., Li, F., and Shi, Z.-L. (2019). Origin and Evolution of Pathogenic Coronaviruses. *Nat. Rev. Microbiol.* 17, 181–192. doi:10.1038/s41579-018-0118-9

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., et al. (2015). Ensembl 2015. *Nucleic Acids Res.* 43, D662–D669. doi:10.1093/nar/gku1010

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A., et al. (2011). The Variant Call Format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330

Gautier, M., and Vitalis, R. (2012). Rehh: an R Package to Detect Footprints of Selection in Genome-wide SNP Data from Haplotype Structure. 28**,** 1176–1177.doi:10.1093/bioinformatics/bts115

Gaziano, L., Giambartolomei, C., Pereira, A. C., Gaulton, A., Posner, D. C., Swanson, S. A., et al. (2021). Actionable Druggable Genome-wide Mendelian Randomization Identifies Repurposing Opportunities for COVID-19. *Nat. Med.* 27, 668–676. doi:10.1038/s41591-021-01310-z

Group, S. C.-G., Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., et al. (2020). Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N. Engl. J. Med.* 383, 1522–1534. doi:10.1056/NEJMoa2020283

Guan, W.-J., Ni, Z.-Y., Hu, Y., Liang, W.-H., Ou, C.-Q., He, J.-X., et al. (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* 382, 1708–1720. doi:10.1056/nejmoa2002032

Hu, J., Li, C., Wang, S., Li, T., and Zhang, H. J. M. (2020). Genetic Variants Are Identified to Increase Risk of COVID-19 Related Mortality from UK Biobank Data. *Hum. Genomics* 15 (1), 10. doi:10.1186/s40246-021-00306-7

Karlsson, E. K., Kwiatkowski, D. P., and Sabeti, P. C. (2014). Natural Selection and Infectious Disease in Human Populations. *Nat. Rev. Genet.* 15, 379–393. doi:10.1038/nrg3734

Kulasinghe, A., Tan, C. W., Dos Santos Miggiolaro, A. F. R., Monkman, J., Sadeghirad, H., Bhuva, D. D., et al. (2021). Profiling of Lung SARS-CoV-2 and Influenza Virus Infection Dissects Virus-specific Host Responses and Gene Signatures. *Eur. Respir. J.*, 2101881. doi:10.1183/13993003.01881-2021

Lopez, L., Sang, P. C., Tian, Y., and Sang, Y. (2020). Dysregulated Interferon Response Underlying Severe COVID-19. *Viruses* 12, 1433. doi:10.3390/v12121433

Maeda, J. M., and Nkengasong, J. N. (2021). The Puzzle of the COVID-19 Pandemic in Africa. *Science* 371, 27–28. doi:10.1126/science.abf8832

Mbow, M., Lell, B., Jochems, S. P., Cisse, B., Mboup, S., Dewals, B. G., et al. (2020). COVID-19 in Africa: Dampening the Storm? *Science* 369, 624–626. doi:10.1126/science.abd3902

Moraga, I., Harari, D., Schreiber, G., Uze´, G., Pellegrini, S., and Biology, C. (2009). Receptor Density Is Key to the Alpha2/beta Interferon Differential Activities. *Mol. Cel Biol* 29, 4778–4787. doi:10.1128/mcb.01808-08

Nature, G. P. C. J., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A Global Reference for Human Genetic Variation. *Nature* 526, 68–74. doi:10.1038/nature15393

Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., Pasko, D., et al. (2020). Genetic Mechanisms of Critical Illness in Covid-19. *Nature* 591 (7848), 92–98.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., et al. (2002). Detecting Recent Positive Selection in the Human Genome from Haplotype Structure. *Nature* 419, 832–837. doi:10.1038/nature01140

Saleh, A. Z. M., Fang, A. T., Arch, A. E., Neupane, D., Fiky, A. E., and Krolewski, J. J. (2004). Regulated Proteolysis of the IFNaR2 Subunit of the Interferon-Alpha Receptor. *Oncogene* 23, 7076–7086. doi:10.1038/sj.onc.1207955

Salem, R. M., Wessel, J., and Schork, N. J. J. H. G. (2005). A Comprehensive Literature Review of Haplotyping Software and Methods for Use with Unrelated Individuals. 2**,** 1–28.doi:10.1186/1479-7364-2-1-39

Scheet, P., and Stephens, M. (2006). A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. *Am. J. Hum. Genet.* 78, 629–644. doi:10.1086/502802

Schneider, W. M., Chevillotte, M. D., and Rice, C. M. (2014). Interferon-stimulated Genes: a Complex Web of Host Defenses. *Annu. Rev. Immunol.* 32, 513–545. doi:10.1146/annurev-immunol-032713-120231

Shen, Z., Xiao, Y., Kang, L., Ma, W., Shi, L., Zhang, L., et al. (2020). Genomic Diversity of Severe Acute Respiratory Syndrome-Coronavirus 2 in Patients with Coronavirus Disease 2019. *Clin. Infect. Dis.* 71, 713–720. doi:10.1093/cid/ciaa203

Siqueira, J. D., Goes, L. R., Alves, B. M., De Carvalho, P. S., Cicala, C., Arthos, J., et al. (2021). SARS-CoV-2 Genomic Analyses in Cancer Patients Reveal Elevated Intrahost Genetic Diversity. *Virus. Evol.* 7 (1), veab013. doi:10.1093/ve/veab013

Smieszek, S. P., Polymeropoulos, V. M., Xiao, C., Polymeropoulos, C. M., and Polymeropoulos, M. H. (2021). Loss-of-function Mutations in IFNAR2 in COVID-19 Severe Infection Susceptibility. *J. Glob. Antimicrob. Resist.* 26, 239–240. doi:10.1016/j.jgar.2021.06.005

Ward, L. D., and Kellis, M. (2012). HaploReg: a Resource for Exploring Chromatin States, Conservation, and Regulatory Motif Alterations within Sets of Genetically Linked Variants. *Nucleic Acids Res.* 40, D930–D934. doi:10.1093/nar/gkr917

Williamson, E. J., Walker, A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., et al. (2020). OpenSAFELY: Factors Associated with COVID-19 Death in 17 Million Patients. doi:10.1101/2020.05.06.20092999

Xue, K. S., Moncla, L. H., Bedford, T., and Bloom, J. D. (2018). Within-host Evolution of Human Influenza Virus. *Trends Microbiol.* 26, 781–793. doi:10.1016/j.tim.2018.02.007

Zehra, R., Abbasi, A. A., and Evolution (2018). Homo Sapiens-specific Binding Site Variants within Brain Exclusive Enhancers Are Subject to Accelerated Divergence across Human Population. 10**,** 956–966.doi:10.1093/gbe/evy052

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

CASE REPORT
published: 12 April 2022
doi: 10.3389/fgene.2022.892682

# Case Report: Omicron BA.2 Subvariant of SARS-CoV-2 Outcompetes BA.1 in Two Co-infection Cases

*Marija Gjorgjievska[1], Sanja Mehandziska[1], Aleksandra Stajkovska[2], Slavica Pecioska-Dokuzovska[1], Anica Dimovska[1], Idriz Durmish[1], Sara Ismail[1], Teodora Pavlovska[1], Antonija Stojchevska[1], Haris Amedi[1], Jasna Andonova[1], Marija Nikolovska[1], Sara Velickovikj[1], Zan Mitrev[1], Ivan Kungulovski[2] and Goran Kungulovski[2]**

[1]Zan Mitrev Clinic, Skopje, North Macedonia, [2]Bio Engineering LLC, Skopje, North Macedonia

Trends from around the world suggest that the omicron BA.2 subvariant is increasing in proportion to the original BA.1 subvariant. Here we report two cases of co-infection with omicron BA.1 and omicron BA.2 in co-exposed individuals. In both individuals, genome sequencing and/or S-gene specific PCR identified omicron BA.1 at early time-points, which was replaced by omicron BA.2 at later time-points of the infection. The timeline of our data supports the proposition that BA.2 outcompetes BA.1 in a real-life scenario, and in time becomes the dominant variant in the upper respiratory tract of the host.

**Keywords: COVID-19, co-infection, omicron, next-generation sequencing, SARS- CoV2**

OPEN ACCESS

***Edited by:***
*Stephen J. Bush,*
*University of Oxford, United Kingdom*

***Reviewed by:***
*Xiang-Jiao Yang,*
*McGill University, Canada*
*Haiguang Liu,*
*Beijing Computational Science Research Center (CSRC), China*
*Tim-Christoph Roloff Handschin,*
*University Hospital of Basel, Switzerland*
*Eugene V. Koonin,*
*National Library of Medicine (NIH), United States*

****Correspondence:***
*Goran Kungulovski*
*goran.kungulovski@zmc.mk*

***Specialty section:***
*This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics*

***Received:*** *09 March 2022*
***Accepted:*** *23 March 2022*
***Published:*** *12 April 2022*

***Citation:***
*Gjorgjievska M, Mehandziska S, Stajkovska A, Pecioska-Dokuzovska S, Dimovska A, Durmish I, Ismail S, Pavlovska T, Stojchevska A, Amedi H, Andonova J, Nikolovska M, Velickovikj S, Mitrev Z, Kungulovski I and Kungulovski G (2022) Case Report: Omicron BA.2 Subvariant of SARS-CoV-2 Outcompetes BA.1 in Two Co-infection Cases. Front. Genet. 13:892682. doi: 10.3389/fgene.2022.892682*

## INTRODUCTION

Coronavirus disease 19 (COVID-19) is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), first identified in Wuhan, China (Zhu et al., 2020). Throughout the pandemic, the SARS-CoV-2 virus has been continuously evolving leading to the emergence of new variants. The ones that posed an increased risk to global public health due to increased transmissibility, increased virulence, or immune evasion have been designated as variants of interest (VOIs) or variants of concern (VOCs) by the World Health Organization (https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/) (Otto et al., 2021). The VOCs Alpha, Beta, Gamma, and Delta were first detected in the second half of 2020, with Delta becoming the dominant variant for most of the second half of 2021. By the end of 2021, the omicron variant (BA.1) began overtaking the delta variant as the dominant strain, and by early 2022 it has become the dominant strain in Europe and USA due to its striking antibody evasion properties (Liu et al., 2022). Several countries, including Denmark, have observed two omicron subvariants (BA.1 and BA.2). Evidence from Denmark suggests that omicron BA.2 leads to 2-3 times increased susceptibility to infection compared to BA.1, and has rapidly replaced BA.1 as the dominant subvariant (Lyngse et al., 2022). In line with this evidence, we report two cases of co-infection with omicron BA.1 and omicron BA.2 in co-exposed individuals. In both individuals, genome sequencing and/or S-gene specific PCR identified omicron BA.1 at early time-points, which was replaced by omicron BA.2 at later time-points of the infection.

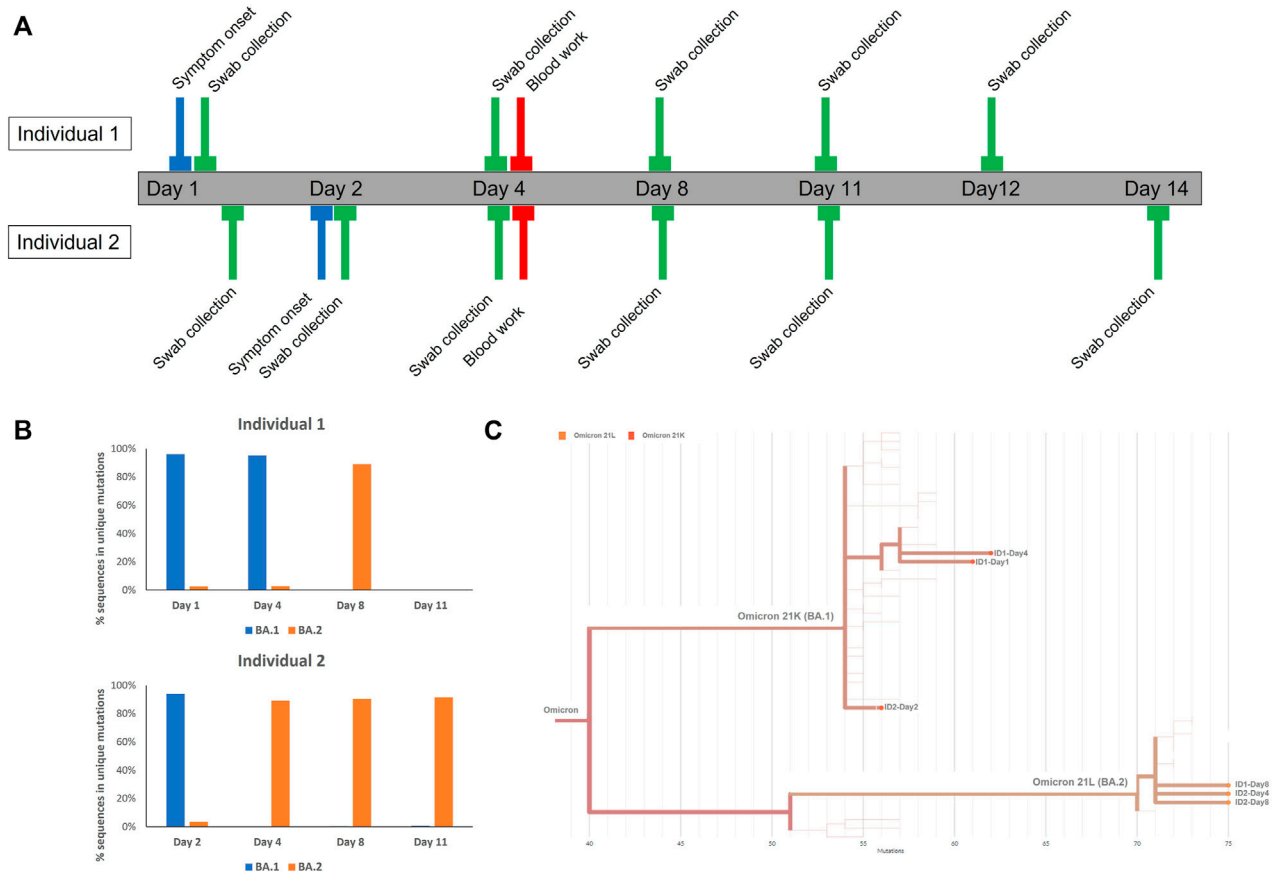



**FIGURE 1 | (A)** Roadmap of sample collection per tested individuals **(B)** Quantification of sequences mapping to unique regions in Omicron BA.1 and omicron BA.2 per day. More information can be found in **Supplementary Datasheet S2**. **(C)** Phylogenetic tree of genome assemblies.

**TABLE 1 |** Timeline of symptom emergence, PCR results, SGTF status and genome sequencing.

| | | Day 1 | Day 2 | Day 4 | Day 8 | Day 11 | Day 12 | Day 14 |
|---|---|---|---|---|---|---|---|---|
| Individual 1 | Ct (ORF1ab) | 25.5 | / | 24.5 | 23 | NEG | NEG | / |
| | SGTF (PCR) | YES | / | YES | NO | NEG | NEG | / |
| | NGS | Omicron BA.1 | / | Omicron BA.1 | Omicron BA.2 | / | / | / |
| | Symptoms | YES | YES | NO | NO | NO | NO | NO |
| Individual 2 | Ct (ORF1ab) | 34 | 27.5 | 19 | 21 | 24.5 | / | 29.5 |
| | SGTF (PCR) | NO | NO | NO | NO | NO | / | NO |
| | NGS | / | Omicron BA.1 | Omicron BA.2 | Omicron BA.2 | / | / | / |
| | Symptoms | NO | YES | NO | NO | NO | NO | NO |

## CASE PRESENTATION

Two individuals with COVID-19 symptoms were confirmed to be positive by PCR and then analyzed by genome sequencing at different time points at our institution (**Figure 1A**). First, a 25-year-old female (individual 1), which experienced throat scratching and low-grade fever on day 1, underwent SARS-CoV-2 PCR confirmatory testing after getting a positive lateral flow test. The assay resulted in a positive PCR test with S-gene target failure (SGTF), followed by another SGTF PCR test on day 4 (**Table 1**; **Supplementary Figure S1**). Interestingly, the follow-up PCR test on day 8 exhibited S-gene amplification (non-SGTF) (**Table 1**, **Supplementary Figure S1**). Consistent with the PCR results, SARS-CoV-2 genome sequencing confirmed the presence of omicron BA.1 in the samples collected on day 1 and day 4, and omicron BA.2 on day 8 (**Figures 1B,C**; **Supplementary Figure S2**). Unfortunately, samples were not collected on days 9 and 10, whilst PCR samples on days 11 and Day 12 were negative.

Her partner, a 35-year-old male (individual 2) tested positive at day 1 for trace amounts of SARS-CoV-2, yielding a non-SGTF

PCR, which was confirmed by follow-up non-SGTF PCRs at day 2, day 4, day 8, day 11, and Day 14 (**Table 1**; **Supplementary Figure S1**). He experienced symptom onset at day 2 manifested as throat discomfort early in the day and fever in the evening. Surprisingly, genome sequencing at day 2 and genome assembly analysis indicated infection with the omicron BA.1 subvariant, in contrast to the PCR result. More focused analysis of selected sequences in regions specific for BA.1 and BA.2 revealed the presence of sequences specific for both genomes, with the majority of sequences belonging to BA.1 (**Figure 1B**; **Supplementary Figure S2**). Due to gaps in genome sequencing and incomplete genome assembly (e.g the 69-70 deletion region was not covered), we were unable to carry out more precise comparative analyses, which might further explain the discrepancy between the sequencing and PCR results; only 5/15 BA.1, and 12/23 BA.2 mutations were covered with >10 sequences in this sample (**Supplementary Material S2**). Further sequencing at day 4 and day 8 unequivocally identified the BA.2 subvariant with none of the BA.1 specific sequences present in our samples (**Figures 1B,C**; **Supplementary Figure S2**).

Both individuals were fully vaccinated with two doses of the BNT162b2, COVID-19 mRNA vaccine, 9 and 11 months prior to the infection, respectively. The most likely route of omicron BA.1 infection for individual 1 was a beauty salon; individual 2 was most likely exposed to omicron BA.2 in a crowded restaurant. Both individuals underwent home isolation together and experienced mild symptoms. Blood work results at day 4 were within reference ranges.

## MATERIALS AND METHODS

### RNA Extraction and RT-PCR

For the detection of viral RNA by RT-PCR and sequencing, nasal and oropharyngeal swabs were collected for each time point. Both swabs were combined and immersed in saline solution and processed immediately. RNA was extracted with the abGenix (AITbiotech, Singapore) automatic DNA/RNA extractor, and PCR was carried out on QuantStudio™ 5 (Thermo Fisher Scientific, MA, USA) thermal cycler using the TaqPath protocol (ThermoFisher Scientific, MA, USA), according to the manufacturer's recommendations. Positive and negative controls were routinely included in each run.

### Viral Whole-Genome Sequencing

Reverse transcription using SuperScript IV Reverse Transcriptase (Invitrogen, Thermo Fisher Scientific, MA, USA) and SARS-CoV-2 genome amplification using the ARTIC panel of primers (Integrated DNA Technologies, IA, USA) was performed as described in the Nanopore protocol "*PCR tiling of COVID-19 virus" (Version: PTC_9096_v109_revD_06February 2020)*. The protocol was modified with rapid barcoding instead of native barcoding using the Rapid Barcoding Sequencing Kit, SQK-RBK004 (Oxford Nanopore Technologies, UK). The samples were sequenced on Flow Cell R9.4.1 using the MinION device.

Demultiplexing of the samples was carried out with the FASTQ Barcoding tool on the EPI2ME platform (Oxford Nanopore Technologies, UK). Multiple FASTQ files were concatenated into one, and genome assembly was conducted with the *medaka consensus pipeline* for creating a consensus sequence using Galaxy platform (Afgan et al., 2018). Consensus sequences were used to generate a phylogenetic tree in Nextstrain (https://clades.nextstrain.org/). Genome coverage was between 23x and 121x (**Supplementary Figure S3**) with two-thirds of bases covered with more than >20 sequences. Finally, editing and gap-filling were done in BioEdit. The coverage of regions specific for omicron BA.1 and omicron BA.2 (**Supplementary Datasheet S2**) was evaluated directly from the BAM files in the Integrative Genomic Viewer, Broad Institute (Robinson et al., 2017).

## DISCUSSION AND CONCLUDING REMARKS

In this report, we illustrate two interesting cases of co-infection with omicron BA.1 and omicron BA.2 in co-exposed individuals in the same household. In both individuals, genome sequencing and/or S-gene specific PCR identified omicron BA.1 at early time-points, which was replaced by omicron BA.2 at later time-points of the infection.

The timeline of our sample collection and symptom onset supports the proposition that individual 1 initially got infected with omicron BA.1 and exposed individual 2. During the incubation phase or early phase of infection individual 2 most likely got exposed to omicron BA.2 independently, which quickly outcompeted omicron BA.1 in his upper respiratory tract. During home isolation with individual 2, individual 1 who was already infected with omicron BA.1 got exposed to omicron BA.2, which seems to have outcompeted omicron BA.1 in the following days. These observations suggest that omicron BA.2 has biological properties allowing it to outcompete omicron BA.1 in the host, at least in the immunological and genetic context of these two individuals. Similar cases of co-infection with two different SARS-CoV-2 variants could serve as an evolutionary substrate for viral recombination events, and the emergence of new variants.

In addition, our study has several limitations. The most obvious one is the small sample size of two non-related individuals; meaning that our observations are not readily translatable for other households or larger populations. Second, we were unable to provide complete ungapped genome assembly and full genome coverage of the sample collected on day 2 from individual 2. Unfortunately, this prevented us to conduct more detailed genomic analyses to evaluate the complete distribution and representation of the omicron subvariants in that sample.

All in all, our observations although limited in nature are consistent with the epidemiological situation in several other countries, where omicron BA.2 has replaced or has been

replacing omicron BA.1. These studies along with our observations suggest that omicron BA.2 has biological features leading to host-specific growth advantage.

## DATA AVAILABILITY STATEMENT

FASTQ files are available at 10.5281/zenodo.6325649. The genome sequences were deposited in the GISAID database under accession IDs EPI_ISL_11166201, EPI_ISL_11166327, EPI_ISL_11166339, EPI_ISL_11166341, EPI_ISL_11166342, EPI_ISL_11166343.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee of ZM Clinic. The patients/participants provided their written informed consent to participate in this study.

## REFERENCES

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Cech, M., et al. (2018). The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2018 Update. *Nucleic Acids Res.* 46, W537–W544. doi:10.1093/nar/gkaa434

Liu, L., Iketani, S., Guo, Y., Chan, J. F., Wang, M., Liu, L., et al. (2022). Striking Antibody Evasion Manifested by the Omicron Variant of SARS-CoV-2. *Nature* 602, 676–681. doi:10.1038/s41586-021-04388-0

Lyngse, F. P., Kirkeby, C. T., Denwood, M., Christiansen, L. E., Mølbak, K., Møller, C. H., et al. 2022. Transmission of SARS-CoV-2 Omicron VOC Subvariants BA.1 and BA.2: Evidence from Danish Households. medRxiv [Preprint]. Available at: doi:10.1101/2022.01.28.22270044(Accesed January 30, 2022)

Otto, S. P., Day, T., Arino, J., Colijn, C., Dushoff, J., Li, M., et al. (2021). The Origins and Potential Future of SARS-CoV-2 Variants of Concern in the Evolving COVID-19 Pandemic. *Curr. Biol.* 31, R918–R929. doi:10.1016/j.cub.2021.06.049

Robinson, J. T., Thorvaldsdottir, H., Wenger, A. M., Zehir, A., and Mesirov, J. P. (2017). Variant Review with the Integrative Genomics Viewer. *Cancer Res.* 77, e31–e34. doi:10.1158/0008-5472.CAN-17-0337

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* 382, 727–733. doi:10.1056/nejmoa2001017

## AUTHOR CONTRIBUTIONS

GK and MG conceived and designed this case study. MG conducted the sequencing data and bioinformatic analysis. SM, AS, SP-D, AD, ID, SI, TP, AS, HA, JA, MN, and SV contributed to routine BSL2 and PCR work. IK and ZM contributed to the recruitment of patients in the hospital and contributed intellectually. GK wrote the manuscript. All authors contributed to the improvement of the manuscript and read the final version of the manuscript.

## ACKNOWLEDGMENTS

We thank the participants of this study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.892682/full#supplementary-material

**Conflict of Interest:** AS, IK, and GK were employed by the company Bio Engineering LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

OPINION
published: 10 May 2022
doi: 10.3389/fgene.2022.909117

# African Genetic Representation in the Context of SARS-CoV-2 Infection and COVID-19 Severity

*Desiree C. Petersen[1*†], Chrystal Steyl[1†], Denise Scholtz[1], Bienyameen Baker[1], Ibtisam Abdullah[2], Caitlin Uren[1,3], Marlo Möller[1,3] and for The COVID-19 Host Genetics Project*

[1]DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research, South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, [2]Division of Haematological Pathology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University and NHLS Tygerberg Hospital, Cape Town, South Africa, [3]Centre for Bioinformatics and Computational Biology, Stellenbosch University, Stellenbosch, South Africa

**Keywords: African genomics, genetic susceptibility, SARS-CoV-2 infection variability, COVID-19 severity, COVID-19 genetic associations, limited African data**

OPEN ACCESS

***Edited by:***
*Zhongshan Cheng, St. Jude Children's Research Hospital, United States*

***Reviewed by:***
*Prashanth N Suravajhala, Amrita Vishwa Vidyapeetham University, India*
*Juan Carlos Fernandez-Lopez, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico*
*Stephen Rawlings, Maine Health, United States*

***Correspondence:**
*Desiree C. Petersen*
*dp3@sun.ac.za*

*†These authors share first authorship*

***Specialty section:***
*This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics*

***Received:*** *31 March 2022*
***Accepted:*** *22 April 2022*
***Published:*** *10 May 2022*

***Citation:***
*Petersen DC, Steyl C, Scholtz D, Baker B, Abdullah I, Uren C and Möller M (2022) African Genetic Representation in the Context of SARS-CoV-2 Infection and COVID-19 Severity. Front. Genet. 13:909117. doi: 10.3389/fgene.2022.909117*

## INTRODUCTION

Towards the end of 2019, the world faced the emergence of the Coronavirus Disease 2019 (COVID-19) pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Up to 28 March 2022, SARS-CoV-2 resulted in over 480 million infections and has been the cause of death in approximately 6.1 million individuals (World Health Organization, 2022a). South Africa has not remained unscathed by the pandemic, having more than 3.7 million COVID-19 cases, and nearing close to 100,000 COVID-19 related deaths (World Health Organization, 2022b) with the introduction of different SARS-CoV-2 variants at various timepoints (**Figure 1A**). The risks of overwhelmed healthcare systems and an increasing mortality rate have urged for a large amount of research devoted to this disease since much remains unknown (Else, 2020). One of the knowledge gaps is the significant inter-individual variability of host responses demonstrated among SARS-CoV-2 infected individuals. This variability ranges from asymptomatic carriers to individuals who develop severe and, in some cases, lethal COVID-19. Although it has been shown that individuals older than 55 years and those with underlying comorbidities are at higher risk of severe disease, it does not explain the full extent of the variability (Meyts et al., 2020; Zhang et al., 2020; Zhou et al., 2020). A small percentage of younger and relatively healthy individuals also appear unable to control SARS-CoV-2 infection and require medical intervention (van der Made et al., 2020; Grolmusz et al., 2021). Therefore, in addition to considering the pathophysiology, transmissibility and disease severity caused by different SARS-CoV-2 variants, host genetic factors have been proposed as a possible explanation for this residual inter-individual variability (Meyts et al., 2020; Zhang et al., 2020; Guilger-Casagrande et al., 2021; Triggle et al., 2021; Velavan et al., 2021). Human genetic studies to date, mainly performed in Eurasian populations, have identified genetic variants associated with severe COVID-19. As is the case with most disease-associated human genetic studies, many first world countries have been at the forefront of publishing on the COVID-19 topic (Ellinghaus et al., 2020; van der Made et al., 2020; Zhang et al., 2020; Pairo-Castineira et al., 2021; Velavan et al., 2021). This is likely attributed to the availability of large existing biobanks making rapid COVID-19 human genetic research possible (COVID-19 Host Genetics Initiative, 2020; Zhang et al., 2020; Pairo-Castineira et al., 2021; Kousathanas et al., 2022). South Africa and the rest of the African continent has, however, not been able to contribute human genetic data at the same pace resulting in limited information being available for

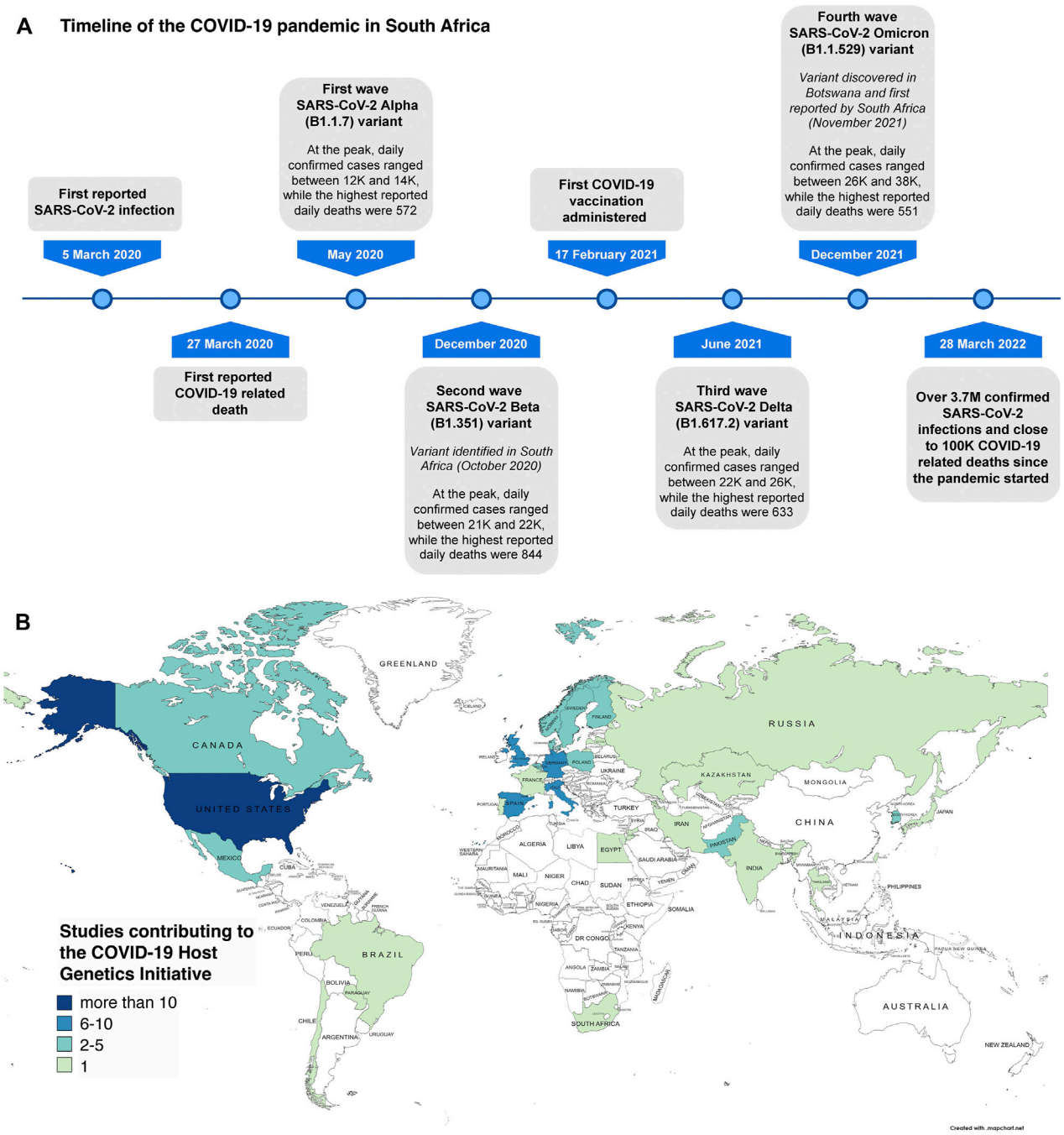



**FIGURE 1 |** SARS-CoV-2 variant timepoints in South Africa and limited human genetic studies in Africa. **(A)** indicates the significant timepoints where SARS-CoV-2 variants emerged and how this shaped the direction of the COVID-19 pandemic in South Africa (created using data from https://covid19.who.int/region/afro/country/za). **(B)** shows the partners of the COVID-19 Host Genetics Initiative (adapted from https://www.covid19hg.org/partners/). Our own research project entitled, "Host genetic factors contributing to susceptibility to COVID-19 in South Africans" is a registered study with this international effort.

local African populations. Furthermore, African populations show the greatest genetic diversity and extrapolating the results obtained from Eurasian population studies might prove to be irrelevant or may result in the exclusion of significant genetic variants when establishing COVID-19 genetic risk profiles for these understudied populations (Martin et al., 2018).

## HUMAN GENETIC STUDIES OF SARS-COV-2 INFECTION AND COVID-19 SEVERITY

Publications focusing on the role of host genetic factors in determining susceptibility to SARS-CoV-2 infection and

COVID-19 severity have included epigenetic, mitochondrial, candidate gene, and genome-wide association studies (GWAS) as well as the use of whole exome sequencing and whole genome sequencing (WGS) (Ellinghaus et al., 2020; Zhang et al., 2020; Chlamydas et al., 2021; Kgatle et al., 2021; Scozzi et al., 2021; Sen et al., 2021; Velavan et al., 2021; Wu et al., 2021). One of the earlier human genetic studies included a GWAS of SARS-CoV-2 respiratory failure, which identified associations with the ABO blood locus and a chromosome 3 gene cluster (*SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6* and *XCR1*) in Italian and Spanish populations (Ellinghaus et al., 2020). Another study of 659 hospitalized COVID-19 patients identified rare and likely pathogenic genetic variants at 13 loci, known to influence immunity to the influenza virus, associated with life-threatening COVID-19 pneumonia (Zhang et al., 2020). Several studies have since confirmed these genetic associations and have identified additional variants in *MUC5B, OAS3, OAS1, TLR7* and *TYK2*, which are associated with critical illness and severity (Pairo-Castineira et al., 2021; Velavan et al., 2021). The largest study to date consists of 49,562 COVID-19 cases representing 19 countries. This includes findings from three GWAS meta-analyses, performed by the COVID-19 Host Genetics Initiative (HGI), showing 13 significant genetic loci to be associated with either susceptibility to SARS-CoV-2 infection or severe outcomes of COVID-19 (COVID-19 Host Genetics Initiative, 2021). A large United Kingdom case-control cohort used WGS and identified novel variants in 16 genes that are associated with critical COVID-19 (Kousathanas et al., 2022). Many of the genes associated with COVID-19 in the above-mentioned studies, are implicated in fundamental pathophysiological processes, with the majority affecting immune response pathways (Ellinghaus et al., 2020; Zhang et al., 2020; COVID-19 Host Genetics Initiative, 2021; Pairo-Castineira et al., 2021; Velavan et al., 2021; Kousathanas et al., 2022).

Previous host genetic research has indicated that SARS-CoV-2 susceptibility and COVID-19 severity seem to be polygenic. It has therefore been proposed that calculating polygenic risk scores (PRS) could be useful as it allows for the detection of individuals at high risk. (Grolmusz et al., 2021; Velavan et al., 2021). A study by Prakrithi *et al.* calculated the PRS of previously identified COVID-19 associated single nucleotide polymorphisms (SNPs) in different Indian sub-populations, which allowed them to identify populations at higher risk of COVID-19-related deaths, and thereby provide support for vaccination prioritization in those specific populations (Prakrithi et al., 2021). In addition, a group from Australia designed a model to predict an individual's COVID-19 severity risk, showing that including genetic and clinical risk factors, as opposed to only using age and sex, increases the accuracy for risk discrimination by 111%. (Dite et al., 2021). Predictive scores and models do have some limitations, including being population specific as the allele frequencies and linkage disequilibrium (LD) patterns used in these calculations could differ significantly across and within populations (Sirugo et al., 2019; Grolmusz et al., 2021; Prakrithi et al., 2021).

## LACK OF COVID-19 GENETIC STUDIES IN AFRICA–WHY THIS IS AN AREA OF CONCERN

In contrast to the rest of the world, people living in Sub-Saharan Africa appear to be less prone to develop severe COVID-19 (Adams et al., 2021). This was surprising as the risk of developing severe COVID-19 was predicted to be elevated in Africa due to the high incidence of other infectious diseases such as HIV/AIDS and tuberculosis (TB), as well as the increased prevalence of non-communicable diseases such as hypertension and type 2 diabetes mellitus in certain African countries, including South Africa (Dave et al., 2021; Jassat et al., 2021). Several main hypotheses, including Sub-Saharan Africa's demographic distribution relating to age and sex; the lack of SARS-CoV-2 testing; the shortage of long-term care facilities that pose a higher risk for transmitting infectious and communicable diseases; existing protection due to previous exposure to locally circulating coronaviruses; and effective public health response supported by African governments, may have resulted in reduced morbidity and mortality rates (Adams et al., 2021). It is also possible that certain diseases or other prior infections may have an unexpected protective effect against severe COVID-19, as has been shown in the case of malaria (Altable and de la Serna, 2021; Osei et al., 2022), however, variable COVID-19 severity could also be explained by genetic differences present in these African populations (Adams et al., 2021).

Populations in Africa are highly diverse and represent some of the oldest extant populations e.g., the Khoe-San (Schuster et al., 2010; Pickrell et al., 2012; Petersen et al., 2013; Uren et al., 2016; Uren et al., 2017). In addition, modern migration routes have allowed for admixture between previously geographically distinct populations and have led to highly genetically heterogeneous populations where in some cases, there are five contributing ancestral populations (Campbell and Tishkoff, 2010; de Wit et al., 2010; Patterson et al., 2010; Petersen et al., 2013; Uren et al., 2016; Uren et al., 2017). Furthermore, African genomes have novel characteristics i.e., a larger number of novel variants and shorter more heterogeneous LD (Sirugo et al., 2019; Vergara-Lope et al., 2019). In addition, best practices as implemented by standard data analysis pipelines lack efficiency and accuracy in African populations (Uren et al., 2020). This diversity and unique genomic characteristics have phenotypic implications resulting from unique genetic factors influencing both simple and complex phenotypes such as altered disease susceptibility (Campbell and Tishkoff, 2010; Patterson et al., 2010; Uren et al., 2017). To date, however, the majority of human genetic data generated, particularly those investigating genotype-phenotype correlations, has been biased towards Eurasian populations, as has also been noted for COVID-19 research (Sirugo et al., 2019; Ellinghaus et al., 2020; van der Made et al., 2020; Zhang et al., 2020; Velavan et al., 2021; Pairo-Castineira et al., 2021). International consortiums, including the COVID-19 HGI, the COVID-19 Human Genetics effort (HGE), and the Genetics Of Mortality In Critical Care (GenOMICC) have promoted the sharing of data to facilitate the inclusion of large study cohorts from multiple populations for ongoing meta-analyses

(COVID-19 Host Genetics Initiative, 2020; COVID Human Genetic Effort, 2022; GenOMICC, 2022). Although the existing international consortiums and several research groups aim to bridge this gap, much more is needed. Only one of the 119 partner studies that contribute to the COVID-19 HGI (**Figure 1B**) (COVID-19 Host Genetics Initiative, 2022) and two of the 276 centers that contribute to the COVID HGE (COVID Human Genetic Effort, 2022) currently include populations from Sub-Saharan Africa. By considering the unique aspects of African genomes, and preliminary findings suggesting novel COVID-19 susceptibility markers in African populations, a larger emphasis needs to be placed on generating and analyzing genetic data that is representative of Africa. There is still a lack of suitable genomic references and statistical tools for interpretation of African genetic data since most of the existing references and other related tools are based on Eurasian populations (Martin et al., 2018; Sirugo et al., 2019). Results from African-based COVID-19 host genetic studies will not only benefit the populations in which they occur, but rather all populations with African ancestral contributions.

## DISCUSSION

Although viral genome sequencing and the rapid discovery of new viral variants were exceptionally successful in South Africa (Tegally et al., 2021; Wilkinson et al., 2021; Viana et al., 2022), the same is not completely true for the host genome sequencing, even though the expertise and infrastructure for human WGS is available (Glanzmann et al., 2021). The relatively cheaper viral sequencing compared to human WGS, and the complexity of human genome data compared to the viral genome data are partly the reason, together with unique ethical considerations for human genetic research in Africa, difficulty in obtaining written informed consent, and the challenges faced with sample collection (Martin et al., 2018; Marshall et al., 2022). Unfortunately, genetic research on the African continent continues to be hindered by limited resources, including delays to obtain ethical approval for human genetic studies and inadequate infrastructure (Martin et al., 2018; Hamdi et al., 2021). Despite the substantial investment made by both local and international funding organizations for human genetic studies in the more recent years (H3Africa Working Group, 2011; Choudhury et al., 2020; Maxmen, 2020), this alone is insufficient to generate large-scale genotyping and sequencing data in Africa. This is partially due to higher technology and reagent costs compared to most first world countries as well as additional expenses for bioinformatics processing and data storage hardware (Hamdi et al., 2021; Mboowa et al., 2021).

In addition to forming international collaborations, Africa needs to establish large biobanks, including the collection of phenotype information and the option to recontact participants for additional informed consent, in the case of broad consent not being favored in certain countries (Moodley and Singh, 2016; Tindana et al., 2019). This will ensure that the continent also swiftly contributes to human genetic studies in the case of future pandemics. Genetic findings for African populations may provide significant insights into the disease pathogenesis, which could lead to developing suitable therapeutic interventions that could assist with the management of COVID-19 in many resource-poor countries. This includes prioritized vaccination of genetically at-risk individuals to avoid unfavorable COVID-19 outcomes. At this stage, due to the many above-mentioned shortcomings, Africa continues to remain behind in matching the host genetic research efforts made by international collaborators on a global scale.

## COVID-19 HOST GENETICS PROJECT

Marlo Möller, Desiree C. Petersen, Craig J. Kinnear, Caitlin Uren, Brigitte Glanzmann, Elouise Kroon, Richard Glashoff, Shane Murray, Judith Hornby Cuff, Hendrik La Grange, Natrisha Damons, Helena Cornelissen, Zivanai Cuthbert Chapanduka, Ibtisam Abdullah, Deepthi Raju Abraham, Helena Rabie, Chrystal Steyl, Denise Scholtz, Annecke Vermeulen, Tongai Maponga, Kate Webb, Sian Hemmings, Gert van Zyl, Aubrey Shoko, Ansie Wichers and Sihaam Boolay from Stellenbosch University (Divisions of Molecular Biology and Human Genetics, Haematological Pathology, Medical Microbiology and Immunology, Rheumatology, and Medical Virology, Departments of Paediatrics and Child Health and Psychiatry), the University of Cape Town (Department of Paediatrics and Child Health), the Centre for Proteomic and Genomic Research (CPGR), Artisan Biomed, the South African Medical Research Council Centre for Tuberculosis Research, the South African Medical Research Council Genomics Centre and the National Health Laboratory Service (NHLS).

## AUTHOR CONTRIBUTIONS

DCP, CS and MM were responsible for the manuscript idea. CS was responsible for the general manuscript outline. DS, CS and DCP were responsible for the figure. All authors actively participated in the manuscript writing and reviewing. DCP and MM were responsible for manuscript editing. All authors read and approved the submitted manuscript.

## ACKNOWLEDGMENTS

We acknowledge the support of the DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research, South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. The COVID-19 Host Genetics Project is further supported by Distributed Platform in Omics (DIPLOMICS), Technology Innovation Agency (TIA) through the CPGR, Artisan Biomed, a South African Medical Research Council Self-Initiated Research Grant and the Michael and Susan Dell Foundation through the South African Medical Research Council Genomics Centre.

## REFERENCES


Adams, J., MacKenzie, M. J., Amegah, A. K., Ezeh, A., Gadanya, M. A., Omigbodun, A., et al. (2021). The Conundrum of Low COVID-19 Mortality Burden in Sub-Saharan Africa: Myth or Reality? *Glob. Health Sci. Pract.* 9, 433–443. doi:10.9745/GHSP-D-21-00172

Altable, M., and de la Serna, J. M. (2021). Protection against COVID-19 in African Population: Immunology, Genetics, and Malaria Clues for Therapeutic Targets. *Virus Res.* 299, 198347. doi:10.1016/j.virusres.2021.198347

Campbell, M. C., and Tishkoff, S. A. (2010). The Evolution of Human Genetic and Phenotypic Variation in Africa. *Curr. Biol.* 20, R166–R173. doi:10.1016/j.cub.2009.11.050

Chlamydas, S., Papavassiliou, A. G., and Piperi, C. (2021). Epigenetic Mechanisms Regulating COVID-19 Infection. *Epigenetics* 16, 263–270. doi:10.1080/15592294.2020.1796896

Choudhury, A., Aron, S., Botigué, L. R., Sengupta, D., Botha, G., Bensellak, T., et al. (2020). High-depth African Genomes Inform Human Migration and Health. *Nature* 586, 741–748. doi:10.1038/s41586-020-2859-7

COVID Human Genetic Effort (2022). Covid Human Genetic Effort. Available at: https://www.covidhge.com (Accessed March 29, 2022).

COVID-19 Host Genetics Initiative (2020). The COVID-19 Host Genetics Initiative, a Global Initiative to Elucidate the Role of Host Genetic Factors in Susceptibility and Severity of the SARS-CoV-2 Virus Pandemic. *Eur. J. Hum. Genet.* 28, 715–718. doi:10.1038/s41431-020-0636-6

COVID-19 Host Genetics Initiative (2021). Mapping the Human Genetic Architecture of COVID-19. *Nature* 600, 472–477. doi:10.1038/s41586-021-03767-x

COVID-19 Host Genetics Initiative (2022). COVID-19 Host Genetics Initiative Partners. Available at: https://www.covid19hg.org/partners/ (Accessed March 29, 2022).

Dave, J. A., Tamuhla, T., Tiffin, N., Levitt, N. S., Ross, I. L., Toet, W., et al. (2021). Risk Factors for COVID-19 Hospitalisation and Death in People Living with Diabetes: A Virtual Cohort Study from the Western Cape Province, South Africa. *Diabetes Res. Clin. Pract.* 177, 108925. doi:10.1016/j.diabres.2021.108925

de Wit, E., Delport, W., Rugamika, C. E., Meintjes, A., Möller, M., van Helden, P. D., et al. (2010). Genome-wide Analysis of the Structure of the South African Coloured Population in the Western Cape. *Hum. Genet.* 128, 145–153. doi:10.1007/s00439-010-0836-1

Dite, G. S., Murphy, N. M., and Allman, R. (2021). An Integrated Clinical and Genetic Model for Predicting Risk of Severe COVID-19: A Population-Based Case-Control Study. *PLoS One* 16, e0247205. doi:10.1371/journal.pone.0247205

Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., et al. (2020). Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N. Engl. J. Med.* 383, 1522–1534. doi:10.1056/NEJMoa2020283

Else, H. (2020). How a Torrent of COVID Science Changed Research Publishing - In Seven Charts. *Nature* 588, 553. doi:10.1038/d41586-020-03564-y

GenOMICC (2022). Data Releases from GenOMICC. Available at: https://genomicc.org/data/ (Accessed March 29, 2022).

Glanzmann, B., Jooste, T., Ghoor, S., Gordon, R., Mia, R., Mao, J., et al. (2021). Human Whole Genome Sequencing in South Africa. *Sci. Rep.* 11, 606. doi:10.1038/s41598-020-79794-x

Grolmusz, V. K., Bozsik, A., Papp, J., and Patócs, A. (2021). Germline Genetic Variants of Viral Entry and Innate Immunity May Influence Susceptibility to SARS-CoV-2 Infection: Toward a Polygenic Risk Score for Risk Stratification. *Front. Immunol.* 12, 653489. doi:10.3389/fimmu.2021.653489

Guilger-Casagrande, M., de Barros, C. T., Antunes, V. A. N., de Araujo, D. R., and Lima, R. (2021). Perspectives and Challenges in the Fight against COVID-19: The Role of Genetic Variability. *Front. Cell. Infect. Microbiol.* 11, 598875. doi:10.3389/fcimb.2021.598875

H3Africa Working Group (2011). Harnessing Genomic Technologies toward Improving Health in Africa: Opportunities and Challenges. Available at: https://h3africa.org/wp-content/uploads/2018/05/h3africa_whitepaper.pdf (Accessed March 29, 2022).

Hamdi, Y., Zass, L., Othman, H., Radouani, F., Allali, I., Hanachi, M., et al. (2021). Human OMICs and Computational Biology Research in Africa: Current Challenges and Prospects. *OMICS A J. Integr. Biol.* 25, 213–233. doi:10.1089/omi.2021.0004

Jassat, W., Cohen, C., Tempia, S., Masha, M., Goldstein, S., Kufa, T., et al. (2021). Risk Factors for COVID-19-Related In-Hospital Mortality in a High HIV and Tuberculosis Prevalence Setting in South Africa: A Cohort Study. *Lancet HIV* 8, e554–e567. doi:10.1016/S2352-3018(21)00151-X

Kgatle, M. M., Lawal, I. O., Mashabela, G., Boshomane, T. M. G., Koatale, P. C., Mahasha, P. W., et al. (2021). COVID-19 is a Multi-Organ Aggressor: Epigenetic and Clinical Marks. *Front. Immunol.* 12, 752380. doi:10.3389/fimmu.2021.752380

Kousathanas, A., Pairo-Castineira, E., Rawlik, K., Stuckey, A., Odhams, C. A., Walker, S., et al. (2022). Whole Genome Sequencing Reveals Host Factors Underlying Critical Covid-19. *Nature*. doi:10.1038/s41586-022-04576-6

Marshall, P., Royal, C. D. M., and Chadwick, R. (2022). Translational Science, DNA Commercialization, and Informed Consent: The Need for Specific Terminology, Insights from a Review of H3Africa Projects. *Public Health Genomics*, 1–8. doi:10.1159/000521371

Martin, A. R., Teferra, S., Möller, M., Hoal, E. G., and Daly, M. J. (2018). The Critical Needs and Challenges for Genetic Architecture Studies in Africa. *Curr. Opin. Genet. Dev.* 53, 113–120. doi:10.1016/j.gde.2018.08.005

Maxmen, A. (2020). The Next Chapter for African Genomics. *Nature* 578, 350–354. doi:10.1038/d41586-020-00454-1

Mboowa, G., Sserwadda, I., and Aruhomukama, D. (2021). Genomics and Bioinformatics Capacity in Africa: No Continent is Left behind. *Genome* 64, 503–513. doi:10.1139/gen-2020-0013

Meyts, I., Bucciol, G., Quinti, I., Neven, B., Fischer, A., Seoane, E., et al. (2020). Coronavirus Disease 2019 in Patients with Inborn Errors of Immunity: An International Study. *J. Allergy Clin. Immunol.* 147, 520–531. doi:10.1016/j.jaci.2020.09.010

Moodley, K., and Singh, S. (2016). "It's All about Trust": Reflections of Researchers on the Complexity and Controversy Surrounding Biobanking in South Africa. *BMC Med. Ethics* 17, 57. doi:10.1186/s12910-016-0140-2

Osei, S. A., Biney, R. P., Anning, A. S., Nortey, L. N., and Ghartey-Kwansah, G. (2022). Low Incidence of COVID-19 Case Severity and Mortality in Africa; Could Malaria Co-infection Provide the Missing Link? *BMC Infect. Dis.* 22, 78. doi:10.1186/s12879-022-07064-4

Pairo-Castineira, E., Clohisey, S., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., et al. (2021). Genetic Mechanisms of Critical Illness in COVID-19. *Nature* 591, 92–98. doi:10.1038/s41586-020-03065-y

Patterson, N., Petersen, D. C., van der Ross, R. E., Sudoyo, H., Glashoff, R. H., Marzuki, S., et al. (2010). Genetic Structure of a Unique Admixed Population: Implications for Medical Research. *Hum. Mol. Genet.* 19, 411–419. doi:10.1093/hmg/ddp505

Petersen, D. C., Libiger, O., Tindall, E. A., Hardie, R.-A., Hannick, L. I., Glashoff, R. H., et al. (2013). Complex Patterns of Genomic Admixture within Southern Africa. *PLoS Genet.* 9, e1003309. doi:10.1371/journal.pgen.1003309

Pickrell, J. K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., et al. (2012). The Genetic Prehistory of Southern Africa. *Nat. Commun.* 3, 1143. doi:10.1038/ncomms2140

Prakrithi, P., Lakra, P., Sundar, D., Kapoor, M., Mukerji, M., Gupta, I., et al. (2021). Genetic Risk Prediction of COVID-19 Susceptibility and Severity in the Indian Population. *Front. Genet.* 12, 714185. doi:10.3389/fgene.2021.714185

Schuster, S. C., Miller, W., Ratan, A., Tomsho, L. P., Giardine, B., Kasson, L. R., et al. (2010). Complete Khoisan and Bantu Genomes from Southern Africa. *Nature* 463, 943–947. doi:10.1038/nature08795

Scozzi, D., Cano, M., Ma, L., Zhou, D., Zhu, J. H., O'Halloran, J. A., et al. (2021). Circulating Mitochondrial DNA is an Early Indicator of Severe Illness and Mortality from COVID-19. *JCI Insight* 6, e143299. doi:10.1172/jci.insight.143299

Sen, R., Garbati, M., Bryant, K., and Lu, Y. (2021). Epigenetic Mechanisms Influencing COVID-19. *Genome* 64, 372–385. doi:10.1139/gen-2020-0135

Sirugo, G., Williams, S. M., and Tishkoff, S. A. (2019). The Missing Diversity in Human Genetic Studies. *Cell* 177, 26–31. doi:10.1016/j.cell.2019.02.048

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., et al. (2021). Detection of a SARS-CoV-2 Variant of Concern in South Africa. *Nature* 592, 438–443. doi:10.1038/s41586-021-03402-9

Tindana, P., Molyneux, S., Bull, S., and Parker, M. (2019). 'It is an Entrustment': Broad Consent for Genomic Research and Biobanks in Sub-Saharan Africa. *Dev. World Bioeth.* 19, 9–17. doi:10.1111/dewb.12178

Triggle, C. R., Bansal, D., Ding, H., Islam, M. M., Farag, E. A. B. A., Hadi, H. A., et al. (2021). A Comprehensive Review of Viral Characteristics, Transmission, Pathophysiology, Immune Response, and Management of SARS-CoV-2 and COVID-19 as a Basis for Controlling the Pandemic. *Front. Immunol.* 12, 631139. doi:10.3389/fimmu.2021.631139

Uren, C., Kim, M., Martin, A. R., Bobo, D., Gignoux, C. R., van Helden, P. D., et al. (2016). Fine-scale Human Population Structure in Southern Africa Reflects Ecogeographic Boundaries. *Genetics* 204, 303–314. doi:10.1534/genetics.116.187369

Uren, C., Möller, M., van Helden, P. D., Henn, B. M., and Hoal, E. G. (2017). Population Structure and Infectious Disease Risk in Southern Africa. *Mol. Genet. Genomics* 292, 499–509. doi:10.1007/s00438-017-1296-2

Uren, C., Hoal, E. G., and Möller, M. (2020). Putting RFMix and ADMIXTURE to the Test in a Complex Admixed Population. *BMC Genet.* 21. doi:10.1186/s12863-020-00845-3

van der Made, C. I., Simons, A., Schuurs-Hoeijmakers, J., van den Heuvel, G., Mantere, T., Kersten, S., et al. (2020). Presence of Genetic Variants Among Young Men with Severe COVID-19. *JAMA* 324, 663–673. doi:10.1001/jama.2020.13719

Velavan, T. P., Pallerla, S. R., Rüter, J., Augustin, Y., Kremsner, P. G., Krishna, S., et al. (2021). Host Genetic Factors Determining COVID-19 Susceptibility and Severity. *eBioMedicine* 72, 103629. doi:10.1016/j.ebiom.2021.103629

Vergara-Lope, A., Jabalameli, M. R., Horscroft, C., Ennis, S., Collins, A., and Pengelly, R. J. (2019). Linkage Disequilibrium Maps for European and African Populations Constructed from Whole Genome Sequence Data. *Sci. Data* 6, 208. doi:10.1038/s41597-019-0227-y

Viana, R., Moyo, S., Amoako, D. G., Tegally, H., Scheepers, C., Althaus, C. L., et al. (2022). Rapid Epidemic Expansion of the SARS-CoV-2 Omicron Variant in Southern Africa. *Nature* 603, 679–686. doi:10.1038/s41586-022-04411-y

Wilkinson, E., Giovanetti, M., Tegally, H., San, J. E., Lessells, R., Cuadros, D., et al. (2021). A Year of Genomic Surveillance Reveals How the SARS-CoV-2 Pandemic Unfolded in Africa. *Science* 374, 423–431. doi:10.1126/science.abj4336

World Health Organization (2022a). South Africa: WHO Coronavirus Disease (COVID-19) Dashboard with Vaccination Data. Available at: https://covid19.who.int/region/afro/country/za (Accessed March 28, 2022).

World Health Organization (2022b). WHO Coronavirus (COVID-19) Dashboard. Available at: https://covid19.who.int (Accessed March 28, 2022).

Wu, Y., Wang, X.-H., Li, X.-H., Song, L.-Y., Yu, S.-L., Fang, Z.-C., et al. (2021). Common mtDNA Variations at C5178a and A249d/T6392C/G10310A Decrease the Risk of Severe COVID-19 in a Han Chinese Population from Central China. *Mil. Med. Res.* 8, 57. doi:10.1186/s40779-021-00351-2

Zhang, Q., Bastard, P., Liu, Z., Le Pen, J., Moncada-Velez, M., Chen, J., et al. (2020). Inborn Errors of Type I IFN Immunity in Patients with Life-Threatening COVID-19. *Science* 370, eabd4570. doi:10.1126/science.abd4570

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., et al. (2020). Clinical Course and Risk Factors for Mortality of Adult Inpatients with COVID-19 in Wuhan, China: A Retrospective Cohort Study. *Lancet* 395, 1054–1062. doi:10.1016/S0140-6736(20)30566-3

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 14 July 2022
doi: 10.3389/fgene.2022.911010

# Genome-Wide Association Study of COVID-19 Outcomes Reveals Novel Host Genetic Risk Loci in the Serbian Population

*Marko Zecevic[1,2], Nikola Kotur[1], Bojan Ristivojevic[1], Vladimir Gasic[1], Vesna Skodric-Trifunovic[3,4], Mihailo Stjepanovic[3,4], Goran Stevanovic[4,5], Lidija Lavadinovic[5], Branka Zukic[1], Sonja Pavlovic[1] and Biljana Stankovic[1]**

[1]*Laboratory for Molecular Biomedicine, Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Belgrade, Serbia,* [2]*Seven Bridges, Boston, MA, United States,* [3]*Clinic of Pulmonology, Clinical Centre of Serbia, Belgrade, Serbia,* [4]*Faculty of Medicine, University of Belgrade, Belgrade, Serbia,* [5]*Clinic for Infectious and Tropical Diseases, Clinical Centre of Serbia, Belgrade, Serbia*

OPEN ACCESS

**Edited by:**
*Hongsheng Gui,*
*Henry Ford Health System,*
*United States*

**Reviewed by:**
*Marlo Möller,*
*Stellenbosch University, South Africa*
*Zhongshan Cheng,*
*St. Jude Children's Research Hospital,*
*United States*
*Julong Wei,*
*Wayne State College, United States*

***Correspondence:**
*Biljana Stankovic*
*biljana.stankovic@imgge.bg.ac.rs*

**Specialty section:**
*This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics*

**Received:** *01 April 2022*
**Accepted:** *21 June 2022*
**Published:** *14 July 2022*

**Citation:**
*Zecevic M, Kotur N, Ristivojevic B, Gasic V, Skodric-Trifunovic V, Stjepanovic M, Stevanovic G, Lavadinovic L, Zukic B, Pavlovic S and Stankovic B (2022) Genome-Wide Association Study of COVID-19 Outcomes Reveals Novel Host Genetic Risk Loci in the Serbian Population. Front. Genet. 13:911010. doi: 10.3389/fgene.2022.911010*

Host genetics, an important contributor to the COVID-19 clinical susceptibility and severity, currently is the focus of multiple genome-wide association studies (GWAS) in populations affected by the pandemic. This is the first study from Serbia that performed a GWAS of COVID-19 outcomes to identify genetic risk markers of disease severity. A group of 128 hospitalized COVID-19 patients from the Serbian population was enrolled in the study. We conducted a GWAS comparing (1) patients with pneumonia ($n$ = 80) against patients without pneumonia ($n$ = 48), and (2) severe ($n$ = 34) against mild disease ($n$ = 48) patients, using a genotyping array followed by imputation of missing genotypes. We have detected a significant signal associated with COVID-19 related pneumonia at locus 13q21.33, with a peak residing upstream of the gene *KLHL1* ($p = 1.91 \times 10^{-8}$). Our study also replicated a previously reported COVID-19 risk locus at 3p21.31, identifying lead variants in *SACM1L* and *LZTFL1* genes suggestively associated with pneumonia ($p = 7.54 \times 10^{-6}$) and severe COVID-19 ($p = 6.88 \times 10^{-7}$), respectively. Suggestive association with COVID-19 pneumonia has also been observed at chromosomes 5p15.33 (*IRX, NDUFS6, MRPL36,* $p = 2.81 \times 10^{-6}$), 5q11.2 (*ESM1,* $p = 6.59 \times 10^{-6}$), and 9p23 (*TYRP1, LURAP1L*, $p = 8.69 \times 10^{-6}$). The genes located in or near the risk loci are expressed in neural or lung tissues, and have been previously associated with respiratory diseases such as asthma and COVID-19 or reported as differentially expressed in COVID-19 gene expression profiling studies. Our results revealed novel risk loci for pneumonia and severe COVID-19 disease which could contribute to a better understanding of the COVID-19 host genetics in different populations.

**Keywords: GWAS, SARS-CoV-2, genetic markers, pneumonia, severe disease**

# 1 INTRODUCTION

The waning effectiveness of vaccines and new SARS-CoV-2 variants of concern indicate that the COVID-19 health threat will likely remain in the future. The course of SARS-CoV-2 infection ranges from asymptomatic and mild to severe disease, which can progress to critical illness and a lethal outcome. Besides acute disease, patients with severe symptoms are more likely to suffer from long term COVID-19 related distress. Several risk factors for the severe form of the COVID-19 disease have been identified, namely old age, male sex, non-Caucasian ethnicity, smoking, low income, obesity, and other preexisting comorbidities (COVID-19 Host Genetics Initiative, 2021; Hu and Wang, 2021; Shelton et al., 2021). Using this data has resulted in establishing prioritization strategies in the prevention and treatment of COVID-19 employed to optimally allocate limited healthcare resources to vulnerable groups.

Besides health and demographic data, variations in the human genome have also been analyzed to find genetic markers related to susceptibility to infection and severity of COVID-19. Identifying genetic markers associated with different COVID-19 outcomes would not only advance patient stratification but also reveal potential mechanisms of disease progression, point out important pathways, and contribute to drug discovery. Using the candidate gene approach, several loci showed association with severe disease or infection rate. These include *APOE4* (Kuo et al., 2020), *ACE1*, *ACE2* and *TMPRSS2* genes (Andolfo et al., 2021; Fink-Baldauf et al., 2022), *HLA* locus (Fricke-Galindo and Falfán-Valencia, 2021), and vitamin D level influencing genes (Kotur et al., 2021). Genome sequencing analysis focused on rare genetic variants implicated type I interferon immunity in COVID-19 progression (Zhang et al., 2020).

Genome-wide association study (GWAS) is a hypothesis-free approach suitable for the discovery of novel, common genetic markers. Genome-wide analyses have been focused on different COVID-19 outcomes, such as susceptibility to infection, severe disease, critical disease, and lethal outcome, as well as hospitalization rate, post-COVID-19 syndrome and response to COVID-19 vaccines (COVID-19 Host Genetics Initiative, 2021; Pairo-Castineira et al., 2021; Wu et al., 2021; Thibord et al., 2022). The results of these studies have implicated several genome loci with a COVID-19 outcome. The most replicated result so far associated variants on chromosome 3 near *LZTFL1* and *SLC6A20* genes with infection, hospitalization, and critical illness, as well as *ABO* gene on chromosome 9 correlating non-O blood type with a higher rate of infection and severe disease (Ellinghaus et al., 2020; COVID-19 Host Genetics Initiative, 2021; Pairo-Castineira et al., 2021; Shelton et al., 2021; Thibord et al., 2022).

Efforts to find genetic variants continue to include more ethnic groups and COVID-19 outcomes. In this study, GWAS was performed on COVID-19 patients of Serbian origin in order to identify genetic markers of different COVID-19 phenotypes.

# 2 MATERIALS AND METHODS

## 2.1 Subjects and Genotyping

A total of 216 participants were recruited for the study. Participants were confirmed positive for COVID-19 according to local clinical testing (SARS-CoV-2 RNA RT-PCR) in tertiary healthcare institutions in Belgrade, Serbia, between April and June of 2020. Whole blood samples were taken from the participants and further used for DNA extraction. All subjects were genotyped using the Illumina Infinium Global Screening Array v.3.0 + Multi-Disease BeadChip (GSAMD-24v3-0-EA), a high-density array that covers over 700,000 variants. The obtained data was used to build a population-specific cluster file.

From the total study group, we selected 128 adult patients that satisfied all criteria needed for the phenotypic classification used in association analysis. Precisely, patients were divided into three groups according to the National Institutes of Health (NIH) Definition of COVID-19 Disease Severity [COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. National Institutes of Health. Available at https://www.covid19treatmentguidelines.nih.gov/. Accessed 01 September 2021]: mild—patients with COVID-19 related symptoms without pneumonia; moderate—patients with evidence of pneumonia based on imaging showing up to 50% of lung involvement and with oxygen saturation ≥94% on room air; and severe—patients with pneumonia with >50% of lung involvement on imaging or had blood oxygen saturation level <94% on room air or required supportive oxygen therapy.

Patients included in the study were not vaccinated against COVID-19 at the time of diagnosis. Informed consent was obtained from each participant or their parents/legal guardians. This study was approved by the Ethics Committee of the Institute of Molecular Genetics and Genetic Engineering University of Belgrade (approval for sample collection and biobank formation O-EO-016/2020, 05 May 2020; approval for the genetic study O-EO-016/2020/1, 03 September 2020).

## 2.2 Data Analysis

### 2.2.1 Data Preprocessing, Variant Calling and Imputation

This study analyzed samples from the Serbian population using the custom GSAMD panel which required generating a population-specific cluster file. Initial genotype calling from GSAMD intensity data (IDAT) files and quality control (QC) analysis before cluster file generation was performed using GenomeStudio v.2.0 software with GSAMD-24v3-0-EA_20034606_A1.bpm manifest file based on the human genome assembly hg19. Details regarding cluster file generation and initial QC analysis are provided in the **Supplementary Methods**.

The cluster file and lists of samples and variants that passed QC filters were further used as inputs for the GWAS analysis pipeline that we have set up. The pipeline encompasses variant calling, phasing, and imputation, as well as QC analysis (described in detail in the **Supplementary Methods**). After the

imputation and prior to the genome-wide association analysis, our dataset contained 12,001,939 variants.

### 2.2.2 Genome-Wide Association Study

The GWAS analysis pipeline was based on the GENESIS v2.10.0 (Gogarten et al., 2019) R/Bioconductor package and the various methods it implements. It has provided us with a mixed model framework that accounts for genetic relatedness and allows for the inclusion of different risk factors as covariates.

We fitted a generalized linear mixed model (GLMM) containing both fixed effects (independent covariates) and random effect which models the genetic correlation between the individuals (kinship matrix), under the null hypothesis of no genotype effects. Previous studies have identified several independent risk factors, including age and male sex (Hu and Wang, 2021). In our model formula, we have included age and sex as well as an interaction term (age*sex) between the two.

The fitted null model was then used for single variant association testing and score tests were performed for all variants with minor allele count (MAC) ≥ 10 individually. We have chosen to apply the saddle point approximation (SPA) to the score test statistic to estimate the null distribution. The two significance thresholds utilized were those of genome-wide significance ($p < 5 \times 10^{-8}$)—representing a Bonferroni-corrected 5% family-wise error rate threshold for the estimated effective number of 1000000 independent common genetic variants given the linkage disequilibrium structure of the human genome (Uffelmann et al., 2021); and suggestive association ($p < 1 \times 10^{-5}$)—a less stringent threshold intended for the identification of SNPs that should be considered in follow-up studies.

Detailed information on the genome-wide association analysis methods that we applied is provided in **Supplementary Methods**.

### 2.2.3 Post-GWAS Analysis

Bayesian fine-mapping from GWAS summary statistics of the loci reaching genome-wide significance was performed using the SuSiE method implemented in the R package susieR (Zou et al., 2021) to determine the posterior inclusion probability (PIP) for each variant being causal as well as to determine a credible set, which is the smallest set of variants that contains all the causal variants with a probability ≥0.95%.

LD clumping and variant annotation were performed using the FUMA v1.3.7 web application (Watanabe et al., 2017). Variant annotation for each locus was made on independent significant variants and all the variants in LD with them that are less than 250 kb away, and with a score test $p$-value < 0.05.

Next, we conducted an exploratory analysis of the annotated variants (independent significant variants and their LD proxies) in order to identify causal variants that have a deleterious gene effect or effects on gene expression, by employing different genome browsers (Ensembl (Howe et al., 2021), The University of California Santa Cruz, UCSC (Lee et al., 2022)) and web-based tools and databases such as those from the LDlink v5.2 suite: LDproxy and LDexpress (Machiela and Chanock, 2015; Lin et al., 2021)—linked to Regulome and GTEx databases, HaploReg v4.1 (Ward and Kellis, 2012)—linked to Roadmap Epigenomics and ENCODE projects as well as different eQTLs studies data, GeneAtlas (Canela-Xandri et al., 2018) and FUMA which uses information from 18 different repositories (Watanabe et al., 2017). For eQTL mapping, the window size was defined as 1 Mb upstream and downstream from the transcription start site.

Through the FUMA web app, by utilizing the Multi-marker Analysis of GenoMic Annotation (MAGMA) (de Leeuw et al., 2015) method, genes and gene-sets associated with COVID-19 severity were analyzed, based on the signals identified at GWAS loci. To perform functional analysis, first, gene-mapping had to be performed, selecting genes located up to 10 kb upstream or downstream from variants that are functionally annotated (i.e., having a functional consequence on gene expression) and showing a suggestive association with the disease severity ($p < 1 \times 10^{-5}$).

### 2.2.4 Comparative Population Analysis

We investigated effect allele frequencies of GWAS identified risk loci in world-wide populations. The effect allele frequencies (AF) were extracted from the 1,000 Genome Project (1kGP) including European populations (Italy, Spain, Finland, Great Britain and USA with European ancestry) as well as Eastern Asians, South Asians, African and Ad-Mixed American (Central and South American populations) (Auton et al., 2015). We examined the level of genetic variability among populations at each risk loci by looking at the maximal global differences in allele frequencies (delta AF, dAF) calculated by subtracting the minimum from the maximum effect allele frequency across analyzed population groups.

# 3 RESULTS

## 3.1 Study Group

A group of 216 participants diagnosed with COVID-19 (102 males, 114 females) was included in the genotyping part of the study in order to create a cluster file specific for the Serbian population. Of the total genotyping group, 16 subjects (8 males and 8 females) were excluded due to the low call rate (<0.95) during initial QC, leaving 200 samples for the cluster file generation.

For the association analysis of the COVID-19 severity, a subgroup of 128 hospitalized COVID-19 patients was selected and classified into three phenotypic groups: mild ($n$ = 48), moderate ($n$ = 46), and severe ($n$ = 34). Demographic and clinical data for the study group are summarized in **Table 1**. In addition, the ancestry of our study group was defined using the principal component analysis: almost all of our participants clustered with the European population (98.4%), and only 2 patients clustered between European and Ad Mixed American populations (**Supplementary Figure S1**).

In the COVID-19 study group, 12 patients (9.4%) required supportive oxygen therapy and 4 (3.1%) had COVID-19 related death outcomes. The age of patients significantly varied between the groups, being the highest in the severe group ($p < 0.0001$). Gender distribution was different among the mild, moderate, and

**TABLE 1 |** COVID-19 patients' demographic and clinical data.

| | Mild | Moderate | Severe | *p* |
|---|---|---|---|---|
| N (%) | 48 (37.5) | 46 (35.9) | 34 (26.6) | |
| Age, median [IQR] | 39.0 [29.0–49.0] | 45.5 [36.0–61.0] | 61.0 [50.0–68.5] | <0.0001 |
| Gender, male n (%) | 12 (25.0) | 24 (52.2) | 22 (64.7) | 0.0009 |
| Obesity, n/available (%) | 6/42 (14.3) | 9/36 (25.0) | 9/27 (33.3) | 0.17 |
| Diabetes, n/available (%) | 1/48 (2.1) | 4/46 (8.7) | 6/33 (18.2) | 0.03 |
| Hypertension, n/available (%) | 8/48 (16.7) | 14/46 (30.4) | 18/33 (54.5) | 0.0015 |
| ACE inhibitors, n/available (%) | 5/48 (10.4) | 7/43 (16.3) | 13/30 (43.3) | 0.002 |
| % SatO2, median [IQR] | 98 [98–99] | 98 [97–99] | 90 [85–96] | <0.0001 |
| CRP, median [IQR] | 1.1 [0.5–5.4] | 10.3 [3.0–33.3] | 128.0 [60.6–200.0] | <0.0001 |
| Febrile, n/available (%) | 15/47 (31.9) | 34/46 (73.9) | 30/33 (90.9) | <0.0001 |
| Lymphopenia, (≤1 × $10^9$/L), n/available (%) | 12/46 (26.1) | 23/46 (50.0) | 24/32 (75.0) | 0.0001 |
| Thrombocytopenia, (<150,000/mm3), n/available (%) | 8/45 (17.8) | 8/46 (17.4) | 17/32 (53.1) | 0.0005 |

*Each count was presented along with the total available number of observations for that category (n/available). Differences between the groups were tested using the Kruskal–Wallis test for continuous data, Chi-square, or Fisher exact test for discrete data. IQR - interquartile range, SatO2 - blood oxygen saturation, CRP - C-reactive protein.*

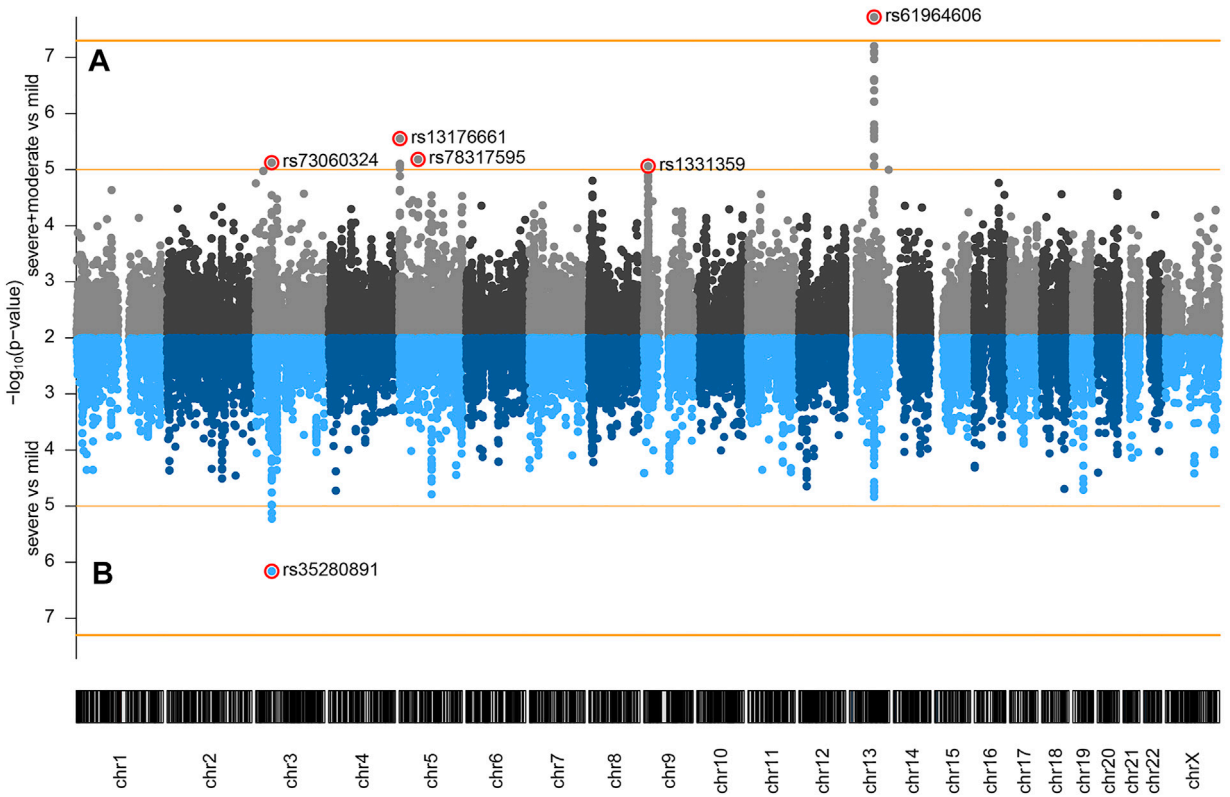



**FIGURE 1 |** Genome-wide association of COVID-19 outcomes in Serbian population: **(A)** severe and moderate ($n$ = 80) versus mild ($n$ = 48); **(B)** severe ($n$ = 34) versus mild ($n$ = 48) comparisons. Genomic coordinates of analyzed SNPs are displayed along the $x$-axis and the observed statistical significance of association with the disease outcome on the negative $\log_{10}$ scale is displayed on the $y$-axis. SNPs with $p < 0.01$ are not shown on the plot. The two horizontal orange lines represent the two significance thresholds used: genome-wide significance ($p < 5 \times 10^{-8}$, y ≈ 7.3) and suggestive association ($p < 1 \times 10^{-5}$, y = 5).

severe groups, including 25%, 52.2%, and 64.7% of male patients, respectively ($p$ = 0.0009). Patients with severe disease more frequently suffered from diabetes and hypertension ($p$ = 0.03, $p$ = 0.0015, respectively). CRP level was prominently higher in the severe group and those patients had a higher percentage of lymphopenia (75%) and thrombocytopenia (53.1%) events compared to patients with moderate and mild disease ($p$ < 0.0001, $p$ = 0.0001 and $p$ = 0.0005, respectively).

**TABLE 2 |** Lead variants: severe and moderate versus mild disease.

| Lead Variant | Cytoband | Position (hg19) | *p*-value | Nearest gene(s) | Location | Ref | Alt (effect allele) | Alt allele frequency | Effect size (beta) | OR | 95% CI | $r^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs61964606 | 13q21.33 | 70763164 | $1.91 \times 10^{-8}$ | *KLHL1, ATXN8, ATXN80S* | intergenic | A | G | 0.824 | 2.314 | 10.115 | 4.458–22.949 | 0.903 |
| rs73060324 | 3p21.31 | 45785915 | $7.54 \times 10^{-6}$ | *SACM1L* | 3'-UTR | T | G | 0.082 | −2.473 | 0.084 | 0.028–0.257 | 0.984 |
| rs13176661 | 5p15.33 | 2191105 | $2.81 \times 10^{-6}$ | *IRX4, NDUFS6, MRPL36* | intergenic | G | A | 0.504 | −1.391 | 0.249 | 0.138–0.449 | 0.974 |
| rs78317595 | 5q11.2 | 54288077 | $6.59 \times 10^{-6}$ | *ESM1* | intronic | T | C | 0.223 | −1.727 | 0.117 | 0.083–0.379 | 0.956 |
| rs1331359 | 9p23 | 12363456 | $8.69 \times 10^{-6}$ | *TYRP1, LURAP1L* | intergenic | G | A | 0.117 | −2.309 | 0.099 | 0.035–0.279 | 0.999 |

*Lead variants representing 5 genomic loci and showing at least a suggestive association ($p < 1 \times 10^{-5}$) with the disease severity when comparing patients with pneumonia (severe and moderate disease groups) versus those without pneumonia (mild disease group). Summary statistics such as p-value, alternative allele frequency in our cohort, effect size estimate for each additional copy of the alternative allele, and odds-ratio (OR) are shown as well the imputation quality metric ($r^2$). Ref – reference allele, Alt – alternative allele, CI – confidence interval.*

## 3.2 Severity Loci Identified by the Genome-Wide Analysis

To assess the genetic component of risk for different COVID-19 outcomes in the Serbian population, we have performed a genome-wide association analysis, testing for genetic variant allele frequency differences between groups of patients classified as either mild ($n$ = 48), moderate ($n$ = 46) or severe ($n$ = 34).

The following comparisons have been made: 1) severe and moderate versus mild, 2) severe versus mild, and 3) severe versus moderate and mild. Genome-wide association analysis of severe and moderate versus mild, which actually compares patients with pneumonia against patients with no pneumonia diagnosed, identified a significant association signal at locus 13q21.33 (**Figure 1A**). In the two other comparisons, we did not observe signals reaching genome-wide significance ($p < 5 \times 10^{-8}$), but several signals showed a suggestive association with disease severity ($p < 1 \times 10^{-5}$). Severe versus mild comparison gave a stronger signal at the already recognized risk locus at chromosome 3 (**Figure 1B**) than severe versus moderate and mild (**Supplementary Figure S2**), probably due to severe and moderate categories not being as distinctly separated as moderate and mild. Hence, we focused our efforts on the first two genotypic-phenotypic comparisons: testing for differences in the allele frequency of genetic variants in (1) patients with pneumonia (severe and moderate patients grouped together) versus those without pneumonia (mild disease group) and (2) in patients diagnosed with pneumonia with above 50% lung involvement or blood oxygen level below 94% (severe disease group) versus patients without pneumonia (mild disease group). In both genotypic-phenotypic comparisons, patients who had a mild disease (without pneumonia) served as a reference or control group in statistical analyses.

### 3.2.1 Genome-Wide Association of COVID-19 Related Pneumonia (Severe and Moderate Versus Mild Disease)

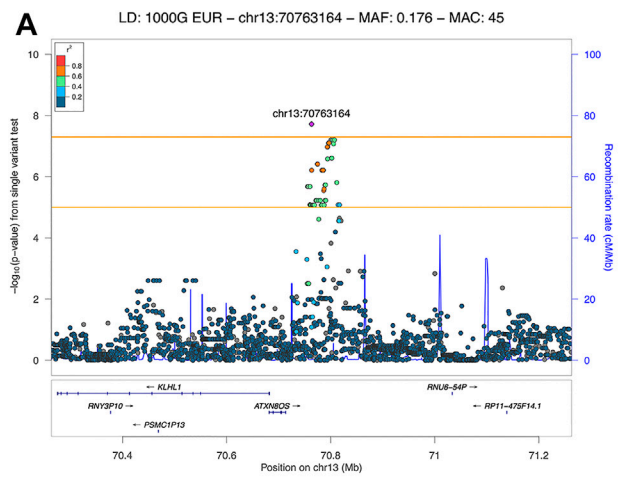
We have performed GWAS comparing 80 patients from the severe and moderate category versus 48 from the mild disease category. The total number of SNPs analyzed by fitting a logistic mixed effect model adjusted for age, sex, and genetic relatedness was 7496155. Genomic inflation factor lambda (λ) was 0.9959 and the quantile-quantile (QQ) plot is shown in **Supplementary Figure S3**.

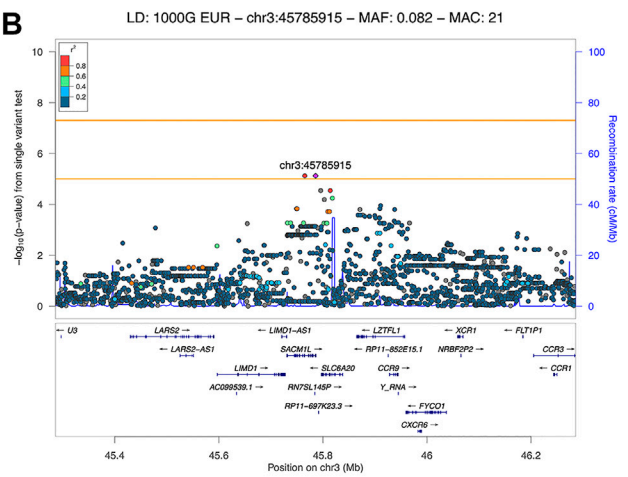
We detected 5 risk loci, namely 13q21.33, 3p21.31, 5p15.33, 5q11.2, and 9p23, showing at least a suggestive association with COVID-19 related pneumonia (**Table 2**). The imputation quality of the associated independent variants was high (allelic $r^2 > 0.82$). Regional association plots for the identified risk loci are shown in **Figure 2**. The annotated independent significant variants together with the variants that they are in linkage disequilibrium with ($r^2 > 0.6$) are presented in **Supplementary Table S2**.

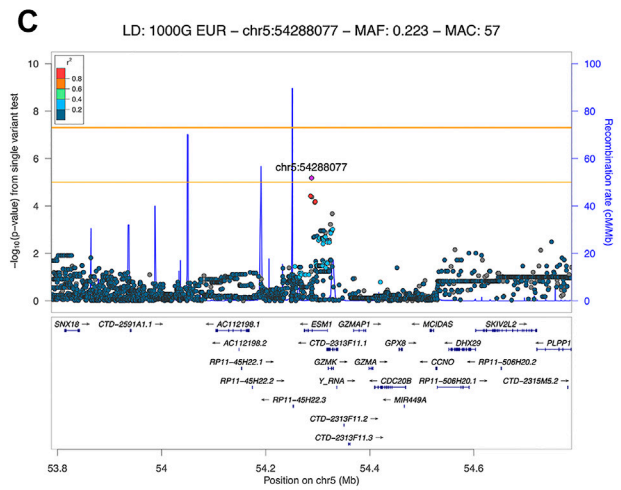
Out of the 5 loci, only 13q21.33 achieved genome-wide significance with rs61964606 as its lead variant. The major G allele of this variant conferred increased risk for COVID-19 related pneumonia (OR = 10.115, 95% CI: 4.458–22.949, $p = 1.91 \times 10^{-8}$). The rs61964606 variant is located in the intergenic region upstream of the *KLHL1* gene (~80 kb), surrounded by regulatory sequences—enhancers, transcriptional and CCCTC-binding factor (CTCF) sites. In the proximity of the lead variant, we detected transcriptional factors JUND and FOXA1 binding sites according to Open Regulatory Annotation database (ORegAnno) which is incorporated into the UCSC genome browser (Lesurf et al., 2016). Linkage disequilibrium ($0.6 > r^2 > 0.5$) was found between our lead variant and several eQTL variants associated with *KLHL1* gene expression in brain tissue (rs7339068, rs9542235, rs7339309, all three $p = 2.2 \times 10^{-7}$, based on PsychENCODE database (Wang et al., 2018)).

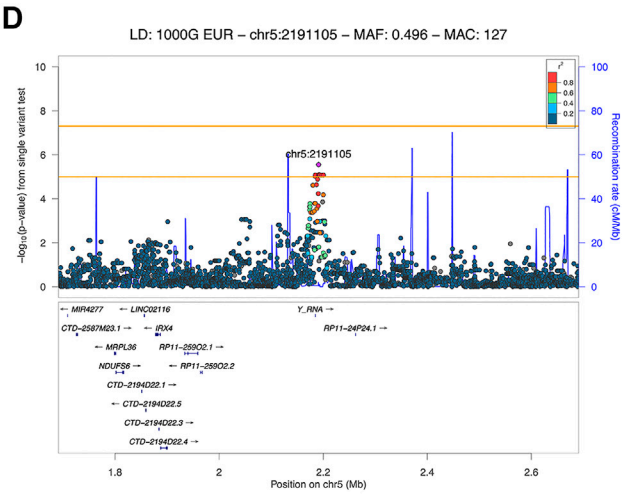
Fine-mapping from summary statistics enabled us to examine the posterior inclusion probability (PIP) for each variant on the genomic risk locus on chromosome 13 (from chr13:70755332 to chr13:70819769). The 95% credible variant set is shown in **Supplementary Table S1**.

Our genome-wide analysis detected a suggestive signal in replicated COVID-19 risk locus at chromosome 3p21.31, in the 3′-UTR of the *SACM1L* gene (OR = 0.084, 95% CI: 0.028–0.257, $p = 7.54 \times 10^{-6}$). Detected lead *SACM1L* rs73060324 variant highly correlates with rs17279437 missense variant ($r^2 = 0.88$) located in the *SLC6A20* gene, a potential causative candidate which has been previously associated with COVID-19. Predicted pathogenicity of rs17279437 was high



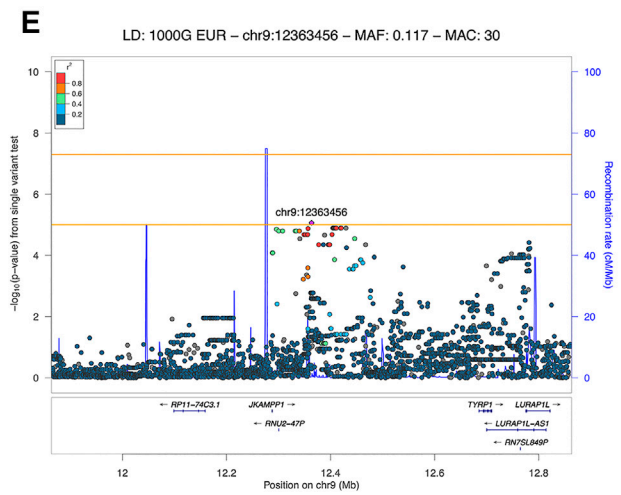




**FIGURE 2 |** LocusZoom plot showing the regional association of GWAS variant enriched loci with COVID-19 severity (severe and moderate versus mild patient group comparison). Variants in linkage disequilibrium with the lead variant are shown in color gradient indicating $r^2$ levels (hg19, 1KGP, Phase 3—November 2014, EUR): **(A)** 13q21.33 with rs61964606 as the lead variant; **(B)** 3p21.31 with rs73060324 as the lead variant; **(C)** 5p15.33 with rs13176661 as the lead variant; **(D)** 5q11.2 with rs78317595 as the lead variant; **(E)** 9p23 with rs1331359 as the lead variant.

**TABLE 3 |** Lead variants: severe versus mild COVID-19.

| Lead Variant | Cytoband | Position (hg19) | *p*-value | Nearest Gene | Location | Ref | Alt (effect allele) | Alt Allele Frequency | Effect Size (beta) | OR | 95% CI | $r^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs35280891 | 3p21.31 | 45951647 | $6.88 \times 10^{-7}$ | *LZTFL1* | intronic | G | A | 0.116 | 2.988 | 19.846 | 5.728–68.761 | 0.762 |

*The lead variant in the 3p21.31 locus showed a suggestive association ($p < 1 \times 10^{-5}$) with the disease severity when comparing patients from the severe versus those from the mild disease group. Summary statistics such as* p-value, *alternative allele frequency in our cohort, effect size estimate for each additional copy of the alternative allele, and odds-ratio (OR) are shown as well the imputation quality metric ($r^2$). Ref – reference allele, Alt –alternative allele, CI – confidence interval*

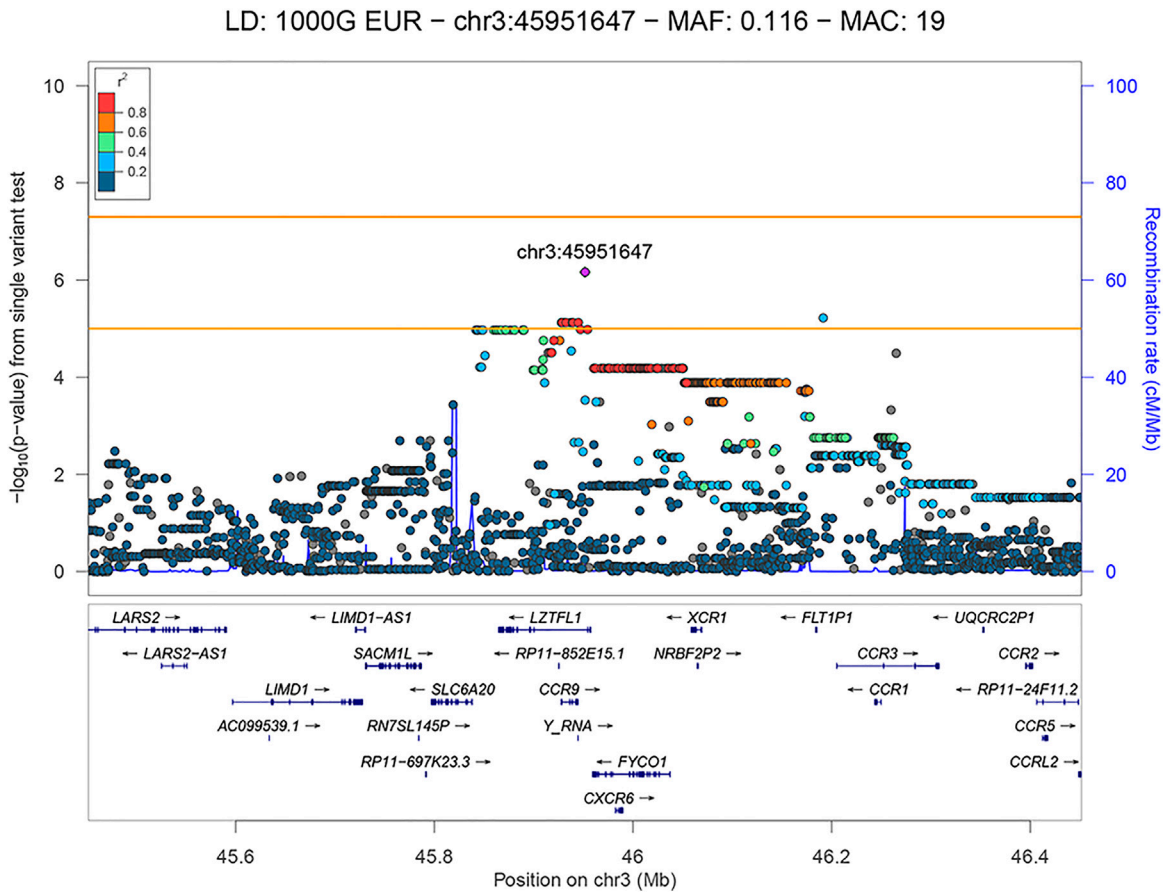



**FIGURE 3 |** LocusZoom plot showing the regional association of 3p21.31 with COVID-19 severity (severe versus mild patient group comparison), with rs35280891 as the lead variant. Variants in linkage disequilibrium with rs35280891 are shown in color gradient indicating $r^2$ levels (hg19, 1KGP, Phase 3—November 2014, EUR).

(CADD score = 25.4, Polyphen = 1). Our results showed that allele A of this variant was associated with a lower risk of pneumonia in COVID-19 patients.

Regarding other suggestive loci, the signal at 5q11.2 was located in the intron of the *ESM1* gene. Functional annotation showed that lead variant rs78317595 was in LD with eQTL rs10076939 ($r^2$ = 0.74) associated with *ESM1* gene expression in human blood vessel tissue ($p = 10^{-4}$, based on GTEx v8). The remaining two signals were found in intergenic regions: 5p15.33 was proximal to *IRX4, NDUFS6,* and *MRPL36,* while 9p23 was close to *TYRP1 and LURAP1L* genes (both at ~ 300–400 kb distance from nearest genes). Intergenic signal near *IRX4, NDUFS6,* and *MRPL36* was located in the region highly enriched with regulatory sequences. According to Roadmap Epigenomics data, the rs13172851 variant (high LD with lead variant rs13176661, $r^2$ = 0.88), lies in the enhancer histone marks (H3K4me1_Enh and H3K27ac_Enh) active in various tissues, including lungs.

### 3.2.2 Genome-Wide Association of Severe COVID-19 (Severe Versus Mild Disease)

The results of genome-wide association analysis of severe ($n = 34$) versus mild ($n = 48$) Serbian COVID-19 patients include only a single locus—3p21.31, that reached the suggestive, but not the genome-wide significance threshold ($p = 6.88 \times 10^{-7}$) (**Table 3**). The model type and the covariates chosen were the same as in the comparison that was described in the previous section. The total number of variants analyzed was 6,695,505; genomic inflation factor lambda ($\lambda$) was 0.9944 and the quantile-quantile (QQ) plot is shown in **Supplementary Figure S4**.

We have observed a total of 10 variants in the 3p21.31 locus that showed a suggestive association, all of which are shown in the regional association plot in **Figure 3**. The variants were found to be physically close together (within a ~260 kb region) and in LD with each other ($r^2 \geq 0.725$; based on the 1kGP Phase 3 data—EU super-population) with the lead variant being rs35280891 (OR = 19.846, 95% CI: 5.728–68.761, $p = 6.88 \times 10^{-7}$). The imputation quality of the associated variants was high ($r^2 > 0.5844$). The annotated independent significant variants together with the variants they are in linkage disequilibrium with ($r^2 > 0.6$) are presented in **Supplementary Table S3**.

The lead variant is in a high LD with three missense variants ($r^2 > 0.84$) located in the *FYCO1* gene: rs13079478, rs13059238 and rs33910087 (all three: OR = 11.107, 95% CI: 3.262–37.817, $p = 6.57 \times 10^{-5}$). Variants rs13079478 and rs13059238 are located at the same codon, so three missense variants affect two amino acids of the resulting polypeptide. According to GeneAtlas, *FYCO1* variant rs33910087 is a highly significant modifier of monocyte percentage ($p = 3.85 \times 10^{-46}$), and based on GTEx v8, this variant is also an eQTL for several protein-coding genes previously associated with COVID-19 (*CXCR6*, *FYCO1*, *SLC6A20*, *CCR1*, *LZTFL1*). Also, independent significant variant rs192311430, located in the intron of the *LZTFL1* gene, is associated with *CCR3* and *FLT1P1* gene expression in whole blood ($p = 3 \times 10^{-5}$ and $p = 2 \times 10^{-5}$, respectively), as well as with *CCR6* gene expression in neural tissue ($p = 2 \times 10^{-4}$), based on GTEx v8.

## 3.3 Functional Analysis

In both comparisons (severe and moderate versus mild, and severe versus mild), the gene set enrichment analysis of genes located up to 10 kb upstream or downstream from a candidate variant showed significant enrichment of chemokine related pathways (*REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES*, FDR adjusted $p$-value = $6.32 \times 10^{-10}$, 8 out of 48 genes from this gene set have been found near our candidate variants: *CCR9, CXCR6, XCR1, CCR3, CCR1, CCR2, CCR5, CCRL2*).

## 3.4 Comparative Population Analysis

Frequencies of the lead variants in the worldwide populations have been extracted from the 1kGP (**Supplementary Table S4**) and visualized along with the Serbian population in **Supplementary Figure S5**.

Examination of entries in the 1kGP database for risk loci showed that the frequency of effect alleles among worldwide populations varied highly (highest dAF, which is a difference between maximal and minimal effect allele frequency) for rs13176661, ranging from 91% in African to 56% in European populations, being the lowest in Serbian population 50%, and rs1331359, ranging from 95% in African to 46% in East Asian populations.

## 3.5 Validation of Lead Variants in Publicly Available COVID-19 Genome-Wide Association Studies

Since another cohort of COVID-19 patients from the Serbian population was not available, we relied on publicly available GWAS datasets of COVID-19 patients for validation of obtained associations. In order to validate lead variants as novel risk markers of COVID-19 severity, we employed UK biobank data of COVID-19 positive patients of European origin. Two phenotypes similar to the ones we analyzed were selected: severe versus non-severe COVID-19 positive and hospitalized positive versus non-hospitalized positive (https://grasp.nhlbi.nih.gov/Covid19GWASResults.aspx) (Thibord et al., 2022). We considered an association validated if at least a nominally significant result ($p < 0.05$) in the UK biobank was noted in the same direction of association. The results of this analysis showed that:

1) Variant rs35280891 at 3p21.31 suggestively significant in our GWAS ($p = 6.88 \times 10^{-7}$) was nominally significant in both UK biobank comparisons—severe versus non-severe and hospitalized positive versus non-hospitalized positive phenotypes ($p = 2.22 \times 10^{-4}$ and $p = 7.57 \times 10^{-7}$, respectively).
2) Variant rs73060324 at 3p21.31 suggestively significant in our GWAS ($p = 7.54 \times 10^{-6}$) showed a statistical trend related to UK biobank hospitalized positive versus non-hospitalized positive phenotype ($p = 0.0748$).
3) Variant rs78317595 at 5q11.2 suggestively significant in our GWAS ($p = 6.59 \times 10^{-6}$) was nominally significant in the UK biobank severe versus non-severe comparison ($p = 0.042$). However, the odds ratio for the effect allele C was below 1 (protective) in our GWAS (OR = 0.177, 95% CI [0.083–0.379]), and above 1 (risk) in the UK biobank comparison (1.130, 95% CI [1.004–1.271]), therefore we cannot consider this result validated by the UK biobank data. Other results related to lead variants from our GWAS, namely rs61964606, rs13176661, and rs1331359 were not validated by the UK biobank data, either ($p > 0.05$).

The results of the analysis are contained in the **Supplementary Table S5**.

## 3.6 Replication Analysis of the Previously Identified Association Signals

Next, we performed a replication analysis of the 12 variants reported to be associated with COVID-19 severity in previous studies. We have chosen associations that emerged in large

GWAS meta-studies (Horowitz et al., 2022; Roberts et al., 2022) and recent studies that analyzed associations of *ACE2* genetic variants with COVID-19 severity (Martínez-Gómez et al., 2022; Sabater Molina et al., 2022) (**Supplementary Table S6**). Our selection approach was focused on variants that showed association with severity and not with susceptibility to COVID-19. Of the 12 selected variants, we replicated 4 variants in phenotype COVID-19 related pneumonia/without pneumonia (the same as severe and moderate versus mild) (*IFNAR2* rs13050728, *LZTFL1* rs35081325, *CCHCR1* rs143334143, *ACE2* rs2106809) and 2 variants in phenotype severe/mild (*LZTFL1* rs35081325, *CCHCR1* rs143334143) at least at nominal *p*-value of 0.05 (**Supplementary Table S6**). We confirmed that all replicated associations have the same direction of effects (beta) as previously reported.

## 4 DISCUSSION

This study aimed to elucidate genetic risk loci associated with different COVID-19 clinical outcomes in the Serbian population. We detected a novel risk locus at chromosome 13q21.33, significantly associated with the SARS-CoV-2 infection-induced pneumonia. Also, we identified other suggestive signals at positions 3p21.31, 5p15.33, 5q11.2, and 9p23 associated with pneumonia and a suggestive signal at 3p21.31, associated with severe COVID-19 disease.

The peak association signal at 13q21.33 was located close to the gene *KLHL1,* in the intergenic region surrounded by the regulatory sequences. KLHL1 (Kelch Like Family Member 1) is a protein belonging to a family of actin-organizing proteins which modulates voltage-gated calcium channels expressed primarily in various brain tissues (Perissinotti et al., 2014). In a previous genome-wide analysis, the interaction of *KLHL1* locus and early life smoke exposure was found to be associated with the onset of childhood asthma (Sugier et al., 2019). Moreover, it was shown that an increased mutation rate in the *KLHL1* gene was associated with lifetime benzo(a) pyrene exposure in patients with air-pollution-related lung cancers (Yu et al., 2015). Transcriptional factors JUND and FOXA1 binding sites are found in the proximity of the lead variant rs61964606 and both have been related to gene expression regulation in previous studies of SARS-CoV-2 infection (Qiao et al., 2020; Ahmed et al., 2021).

The same locus harbors bidirectional transcripts of the *ATXN8* gene which codes for almost pure polyglutamine protein and the long non-coding antisense *ATXN80S* gene. Both genes contain expanded trinucleotide repeats associated with spinocerebellar ataxia type 8 (Batra et al., 2010). Although COVID-19 is primarily a respiratory disease, it can cause neurological manifestations as well. With COVID-19 worldwide expansion, there is an increased number of studies reporting cases of coronavirus induced acute cerebellar ataxia (Chan et al., 2021; Povlow and Auerbach, 2021; Werner et al., 2021). To the best of our knowledge, no previous study connected 13q21.33 nor any of the genes found in this locus to COVID-19. Therefore, we cannot reliably implicate causality without further investigation.

Although 13q21.33 was the only locus to show significant association at the genome-wide level in our study, other loci showed suggestive association. They harbor potentially relevant candidates with a biological function that could be important to the COVID-19 severity.

Our genome-wide analysis of pneumonia and severe COVID-19 confirmed findings from previous studies and showed suggestive signals at chromosome 3p21.31. Locus 3p21.31 has been discovered as the most strongly associated with COVID-19 in a study of severely ill Italian and Spanish patients by the Severe Covid-19 GWAS group (Ellinghaus et al., 2020). This finding was confirmed in another GWAS focused on critically ill patients from the UK intensive care units (Pairo-Castineira et al., 2021). A GWAS by the COVID-19 host genetics initiative (HGI) organized on a worldwide level which included genomic profiles of almost 50,000 SARS-CoV-2 positive, hospitalized, or critically ill COVID-19 patients also showed a strong association of the 3p21.31 locus with severe disease, as well as infection rate (COVID-19 Host Genetics Initiative, 2021). This finding was subsequently confirmed in another large GWAS (Shelton et al., 2021), and also a study that employed a different methodology (Rescenko et al., 2021).

Locus 3p21.31 is rich in protein-coding genes, some of which could impact COVID-19 disease severity. Our genome-wide analysis of pneumonia and severe COVID-19 point to two overlapping regions (defined by linkage disequilibrium analysis), both of which include *SLC6A20* and *LZTFL1* genes. These two genes have been previously considered causative factors related to COVID-19 susceptibility and severity (Ellinghaus et al., 2020; Pairo-Castineira et al., 2021; Shelton et al., 2021). SLC6A20 is an imino acid transporter co-expressed in the intestine and lungs with ACE2 membrane enzyme. Notably, a heterodimer of ACE2 and either SLC6A19 or SLC6A20 serves as a binding site for the SARS-CoV-2 spike protein, which may facilitate viral infection (Camargo et al., 2020). *LZTFL1* gene is implicated in ciliary function in the lungs important for airway viral clearance (Fink-Baldauf et al., 2022). Eight genes (*CCR9, CXCR6, XCR1, CCR3, CCR1, CCR2, CCR5,* and *CCRL2*) related to chemokine pathways, involved in the migration of leukocytes, are located near locus 3p21.31. Elevated levels of chemokines can cause acute respiratory disease syndrome in COVID-19 patients, which is associated with poor outcomes (Khalil et al., 2021).

*EMS1* gene, located at suggestive risk locus 5q11.2, encodes a dermatan sulfate proteoglycan called endocan that is mainly secreted by pulmonary and kidney vascular endothelial cells in response to inflammatory cytokines (Lassalle et al., 1996). Its level can predict multiple organ dysfunctions and mortality in patients with acute respiratory distress (Tang et al., 2014). In one proteomic study, endocan was reported in the top 50 plasma proteins found elevated in the SARS-CoV-2 infected patients with mild to moderate disease (Zhong et al., 2021). Moreover, the *GZMK* gene which codes for granzyme K, a serine protease found in cytoplasmic granules of cytotoxic lymphocytes, is located next to the *EMS1* gene. Decreased level of *GZMK* mRNA, as well as a decreased proportion of

effector memory $CD8^+$ T cells that produce GZMK, was observed in the peripheral blood of COVID-19 patients compared to healthy subjects (Ramljak et al., 2021). Consistently, patients with severe disease had lower proportions of $CD8^+$ T cells that express the *GZMK* gene compared to moderate patients in the study that analyzed COVD-19 single-cell landscape of bronchoalveolar immune cells (Liao et al., 2020).

We identified another suggestive signal at chromosome 5p15.33, ~300–400 kb distant from genes *IRX4*, *NDUFS6* and *MRPL36*. Exploration of this locus indicated that it may function as an enhancer in different tissues, including the lungs. The *NDUFS6* and *MRPL36* are both nuclear-encoded mitochondrial genes. *NDUFS6* encodes the subunit of the NADH:ubiquinone oxidoreductase (Complex I) while *MRPL36* encodes mitochondrial ribosomal protein. A study that analyzed RNA-Seq data derived from primary cells, cell lines, as well as lung and bronchoalveolar lavage fluid of COVID-19 patients showed that SARS-CoV-2 significantly downregulated nuclear-encoded mitochondrial genes related to cellular respiration and Complex I across all models, while mitochondrial ribosomal protein genes' expression was particularly downregulated in primary cells (Miller et al., 2021). It has been shown that SARS-CoV-2 proteins directly interact with several Complex I subunits (Gordon et al., 2020). Reports on other respiratory viruses suggested that inhibition of Complex I could promote viral replication (Hu et al., 2019).

The suggestive signal at chromosome 9p23 is located close to genes *TYRP1*, *LURAP1L*, and the antisense transcript *LURAP1L-AS1*. LURAP1L (Leucine Rich Adaptor Protein 1 Like) is a protein predicted to be involved in inflammatory signaling since it is a paralog to LURAP1 which acts as an activator of the canonical NF-κB signaling pathway (Jing et al., 2010). Previous GWA studies identified suggestive associations of *LURAP1L* genetic variants with juvenile idiopathic arthritis (Li et al., 2015) and pulmonary function in smokers (Lutz et al., 2015). *LURAP1L* gene expression was found significantly increased in the $CD4^+$ T-cells of obese compared to normal-weighted children with asthma (Rastogi et al., 2018). One study showed that *LURAP1L* was among genes with altered gene expression in lung tissue of deceased COVID-19 patients that could be targeted with the anti-inflammatory activities of glucocorticoid drugs (Sharma, 2021).

The main limitation of this study is its relatively small sample size. This limitation is mitigated by (1) good quality clinical data collected by health professionals, so we did not have to rely on self-reported information from patients, and (2) our study group included only patients hospitalized in the first 3 months of the pandemic, so they were all unvaccinated and likely contracted the same variant of SARS-CoV-2 virus, which limited the number of cofounders. An additional limitation of this study is the absence of the replication cohort. Identified candidate genetic variants need to be validated in an independent group of patients before being recognized as reliable disease markers. If they do get validated, further functional studies can be performed to decipher their biological role in COVID-19 infection.

We have observed that the frequencies of effect alleles of risk loci identified in our study were highly variable among world-wide populations. In order to identify important genetic patterns underlying disease outcomes, it is essential to analyze different ethnicities. Important genetic patterns might be difficult to detect if they are not sufficiently represented in the analyzed population, although they may be much more frequent in other populations. GWAS analysis is not suitable for the identification of rare genetic variants, and this approach has been the most successful so far in the detection of genetic risk factors related to COVID-19.

The current study was the first GWAS to include COVID-19 patients of Serbian origin. In fact, until now, no comprehensive genetic study of COVID-19 patients included patients from southeastern Europe, a region generally underrepresented in genomic research. This study confirms the previously validated genetic locus at 3p21.31 as a marker of severe disease. In addition, our results point to novel genomic loci at 5p15.33, 5q11.2, 9p23, and 13q21.33 potentially implicated in the development of pneumonia and more severe COVID-19 disease.

## CODE AVAILABILITY

We provided two Common Workflow Language (CWL) (Crusoe et al., 2021) workflows that were used on the Seven Bridges Cancer Genome Cloud platform (Lau et al., 2017): (1) a preprocessing pipeline that includes variant calling from the raw microarray data, phasing, and imputation and (2) the GWAS analysis. Both CWL workflows are available as JSON format files in the GitHub repository (https://github.com/markozecevic/covid19gwas).

## DATA AVAILABILITY STATEMENT

GWAS summary data of the current study has been deposited in the National Human Genome Research Institute-European Bioinformatics Institute (NHGRI-EBI) GWAS Catalog database (study accession numbers: severe and moderate/mild - GCST90104347, severe/mild - GCST90104348).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Institute of Molecular Genetics and Genetic Engineering University of Belgrade. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

MZ: Data processing and statistical analysis, investigation, results interpretation, writing—draft preparation and editing. NK: Data analysis, methodology, results interpretation, writing—draft preparation and editing. BR: Sample processing, investigation, methodology. VG: Sample processing, investigation, methodology. VS-T, MS, GS, and LL: Methodology, sample

collection, and clinical data analysis. BZ: Methodology, investigation, writing—review and editing, SP: Concept and design of the study, investigation, writing—review and editing, BS: Concept and design of the study, data analysis, investigation, results interpretation, writing—draft preparation and review of final manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was funded by the Ministry of Education, Science and Technological Development, Republic of Serbia (Grant No. 451-03-68/2022-14/200042).

## ACKNOWLEDGMENTS

Genotyping of the samples was supported by COVID-19 Host Genetics Initiative. It was performed by the Human Genomics Facility of the Genetic Laboratory of the Department of Internal Medicine at Erasmus MC. Computational resources were provided by The Cancer Genomics Cloud, powered by Seven Bridges, a component of the NCI Cancer Research Data Commons (datacommons.cancer.gov), funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN261201400008C and ID/IQ Agreement No. 17X146 under Contract No. HHSN261201500003I.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.911010/full#supplementary-material

**Data Sheet 1 |** Supplementary Methods.

**Data Sheet 2 |** Supplementary Results.

## REFERENCES

Ahmed, S. S. S. J., Paramasivam, P., Kamath, M., Sharma, A., Rome, S., and Murugesan, R. (2021). Genetic Exchange of Lung-Derived Exosome to Brain Causing Neuronal Changes on COVID-19 Infection. *Mol. Neurobiol.* 58, 5356–5368. doi:10.1007/s12035-021-02485-9

Andolfo, I., Russo, R., Lasorsa, V. A., Cantalupo, S., Rosato, B. E., Bonfiglio, F., et al. (2021). Common Variants at 21q22.3 Locus Influence MX1 and TMPRSS2 Gene Expression and Susceptibility to Severe COVID-19. *iScience* 24, 102322. doi:10.1016/j.isci.2021.102322

Auton, A., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A Global Reference for Human Genetic Variation. *Nature* 526, 68–74. doi:10.1038/nature15393

Batra, R., Charizanis, K., and Swanson, M. S. (2010). Partners in Crime: Bidirectional Transcription in Unstable Microsatellite Disease. *Hum. Mol. Genet.* 19, R77–R82. doi:10.1093/hmg/ddq132

Camargo, S. M. R., Vuille-Dit-Bille, R. N., Meier, C. F., and Verrey, F. (2020). ACE2 and Gut Amino Acid Transport. *Clin. Sci.* 134, 2823–2833. doi:10.1042/CS20200477

Canela-Xandri, O., Rawlik, K., and Tenesa, A. (2018). An Atlas of Genetic Associations in UK Biobank. *Nat. Genet.* 50, 1593–1599. doi:10.1038/s41588-018-0248-z

Chan, J. L., Murphy, K. A., and Sarna, J. R. (2021). Myoclonus and Cerebellar Ataxia Associated with COVID-19: A Case Report and Systematic Review. *J. Neurol.* 268, 3517–3548. doi:10.1007/s00415-021-10458-0

COVID-19 Host Genetics Initiative (2021). Mapping the Human Genetic Architecture of COVID-19. *Nature* 600, 472–477. doi:10.1038/s41586-021-03767-x

Crusoe, M. R., Abeln, S., Iosup, A., Amstutz, P., Chilton, J., Tijanić, N., et al. (2021). Methods Included: Standardizing Computational Reuse and Portability with the Common Workflow Language. arXiv, 2105.07028. doi:10.1145/3486897

de Leeuw, C. A., Mooij, J. M., Heskes, T., and Posthuma, D. (2015). MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* 11, e1004219. doi:10.1371/journal.pcbi.1004219

Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., et al. (2020). Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N. Engl. J. Med.* 383, 1522–1534. doi:10.1056/NEJMoa2020283

Fink-Baldauf, I. M., Stuart, W. D., Brewington, J. J., Guo, M., and Maeda, Y. (2022). CRISPRi Links COVID-19 GWAS Loci to LZTFL1 and RAVER1. *eBioMedicine* 75, 103806. doi:10.1016/j.ebiom.2021.103806

Fricke-Galindo, I., and Falfán-Valencia, R. (2021). Genetics Insight for COVID-19 Susceptibility and Severity: A Review. *Front. Immunol.* 12, 622176. doi:10.3389/fimmu.2021.622176

Gogarten, S. M., Sofer, T., Chen, H., Yu, C., Brody, J. A., Thornton, T. A., et al. (2019). Genetic Association Testing Using the GENESIS R/Bioconductor Package. *Bioinformatics* 35, 5346–5348. doi:10.1093/bioinformatics/btz567

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. *Nature* 583, 459–468. doi:10.1038/s41586-020-2286-9

Horowitz, J. E., Kosmicki, J. A., Damask, A., Sharma, D., Roberts, G. H. L., Justice, A. E., et al. (2022). Genome-Wide Analysis Provides Genetic Evidence that ACE2 Influences COVID-19 Risk and Yields Risk Scores Associated with Severe Disease. *Nat. Genet.* 54, 382–392. doi:10.1038/s41588-021-01006-7

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., et al. (2021). Ensembl 2021. *Nucleic Acids Res.* 49, D884–D891. doi:10.1093/nar/gkaa942

Hu, J., and Wang, Y. (2021). The Clinical Characteristics and Risk Factors of Severe COVID-19. *Gerontology* 67, 255–266. doi:10.1159/000513400

Hu, M., Bogoyevitch, M. A., and Jans, D. A. (2019). Subversion of Host Cell Mitochondria by RSV to Favor Virus Production Is Dependent on Inhibition of Mitochondrial Complex I and ROS Generation. *Cells* 8, 1417. doi:10.3390/cells8111417

Jing, Z., Yuan, X., Zhang, J., Huang, X., Zhang, Z., Liu, J., et al. (2010). Chromosome 1 Open Reading Frame 190 Promotes Activation of NF-κB Canonical Pathway and Resistance of Dendritic Cells to Tumor-Associated Inhibition *In Vitro*. *J. Immunol.* 185, 6719–6727. doi:10.4049/jimmunol.0903869

Khalil, B. A., Elemam, N. M., and Maghazachi, A. A. (2021). Chemokines and Chemokine Receptors during COVID-19 Infection. *Comput. Struct. Biotechnol. J.* 19, 976–988. doi:10.1016/j.csbj.2021.01.034

Kotur, N., Skakic, A., Klaassen, K., Gasic, V., Zukic, B., Skodric-Trifunovic, V., et al. (2021). Association of Vitamin D, Zinc and Selenium Related Genetic Variants with COVID-19 Disease Severity. *Front. Nutr.* 8, 689419. doi:10.3389/fnut.2021.689419

Kuo, C.-L., Pilling, L. C., Atkins, J. L., Masoli, J. A. H., Delgado, J., Kuchel, G. A., et al. (2020). APOE e4 genotype predicts severe COVID-19 in the UK Biobank Community Cohort. *J. Gerontol. - Ser. A Biol. Sci. Med. Sci.* 75, 2231–2232. doi:10.1093/gerona/glaa131

Lassalle, P., Molet, S., Janin, A., Van der Heyden, J., Tavernier, J., Fiers, W., et al. (1996). ESM-1 is a Novel Human Endothelial Cell-Specific Molecule Expressed in Lung and Regulated by Cytokines. *J. Biol. Chem.* 271, 20458–20464. doi:10.1074/jbc.271.34.20458

Lau, J. W., Lehnert, E., Sethi, A., Malhotra, R., Kaushik, G., Onder, Z., et al. (2017). The Cancer Genomics Cloud: Collaborative, Reproducible, and Democratized-A New Paradigm in Large-Scale Computational Research. *Cancer Res.* 77, e3–e6. doi:10.1158/0008-5472.CAN-17-0387

Lee, B. T., Barber, G. P., Benet-Pagès, A., Casper, J., Clawson, H., DIekhans, M., et al. (2022). The UCSC Genome Browser Database: 2022 Update. *Nucleic Acids Res.* 50, D1115–D1122. doi:10.1093/nar/gkab959

Lesurf, R., Cotto, K. C., Wang, G., Griffith, M., Kasaian, K., Jones, S. J. M., et al. (2016). ORegAnno 3.0: A Community-Driven Resource for Curated Regulatory Annotation. *Nucleic Acids Res.* 44, D126–D132. doi:10.1093/nar/gkv1203

Li, Y. R., Li, J., Zhao, S. D., Bradfield, J. P., Mentch, F. D., Maggadottir, S. M., et al. (2015). Meta-Analysis of Shared Genetic Architecture across Ten Pediatric Autoimmune Diseases. *Nat. Med.* 21, 1018–1027. doi:10.1038/nm.3933

Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., et al. (2020). Single-Cell Landscape of Bronchoalveolar Immune Cells in Patients with COVID-19. *Nat. Med.* 26, 842–844. doi:10.1038/s41591-020-0901-9

Lin, S.-H., Thakur, R., and Machiela, M. J. (2021). LDexpress: An Online Tool for Integrating Population-Specific Linkage Disequilibrium Patterns with Tissue-Specific Expression Data. *BMC Bioinforma.* 22, 608. doi:10.1186/s12859-021-04531-8

Lutz, S. M., Cho, M. H., Cho, M. H., Young, K., Hersh, C. P., Castaldi, P. J., et al. (2015). A Genome-Wide Association Study Identifies Risk Loci for Spirometric Measures Among Smokers of European and African Ancestry. *BMC Genet.* 16, 138. doi:10.1186/s12863-015-0299-4

Machiela, M. J., and Chanock, S. J. (2015). LDlink: A Web-Based Application for Exploring Population-Specific Haplotype Structure and Linking Correlated Alleles of Possible Functional Variants. *Bioinformatics* 31, 3555–3557. doi:10.1093/bioinformatics/btv402

Martínez-Gómez, L. E., Herrera-López, B., Martinez-Armenta, C., Ortega-Peña, S., Camacho-Rea, M. D. C., Suarez-Ahedo, C., et al. (2022). ACE and ACE2 Gene Variants are Associated with Severe Outcomes of COVID-19 in Men. *Front. Immunol.* 13, 812940. doi:10.3389/fimmu.2022.812940

Miller, B., Silverstein, A., Flores, M., Cao, K., Kumagai, H., Mehta, H. H., et al. (2021). Host Mitochondrial Transcriptome Response to SARS-CoV-2 in Multiple Cell Models and Clinical Samples. *Sci. Rep.* 11, 3. doi:10.1038/s41598-020-79552-z

Pairo-Castineira, E., Clohisey, S., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., et al. (2021). Genetic Mechanisms of Critical Illness in COVID-19. *Nature* 591, 92–98. doi:10.1038/s41586-020-03065-y

Perissinotti, P. P., Ethington, E. G., Cribbs, L., Koob, M. D., Martin, J., and Piedras-Rentería, E. S. (2014). Down-Regulation of Endogenous KLHL1 Decreases Voltage-Gated Calcium Current Density. *Cell Calcium* 55, 269–280. doi:10.1016/j.ceca.2014.03.002

Povlow, A., and Auerbach, A. J. (2021). Acute Cerebellar Ataxia in COVID-19 Infection: A Case Report. *J. Emerg. Med.* 60, 73–76. doi:10.1016/j.jemermed.2020.10.010

Qiao, Y., Wang, X.-M., Mannan, R., Pitchiaya, S., Zhang, Y., Wotring, J. W., et al. (2020). Targeting Transcriptional Regulation of SARS-CoV-2 Entry Factors ACE2 and TMPRSS2. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2021450118. doi:10.1073/pnas.2021450118

Ramljak, D., Vukoja, M., Curlin, M., Vukojevic, K., Barbaric, M., Glamoclija, U., et al. (2021). Early Response of CD8+ T Cells in COVID-19 Patients. *J. Pers. Med.* 11, 1291. doi:10.3390/jpm11121291

Rastogi, D., Nico, J., Johnston, A. D., Tobias, T. A. M., Jorge, Y., Macian, F., et al. (2018). CDC42-Related Genes are Upregulated in Helper T Cells from Obese Asthmatic Children. *J. Allergy Clin. Immunol.* 141, 539–548. e7. doi:10.1016/j.jaci.2017.04.016

Rescenko, R., Peculis, R., Briviba, M., Ansone, L., Terentjeva, A., Litvina, H. D., et al. (2021). Replication of LZTFL1 Gene Region as a Susceptibility Locus for COVID-19 in Latvian Population. *Virol. Sin.* 36, 1241–1244. doi:10.1007/s12250-021-00448-x

Roberts, G. H. L., Partha, R., Rhead, B., Knight, S. C., Park, D. S., Coignet, M. V., et al. (2022). Expanded COVID-19 Phenotype Definitions Reveal Distinct Patterns of Genetic Association and Protective Effects. *Nat. Genet.* 54, 374–381. doi:10.1038/s41588-022-01042-x

Sabater Molina, M., Nicolás Rocamora, E., Bendicho, A. I., Vázquez, E. G., Zorio, E., Rodriguez, F. D., et al. (2022). Polymorphisms in ACE, ACE2, AGTR1 Genes and Severity of COVID-19 Disease. *PLoS One* 17, e0263140. doi:10.1371/journal.pone.0263140

Sharma, A. (2021). Inferring Molecular Mechanisms of Dexamethasone Therapy in Severe COVID-19 from Existing Transcriptomic Data. *Gene* 788, 145665. doi:10.1016/j.gene.2021.145665

Shelton, J. F., Shastri, A. J., Ye, C., Weldon, C. H., Filshtein-Sonmez, T., Coker, D., et al. (2021). Trans-Ancestry Analysis Reveals Genetic and Nongenetic Associations with COVID-19 Susceptibility and Severity. *Nat. Genet.* 53, 801–808. doi:10.1038/s41588-021-00854-7

Sugier, P. E., Sarnowski, C., Granell, R., Laprise, C., Ege, M. J., Margaritte-Jeannin, P., et al. (2019). Genome-Wide Interaction Study of Early-Life Smoking Exposure on Time-to-Asthma Onset in Childhood. *Clin. Exp. Allergy* 49, 1342–1351. doi:10.1111/cea.13476

Tang, L., Zhao, Y., Wang, D., Deng, W., Li, C., Li, Q., et al. (2014). Endocan Levels in Peripheral Blood Predict Outcomes of Acute Respiratory Distress Syndrome. *Mediat. Inflamm.* 2014, 625180. doi:10.1155/2014/625180

Thibord, F., Chan, M. V., Chen, M.-H., and Johnson, A. D. (2022). A Year of COVID-19 GWAS Results from the GRASP Portal Reveals Potential Genetic Risk Factors. *Hum. Genet. Genomics Adv.* 3, 100095. doi:10.1016/j.xhgg.2022.100095

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., et al. (2021). Genome-Wide Association Studies. *Nat. Rev. Methods Prim.* 1, 59. doi:10.1038/s43586-021-00056-9

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F. C. P., et al. (2018). Comprehensive Functional Genomic Resource and Integrative Model for the Human Brain. *Science* 362, eaat8464. doi:10.1126/science.aat8464

Ward, L. D., and Kellis, M. (2012). HaploReg: A Resource for Exploring Chromatin States, Conservation, and Regulatory Motif Alterations within Sets of Genetically Linked Variants. *Nucleic Acids Res.* 40, D930–D934. doi:10.1093/nar/gkr917

Watanabe, K., Taskesen, E., Van Bochoven, A., and Posthuma, D. (2017). Functional Mapping and Annotation of Genetic Associations with FUMA. *Nat. Commun.* 8, 1826. doi:10.1038/s41467-017-01261-5

Werner, J., Reichen, I., Huber, M., Abela, I. A., Weller, M., and Jelcic, I. (2021). Subacute Cerebellar Ataxia Following Respiratory Symptoms of COVID-19: A Case Report. *BMC Infect. Dis.* 21, 298. doi:10.1186/s12879-021-05987-y

Wu, P., Ding, L., Li, X., Liu, S., Cheng, F., He, Q., et al. (2021). Trans-Ethnic Genome-Wide Association Study of Severe COVID-19. *Commun. Biol.* 4, 1034. doi:10.1038/s42003-021-02549-5

Yu, X.-J., Yang, M.-J., Zhou, B., Wang, G.-Z., Huang, Y.-C., Wu, L.-C., et al. (2015). Characterization of Somatic Mutations in Air Pollution-Related Lung Cancer. *EBioMedicine* 2, 583–590. doi:10.1016/j.ebiom.2015.04.003

Zhang, Q., Bastard, P., Liu, Z., Le Pen, J., Moncada-Velez, M., Chen, J., et al. (2020). Inborn Errors of Type I IFN Immunity in Patients with Life-Threatening COVID-19. *Science* 370, eabd4570. doi:10.1126/science.abd4570

Zhong, W., Altay, O., Arif, M., Edfors, F., Doganay, L., Mardinoglu, A., et al. (2021). Next Generation Plasma Proteome Profiling of COVID-19 Patients with Mild to Moderate Symptoms. *EBioMedicine* 74, 103723. doi:10.1016/j.ebiom.2021.103723

Zou, Y., Carbonetto, P., Wang, G., and Stephens, M. (2021). Fine-Mapping from Summary Data with the "Sum of Single Effects" Model. bioRxiv. doi:10.1101/2021.11.03.467167

**Conflict of Interest:** Author MZ was employed by the company Seven Bridges.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

BRIEF RESEARCH REPORT
published: 18 July 2022
doi: 10.3389/fgene.2022.931562

# Evaluation of a Functional Single Nucleotide Polymorphism of the SARS-CoV-2 Receptor *ACE2* That Is Potentially Involved in Long COVID

*Yu-Si Luo[1,2†], Lei Luo[3†], Wei Li[4], Yan Chen[5], Guo-Feng Wu[1], Fang Chen[2], Hu-Yan Shen[2], Hong-Man Li[6], Ming-Yang Guo[2], Sha Yin[3], Ke Zhang[2]* and Zhong-Shan Cheng[7]**

[1]Department of Emergency, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, [2]The Key and Characteristic Laboratory of Modern Pathogenicity Biology, School of Basic Medical Sciences, Guizhou Medical University, Guiyang, China, [3]Good Clinical Practice Center, Guizhou Provincial People's Hospital, Guiyang, China, [4]Department of Cardiovascular Medicine, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, [5]The High Efficacy Application of Natural Medicinal Resources Engineering Center of Guizhou Province, School of Pharmaceutical Sciences, Guizhou Medical University, Guiyang, China, [6]Department of Hypertension, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, [7]Center for Applied Bioinformatics, St. Jude Children's Research Hospital, Memphis, TN, United States

Since the occurrence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in December 2019, SARS-CoV-2 has led to a global coronavirus disease 2019 (COVID-19) pandemic. A better understanding of the SARS-CoV-2 receptor *ACE2* at the genetic level would help combat COVID-19, particularly for long COVID. We performed a genetic analysis of *ACE2* and searched for its common potential single nucleotide polymorphisms (SNPs) with minor allele frequency >0.05 in both European and Chinese populations that would contribute to *ACE2* gene expression variation. We thought that the variation of the *ACE2* expression would be an important biological feature that would strongly affect COVID-19 symptoms, such as "brain fog", which is highlighted by the fact that ACE2 acts as a major cellular receptor for SARS-CoV-2 attachment and is highly expressed in brain tissues. Based on the human GTEx gene expression database, we found rs2106809 exhibited a significant correlation with the *ACE2* expression among multiple brain and artery tissues. This expression correlation was replicated in an independent European brain eQTL database, Braineac. rs2106809*G also displays significantly higher frequency in Asian populations than in Europeans and displays a protective effect ($p$ = 0.047) against COVID-19 hospitalization when comparing hospitalized COVID-19 cases with non-hospitalized COVID-19 or SARS-CoV-2 test-negative samples with European ancestry from the UK Biobank. Furthermore, we experimentally demonstrated that rs2106809*G could upregulate the transcriptional activity of *ACE2*. Therefore, integrative analysis and functional experiment strongly support that *ACE2* SNP rs2106809 is a functional brain eQTL and its potential involvement in long COVID, which warrants further investigation.

**Keywords: SARS-CoV-2, COVID-19, long COVID, ACE2, rs2106809**

OPEN ACCESS

***Edited by:***
*Ramcés Falfán-Valencia, Instituto Nacional de Enfermedades Respiratorias-México (INER), Mexico*

***Reviewed by:***
*Debmalya Barh, Institute of Integrative Omics and Applied Biotechnology (IIOAB), India*
*José Manuel Rodríguez-Pérez, Instituto Nacional de Cardiologia Ignacio Chavez, Mexico*

***Correspondence:**
*Ke Zhang*
*kezhang@gmc.edu.cn*
*Zhong-Shan Cheng*
*cheng.zhong.shan@gmail.com*

*†These authors have contributed equally to this work and share first authorship*

***Specialty section:***
*This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics*

***Received:*** *29 April 2022*
***Accepted:*** *13 June 2022*
***Published:*** *18 July 2022*

***Citation:***
*Luo Y-S, Luo L, Li W, Chen Y, Wu G-F, Chen F, Shen H-Y, Li H-M, Guo M-Y, Yin S, Zhang K and Cheng Z-S (2022) Evaluation of a Functional Single Nucleotide Polymorphism of the SARS-CoV-2 Receptor ACE2 That Is Potentially Involved in Long COVID. Front. Genet. 13:931562. doi: 10.3389/fgene.2022.931562*

## INTRODUCTION

*ACE2* encodes the protein angiotensin-converting enzyme (ACE) 2, which is a receptor of SARS-CoV-2 [1-3]. Before ACE2 was identified as the SARS-CoV-2 receptor, it was well known as a negative regulator of the renin–angiotensin system (RAS). The functions of ACE2 in the RAS are to hydrolyze angiotensin (Ang) I into Ang (1–9) and to directly cleave Ang II, a powerful vasoconstrictor, to Ang (1–7) [4]. Therefore, the hydrolysis of Ang I and Ang II by ACE2 strictly controls the deleterious functions of Ang II and Ang I to cardiovascular system by limiting oxidative stress and then inducing antifibrotic and vasodilatory actions [5]. Recently, highly expressed Ang II-induced severe complications have been observed in COVID-19 [6]. However, not only mainly expressed in the cardiovascular system, *ACE2* is also discovered to be highly decoded in brain regions such as the cerebral cortex, amygdala, and the brainstem of humans. The brainstem is a sort of key life control center for the maintenance of cardiorespiratory, cardiovascular, gastrointestinal, and neurological processes. The pons and medulla of the brainstem have the highest *ACE2* expression level [7].

Based on the sizeable scientific literature reports, ACE2 was hijacked as a functional cell receptor for virus attachment and entry by lineage B β-coronaviruses, including SARS-CoV-2 [1-3]. SARS-CoV-2 was depicted as a neurotropic virus, in line with the evidence such as the capacity of SARS-CoV-2 to infect and replicate in neuro cells [8], and COVID-19 was also thought as an endothelial but not an epithelial disease[9], consistent with the pieces of evidence that *ACE2* is broadly expressed on the membrane of the endothelium and pericytes which are composed of the capillaries of vascular microcirculation of all organs, including the cerebrum [10]. Since SARS-CoV-2 infection reduces *ACE2* expression [6] and brainstem has the relatively higher *ACE2* expression level than other cerebral regions, the dysregulation of *ACE2* in the brainstem after SARS-CoV-2 infection might be closely associated with a novel conception: long COVID.

The British National Institute for Health and Care Excellence defines long COVID as "signs and symptoms that develop during or after an infection consistent with COVID-19, continue for more than 12 weeks, and are not explained by an alternative diagnosis" [11]. The current description of signs and symptoms for long COVID includes fatigue, dyspnea, headache, anxiety, depression, cognitive disturbances (brain fog), cough, joint and chest pains, smell and taste dysfunction, and myalgia that persist for at least 4 weeks after symptom onset or hospital discharge [12,13]. Further study showed 30–80% COVID-19 survivors suffered from long COVID lasting for 1–6 months [14].

Remarkably, as one potent explanation for "brain fog," anxiety, and depression of long COVID symptoms, SARS-CoV-2 was demonstrated to infect astrocytes, one of the neuro cells distributed in the brainstem with a vigorous *ACE2* expression, and to subsequently impede the transfer of glucose and lactate from astrocytes to neurons [15]. Furthermore, infection of astrocytes could lead to disruption of the blood–brain barrier; then, the systemic "cytokine storm" swarming, neuro-inflammation, and microglial activation happened [16]. Such pathophysiological basis of neurological symptoms for long COVID could be led by the disrupted *ACE2* expression in brain regions, especially in the brainstem. Therefore, *ACE2* gene polymorphism might play a key role in long COVID.

As it is urgent to understand the novel neurological issue, such as "brain fog," of long COVID, we performed an integrative genetic analysis of the SARS-CoV-2 receptor *ACE2* and prioritized a promoter SNP of *ACE2*, rs2106809, which could regulate the expression of *ACE2* in brain tissues. This study may shed light on genetics factors that are involved in long COVID.

## METHODS AND MATERIALS

### *ACE2* Expression Analyses

The *ACE2* expression in diverse tissues of human body was depicted *via* the Genotype-Tissue Expression (GTEx) project. The GTEx project was established for sharing characteristic human transcriptomes within and across individuals for a wide variety of primary tissues and cell types. The expression levels of *ACE2* in all donors across 49 tissues were retrieved from GTEx V8 [17] (https://storage.googleapis.com/gtex_analysis_v8/rna_seq_data/GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_tpm.gct.gz). Sex information for GTEx samples was obtained from the phenotype file "GTEx_Analysis_v8_Annotations_SubjectPhenotypesDS.txt" (https://storage.googleapis.com/gtex_analysis_v8/annotations/GTEx_Analysis_v8_Annotations_SubjectPhenotypesDS.txt). Tissue information for GTEx samples was collected from the file "GTEx_Analysis_v8_Annotations_SampleAttributesDS.txt" (https://storage.googleapis.com/gtex_analysis_v8/annotations/GTEx_Analysis_v8_Annotations_SampleAttributesDS.txt).

### Prioritization of the *ACE2* Promoter Single Nucleotide Polymorphism rs2106809 for the Functional Study

In order to obtain common functional SNPs regulating the *ACE2* expression in both European and Chinese populations, we downloaded all *ACE2* eQTLs from the GTEx database, where most samples are derived from European ancestry. In detail, we used *ACE2* to search for all eQTLs in the GTEx portal (https://gtexportal.org/home/); by clicking the option of "Significant Single-Tissue eQTLs for *ACE2* (ENSG00000130234.10) in all tissues," we exported all eQTLs of *ACE2* from the GTEx portal. The total number of *ACE2* eQTLs was 215, all of which were associated with the *ACE2* expression with $p < 0.001$ across any of 49 GTEx tissues.

As our purpose was to study functional SNPs that are common between European and Chinese samples, we kept these *ACE2* eQTLs based on their minor allele frequencies (MAFs) > 0.05 in both European and Chinese populations, and the MAFs of these SNPs in the two populations were determined by PLINK2.0 [18] with the genotyping data across 503 European individuals and 103 Chinese Han in Beijing (CHB) obtained from the

1000 Genomes Project released in June 2011 [19]. This filtering resulted in 140 *ACE2* eQTLs with MAF > 0.05 in both populations.

With genotyping data for these *ACE2* eQTLs of European and CHB populations from the 1000 Genomes Project, we used Haploview [20] to analyze and visualize the linkage disequilibrium (LD) pattern of these common *ACE2* eQTLs between European and Chinese samples. By evaluating the genomic region of *ACE2* in the UCSC Genome Browser, we specifically focused on *ACE2* eQTLs located in the *ACE2* promoter (chrX: 15 597 069–15 607 069; hg38), and only the SNP rs2106809 is common with MAF > 0.05 in both European and Chinese populations. Thus, this *ACE2* promoter SNP was selected for the functional study.

## Association of the *ACE2* Promoter Single Nucleotide Polymorphisms With COVID-19 Hospitalization in European Population

To determine whether the promoter SNP rs2106809 is associated with COVID-19, we evaluated the COVID-19 genome-wide association study (GWAS) summary statistics freely available at GRASP (https://grasp.nhlbi.nih.gov/Covid19GWASResults.aspx) [21]. We prioritized the COVID-19 GWAS with sample sizes of cases > 1 000, controls required to be tested for SARS-CoV-2 infection, conducted in a single population without a potential ancestry effect, and COVID-19 phenotypes highly related to COVID-19 severity. We found the COVID-19 hospitalization GWAS of "Hospitalized COVID-19–positive vs. non-hospitalized COVID-19–positive or COVID-19–negative in EUR UK Biobank tested samples" met these criteria. This European COVID-19 hospitalization GWAS has 1,712 cases and 56,988 controls, the summary statistics of which were downloaded from GRASP (https://grasp.nhlbi.nih.gov/downloads/COVID19GWAS/02242021/UKBB_hsptl_EURtested_022421.txt.gz).

## Cells

The A549 (ATCC, United States, CCL-185) and HEK293T (ATCC, United States, CRL-1573) cells were, respectively, grown in T75 tissue culture flasks using DMEM (Gibco Ltd., c11995500bt) supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin in a 37°C incubator with 5% $CO_2$. The cells were passaged every 3 days at a confluence of 70–80% approximately.

## Reporter Vector Construction, Transfection, and Luciferase Assay

The putative promoter segment of 1 328 bps in the *ACE2* promoter region (chrX: 15 599 868—15 601 196; hg38) harboring rs2106809*A and rs2106809*G was directly synthesized by Sangon Biotech (Shanghai) Co., Ltd. The endonucleases HindIII and KpnI were involved in the putative promoter segment synthesis and cloned into the pGL3-Basic vector (Promega Ltd., E1754) to generate two luciferase vectors with rs2106809*A and rs2106809*G, respectively. The correctness of the constructs was verified by direct sequencing. The equal amounts of the two luciferase vectors were, respectively, transfected into A549 and HEK293T cells by Lipofectamine 3000 (Thermo Fisher Scientific Ltd., L3000015). A luciferase assay was performed according to the instruction of the Dual Luciferase Reporter Gene Assay Kit (YEASEN Biotechnology Co., Ltd., 11402ES60). The bioluminescence of firefly luciferase promoted by the putative promoter element of *ACE2* and of *Renilla* luciferase as the internal control was read by using a GloMax® 20/20 Luminometer (Promega Ltd., E5311) at 560 and 480 nm, respectively. The final results were presented as the ratio between bioluminescence values of firefly luciferase and *Renilla* luciferase, as described before [22]. The luciferase assay results were analyzed using the Student's t-test, and the significance level was set at $p < 0.05$.

# RESULTS

## Identification of the Putative Functional Single Nucleotide Polymorphism rs2106809 of *ACE2*

*ACE2* expression levels were determined across 49 human tissues *via* the GTEx database (**Figure 1A**). The testis, small intestine, adipose, kidney (cortex), and heart showed the highest *ACE2* expression levels. We also noticed that arterial and brain tissues presented higher *ACE2* expression levels than other tissues. LD analysis of *ACE2* eQTLs in EUR and CHB populations showed similar LD patterns (**Figure 1B**). In addition, the minor allele G of rs2106809 is a protective allele associated with COVID-19 hospitalization in the European population ($p = 0.047$; **Figure 1C**). The normalized effect size (NES) of the single-tissue eQTLs of the *ACE2* SNP rs2106809 across 49 human tissues in the GTEx database (**Figure 2**) revealed that most brain tissues display a correlation between *ACE2* expression and rs2106809. This association was replicated in the UK brain eQTL database, Braineac (**Figure 3**). Taken together, the promoter SNP rs2106809*G is tightly associated with a higher *ACE2* expression in multiple brain tissues, which also shows protective association with COVID-19 hospitalization in European population.

## rs2106809 G Allele Increasing the Promoter Activity in Luciferase Assay

Based on genetic analysis, we predicted rs2106809 would affect the promoter activity of *ACE2*. In order to verify such prediction, we separately cloned two luciferase vectors with the genotype GG and AA of rs2106809 into the pGL3-Basic vector and confirmed it by sequencing (**Supplementary Figure S1**). The two luciferase vectors were, respectively, transfected into HEK293T cells and A549 cells for luciferase activity evaluation. The dual luciferase-reporter gene assay showed rs2106809 GG increased the promoter activity than the AA genotype in HEK293T cells and significantly enhanced the promoter activity than the AA genotype in A549 cells (**Figure 4**, $p < 0.05$). The results of

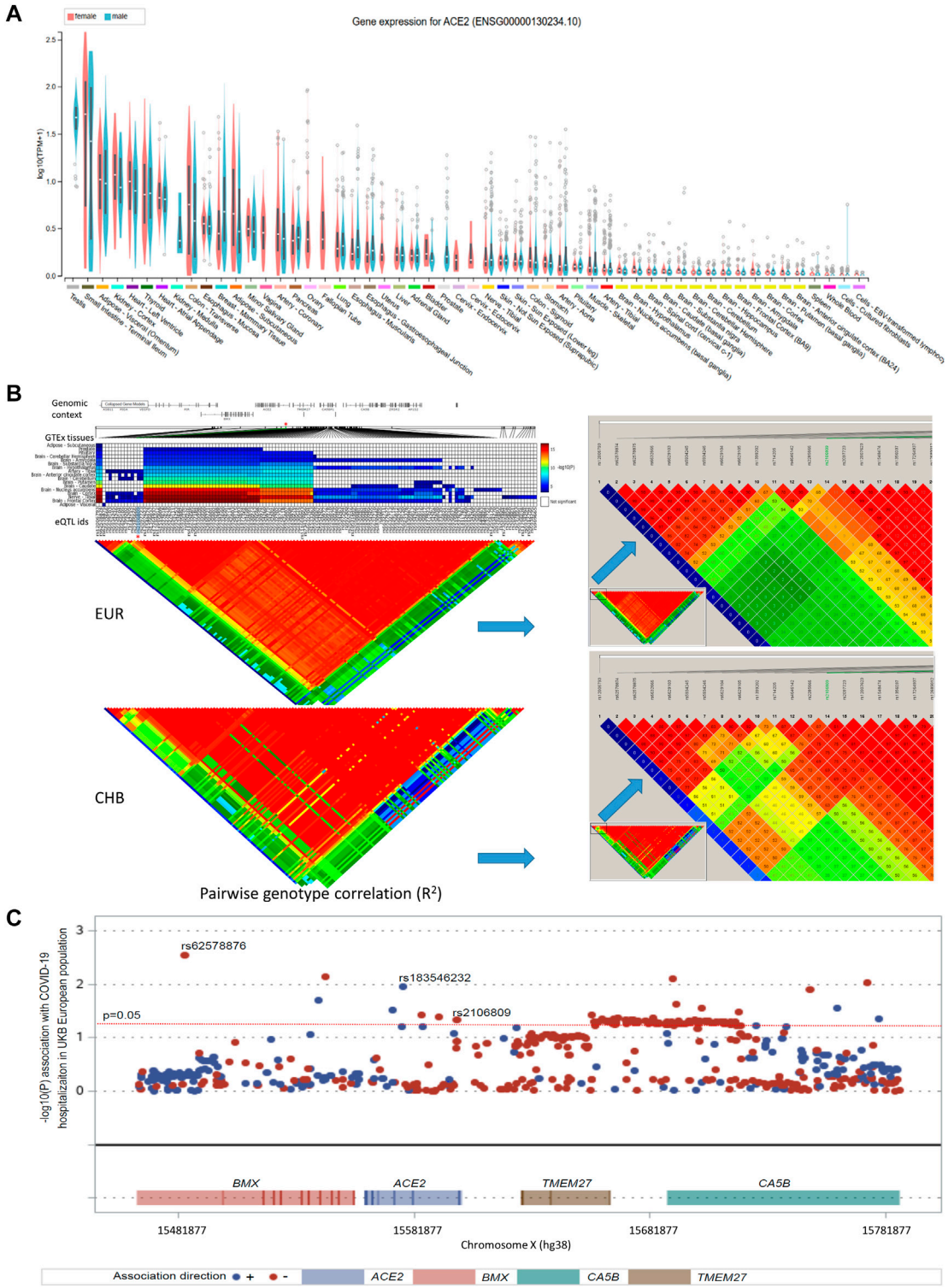



**FIGURE 1 |** *ACE2* expression among multiple GTEx tissues **(A)**, comparison of the LD pattern of *ACE2* GTEx expression quantitative trait loci (eQTLs) (minor allele frequency > 0.05) between EUR and CHB populations **(B)**, and the association between rs2106809 with COVID-19 hospitalization in European population **(C)**. In panel A, the *ACE2* expression by sex among multiple GTEx tissues are demonstrated in violin plots. In adipose-subcutaneous and whole blood, the *ACE2* expression is higher in females than in males; while in breast mammary tissue, the *ACE2* expression is lower in females than in males (all *p* values < 0.01; ANOVA test). In panel B, there are 140 common eQTLs located in the genomic region of chrX: 15 284 068–16 013 888 (hg38). Only rs2106809 is a common SNP located in the promoter region of *ACE2* (chrX: 15 597 069–15 607 069; hg38), which is highlighted by "*". rs2106809 displays the similar LD pattern with other GTEx eQTLs in both EUR and CHB populations. The pairwise genotype correlation between all GTEx eQTLs was determined using Haploview, and their corresponding $R^2$ values are shown with different colors, including red ($R^2$ > 0.8), yellow ($R^2$ between 0.5 and 0.8), blue ($R^2$ between 0.2 and 0.5), and green ($R^2$ < 0.2). In Panel C, rs2106809 along with two rare SNPs

(*Continued*)

**FIGURE 1 |** (rs62578876 and rs183546232) that are specific to EUR show an association with COVID-19 in the UK Biobank COVID-19 hospitalization GWAS of hospitalized cases ($n$ = 1 712) vs. not hospitalized or tested negative samples ($n$ = 56,988) in the EUR ancestry. rs2106809 minor allele G is protective against COVID-19 hospitalization ($p$ = 0.047; beta = –0.07; se = 0.035), while the minor alleles of other two rare SNPs, including rs62578876 and rs183546232 (an intronic SNP of *ACE2*), are protective ($p$ = 0.003; beta = –0.40; se = 0.13) and predisposing ($p$ = 0.011; beta = 0.32; se = 0.12) to COVID-19 hospitalization, respectively. EUR: European; CHB: Chinese Han in Beijing.

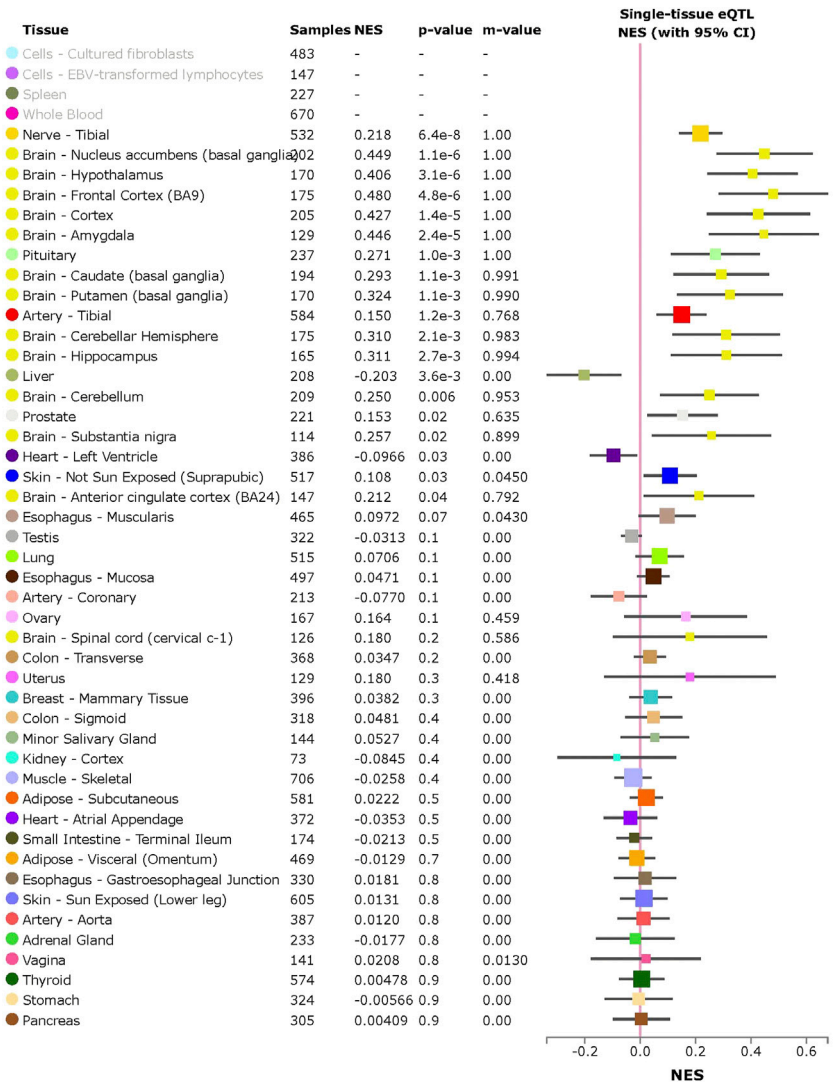



**FIGURE 2 |** Single-tissue eQTL analysis of the *ACE2* promoter SNP rs2106809 across 49 human tissues in GTEx. rs2106809 is a strong eQTL of *ACE2* across multiple brain tissues, as well as in nerve and artery tissues. In lung tissue, rs2106809 is marginally associated with the *ACE2* expression ($p$ = 0.07). NES: normalized effect size.

luciferase assay suggested that rs2106809 was closely correlated with the differential *ACE2* expression.

## DISCUSSION

We mapped the eQTLs of *ACE2* that existed in the *ACE2* promoter in EUR and CHB populations and verified the possible regulatory function of one candidate SNP rs2106809, which is a brain eQTL of *ACE2*. In our detailed examination of the COVID-19 hospitalization GWAS of the EUR ancestry from the UK Biobank (**Figure 1C**), we revealed the minor allele G of rs2106809 is protective against COVID-19 hospitalization, although its association was only nominally significant ($p$ = 0.047). A recent epidemiological investigation from China demonstrated that among COVID-19 cases, there were 37.6%

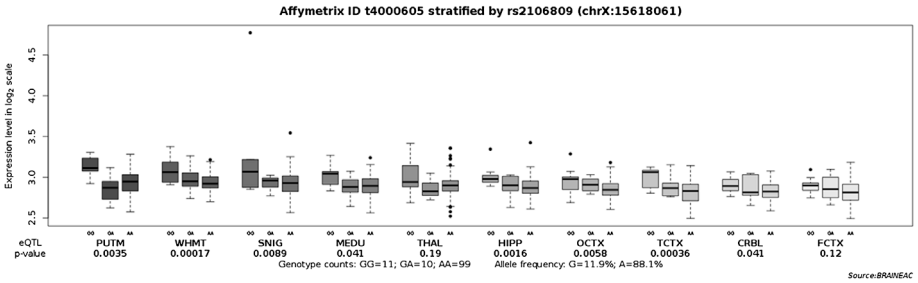



**FIGURE 3 |** Replication of the associations of the *ACE2* promoter SNP rs2106809 with the *ACE2* expression across multiple brain tissues in the UK brain eQTL database, Braineac.

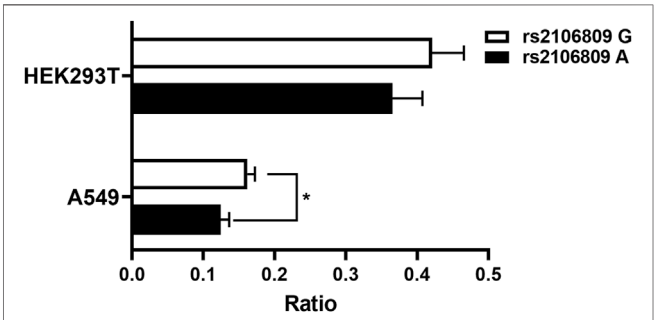



**FIGURE 4 |** Luciferase assay for the putative promoter of *ACE2* in HEK293T and A549 cells. Two constructs representing rs2106809 genotype GG and genotype AA harbored in the promoter region (~1.33kb; chrX: 15 599 868–15 601 196; hg38) were, respectively, inserted into the pGL3-Basic vector and then transfected into HEK293T and A549 cells separately. A *Renilla* luciferase vector was co-transfected as an internal control. The luciferase activity was measured at 24 h post transfection. The ratio of firefly to *Renilla* bioluminescence is presented as the mean + SD of triplicated transfection in one representative experiment (*: $p < 0.05$).

hospitalized COVID-19 patients diagnosed with long COVID [23], and another study from Faroe Islands, part of the Kingdom of Denmark, reported that around 50% of non-hospitalized COVID-19 patients developed into long COVID [24]. We hypothesized that severe COVID-19 and long COVID are two different phenotypes; given the high frequency of the rs2106809 major allele A in EUR and CHB populations, further investigation is needed to determine whether the A allele of rs2106809 is associated with long COVID symptoms, such as "brain fog".

There were several interesting scientific literature reports focused on the *ACE2* SNP rs2106809. A multicentral clinical trial of 3,408 patients found that the rs2106809 major allele A conferred a 1.6-fold risk for hypertension in Chinese women [25]. The subsequent research that enrolled 647 Chinese Han patients concluded that the rs2106809 A allele was statistically associated with left ventricular hypertrophy with 2.0-fold risk [26]. The other report showed a remarkable relationship between the rs2106809 polymorphism and essential hypertension (EH) in 246 hypertensives and 274 normotensives from Odisha, India [27]. Notably, a clinical survey of 96 Chinese female EH patients found the circulating Ang (1–7) levels were significantly higher in patients carrying the rs2106809 G allele than those carrying the A allele [28], in line with the fact that rs2106809*G associates with a higher *ACE2* expression, and a higher ACE2 activity increases the amount of Ang [6]. Such a series of epidemiological investigations provided various solid pieces of evidence that the *ACE2* SNP rs2106809 possessed a tight relationship with EH in Chinese and Indian populations. The A allele preferred to confer a higher risk for EH and EH-related cardiovascular disease. On the contrary, the G allele preferred to confer ameliorated EH. However, a case study of 155 COVID-19 patients of Turkey demonstrated the *ACE2* rs2106809 polymorphism was not associated with the clinical severity of COVID-19 infection [29]. In our analysis of the European COVID-19 hospitalization GWAS performed among all SARS-CoV-2 tested samples (cases = 1 712 and controls = 56,988), rs2106809*G is protective against COVID-19 hospitalization. Since ACE2 is a cellular receptor of SARS-CoV-2 and is also involved in hypertension that delays viral clearance and exacerbates airway hyperinflammation in patients with COVID-19 [30], this conclusion may need to be verified with more COVID-19 cases and among multiple human populations, such as CHB, EUR, and Yoruba in Ibadan, Nigeria.

There are several limitations that should be mentioned. First, our current study did not present an epidemiological investigation on the relationship between severe COVID-19 and long COVID in CHB, which needs to be investigated further. Second, our data did not include the COVID-19 or long COVID patients who carry this SNP; thus, we could not examine the association between the SNP with the two COVID-19 phenotypes in CHB. Third, the 1,328-bp promoter region of *ACE2* used in our study was synthesized based on reference sequence from GenBank (chrX: 15 599 868—15 601 196; hg38); however, such a segment from real samples of patients with EH, severe COVID-19, or long COVID, might offer more valuable information to this study. Fourth, the cells used in luciferase assay did not include the astrocyte cell line such as HMC3 [31]. The result of the dual luciferase reporter gene assay could not fully represent the real *ACE2* expression across different tissues. Finally, more SNPs of *ACE2*, particularly for population-specific rare SNPs, should be assessed to find out whether they are associated with severe COVID-19 or long COVID.

Despite these limitations, our findings provide solid information on the correlation between rs2106809*G and the remarkably higher expression level of *ACE2* in cerebral tissue, the

site where brain fog was characterized in long COVID. We integrated the data from GTEx and Braineac to interpret the potential function of rs2106809 to the *ACE2* expression. Importantly, we experimentally confirmed rs2106809 regulates the *ACE2* expression. In conclusion, our results are based on solid evidence *via* bioinformatics analyses and the dual luciferase reporter gene assay that demonstrated the rs2106809 G allele was able to significantly increase the expression of *ACE2*. The identification of rs2106809 as an *ACE2* eQTL among multiple brain tissues supports that rs2106809 may be involved in long COVID, but further investigations are warranted.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## AUTHOR CONTRIBUTIONS

Y-SL, WL, KZ, and Z-SC conceived the study. Y-SL, FC, H-YS, H-ML, M-YG, and KZ performed the experiments. WL, LL, SY, KZ, and Z-SC processed the data. WL, LL, YC, KZ, and Z-SC analyzed and interpreted the data. Y-SL, LL, KZ, and Z-SC wrote and revised the manuscript.

## FUNDING

This project was partly supported by The Science and Technology Supporting Plan of Scientific and Technological Department of Guizhou Province ([2020]4Y163 and [2020] 4Y164), the Science and Technology Program of Guiyang Science and Technology Bureau ([2020]-10-2), the Science and Technology Project of Guizhou Provincial Health Commission (GZWKJ2021-428), the Cultivation Project of the National Natural Science Foundation of China of the Affiliated Hospital of Guizhou Medical University, and Guizhou Medical University (I-2020-06 and 20NSP010).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.931562/full#supplementary-material

## REFERENCES

Letko, M., Marzi, A., and Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* 5 (4), 562–569. doi:10.1038/s41564-020-0688-y

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., and Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579 (7798), 270–273. doi:10.1038/s41586-020-2012-7

Tai, W., He, L., Zhang, X., Pu, J., Voronin, D., Jiang, S., Zhou, Y., and Du, L. (2020). Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol* 17 (6), 613–620. doi:10.1038/s41423-020-0400-4

Crackower, M. A., Sarao, R., Oudit, G. Y., Yagil, C., Kozieradzki, I., Scanga, S. E., Oliveira-dos-Santos, A. J., da Costa, J., Zhang, L., Pei, Y., Scholey, J., Ferrario, C. M., Manoukian, A. S., Chappell, M. C., Backx, P. H., Yagil, Y., and Penninger, J. M. (2002). Angiotensin-converting enzyme 2 is an essential regulator of heart function. *Nature* 417 (6891), 822–828. doi:10.1038/nature00786

Patel, V. B., Zhong, J.-C., Grant, M. B., and Oudit, G. Y. (2016). Role of the ACE2/Angiotensin 1-7 Axis of the Renin-Angiotensin System in Heart Failure. *Circ Res* 118 (8), 1313–1326. doi:10.1161/circresaha.116.307708

Miesbach, W (2020). Pathological role of Angiotensin II in severe COVID-19. *TH Open* 4 (2), e138–e44. doi:10.1055/s-0040-1713678

Lukiw, W. J., Pogue, A., and Hill, J. M. (2022). SARS-CoV-2 infectivity and neurological targets in the brain. *Cell Mol Neurobiol* 42 (1), 217–224. doi:10.1007/s10571-020-00947-7

Zhang, B.-Z., Chu, H., Han, S., Shuai, H., Deng, J., Hu, Y.-f., Gong, H.-r., Lee, A. C.-Y., Zou, Z., Yau, T., Wu, W., Hung, I. F.-N., Chan, J. F.-W., Yuen, K.-Y., and Huang, J.-D. (2020). SARS-CoV-2 infects human neural progenitor cells and brain organoids. *Cell Res* 30 (10), 928–931. doi:10.1038/s41422-020-0390-x

Libby, P., and Lüscher, T. (2020). COVID-19 is, in the end, an endothelial disease. *Eur Heart J* 41 (32), 3038–3044. doi:10.1093/eurheartj/ehaa623

Ackermann, M., Verleden, S. E., Kuehnel, M., Haverich, A., Welte, T., Laenger, F., Vanstapel, A., Werlein, C., Stark, H., Tzankov, A., Li, W. W., Li, V. W., Mentzer, S. J., and Jonigk, D. (2020). Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. *N Engl J Med* 383 (2), 120–128. doi:10.1056/nejmoa2015432

Jarrott, B, Head, R, Pringle, KG, Lumbers, ER, and Martin, JH (2022). "LONG COVID"—A hypothesis for understanding the biological basis and pharmacological treatment strategy. *Pharmacol Res Perspect* 10 (1), e00911. doi:10.1002/prp2.911

Yong, S. J. (2021). Long COVID or post-COVID-19 syndrome: putative pathophysiology, risk factors, and treatments. *Infectious Diseases* 53 (10), 737–754. doi:10.1080/23744235.2021.1924397

Datta, SD, Talwar, A, and Lee, JT (2020). A proposed framework and timeline of the spectrum of disease due to SARS-CoV-2 infection: Illness beyond acute infection and public health implications. *JAMA* 324 (22), 2251–2. doi:10.1001/jama.2020.22717

Yong, S. J. (2021). Persistent Brainstem Dysfunction in Long-COVID: A Hypothesis. *ACS Chem. Neurosci.* 12 (4), 573–580. doi:10.1021/acschemneuro.0c00793

Marshall, M. (2021). COVID and the brain: researchers zero in on how damage occurs. *Nature* 595 (7868), 484–485. doi:10.1038/d41586-021-01693-6

Ribeiro, D. E., Oliveira-Giacomelli, Á., Glaser, T., Arnaud-Sampaio, V. F., Andrejew, R., Dieckmann, L., Baranova, J., Lameu, C., Ratajczak, M. Z., and Ulrich, H. (2020). Hyperactivation of P2X7 receptors as a culprit of COVID-19 neuropathology. *Mol Psychiatry* 26 (4), 1044–1059. doi:10.1038/s41380-020-00965-3

The GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369 (6509), 1318–1330. doi:10.1126/science.aaz1776

Chang, CC, Chow, CC, Tellier, LC, Vattikuti, S, Purcell, SM, and Lee, JJ (2014). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4, 7. doi:10.1186/s13742-015-0047-8

Clarke, L., Zheng-Bradley, X, Zheng-Bradley, X., Smith, R., Kulesha, E., Xiao, C., Toneva, I., Vaughan, B., Preuss, D., Leinonen, R., Shumway, M., Sherry, S., and Flicek, P. (2012). The 1000 Genomes Project: data management and community access. *Nat Methods* 9 (5), 459–462. doi:10.1038/nmeth.1974

Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21 (2), 263–265. doi:10.1093/bioinformatics/bth457

Thibord, F., Chan, M. V., Chen, M.-H., and Johnson, A. D. (2022). A year of COVID-19 GWAS results from the GRASP portal reveals potential genetic risk factors. *Human Genetics and Genomics Advances* 3 (2), 100095. doi:10.1016/j.xhgg.2022.100095

Cheng, Z., Zhou, J., To, K. K.-W., Chu, H., Li, C., Wang, D., Yang, D., Zheng, S., Hao, K., Bossé, Y., Obeidat, M. e., Brandsma, C.-A., Song, Y.-Q., Chen, Y., Zheng, B.-J., Li, L., and Yuen, K.-Y. (2015). Identification ofTMPRSS2as a Susceptibility Gene for Severe 2009 Pandemic A(H1N1) Influenza and A(H7N9) Influenza. *J Infect Dis.* 212 (8), 1214–1221. doi:10.1093/infdis/jiv246

Xiong, Q, Xu, M, Li, J, Liu, Y, Zhang, J, Xu, Y, and Dong, W (2020). Clinical sequelae of COVID-19 survivors in Wuhan, China: a single-centre longitudinal study. *Clin Microbiol Infect* 27 (1), 89–95. doi:10.1016/j.cmi.2020.09.023

Petersen, M. S., Kristiansen, M. F., Hanusson, K. D., Danielsen, M. E., á Steig, B., Gaini, S., Strøm, M., and Weihe, P. (2020). Long COVID in the Faroe Islands: a longitudinal study among nonhospitalized patients. *Clin Infect Dis* 73 (11), e4058–e4063. doi:10.1093/cid/ciaa1792

Fan, X., Wang, Y., Sun, K., Zhang, W., Yang, X., Wang, S., Zhen, Y., Wang, J., Li, W., Han, Y., Liu, T., Wang, X., Chen, J., Wu, H., and Hui, R. (2007). Polymorphisms of *ACE2* gene are associated with essential hypertension and antihypertensive effects of Captopril in women. *Clin Pharmacol Ther* 82 (2), 187–196. doi:10.1038/sj.clpt.6100214

Fan, Z., Wu, G., Yue, M., Ye, J., Chen, Y., Xu, B., Shu, Z., Zhu, J., Lu, N., and Tan, X. (2019). Hypertension and hypertensive left ventricular hypertrophy are associated with ACE2 genetic polymorphism. *Life Sciences* 225, 39–45. doi:10.1016/j.lfs.2019.03.059

Patnaik, M., Pati, P., Swain, S. N., Mohapatra, M. K., Dwibedi, B., Kar, S. K., and Ranjit, M. (2013). Association of angiotensin-converting enzyme and angiotensin-converting enzyme-2 gene polymorphisms with essential hypertension in the population of Odisha, India. *Annals of Human Biology* 41 (2), 145–152. doi:10.3109/03014460.2013.837195

Liu, D., Chen, Y., Zhang, P., Zhong, J., Jin, L., Zhang, C., Lin, S., Wu, S., and Yu, H. (2016). Association between circulating levels of ACE2-Ang-(1-7)-MAS axis and ACE2 gene polymorphisms in hypertensive patients. *Medicine (Baltimore)* 95 (24), e3876. doi:10.1097/md.0000000000003876

Çelik, SK, GenÇ, GÇ, Pişkin, N, Açikgöz, B, Altinsoy, B, İşsiz, BK, et al. (2021). Polymorphisms of ACE (I/D) and ACE2 receptor gene (Rs2106809, Rs2285666) are not related to the clinical course of COVID-19: A case study. *J Med Virol* 93 (10), 5947–52. doi:10.1002/jmv.27160

Trump, S., Lukassen, S., Anker, M. S., Chua, R. L., Liebig, J., Thürmann, L., Corman, V. M., Binder, M., Loske, J., Klasa, C., Krieger, T., Hennig, B. P., Messingschlager, M., Pott, F., Kazmierski, J., Twardziok, S., Albrecht, J. P., Eils, J., Hadzibegovic, S., Lena, A., Heidecker, B., Bürgel, T., Steinfeldt, J., Goffinet, C., Kurth, F., Witzenrath, M., Völker, M. T., Müller, S. D., Liebert, U. G., Ishaque, N., Kaderali, L., Sander, L.-E., Drosten, C., Laudi, S., Eils, R., Conrad, C., Landmesser, U., and Lehmann, I. (2020). Hypertension delays viral clearance and exacerbates airway hyperinflammation in patients with COVID-19. *Nat Biotechnol* 39 (6), 705–716. doi:10.1038/s41587-020-00796-1

Jeong, GU, Lyu, J, Kim, KD, Chung, YC, Yoon, GY, Lee, S, et al. (2022). SARS-CoV-2 infection of microglia elicits proinflammatory activation and apoptotic cell death. *Microbiol Spectr*, e0109122. doi:10.1128/spectrum.01091-22

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 16 August 2022
DOI 10.3389/fgene.2022.928466



OPEN ACCESS

EDITED BY
Zhongshan Cheng,
St. Jude Children's Research Hospital, United States

REVIEWED BY
Andrei Lobiuc,
Ștefan cel Mare University of Suceava, Romania
Yi Cai,
Shenzhen University, China

*CORRESPONDENCE
Joseph T. Glessner,
glessner@chop.edu

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 25 April 2022
ACCEPTED 13 July 2022
PUBLISHED 16 August 2022

CITATION
Glessner JT, Chang X, Mentch F, Qu H, Abrams DJ, Thomas A, Sleiman PMA and Hakonarson H (2022), COVID-19 in pediatrics: Genetic susceptibility. *Front. Genet.* 13:928466. doi: 10.3389/fgene.2022.928466



# COVID-19 in pediatrics: Genetic susceptibility

Joseph T. Glessner[1,2]*, Xiao Chang[1], Frank Mentch[1], Huiqi Qu[1], Debra J. Abrams[1], Alexandria Thomas[1], Patrick M. A. Sleiman[1,2] and Hakon Hakonarson[1,2]

[1]Children's Hospital of Philadelphia, Philadelphia, PA, United States, [2]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

The uptick in SARS-CoV-2 infection has resulted in a worldwide COVID-19 pandemic, which has created troublesome health and economic problems. We performed case–control meta-analyses in both African and European ethnicity COVID-19 disease cases based on laboratory test and phenotypic criteria. The cases had laboratory-confirmed SARS-CoV-2 infection. We uniquely investigated COVID infection genetics in a pediatric population. Our cohort has a large African ancestry component, also unique to our study. We tested for genetic variant association in 498 cases vs. 1,533 controls of African ancestry and 271 cases vs. 855 controls of European ancestry. We acknowledge that the sample size is relatively small, owing to the low prevalence of COVID infection among pediatric individuals. COVID-19 cases averaged 13 years of age. Pediatric genetic studies enhance the ability to detect genetic associations with a limited possible environment impact. Our findings support the notion that some genetic variants, most notably at the SEMA6D, FMN1, ACTN1, PDS5B, NFIA, ADGRL3, MMP27, TENM3, SPRY4, MNS1, and RSU1 loci, play a role in COVID-19 infection susceptibility. The pediatric cohort also shows nominal replication of previously reported adult study results: CCR9, CXCR6, FYCO1, LZTFL1, TDGF1, CCR1, CCR2, CCR3, CCR5, MAPT-AS1, and IFNAR2 gene variants. Reviewing the biological roles of genes implicated here, NFIA looks to be the most interesting as it binds to a palindromic sequence observed in both viral and cellular promoters and in the adenovirus type 2 origin of replication.

KEYWORDS

**GWAS, genome-wide association study, diverse populations, pediatrics, statistical genetics and genomics, COVID-19**

## Introduction

The ongoing coronavirus disease 2019 (COVID-19) pandemic has posed an extraordinary threat to global public health. COVID-19 is caused by the infection of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Wu and McGoogan, 2020). SARS-COV-2 is not as virulent as severe acute respiratory syndrome (SARS), and a large number of patients are asymptomatic or suffer only mild symptoms (Bai et al., 2020). The first genome-wide association study (GWAS) of COVID-19 reported two genomic loci associated with severe COVID-19, indicating a strong genetic influence on the

TABLE 1 COVID-19 pediatric case–control cohorts analyzed.

| | Total queried phenotype | Total (pre-QC) | EUR (pre-QC) | AFR (pre-QC) | Total (post-QC) | EUR (post-QC) | AFR (post-QC) |
|---|---|---|---|---|---|---|---|
| Case | 994 | 841 | 286 | 555 | 769 | 271 | 498 |
| Control | 2965 | 2490 | 873 | 1617 | 2388 | 855 | 1533 |

severity of COVID-19 (Severe Covid et al., 2020). COVID-19 Host Genetics Initiative performed the largest GWAS in adults to date including 49,562 patients from 46 studies across 19 countries (Initiative, 2020; Kousathanas et al., 2022). They reported 13 genome-wide significant loci that are associated with SARS-CoV-2 infection or severe manifestations of COVID-19.

To date, a number of GWASs on COVID-19 have been reported (Initiative, 2020; Severe Covid et al., 2020; Hu et al., 2021a; Kosmicki et al., 2021a; Pairo-Castineira et al., 2021a; Shelton et al., 2021a; Dubé et al., 2021; Ma et al., 2021; Mousa et al., 2021; Patrick et al., 2021; Peloso et al., 2021; Chamnanphon et al., 2022; Horowitz et al., 2022; Kousathanas et al., 2022; Roberts et al., 2022). The research subjects included European (Initiative, 2020; Severe Covid et al., 2020; Hu et al., 2021a; Kosmicki et al., 2021a; Pairo-Castineira et al., 2021a; Shelton et al., 2021a; Dubé et al., 2021; Ma et al., 2021; Mousa et al., 2021; Patrick et al., 2021; Peloso et al., 2021; Horowitz et al., 2022; Kousathanas et al., 2022; Roberts et al., 2022), African (Kosmicki et al., 2021a; Shelton et al., 2021a; Peloso et al., 2021; Horowitz et al., 2022), East Asian (Mousa et al., 2021; Horowitz et al., 2022), South Asian (Kosmicki et al., 2021a; Mousa et al., 2021; Chamnanphon et al., 2022; Horowitz et al., 2022), and Latin American (Shelton et al., 2021a; Horowitz et al., 2022) populations. The reported studies were all performed on adult populations. In contrast to adults, most of the children infected with COVID-19 presented only mild or moderate symptoms (De Souza et al., 2020), suggesting that different genetic mechanisms from adults may exist. As observed in the GWAS on asthma, 20% of susceptibility loci are pediatric-specific (Ferreira et al., 2019). Due to the gene–environment interaction, some genetic factors may affect sensitivity to environmental factors and vice versa (D'amato et al., 2005). In addition, environmental exposures change over years of life. To date, GWAS of COVID-19 has not been conducted on pediatric populations.

Here, we developed sensitive phenotyping query methods and matched suitable samples to genotyping data pre-QC and post-QC (Table 1). Variants quality controlled with an allele frequency >1%, SNP call rate genotype missingness <0.05, Hardy–Weinberg equilibrium deviation $p$-value > 1e-6, and imputation quality R2 > 0.3 were further assessed in African and European studies. Despite many active studies, the genetics impacting SARS-CoV-2 infection risk and progression severity remains poorly understood. The SNP-based associations were refined based on peaks of significance for contiguous SNPs and linkage disequilibrium (LD) of top significant SNP regions. Further work on larger cohorts is needed to better understand which traits (disease, health, and neuropsychiatric phenotypes) are genetically correlated and potentially causally associated with the infection of SARS-CoV-2. Tremendous worldwide COVID-19 genotype aggregation efforts have launched sample sizes of 49,562 patients with COVID-19 and 2 million controls (Niemi et al., 2021). PLINK23 (Purcell et al., 2007) software was leveraged for efficient quality filtering, statistical association, and review of results.

## Results

To limit the chance of spurious associations, implicated disease phenotypes associated with SNPs in LD (r2 > 0.8) with the top significant COVID-19 variants were reviewed. The inclusion of pediatric cases and controls from both European and African ancestries demonstrates the value of including data from diverse populations for characterizing genetic associations. Environmental, clinical, and social factors contribute to exposure and severity of COVID-19 (Docherty et al., 2020; Zhou et al., 2020) with host genetics also playing an important role. Here, we show genome-wide association meta-analysis results that consist of 498 pediatric cases vs. 1,533 controls of African ancestry and 271 pediatric cases vs. 855 controls of European ancestry (Tables 2, 3).

Details of genomic loci and observed significance are provided in LocusZoom (Pruim et al., 2010) plots (Figures 1, 2). Replicating a previously reported study (Roberts et al., 2020), a top significant locus in our results was within the 3p21.31 region associated with SARS-CoV-2 infection susceptibility (Table 2). We referenced cis-protein QTLs (Sun et al., 2018) to more deeply characterize the top significant loci. We used the European and African reference panel from TOPMed and the 1000 Genomes Project (Abecasis et al., 2010) to show LD between genetic variants. Genetic variants underlying COVID-19 susceptibility holds the potential to glean models of disease biology for therapeutic development, to extend new prevention and treatment options beyond the recent release of vaccines. Some of the most significantly associated SNPs (Table 2) overlap previously confirmed genetic associations as

TABLE 2 Replication of previous findings.

| Chr:Start–end (hg38/GRCh38) ($P < 0.05$) | Gene | Cohort | Lead SNP | P (E) | OR | Broad cohort | Broad lead SNP | Broad P | Broad beta |
|---|---|---|---|---|---|---|---|---|---|
| 3:45848457–45976785 | *CCR9*, *CXCR6*, *FYCO1*, and *LZTFL1* | EUR | 3:45961470:T:C | 2.55–3 | 0.729 | AFR + EUR META | 3:45848457:C:T | 4.25E-81 | 0.588 |
| 3:46610496–46610496 | *TDGF1* | EUR | 3:46610496:A:G | 5.02–3 | 3.006 | EUR META | 3:46610496:A:G | 4.99E-8 | 0.427 |
| 3:46108627–46374725 | *CCR1*, *CCR2*, *CCR3*, and *CCR5* | EUR | 3:46108627:C:T | 5.84–3 | 0.750 | AFR + EUR META | 3:46231218:A:C | 3.47E-20 | 0.304 |
| 17:45880713–45880713 | *MAPT-AS1* | AFR | 17:45880713:C:T | 5.92–3 | 1.331 | AFR META | 17:45880713:C:T | 2.68E-8 | -0.127 |
| 21:33238182–33238182 | *IFNAR2* | EUR | 21:33238182:T:C | 3.36–2 | 1.244 | EUR META | 21:33238182:T:C | 1.01E-11 | 0.128 |

TABLE 3 Novel findings of this study.

| Chr:Start–end (hg38/GRCh38) ($P < 5e-5$) | Gene | AFR_SNP | EUR_SNP | META_SNP | AFR_P | EUR_P | META_P | AFR OR | EUR OR | Meta OR |
|---|---|---|---|---|---|---|---|---|---|---|
| 15:47504866–47504920 | *SEMA6D* | 15:47504920:T:TG | NA | 15:47504866:C:G | 9.80E-06 | NA | NA | 2.300 | NA | NA |
| 15:33036296–33036318 | *FMN1* | 15:33036296:T:C | 15:33036296:T:C | 15:33036296:T:C | 3.17E-04 | 1.07E-02 | 1.04E-05 | 1.337 | 1.305 | 1.325 |
| 14:68910449–68910548 | *ACTN1* | 14:68910548:A:G | 14:68910449:T:C | 14:68910548:A:G | 2.61E-01 | 1.09E-05 | 1.09E-02 | 1.169 | 5.140 | 1.389 |
| 13:32665329–32665331 | *PDS5B* | 13:32665331:T:C | NA | 13:32665329:A:C | 1.19E-05 | NA | NA | 2.349 | NA | NA |
| 1:61414689–61414750 | *NFIA* | 1:61414689:T:C | 1:61414750:A:G | 1:61414689:T:C | 1.80E-05 | 4.63E-01 | 1.24E-03 | 0.692 | 1.103 | 0.792 |
| 4:61421195–61421214 | *ADGRL3* | 4:61421195:T:A | NA | 4:61421195:T:A | 2.14E-05 | NA | NA | 2.783 | NA | NA |
| 11:102697419–102697493 | *MMP27* | 11:102697419:G:A | 11:102697419:G:A | 11:102697419:G:A | 1.90E-03 | 3.66E-03 | 2.23E-05 | 1.299 | 1.357 | 1.321 |
| 4:182739578–182739648 | *TENM3* | NA | 4:182739648:G:A | 4:182739578:C:A | NA | 2.43E-05 | NA | NA | 1.597 | NA |
| 5:142320157–142320171 | *SPRY4* | 5:142320157:T:C | NA | 5:142320157:T:C | 3.08E-05 | NA | NA | 1.999 | NA | NA |
| 15:56458970–56459025 | *MNS1* | 15:56458970:A:G | 15:56458970:A:G | 15:56458970:A:G | 2.05E-03 | 5.58E-03 | 3.39E-05 | 0.739 | 0.736 | 0.738 |
| 10:16597265–16603587 | *RSU1* | 10:16597265:G:A | NA | 10:16597265:G:A | 3.76E-05 | NA | NA | 2.193 | NA | NA |

mentioned previously (David et al., 2020; Pairo-Castineira et al., 2021b).

## Discussion

Among the reported genes by the previous GWASs (Initiative, 2020; Severe Covid et al., 2020; Hu et al., 2021a; Kosmicki et al., 2021a; Pairo-Castineira et al., 2021a; Shelton et al., 2021a; Dubé et al., 2021; Ma et al., 2021; Mousa et al., 2021; Patrick et al., 2021; Peloso et al., 2021; Chamnanphon et al., 2022; Horowitz et al., 2022; Kousathanas et al., 2022; Roberts et al., 2022), interestingly, genes related to cytokine receptor activity (GO:0004896) are significantly enriched (FDR-corrected $p$ = 0.017) by gene-set enrichment analysis using the WebGestalt (WEB-based Gene SeT AnaLysis Toolkit) web tool (Wang et al.,

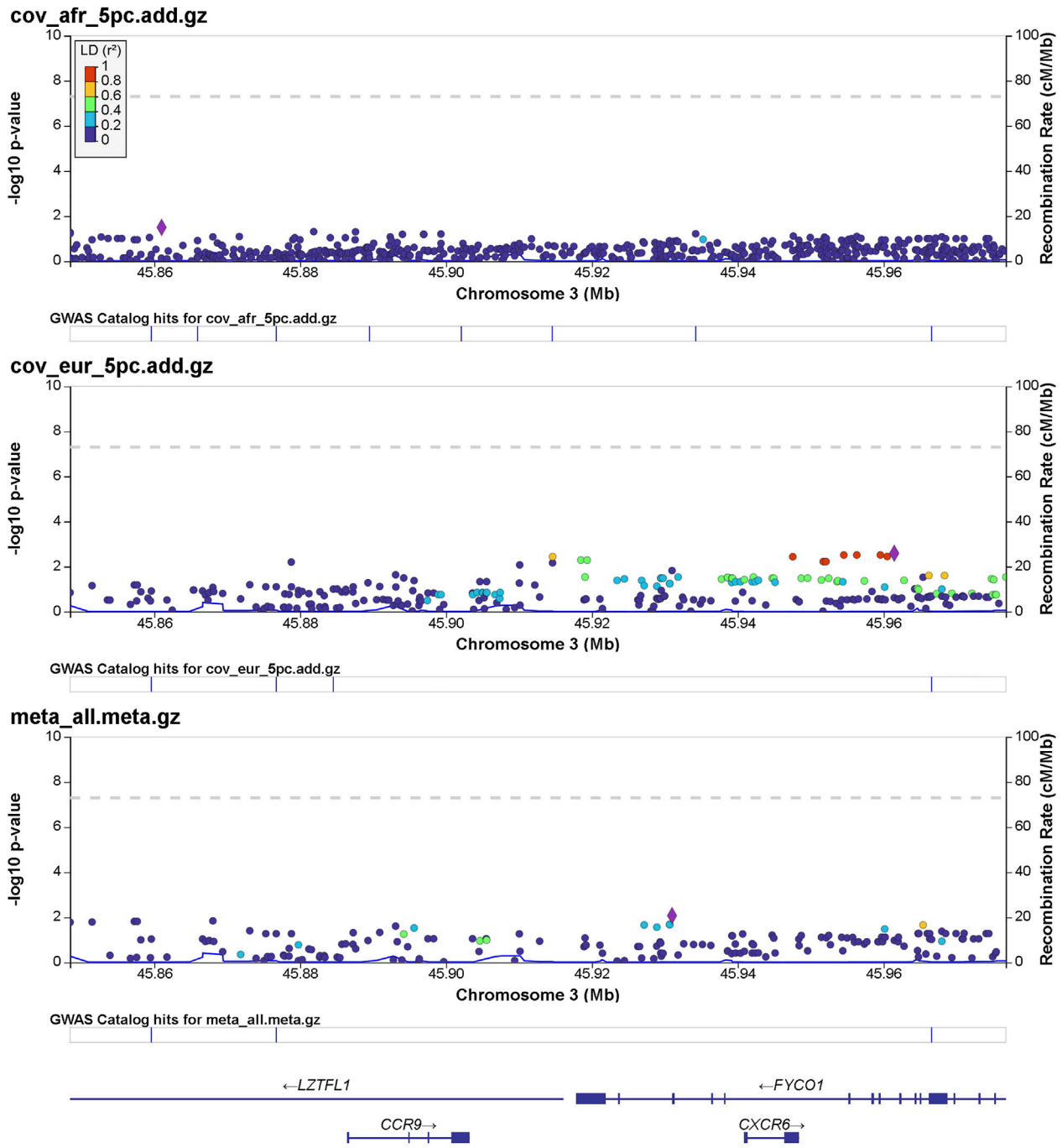



FIGURE 1
Regional significance plot of the 3p21.31 region by LocusZoom (Pruim et al., 2010). The genes *CCR9*, *CXCR6*, *FYCO1*, and *LZTFL1* are included in this region. The peak of significance is from the SNP 3:45961470:T:C (rs1601867) at the intronic region of *FYCO1*.

2013). The genes include C-C motif chemokine receptor 1 (*CCR1*), C-C motif chemokine receptor 3 (*CCR3*), C-C motif chemokine receptor 9 (*CCR9*), C-X-C motif chemokine receptor 6 (*CXCR6*), interferon alpha and beta receptor subunit 2 (*IFNAR2*), interleukin 10 receptor subunit beta (*IL10RB*), LIF receptor alpha (*LIFR*), and X-C motif chemokine receptor 1 (*XCR1*). As shown in Table 2, the genes related to cytokine receptor activity, including *CCR1*, *CCR3*, *CCR9*, *CXCR6*, and *IFNAR2*, are also identified in this study. Chemokine receptors are G protein-coupled receptors and bind chemokines to mediate

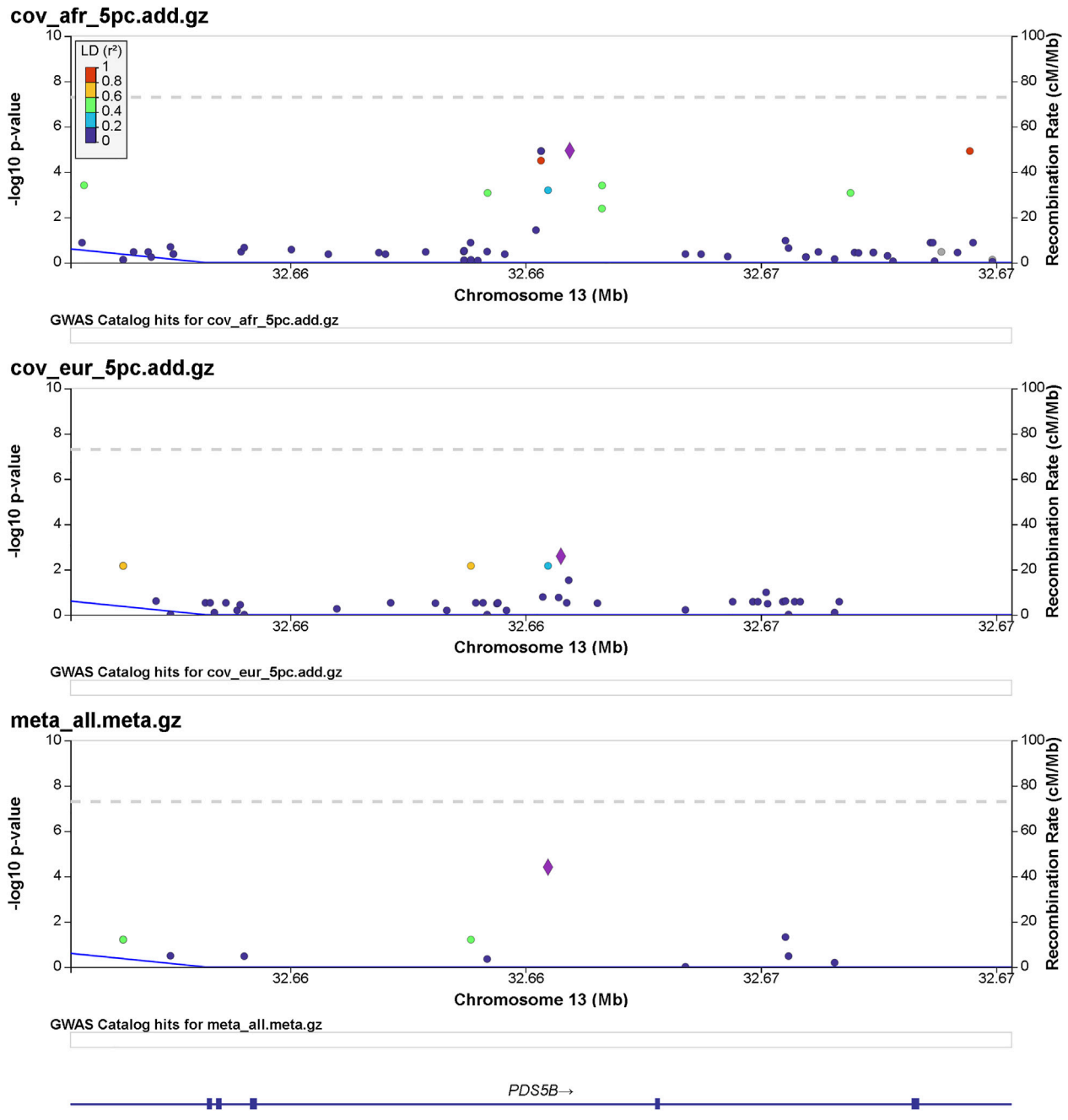



FIGURE 2
Regional significance plot of the 13q13.1 region by LocusZoom (Pruim et al., 2010). The PDS5 cohesin-associated factor B gene (*PDS5B*) maps to this region. The peak of significance by the meta-analysis is from the SNP 13:32665329:A:C (rs144965594) at the intronic region of *PDS5B*.

cell migration in immune surveillance and inflammation (Allen et al., 2007). *CCR1*, *CCR3*, and *CCR9* encode receptors of the C-C family chemokines with two adjacent N-terminal cysteine residues. There are 28 C-C chemokines and 10 C-C chemokine receptors identified to date (White et al., 2013). CCR1 and CCR3 bind to multiple CC chemokines with critical roles in inflammation (Pakianathan et al., 1997). *CCR9* encodes the receptor of C-C motif chemokine ligand 25 (CCL25), with a role in the development of T cell in thymus (Vicari et al., 1997). CXCR6 has a protein structure close to CCRs and binds to the ligand CXCL16 of the CXC family chemokines with one amino acid between the two N-terminal cysteine residues

(Day et al., 2009). CXCR6 may have important roles in T-cell recruitment to the lung in COVID-19 infection, as suggested by its high expression in the lung (Day et al., 2009). *IFNAR2* encodes subunit 2 of the interferon-α/β receptor (IFNAR) (Lutfalla et al., 1995), mediating the roles of type 1 interferons α and β in innate immune response to viral infections (Biron, 1998). In addition to the roles of the cytokines in anti-viral immunity and inflammation (Bartee and McFadden, 2013), these genes may also be involved in cytokine storm in severe COVID-19 (Hu et al., 2021b).

We show here 13 ethnicity-specific and/or meta-analysis variants that pass the top rank and nominal significance threshold ($p$ < 5e-5). Several genome-wide association studies investigating case and control samples with many SNP genotypes, which have associated certain SNPs (David et al., 2020; Roberts et al., 2020; Pairo-Castineira et al., 2021b; Shelton et al., 2021b) to COVID-19, have indicated support for several genomic loci associated with COVID-19 susceptibility and severity; the strongest association related to severity is at the 3p21.31 locus (David et al., 2020; Roberts et al., 2020; Kosmicki et al., 2021b; Pairo-Castineira et al., 2021b; Shelton et al., 2021b). Two separate loci in the 3p21.31 region include genes prioritized from different methods and signals.

A number of loci identified in this study have not been reported in the previous GWASs on adults (Initiative, 2020; Severe Covid et al., 2020) (Table 3). Interestingly, five of these loci, i.e., *ACTN1*, *PDS5B*, *SEMA6D*, *SPRY4*, and *TENM3*, have been reported of association with the genetic susceptibility of asthma (Yucesoy et al., 2015; Almoguera et al., 2017; Demenais et al., 2018; Olafsdottir et al., 2020). As reviewed by Adir et al. (2021), asthma may impose important factors related to SARS-CoV-2 infection and disease severity, for e.g., Th2-high inflammation in asthma may reduce the risk of SARS-Cov-2 infection and chronic use of systemic corticosteroids (ICS) is associated with poor outcomes of COVID-19.

Further population sampling and genotyping of COVID-19 and related phenotypes is warranted to further characterize susceptibility, severity, or mortality in the future, guided by Centers for Disease Control enumeration of prior medical conditions linked with COVID-19 severity (CDC, 2021) or traits linked with risk of COVID-19 mortality by OpenSAFELY (Williamson et al., 2020).

## Limitations

This study has limitations. First, the controls were determined based on the records from our EMR data collected in October 2021. The controls might get infected at a later time point. As COVID-19 is an infectious disease, this limitation exits in all COVID GWASs. Second, the sample size is relatively small. Future studies with a larger sample size may identify genetic loci of COVID-19, especially associated with pediatric populations. Third, this study was performed on a unique pediatric cohort of COVID-19. However, we acknowledge that follow-up analyses for the novel loci described in this study are warranted.

# Conclusion

Our analyses report 17 independent genome-wide nominal significance SNPs with neighboring higher than expected $p$-value SNPs (6 were replication of previous findings and 11 were novel findings), defining COVID-19 loci with a threshold of $p$ < 5 E-5 (unadjusted for multiple testing). A unique and challenging aspect is variable progression of SARS-CoV-2 infection, ranging from acute to severe clinical presentations of viral pneumonia or acute respiratory distress syndrome (Buitrago-Garcia et al., 2020). Additional cohorts and studies will be needed to effectively leverage biological and clinical yield potential of these genetic associations. We applied covariates including age, sex, and the five first principal components to properly account for these population characteristics in addition to the SNP genotypes. For all 13 loci, we compared the lead variant (strongest association $p$-value) and odds ratios (ORs) for the risk allele across different ethnic groups. Four of the thirteen genome-wide nominal significant loci showed similar trends in SARS-CoV-2 infection (i.e., disease susceptibility). Host-specific genetic variants identified here hold the potential to characterize biological interaction and function, informing therapeutic possibilities, and delineate causal link of risk factors in the environment for SARS-CoV-2 infection and prognosis. These findings indicate a multi-gene and multi-function mechanism to be more fully characterized by future studies.

# Methods

## Subjects

All subjects were recruited using CHOP Institutional Review Board-approved protocols. The SARS-CoV-2 infection-positive group had laboratory-confirmed SARS-CoV-2 infection, electronic health record ICD coding, or was self-reported by a survey, along with the annotation whether symptoms of severity were observed. The Diagnosis and Treatment Protocol for Novel Coronavirus Pneumonia was used to classify illness severity and hospitalization observations (Wei, 2020). Controls were populations based on data of negative SARS-CoV-2 infection and negative COVID-19 status. Genetic-ancestry-matched control individuals for the COVID-19-positive cases were matched with population-based cohorts based on nearest PCA distance. Control individuals were infection-negative based on questionnaire/electronic health record-based database queries.

## Genotyping

Samples were genotyped using the Illumina Infinium BeadChip Global Screening Array (GSA). SNP genotypes and variant allele naming were coordinated to human genome build hg38/GRCh38 and referenced with respect to gnomAD 3.0 genomes (Karczewski et al., 2020) by matching SNPs via variant matching by testing strand flip and allele order switches. To gain additional resolution of genotyping data for these samples, we performed imputation on the TOPMed Imputation Server at https://imputation.biodatacatalyst.nhlbi.nih.gov/.

## African ancestry COVID case–control

A total of 367,556 genetic variants passed filters and quality control and thus were tested for association to COVID-19-infected phenotype individuals. A total of 2,172 individuals (1,017 males and 1,155 females) were included. The total genotyping rate in samples remaining after quality control was 0.997553. The number of individuals who passed filters and QC was 2,031. Among the remaining phenotypes, 498 were cases and 1,533 were controls.

## European ancestry COVID case–control

A total of 486,109 variants were assessed that met filter and QC standards. A total of 1,159 individuals (643 males and 516 females) were included. The total genotyping rate in the remaining samples was 0.998073. Altogether, 486,109 variants and 1,126 individuals passed filters and QC. Among the remaining phenotypes, 271 were cases and 855 were controls.

## African ancestry and European ancestry COVID meta-analysis

A meta-analysis including 14,336,851 variants was processed, and 3,854,317 variants had non-NA *p*-values. Several known clinical factors of the host track closely to disease severity such as older age, being male, and larger body mass index (Docherty et al., 2020), but these factors are not sufficient to model disease severity variability. These findings support prioritizing candidate genes along with future functional characterization to refine the genes. Control samples were chosen based on principal component analysis-driven genetic ancestry-matching samples without known SARS-CoV-2 infection.

## Data analysis

To prioritize candidate gene regions reported in this study, we used both locus-based and similarity-based methods. We report the raw *p*-values and odds ratios for each lead variant with closely adjacent nominal significance variants along with the nearest gene. In an effort to better characterize the biological mechanism of observed variants at each locus, we prioritized candidate genes and referenced knowledge from results from related diseases and traits. The relevant stage of disease from SARS-CoV-2 infection to progression and outcome was a factor considered in the modeling of gene roles in associated loci.

We used PLINK2 (Chang et al., 2015) to apply sample and SNP quality control thresholds, in association with an additive effect model, applying the top five principal components as covariates and conducting meta-analysis.

We conducted GWAS statistical analyses with the tool Scalable and Accurate Implementation of GEneralized (SAIGE) mixed model (Zhou et al., 2018) on all autosomes and chromosome X. Our 17q21.31 replication top finding overlapping MAPT-AS1 (KANSL1 150 kb away) coincides with a deeply studied locus with structural variants including a large megabase recurrent inversion deviating from the reference H1 to the inverted H2 form that has been selected positively in European ancestry persons (Stefansson et al., 2005; Boettger et al., 2012). SAIGE features robust modeling of sample relatedness and case–control count differences. The genetic identity of a person influences viral infection susceptibility and response. We sought to characterize the 13 nominal significant loci for potential to fulfill roles in risk and progression following infection. We used the Cochran's Q measure (Cochran, 1954; Evangelou and Ioannidis, 2013) using the two analyses effect sizes vs. the meta-analysis effect size (weighted by inverse variance of effect sizes) sum of squared differences.

# Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ebi.ac.uk/gwas/studies/, GCST011074 and GCP000381.

# Ethics statement

The studies involving human participants were reviewed and approved by the Children's Hospital of Philadelphia Institutional Review Board. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

# Author contributions

HH and PS contributed to the conception and design of the study. DA, AT, and FM organized and queried the phenotype database. XC performed the statistical analysis. JG wrote the first

draft of the manuscript. HQ, HH, and XC wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## Funding

This work was supported in part by U01-HG006830 (NHGRI-sponsored eMERGE Network), CHOP's Endowed Chair in Genomic Research and Institutional Development Award (Hakonarson), and by sponsored research agreement from Aevi Genomic Medicine, Inc. (HH).

## Acknowledgments

The authors thank the patients recruited in the study for their consent for biobank inclusion, allowing derived genotypes and phenotypes for conducting the analysis. They also thank the developers of the statistical model, which helped in efficient computational tools used.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., and Gibbs, R. A. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073. doi:10.1038/nature09534

Adir, Y., Saliba, W., Beurnier, A., and Humbert, M. (2021). Asthma and COVID-19: An update. *Eur. Respir. Rev.* 30 (162), 210152. doi:10.1183/16000617.0152-2021

Allen, S. J., Crown, S. E., and Handel, T. M. (2007). Chemokine: Receptor structure, interactions, and antagonism. *Annu. Rev. Immunol.* 25, 787–820. doi:10.1146/annurev.immunol.24.021605.090529

Almoguera, B., Vazquez, L., Mentch, F., Connolly, J., Pacheco, J. A., Sundaresan, A. S., et al. (2017). Identification of four novel loci in asthma in European American and african American populations. *Am. J. Respir. Crit. Care Med.* 195 (4), 456–463. doi:10.1164/rccm.201604-0861OC

Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D-Y., Chen, L., et al. (2020). Presumed asymptomatic carrier transmission of COVID-19. *Jama* 323, 1406–1407. doi:10.1001/jama.2020.2565

Bartee, E., and McFadden, G. (2013). Cytokine synergy: An underappreciated contributor to innate anti-viral immunity. *Cytokine* 63 (3), 237–240. doi:10.1016/j.cyto.2013.04.036

Biron, C. A. (1998). Role of early cytokines, including alpha and beta interferons (IFN-α\β), in innate and adaptive immune responses to viral infections. *Seminars Immunol.* 10 (5), 383–390. doi:10.1006/smim.1998.0138

Boettger, L. M., Handsaker, R. E., Zody, M. C., and McCarroll, S. A. (2012). Structural haplotypes and recent evolution of the human 17q21.31 region. *Nat. Genet.* 44, 881–885. doi:10.1038/ng.2334

Buitrago-Garcia, D., Egli-Gany, D., Counotte, M. J., Hossmann, S., Imeri, H., Ipekci, A. M., et al. (2020). Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLoS Med.* 17 (9), e1003346. doi:10.1371/journal.pmed.1003346

CDC (2021). *People with certain medical conditions.* Available at: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html.

Chamnanphon, M., Pongpanich, M., Suttichet, T. B., Jantarabenjakul, W., Torvorapanit, P., Putcharoen, O., et al. (2022). Host genetic factors of COVID-19 susceptibility and disease severity in a Thai population. *J. Hum. Genet.* 67 (5), 295–301. doi:10.1038/s10038-021-01009-6

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. doi:10.1186/s13742-015-0047-8

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics* 10, 101. doi:10.2307/3001666

D'amato, G., Liccardi, G., D'amato, M., and Holgate, S. (2005). Environmental risk factors and allergic bronchial asthma. *Clin. Exp. Allergy* 35 (9), 1113–1124. doi:10.1111/j.1365-2222.2005.02328.x

David, E., Frauke, D., Luis, B., Maria, B., Agustín, A., and Pietro, I. (2020). Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* 383 (16), 1522–1534. doi:10.1056/NEJMoa2020283

Day, C., Patel, R., Guillen, C., and Wardlaw, A. J. (2009). The chemokine CXCL16 is highly and constitutively expressed by human bronchial epithelial cells. *Exp. Lung Res.* 35 (4), 272–283. doi:10.1080/01902140802635517

De Souza, T. H., Nadal, J. A., Nogueira, R. J., Pereira, R. M., and Brandão, M. B. (2020). Clinical manifestations of children with COVID-19: A systematic review. *Pediatr. Pulmonol.* 55 (8), 1892–1899. doi:10.1002/ppul.24885

Demenais, F., Margaritte-Jeannin, P., Barnes, K. C., Cookson, W. O. C., Altmüller, J., Ang, W., et al. (2018). Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat. Genet.* 50 (1), 42–53. doi:10.1038/s41588-017-0014-7

Docherty, A. B., Harrison, E. M., Green, C. A., Hardwick, H. E., Pius, R., Norman, L., et al. (2020). Features of 20 133 UK patients in hospital with COVID-19 using the ISARIC WHO clinical characterisation protocol: Prospective observational cohort study. *Br. Med. J.* 369, m1985. doi:10.1136/bmj.m1985

Dubé, M. P., Lemaçon, A., Barhdadi, A., Lemieux Perreault, L. P., Oussaïd, E., Asselin, G., et al. (2021). Genetics of symptom remission in outpatients with COVID-19. *Sci. Rep.* 11 (1), 10847. doi:10.1038/s41598-021-90365-6

Evangelou, E., and Ioannidis, J. P. A. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* 14, 379–389. doi:10.1038/nrg3472

Ferreira, M. A., Mathur, R., Vonk, J. M., Szwajda, A., Brumpton, B., Granell, R., et al. (2019). Genetic architectures of childhood-and adult-onset asthma are partly distinct. *Am. J. Hum. Genet.* 104 (4), 665–684. doi:10.1016/j.ajhg.2019.02.022

Horowitz, J. E., Kosmicki, J. A., Damask, A., Sharma, D., Roberts, G. H. L., Justice, A. E., et al. (2022). Genome-wide analysis provides genetic evidence that ACE2 influences COVID-19 risk and yields risk scores associated with severe disease. *Nat. Genet.* 54 (4), 382–392. doi:10.1038/s41588-021-01006-7

Hu, B., Huang, S., and Yin, L. (2021). The cytokine storm and COVID-19. *J. Med. Virol.* 93 (1), 250–256. doi:10.1002/jmv.26232

Hu, J., Li, C., Wang, S., Li, T., and Zhang, H. (2021). Genetic variants are identified to increase risk of COVID-19 related mortality from UK Biobank data. *Hum. Genomics* 15 (1), 10. doi:10.1186/s40246-021-00306-7

Initiative, C-H. G. (2020). The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of

the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet.* 28 (6), 715–718. doi:10.1038/s41431-020-0636-6

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., et al. (2020). The mutational constraint spectrum quantified from variation in 141, 456 humans. *Nature* 581 (7809), 434–443. doi:10.1038/s41586-020-2308-7

Kosmicki, J. A., Horowitz, J. E., Banerjee, N., Lanche, R., Marcketta, A., Maxwell, E., et al. (2021). Pan-ancestry exome-wide association analyses of COVID-19 outcomes in 586, 157 individuals. *Am. J. Hum. Genet.* 108 (7), 1350–1355. doi:10.1016/j.ajhg.2021.05.017

Kosmicki, J. A., Horowitz, J. E., Banerjee, N., Lanche, R., Marcketta, A., Maxwell, E., et al. (2021). *A catalog of associations between rare coding variants and COVID-19 outcomes*. medRxiv. Preprint. doi:10.1101/2020.10.28.20221804

Kousathanas, A., Pairo-Castineira, E., Rawlik, K., Stuckey, A., Odhams, C. A., Walker, S., et al. (2022). Whole-genome sequencing reveals host factors underlying critical COVID-19. *Nature* 607 (7917), 97–103. doi:10.1038/s41586-022-04576-6

Lutfalla, G., Holland, S. J., Cinato, E., Monneron, D., Reboul, J., Rogers, N. C., et al. (1995). Mutant U5A cells are complemented by an interferon-alpha beta receptor subunit generated by alternative processing of a new member of a cytokine receptor gene cluster. *EMBO J.* 14 (20), 5100–5108. doi:10.1002/j.1460-2075.1995.tb00192.x

Ma, Y., Huang, Y., Zhao, S., Yao, Y., Zhang, Y., Qu, J., et al. (2021). Integrative genomics analysis reveals a 21q22.11 locus contributing risk to COVID-19. *Hum. Mol. Genet.* 30 (13), 1247–1258. doi:10.1093/hmg/ddab125

Mousa, M., Vurivi, H., Kannout, H., Uddin, M., Alkaabi, N., Mahboub, B., et al. (2021). Genome-wide association study of hospitalized COVID-19 patients in the United Arab Emirates. *EBioMedicine* 74, 103695. doi:10.1016/j.ebiom.2021.103695

Niemi, M. E. K., Karjalainen, J., Liao, R. G., Neale, B. M., Daly, M., Ganna, A., et al. (2021). Mapping the human genetic architecture of COVID-19. *Nature* 600 (7889), 472–477.

Olafsdottir, T. A., Theodors, F., Bjarnadottir, K., Bjornsdottir, U. S., Agustsdottir, A. B., Stefansson, O. A., et al. (2020). Eighty-eight variants highlight the role of T cell regulation and airway remodeling in asthma pathogenesis. *Nat. Commun.* 11 (1), 393. doi:10.1038/s41467-019-14144-8

Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., Pasko, D., et al. (2021). Genetic mechanisms of critical illness in COVID-19. *Nature* 591 (7848), 92–98. doi:10.1038/s41586-020-03065-y

Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., Pasko, D., et al. (2021). Genetic mechanisms of critical illness in COVID-19. *Nature* 591 (7848), 92–98. doi:10.1038/s41586-020-03065-y

Pakianathan, D. R., Kuta, E. G., Artis, D. R., Skelton, N. J., and Hébert, C. A. (1997). Distinct but overlapping epitopes for the interaction of a CC-chemokine with CCR1, CCR3, and CCR5. *Biochemistry* 36 (32), 9642–9648. doi:10.1021/bi970593z

Patrick, M. T., Zhang, H., Wasikowski, R., Prens, E. P., Weidinger, S., Gudjonsson, J. E., et al. (2021). Associations between COVID-19 and skin conditions identified through epidemiology and genomic studies. *J. Allergy Clin. Immunol.* 147 (3), 857–869.e7. e7. doi:10.1016/j.jaci.2021.01.006

Peloso, G. M., Tcheandjieu, C., McGeary, J. E., Posner, D. C., Ho, Y. L., Zhou, J. J., et al. (2021). Genetic loci associated with COVID-19 positivity and hospitalization in white, black, and hispanic veterans of the VA million veteran program. *Front. Genet.* 12, 777076. doi:10.3389/fgene.2021.777076

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., et al. (2010). LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics* 26 (18), 2336–2337. doi:10.1093/bioinformatics/btq419

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795

Roberts, G. H. L., Partha, R., Rhead, B., Knight, S. C., Park, D. S., Coignet, M. V., et al. (2022). Expanded COVID-19 phenotype definitions reveal distinct patterns of genetic association and protective effects. *Nat. Genet.* 54 (4), 374–381. doi:10.1038/s41588-022-01042-x

Roberts, G. H. L., Park, D. S., Coignet, M. V., McCurdy, S. R., Knight, S. C., Partha, R., et al. (2020). AncestryDNA COVID-19 host genetic study identifies three novel loci. *Prepr. A. T.* doi:10.1101/2020.10.06.20205864

Severe Covid, G. G., Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., et al. (2020). Genomewide association study of severe covid-19 with respiratory failure. *N. Engl. J. Med.* 383 (16), 1522–1534. doi:10.1056/NEJMoa2020283

Shelton, J. F., Shastri, A. J., Ye, C., Weldon, C. H., Filshtein-Sonmez, T., Coker, D., et al. (2021). Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat. Genet.* 53 (6), 801–808. doi:10.1038/s41588-021-00854-7

Shelton, J. F., Shastri, A. J., Ye, C., Weldon, C. H., Filshtein-Sonmez, T., Coker, D., et al. (2021). Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat. Genet.* 53, 801–808. doi:10.1038/s41588-021-00854-7

Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., et al. (2005). A common inversion under selection in Europeans. *Nat. Genet.* 37, 129–137. doi:10.1038/ng1508

Sun, B. B., Maranville, J. C., Peters, J. E., Stacey, D., Staley, J. R., Blackshaw, J., et al. (2018). Genomic atlas of the human plasma proteome. *Nature* 558, 73–79. doi:10.1038/s41586-018-0175-2

Vicari, A. P., Figueroa, D. J., Hedrick, J. A., Foster, J. S., Singh, K. P., Menon, S., et al. (1997). Teck: A novel CC chemokine specifically expressed by thymic dendritic cells and potentially involved in T cell development. *Immunity* 7 (2), 291–301. doi:10.1016/s1074-7613(00)80531-2

Wang, J., Duncan, D., Shi, Z., and Zhang, B. (2013). WEB-Based GEne SeT AnaLysis Toolkit (WebGestalt): Update 2013. *Nucleic Acids Res.* 41 (W1), W77–W83. doi:10.1093/nar/gkt439

Wei, P. F. (2020). Diagnosis and treatment protocol for novel coronavirus pneumonia (trial version 7). *Chin. Med. J. Engl.* 133 (9), 1087–1095. doi:10.1097/CM9.0000000000000819

White, G. E., Iqbal, A. J., and Greaves, D. R. (2013). CC chemokine receptors and chronic inflammation—Therapeutic opportunities and pharmacological challenges. *Pharmacol. Rev.* 65 (1), 47–89. doi:10.1124/pr.111.005074

Williamson, E. J., Walker, A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., et al. (2020). Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 584 (7821), 430–436. doi:10.1038/s41586-020-2521-4

Wu, Z., and McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72 314 cases from the Chinese center for disease control and prevention. *JAMA* 323, 1239–1242. doi:10.1001/jama.2020.2648

Yucesoy, B., Kaufman, K. M., Lummus, Z. L., Weirauch, M. T., Zhang, G., Cartier, A., et al. (2015). Genome-wide association study identifies novel loci associated with diisocyanate-induced occupational asthma. *Toxicol. Sci.* 146 (1), 192–201. doi:10.1093/toxsci/kfv084

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in wuhan, China: A retrospective cohort study. *Lancet* 395, 1054–1062. doi:10.1016/S0140-6736(20)30566-3

Zhou, W., Nielsen, J. B., Fritsche, L. G., Dey, R., Gabrielsen, M. E., Wolford, B. N., et al. (2018). Efficiently controlling for case–control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* 50, 1335–1341. doi:10.1038/s41588-018-0184-y

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 05 September 2022
DOI 10.3389/fgene.2022.955965



OPEN ACCESS

EDITED BY
Zhongshan Cheng,
St. Jude Children's Research Hospital, United States

REVIEWED BY
Elisa Zavattaro,
University of Eastern Piedmont, Italy
Yu-si Luo,
Affiliated Hospital of Guizhou Medical University, China

*CORRESPONDENCE
Mahmoud Mahmoudi,
mahmoudim@mums.ac.ir
Tooba Ghazanfari,
Tghazanfari@yahoo.com

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 29 May 2022
ACCEPTED 08 August 2022
PUBLISHED 05 September 2022

CITATION
Faridzadeh A, Mahmoudi M, Ghaffarpour S, Zamani MS, Hoseinzadeh A, Naghizadeh MM and Ghazanfari T (2022), The role of ACE1 I/D and ACE2 polymorphism in the outcome of Iranian COVID-19 patients: A case-control study.
*Front. Genet.* 13:955965.
doi: 10.3389/fgene.2022.955965



# The role of ACE1 I/D and ACE2 polymorphism in the outcome of Iranian COVID-19 patients: A case-control study

Arezoo Faridzadeh[1,2], Mahmoud Mahmoudi[1,2]*, Sara Ghaffarpour[3], Mohammad Saber Zamani[3], Akram Hoseinzadeh[1,4], Mohammad Mehdi Naghizadeh[5] and Tooba Ghazanfari[3,6]*

[1]Immunology Research Center, Mashhad University of Medical Sciences, Mashhad, Iran, [2]Department of Immunology and Allergy, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran, [3]Immunoregulation Research Center, Shahed University, Tehran, Iran, [4]Department of Immunology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran, [5]Noncommunicable Diseases Research Center, Fasa University of Medical Science, Fasa, Iran, [6]Department of Immunology, Shahed University, Tehran, Iran

**Background:** Since the beginning of the pandemic of coronavirus disease 2019 (COVID-19), many countries have experienced a considerable number of COVID-19 cases and deaths. The etiology of a broad spectrum of symptoms is still debated. Host genetic variants might also significantly influence the outcome of the disease. This study aimed to evaluate the association of angiotensin-converting enzyme (ACE1) gene Insertion/Deletion (I/D) polymorphism (rs1799752) and ACE2 gene rs1978124 single nucleotide polymorphism with the COVID-19 severity.

**Methods:** This study was conducted on 470 COVID-19 patients and a control group of 56 healthy individuals across several major cities in Iran. The blood sample and clinical data were collected from the participants, and their ACE1 I/D and ACE2 rs1978124 polymorphisms were determined using polymerase chain reaction and PCR-RFLP, respectively. Serum levels of C-reactive protein (CRP), interleukin 6 (IL-6), and ACE1 were measured in the blood samples.

**Results:** We found that the ACE1 DD genotype frequency was inversely correlated with the risk of intubation ($p$ = 0.017) and mortality in COVID-19 patients ($p$ = 0.049). Even after adjustment, logistic regression demonstrated that this significant inverse association remained constant for the above variables at odds ratios of (OR) = 0.35 and Odds Ratio = 0.49, respectively. Also, in the expired ($p$ = 0.042) and intubated ($p$ = 0.048) groups with II + ID genotypes, the mean level of CRP was significantly higher than in the DD genotype group. Furthermore, in both intubated and expired groups, the mean serum level of ACE1 was higher compared with non-intubated and survived groups with II or II + ID genotypes. The results also indicated that ACE2 rs1978124 TT + CT genotypes in females have a significant positive role in susceptibility to COVID-19; however, in females, the TT + CT genotypes had a protective effect (OR = 0.098) against the severity of COVID-19.

**Conclusion:** These findings suggest that ACE1 I/D and ACE2 rs1978124 polymorphism could potentially influence the outcome of COVID-19 in the Iranian population.

KEYWORDS

polymorphism, insertion/deletion, angiotensin-converting enzyme (ACE), coronavirus disease 2019 (COVID-19), severity, SNP

# 1 Introduction

The current pandemic results from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), causing coronavirus disease 2019 (COVID-19). It has turned into a full-blown global crisis since its onset, prompting researchers worldwide to seek solutions for this problem. So far, COVID-19 has affected over 566 million individuals worldwide, resulting in 6.3 million deaths (WHO, 2022). The clinical spectrum of SARS-CoV-2 patients ranges from asymptomatic and mild flu-like symptoms to severe acute respiratory distress syndrome (ARDS). The prevalence and mortality rates of COVID-19 vary considerably from country to country, which raises the question of whether geographical origin and host genetic variations play a role in the severity and mortality of COVID-19 infection.

Angiotensin-converting enzyme-2 (ACE2) is the receptor of SARS-CoV-2, and transmembrane protease serine 2 (TMPRSS2) facilitates the virus entry. ACE2 is mainly expressed in the lung, intestine, cardiovascular system, kidney, adipose tissue, and central nervous system (Gheblawi et al., 2020). ACE1 converts Angiotensin I into Angiotensin II, which promotes inflammation, thrombosis, and vasoconstriction. ACE2 converts Angiotensin II into Angiotensin 1–7 and hence promotes vasodilation (D'Ardes et al., 2020). Downregulation of ACE2 expression due to SARS-CoV-2 infection may prevent viral infection. However, it also diminishes the beneficial impacts of ACE2 in the lungs and other organs (Gao et al., 2022). Therefore, COVID-19 may lead to ACE1/ACE2 imbalance and increase angiotensin II levels because it activates the renin-angiotensin-aldosterone system (RAAS) and thus the progression of COVID-19, especially in patients with underlying diseases such as high blood pressure (HTN), cardiovascular disease (CVD) and diabetes (DM) (Adamzik et al., 2007; Beyerstedt et al., 2021a). Moreover, factors like sex, age, smoking habit, obesity, blood group, HTN, DM, CVD, and genetics might be important in COVID-19 infection (Gard, 2010; Cai et al., 2020; Ejaz et al., 2020; Ovsyannikova et al., 2020; Sattar et al., 2020; Zeberg and Pääbo, 2020; Goel et al., 2021). On the other hand, genetic variation of a gene likely modifies the function and expression of an encoded product, which could be considered the interindividual differences in susceptibility to several infectious diseases. So far, few studies have shown the roles of angiotensin-converting enzyme 1 (ACE1), ACE2, and transmembrane protease serine 2 (TMPRSS2) gene variants in the COVID-19 severity (Gorbalenya et al., 2020; Hou et al., 2020; de Araújo et al., 2022; Gintoni et al., 2022).

## 1.1 ACE1

The ACE1 gene is located on chromosome 17q35 with 26 exons and 25 introns. The insertion/deletion (I/D) ACE1 polymorphism (rs1799752) is described by an insertion (allele I) or deletion (allele D) of a 287-base pair (bp) Alu repeat sequence in the 16th intron of the ACE1 gene, which accounts for most of the interindividual variability in circulating ACE activity and shows significant geographic variability. Therefore, I/D polymorphism has three different genotypes: II, ID, and DD (Rigat et al., 1990; Rieder et al., 1999; Sayed-Tabatabaei et al., 2006). Some recent studies suggest that the ACE1 ID polymorphism as a main geographical variation could be one of the genetic markers of susceptibility and pathogenicity of COVID-19 (Pati et al., 2020; Yamamoto et al., 2020). A review on I/D polymorphism suggested that the DD genotype in COVID-19 patients might cause severe lung injury (Zheng and Cao, 2020). However, a meta-analysis by Delanghe et al. demonstrated a negative association between COVID-19 mortality and D alleles frequency from an evaluation of 25 countries in the Middle East, North Africa, and Europe (Delanghe et al., 2020a). An ecological study demonstrated the distribution of II genotype is highest in Asian countries and lower among the African and European countries across 25 countries (Aung et al., 2020). A case-control study in one of the southeastern cities of Iran with a smaller sample size has shown that the II genotype decreases the risk of COVID-19 infection (Kouhpayeh et al., 2021).

## 1.2 ACE2

ACE2 rs1978124 SNP is located on chromosome Xp22 in intron one, suggesting that this variant may affect the expression of the ACE2 gene (Zhao et al., 2010; Patel et al., 2014). Also, some studies demonstrated that the ACE2 rs1978124 SNP was associated with the severity of COVID-19, the risk of diabetes-related left ventricular remodeling, and dyslipidemia (Liu et al., 2018; Sabater Molina et al., 2022).

The present study is the first to determine the potential role of ACE1/ACE2 gene polymorphisms in susceptibility to COVID-

19 and the disease outcome of COVID-19 with a larger sample size compared to previous studies, representing the entire Iranian population.

# 2 Participants and methods

## 2.1 Study subjects

This case-control study was conducted on 470 patients with COVID-19 and 56 healthy controls referred to hospitals between 2020 and 2021 across several major cities in Iran. COVID-19 infections were confirmed by real-time reverse transcription-polymerase chain reaction (RT-PCR) or chest CT scan findings. COVID-19 severity was classified into mild, moderate, severe, and critical as defined by the World Health Organization (WHO) (WHO, 2021). The research protocol was approved by the National Institute for Medical Research Development (IR.NIMAD.REC.1399.041).

## 2.2 Data collection and blood sampling

The relevant personal information and medical history of most subjects, including their sex, age, smoking status, and comorbidities, were obtained by a patient checklist. Informed consent was acquired from all individuals or their family members before collecting blood samples. The selection method of patients was not probabilistic. At the beginning of hospitalization, all patients and healthy control individuals donated 5 ml blood samples collected in the clot activator tubes and tubes containing ethylene diamine tetraacetic acid (EDTA).

## 2.3 Interleukin-6, C-reactive protein, and ACE1 assessment

Interleukin-6 (IL-6) was measured in serum samples using an automated immunoassay (IMMULITE 2000; Siemens Healthcare Diagnostics, The United Kingdom). Serum level of C-reactive protein (CRP) and the quantitative enzymatic determination of ACE1 were done in serum samples with 7180 clinical analyzers (Hitachi, Japan) using ACE BIOLIS (Genbio, Ireland).

## 2.4 DNA extraction and genotyping

DNA was extracted from the buffy coat samples of all subjects using a spin column kit (GenAll Exgene Cell SV mini kit, GenAll Biotechnology, South Korea).

### 2.4.1 Detection of ACE I/D gene polymorphism

The ACE1 intronic Alu insertion (I) or deletion (D) polymorphism (rs1799752) was determined by polymerase chain reaction (PCR) and agarose gel electrophoresis methods with specific primers (forward primer- 5′ -CTGGAGACCACT CCCATCCTTTCT-3′ and reverse primer- 5′ -GATGTGGCC ATCACATTCGTCAGAT-3′). PCR reactions were performed in a final volume of 25 μL comprising TEMPase Hot Start 2x Master Mix A (Ampliqon, Denmark), ten pmol of each primer (TAG Copenhagen A/S, Denmark), 20–100 ng genomic DNA, and distilled water. After the initial denaturation step at 95°C for 15 min, the reaction mixtures were subjected to 40 cycles of 95°C for 30 s, 58°C for 45 s, 72°C for 1 min, and a final extension at 72°C for 5 min. The PCR products were electrophoresed and visualized in 2% agarose gels containing DNA-safe stains. This technique provided amplification products of 191 bp for the DD genotype, 480 base pairs (bp) for the II genotype, and 480 bp + 191 bp for the ID genotype (Figure 1).

### 2.4.2 Determination of rs1978124 SNP of the ACE2 gene

The ACE2 rs1978124 polymorphism was assessed by PCR–restriction fragment length polymorphism (PCR-RFLP). The PCR product of 553 bp was generated using the forward primer-5′- CAACCACACATACCACAAT-3′and reverse primer-5′- TTTCCTTTAGCCTACAATATCAAT -3′, and were incubated with 1 μL of *Echo471* (*Ava II*) restriction enzyme at 37°C overnight. After digestion, two 464 bp and 89 bp products identify the C allele, and a 553 bp band identifies the T allele on agarose gel (Figure 2). About 10% of the samples were directly Sanger sequenced to ensure PCR-RFLP for SNP rs1978124. Using PCR primers, sanger sequencing was performed on 10% of the resulting samples. The sequencing result of rs1978124 SNP after alignment is shown in Supplementary Figure S1.

## 2.5 Statistical analysis

The numerical variables of each group were presented as mean ± standard deviation, and the Mann-Whitney test compared continuous data. The genotypes frequencies were reported as number (percentages) or n (%) in each group and assessed using the Chi-square. Also, Hardy–Weinberg equilibrium (HWE) was calculated and tested by the Chi-square test. The associations of ACE1 insertion/deletion polymorphism and ACE2 rs1978124 SNP with susceptibility and severity for SARS-CoV-2 infection at both the multiple and univariate levels were assessed by multinomial or binary logistic regressions to calculate odds ratios (ORs) (adjusted and unadjusted) with 95% confidence intervals (CI). Individuals were included in the comparison among groups after adjustment for sex, age, HTN, diabetes mellitus, CVD, renal disease, and cigarette smoking. For rs1978124, males and females were analyzed

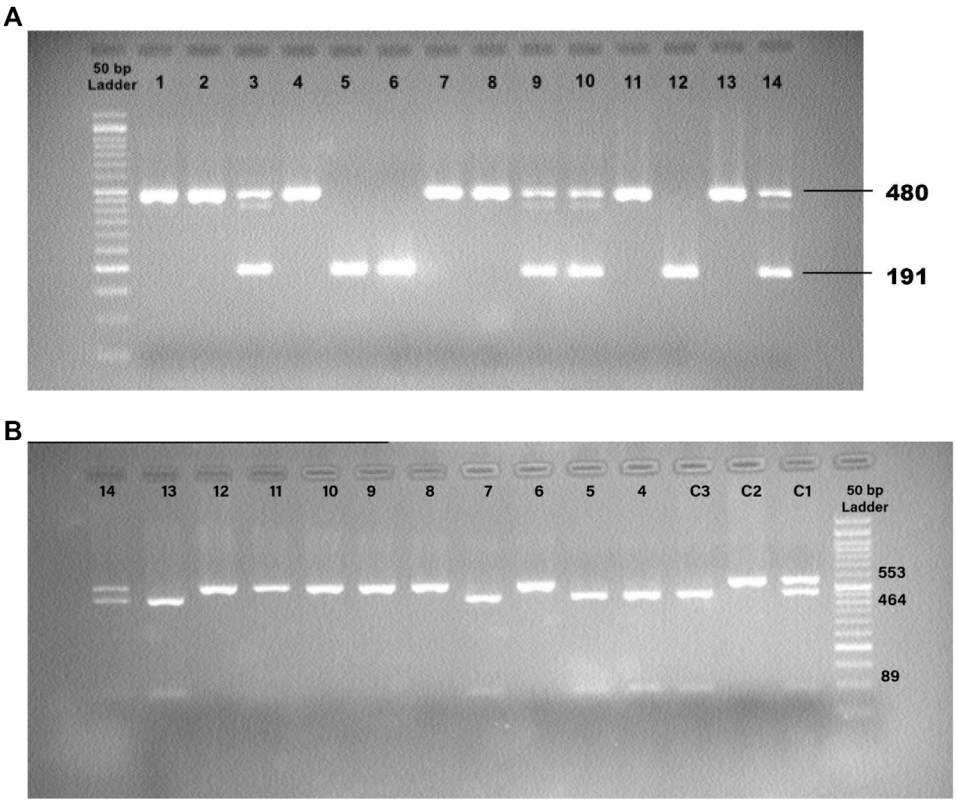



FIGURE 1
**(A)** Detection of the PCR Products for ACE1 Insertion/Deletion (I/D) Polymorphism. 12: Control, DD; 13: Control, II; 14: Control, ID; 1,2,4,7,8,11: II; 5,6: DD; 3,9,10: ID. **(B)** Detection of the PCR-RFLP Products for ACE2 rs1978124 SNP. C1: Control, CT; C2: Control, TT; C3: Control, CC; 4, 5, 7, 13: CC; 6, 8, 9, 10, 11, 12: TT; 14: CT.

separately since the ACE2 gene is on the X chromosome. A $p$-value less than 0.05 was regarded to be significant.

# 3 Results

## 3.1 Demographic characteristics of the study on Iranian population

Patients were divided into groups based on the disease severity (COVID-19 Treatment Guidelines, 2019). The frequencies of comorbidities such as HTN, diabetes mellitus, CVD, renal disease, and cigarette smoking are presented in Table 1 and Supplementary Table S1. The results of this study demonstrated that HTN ($p < 0.001$), diabetes mellitus ($p < 0.001$), CVD ($p < 0.001$), and renal disease ($p = 0.024$) were significantly higher in COVID-19 patients compared to controls. Also, HTN ($p = 0.008$), diabetes mellitus ($p = 0.007$), and CVD ($p = 0.009$) were significantly correlated with COVID-19 mortality. There was no significant gender difference between the groups. The mean age was associated with increased severity and disease mortality in COVID-19. The mean age of healthy control, outpatients, inpatients, ICU admitted patients, intubated, and expired patients were 45.3 ± 13.3, 44.6 ± 14.1, 57.5 ± 16.5, 61.8 ± 17, 62 ± 14.9, and 65.8 ± 13.9 years, respectively.

## 3.2 ACE1 I/D and ACE2 rs1978124 genotypes

The genotype frequencies of the ACE1 I/D and ACE2 rs1978124 polymorphisms in the patients and control groups agreed with the Hardy-Weinberg equilibrium using the Chi-square analysis. (Supplementary Table S2).

## 3.3 Statistical comparisons between controls and patients (P value) for ACE1 and ACE2 genotypes/alleles

We found no significant relationship between different ACE1 I/D and ACE2 rs1978124 genotypes/alleles frequencies

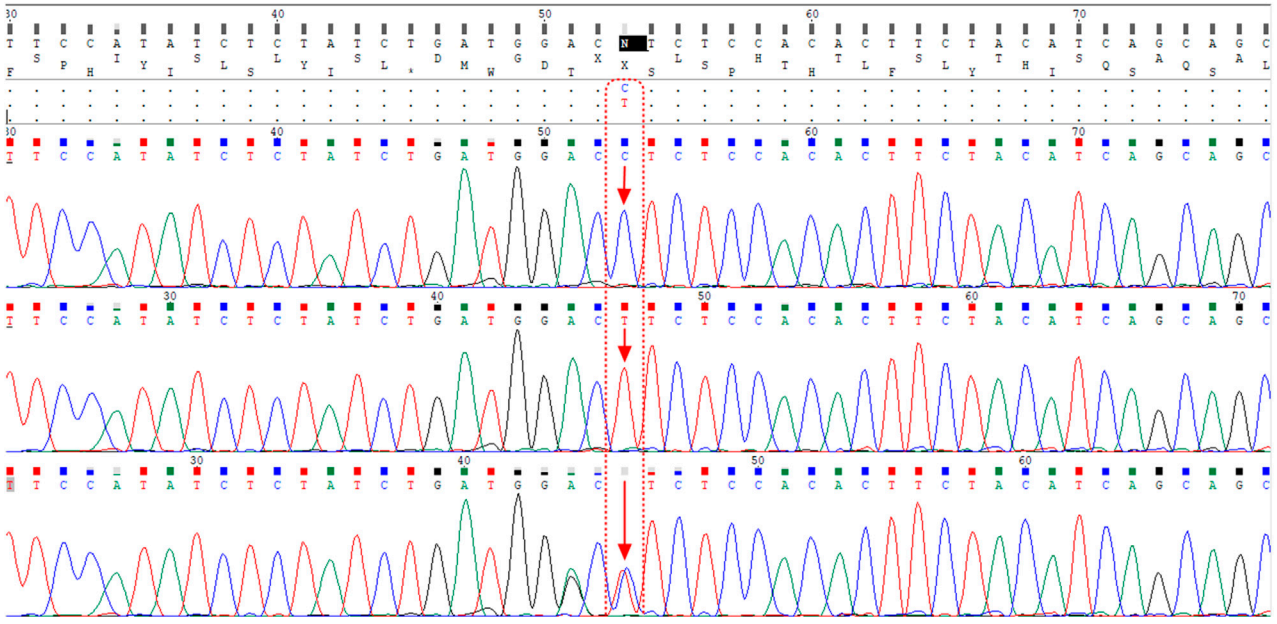


FIGURE 2
Alignment of the sequencing results of rs1978124 SNP located in ACE2 gene; the first, second, and third rows are related to the sample with CC genotype, TT genotype, and CT genotype, respectively. The presence of both peaks for both alleles is evident in the third row.

TABLE 1 Demographic characteristics of the study participants.

| | Control (n = 56) | COVID-19 | | | | |
|---|---|---|---|---|---|---|
| | | Outpatient (n = 207) | Inpatient (n = 263) | Admitted to ICU (n = 76) | Intubated (n = 47) | Expired (n = 57) |
| Sex (Female ratio) | 31 (55.4%) | 83 (40.1%) | 109 (41.4%) | 31 (40.8%) | 18 (38.3%) | 21 (36.8%) |
| Cigarette smoking | 2 (3.6%) | 10 (4.8%) | 25 (9.5%) | 8 (10.5%) | 4 (8.5%) | 4 (7%) |
| HTN | 2 (3.6%) | 24 (11.6%) | 105 (39.9%) | 30 (39.5%) | 17 (36.2%) | 24 (42.1%) |
| DM | 1 (1.8%) | 16 (7.7%) | 84 (31.9%) | 26 (34.2%) | 17 (36.2%) | 20 (35.1%) |
| CVD | 0 (0%) | 5 (2.4%) | 70 (26.6%) | 20 (26.3%) | 9 (19.1%) | 14 (24.6%) |
| Renal disease | 0 (0%) | 8 (3.9%) | 21 (8%) | 9 (11.8%) | 5 (10.6%) | 6 (10.5%) |

N: number, %: frequency. Abbreviations: COVID-19, Coronavirus disease 2019; HTN, hypertension; DM, diabetes mellitus; CVD, cardiovascular disease.

with comorbidities. The relevant statistical details are presented in Supplementary Table S3.

### 3.3.1 Susceptibility to COVID-19 infection

There was no significant association between ACE1 I/D (rs1799752) and ACE2 rs1978124 SNP genotypes/allele frequencies and susceptibility to COVID-19, but ACE2 rs1978124 T allele and TT+CT genotypes frequencies were higher in women with COVID-19 than in female controls. After adjusting for sex, age, HTN, diabetes mellitus, CVD, renal disease, and cigarette smoking, the comparison of 470 COVID-19 patients and 56 healthy controls through logistic regression revealed no significant association between genotypes of ACE1 and susceptibility to COVID-19 infection. However, for ACE2 CT + TT genotypes, this significant association remained even after adjustment ($p$ = 0.008, 95%CI = 5.99) (Table 2).

TABLE 2 Association of ACE1 I/D and ACE2 rs1978124 Genotypes/Alleles Distribution with Susceptibility to COVID-19, Adjusted by Age, Sex, Cigarette Smoking, Diabetes Mellitus, HTN, CVD, and renal diseases.

| Genotypes alleles N (%) | | Study group | | | Unadjusted | | Adjusted | |
|---|---|---|---|---|---|---|---|---|
| | | Control (n = 56) | COVID-19 (n = 470) | *p*-value (Chi-square) | *p*-value | OR-95%CI-(L-U) | *p*-value | OR-95%CI-(L-U) |
| **ACE1 I/D** | II | 9 (16.1%) | 107 (22.8%) | — | — | — | — | — |
| | ID | 30 (53.6%) | 215 (45.7%) | 0.428 | 0.204 | 0.603 (0.276–1.315 | 0.435 | 0.722 (0.318–1.636) |
| | DD | 17 (30.4%) | 148 (31.5%) | — | 0.470 | 0.732 (0.314–1.705) | 0.958 | 1.024 (0.417–2.515) |
| | II | 9 (16.1%) | 107 (22.8%) | — | — | — | — | — |
| | ID + DD | 47 (83.9%) | 363 (77.2%) | 0.253 | 0.256 | 0.650 (0.308–1.368) | 0.262 | 0.823 (0.376–1.802) |
| | II + ID | 39 (69.6%) | 322 (68.5%) | — | — | — | — | — |
| | DD | 17 (30.4%) | 148 (31.5%) | 0.863 | 0.863 | 1.054 (0.578–1.925) | 0.418 | 1.306 (0.684–2.491) |
| | I | 48 (42.9%) | 429 (45.6%) | — | — | — | — | — |
| | D | 64 (68.8%) | 511 (54.4%) | 0.576 | 0.576 | 0.893 (0.601–1.327) | — | — |
| **ACE2 rs1978124** | **Female** | — | — | — | — | — | — | — |
| | CC | 6 (19.4%) | 7 (3.6%) | — | — | — | — | — |
| | CT | 14 (45.2%) | 72 (37.5%) | **<0.001** | **0.018** | 4.408 (1.285–15.105) | 0.051 | 3.968 (0.992–15.871) |
| | TT | 11 (35.5%) | 113 (58.9%) | — | **<0.001** | 8.805 (2.513–30.853) | **0.003** | 8.452 (2.087–34.236) |
| | CC | 6 (19.4%) | 7 (3.6%) | — | — | — | — | — |
| | TT + CT | 25 (80.6%) | 185 (96.4%) | **<0.001** | **0.002** | 6.343 (1.973–20.389) | **0.008** | 5.994 (1.587–22.25) |
| | CC + CT | 20 (64.5%) | 79 (41.1%) | — | — | — | — | — |
| | TT | 11 (35.5%) | 113 (58.9%) | **0.015** | **0.018** | 2.601 (1.180–5.730) | **0.017** | 2.740 (1.197–6.273) |
| | C | 26 (41.9%) | 86 (22.4%) | — | — | — | — | — |
| | T | 36 (62.9%) | 298 (77.6%) | **<0.001** | **0.001** | 2.503 (1.432–4.375) | — | — |
| | **Male** | — | — | — | — | — | — | — |
| | C | 6 (24%) | 73 (26.3%) | — | — | — | — | — |
| | T | 19 (76%) | 205 (73.7%) | 0.805 | 0.807 | 0.887 (0.341–2.307) | — | — |

The significant *p* values are in bold, N (%): number (percentage). Abbreviations: ACE, Angiotensin-converting enzyme; I, insertion; D, deletion; OR, odds ratio; CI, confidence interval; L, lower; U, upper; COVID-19, Coronavirus disease 2019.

### 3.3.2 Severity of COVID-19

Different genotypes/alleles of ACE1 rs1799752 and ACE2 rs1978124 are not associated with COVID-19 hospitalization. After adjustment, binary logistic regression showed that patients with the TT + CT genotypes of rs1978124 carried a lower risk of hospitalization (94.5% vs 98.8%; $p = 0.042$, OR = 0.099; Supplementary Table S4).

Also, we observed a significant inverse association between the frequency of DD genotype and the risk for ICU admission (21.1% vs 33.2%, $p = 0.049$, OR = 0.53; Supplementary Table S5). Moreover, there was a significant inverse association between the frequency of DD genotype and intubation (84% vs 66.8, $p = 0.017$). While, the frequency of DD genotype and D allele significantly decreased in intubated patients (14.9% vs 32.9%, $p = 0.014$; 42.6% vs 55.1%, $p = 0.027$). After adjustment, patients with the DD genotype were at a reduced risk of intubation (OR = 0.35, $p = 0.018$; Table 3 and Figure 3).

### 3.3.3 Mortality of COVID-19

Of the 470 patients, 413 (88.3%) survived, and 57 (12.1%) expired. The D allele and DD genotype frequency were lower in the expired group than in the surviving group (19.3% vs 32.3%, $p = 0.035$; 44.7% vs 55.7%, $p = 0.028$, respectively). After adjustment, patients with DD genotype were at a decreased risk of mortality (OR = 0.49; $p = 0.049$; Table 4 and Figure 3). However, there was no significant relationship between genotype/allele frequency of ACE2 rs1978124 and mortality, even after adjustment (Table 4).

### 3.3.4 Clinical laboratory data

#### 3.3.4.1 C-reactive protein

C-reactive protein (CRP) increased along with COVID-19 severity, as shown in Supplementary Tables S6-S8. The mean ± SD serum levels of CRP in intubated patients with DD and II + ID genotypes were 16.057 ± 7.69 and 23.91 ± 8.05, respectively. Also, it

TABLE 3 Association of ACE1 Genotypes Distribution with Intubation of COVID-19 Patients, Adjusted by Age, Sex, Cigarette Smoking, DM, HTN, CVD, and renal diseases.

| Genotypes alleles N (%) | | Intubation | | | Unadjusted | | Adjusted | |
|---|---|---|---|---|---|---|---|---|
| | | No (n = 216) | Yes (n = 47) | *p*-value (Chi-square) | *p*-value | OR-95%CI-(L-U) | *p*-value | OR-95%CI-(L-U) |
| **ACE1 I/D** | II | 49 (22.7%) | 14 (29.8%) | — | — | — | — | — |
| | ID | 96 (44.4%) | 26 (55.3%) | **0.050** | 0.887 | 0.948 (0.454–1.977) | 0.808 | 1.100 (0.510–2.375) |
| | DD | 71 (32.9%) | 7 (14.9%) | — | **0.033** | 0.345 (0.130–0.917) | 0.053 | 0.376 (0.140–1.015) |
| | II | 49 (22.7%) | 14 (29.8%) | — | — | — | — | — |
| | ID + DD | 167 (77.3%) | 33 (70.2%) | 0.301 | 0.303 | 0.692 (0.343–1.395) | 0.475 | 0.767 (0.372–1.585) |
| | II + ID | 145 (67.1%) | 40 (85.1%) | — | — | — | — | — |
| | DD | 71 (32.9%) | 7 (14.9%) | **0.014** | **0.014** | 0.357 (0.152–0.838) | **0.018** | 0.354 (0.150–0.837) |
| | I | 194 (44.9%) | 54 (57.4%) | — | — | — | — | — |
| | D | 238 (55.1%) | 40 (42.6%) | **0.027** | **0.027** | 0.604 (0.385–0.948) | — | — |
| **ACE2 rs1978124** | **Female** | — | — | — | — | — | — | — |
| | CC | 5 (5.5%) | 1 (5.6%) | — | — | — | — | — |
| | CT | 30 (33.0%) | 8 (44.4%) | 0.637 | 0.895 | 1.333 (0.136–13.092) | 0.727 | 1.527 (0.142–16.421) |
| | TT | 56 (61.5%) | 9 (50.0%) | — | 0.850 | 0.804 (0.804–7.697) | 0.941 | 0.915 (0.087–9.677) |
| | CC | 5 (5.5%) | 1 (5.6%) | — | — | — | — | — |
| | TT + CT | 86 (94.5%) | 17 (94.4%) | 0.992 | 0.992 | 0.988 (0.109–9.002) | 0.911 | 1.141 (0.114–11.449) |
| | CC + CT | 35 (38.5%) | 9 (50.0%) | — | — | — | — | — |
| | TT | 56 (61.5%) | 9 (50.0%) | 0.362 | 0.365 | 0.625 (0.226–1.726) | 0.391 | 0.631 (0.221–1.806) |
| | C | 40 (22.0%) | 10 (27.8%) | — | — | — | — | — |
| | T | 142 (78.0%) | 26 (72.2%) | 0.449 | 0.451 | 0.732 (0.326–1.645) | — | — |
| | **Male** | — | — | — | — | — | — | — |
| | C | 35 (28.0%) | 6 (20.7%) | — | — | — | — | — |
| | T | 90 (72.0%) | 23 (79.3%) | 0.422 | 0.259 | 1.491 (0.560–3.971) | — | — |

The significant *p* values are in bold, N (%): number (percentage). Abbreviations: ACE, Angiotensin-converting enzyme; I, insertion; D, deletion; OR, odds ratio; CI, confidence interval; L, lower; U, upper; HTN, hypertension; DM, diabetes mellitus; CVD, cardiovascular disease.

was observed that the mean ± SD serum level of CRP in the intubated groups with DD genotype was significantly lower than in the II + ID genotypes group. Also, the serum level of CRP decreased in the expired group with DD genotype compared to II + ID genotype carriers.

In comparing COVID-19 patients with the control group, the increase in the serum level of CRP in the ACE2 TT genotype was significant compared to CC + CT genotypes (12.05 ± 11.39 vs 3.14 ± 1.69; $p = 0.029$).

### 3.3.4.2 Interleukin-6

In all carriers of genotypes ACE1 rs1799752 and ACE2 rs1978124, the amount of IL-6 was noticeably increased in the control group compared with the COVID-19 group, as well as outpatients compared with inpatients, and in the survived compared with the expired (Supplementary Tables S9-S11). The IL-6 serum levels in patients admitted to the ICU as opposed to those admitted to the non-ICU ward were considerably increased only in groups with ID, ID + II, and ID + DD genotypes. This finding was also detected when comparing intubated with non-intubated patients. The serum level of IL-6 in all rs1978124 genotypes increased with disease severity, except for the CC genotype.

### 3.3.4.3 ACE1

The mean serum ACE1 level was significantly higher amongst DD and ID genotypes compared to II genotype in control, COVID-19, outpatients, inpatients, and survived groups. However, this association between ACE1 genotypes and serum ACE1 levels was not observed in ICU, intubated, and expired groups. Even in these groups, the level of ACE1 in the DD genotype was lower than II + ID. Notably, the serum level of ACE1 was significantly decreased in the intubated patients with DD genotype compared to the non-intubated patients with DD genotype. In addition, serum levels of ACE1 in patients with II or II + ID genotypes were increased in both intubated and expired groups compared to non-intubated and survived groups.

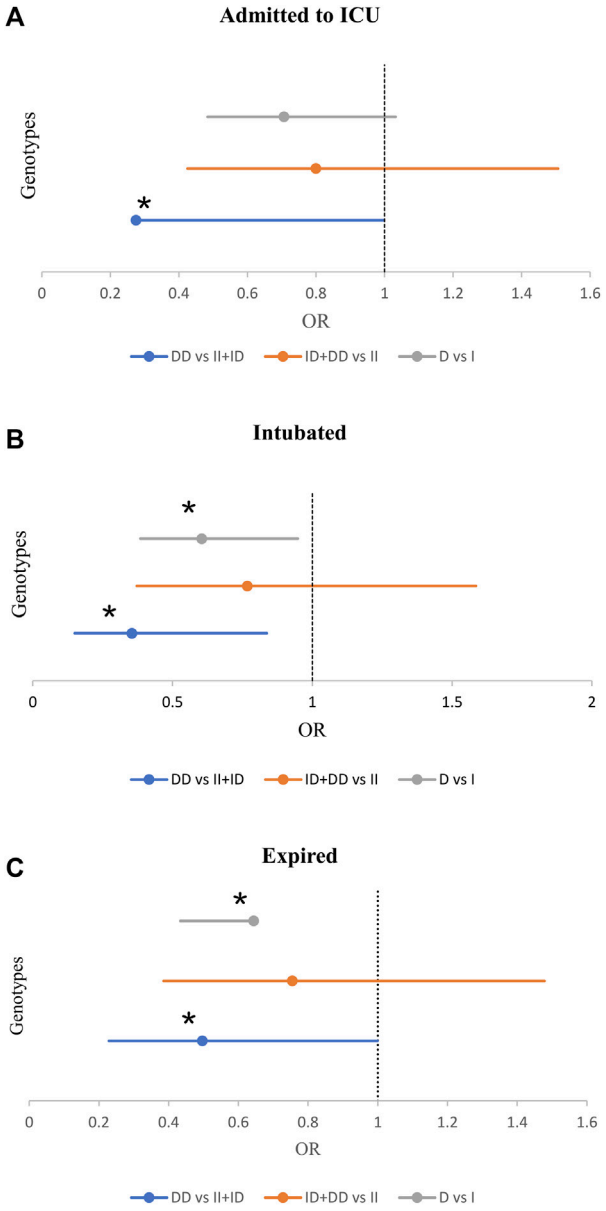



FIGURE 3
The Forest Plot displays Adjusted Odds Ratio (OR) ± 95% Confidence Intervals (CI) of ACE1 Polymorphism for ICU Admission **(A)**, Intubated **(B)**, and Expired **(C)** Groups; *$p < 0.05$. (Adjusted by Age, Sex, Cigarette Smoking, Diabetes Mellitus, Hypertension, Cardiovascular Disease, and Renal Diseases).

Supplementary Table S12-S14 summarize the relationship between the serum level of ACE1 and COVID-19 severity.

## 4 Discussion

In this case-control study, the ACE1 I/D and ACE2 (rs1978124) genotypes were found to be distributed in accordance with the Hardy-Weinberg equilibrium, indicating that the chosen samples were representative of the society. This study is the first to evaluate the possible associations between these genetic factors (ACE1 rs1799752 and ACE2 rs1978124) and COVID-19 severity in several major cities in Iran.

The present study showed that HTN, DM, CVD, and renal disease were significantly associated with COVID-19 infection. Also, HTN, DM, and CVD were significantly correlated with the COVID-19 mortality rate. These findings confirm the results of previous studies (Docherty et al., 2020; Ejaz et al., 2020; Garg et al., 2020; Richardson et al., 2020).

Since ACE2 is the cellular receptor for SARS-COV-2 entry and a main component of the RAAS system, the downregulation of ACE2 expression following a viral infection promotes ACE1/ACE2 imbalance (Beyerstedt et al., 2021b). Therefore, studying ACE1 and ACE2 polymorphisms promises to be an effective way of regulating RAAS activity, which can improve the prognosis of COVID-19 patients. Accordingly, numerous studies have been conducted on the association of ACE1 I/D (rs1799752) and ACE2 rs1978124 polymorphisms with different diseases (Rigat et al., 1990; Tiret et al., 1992; Danser et al., 1995; Luo et al., 2019). However, the location of the ACE1 I/D rs1799752 polymorphism and ACE2 rs1978124 SNP in a non-coding region of the genes means they are unlikely to be functional variants. But previous studies have pointed to an association between serum and tissue levels of the ACE1 protein with rs1799752 polymorphism, which can affect the balance of ACE1/ACE2. The rs1799752 polymorphism can describe approximately 50% of ACE activity (Sayed-Tabatabaei et al., 2006). Higher activity of ACE can lead to an increase in the concentration of angiotensin II, which plays an essential role in inflammation (Osadnik et al., 2016).

A recent study showed the effect of a family history of hypertension and ACE1 I/D polymorphism (rs1799752) on cardiac autonomic modulation in adolescents. The D allele is a prognostic factor associated with increased serum ACE levels (Dias-Filho et al., 2021). Another study showed no association between the rs1799752 ACE I/D polymorphism and diabetic retinopathy (Li et al., 2013). The ACE2 rs1978124 SNP is a common genetic factor for cardiovascular disease (Yang et al., 2006; Palmer et al., 2008; Chaoxin et al., 2013), diabetes (Patel et al., 2012), and hypertension (Benjafield et al., 2004). Since the results of relevant studies in different populations display considerable variation, it is necessary that other ethnic groups should be evaluated separately. The present study showed that ACE1 I/D and ACE2 rs1978124 genotypes/alleles frequencies are not significantly correlated with diabetes mellitus, HTN, CVD, and renal disease.

That CRP and IL-6 levels reflect disease severity (Rostamian et al., 2020; Wang, 2020; Luan et al., 2021; Santa Cruz et al., 2021). Accordingly, we also observed that CRP levels were higher in intubated patients with genotype II + ID compared to those with genotype DD. Also, comparing the two groups of COVID-19 and

TABLE 4 Association of ACE1 Genotypes Distribution with COVID-19 Mortality, Adjusted by Age, Sex, Cigarette Smoking, DM, HTN, CVD, and renal diseases.

| Genotypes alleles N (%) | | Mortality | | | Unadjusted | | Adjusted | |
|---|---|---|---|---|---|---|---|---|
| | | Survived (n = 413) | Expired (n = 57) | *p*-value (Chi-square) | *p*-value | OR-95%CI-(L-U) | *p*-value | OR-95% CI- (L-U) |
| **ACE1 I/D** | II | 90 (21.8%) | 17 (29.8%) | — | — | — | — | — |
| | ID | 186 (45%) | 29 (50.9%) | 0.088 | 0.563 | 0.825 (0.431–1.580) | 0.938 | 0.972 (0.477–1.982) |
| | DD | 137 (33.2%) | 11 (19.3%) | | **0.037** | 0.425 (0.190–0.950) | 0.088 | 0.470 (0.197–1.120) |
| | II | 90 (21.8%) | 17 (29.8%) | | | | | |
| | ID + DD | 323 (78.2%) | 40 (70.2%) | 0.175 | 0.175 | 0.656 (0.355–1.211) | 0.413 | 0.755 (0.386–1.478) |
| | II + ID | 276 (66.8%) | 46 (80.7%) | | | | | |
| | DD | 137 (33.2%) | 11 (19.3%) | **0.035** | **0.035** | 0.482 (0.242–0.959) | **0.049** | 0.497 (0.229–0.999) |
| | I | 366 (44.3%) | 63 (55.3%) | | | | | |
| | D | 460 (55.7%) | 51 (44.7%) | **0.028** | **0.028** | 0.644 (0.434–0.644) | | |
| **ACE2 rs1978124** | **Female** | — | — | — | — | — | — | — |
| | CC | 6 (3.5%) | 1 (4.8%) | — | — | — | — | — |
| | CT | 65 (38%) | 7 (33.3%) | 0.893 | 0.704 | 0.646 (0.068–6.167) | 0.443 | 0.394 (0.036–4.263) |
| | TT | 100 (58.5%) | 13 (61.9%) | — | 0.824 | 0.780 (0.087–7.001) | 0.593 | 0.530 (0.052–5.417) |
| | CC | 6 (3.5%) | 1 (4.8%) | — | — | — | — | — |
| | TT + CT | 165 (96.5%) | 20 (95.2%) | 0.772 | 0.773 | 0.727 (0.083–6.352) | 0.518 | 0.471 (0.048–4.622) |
| | CC + CT | 71 (41.5%) | 8 (38.1%) | — | — | — | — | — |
| | TT | 100 (58.5%) | 13 (61.9%) | 0.763 | 0.764 | 1.154 (0.454–2.9229) | 0.682 | 1.235 (0.450–3.391) |
| | C | 77 (22.5%) | 9 (21.4%) | — | — | — | — | — |
| | T | 265 (77.5%) | 33 78.6%) | 0.873 | 0.873 | 1.065 (0.489–2.323) | | |
| | **Male** | — | — | — | — | — | — | — |
| | C | 64 (26.4%) | 9 (25%) | — | — | — | — | — |
| | T | 178 (73.6%) | 27 (75%) | 0.854 | 0.857 | 1.079 (0.481–2.417) | — | — |

The significant *p* values are in bold, N (%): number (percentage). Abbreviations: ACE, Angiotensin-converting enzyme; I, insertion; D, deletion; OR, odds ratio; CI, confidence interval; L, lower; U, upper; HTN, hypertension; DM, diabetes mellitus; CVD, cardiovascular disease.

control in terms of ACE2, the CRP level was found to be significantly higher in carriers of TT genotype, which is positively correlated with COVID-19 susceptibility.

Recent studies have demonstrated that ACE1 I/D polymorphism could have a significant role in the prognosis of COVID-19, but these results are controversial (Delanghe et al., 2020a; Bellone and Calvisi, 2020; Saadat, 2020; Yamamoto et al., 2020; Hubacek et al., 2021). Even though numerous studies have reported that the ACE1 DD genotype is associated with COVID-19 severity, they are mostly limited to epidemiological studies and in silico analyses (Gómez et al., 2020; Pati et al., 2020; Verma et al., 2021). Also, a study on the population of Turkey indicated that ACE1 I/D polymorphism was not associated with the severity of COVID-19 (Karakaş Çelik et al., 2021). Overall, we found no significant relationship between ACE1 I/D genotypes and the susceptibility to COVID-19.

Interestingly, we found that the frequencies of DD genotype and D allele were significantly low in the group of patients admitted to the ICU, intubated, and expired. Notably, this association remained significant, even after adjustment. Our results are consistent with the findings of Delanghe et al. They found a significant inverse association between the frequency of the D allele and mortality

rate in a study that spanned more than 25 countries (Delanghe et al., 2020b; Delanghe et al., 2020c). Saad H et al. found a positive correlation between ACE1 II genotype and a heightened risk of contracting COVID-19 (Saad et al., 2021). In addition, Hubacek JA et al. showed that ACE1 II genotypes increased the risk of symptomatic COVID-19 in the Czech Republic (Hubacek et al., 2021). Furthermore, Jacobs found a significant elevation in the level of ACE2 protein in the alveolar epithelium cells when patients had II genotype of the rs1799752 polymorphism, as it can facilitate host cell entry of SARS-CoV-2 (Jacobs et al., 2021).

In the COVID-19 outpatients and healthy control, the serum level of ACE1 was significantly higher in DD compared with II + ID genotypes carriers, as expected. Furthermore, the ACE1 level was considerably higher in COVID-19 outpatients than in control subjects. Nevertheless, this significant association was not seen in patients admitted to ICU, intubated, and expired compared to those admitted to the non-ICU ward, not intubated, and who survived. In fact, in these groups, the ACE1 level had even decreased among the DD genotype and increased among II + ID genotypes. Therefore, the serum ACE1 level can be an influential factor in disease prognosis. Similar to our results, Annoni F et al. demonstrated that ACE1 levels were higher in non-survivors compared with survivors of ARDS (Annoni et al., 2019). Therefore, elevated ACE1 levels are associated with poor prognosis in ARDS patients (Sriram and Insel, 2020), which might point to endothelial activation and can prove to be a therapeutic target. Cambien F et al. demonstrated that plasma level of ACE was higher in myocardial infarction patients compared with the control group among subjects with II and ID genotypes (Cambien et al., 1994).

In conclusion, we observed that the ACE I/D polymorphism might alter ACE, IL-6, and CRP expression levels.

Previous studies have demonstrated a relation between ACE2 rs1978124 gene polymorphism and underlying comorbidities affecting the severity of COVID-19. A case-control study was conducted in West China to evaluate the association of rs1978124 ACE2 polymorphism with diabetes. It was revealed that there is found a significant relationship between the frequency of TT + CT genotypes and diabetes (OR = 2.2) (Liu et al., 2018). Another study in China showed a significant relationship between the TT + CT genotypes of rs1978124 polymorphism and dyslipidemia (Pan et al., 2018). Barry R. Palmer also found that the T allele of the ACE2 SNP rs1978124 was associated with higher mortality in an acute coronary syndrome cohort of European ancestry (Palmer et al., 2008).

A recent study on 318 Spanish COVID-19 patients aimed to evaluate the association of rs1978124 polymorphism with the severity of the disease. This study showed that CT genotype in women has a protective role (OR = 0.32) against the severity of COVID-19 (Sabater Molina et al., 2022). Our results also showed that the TT + CT genotypes have a protective effect (OR = 0.098) against the severity of COVID-19 in females. However, we observed that TT + CT genotypes in females have a significant positive role in the susceptibility to COVID-19 (infectivity), and even after adjusting for age and underlying diseases, this association remained significant. Also, the serum level of CRP was higher in patients with TT + CT genotypes compared to the control group. The functional mechanism by which rs1978124 SNP, a noncoding region of the ACE2 gene, affects the outcome of COVID-19 is unclear and needs further investigation. One possible explanation is that polymorphism affects the stability of ACE2 mRNA (i.e., splicing), post-transcriptional regulation by microRNA, and the efficiency of mRNA splicing (for example, silencing element or enhancer of intron splicing).

## 5 Study limitations

One of the limitations of this study was that the blood sample volume obtained from some patients was insufficient for laboratory tests, such as the ACE1, CRP, and IL-6 serum levels tests. Moreover, even though the assessment of angiotensin II and ACE2 serum levels can provide additional evidence for predicting the outcome of COVID-19, unfortunately, no blood sample was left to perform these tests.

## 6 Conclusion

We was demonstrated that the ACE1 I/D and ACE2 rs1978124 polymorphisms are relevant prognostic factors for the outcome of COVID-19. Patients with the II + ID genotype might have a significantly worse prognosis than those with the DD genotype. The T allele of SNP rs1978124 could affect susceptibility to COVID-19; ACE1/ACE2 polymorphism-mediated pathology is relevant, at least in the Iranian population. Consequently, further genetic studies on COVID-19 patients in different countries are required to clarify the mechanisms of lung injury and determine new therapeutic approaches.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ Supplementary Material.

## Ethics statement

The studies involving human participants were reviewed and approved by The research protocol was approved by the National Institute for Medical Research Development (IR.NIMAD.REC.1399.041). The patients/participants provided their written informed consent to participate in this study.

## Author contributions

All authors contributed to the conception and design of this paper. MM and TH were responsible for the organizing and management of the project and also the final edition of the paper. AA, SH, MZ, and MN had an essential role in collecting blood samples and questionnaires, performing the experiments, analyzing the data, and writing papers. AH had a role in collecting blood samples and writing part of the draft article. All authors read and approved the final manuscript.

## Funding

This work was supported by the Immunoregulation Research Center of Shahed University and the Ministry of Health and Medical Education of Iran.

## Acknowledgments

The authors highly appreciate all the people for their participation in this project. The facilities provided by the universities of Mashhad, Tehran, Esfahan, Babol, and Zahedan are gratefully acknowledged.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.955965/full#supplementary-material

## References

Adamzik, M., Frey, U., Sixt, S., Knemeyer, L., Beiderlinden, M., Peters, J., et al. (2007). ACE I/D but not AGT (-6)A/G polymorphism is a risk factor for mortality in ARDS. *Eur. Respir. J.* 29 (3), 482–488. doi:10.1183/09031936.00046106

Annoni, F., Orbegozo, D., Rahmania, L., Irazabal, M., Mendoza, M., De Backer, D., et al. (2019). Angiotensin-converting enzymes in acute respiratory distress syndrome. *Intensive Care Med.* 45 (8), 1159–1160. doi:10.1007/s00134-019-05600-6

Aung, A. K., Aitken, T., Teh, B. M., Yu, C., Ofori-Asenso, R., Chin, K. L., et al. (2020). Angiotensin converting enzyme genotypes and mortality from COVID-19: An ecological study. *J. Infect.* 81 (6), 961–965. doi:10.1016/j.jinf.2020.11.012

Bellone, M., and Calvisi, S. L. (2020). ACE polymorphisms and COVID-19-related mortality in Europe. *J. Mol. Med.* 98 (11), 1505–1509. doi:10.1007/s00109-020-01981-0

Benjafield, A. V., Wang, W. Y., and Morris, B. J. (2004). No association of angiotensin-converting enzyme 2 gene (ACE2) polymorphisms with essential hypertension. *Am. J. Hypertens.* 17 (7), 624–628. doi:10.1016/j.amjhyper.2004.02.022

Beyerstedt, S., Casaro, E. B., and Rangel É, B. (2021). COVID-19: Angiotensin-converting enzyme 2 (ACE2) expression and tissue susceptibility to SARS-CoV-2 infection. *Eur. J. Clin. Microbiol. Infect. Dis.* 40 (5), 905–919. doi:10.1007/s10096-020-04138-6

Beyerstedt, S., Casaro, E. B., and Rangel É, B. (2021). COVID-19: Angiotensin-converting enzyme 2 (ACE2) expression and tissue susceptibility to SARS-CoV-2 infection. *Eur. J. Clin. Microbiol. Infect. Dis.* 40 (5), 905–919. doi:10.1007/s10096-020-04138-6

Cai, G., Bossé, Y., Xiao, F., Kheradmand, F., and Amos, C. I. (2020). Tobacco smoking increases the lung gene expression of ACE2, the receptor of SARS-CoV-2. *Am. J. Respir. Crit. Care Med.* 201 (12), 1557–1559. doi:10.1164/rccm.202003-0693LE

Cambien, F., Costerousse, O., Tiret, L., Poirier, O., Lecerf, L., Gonzales, M. F., et al. (1994). Plasma level and gene polymorphism of angiotensin-converting enzyme in relation to myocardial infarction. *Circulation* 90 (2), 669–676. doi:10.1161/01.cir.90.2.669

Chaoxin, J., Daili, S., Yanxin, H., Ruwei, G., Chenlong, W., and Yaobin, T. (2013). The influence of angiotensin-converting enzyme 2 gene polymorphisms on type 2 diabetes mellitus and coronary heart disease. *Eur. Rev. Med. Pharmacol. Sci.* 17 (19), 2654–2659.

COVID-19 Treatment Guidelines (2019). COVID-19 treatment Guidelines. Available at: https://www.covid19treatmentguidelines.nih.gov/.

D'Ardes, D., Boccatonda, A., Rossi, I., Guagnano, M. T., Santilli, F., Cipollone, F., et al. (2020). COVID-19 and RAS: Unravelling an unclear relationship. *Int. J. Mol. Sci.* 21 (8), E3003. doi:10.3390/ijms21083003

Danser, A. H., Schalekamp, M. A., Bax, W. A., van den Brink, A. M., Saxena, P. R., Riegger, G. A., et al. (1995). Angiotensin-converting enzyme in the human heart. Effect of the deletion/insertion polymorphism. *Circulation* 92 (6), 1387–1388. doi:10.1161/01.cir.92.6.1387

de Araújo, J. L. F., Menezes, D., Aguiar, R. S. D., and Souza, R. P. D. (2022). IFITM3, FURIN, ACE1, and TNF-α genetic association with COVID-19 outcomes: Systematic review and meta-analysis. *Front. Genet.* 13, 775246. doi:10.3389/fgene.2022.775246

Delanghe, J. R., Speeckaert, M. M., and De Buyzere, M. L. (2020). ACE polymorphism and COVID-19 outcome. *Endocrine* 70 (1), 13–14. doi:10.1007/s12020-020-02454-7

Delanghe, J. R., Speeckaert, M. M., and De Buyzere, M. L. (2020). COVID-19 infections are also affected by human ACE1 D/I polymorphism. *Clin. Chem. Lab. Med.* 58 (7), 1125–1126. doi:10.1515/cclm-2020-0425

Delanghe, J. R., Speeckaert, M. M., and De Buyzere, M. L. (2020). The host's angiotensin-converting enzyme polymorphism may explain epidemiological findings in COVID-19 infections. *Clin. Chim. Acta.* 505, 192–193. doi:10.1016/j.cca.2020.03.031

Dias-Filho, C. A. A., Soares, N. J. S. J., Bomfim, M. R. Q., José Dias, C., Vidal, F. C. B., Magalhães, B. C., et al. (2021). The effect of family history of hypertension and polymorphism of the ACE gene (rs1799752) on cardiac autonomic modulation in adolescents. *Clin. Exp. Pharmacol. Physiol.* 48 (2), 177–185. doi:10.1111/1440-1681.13426

Docherty, A. B., Harrison, E. M., Green, C. A., Hardwick, H. E., Pius, R., Norman, L., et al. (2020). Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: Prospective observational cohort study. *BMJ Clin. Res. ed)* 369, m1985. doi:10.1136/bmj.m1985

Ejaz, H., Alsrhani, A., Zafar, A., Javed, H., Junaid, K., Abdalla, A. E., et al. (2020). COVID-19 and comorbidities: Deleterious impact on infected patients. *J. Infect. Public Health* 13 (12), 1833–1839. doi:10.1016/j.jiph.2020.07.014

Gao, X., Zhang, S., Gou, J., Wen, Y., Fan, L., Zhou, J., et al. (2022). Spike-mediated ACE2 down-regulation involved in the pathogenesis of SARS-CoV-2 infection. *J. Infect.* doi:10.1016/j.jinf.2022.06.030

Gard, P. R. (2010). Implications of the angiotensin converting enzyme gene insertion/deletion polymorphism in health and disease: A snapshot review. *Int. J. Mol. Epidemiol. Genet.* 1 (2), 145–157.

Garg, S., Kim, L., Whitaker, M., O'Halloran, A., Cummings, C., Holstein, R., et al. (2020). Hospitalization rates and Characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019 - COVID-NET, 14 States, march 1-30, 2020. *MMWR. Morb. Mortal. Wkly. Rep.* 69 (15), 458–464. doi:10.15585/mmwr.mm6915e3

Gheblawi, M., Wang, K., Viveiros, A., Nguyen, Q., Zhong, J. C., Turner, A. J., et al. (2020). Angiotensin-converting enzyme 2: SARS-CoV-2 receptor and regulator of the renin-angiotensin system: Celebrating the 20th anniversary of the discovery of ACE2. *Circ. Res.* 126 (10), 1456–1474. doi:10.1161/CIRCRESAHA.120.317015

Gintoni, I., Adamopoulou, M., and Yapijakis, C. (2022). The impact of *ACE* and *ACE2* gene polymorphisms in pulmonary diseases including COVID-19. *Vivo* 36 (1), 13–29. doi:10.21873/invivo.12672

Goel, R., Bloch, E. M., Pirenne, F., Al-Riyami, A. Z., Crowe, E., Dau, L., et al. (2021). ABO blood group and COVID-19: A review on behalf of the ISBT COVID-19 working group. *Vox Sang.* 116 (8), 849–861. doi:10.1111/vox.13076

Gómez, J., Albaiceta, G. M., García-Clemente, M., López-Larrea, C., Amado-Rodríguez, L., Lopez-Alonso, I., et al. (2020). Angiotensin-converting enzymes (ACE, ACE2) gene variants and COVID-19 outcome. *Gene* 762, 145102. doi:10.1016/j.gene.2020.145102

Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., et al. (2020). The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5 (4), 536–544. doi:10.1038/s41564-020-0695-z

Hou, Y., Zhao, J., Martin, W., Kallianpur, A., Chung, M. K., Jehi, L., et al. (2020). New insights into genetic susceptibility of COVID-19: An ACE2 and TMPRSS2 polymorphism analysis. *BMC Med.* 18 (1), 216. doi:10.1186/s12916-020-01673-z

Hubacek, J. A., Dusek, L., Majek, O., Adamek, V., Cervinkova, T., Dlouha, D., et al. (2021). ACE I/D polymorphism in Czech first-wave SARS-CoV-2-positive survivors. *Clin. Chim. Acta.* 519, 206–209. doi:10.1016/j.cca.2021.04.024

Jacobs, M., Lahousse, L., Van Eeckhoutte, H. P., Wijnant, S. R. A., Delanghe, J. R., Brusselle, G. G., et al. (2021). Effect of ACE1 polymorphism rs1799752 on protein levels of ACE2, the SARS-CoV-2 entry receptor, in alveolar lung epithelium. *ERJ Open Res.* 7 (2), 00940–02020. doi:10.1183/23120541.00940-2020

Karakaş Çelik, S., Çakmak Genç, G., Pişkin, N., Açikgöz, B., Altinsoy, B., Kurucu İşsiz, B., et al. (2021). Polymorphisms of ACE (I/D) and ACE2 receptor gene (Rs2106809, Rs2285666) are not related to the clinical course of COVID-19: A case study. *J. Med. Virol.* 93 (10), 5947–5952. doi:10.1002/jmv.27160

Kouhpayeh, H. R., Tabasi, F., Dehvari, M., Naderi, M., Bahari, G., Khalili, T., et al. (2021). Association between angiotensinogen (AGT), angiotensin-converting enzyme (ACE) and angiotensin-II receptor 1 (AGTR1) polymorphisms and COVID-19 infection in the southeast of Iran: A preliminary case-control study. *Transl. Med. Commun.* 6 (1), 26. doi:10.1186/s41231-021-00106-0

Li, N., Yang, X. F., Gu, H., Deng, Y., Xu, J., Ma, K., et al. (2013). Relationship of angiotensin converting enzyme gene polymorphism with diabetic retinopathy. *Zhonghua. Yan Ke Za Zhi.* 49 (1), 52–57.

Liu, C., Li, Y., Guan, T., Lai, Y., Shen, Y., Zeyaweiding, A., et al. (2018). ACE2 polymorphisms associated with cardiovascular risk in Uygurs with type 2 diabetes mellitus. *Cardiovasc. Diabetol.* 17 (1), 127. doi:10.1186/s12933-018-0771-3

Luan, Y-y., Yin, C-h., and Yao, Y-m. (2021). Update advances on C-reactive protein in COVID-19 and other viral infections. *Front. Immunol.* 12, 720363. doi:10.3389/fimmu.2021.720363

Luo, Y., Liu, C., Guan, T., Li, Y., Lai, Y., Li, F., et al. (2019). Association of ACE2 genetic polymorphisms withhypertension-related target organ damages in south Xinjiang. *Hypertens. Res.* 42 (5), 681–689. doi:10.1038/s41440-018-0166-6

Osadnik, T., Strzelczyk, J. K., Fronczek, M., Bujak, K., Reguła, R., Gonera, M., et al. (2016). Relationship of the rs1799752 polymorphism of the angiotensin-converting enzyme gene and the rs699 polymorphism of the angiotensinogen gene to the process of in-stent restenosis in a population of Polish patients with stable coronary artery disease. *Adv. Med. Sci.* 61 (2), 276–281. doi:10.1016/j.advms.2016.03.006

Ovsyannikova, I. G., Haralambieva, I. H., Crooke, S. N., Poland, G. A., and Kennedy, R. B. (2020). The role of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility and severity. *Immunol. Rev.* 296 (1), 205–219. doi:10.1111/imr.12897

Palmer, B. R., Jarvis, M. D., Pilbrow, A. P., Ellis, K. L., Frampton, C. M., Skelton, L., et al. (2008). Angiotensin-converting enzyme 2 A1075G polymorphism is associated with survival in an acute coronary syndromes cohort. *Am. Heart J.* 156 (4), 752–758. doi:10.1016/j.ahj.2008.06.013

Pan, Y., Wang, T., Li, Y., Guan, T., Lai, Y., Shen, Y., et al. (2018). Association of ACE2 polymorphisms with susceptibility to essential hypertension and dyslipidemia in Xinjiang, China. *Lipids Health Dis.* 17 (1), 241. doi:10.1186/s12944-018-0890-6

Patel, S. K., Velkoska, E., Freeman, M., Wai, B., Lancefield, T. F., and Burrell, L. M. (2014). From gene to protein-experimental and clinical studies of ACE2 in blood pressure control and arterial hypertension. *Front. Physiol.* 5, 227. doi:10.3389/fphys.2014.00227

Patel, S. K., Wai, B., Ord, M., MacIsaac, R. J., Grant, S., Velkoska, E., et al. (2012). Association of ACE2 genetic variants with blood pressure, left ventricular mass, and cardiac function in Caucasians with type 2 diabetes. *Am. J. Hypertens.* 25 (2), 216–222. doi:10.1038/ajh.2011.188

Pati, A., Mahto, H., Padhi, S., and Panda, A. K. (2020). ACE deletion allele is associated with susceptibility to SARS-CoV-2 infection and mortality rate: An epidemiological study in the Asian population. *Clin. Chim. Acta.* 510, 455–458. doi:10.1016/j.cca.2020.08.008

Richardson, S., Hirsch, J. S., Narasimhan, M., Crawford, J. M., McGinn, T., Davidson, K. W., et al. (2020). Presenting Characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York city area. *Jama* 323 (20), 2052–2059. doi:10.1001/jama.2020.6775

Rieder, M. J., Taylor, S. L., Clark, A. G., and Nickerson, D. A. (1999). Sequence variation in the human angiotensin converting enzyme. *Nat. Genet.* 22 (1), 59–62. doi:10.1038/8760

Rigat, B., Hubert, C., Alhenc-Gelas, F., Cambien, F., Corvol, P., and Soubrier, F. (1990). An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J. Clin. Invest.* 86 (4), 1343–1346. doi:10.1172/JCI114844

Rostamian, A., Ghazanfari, T., Arabkheradmand, J., Edalatifard, M., Ghaffarpour, S., Salehi, M. R., et al. (2020). Interleukin-6 as a potential predictor of COVID-19 disease severity in hospitalized patients and its association with clinical laboratory routine tests. *Immunoregulation* 3 (1), 29–36. doi:10.32598/immunoregulation.3.1.4

Saad, H., Jabotian, K., Sakr, C., Mahfouz, R., Akl, I. B., and Zgheib, N. K. (2021). The role of angiotensin converting enzyme 1 insertion/deletion genetic polymorphism in the risk and severity of COVID-19 infection. *Front. Med.* 8, 798571. doi:10.3389/fmed.2021.798571

Saadat, M. (2020). No significant correlation between ACE Ins/Del genetic polymorphism and COVID-19 infection. *Clin. Chem. Lab. Med.* 58 (7), 1127–1128. doi:10.1515/cclm-2020-0577

Sabater Molina, M., Nicolás Rocamora, E., Bendicho, A. I., Vázquez, E. G., Zorio, E., Rodriguez, F. D., et al. (2022). Polymorphisms in ACE, ACE2, AGTR1 genes and severity of COVID-19 disease. *PLoS One* 17 (2), e0263140. doi:10.1371/journal.pone.0263140

Santa Cruz, A., Mendes-Frias, A., Oliveira, A. I., Dias, L., Matos, A. R., Carvalho, A., et al. (2021). Interleukin-6 is a biomarker for the development of fatal severe acute respiratory syndrome coronavirus 2 pneumonia. *Front. Immunol.* 12, 613422. doi:10.3389/fimmu.2021.613422

Sattar, N., McInnes, I. B., and McMurray, J. J. V. (2020). Obesity is a risk factor for severe COVID-19 infection: Multiple potential mechanisms. *Circulation* 142 (1), 4–6. doi:10.1161/CIRCULATIONAHA.120.047659

Sayed-Tabatabaei, F. A., Oostra, B. A., Isaacs, A., van Duijn, C. M., and Witteman, J. C. (2006). ACE polymorphisms. *Circ. Res.* 98 (9), 1123–1133. doi:10.1161/01.RES.0000223145.74217.e7

Sriram, K., and Insel, P. A. (2020). A hypothesis for pathobiology and treatment of COVID-19: The centrality of ACE1/ACE2 imbalance. *Br. J. Pharmacol.* 177 (21), 4825–4844. doi:10.1111/bph.15082

Tiret, L., Rigat, B., Visvikis, S., Breda, C., Corvol, P., Cambien, F., et al. (1992). Evidence, from combined segregation and linkage analysis, that a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. *Am. J. Hum. Genet.* 51 (1), 197–205.

Verma, S., Abbas, M., Verma, S., Khan, F. H., Raza, S. T., Siddiqi, Z., et al. (2021). Impact of I/D polymorphism of angiotensin-converting enzyme 1 (ACE1) gene on the severity of COVID-19 patients. *Infect. Genet. Evol.* 91, 104801. doi:10.1016/j.meegid.2021.104801

Wang, L. (2020). C-reactive protein levels in the early stage of COVID-19. *Med. Mal. Infect.* 50 (4), 332–334. doi:10.1016/j.medmal.2020.03.007

WHO (2022). *Coronavirus (COVID-19) dashboard*. Geneva: World Health Organization. Available at: https://covid19.who.int/.

WHO (2021). *COVID-19 clinical management: Living guidance, 25 january 2021*. Geneva: World Health Organization. Contract No.: WHO/2019-nCoV/clinical/2021.1.

Yamamoto, N., Ariumi, Y., Nishida, N., Yamamoto, R., Bauer, G., Gojobori, T., et al. (2020). SARS-CoV-2 infections and COVID-19 mortalities strongly correlate with ACE1 I/D genotype. *Gene* 758, 144944. doi:10.1016/j.gene.2020.144944

Yang, W., Huang, W., Su, S., Li, B., Zhao, W., Chen, S., et al. (2006). Association study of ACE2 (angiotensin I-converting enzyme 2) gene polymorphisms with coronary heart disease and myocardial infarction in a Chinese Han population. *Clin. Sci.* 111 (5), 333–340. doi:10.1042/CS20060020

Zeberg, H., and Pääbo, S. (2020). The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* 587 (7835), 610–612. doi:10.1038/s41586-020-2818-3

Zhao, Q., Gu, D., Kelly, T. N., Hixson, J. E., Rao, D. C., Jaquish, C. E., et al. (2010). Association of genetic variants in the apelin-APJ system and ACE2 with blood pressure responses to potassium supplementation: The GenSalt study. *Am. J. Hypertens.* 23 (6), 606–613. doi:10.1038/ajh.2010.36

Zheng, H., and Cao, J. J. (2020). Angiotensin-converting enzyme gene polymorphism and severe lung injury in patients with coronavirus disease 2019. *Am. J. Pathol.* 190 (10), 2013–2017. doi:10.1016/j.ajpath.2020.07.009

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 08 September 2022
DOI 10.3389/fgene.2022.933148



OPEN ACCESS

EDITED BY
Zhongshan Cheng,
St. Jude Children's Research Hospital, United States

REVIEWED BY
Chunyu Li,
Sichuan University, China
Dongqing Gu,
Army Medical University, China

*CORRESPONDENCE
Ancha Baranova,
abaranov@gmu.edu
Lei Cai,
lcai@sjtu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 30 April 2022
ACCEPTED 08 July 2022
PUBLISHED 08 September 2022

CITATION
Cai L, He C, Liu Y, Sun Y, He L and Baranova A (2022), Inflammation and immunity connect hypertension with adverse COVID-19 outcomes. *Front. Genet.* 13:933148. doi: 10.3389/fgene.2022.933148



# Inflammation and immunity connect hypertension with adverse COVID-19 outcomes

Lei Cai[1*†], Chuan He[1†], Yonglin Liu[2†], Yanlan Sun[1], Lin He[1,3] and Ancha Baranova[4*]

[1]Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Collaborative Innovation Center for Genetics and Development, Shanghai Mental Health Center, Shanghai Jiaotong University, Shanghai, China, [2]Sanya Women and Children's Hospital managed by Shanghai Children's Medical Center, Sanya, China, [3]Shanghai Center for Women and Children's Health, Shanghai, China, [4]School of Systems Biology, George Mason University, Fairfax, VA, United States

**Objectives:** To explore the connection of hypertension and severe COVID-19 outcomes.

**Methods:** A total of 68 observational studies recording mortality and/or general severity of COVID-19 were pooled for meta-analyses of the relationship of severe COVID-19 outcomes with hypertension as well as systolic and diastolic blood pressure. Genome-wide cross-trait meta-analysis (GWCTM) was performed to explore the genes linking between hypertension and COVID-19 severity.

**Results:** The results of meta-analysis with the random effect model indicated that pooled risk ratios of hypertension on mortality and severity of COVID-19 were 1.80 [95% confidence interval (CI) 1.54–2.1] and 1.78 (95% confidence interval 1.56–2.04), respectively, although the apparent heterogeneity of the included studies was detected. In subgroup analysis, cohorts of severe and mild patients of COVID-19 assessed in Europe had a significant pooled weighted mean difference of 6.61 mmHg (95% CI 3.66–9.55) with no heterogeneity found ($p$ = 0.26). The genes in the shared signature of hypertension and the COVID-19 severity were mostly expressed in lungs. Analysis of molecular networks commonly affected both by hypertension and by severe COVID-19 highlighted CCR1/CCR5 and IL10RB signaling, as well as Th1 and Th2 activation pathways, and also a potential for a shared regulation with multiple sclerosis.

**Conclusion:** Hypertension is significantly associated with the severe course of COVID-19. Genetic variants within inflammation- and immunity-related genes may affect their expression in lungs and confer liability to both elevated blood pressure and to severe COVID-19.

KEYWORDS

**COVID-19, hypertension, DBP, SBP, meta-analysis, multiple sclerosis**

# Introduction

The coronavirus disease 2019 (COVID-19) pandemic, which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has created a public health crisis worldwide. This disease is associated with a wide spectrum of clinical manifestations ranging from mild to life-threatening, with a possibility of adverse outcomes. Across 2 years, the prevention and therapy went through a few rounds of optimization, which decreased the global fatality rate for COVID-19 to 1.22% (COVID-19. who.int/). Further improvements require comprehensive understanding of COVID-19 pathogenesis through elucidating both the viral and the host determinants of the severe course of COVID-19.

Hypertension is a common disease defined as the systolic blood pressure (BP) readings on two different days being over 140 mmHg, and/or the diastolic BP readings ≥90 mmHg (Organization, 2019). While hypertension is repeatedly reported as one of the predictors of adverse SARS-CoV-2 infection outcomes in various cohorts (Manohar, 2021; Stanetic et al., 2021; Kumaran et al., 2022), the connection of this multifactorial condition and severe COVID-19 remains not very well characterized, especially when compared to that of obesity or diabetes (Madjid et al., 2020; Schiffrin, 2020). Relative lack of systemic efforts in clinical dissection of interactions between hypertension and COVID-19 is surprising, especially in light of an active involvement of ACE2, a receptor for SARS-CoV-2, in the hypertension control (Verano-Braga et al., 2020). On the other hand, an importance of hypertension as a factor contributing to COVID-19 outcomes is underlined by notions that adequate control of BP is one of the prerequisites for alleviating COVID-19-associated organ damage (Zheng et al., 2020; Hessami et al., 2021; Wu et al., 2021).

The genome-wide association study (GWAS) is a powerful tool to identify genetic variants contributing to complex phenotypic traits, which are influenced by many factors at once. Under the assumption that many gene variants may be associated with multiple traits, cross-phenotype (CP) associations analyses were made possible recently by introducing the CPASSOC package that uses summary-level data from GWAS to analyze multiple phenotypes for each SNP by accounting for their correlations among traits and among cohorts (Park et al., 2016; Li and Zhu, 2017). We mined multiple existing GWASs that have reported various outcomes related to COVID-19, including hospitalization, severe respiratory problems, and even death in an attempt to extract a list of genes possibly contributing to hypertension and the severity of COVID-19. We analyzed the functions of these genes and extracted additional insights into common pathophysiology of these two diseases.

# Methods and materials

## Literature search strategy

An extensive search was performed within the databases of PubMed, COVID-19 Portfolio, Embase, Scopus, and China National Knowledge Infrastructure (CNKI) by using the keywords of (COVID-19 OR SARS-CoV-2 OR coronavirus) AND (severity OR clinical outcome) AND (hypertension OR blood pressure). All relevant sources from 30 December 2019 to 20 June 2021 were retrieved without language restrictions. Two authors (CH and YS) independently reviewed the collected literature works. Furthermore, the reference lists of all relevant studies were also manually checked for additional entries.

## Study selection and evaluation

Eligible studies were considered when they met the following criteria: 1. there was properly established COVID-19 diagnosis; 2. complete data for survivor/non-survivor or severe COVID-19 infection/mild; 3. complete data for hypertension prevalence or BP measurement; and 4. complete information for subject characteristics. Here, COVID-19 diagnosis was based on SAR-CoV-2 reverse transcription-polymerase chain reaction (RT-PCR) testing of nasopharyngeal (NP) swab, throat swab, or other types of respiratory sampling, and the severe COVID-19 infection was detected in patients requiring oxygen support, or admitted to intensive care, or reported as dead. In case of any degree of non-clarity in the data, the corresponding authors were contacted for full information. The exclusion criteria were as follows: 1. duplicated studies; 2. meta-analyses, reviews, case reports, and nonhuman studies; 3. containing COVID-19 patients with pneumonia or other lung diseases; 4. containing pediatric patients; and 5. the quality assessment scores of studies were below 7. The quality score of each study was less than 7 based on the principles of AHRQ (Rostom, 2004) and QUADAS (JB, 2003), which was assessed based on five items (Supplementary Table S1). The third party (LC and YL) took part in discussion to solve the disagreement of the evaluation result.

## Data extraction

The following information was extracted from each study: first author's name, publication year, period of patients' admission, study type, country, sample size, age, gender, hypertension prevalence, systolic blood pressure, diastolic blood pressure, cases of non-survivors or severe COVID-19 infection, and controls of non-survivor of mild COVID-19 infection were recorded into a standardized information sheet.

## Genome-wide cross-trait meta-analysis

The cross-phenotype association (CPASSOC) approach (Zhu et al., 2015) was employed to identify genetic variants shared between COVID-19 and hypertension. CPASSOC allows for the presence of heterogeneous effects across traits and provides $S_{Het}$ statistics and *p*-values weighted by a sample size. In two-step CPASSOC analysis, the correlation matrix was built with SNPs whose summary statistics Z-scores were greater than 1.96 or less than −1.96 and which had a linkage disequilibrium (LD) pattern from 1,000 Genomes Project phase 3, and then $S_{Hom}$ and $S_{Het}$ tests were performed. $S_{Hom}$ is more powerful when heterogeneity is not present, while $S_{Het}$ allows for trait heterogeneity. Significant levels of less than $5 \times 10^{-8}$ were used as cut-off values. Here, λmeta statistics were calculated to test the possibility of sample overlap through measuring concordance of effect sizes (Chen et al., 2016). Under the null hypothesis, λmeta = 1 means that the pair of cohorts are completely independent, while λmeta <1 indicates samples overlap between cohorts.

## Identification of shared genes and enrichment analysis

The FUMA (v1.3.5 d) (https://fuma.ctglab.nl/), a platform for annotation, visualization, and interpretation of GWAS results, was utilized for functional mapping of the genes found by GWCTM (Watanabe et al., 2017), with the summary statistics obtained from GWCTM treated as input. In the beginning, positional gene mapping of SNP2GENE function within FUMA was carried out. Then, the tissue specificity of these genes was explored with GENE2FUNC within FUMA and GTEx v8 54 tissue-types data. Finally, the eQTL gene mapping was carried out within the aforementioned identified specific tissue of the GTEx v8 dataset.

IPA (Ingenuity Pathway Analysis) software (Ingenuity Systems; Qiagen China Co., Ltd.) was employed to perform core enrichment analysis in human lung tissue with genes identified for COVID-19 severity and its adverse outcomes, respectively. The Gene Ontology (GO) enrichment analysis was performed as previously described (Cai et al., 2018).

## Statistical analysis

For studies reporting the interquartile range, the standard deviation values were obtained as described in the Cochrane Handbook for Systematic Reviews (Cumpston et al., 2019). The weighted mean differences (WMDs) were calculated with 95% confidence intervals (CIs) (Cai et al., 2015). The heterogeneity was evaluated using the Q-test and $I^2$ statistic (Cai et al., 2013). $I^2 > 50\%$ or $p < 0.1$ were considered significantly heterogeneous. A fixed-effects model was used when the result showed no significant heterogeneity; otherwise, a random-effects model was applied. The publication bias was evaluated by Egger's regression test (Lei et al., 2005). All the analyses were conducted in accordance with the Cochrane Handbook for Systematic Reviews (Version 5.0) or R software, with two-sided $p < 0.05$ set for statistical significance.

# Result

## Data collection

The process of literature selection is shown in Supplementary Figure S1. Through electronic-database searching and manual examination of the reference lists, we collected a total of 4,284 records. After three rounds of filtering, a total of 68 studies were retained for further statistics analysis, among which there were 21 studies containing 32,015 COVID-19 patients with the outcome of survivor/non-survivor, and 47 studies including 230,941 COVID-19 patients with the outcome of clinical severity or mildness. The main characteristics of studies included in the current meta-analysis are summarized in Table 1.

GWAS summary datasets for hypertension and COVID-19 are listed in Table 2. Briefly, we utilized one GWAS dataset on hypertension with the largest sample size as found in the MR-base database (Elsworth, 2020) and three GWAS datasets on COVID-19 of varying severity, including recorded cases of death, confirmed very severe respiratory COVID-19, and hospitalized COVID-19.

## Association of severity of COVID-19 and hypertension

In mortality meta-analysis of 16,548 survivors and 3,297 non-survivors, the random-effects model was optimal. Using this model, hypertension had an estimated pooled RR of 1.80 (95% CI 1.54–2.1) for COVID-19-related death, with heterogeneity being detected ($I^2$ = 91%, $p < 0.001$). With univariable meta-regression model and 1,000 permutations in the permutation test, both age and gender displayed significant influence on the pooled RR, $p$ = 0.001 and 0.015, respectively. Taking separately, age and gender regression models explained 66.76 and 26.29% of heterogeneity, respectively. Interestingly, in each sequentially increasing age bracket, the RR of hypertension was decreasing by 0.03 (95% CI −0.043 to −0.019).

In meta-analysis of 19,011 severe and 211,930 mild cases of COVID-19, the random-effects model detected hypertension-related pooled RR of 1.78 (95% CI 1.56–2.04) for severe course of COVID-19, also with evidence of heterogeneity ($I^2$ = 97.2%, $p <$ 0.001). With univariable meta-regression model and 1,000 permutations in the permutation test, age, but not gender had significant influence on the pooled RR, with $p$ =

TABLE 1 Characteristics of studies included in the current meta-analysis.

| Study | Period | Outcome | Country | Number of subjects | | Age (year) | | Gender (male %) | | Number of hypertension | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Case | Control | Case | Control | Case | Control | Case | Control |
| Souza, et al., 2021 | March–May 2020 | Dead/survivor | India | 156 | 533 | 55 ± 29.9 | 38 ± 29.7 | 60.26 | 45.78 | 33 | 33 |
| Surendra, et al., 2021 | March–July 2020 | Dead/survivor | Indonesia | 497 | 3,768 | 58.7 ± 11.9 | 43.3 ± 18 | 61 | 51 | 184 | 594 |
| Pareek, et al., 2021 | March–May 2020 | Dead/survivor | United States of America | 82 | 504 | 78.8 ± 15.8 | 65.7 ± 17.8 | 63 | 50.8 | 60 | 293 |
| Muhammad, et al., 2021 | March–May 2020 | Dead/survivor | United States of America | 45 | 155 | 67.1 ± 13.7 | 58 ± 14.9 | 64.4 | 59.4 | 35 | 95 |
| Marcolino, et al., 2021 | March–September 2020 | Dead/survivor | Brazil | 439 | 1,551 | 70 ± 16.4 | 56 ± 17.8 | 54.4 | 51.8 | 310 | 746 |
| Kim, et al., 2021 | February–July 2020 | Dead/survivor | Korea | 179 | 2075 | 78.3 ± 10.5 | 54.2 ± 21.5 | 53.1 | 34.3 | 112 | 534 |
| Halem, et al., 2020 | March–April 2020 | Dead/survivor | Belgium | 81 | 238 | 81 ± 7.5 | 69.6 ± 14.9 | 61.7 | 57.98 | 61 | 101 |
| Thompson, et al., 2020 | March–May 2020 | Dead/survivor | United Kingdom | 169 | 301 | 78.3 ± 12.1 | 63.4 ± 17.6 | 54.4 | 54.1 | 101 | 117 |
| Rodriguez-Nava, et al., 2021 | March–May 2020 | Dead/survivor | United States of America | 101 | 212 | 74.4 ± 12.8 | 65.8 ± 14.2 | 64.4 | 55.2 | 79 | 143 |
| Huang, et al., 2020 | January–April 2020 | Dead/survivor | China | 140 | 536 | 66.3 ± 13.5 | 50.3 ± 20.8 | 69.3 | 40.5 | 84 | 142 |
| Bonnet, et al., 2021 | February–April 2020 | Dead/survivor | France | 360 | 2,503 | 80.4 ± 12 | 64.6 ± 16.7 | 61.7 | 57.3 | 261 | 1,186 |
| Diebold, et al., 2021 | February–May 2020 | Dead/survivor | Switzerland | 88 | 855 | 75.4 ± 11.3 | 63.4 ± 17.1 | 72 | 72 | 59 | 388 |
| Basu, et al., 2021 | March–May 2020 | Dead/survivor | United Kingdom | 361 | 546 | 77.8 ± 12.4 | 66.3 ± 17.6 | 62 | 50 | 287 | 343 |
| Novelli, et al., 2021 | February–March 2020 | Dead/survivor | Italy | 171 | 337 | 75.2 ± 9.2 | 61.7 ± 16.5 | 79.5 | 68.8 | 120 | 151 |
| Wang, et al., 2020 | January–February 2020 | Dead/survivor | China | 116 | 177 | 72.7 ± 12.1 | 49.5 ± 22.2 | 56 | 41.2 | 66 | 26 |
| Zhou, et al., 2020 | December 2019–January 2020 | Dead/survivor | China | 54 | 137 | 69.4 ± 9.9 | 51.6 ± 9.7 | 70 | 69 | 26 | 32 |
| Huang, et al., 2020 | January–March 2020 | Dead/survivor | China | 16 | 283 | 69.2 ± 9.7 | 52.5 ± 16.6 | 68.8 | 52.7 | 11 | 63 |
| Iaccarino, et al., 2020 | March–April 2020 | Dead/survivor | Italy | 188 | 1,403 | 79.6 ± 0.8 | 64.7 ± 0.4 | 66.5 | 63.6 | 137 | 737 |
| Russo, et al., 2020 | February–April 2020 | Dead/survivor | Italy | 35 | 157 | 77 ± 8.3 | 65.5 ± 15.6 | 57.1 | 60.5 | 27 | 84 |
| Wang, et al., 2020 | January–February 2020 | Dead/survivor | China | 19 | 277 | 65.6 ± 12.6 | 46.0 ± 14.4 | 57.9 | 46.6 | 9 | 33 |
| Rodilla, et al., 2021 | March–June 2020 | Dead/survivor | Spain | 2,606 | 9,564 | 79.7 ± 10.5 | 64.1 ± 15.7 | 62.2 | 55.2 | 1842 | 4,352 |
| Thibeault, et al., 2021 | March–June 2020 | Severe/mild | German | 71 | 90 | 57.6 ± 22.6 | 62.7 ± 13.6 | 70.4 | 61.1 | 45 | 38 |
| Shen, et al., 2021 | January–February 2020 | Severe/mild | China | 32 | 291 | 49.78 ± 17 | 62.22 ± 14.68 | 56.25 | 52.58 | 11 | 51 |
| Pouw, et al., 2021 | March–May 2020 | Severe/mild | Netherland | 476 | 476 | 67.6 ± 16.4 | 29.3 ± 11.9 | 68.3 | 58.8 | 195 | 179 |
| Otoshi, et al., 2021 | April–November 2020 | Severe/mild | Japan | 46 | 254 | 62.6 ± 17.3 | 72.2 ± 10.3 | 30.3 | 42.9 | 25 | 96 |

(Continued on following page)

TABLE 1 (*Continued*) Characteristics of studies included in the current meta-analysis.

| Study | Period | Outcome | Country | Number of subjects | | Age (year) | | Gender (male %) | | Number of hypertension | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Case | Control | Case | Control | Case | Control | Case | Control |
| Mollinedo-Gajate, et al., 2021 | March–April 2020 | Severe/mild | Spain | 62 | 131 | 62.6 ± 15 | 67.5 ± 16.3 | 64.5 | 52.7 | 30 | 56 |
| | July–August 2020 | Severe/mild | Spain | 24 | 59 | 69.9 ± 28.1 | 63.6 ± 27 | 70.8 | 52.5 | 14 | 26 |
| Zhang, et al., 2021 | January–February 2020 | Severe/mild | China | 51 | 121 | 42.3 ± 17.2 | 60.6 ± 13.7 | 58.8 | 51.2 | 16 | 16 |
| Scho¨nfeld, et al., 2021 | March–October 2020 | Severe/mild | Argentina | 13,389 | 193,690 | 40 ± 18.5 | 70.65 ± 15.6 | 58.1 | 50.6 | 6,981 | 32,852 |
| Vlachos, et al., 2021 | February–March 2020 | Severe/mild | United Kingdom | 76 | 353 | 68 ± 20.8 | 55.9 ± 11.3 | 66 | 52 | 38 | 187 |
| Sim, et al., 2020 | February–May 2020 | Severe/mild | Malaysia | 471 | 5,418 | 31.3 ± 10.4 | 57.6 ± 12.6 | 71.5 | 71.7 | 229 | 702 |
| Matangila, et al., 2020 | March–July 2020 | Severe/mild | Congo | 19 | 92 | 48.9 ± 19.6 | 51.4 ± 22.4 | 53 | 45 | 10 | 24 |
| Vial, et al., 2020 | March–April 2020 | Severe/mild | Chile | 18 | 70 | 47.4 ± 15.1 | 66.3 ± 10.5 | 83.3 | 40 | 10 | 15 |
| Mutair, et al., 2020 | April–May 2020 | Severe/mild | Saudi Arabia | 160 | 241 | 37.32 ± 13.6 | 39.43 ± 13.1 | 81.3 | 79.3 | 33 | 26 |
| He, et al., 2020 | January–March 2020 | Severe/mild | China | 501 | 530 | 57.5 ± 15.6 | 65.3 ± 11.9 | 59.3 | 52.2 | 237 | 146 |
| Xiong, et al., 2020 | February–March 2020 | Severe/mild | China | 55 | 61 | 52.1 ± 20.5 | 64.4 ± 17.5 | 69.1 | 68.9 | 26 | 19 |
| Jourdes, et al., 2020 | March–April 2020 | Severe/mild | France | 50 | 213 | 64.4 ± 17.1 | 65.2 ± 13 | 66 | 57.3 | 19 | 85 |
| Popov, et al., 2020 | March–June 2020 | Severe/mild | Bulgaria | 43 | 95 | 48.3 ± 15.7 | 63 ± 12.8 | 76.8 | 56.9 | 32 | 37 |
| Nachega, et al., 2020 | March–July 2020 | Severe/mild | Congo | 191 | 575 | 42.7 ± 16.4 | 55.9 ± 14.4 | 71.1 | 63.8 | 87 | 107 |
| Wei, et al., 2020 | January–March 2020 | Severe/mild | China | 14 | 262 | 48.6 ± 13.4 | 66 ± 10.5 | 71.4 | 55.3 | 8 | 39 |
| Guan, et al., 2020 | January 2020 | Severe/mild | China | 173 | 926 | 45.4 ± 17.1 | 52.4 ± 16.7 | 57.8 | 58.2 | 41 | 124 |
| Yue, et al., 2020 | January–February 2020 | Severe/mild | China | 44 | 42 | 41.7 ± 20.2 | 42.5 ± 17.2 | 52.3 | 35.7 | 5 | 1 |
| Charlotte, et al., 2020 | March–April 2020 | Severe/mild | Switzerland | 49 | 147 | 72.6 ± 16.5 | 63.9 ± 11.5 | 61 | 60 | 27 | 91 |
| Wang, et al., 2020 | January–February 2020 | Severe/mild | China | 45 | 230 | 46.1 ± 21.6 | 62.1 ± 13 | 57.8 | 44.3 | 18 | 36 |
| Wang, et al., 2021 | January–February 2020 | Severe/mild | China | 25 | 122 | 40.9 ± 13.5 | 52.5 ± 16.5 | 76 | 58.2 | 8 | 11 |
| Petrilli, et al., 2020 | March–April 2020 | Severe/mild | United States of America | 990 | 1739 | 59.6 ± 17.1 | 68 ± 14.8 | 66.3 | 58.4 | 680 | 1,013 |
| Taha, et al., 2021 | March–April 2021 | Severe/mild | Egypt | 50 | 130 | 51.1 ± 59.2 | 57.5 ± 52.7 | 44 | 31.5 | 29 | 36 |
| Huang, et al., 2020 | January 2020 | Severe/mild | China | 13 | 28 | 49.2 ± 12.9 | 50.5 ± 16.6 | 85 | 68 | 2 | 4 |
| Krishna, et al., 2021 | March–August 2020 | Severe/mild | United States of America | 70 | 109 | 58.6 ± 17.3 | 62.8 ± 20.4 | 46 | 60 | 37 | 47 |
| Padmaprakash, et al., 2021 | April–August 2020 | Severe/mild | India | 175 | 1,361 | 57 | 35 | 80.5 | 89.3 | 59 | 71 |

(Continued on following page)

TABLE 1 (*Continued*) Characteristics of studies included in the current meta-analysis.

| Study | Period | Outcome | Country | Number of subjects | | Age (year) | | Gender (male %) | | Number of hypertension | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Case | Control | Case | Control | Case | Control | Case | Control |
| Pandita, et al., 2021 | February–May 2020 | Severe/mild | United States of America | 91 | 168 | 60.4 ± 17.2 | 65.4 ± 13.8 | 55 | 52.4 | 63 | 101 |
| Alhumaid, et al., 2021 | March–July 2020 | Severe/mild | Saudi Arabia | 205 | 809 | 45.3 ± 19.3 | 52.9 ± 17.3 | 56.5 | 57.6 | 111 | 165 |
| Nasir, et al., 2021 | February–June 2020 | Severe/mild | Pakistan | 137 | 156 | 42.3 ± 16.2 | 59.4 ± 14.1 | 76.6 | 55.1 | 72 | 30 |
| Neto, et al., 2021 | March–May 2020 | Severe/mild | Brazil | 300 | 206 | 58.4 ± 15.5 | 61.2 ± 14.7 | 58.7 | 55.3 | 168 | 63 |
| Ser, et al., 2021 | April 2020 | Severe/mild | Spain | 20 | 42 | 80.5 ± 3.2 | 82.9 ± 4.2 | 60 | 28.6 | 12 | 19 |
| Khan, et al., 2020 | March 2020 | Severe/mild | Saudi Arabia | 77 | 571 | 33 ± 18 | 37 ± 27 | 67.5 | 50.8 | 16 | 59 |
| Sami, et al., 2020 | February–December 2020 | Severe/mild | Iran | 90 | 400 | 55.52 ± 14.45 | 61.32 ± 16.99 | 73 | 58 | 36 | 135 |
| Xiong, et al., 2020 | January–March 2020 | Severe/mild | China | 65 | 407 | 41.6 ± 14.9 | 51 ± 21.2 | 58.5 | 52.1 | 22 | 49 |
| Mughal, et al., 2020 | March–April 2020 | Severe/mild | United States of America | 30 | 99 | 56.8 ± 24.8 | 64.4 ± 10.9 | 83.3 | 56.6 | 14 | 42 |
| Claudia, et al., 2020 | February–April 2020 | Severe/mild | Switzerland | 35 | 64 | 65.3 ± 15.2 | 66.9 ± 13.9 | 80 | 53 | 19 | 37 |
| Li, et al., 2020 | January–February 2020 | Severe/mild | China | 269 | 279 | 55 ± 16.4 | 63.6 ± 13.4 | 56.9 | 45.2 | 104 | 62 |
| Ren, et al., 2020 | January–February 2020 | Severe/mild | China | 62 | 89 | 53.9 ± 16.2 | 67.6 ± 11.6 | 64.5 | 42.7 | 35 | 25 |
| Cheng, et al., 2020 | January–March 2020 | Severe/mild | China | 52 | 200 | 45.4 ± 15.5 | 62.9 ± 16 | 59.6 | 55 | 20 | 28 |
| Huang, et al., 2020 | January–February 2020 | Severe/mild | China | 23 | 179 | 43.3 ± 14.9 | 47.6 ± 19 | 73.9 | 55.3 | 2 | 29 |
| Chen, et al., 2020 | January–March 2020 | Severe/mild | China | 43 | 102 | 45.3 ± 13.6 | 52.8 ± 15.5 | 53.5 | 54.9 | 9 | 13 |
| Wang, et al., 2020 | January 2020 | Severe/mild | China | 36 | 102 | 50.8 ± 5 | 67.1 ± 16.2 | 54.3 | 52 | 21 | 22 |
| Zhang, et al., 2020 | January–February 2020 | Severe/mild | China | 58 | 82 | 51.6 ± 10.7 | 62.7 ± 13.5 | 56.9 | 46.3 | 22 | 20 |
| Gao, et al., 2020 | January–February 2020 | Severe/mild | China | 15 | 28 | 52.96 ± 14 | 45.2 ± 7.68 | 60 | 60.7 | 6 | 7 |
| Lendorf, et al., 2020 | March–May 2020 | Severe/mild | Denmark | 20 | 91 | 69.8 ± 16.6 | 62.9 ± 15.2 | 85 | 55 | 9 | 29 |

TABLE 2 GWAS summary dataset information.

| Data | Source | Trait | Population | Case | Control | N |
|---|---|---|---|---|---|---|
| Hypertension | https://gwas.mrcieu.ac.uk/datasets/ukb-b-12493/ | Essential (primary) hypertension, SBP >140 or DBP >90 vs. population | European | 54,358 | 408,652 | 463,010 |
| COVID-19 death | https://grasp.nhlbi.nih.gov/ | Positive and dead COVID-19 vs. population | European | 1,001 | 458,249 | 459,250 |
| COVID-19 severe respiration | https://www.covid19hg.org/ | Very severe respiratory confirmed COVID-19 vs. population | European | 4,792 | 1,054,664 | 1,059,456 |
| COVID-19 hospitalization | https://www.covid19hg.org/ | Hospitalized COVID-19 vs. population | European | 9,986 | 1,877,672 | 1,887,658 |

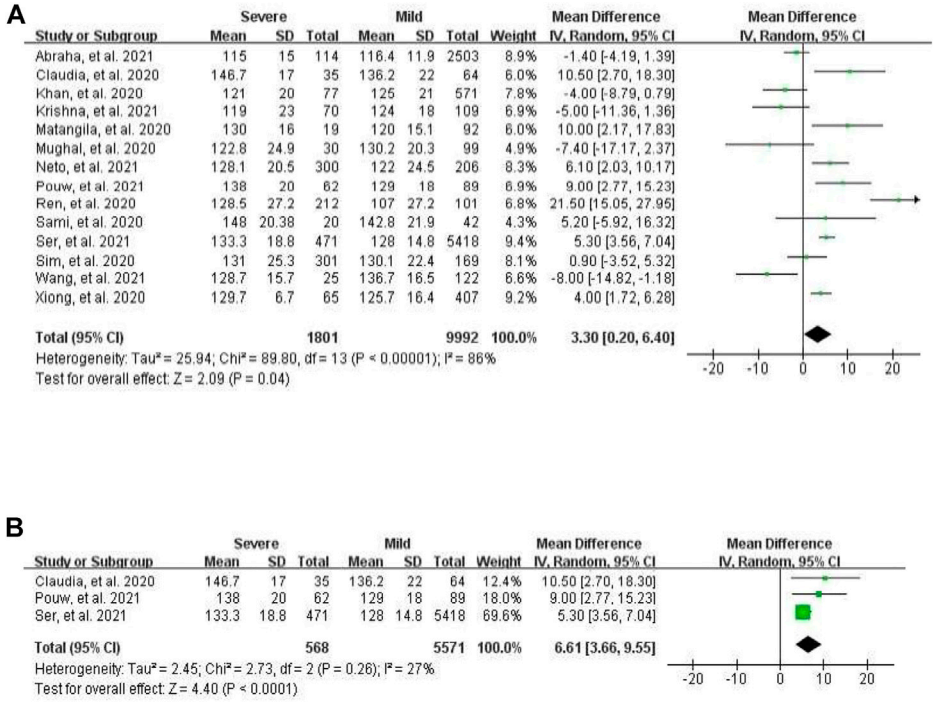



FIGURE 1
Forest plot for the meta-analysis of mean difference of SBP between severe and mild cases. **(A)** In all populations; **(B)** in the European population.

0.009 and 0.25, respectively. Taking separately, age and gender regression models explained 18.00 and 2.92% of heterogeneity, respectively. Similar to that observed in mortality analysis, in each sequentially increasing age bracket, the RR of hypertension was decreasing by 0.02 (95% CI −0.036 to −0.007).

For both types of analyses, mortality, and severity, the likelihood ratio tests pointed at better fits provided by the univariable meta-regression models than multivariable meta-regression models. For the mortality study, the sensitivity analysis confirmed that omitting any one study had no effect on the pooled RR (Supplementary Figure S2A), with no risk of publication bias ($p > 0.05$). For the severity study, Schönfeld's study performed differently than others, and its omission affected RR (Supplementary Figure S2B), with risk of publication bias detected ($p = 0.03$).

## Association of severity of COVID-19 with blood pressure

For systolic blood pressure (SBP), a total of 14 studies, which profiled cohorts with either severe or mild course of COVID-19, were analyzed. The pooled weighted mean difference between these two cohorts was significant at 3.30 mmHg (95% CI 0.20–6.40), with the random-effects model pointing at evident heterogeneity ($I^2 = 86\%$, $p < 0.0001$, Figure 1A). In further subgroup analysis of three European studies, there was significant pooled weighted mean difference of 6.61 mmHg between severe and mild patients (95% CI 3.66–9.55), with no apparent heterogeneity ($I^2 = 27\%$, $p = 0.26$, Figure 1B). When a total of nine mortality studies were analyzed, the pooled weighted mean difference between survivors and non-survivors was at -1.54 mmHg (95% CI -4.33 to 1.24), with moderate heterogeneity ($I^2 = 58\%$, $p = 0.01$).

For diastolic blood pressure (DBP), a total of 13 studies, which profiled cohorts of patients with either severe or milder form of COVID-19, were analyzed. The pooled mean difference between mild and severe groups was −0.44 mmHg and non-significant (95% CI −1.52 to 0.64), with moderate heterogeneity ($I^2 = 45\%$, $p = 0.04$). In four studies presenting DBP values of deceased sufferers and survivors, the pooled weighted mean difference was 2.95 mmHg and non-significant (95 percent CI −3.28–9.18), with evidence of heterogeneity ($I^2 = 91\%$, $p < 0.0001$).

## The shared genes between severe course of COVID-19 and hypertension

In the analyzed GWAS dataset, λmeta values were 1.06 ± 0.01, 1.04 ± 0.01, and 1.10 ± 0.01 between hypertension and three adverse COVID-19 outcomes, i.e., death, very severe respiratory problems, and hospitalization. CPASSOC-driven genome-wide

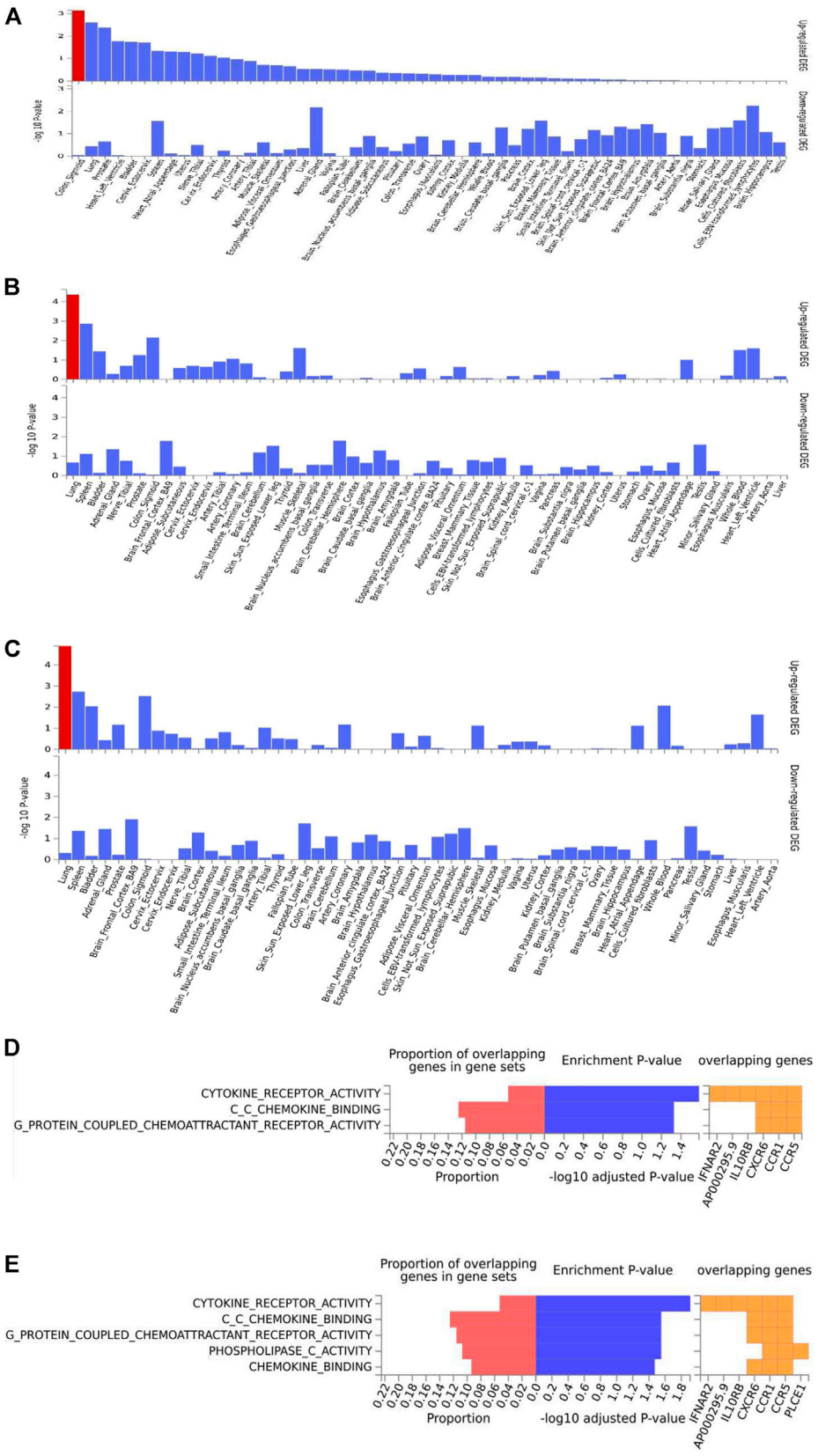



FIGURE 2
Enrichment analysis of shared genes between hypertension and adverse outcomes of COVID-19, i.e., death, very severe respiration, and hospitalization. **(A–C)** Tissue enrichment of shared genes between hypertension and COVID-19 death, very severe respiration, and hospitalization, respectively. **(D)** and (E) GO function enrichment of shared genes between hypertension and COVID-19 very severe respiration and hospitalization, respectively.

TABLE 3 Top Canonical pathways identified with IPA.

| Name | *p*-value | Overlap | Gene |
|---|---|---|---|
| Genes for hypertension with COVID-19 hospitalization | | | |
| Pathogenesis of multiple sclerosis | 4.38E-04 | 22.2% (2/9) | CCR1 and CCR5 |
| IL-10 signaling | 1.65E-03 | 4.5% (3/66) | CCR1, CCR5, and IL10RB |
| Th1 and Th2 activation pathway | 1.95E-03 | 2.7% (4/150) | CCR1, CCR5, CXCR6, and IL10RB |
| G-protein-coupled receptor signaling | 2.37E-03 | 1.3% (7/523) | CCR1, CCR5, CHP1, CXCR6, MAP3K11, NPR3, and PLCE1 |
| Cardiac hypertrophy signaling (enhanced) | 5.29E-03 | 1.3% (6/455) | ACE, CHP1, IL10RB, MAP3K11, PLCE1, and WNT2B |
| Genes for hypertension with COVID-19 very severe respiration | | | |
| Pathogenesis of multiple sclerosis | 5.61E-04 | 22.2% (2/9) | CCR1 and CCR5 |
| IL-10 signaling | 2.36E-03 | 4.5% (3/66) | CCR1, CCR5, and IL10RB |
| Th1 and Th2 activation pathway | 3.08E-03 | 2.7% (4/150) | CCR1, CCR5, CXCR6, and IL10RB |
| Th2 pathway | 1.24E-02 | 2.5% (3/120) | CCR1, CCR5, and CXCR6 |
| Oxytocin signaling pathway | 1.46E-02 | 1.7% (4/235) | CHP1, MFN2, NOS3, and NPR3 |

cross-trait meta-analysis was performed to search for a common variant contributing both to hypertension and to adverse COVID-19 outcomes, including death, very severe respiratory problems, and hospitalization.

The summary statistics from the aforementioned GWCTM results were further explored by GENE2FUNC implemented in the platform of FUMA. A total of 149 genes were identified as shared between hypertension and COVID-19 death by positional gene mapping of SNP2GENE function (Supplementary Figure S3A); a total of 222 genes were identified as shared between hypertension and very severe, respiratory confirmed COVID-19 by positional gene mapping of SNP2GENE function (Supplementary Figure S3B); and a total of 187 genes were identified as shared between hypertension and the outcomes of the requirement of oxygen support and hospitalization (Supplementary Figure S3C). In both shared gene lists, the top positions were occupied by *CLCN6*, *MTHFR*, *C10orf107*, *FES*, and *FURIN*. Gene expression analysis of shared gene sets highlighted the sigmoid colon as the most relevant to the association of COVID-19 death (Figure 2A), suggesting that the disruption of gut homeostasis in the course of COVID-19 (Varshney et al., 2021) may contribute to COVID-19 mortality disproportionally.

On the other hand, a similarly executed analysis pointed at the lung tissue as the most relevant association of hypertension with other severe COVID-19 outcomes (Figures 2B,C). After integrating GWAS results with lung eQTL data from GTEx (version 8), a total of 67 protein-coding genes were pinpointed as shared between hypertension and the severe respiratory involvement in COVID-19 (Supplementary Table S2) and a total of 57 genes were shared between hypertension and the hospitalization due to COVID-19 (Supplementary Table S3). A majority of these genes were involved in immune function, with cytokine receptor pathways, chemokine binding, and the signaling by G-protein-coupled chemoattractant receptors being particularly highlighted (Figures 2C,D).

The analysis of top canonical pathways identified three common sets of shared genes, namely, pathogenesis of multiple sclerosis, IL-10 signaling, and Th1 and Th2 activation (Table 3). Furthermore, these three sets of genes were enriched within the sets describing some other relevant pathophysiological conditions, including "infectious diseases," "organismal injury and abnormalities," "cellular function of altering cell morphology," and "connective tissue development."

## Discussion

Blood pressure is commonly recorded as two numbers: one for systolic BP representing the pressure in blood vessels when the heart contracts or beats, and the other for diastolic BP representing the pressure in the vessels when the heart rests between beats. Hypertension, i.e., elevated blood pressure, is a chronic and serious medical condition that significantly increases the risks of disease of heart, brain, kidney, and other organs. Since hypertension co-occurring with COVID-19 is reported as a risk factor for severe clinical outcomes (Hessami et al., 2021), understanding the association between these two conditions and their mechanistic underpinnings remains a priority. Due to confounding factors, observational studies with a limited sample size may not produce robust results. Thus, we explored the relationship between blood pressure and two types of COVID-19 outcomes, i.e., mortality and severe course of the disease, in a meta-analysis of the existing literature. Despite apparent heterogeneity of the studies included, the random-effects model of the meta-analysis suggested that both for mortality and for severity of COVID-19, their associations with hypertension were significant and substantial, with RR of 1.80 and 1.78, respectively. In mortality meta-analysis, age and gender explained the major share of heterogeneity, while in the meta-analysis of severe COVID-19, the heterogeneity was majorly due to other confounding factors rather than age and gender, thus

suggesting that some other parameters should be recorded for studies aiming at exploring relationships of elevated BP with outcomes of COVID-19. In the subgroup meta-analysis of the European populations using the random-effects model, relationships between the severe course of COVID-19 and SBP were significant, with a pooled weighted mean difference of 6.61 mmHg between COVID-19 cohorts with the severe and mild course of illness and no heterogeneity. These findings were in contrast to the analysis of the entire dataset, which detected a significant pooled weighted mean difference of 3.30 mmHg between COVID-19 cohorts with the severe and mild course of illness, with apparent heterogeneity. These observations suggest that elevated BP is significantly associated with the COVID-19 severity, in presence of complex, yet-to-be identified confounding factors.

In further analysis of the gene sets shared between hypertension and severe outcomes of COVID-19, the mainly affected tissue compartment was in the lungs. Thus, the relevant genes were identified in the lungs by integrating GWCTM results with lung eQTL. In the pathway enrichment analysis, the only pathway directly related to the regulation of BP was one involved in cardiac hypertrophy, with its fifth place among top shared pathways connecting hypertension with COVID-19 severity. Instead, predominant involvement of the signaling related to immune function was highlighted, with an emphasis on CCR1 and CCR5, the members of the beta chemokine receptor family, as well as IL10RB, an accessory chain essential for the active interleukin 10 (IL10) receptor complex.

Notably, both CCR1 and CCR5 serve as receptors for the same set of cytokine/chemokine ligands, including macrophage inflammatory protein 1 alpha (MIP-1 alpha), monocyte chemoattractant protein 3 (MCP-3), myeloid progenitor inhibitory factor-1 (MPIF-1), and regulated on activation normal T expressed and secreted protein (RANTES). The latter is well-known as a biomarker of vascular dysfunction in the pulmonary interface and a major driver of hypertension (Nosalski and Guzik, 2017; Funk-Hilsdorf et al., 2022). Its receptor CCR5 is expressed at the surface of T cells and macrophages and serves as a co-receptor for the cell entrance for macrophage-tropic viruses. One single-cell RNA sequencing study of immune-epithelial interactions within the lung tissue (Chua et al., 2020) indicated that during the infection of SARS-CoV-2, the activated resident macrophages in general and inducible ligands for CCR1 and CCR5 in particular, contribute to inflammatory tissue damage, lung injury, and respiratory failure. In addition to CCR1 and CCR5, our study highlighted involvement of IL10RB, whose co-expression with IL10RA is required for IL10-induced signal transduction. IL10 has pleiotropic effects on immunoregulation and inflammation of mucosal tissues. Notably, mucosal integrity and immunity are indispensable for the prevention of symptomatic COVID-19 (Arrieta et al., 2009; Cortes et al., 2022). In a recent genome-wide study, IL10RB was identified as the top key regulator of COVID-19 host susceptibility, with higher IL10RB expression in patient blood being associated with worse COVID-19 outcomes (Voloudakis et al., 2021). Furthermore, in multiple rodent models, the recombinant IL10 can exert direct antihypertensive action by increasing the production of nitric oxide (NO), a kind of potent vasodilator (Lima et al., 2016; Gillis et al., 2020). Incidentally, the same sets of molecular networks have been reported to be involved in the pathogenesis of multiple sclerosis (Szczucinski and Losy, 2007; Porro et al., 2020), one of the diseases highlighted by our ingenuity-driven analysis of canonical pathways.

The strengths of this study include its multi-pronged design, which included two different definitions of severe COVID-19 outcomes as well as three types of hypertension-related parameters. We should also stress that all or a majority of study participants were of European ancestry, thus reducing the potential population heterogeneity. Several limitations should be acknowledged, one is the heterogeneity of our findings. Observed levels of heterogeneity indicate inherent complexity of relationships between hypertension and severe COVID-19. In part, heterogeneity may be explained by differing age and gender structure of studied populations. In addition, severity of COVID-19 may be affected by a large number of other confounders such as BMI, smoking, history of medications, and presence of comorbidities (Cai and He, 2019; Cai and He, 2021). We have also detected a moderate publication bias, possibly explained by a lack of interest in publishing largely negative association trends. A further meta-analysis of larger sets of independent publications is warranted.

In conclusion, our results suggest that elevated blood pressure is significantly associated with the COVID-19 severity. The genetic liabilities to elevated blood pressure and to severe COVID-19 intertwine, and among them, the immune-regulating receptors CCR1/CCR5 and IL10RB signaling pathway are highlighted in the common effective tissue of the lung.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

## Ethics statement

The studies involving human participants were reviewed and approved by the Bioethics Committee of Bio-X Institutes of Shanghai Jiaotong University. The Ethics Committee waived the requirement of written informed consent for participation.

## Author contributions

LC, LH and AB designed the study and prepared the manuscript, CH and YS collected the primary data for the meta-analysis, LC and YL performed the primary analyses,

LC, YL, LH and AB contributed to the interpretations of the findings and the critical revision of the manuscript.

## Funding

This study was supported by the grants from the Natural Science Foundation of Shanghai (No. 19ZR1427700), and cooperation foundation for Sanya Maternal and Child Health Hospital.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.933148/full#supplementary-material

## References

Arrieta, M. C., Madsen, K., Doyle, J., and Meddings, J. (2009). Reducing small intestinal permeability attenuates colitis in the IL10 gene-deficient mouse. *Gut* 58 (1), 41–48. doi:10.1136/gut.2008.150888

Cai, L., Chen, T., Yang, J., Zhou, K., Yan, X., Chen, W., et al. (2015). Serum trace element differences between Schizophrenia patients and controls in the han chinese population. *Sci. Rep.* 5, 15013. doi:10.1038/srep15013

Cai, L., Deng, S. L., Liang, L., Pan, H., Zhou, J., Wang, M. Y., et al. (2013). Identification of genetic associations of SP110/MYBBP1A/RELA with pulmonary tuberculosis in the chinese han population. *Hum. Genet.* 132 (3), 265–273. doi:10.1007/s00439-012-1244-5

Cai, L., and He, L. (2019). Placebo effects and the molecular biological components involved. *Gen. Psychiatr.* 32 (5), e100089. doi:10.1136/gpsych-2019-100089

Cai, L., and He, L. (2021). The novel coronavirus and humans: who can dominate who? *J. BioX. Res.* 4 (2), 45. doi:10.1097/JBR.0000000000000098

Cai, L., Huang, T., Su, J., Zhang, X., Chen, W., Zhang, F., et al. (2018). Implications of newly identified brain eQTL genes and their interactors in schizophrenia. *Mol. Ther. Nucleic Acids* 12, 433–442. doi:10.1016/j.omtn.2018.05.026

Chen, G. B., Lee, S. H., Zhu, Z. X., Benyamin, B., and Robinson, M. R. (2016). EigenGWAS: finding loci under selection through genome-wide association studies of eigenvectors in structured populations. *Hered. (Edinb)* 117 (1), 51–61. doi:10.1038/hdy.2016.25

Chua, R. L., Lukassen, S., Trump, S., Hennig, B. P., Wendisch, D., Pott, F., et al. (2020). COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* 38 (8), 970–979. doi:10.1038/s41587-020-0602-4

Cortes, G. M., Marcialis, M. A., Bardanzellu, F., Corrias, A., Fanos, V., and Mussap, M. (2022). Inflammatory bowel disease and COVID-19: how microbiomics and metabolomics depict two sides of the same coin. *Front. Microbiol.* 13, 856165. doi:10.3389/fmicb.2022.856165

Cumpston, M., Li, T., Page, M. J., Chandler, J., Welch, V. A., Higgins, J. P., et al. (2019). Updated guidance for trusted systematic reviews: a new edition of the cochrane handbook for systematic reviews of interventions. *Cochrane Database Syst. Rev.* 10, ED000142. doi:10.1002/14651858.ED000142

Elsworth, B. (2020). The MRC IEU OpenGWAS data infrastructure. *BioRxiv*. doi:10.1002/14651858.ED000142

Funk-Hilsdorf, T. C., Behrens, F., Grune, J., and Simmons, S. (2022). Dysregulated immunity in pulmonary hypertension: From companion to composer. *Front. Physiol.* 13, 819145. doi:10.3389/fphys.2022.819145

Gillis, E. E., Musall, J. B., Baban, B., and Sullivan, J. C. (2020). IL-10 treatment decreases blood pressure in male, but not female, spontaneously hypertensive rats. *Am. J. Physiol. Ren. Physiol.* 319 (3), F359-F365. doi:10.1152/ajprenal.00206.2020

Hessami, A., Shamshirian, A., Heydari, K., Pourali, F., Alizadeh-Navaei, R., Moosazadeh, M., et al. (2021). Cardiovascular diseases burden in COVID-19: systematic review and meta-analysis. *Am. J. Emerg. Med.* 46, 382–391. doi:10.1016/j.ajem.2020.10.022

Jb, W. P. R. A. R. (2003). Bossuyt PM kleijnen J the development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med. Res. Methodol.* 3, 25.

Kumaran, M., Pham, T. M., Wang, K., Usman, H., Norris, C. M., MacDonald, J., et al. (2022). Predicting the risk factors associated with severe outcomes among COVID-19 patients-decision tree modeling approach. *Front. Public Health* 10, 838514. doi:10.3389/fpubh.2022.838514

Lei, C., Dongqing, Z., Yeqing, S., Oaks, M. K., Lishan, C., Jianzhong, J., et al. (2005). Association of the CTLA-4 gene with rheumatoid arthritis in chinese han population. *Eur. J. Hum. Genet.* 13 (7), 823–828. doi:10.1038/sj.ejhg.5201423

Li, X., and Zhu, X. (2017). Cross-phenotype association analysis using summary statistics from GWAS. *Methods Mol. Biol.* 1666, 455–467. doi:10.1007/978-1-4939-7274-6_22

Lima, V. V., Zemse, S. M., Chiao, C. W., Bomfim, G. F., Tostes, R. C., Clinton Webb, R., et al. (2016). Interleukin-10 limits increased blood pressure and vascular RhoA/Rho-kinase signaling in angiotensin II-infused mice. *Life Sci.* 145, 137–143. doi:10.1016/j.lfs.2015.12.009

Manohar, J., Abedian, S., Martini, R., Kulm, S., Salvatore, M., Ho, K., et al. (2021). *Social and clinical determinants of COVID-19 outcomes: modeling real-world data from a pandemic epicenter*. medRxiv. doi:10.1101/2021.04.06.21254728

Madjid, M., Safavi-Naeini, P., Solomon, S. D., and Vardeny, O. (2020). Potential effects of coronaviruses on the cardiovascular system: a review. *JAMA Cardiol.* 5 (7), 831–840. doi:10.1001/jamacardio.2020.1286

Nosalski, R., and Guzik, T. J. (2017). Perivascular adipose tissue inflammation in vascular disease. *Br. J. Pharmacol.* 174 (20), 3496–3513. doi:10.1111/bph.13705

Organization, W. H. (2019). Hypertension. Fact sheets. WHO, 16. Dostęp. Dostępny w internecie Available at: https://www.who.int/news-room/fact–sheets/detail/hypertension.

Park, H., Li, X., Song, Y. E., He, K. Y., and Zhu, X. (2016). Multivariate analysis of anthropometric traits using summary statistics of genome-wide association studies from GIANT consortium. *PLoS One* 11 (10), e0163912. doi:10.1371/journal.pone.0163912

Porro, C., Cianciulli, A., and Panaro, M. A. (2020). The regulatory role of IL-10 in neurodegenerative diseases. *Biomolecules* 10 (7), E1017. doi:10.3390/biom10071017

Rostom, A. (2004). *Celiac disease. Appendix D. quality assessment forms*. Rockville (MD): Agency for Healthcare Research and Quality (US).

Schiffrin, E. L., *Hypertension and COVID-19. 2020*. Oxford University Press US, 373–374.

Stanetic, K., Stanetic, B., Petrovic, V., Markovic, B., Kevic, V., Todorovic, N., et al. (2021). The influence of different risk factors on COVID-19 outcomes in adult patients - an observational-descriptive study. *Acta Med. Acad.* 50 (2), 308–316. doi:10.5644/ama2006-124.348

Szczucinski, A., and Losy, J. (2007). Chemokines and chemokine receptors in multiple sclerosis. potential targets for new therapies. *Acta Neurol. Scand.* 115 (3), 137–146. doi:10.1111/j.1600-0404.2006.00749.x

Varshney, R., Bansal, N., Khanduri, A., Gupta, J., and Gupta, R. (2021). Colonic gangrene: a sequela of coronavirus disease 2019. *Cureus* 13 (4), e14687. doi:10.7759/cureus.14687

Verano-Braga, T., Martins, A. L. V., Motta-Santos, D., Campagnole-Santos, M. J., and Santos, R. A. S. (2020). ACE2 in the renin-angiotensin system. *Clin. Sci.* 134 (23), 3063–3078. doi:10.1042/CS20200478

Voloudakis, G., Hoffmann, G., Venkatesh, S., Lee, K. M., Dobrindt, K., Vicari, J. M., et al. (2021). *IL10RB as a key regulator of COVID-19 host susceptibility and severity*. medRxiv. doi:10.1101/2021.05.31.21254851

Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* 8 (1), 1826. doi:10.1038/s41467-017-01261-5

Wu, C., Qu, G., Wang, L., Cao, S., Xia, D., Wang, B., et al. (2021). Clinical characteristics and inflammatory immune responses in COVID-19 patients with hypertension: a retrospective study. *Front. Pharmacol.* 12, 721769. doi:10.3389/fphar.2021.721769

Zheng, Z., Peng, F., Xu, B., Zhao, J., Liu, H., Peng, J., et al. (2020). Risk factors of critical & mortal COVID-19 cases: a systematic literature review and meta-analysis. *J. Infect.* 81 (2), e16–e25. doi:10.1016/j.jinf.2020.04.021

Zhu, X., Feng, T., Tayo, B. O., Liang, J., Young, J. H., Franceschini, N., et al. (2015). Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am. J. Hum. Genet.* 96 (1), 21–36. doi:10.1016/j.ajhg.2014.11.011

frontiers | Frontiers in Genetics

TYPE Brief Research Report
PUBLISHED 24 October 2022
DOI 10.3389/fgene.2022.1014191



OPEN ACCESS

EDITED BY
Ramcés Falfán-Valencia,
Instituto Nacional de Enfermedades Respiratorias-México (INER), Mexico

REVIEWED BY
Ingrid Fricke-Galindo,
Laboratorio de HLA, Instituto Nacional de Enfermedades Respiratorias-México (INER), Mexico
Antonino Zito,
Harvard Medical School, United States
Craig Kinnear,
Stellenbosch University, South Africa

*CORRESPONDENCE
Ke Zhang,
kezhang@gmc.edu.cn
Zhong-Shan Cheng,
cheng.zhong.shan@gmail.com

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 08 August 2022
ACCEPTED 29 September 2022
PUBLISHED 24 October 2022

CITATION
Luo Y-S, Li W, Cai Y, Zhang J, Gui H, Zhang K and Cheng Z-S (2022), Genome-wide screening of sex-biased genetic variants potentially associated with COVID-19 hospitalization. *Front. Genet.* 13:1014191. doi: 10.3389/fgene.2022.1014191



# Genome-wide screening of sex-biased genetic variants potentially associated with COVID-19 hospitalization

Yu-Si Luo[1,2], Wei Li[3], Yi Cai[4], Jingxuan Zhang[5], Hongsheng Gui[6], Ke Zhang[2]* and Zhong-Shan Cheng[7]*

[1]Department of Emergency, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, [2]The Key and Characteristic Laboratory of Modern Pathogenicity Biology, School of Basic Medical Sciences, Guizhou Medical University, Guiyang, China, [3]Department of Cardiovascular, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, [4]Guangdong Key Laboratory of Regional Immunity and Diseases, Department of Pathogen Biology, Shenzhen University School of Medicine, Shenzhen, China, [5]Hubei Key Laboratory of Embryonic Stem Cell Research, School of Basic Medical Sciences, Hubei University of Medicine, Shiyan, China, [6]Center for Health Policy and Health Services Research, Behavioral Health Services and Department of Psychiatry, Henry Ford Health System, Detroit, MI, United States, [7]Center for Applied Bioinformatics, St. Jude Children's Research Hospital, 262 Danny Thomas Hospital, Memphis, TN, United States

Sex-biased difference in coronavirus disease 2019 (COVID-19) hospitalization has been observed as that male patients tend to be more likely to be hospitalized than female patients. However, due to the insufficient sample size and existed studies that more prioritized to sex-stratified COVID-19 genome-wide association study (GWAS), the searching for sex-biased genetic variants showing differential association signals between sexes with COVID-19 hospitalization was severely hindered. We hypothesized genetic variants would show potentially sex-biased genetic effects on COVID-19 hospitalization if they display significant differential association effect sizes between male and female COVID-19 patients. By integrating two COVID-19 GWASs, including hospitalized COVID-19 patients vs. general population separated into males (case = 1,917 and control = 221,174) and females (case = 1,343 and control = 262,886), we differentiated the association effect sizes of each common single nucleotide polymorphism (SNP) within the two GWASs. Twelve SNPs were suggested to show differential COVID-19 associations between sexes. Further investigation of genes (n = 58) close to these 12 SNPs resulted in the identification of 34 genes demonstrating sex-biased differential expression in at least one GTEx tissue. Finally, 5 SNPs are mapped to 8 genes, including rs1134004 (*GADD45G*), rs140657166 (*TRIM29* and *PVRL1*), rs148143613 (*KNDC1* and *STK32C*), rs2443615 (*PGAP2* and *TRIM21*), and rs2924725 (*CSMD1*). The 8 genes display significantly differential gene expression in blood samples derived from COVID-19 patients compared to healthy controls. These genes are potential genetic factors contributing to sex differences in COVID-19 hospitalization and warranted for further functional studies.

KEYWORDS

**GWAS, single nucleotide polymorphisms, COVID-19 hospitalization, European population, sex-biased**

## Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first emerged in Wuhan, China, in October 2019, and then it was spread to almost all countries in the world. Until now, SARS-CoV-2 induced coronavirus disease 2019 (COVID-19) has led to ~6.32 million deaths worldwide according to COVID-19 Dashboard from Johns Hopkins University (Dong et al., 2020). Genetic factors have been suggested to contribute to different variation of COVID-19 severity, and two large-scale genome-wide association studies (GWASs) related to the severe COVID-19 phenotype (The Severe Covid-19 GWAS Group, 2020; Pairo-Castineira et al., 2021) have uncovered several loci in human genome predisposing to COVID-19 severity. Significant single nucleotide polymorphisms (SNPs) have been reported to associate with the risk of severe COVID-19. These SNPs are rs11385942 located in the 3p21.31 region harboring the genes *SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6*, and *XCR1*, rs10735079 mapped to a gene cluster comprising *OAS1*, *OAS2*, and *OAS3*, and three other SNPs, including rs74956615, rs2109069, and rs2236757, mapped to *TYK2*, *DPP9*, and *IFNAR2*, respectively (Pairo-Castineira et al., 2021). In addition, increasing studies have been reported that COVID-19 leads to more severe symptoms and higher mortality in men than in women (Alwani et al., 2021). A large-scale cohort comprising 17 million European adults discovered that male sex was more predisposing to COVID-19 induced death (hazard ratio 1.59) (Williamson et al., 2020). A recent study claimed that male tends to have weaker immune responses, particularly T cell response against SARS-CoV-2 infection compared to female. However, the study was based on a dataset with relatively small sample size ($n < 100$) and did not adjust for confounding factors in the analyses, the conclusion of which was criticized by researchers (Takahashi et al., 2020; Shattuck-Heidorn et al., 2021; Takahashi et al., 2021; Danielsen et al., 2022). Therefore, whether or how sex differences contribute to COVID-19 severity/hospitalization is still not clear. Furthermore, previously published COVID-19 GWASs mainly focused on both sexes altogether or considered males and females separately (Tu et al., 2020; Viveiros et al., 2021). Recently, several COVID-19 risk SNPs have been reported by sex-stratified GWASs, such as rs17763742, rs4443214, and rs35477280, which are three risk SNPs located in the famous 3p21.31 region and are predominantly associated with COVID-19 in males (Cruz et al., 2022). However, due to the requirement of large sample sizes for both sexes, the determination of the interaction between SNPs and sexes with COVID-19 severity/hospitalization was hindered. As few genome-wide significant sex-biased COVID-19 SNPs have been reported, there is an urgent need to search for sex-biased genetic factors associated with COVID-19, given that substantial sex differences in inflammation, immunity, and aberrant renin-angiotensin system (RAS) activity have been observed in the pathogenesis of COVID-19. Many immune and inflammation association genes are X-linked, including but not limited to *TLR7*, *TLR8*, and *IRAK1* (Viveiros et al., 2021). Females have an increased IFN-α secretion, early virus sensing, and prompt antiviral response upon *TLR7* stimulation in dendritic cells. In addition, the *ACE2* gene resides in the Xp22.2 region of the X chromosome and is recognized as an escape gene. Females theoretically have a double dose of *ACE2*, which may compensate for SARS-CoV-2 mediated loss of membrane ACE2 and alleviate aberrant RAS activity and related cardiovascular diseases (Tu et al., 2020; Viveiros et al., 2021). To search for novel sex-biased loci, i.e., loci displaying significantly differential associations with COVID-19 hospitalization between sexes, we performed an integrative study of sex-biased genetic factors on COVID-19 by comparing male- and female-stratified COVID-19 hospitalization GWAS summary statistics from UK Biobank. Here the sex-biased variants represent variants showing differential association signals based on the Z-test between the effect sizes of each SNP from sex-stratified COVID-19 hospitalization GWASs. Genes close to these SNPs were evaluated for its potential sex-biased differential expression as well as its involvements in COVID-19.

## Materials and methods

Summary statistics of two COVID-19 hospitalization GWASs by males and females were obtained from GRASP (Thibord et al., 2022) (https://grasp.nhlbi.nih.gov/Covid19GWASResults.aspx) that were generated based on UK Biobank data. The two COVID-19 hospitalization GWASs were conducted on samples with European descent and designed by comparing hospitalized COVID-19 cases with not hospitalized COVID-19 patients for males and females individually (1,917 hospitalized COVID-19 males vs. 221,174 not hospitalized COVID-19 males; 1,343 hospitalized COVID-19 females vs. 262,886 not hospitalized COVID-19 females). We used the statistical method introduced by Thibord et al. (Thibord et al., 2022) to compare the difference of effect size beta for each common SNP (MAF >0.01, imputation score >0.6) between the two GWASs by calculating its delta Z-score (ΔZ-score) and its corresponding $p$ value, the formulas of which are as follows:

$$\Delta Z - score = \frac{female.gwas.\beta - male.gwas.\beta}{\sqrt{\left(female.gwas.se^2 + male.gwas.se^2\right)}}$$

$$p = pnorm\left(-|\Delta Z - score|\right)*2$$

Common SNPs were kept with MAF >0.01 and imputation score >0.6 in both male- and female-stratified COVID-19

GWASs. The genome-wide significance threshold of ΔZ-score $p$ values was set as $p < 5.0 \times 10^{-8}$; a relaxed $p$-value threshold for ΔZ-score $p$ values, $p < 5 \times 10^{-6}$, was applied to select SNPs with suggestive evidence of sex-biased, differential COVID-19 associations. Only top SNPs passing the relaxed ΔZ-score $p$-value threshold and representing independent, differential COVID-19 association signals were selected for further evaluation.

Genes adjacent to the selected top SNPs were subject to sex-biased differential expression analysis in GTEx tissues (The GTEx Consortium, 2020) and differential gene expression (DGE) analysis between COVID-19 patients and healthy controls. We located genes within the 1-Mbp genomic window with one of the top SNP centered. To gain biological insights about these top SNPs and genes close to them, we determined whether these top SNPs are also cis-expression quantitative trait loci (cis-eQTL) in GTEx database (Oliva et al., 2020) and eQTLGen database (Võsa et al., 2018). We further evaluated whether these SNPs adjacent genes showing DGE between the sexes across GTEx tissues (Oliva et al., 2020). In detail, we downloaded summary statistics of DGE between sexes from GTEx portal (https://storage.googleapis.com/gtex_analysis_v8/sex_biased_genes_and_sbeqtl_data/GTEx_Analysis_v8_sbgenes.tar.gz) and checked the sex-biased DGE $p$ values and its corresponding effect sizes for these genes located within a 1-Mbp window of the top SNPs showing sex-biased COVID-19 associations. To determine the fold change of median gene expression between sexes for these genes in GTEx tissue, we downloaded GTEx gene expression Transcripts Per Million (TPM) matrix and sex information for all samples from these links (TPM matrix: https://storage.googleapis.com/gtex_analysis_v8/rna_seq_data/GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_tpm.gct.gz; tissue and sample information: https://storage.googleapis.com/gtex_analysis_v8/annotations/GTEx_Analysis_v8_Annotations_SubjectPhenotypesDS.txt; sample sex, age, and other clinical information: https://storage.googleapis.com/gtex_analysis_v8/annotations/GTEx_Analysis_v8_Annotations_SampleAttributesDS.txt). After extraction of TPMs for these SNP adjacent genes, we determined the fold change of median gene expression using the formula log2 (female/male). In addition, we queried these genes in COVID-19 expression database COVID19db, a gene expression database related to SARS-CoV-2 infection (Zhang et al., 2022), and determined whether these genes were differentially expressed in blood samples of COVID-19 patients compared to healthy controls (Thair et al., 2021) with one-way ANOVA.

## Results

To search for SNPs displaying sex-biased association signals in hospitalized COVID-19 patients, we compared the association effect size beta of each SNP between two COVID-19 hospitalization GWASs stratified by sexes among samples with European ancestry, which were publicly available from GRASP COVID-19 database (Thibord et al., 2022). The two GWASs were conducted by comparing hospitalized COVID-19 with general population for males and females individually, adjusted by age and 10 principal components, with all samples from European ancestry. When considering the two GWASs individually, only one independent genome-wide significant SNP, rs13071258, mapped to the gene cluster of *SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6*, and *XCR1*, with $p = 9.97 \times 10^{-10}$ and beta = 0.43, emerged from the male GWAS, while in the corresponding female GWAS, the SNP displayed moderate association signal with COVID-19 hospitalization ($p = 4.12 \times 10^{-5}$; beta = 0.34). Because the SNP demonstrated similar effect sizes, there was no significant difference in terms of ΔZ-score (value = −0.84; ΔZ-score $p = 0.40$). We further compared SNP association effect size beta of the two GWASs and determined the ΔZ-scores genome-wide, resulting in 12 independent, differential association signals (all ΔZ-score $p$ values $<5 \times 10^{-6}$) between the female and male COVID-19 GWASs, including rs8116534, rs140657166, rs148143613, rs2443615, rs35239301, rs1134004, rs200808810, rs1965385, rs2924725, rs5927942, rs555336963, and rs472481 (Figure 1 and Table 1). The estimated skewness for the ΔZ-scores genome-wide is 0.016, suggesting the close normal distribution of ΔZ-scores. This is in line with the GWAS inflation lambda, which is 0.996, indicating weak inflation of the sex-biased GWAS based on ΔZ-score $p$ values. Taken together, only one independent SNP emerged as genome-wide significant when considering the two sex-stratified GWASs individually, with the SNP does not show significant differential effect size between sexes; further genome-wide screening of sex-biased COVID-19 risk SNPs harvested 12 candidate SNPs.

We further evaluated previously published COVID-19 risk SNPs showing genome-wide significant association signals in either females/males or both sexes published by Cruz et al. (Cruz et al, 2022) in the currently used UK Biobank COVID-19 hospitalization GWASs. We obtained summary statistics of nine genome-wide significant SNPs emerged in sex-unstratified or sex-stratified COVID-19 GWASs from Cruz and colleagues (Cruz et al., 2022). After performing Z-test to determine differential association signals between females and males for these SNPs, we only found three SNPs passed the ΔZ-score $p$ value threshold of $p < 0.05$ (Supplementary Figure S1) in the data set from Cruz et al. (Cruz et al., 2022). In the currently used UK Biobank COVID-19 GWASs of females and males, none of these nine SNPs demonstrated differential association between sexes with COVID-19 (all ΔZ-score $p > 0.1$; Supplementary Figure S1). However, six out of these nine SNPs were nominally significant in the current UK Biobank sex-stratified COVID-19 GWASs, including these three SNPs (rs17763742, rs115679256, and rs35477280) located in the 3p21.31 region. We further conducted (LD) analysis between these three SNPs and the genome-wide significant SNP rs13071258 from the male only COVID-19 hospitalization GWAS from UK Biobank. We

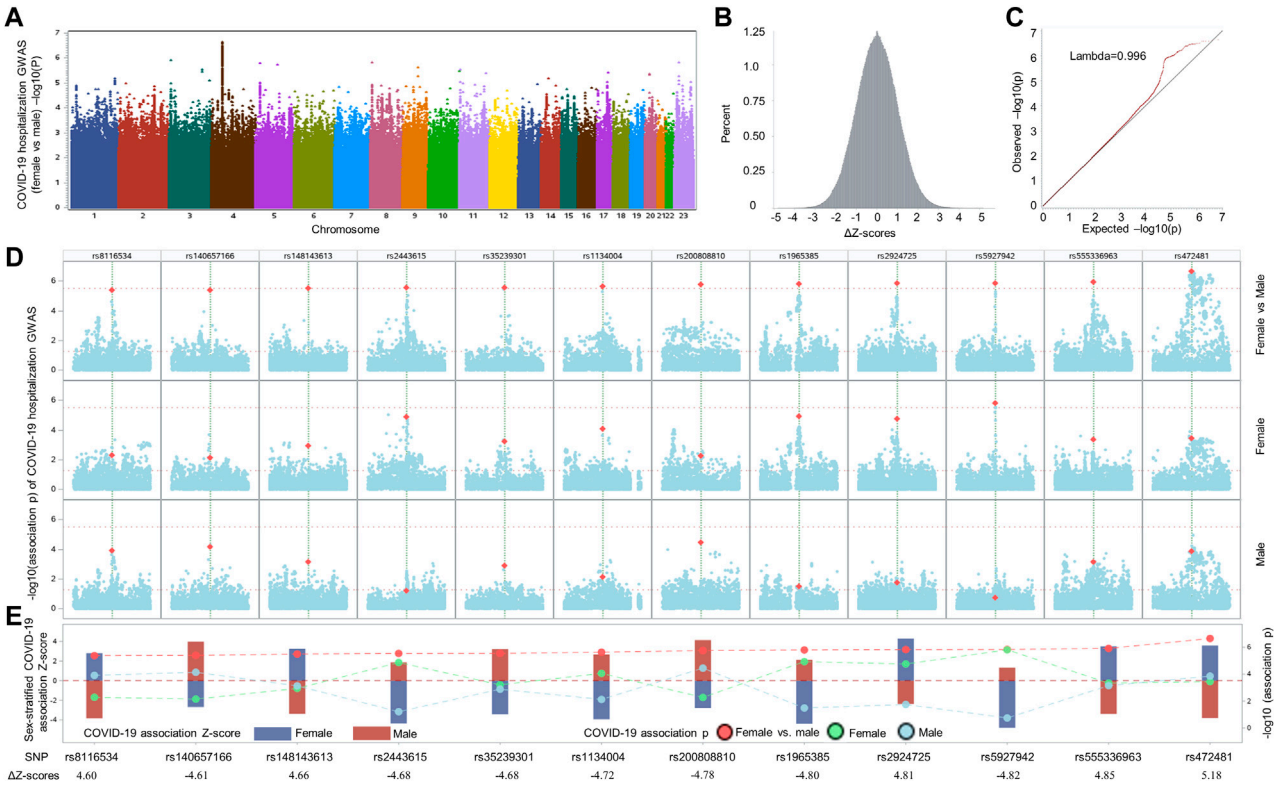



FIGURE 1
Comparative analysis of sex-stratified genome-wide association studies of hospitalized COVID-19 cases vs. general population. **(A)** Manhattan plot illustrating the comparison of effect size beta of each single nucleotide polymorphism (SNP) between the two COVID-19 hospitalization GWASs with European ancestry from UK Biobank. **(B)** Delta Z-score (ΔZ-score) histogram (see method section for the calculation of ΔZ-score). **(C)** QQ plot of ΔZ-score *p* values. **(D)** Local Manhattan plot covering a 1-Mbp region harboring each of 12 top SNPs (highlighted in dark yellow color) that were passed the nominal ΔZ-score *p* value threshold of $p < 5 \times 10^{-6}$ (upper panel), followed by another two local Manhattan plots demonstrating sex-stratified COVID-19 association signals. **(E)** Needle plot at the bottom illustrating the sex-stratified and sex-biased COVID-19 association signals. The sex-stratified SNP Z-scores are displayed according to the left *y*-axis for both sexes (female: dark blue bar; male: dark red bar). The SNP *p* values of sex-stratified association signals for female, represented by light green dots, and male, indicated by light blue dots, as well as the corresponding *p* values of ΔZ-scores, highlighted by red dots, are displayed in accordance to the right *y*-axis. To separate these *p* values and visualize the differences of Z-scores and association *p* values in both sexes, dashed lines are utilized to link and separate these dots representing association *p* values derived from sex-stratified analysis and ΔZ-score *p* values generated by differential Z-test between sexes. The ΔZ-score for each SNP is listed at bottom of panel E.

revealed that rs17763742 and rs13071258 are in complete LD ($R^2$ = 1; D' = 1) in European population; the other two SNPs are in weak LD to rs13071258 in European population, with $R^2$ = 0.08 for rs13071258 vs. rs115679256 and $R^2$ = 0.43 for rs13071258 vs. rs17763742. We confirmed that most of these published COVID-19 SNPs by Cruz et al. (Cruz et al., 2022) are replicable in the current sex-stratified COVID-19 GWASs in either males or females. The lack of replication of sex-biased differential associations for these SNPs in current study may attribute to the relatively small sample sizes of current UK Biobank GWASs. In conclusion, in the current sex-biased COVID-19 GWAS with relatively small sample sizes, previously published genome-wide significant SNPs do not show significant difference of effect size between sexes, which is in line with the requirement of GWAS with large sample size to determine sex-biased COVID-19 risk SNPs.

To investigate the potential regulatory roles played by these top 12 SNPs emerged from sex-biased COVID-19 GWAS, we queried them in GTEx Portal (The GTEx Consortium, 2020) and eQTLGen (Võsa et al., 2018). We identified 4 cis-eQTLs; among them, rs140657166 on chromosome 6, is a cis-eQTL of *THY1* in cells-cultured fibroblasts, which is also close to *PVRL1* (alias *NECTIN1*) and *TRIM29*. In addition, another variant rs1134004, on chromosome 9, is strongly associated with *SEMA4D* expression among multiple GTEx tissues, including cells-cultured fibroblasts, artery-tibial, testis, breast-mammary tissue, and other tissues. eQTLGen also reported the SNP as eQTL for *SEMA4D* in blood samples. Additionally, rs8116534 is an eQTL of the non-coding gene *LINC00652* on chromosome 20 in muscle skeletal and ovary, and three brain tissues, including amygdala, hypothalamus, and cerebellum. One indel, rs555336963, which is a strong cis-eQTL of the RNA gene *AC069277* on chromosome 3 in

TABLE 1 Top variants displaying sex-biased associations with COVID-19 hospitalization in UK Biobank.

| Chr* | Position* | A1/A2^ | | rsid | ΔZ-score *p* value | ΔZ-score | GWAS *p* | | Z-score | | Beta | | Adjacent sex-biased DGE# |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Female | Male | Female | Male | Female | Male | |
| 4 | 43,498,824 | C | T | rs472481 | $2.2 \times 10^{-7}$ | 5.18 | $3.6 \times 10^{-4}$ | $1.4 \times 10^{-4}$ | 3.57 | −3.81 | 0.27 | −0.24 | |
| 3 | 6,540,177 | T | TT | rs555336963 | $1.2 \times 10^{-6}$ | 4.85 | $4.7 \times 10^{-4}$ | $7.2 \times 10^{-4}$ | 3.50 | −3.38 | 0.21 | −0.17 | *GRM7* |
| 23 | 32,200,753 | T | C | rs5927942 | $1.4 \times 10^{-6}$ | −4.82 | $1.5 \times 10^{-6}$ | 0.18 | −4.80 | 1.33 | −0.21 | 0.03 | |
| 8 | 4,841,239 | C | T | rs2924725 | $1.5 \times 10^{-6}$ | 4.81 | $1.8 \times 10^{-5}$ | $1.8 \times 10^{-2}$ | 4.29 | −2.36 | 0.20 | −0.09 | *CSMD1* |
| 5 | 17,633,867 | T | C | rs1965385 | $1.6 \times 10^{-6}$ | −4.80 | $1.2 \times 10^{-5}$ | $3.3 \times 10^{-2}$ | −4.38 | 2.13 | −1.49 | 0.52 | *BASP1* |
| 5 | 105,404,915 | G | A | rs200808810 | $1.7 \times 10^{-6}$ | −4.78 | $5.4 \times 10^{-3}$ | $3.6 \times 10^{-5}$ | −2.78 | 4.13 | −0.15 | 0.18 | |
| 9 | 91,981,877 | C | T | rs1134004 | $2.3 \times 10^{-6}$ | −4.72 | $8.8 \times 10^{-5}$ | $7.6 \times 10^{-3}$ | −3.92 | 2.67 | −0.19 | 0.11 | *S1PR3*, *GADD45G*, *SHC3*, *SECISBP2*, *SEMA4D*, *C9orf47* |
| 3 | 159,373,186 | G | T | rs35239301 | $2.8 \times 10^{-6}$ | −4.68 | $6.2 \times 10^{-4}$ | $1.3 \times 10^{-3}$ | −3.42 | 3.21 | −0.25 | 0.19 | *SCHIP1* |
| 11 | 4,249,031 | C | T | rs2443615 | $2.8 \times 10^{-6}$ | −4.68 | $1.4 \times 10^{-5}$ | $6.3 \times 10^{-2}$ | −4.35 | 1.86 | −3.64 | 0.96 | *RRM1*, *STIM1*, *PGAP2*, *TRIM21* |
| 10 | 134,578,626 | T | C | rs148143613 | $3.2 \times 10^{-6}$ | 4.66 | $1.2 \times 10^{-3}$ | $7.4 \times 10^{-4}$ | 3.25 | −3.37 | 0.50 | −0.44 | *C10orf91*, *PWWP2B*, *KNDC1*, *LRRC27*, *STK32C*, *INPP5A* |
| 11 | 119,728,423 | A | C | rs140657166 | $4.0 \times 10^{-6}$ | −4.61 | $7.2 \times 10^{-3}$ | $7.3 \times 10^{-5}$ | −2.69 | 3.97 | −0.31 | 0.38 | *TRIM29*, *OAF*, *THY1*, *PVRL1*, *AP000679.2* |
| 20 | 18,068,955 | C | T | rs8116534 | $4.2 \times 10^{-6}$ | 4.60 | $5.3 \times 10^{-3}$ | $1.2 \times 10^{-4}$ | 2.79 | −3.84 | 0.16 | −0.19 | *RBBP9*, *DZANK1*, *ZNF133*, *CSRP2BP*, *RRBP1*, *MGME1*, *OVOL2*, *POLR3F*, *SEC23B* |

Notes: *: chromosomal position based on the human reference hg19 build.#: genes located within the 1-Mbp window where the SNP at the center were determined for sex-biased differential gene expression analysis by Olivia et al. (Oliva et al., 2020).
^ indicates the risk allele.

multiple tissues (thyroid, spleen, pituitary, small intestine-terminal ileum, cells-EBV-transformed lymphocytes, colon-transverse, and esophagus-mucosa). Furthermore, rs5927942 is the only SNP located on chromosome X, but it is not an eQTL for any adjacent genes according to GTEx database and eQTLGen database. For other top SNPs, including rs472481, rs2924725, rs2443615, and rs1965385, they are not cis-eQTLs in both GTEx database and eQTLGen database; rs148143613, rs35239301, and rs200808810 are not included in GTEx database with no eQTL records for them in eQTLGen database.

We further examined protein codes genes located within a 1-Mbp genomic region with each of these top 12 COVID-19 SNPs located at the center. Based on the DGE analysis performed between sexes across 43 GTEx tissues by Oliva et al. (Oliva et al., 2020), there were 58 genes identified close to these 12 SNPs, among which 34 genes adjacent to eight SNPs displaying sex-biased differential expression at least in one of 43 GTEx tissues (Figure 2). Details of these eight SNPs mapped to 34 sex-biased genes are included in Table 1. They are rs1134004 (*S1PR3*, *GADD45G*, *SHC3*, *SECISBP2*, *SEMA4D*, and *C9orf47*), rs140657166 (*TRIM29*, *OAF*, *THY1*, *PVRL1*, and *AP000679.2*), rs148143613 (*C10orf91*, *PWWP2B*, *KNDC1*, *LRRC27*, *STK32C*, and *INPP5A*), rs1965385 (*BASP1*), rs2443615 (*RRM1*, *STIM1*, *PGAP2*, and *TRIM21*), rs2924725 (*CSMD1*), rs35239301 (*SCHIP1*), rs555336963 (*GRM7*), and rs8116534 (*RBBP9*, *DZANK1*, *ZNF133*, *CSRP2BP*, *RRBP1*, *MGME1*, *OVOL2*, *POLR3F*, and *SEC23B*). Among these tissues displayed significantly sex-biased DGE, there are nine genes (*SEMA4D*, *STK32C*, *INPP5A*, *BASP1*, *RRM1*, *PGAP2*, *CSMD1*, *SCHIP1*, and *CSRP2BP*) showing constantly, slightly higher expression [log2 (fold change) between 0 and 1] in females, while other six genes, including *S1PR3*, *OAF*, *TRIM29*, *PVRL1*, *KNDC1*, and *STIM1*, display opposite effect sizes across different tissues between sexes. For example, both the expression of *PVRL1* and *TRIM29* are higher in females than in males among artery-tibial, artery-coronary and artery-aorta, with the expression of *TRIM29* is higher in skin not sun exposed suprapubic of females and the expression of *PVRL1* is higher in pituitary of females; additionally, lower expression of *PVRL1* is observed in female tissues of whole blood, brain-caudate basal ganglia, brain-hippocampus, brain-putamen basal ganglia, brain-nucleus accumbens basal ganglia, and muscle skeletal; in contrary, higher expression of *TRIM29* is observed in male adipose subcutaneous and muscle skeletal compared to that of female. Other 15 sex-biased genes display

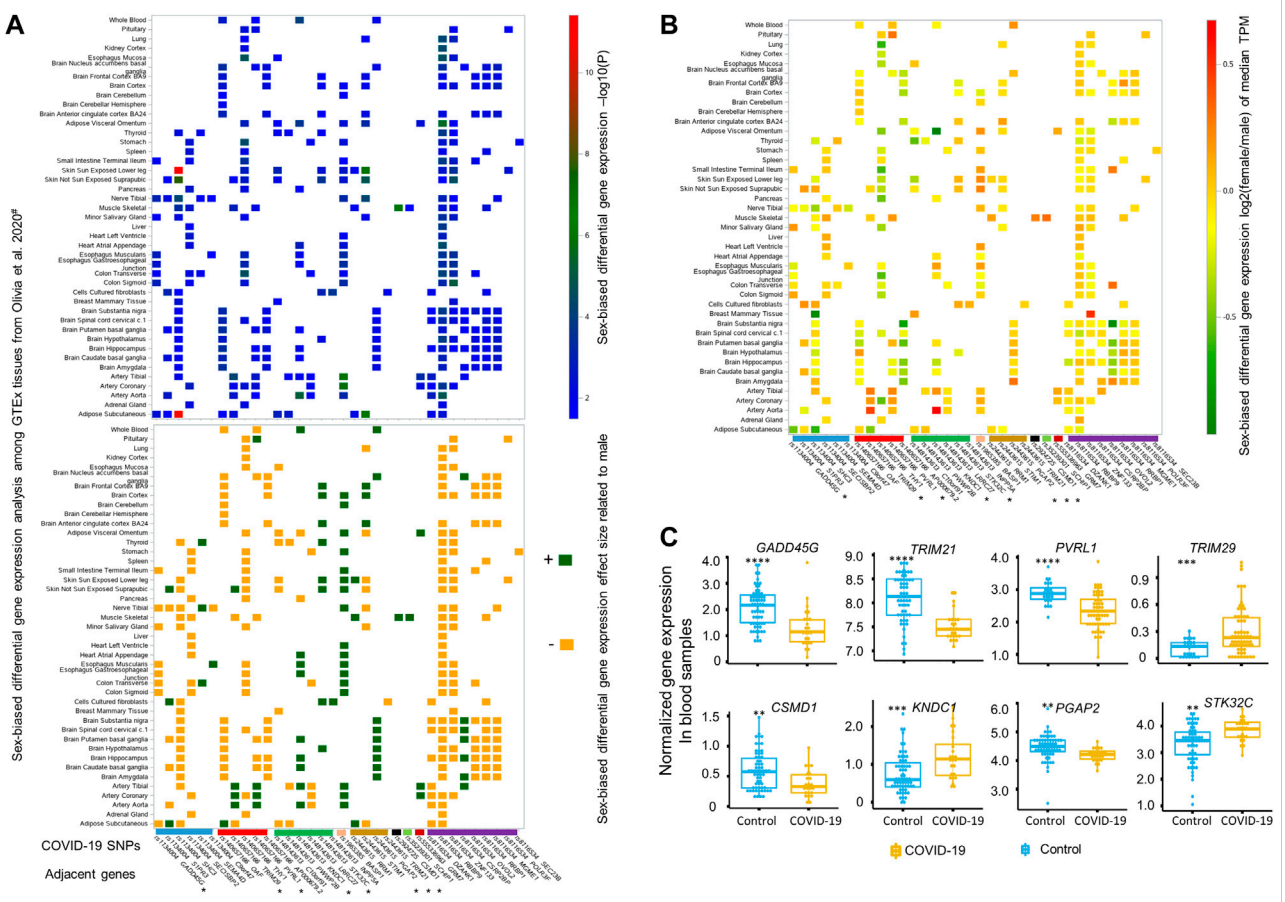



FIGURE 2
Differential gene expression analyses by sex or upon SARS-CoV-2 infection for genes close to these 12 top sex-biased COVID-19 risk SNPs. **(A)** Summary statistics of sex-biased differential gene expression analysis, including -log10(P) (top panel) and its corresponding direction of effect size (lower panel), for 34 sex-biased genes are illustrated based on the data generated by Olivia et al., across 43 different GTEx tissues (Oliva et al., 2020). The directions of effect size are represented by dark green ('+', higher expression in female) and dark yellow ('−', lower expression in female). The median gene expression fold change derived by comparing females to males for these 34 sex-biased genes is depicted in **(B)**, with a color bar on the right side indicating the magnitude of log2 (fold change) of median gene expression of Transcripts Per Million (TPM) between females and males. Among these 34 sex-biased genes, eight genes (marked with * in panel A and B) adjacent to five top sex-biased COVID-19 risk SNPs displaying up- or down-regulated expression **(C)** and also passed the multiple adjusted significance threshold of $p < 0.015$ in blood samples of COVID-19 patients (n = 62) compared to uninfected healthy controls (n = 23) (Thair et al., 2021), with statistical significance determined by one-way ANOVA test and denoted by **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$.

slightly lower expression [log2 (fold change) between −1 and 0] in females compared to males (see effect sizes and fold changes of these genes in Figures 2A,B). These genes are *GADD45G*, *SHC3*, *SECISBP2*, *THY1*, *AP0000679.2*, *C10orf91*, *TRIM21*, *DZANK1*, *RBBP9*, *ZNF133*, *OVOL2*, *RRBP1*, *MGME1*, *POLR3F*, and *SEC23B*, among which the expression of *TRIM21* is only significantly lower in cell cultured fibroblasts of females. Surprisingly, when comparing the expression of these genes in blood samples from COVID-19 patients with healthy controls in the COVID-19 expression database COVID19 db (Thair et al., 2021; Zhang et al., 2022), we found eight genes adjacent to five SNPs, including rs1134004 (*GADD45G*), rs140657166 (*TRIM29* and *PVRL1*), rs148143613 (*KNDC1* and *STK32C*), rs2443615 (*PGAP2* and *TRIM21*), and rs2924725 (*CSMD1*), are DEGs between blood of COVID-19 patients and healthy controls (multiple adjusted significance $p < 0.015$; one-way ANOVA test; Figure 2C). Collectively, our integrative analyses strongly support the potential involvements of these eight genes adjacent to five sex-biased COVID-19 SNPs in COVID-19 hospitalization.

## Discussion

By calculating the ΔZ-score of each SNP in the summary statistics of the sex-stratified COVID-19 hospitalization GWASs,

we identified 12 top SNPs showing suggestive, sex-biased differential associations with COVID-19 hospitalization between males and females with European ancestry. Among these 12 COVID-19 SNPs, we found five SNPs have eight adjacent genes that are DEGs between sexes and up- or down-regulated gene expression in blood samples from COVID-19 patients when compared with healthy controls. These eight genes include *PVRL1*, *TRIM29, TRIM21*, *GADD45G*, *CSMD1*, *KNDC1*, *PGAP2*, and *STK32C*.

Previous studies (Lopez et al., 1995; Takekawa and Saito, 1998; Warner et al., 1998; Kraus et al., 2006; Higgs et al., 2008; Hansen et al., 2013; Dempster et al., 2014; Zhang et al., 2014; Kimura et al., 2015; Hayashi et al., 2017; Starnawska et al., 2017; Dou et al., 2019; Jones et al., 2021; Thair et al., 2021; Zhang et al., 2022) related to these eight genes strongly suggest their potential involvements in COVID-19. Among them, the most promising genes are *PVRL1, TRIM29,* and *TRIM21,* which are all located on chromosome 11. Interestingly, *PVRL1* and *TRIM29* are both close to the sex-biased COVID-19 association SNP rs140657166, and *TRIM29* is adjacent to rs2443615, with the two SNPs 115.5-Mbp far away from each other. According to GeneCards (Safran et al., 2010), *PVRL1* encodes a protein involved in the organization of adherents junctions and tight junctions in epithelial and endothelial cells, and it is also reported as a receptor for herpes simplex virus 1, herpes simplex virus 2, and pseudorabies virus (Lopez et al., 1995; Warner et al., 1998). In our DGE analysis between sexes, *PVRL1* is highly expressed in multiple tissues of females, including artery-tibial, artery-coronary, and artery-aorta (see Figures 2A,B). Furthermore, *PVRL1* was down-regulated in the blood samples of COVID-19 patients according to the COVID-19 expression database COVID-19db (Figure 2C). Interestingly, *TRIM29,* which is close to *PVRL1*, shows higher expression in most female tissues, including artery-tibial, artery-coronary, artery-aorta, and skin not sun exposed suprapubic, except for adipose subcutaneous and muscle skeletal where the expression of *TRIM29* is higher in male and *TRIM29* is up-regulated in blood samples of COVID-19 patient compared to healthy controls. TRIM29 is involved in IFN-γ signaling and cytokine signaling by playing an essential role in activating macrophage upon viral or bacterial infections within the respiratory tract (Li et al., 2018; Dou et al., 2019). Additionally, we found *TRIM21* is another potential target, the expression of which is only higher in cell cultured fibroblasts of males than that of females. *TRIM21* encodes an E3 ubiquitin-protein ligase, which plays key roles in immune host defense, signal transduction, and possibly cell cycle regulation (Jones et al., 2021). Specifically, TRIM21 is involved in multiple immune responses, including the negative regulation of IFN-β production post-pathogen recognition (Higgs et al., 2008), the promotion of IRF8 ubiquitination that subsequently enhances the ability of IRF8 to stimulate cytokine gene transcription in macrophages (Jones et al., 2021), and the attenuation of type I IFN-dependent immune responses (Kimura et al., 2015). Taken together, all pieces of evidence strongly support the involvement of *PVRL1*, *TRIM29*, and *TRIM21* in COVID-19, but further investigation is warranted.

In our integrated analyses, we also revealed other five candidate genes potentially involved in COVID-19, including *GADD45G*, *PGAP2*, *CSMD1*, *KNDC1*, and *STK32C*. In the sex-biased DGE analysis, we revealed that the expression of *GADD45G* and *PGAP2* are specifically higher across multiple GTEx tissues of males and females. *CSMD1* only shows significantly higher expression in female muscle skeletal than that of male. The three genes are all down-regulated upon SARS-CoV-2 infection in blood samples of COVID-19 patients. Another two genes, *KNDC1* and *STK32C*, adjacent genes for the same SNP rs148148613, tend to be expressed higher in most of GTEx tissues of females than males, and they are stimulated in blood of COVID-19 patients. Further functional annotations for these genes also support their potential involvements in COVID-19. Among these genes, *GADD45G*, a stress and DNA-damaging responsive gene, encodes the protein GADD45G that is a mediator in activating the p38/JNK pathway *via* MTK1/MEKK4 kinase and consequently regulates cell growth and apoptosis (Takekawa and Saito, 1998). *CSMD1* encoded protein CSMD1 is suggested to affect learning or memory, mammary gland branching during pregnancy, and development of reproductive structure (Kraus et al., 2006). The protein PGAP2 encoded by the gene *PGAP2* influences the maturation of glycosylphosphatidylinositol (GPI) anchors on GPI-anchored proteins, and protein coding mutations of *PGAP2* were suggested to lead to an autosomal recessive syndrome characterized as hyperphosphatasia and intellectual disability (Hansen et al., 2013). KNDC1 is encoded by the gene *KNDC1*, and it is a Ras guanine nucleotide exchange factor that was suggested to be involved in the dendritic growth in the brain (Hayashi et al., 2017); it was also reported to be involved in aging *via* its roles in the senescence of umbilical vein endothelial cells (Zhang et al., 2014). *STK32C* encodes the protein STK32C that is a member of the serine/threonine protein kinase family. According to GeneCards (Safran et al., 2010), *STK32C* is known to be highly expressed in the human brain, with highest expression levels reported in the cerebellum and frontal cortex. Further studies displayed that the top differentially methylated probes were located within *STK32C* and hypomethylation of a CpG site in an intron of *STK32C* in individuals may contribute to depression disorder (Dempster et al., 2014; Starnawska et al., 2017). Furthermore, *STK32C* is involved in sweet taste signaling. Given their important functions in multiple tissues and sex-biased expression patterns in specific tissues, the interrupted expressions of these five genes upon SARS-CoV-2 infection may lead to

different COVID-19 symptoms between sexes. This may be especially true for *KNDC1* and *STK32C*, both of which are important to brain functions and show higher expression in female brain tissues; the interrupted expressions of the two genes may contribute to long COVID in a sex-biased manner. Taken together, the involvement of these five genes in COVID-19 hospitalization are needed for further investigations.

Our study has potential limitations. One limitation is the potential influence of the interaction between age and sex on the sex-biased COVID-19 association signals in the current study. However, the summary statistics of sex-stratified COVID-19 GWASs were generated by including age and 10 principal components as covariates. Determining the interaction of SNPs with sex and age on COVID-19 hospitalization requests us to perform COVID-19 GWAS from scratch by having raw genotyping data and clinical information, i.e., sex and age. It is well known that even larger sample sizes for both female and male GWASs are needed to detect the potential interactions between SNP and age / sex or age + sex. The sample sizes used by the current two COVID-19 hospitalization GWASs are underpowered to detect the above interactions. Another limitation is that these eight candidate genes display rather small fold change [absolute log2 (fold change) < 1] between sexes, and they were selected because of their close distances to 5 SNPs showing sex-biased associations with COVID-19 hospitalization. Specifically, we found the expression of *TRIM21* is relatively low with median TPM < 1 in muscle skeletal, although in which *TRIM21* demonstrates significant, sex-biased differential expression. We thus conclude that sex-biased differentiation of a gene need to be interpreted with caution if the gene showing lower expression in a specific tissue, especially when there are outliers with extremely high expression. The final limitation is about the re-analyzing of RNAseq data of blood of COVID-19 patients and healthy controls, in which we performed DGE analysis for genes adjacent to the candidate COVID-19 SNPs. Due to the relatively small sample size of the RNAseq data set and the unavailability of sex, age, and other clinical information for these samples, we could not exclude the potential effects of these confounding factors on the DEGs between COVID-19 patients and healthy controls, particularly for these candidate genes. We realize that it is necessary to conduct DGE analysis by including the interaction between sex and COVID-19 status, as well as other clinical information, into a linear regression model. If the interaction of sex and COVID-19 status for a candidate gene is significant, it would further support the potential involvement of the candidate gene in COVID-19.

In summary, our integrative comparison of male- and female-separated COVID-19 hospitalization GWASs provides new insights into the potential sex differences in COVID-19 hospitalization. Although no genome-wide significant SNPs showing differential COVID-19 associations between males and females, by conducting eQTL and DGE analysis between sexes among multiple GTEx tissues, and DGE analysis of blood samples between COVID-19 patients and healthy controls, we prioritize five SNPs with suggestive sex-biased differential association with COVID-19. We highlight eight genes close to these five candidate SNPs representing differential COVID-19 hospitalization associations between males and females. These genes, especially for *PVLR1*, *TRIM21*, and *TRIM29*, are strongly related to immune responses to infection, which are warranted for further cellular and animal studies.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

## Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## Author contributions

Y-SL, KZ, and Z-SC conceived the study. Z-SC, Y-SL, WL, YC, JZ, HG, and KZ proceeded the data statistics. Z-SC analyzed and interpreted the data. Z-SC, KZ, and Y-SL wrote and revised the manuscript.

## Funding

This project was funded by Science and Technology Supporting Plan of Scientific and Technological Department of Guizhou Province [(2020)4Y163], Science and Technology Program of Guiyang Science and Technology Bureau [(2020)-10-2], Cultivation Project of National Natural Science Foundation of China of the Affiliated Hospital of Guizhou Medical University and Guizhou Medical University (I-2020–06, 20NSP010).

## Acknowledgments

We are thankful to all researchers who participated and contributed to the GRASP project. As without their generosities in freely sharing the two COVID-19 hospitalization GWASs to public, we might not be able to have current finding. We also thank for the constructive suggestions provided by three reviewers.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1014191/full#supplementary-material

## References

Alwani, M., Yassin, A., Al-Zoubi, R. M., Aboumarzouk, O. M., Nettleship, J., Kelly, D., et al. (2021). Sex-based differences in severity and mortality in COVID-19. *Rev. Med. Virol.* 31 (6), e2223. doi:10.1002/rmv.2223

Cruz, R., Almeida, S. D. d., Heredia, M. L., Quintela, I., Ceballos, F. C., Pita, G., et al. (2022). Novel genes and sex differences in COVID-19 severity. *Hum. Mol. Genet.*, ddac132. ddac132. doi:10.1093/hmg/ddac132

Danielsen, A. C., Lee, K. M., Boulicault, M., Rushovich, T., Gompers, A., Tarrant, A., et al. (2022). Sex disparities in COVID-19 outcomes in the United States: Quantifying and contextualizing variation. *Soc. Sci. Med.* 294, 114716. doi:10.1016/j.socscimed.2022.114716

Dempster, E. L., Wong, C. C. Y., Lester, K. J., Burrage, J., Gregory, A. M., Mill, J., et al. (2014). Genome-wide methylomic analysis of monozygotic twins discordant for adolescent depression. *Biol. Psychiatry* 76 (12), 977–983. doi:10.1016/j.biopsych.2014.04.013

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *Lancet. Infect. Dis.* 20 (5), 533–534. doi:10.1016/S1473-3099(20)30120-1

Dou, Y., Xing, J., Kong, G., Wang, G., Lou, X., Xiao, X., et al. (2019). Identification of the E3 ligase TRIM29 as a critical checkpoint regulator of NK cell functions. *J. Immunol.* 203 (4), 873–880. doi:10.4049/jimmunol.1900171

Hansen, L., Tawamie, H., Murakami, Y., Mang, Y., ur Rehman, S., Buchert, R., et al. (2013). Hypomorphic mutations in PGAP2, encoding a GPI-anchor-remodeling protein, cause autosomal-recessive intellectual disability. *Am. J. Hum. Genet.* 92 (4), 575–583. doi:10.1016/j.ajhg.2013.03.008

Hayashi, K., Furuya, A., Sakamaki, Y., Akagi, T., Shinoda, Y., Sadakata, T., et al. (2017). The brain-specific RasGEF very-KIND is required for normal dendritic growth in cerebellar granule cells and proper motor coordination. *PLoS One* 12 (3), e0173175. doi:10.1371/journal.pone.0173175

Higgs, R., Ni Gabhann, J., Ben Larbi, N., Breen, E. P., Fitzgerald, K. A., and Jefferies, C. A. (2008). The E3 ubiquitin ligase Ro52 negatively regulates IFN-beta production post-pathogen recognition by polyubiquitin-mediated degradation of IRF3. *J. Immunol.* 181 (3), 1780–1786. doi:10.4049/jimmunol.181.3.1780

Jones, E. L., Laidlaw, S. M., and Dustin, L. B. (2021). TRIM21/Ro52 - roles in innate immunity and autoimmune disease. *Front. Immunol.* 12, 738473. doi:10.3389/fimmu.2021.738473

Kimura, T., Jain, A., Choi, S. W., Mandell, M. A., Schroder, K., Johansen, T., et al. (2015). TRIM-mediated precision autophagy targets cytoplasmic regulators of innate immunity. *J. Cell Biol.* 210 (6), 973–989. doi:10.1083/jcb.201503023

Kraus, D. M., Elliott, G. S., Chute, H., Horan, T., Pfenninger, K. H., Sanford, S. D., et al. (2006). CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. *J. Immunol.* 176 (7), 4419–4430. doi:10.4049/jimmunol.176.7.4419

Li, Q., Lin, L., Tong, Y., Liu, Y., Mou, J., Wang, X., et al. (2018). TRIM29 negatively controls antiviral immune response through targeting STING for degradation. *Cell Discov.* 4, 13. doi:10.1038/s41421-018-0010-9

Lopez, M., EberleF.,Mattei, M. G., Gabert, J., BirgF.BardinF., et al. (1995). Complementary DNA characterization and chromosomal localization of a human gene related to the poliovirus receptor-encoding gene. *Gene* 155 (2), 261–265. doi:10.1016/0378-1119(94)00842-g

Oliva, M., Munoz-Aguirre, M., Kim-Hellmuth, S., Wucher, V., Gewirtz, A. D. H., Cotter, D. J., et al. (2020). The impact of sex on gene expression across human tissues. *Science* 369 (6509), eaba3066. doi:10.1126/science.aba3066

Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., Pasko, D., et al. (2021). Genetic mechanisms of critical illness in COVID-19. *Nature* 591 (7848), 92–98. doi:10.1038/s41586-020-03065-y

Safran, M., Dalah, I., Alexander, J., Rosen, N., Shmoish, M., Nati, N., et al. (2010). GeneCards version 3: The human gene integrator. *Database (Oxford)* 2010, baq020. doi:10.1093/database/baq020

Shattuck-Heidorn, H., Danielsen, A. C., Gompers, A., Bruch, J. D., Zhao, H., Boulicault, M., et al. (2021). A finding of sex similarities rather than differences in COVID-19 outcomes. *Nature* 597 (7877), E7–E9. doi:10.1038/s41586-021-03644-7

Starnawska, A., Sparso, T., Mazin, W., Olsen, L., BertalanM., et al. (2017). Differential DNA methylation at birth associated with mental disorder in individuals with 22q11.2 deletion syndrome. *Transl. Psychiatry* 7 (8), e1221. doi:10.1038/tp.2017.181

Takahashi, T., Ellingson, M. K., Wong, P., Israelow, B., Lucas, C., Klein, J., et al. (2021). Reply to: A finding of sex similarities rather than differences in COVID-19 outcomes. *Nature* 597 (7877), E10–E11. doi:10.1038/s41586-021-03645-6

Takahashi, T., Ellingson, M. K., Wong, P., Israelow, B., Lucas, C., Klein, J., et al. (2020). Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature* 588 (7837), 315–320. doi:10.1038/s41586-020-2700-3

Takekawa, M., and Saito, H. (1998). A family of stress-inducible GADD45-like proteins mediate activation of the stress-responsive MTK1/MEKK4 MAPKKK. *Cell* 95 (4), 521–530. doi:10.1016/s0092-8674(00)81619-0

Thair, S. A., He, Y. D., Hasin-Brumshtein, Y., Sakaram, S., Pandya, R., Toh, J., et al. (2021). Transcriptomic similarities and differences in host response between SARS-CoV-2 and other viral infections. *iScience* 24 (1), 101947. doi:10.1016/j.isci.2020.101947

Thibord, F., Chan, M. V., Chen, M. H., and Johnson, A. D. (2022). A year of Covid-19 GWAS results from the GRASP portal reveals potential genetic risk factors. *HGG Adv.* 3, 100095. doi:10.1016/j.xhgg.2022.100095

The GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369 (6509), 1318–1330. doi:10.1126/science.aaz1776

The Severe Covid-19 GWAS Group (2020). Genomewide association study of severe covid-19 with respiratory failure. *N. Engl. J. Med.* 383 (16), 1522–1534. doi:10.1056/NEJMoa2020283

Tu, H., Vermunt, J. V., Abeykoon, J., Ghamrawi, R., Gunaratne, M., Jayachandran, M., et al. (2020). COVID-19 and sex differences: Mechanisms and biomarkers. *Mayo Clin. Proc.* 95 (10), 2189–2203. doi:10.1016/j.mayocp.2020.07.024

Viveiros, A., Rasmuson, J., Vu, J., Mulvagh, S. L., Yip, C. Y. Y., Norris, C. M., et al. (2021). Sex differences in COVID-19: Candidate pathways, genetics of ACE2, and sex hormones. *Am. J. Physiol. Heart Circ. Physiol.* 320 (1), H296–h304. doi:10.1152/ajpheart.00755.2020

Võsa, U., Claringbould, A., Bonder, M. J., Deelen, P., Zeng, B., Kirsten, H., et al. (2018). *Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis*. bioRxiv preprint.

Warner, M. S., Geraghty, R. J., Martinez, W. M., Montgomery, R. I., Whitbeck, J. C., Xu, R., et al. (1998). A cell surface protein with herpesvirus entry activity (HveB) confers susceptibility to infection by mutants of herpes simplex virus type 1, herpes simplex virus type 2, and pseudorabies virus. *Virology* 246 (1), 179–189. doi:10.1006/viro.1998.9218

Williamson, E. J., Walker, A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., et al. (2020). Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 584 (7821), 430–436. doi:10.1038/s41586-020-2521-4

Zhang, C., Zhen, Y. Z., Lin, Y. J., Liu, J., Wei, J., Xu, R., et al. (2014). KNDC1 knockdown protects human umbilical vein endothelial cells from senescence. *Mol. Med. Rep.* 10 (1), 82–88. doi:10.3892/mmr.2014.2201

Zhang, W., Zhang, Y., Min, Z., Mo, J., Ju, Z., Guan, W., et al. (2022). COVID19db: A comprehensive database platform to discover potential drugs and targets of COVID-19 at whole transcriptomic scale. *Nucleic Acids Res.* 50 (D1), D747–D757. doi:10.1093/nar/gkab850

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 30 November 2022
DOI 10.3389/fgene.2022.1073880



OPEN ACCESS

EDITED BY
Zhongshan Cheng,
St. Jude Children's Research Hospital, United States

REVIEWED BY
Valerio Caputo,
University of Rome Tor Vergata, Italy
Yiyun Lin,
University of Texas MD Anderson Cancer Center, United States

*CORRESPONDENCE
Changjun Wang,
science2008@hotmail.com
Yan Zhang,
zany1983@163.com
Zhihu Zhao,
zhaozh@bmi.ac.cn

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 19 October 2022
ACCEPTED 18 November 2022
PUBLISHED 30 November 2022

CITATION
Li P, Ke Y, Shen W, Shi S, Wang Y, Lin K, Guo X, Wang C, Zhang Y and Zhao Z (2022), Targeted screening of genetic associations with COVID-19 susceptibility and severity. *Front. Genet.* 13:1073880. doi: 10.3389/fgene.2022.1073880



# Targeted screening of genetic associations with COVID-19 susceptibility and severity

Ping Li[1†], Yuehua Ke[2†], Wenlong Shen[1†], Shu Shi[1], Yahao Wang[1], Kailin Lin[3], Xinjie Guo[1], Changjun Wang[2*], Yan Zhang[1*] and Zhihu Zhao[1*]

[1]Beijing Institute of Biotechnology, Beijing, China, [2]Center for Disease Control and Prevention of PLA, Beijing, China, [3]College of Life Science, Henan Normal University, Xinxiang, China

The COVID-19 pandemic has resulted in great morbidity and mortality worldwide and human genetic factors have been implicated in the susceptibility and severity of COVID-19. However, few replicate researches have been performed, and studies on associated genes mainly focused on genic regions while regulatory regions were a lack of in-depth dissection. Here, based on previously reported associated variants and genes, we designed a capture panel covering 1,238 candidate variants and 25 regulatory regions of 19 candidate genes and targeted-sequenced 96 mild and 145 severe COVID-19 patients. Genetic association analysis was conducted between mild and severe COVID-19 patients, between all COVID-19 patients and general population, or between severe COVID-19 patients and general population. A total of 49 variants were confirmed to be associated with susceptibility or severity of COVID-19 ($p < 0.05$), corresponding to 18 independent loci. Specifically, rs1799964 in the promoter of inflammation-related gene *TNF*, rs9975538 in the intron of interferon receptor gene *IFNAR2*, rs429358 in the exon of *APOE*, rs1886814 in the intron of *FOXP4-AS1* and a list of variants in the widely reported 3p21.31 and *ABO* gene were confirmed. It is worth noting that, for the confirmed variants, the phenotypes of the cases and controls were highly consistent between our study and previous reports, and the confirmed variants identified between mild and severe patients were quite different from those identified between patients and general population, suggesting the genetic basis of susceptibility and severity of SARS-CoV-2 infection might be quite different. Moreover, we newly identified 67 significant associated variants in the 12 regulatory regions of 11 candidate genes ($p < 0.05$). Further annotation by RegulomeDB database and GTEx eQTL data filtered out two variants (rs11246060 and rs28655829) in the enhancer of broad-spectrum antiviral gene *IFITM3* that might affect disease severity by regulating the gene expression. Collectively, we confirmed a list of previously reported variants and identified novel regulatory variants associated with susceptibility and severity of COVID-19, which might

provide biological and clinical insights into COVID-19 pathogenesis and treatment.

KEYWORDS

COVID-19, SARS-CoV-2, targeted capture sequencing, SNP, genetic variant, enhancer, suscepibility

## Introduction

Coronavirus disease 2019 (COVID-19), an infectious disease caused by Severe Acute Respiratory Syndrome-Coronavirus 2 (SARS-CoV-2) (Zhu et al., 2020), has spread worldwide, resulting more than 460 million infections and six million deaths up to 16 March 2022 (https://covid19.who.int/). The occurrence and clinical outcomes of COVID-19 have been revealed great heterogeneity, ranging from insensitive, asymptomatic, mild, moderate to severe, critical or even death (Wu and McGoogan, 2020). Host factors such as age, gender, comorbidities were reported to be associated with this heterogeneity (Zhou et al., 2020a; Richardson et al., 2020).

Host genetic variants might also affect susceptibility and severity of coronavirus infection, as indicated by previous studies of SARS, Middle East Respiratory Syndrome (MERS) and emerging studies of COVID-19 (Di Maria et al., 2020). The first genome-wide association study (GWAS) of COVID-19 reported two severity-associated loci in Italians and Spanish: the 3p21.31 locus containing several immune genes and *ABO* locus determining *ABO* blood groups (Ellinghaus et al., 2020). The COVID-19 Host Genetics Initiative (HGI) was established to bring together global effort to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic (Initiative, 2020). However, current reported genetic studies of COVID-19 are mainly based on European populations. Whether these findings could apply to other populations was unknown.

Besides GWAS studies, several candidate gene studies indicated that certain variants in the type I interferon (IFN) pathway genes and SARS-CoV-2 receptor/coreceptor genes were associated with susceptibility and severity of COVID-19 (Zhang et al., 2020a; Benetti et al., 2020; Zhang et al., 2020b; Kuo et al., 2020; Latini et al., 2020; Novelli et al., 2020). For example, rs12252, variant of *IFITM3*, was reported to be associated with severe COVID-19 (Zhang et al., 2020b). Many rare loss-of-function variants in IFN-pathway genes such as *TLR3*, *IRF3*, *IRF7*, *IFNAR1,* and *IFNAR2* were identified to be associated with severe COVID-19 through impairing IFN immunity (Zhang et al., 2020a). Furthermore, in the *ACE2* and *TMPRSS2* genes, receptor and coreceptor gene for SARS-CoV-2 respectively, certain variants showed significantly different allele frequencies between COVID-19 patients and general population (Benetti et al., 2020; Latini et al., 2020; Novelli et al., 2020). However, these studies only focus on the genic region. The regulatory regions of these important genes are a lack of attention. Variants in regulatory regions, especially enhancers, could disrupt regulatory function, affect gene expression and thus contribute to susceptibility and severity of virus infection (Li et al., 2017; Downes et al., 2021). Therefore, it is necessary to pay more attention to variants in enhancers of important genes.

Here, we incorporated results of previous studies and designed a capture panel covering previously reported associated variants and regulatory regions of key genes. Using this panel, we targeted sequenced 96 mild and 145 severe COVID-19 patients. Genetic association analysis was conducted between mild and severe cases as well as ancestry-matched populations from 1000 Genomes Project. A list of previously reported associated variants was confirmed and two variants in the enhancers of *IFITM3* genes that might affect disease severity by regulating the gene expression were newly identified, which will lead to a better understanding of the host genetic factors at play in COVID-19.

## Materials and methods

### Study participants and recruitment

This study included 241 hospitalized COVID-19 patients recruited from Huoshenshan hospital at Wuhan city, Hubei province, China between 11 January 2020 and 11 March 2020. COVID-19 was diagnosed based on chest computed tomography (CT) manifestations and/or reverse transcription-polymerase chain reaction (RT-PCR) following the criteria of the New Coronavirus Pneumonia Prevention and Control Program (5th edition). In this study, the mild COVID-19 patients were those with no obvious clinical symptoms or with fever, respiratory symptoms, and radiological evidence of pneumonia. The severe COVID-19 patients were those having at least one of the following conditions: respiratory distress, respiratory rate ≤30 beats/minute; mean oxygen saturation ≤93% in a resting state; arterial blood oxygen partial pressure/oxygen concentration ≤300 mm·Hg; respiratory failure and requiring mechanical ventilation; shock; and admission to intensive care unit (ICU) with other organ function failure. Classifications of COVID-19 severity were taken as the worst classification during the patient's hospital stay.

The clinical characteristics of the patients were extracted from the electronic medical records. We collected three broad classes of characteristics: 1) demographic variables (age, sex, and ethnicity); 2) symptoms (fever and diarrhea); and 3) comorbid

conditions (hypertension, diabetes, cardiac disease, chronic bronchitis, chronic liver disease, chronic obstructive pulmonary disease, cerebrovascular disease, and cancer).

The Ethics Committee of Huoshenshan Hospital approved the study (HSSLL036). Given the urgency of the COVID-19 pandemic, the need for informed consent was waived by the ethnics boards of the hospital.

## Candidate variants and enhancers selection and probe design

We collected 30 unique variants associated with susceptibility or severity of SARS-CoV/SARS-CoV-2 infection from 21 papers (up to 29 November 2020), referred as "literature dataset". In addition, we downloaded COVID19-hg GWAS meta-analysis results (release 4) produced by COVID-19 host genetics initiative (HGI) from https://www.covid19hg.org/results/r4/, and 1420 unique variants were selected that meet one of the following four requirements: 1) variants with $p$ < 1E-5 in "A1_ALL" group, that is, phenotype of very severe respiratory confirmed covid vs. not hospitalized covid; 2) variants with $p$ < 1E-5 in "B1_ALL" group, that is, phenotype of hospitalized covid vs. not hospitalized covid; 3) variants with $p$ < 5E-7 in any one of group; 4) variants with $p$ < 1E-5 in any three of groups except "A1_ALL" and "B1_ALL". This was referred as "HGI dataset". Two variants were overlapped in the two datasets, resulting in a total of 1448 unique variants collected.

IFNs play a central role in innate immunity against virus infection (Zhang et al., 2020a; Bastard et al., 2020) and cell receptors for virus are key determinant for viral entry (Hoffmann et al., 2020a; Shang et al., 2020). Therefore, We collected 19 IFN-pathway genes previously implicated in SARS-CoV/SARS-CoV-2 susceptibility and severity (Hamano et al., 2005; He et al., 2006; Ching et al., 2010; Zhang et al., 2020a; Zhang et al., 2020b) as well as 3 human cell receptors/co-receptors for SARS-CoV-2 (Hoffmann et al., 2020a; Hoffmann et al., 2020b; Zhou et al., 2020b; Shang et al., 2020). A total of 25 potential enhancers for these genes were obtained from GeneHancer database (GH score > 1, Gene Association > 100) (Fishilevich et al., 2017).

RNA probes were designed to cover these variants and gene enhancers. All the 25 enhancers and 1238 unique candidate variants were covered, including 29 variants in the literature dataset and 1211 variants in the HGI dataset (Supplementary Tables S1–S3).

## Targeted capture sequencing

Peripheral whole blood samples were collected from all participants. Genomic DNAs were extracted from 1 ml of peripheral whole blood, according to the manufacturer's instructions (QIAamp DNA blood kits). The quality of the isolated genomic DNA was verified by the following two methods: 1) the DNA degradation and contamination were monitored in 1% agarose gels; and 2) the DNA concentration was measured using a Qubit DNA Assay Kit and a Qubit 3.0 Fluorometer (Life Technologies).

The targeted capture sequencing was conducted by iGeneTech Bioscience Corporation (Beijing, China). Briefly, Human genomic DNA was sheared to 150–200 bp by Bioruptor Pico (Diagenode). Then end repair, dA-tailing, and adapter ligation were performed. The ligation product was cleaned up and size-selected by using Beckman Ampure XP Beads (Beckman). The purified ligated product was amplified by using PCR. Then the library was in solution hybrid with biotinylated RNA probes, captured with Dynabeads MyOne Streptavidin T1 (Invitrogen), and amplified with PCR. The library was quantified by Qubit and fragment-size measured by Agilent 2100 Bioanalyzer system before high-throughput sequenced by NovaSeq.

## Variant calling and genetic association analysis

Raw reads were firstly quality trimmed with Trimmomatic (Bolger et al., 2014). Clean reads were then aligned to the human reference genome (hg38) using BWA algorithm (Li and Durbin, 2009). PCR duplicates were removed using samtools (Li et al., 2009), and GATK software (McKenna et al., 2010) was used to call SNPs and indels. The detected variants were finally saved as VCF files. Data of autosomal biallelic variants for Han Chinese of 1000 Genomes Project (Auton et al., 2015) were downloaded from https://www.internationalgenome.org/. Genetic association analysis was conducted using PLINK 1.9 software (Chang et al., 2015).

## Statistical analysis

$p$-values comparing demographics severe and mild disease groups were calculated by means of $\chi 2$ or Fisher exact test as appropriate, except for $p$-value for age which was calculated using student's t test. Minor allele frequency (MAF) of variants was compared between case and control groups using Fisher exact test. Logistic regression analysis was also used to compute the contributing variants to severity with adjusting age, sex and comorbidities. Statistic power was calculated using G*Power 3 software (Faul et al., 2007). $p$ < 0.05 was considered statistically significant.

# Results

## Clinical features of the COVID-19 patients

A total of 241 COVID-19 patients were recruited, with 96 mild cases and 145 severe cases. Demographic and

TABLE 1 Patient demographics by clinical phenotype.

| Characteristic | | Patients hospitalized with COVID-19, no. (%)[a] | | |
|---|---|---|---|---|
| | | Mild ($N$ = 96) | Severe ($N$ = 145) | $p$-value* |
| Age, median (IQR) | | 58 (50.5–66.5) | 67 (59–74) | <0.0001*** |
| Male | | 45 (46.88) | 81 (55.86) | 0.17 |
| Comorbidities | | 32 (35.45) | 69 (50.83) | 0.03* |
| | Hypertension | 23 (22.73) | 48 (32.04) | 0.13 |
| | Diabetes | 9 (8.18) | 28 (20.44) | 0.04* |
| | Cardiac disease | 5 (4.55) | 13 (8.84) | 0.33 |
| | Chronic bronchitis | 0 (0.91) | 4 (2.76) | 0.15 |
| | COPD | 0 (0) | 2 (1.1) | 0.52 |
| | Chronic liver disease | 3 (2.73) | 3 (1.66) | 0.68 |
| | Cancer | 1 (0.91) | 6 (3.31) | 0.25 |
| | Cerebrovascular disease | 1 (0.91) | 11 (6.63) | 0.03* |
| Presenting symptoms | | | | |
| | Fever | 71 (71.82) | 106 (71.82) | 0.88 |
| | Diarrhea | 1 (1.82) | 7 (4.42) | 0.15 |

[a]Data represent no. (%) of patients unless otherwise specified.
*$p$-values comparing severe and mild disease groups were calculated by means of $\chi^2$ or Fisher exact tests as appropriate, except for $p$-value for age which was calculated using student's $t$ test.
Abbreviations: IQR, interquartile range; COPD, chronic obstructive pulmonary disease.

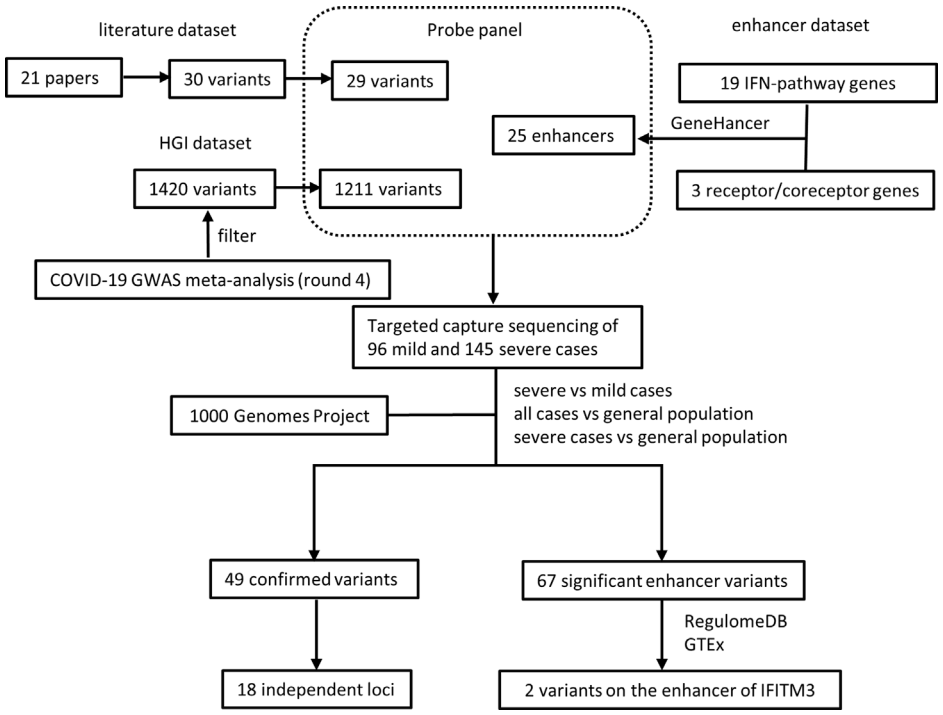



FIGURE 1
Illustration of experimental design and analysis.

TABLE 2 Independent lead variants that were confirmed between mild and severe cases.

| SNP | Position | Allele (major/minor) | MAF (case/control) | *p* | OR (95%CI) | Mapped gene | Previous study | Previous study phenotype |
|---|---|---|---|---|---|---|---|---|
| rs13062942 | 3:62951091 | A/G | 0.51/0.39 | 1.17E-02 | 1.63 (1.12–2.35) | Intron of *LINC00698* | HGI | Hospitalized covid vs. not hospitalized covid |
| rs28373011 | 15:66669812 | G/C | 0.36/0.47 | 1.78E-02 | 0.63 (0.43–0.91) | Intron of *LINC01169* | HGI | Very severe respiratory confirmed covid vs. not hospitalized covid |
| rs1799964 | 6:31574531 | T/C | 0.15/0.23 | 2.92E-02 | 0.59 (0.37–0.93) | Upstream of *TNF* | PMID: 18312678 | SARS cases with femoral head necrosis vs. SARS cases without femoral head necrosis |
| rs2224986 | 1:152712390 | C/T | 0.13/0.07 | 3.75E-02 | 1.98 (1.04–3.75) | Downstream of *LCE4A* | HGI | Hospitalized covid vs. not hospitalized covid |

phenotypic data are shown in Table 1. Comparing mild patients with severe ones, the median age increased from 58 to 67 years ($p < 0.0001$). In addition, we observed a greater percentage of severe patients with comorbidities than mild patients ($p = 0.03$), specifically with diabetes ($p = 0.04$) and cerebrovascular disease ($p = 0.03$). This is in accordance with previous report that older COVID-19 patients and those with comorbidities were more likely to be severe in disease (Zhou et al., 2020a; Richardson et al., 2020). Previous report also indicated that severe patients had a greater percentage of male cases (Richardson et al., 2020). However, no obvious sex difference was found between mild and severe patients in our data ($p = 0.17$).

## Targeted capture sequencing

The experimental design was illustrated in Figure 1. To validate previously reported variants associated with COVID-19 susceptibility and severity, we collected 30 unique variants from 21 papers (referred as "literature dataset") and selected 1420 unique variants from COVID-19 host genetics initiative (HGI) release 4 (Initiative, 2020) (referred as "HGI dataset"). Additionally, to identify regulatory variants associated with COVID-19 susceptibility and severity in the enhancers of previously reported associated genes, we collected 19 IFN-pathway genes previously implicated in coronavirus susceptibility and severity (Hamano et al., 2005; He et al., 2006; Ching et al., 2010; Zhang et al., 2020a; Zhang et al., 2020b) as well as 3 human cell receptors/co-receptors for SARS-CoV-2 (Hoffmann et al., 2020a; Hoffmann et al., 2020b; Zhou et al., 2020b; Shang et al., 2020) and obtained 25 potential enhancers for these genes from GeneHancer database (Fishilevich et al., 2017) (referred as "enhancer dataset").

A panel of RNA probes was designed to capture these variant and enhancers, resulting in 1238 unique variants and all the 25 potential enhancers covered (Supplementary Tables S1–S3). The 1238 unique variants included 29 variants in the literature dataset and 1211 variants in the HGI dataset, with two variants overlapped between the two datasets. Variants that fail to design probes might be due to GC content, repetitive sequence, dimer or secondary structure of probes.

The probe panel was used to targeted capture sequence the 241 COVID-19 patients. A median of 8.5 million raw reads were obtained for each sample. After filtering, a median of 8.3 million reads was kept as clean reads, with a median of average insert size of 173 bp. The reads were mapped to hg38 genome. The median mapping rate was 99.56%, with a median duplication rate of 27.47% (Supplementary Figure S1). Specifically, the median number of target mapped reads was 4.5 million, with a median target coverage rate of 99.57% and median target mean depth of 1472× (Supplementary Figure S2).

## Confirmation of genetic variants associated susceptibility and severity

Out of the 1238 candidate variants in the panel, a total of 1006 variants in the panel were identified in the 241 COVID-19 patients, including 26 variants in the literature dataset and 982 variants in the HGI dataset.

Comparing 96 mild and 145 severe COVID-19 patients, seven variants were confirmed to be significantly different in minor allele frequency (MAF) ($p < 0.05$, Supplementary Table S2), corresponding to four independent loci. The four lead variants were shown in Table 2. Specifically, rs1799964 in the promoter of inflammation-related gene *TNF* was found to be associated with COVID-19 severity. CC genotype of this variant has been reported to be associated with femoral head necrosis after SARS-CoV infection (Wang et al., 2008). In addition, rs2224986 and rs13062942 were still associated with COVID-19 severity after adjusting age, sex and comorbidities on regression analysis. Both variants had a significant difference between hospitalized and non-hospitalized patients in the HGI dataset.

When comparing all COVID-19 patients with ancestry-matched general population from 1000 Genomes Project, we

TABLE 3 Independent lead variants that were confirmed between all COVID-19 patients and general population.

| SNP | Position | Allele (major/minor) | MAF (case/control) | *p* | OR (95%CI) | Mapped gene | Previous study | Previous study phenotype |
|---|---|---|---|---|---|---|---|---|
| rs1291122587 | 9:133270497 | GA/G | 0.26/0.59 | 6.68E-24 | 0.24 (0.18–0.32) | Intron of *ABO* | HGI | Hospitalized covid vs. population |
| | | | | | | | | Hospitalized covid vs. population, leave out 23andMe (EUR) |
| | | | | | | | | Covid vs. population |
| | | | | | | | | Covid vs. population, leave out 23andMe (EUR) |
| rs429358 | 19:44908684 | T/C | 0.1/0 | 6.67E-15 | NA | Missense of *APOE* | PMID: 32451547 | Positive COVID-19 patients vs. participants negative or not tested |
| rs71327056 | 3:46348485 | C/G | 0.01/0.04 | 2.74E-03 | 0.21 (0.07–0.64) | Intergenic of *UQCRC2P1-CCR2* | HGI | Very severe respiratory confirmed covid vs. population |
| | | | | | | | | Hospitalized covid vs. population |
| | | | | | | | | Hospitalized covid vs. population, leave out 23andMe (EUR) |
| rs9411475 | 9:133251881 | T/C | 0.33/0.26 | 1.08E-02 | 1.46 (1.09–1.95) | Downstream of *ABO* | HGI | Covid vs. population |
| rs495828 | 9:133279294 | G/T | 0.28/0.2 | 1.27E-02 | 1.49 (1.09–2.03) | Upstream of *ABO* | HGI | Hospitalized covid vs. population |
| | | | | | | | | Covid vs. lab/self-reported negative |
| | | | | | | | | Covid vs. population |
| | | | | | | | | Very severe respiratory confirmed covid vs. population |
| | | | | | | | | Hospitalized covid vs. population, leave out 23andMe (EUR) |
| rs10734222 | 11:14244913 | G/A | 0.18/0.24 | 2.19E-02 | 0.68 (0.49–0.94) | Intron of *SPON1* | HGI | Hospitalized covid vs. population, leave out 23andMe (EUR) |
| rs1886814 | 6:41534945 | A/C | 0.45/0.37 | 2.55E-02 | 1.36 (1.04–1.77) | Intron of *FOXP4-AS1* | HGI | Hospitalized covid vs. population |
| rs973579 | 19:48738719 | A/G | 0/0.02 | 2.88E-02 | 0.12 (0.02–1.01) | Upstream of *RASIP1* | HGI | Covid vs. lab/self-reported negative |
| rs74609750 | 21:33896073 | C/A | 0.3/0.23 | 2.93E-02 | 1.40 (1.04–1.89) | 3′ UTR of *ITSN1* | HGI | Covid vs. population |
| rs898467998 | 9:133260743 | G/A | 0.27/0.34 | 3.58E-02 | 0.74 (0.55–0.98) | Intron of *ABO* | HGI | Covid vs. lab/self-reported negative |
| | | | | | | | | Covid vs. population |
| | | | | | | | | Hospitalized covid vs. population |

identified 39 significant variants ($p < 0.05$, Supplementary Table S5), corresponding 10 independent loci. The independent lead variants were shown in Table 3. Specifically, multiple genetic variants in the *ABO* gene locus and 3p21.31 region have been validated. In addition, the missense variant rs429358 in exon of *APOE* gene, which has been reported to be associated with COVID-19-positive in the UK Biobank data (Kuo et al., 2020), was confirmed in Chinese population. As for the confirmed variant rs1886814 in the intron of *FOXP4-AS1* (forkhead box P4 antisense RNA 1), recent trans-ethnic genome-wide association study of severe COVID-19 that incorporated Chinese population and HGI results also revealed a significant variant nearby, rs1853837 (Wu et al., 2021), which is LD with rs1886814 (1000 Genomes Project CHB, $r^2 = 0.68$).

When comparing severe COVID-19 patients with ancestry-matched general population from 1000 Genomes Project, we identified 23 significant variants ($p < 0.05$, Supplementary Table S6), corresponding 7 independent loci. The independent lead variants were shown in Table 4. Again, multiple genetic variants in the *ABO* gene locus and 3p21.31 region, rs429358 in exon of *APOE* gene and rs1886814 in the intron of *FOXP4-AS1* have been validated. What's more, rs9975538 in the intron of gene *IFNAR2* was also confirmed. *IFNAR2* gene, along with *IFNAR1* encodes type I interferon receptor, thus the rs9975538 might affect type I interferon pathway.

TABLE 4 Independent lead variants that were confirmed between severe COVID-19 patients and general population.

| SNP | Position | Allele (major/minor) | MAF (case/control) | *p* | OR (95%CI) | Mapped gene | Previous study | Previous study phenotype |
|---|---|---|---|---|---|---|---|---|
| rs1291122587 | 9:133270497 | GA/G | 0.26/0.59 | 7.02E-18 | 0.24 (0.18–0.34) | Intron of *ABO* | HGI | Hospitalized covid vs. population |
| | | | | | | | | Hospitalized covid vs. population, leave out 23 and Me (EUR) |
| | | | | | | | | Covid vs. population |
| | | | | | | | | Covid vs. population, leave out 23 and Me (EUR) |
| rs8111981 | 19:977763 | G/A | 0.28/0.35 | 3.37E-02 | 0.70 (0.50–0.97) | Downstream of *ARID3A* | HGI | Very severe respiratory confirmed covid vs. population |
| rs9975538 | 21:33254549 | C/T | 0.29/0.37 | 4.37E-02 | 0.71 (0.52–0.98) | Intron of *IFNAR2* | HGI | Very severe respiratory confirmed covid vs. population |
| | | | | | | | | Hospitalized covid vs. population |
| | | | | | | | | Hospitalized covid vs. population, leave out 23 and Me (EUR) |
| rs429358 | 19:44908684 | T/C | 0.11/0 | 4.29E-14 | NA | Missense of *APOE* | PMID: 32451547 | Positive COVID-19 patients vs. participants negative or not tested |
| rs10734222 | 11:14244913 | G/A | 0.17/0.24 | 1.19E-02 | 0.61 (0.42–0.90) | Intron of *SPON1* | HGI | Hospitalized covid vs. population, leave out 23 and Me (EUR) |
| rs59492037 | 6:4571951 | A/G | 0.27/0.35 | 2.63E-02 | 0.68 (0.49–0.95) | Intron of *AL162718.1* | HGI | Very severe respiratory confirmed covid vs. not hospitalized covid |
| rs4302292 | 21:33919239 | G/A | 0.31/0.23 | 3.07E-02 | 1.47 (1.05–2.05) | Upstream of *ATP5PO* | HGI | Covid vs. population |
| rs71327056 | 3:46348485 | C/G | 0.01/0.04 | 3.12E-02 | 0.27 (0.08–0.92) | Intergenic of *UQCRC2P1-CCR2* | HGI | Very severe respiratory confirmed covid vs. population |
| | | | | | | | | Hospitalized covid vs. population |
| | | | | | | | | Hospitalized covid vs. population, leave out 23 and Me (EUR) |
| rs1886814 | 6:41534945 | A/C | 0.45/0.37 | 3.60E-02 | 1.39 (1.03–1.88) | Intron of *FOXP4-AS1* | HGI | Hospitalized covid vs. population |

In total, 49 unique variants were confirmed to be associated with susceptibility or severity of COVID-19 ($p < 0.05$), corresponding 18 independent loci. It is worth noting that, for the confirmed variants, the phenotypes of the cases and controls were highly consistent between our study and previous reports, and the confirmed variants identified between mild and severe patients were quite different from those identified between patients and general population, suggesting the genetic basis of susceptibility and severity of SARS-CoV-2 infection might be quite different.

## Identification of associated genetic variants in the enhancers of key genes

Previous candidate gene study of genetic association with COVID-19 mainly focused on the exon region of important genes (Zhang et al., 2020a; Novelli et al., 2020). However, regulatory regions, particularly enhancers, play a vital role in gene expression and may affect disease susceptibility and severity when misfunction (Claringbould and Zaugg, 2021). To investigate the role of enhancer variants of key genes in COVID-19, we also included 25 enhancers of 16 IFN-pathway genes and 3 SARS-CoV-2 receptor/co-receptor genes in the probe panel. These enhancers were predicted by GeneHancer database. Genetic association analysis was conducted between mild and severe COVID-19 patients, between all COVID-19 patients and ancestry-matched general population from 1000 Genomes Project, or between severe COVID-19 patients and ancestry-matched general population from 1000 Genomes Project.

In total, we identified 67 variants in the enhancer region of the panel that were associated with susceptibility or severity of COVID-19, relating to 12 enhancers of 11 genes ($p < 0.05$, Supplementary Tables S7–S9). Further annotation by RegulomeDB database filtered out five potential regulatory variants for broad-spectrum antiviral gene *IFITM3* and one for SARS-CoV-2 co-receptor gene *TMPRSS2* (probability score > 0.8, Table 5). Specifically, among the six variants, GTEx revealed two variants affecting the expression of *IFITM3*, that is, T allele of rs11246060 in the enhancer of *IFITM3*, which protected COVID-19 patients from severe

TABLE 5 Significant enhancer variants that were associated with susceptibility and severity of COVID-19 and predicted regulatory by RegulomeDB.

| Enhancer | Gene | SNP | Position | Allele (major/minor) | MAF (case/control) | *p* | OR (95%CI) | Phenotype | RegulomeDB probability |
|---|---|---|---|---|---|---|---|---|---|
| 11:304431-322801 | *IFITM3* | rs11246060 | 11:306877 | C/T | 0.03/0.08 | 2.38E-02 | 0.39 (0.17–0.89) | Severe covid vs. mild covid | 0.9224 |
| 11:304431-322801 | *IFITM3* | rs11246060 | 11:306877 | C/T | 0.05/0.03 | 4.27E-02 | 2.13 (1.04–4.37) | All covid vs. population | 0.9224 |
| 11:304431-322801 | *IFITM3* | rs79196191 | 11:316475 | G/T | 0.07/0.03 | 8.33E-03 | 2.43 (1.24–4.78) | All covid vs. population | 0.99633 |
| 11:304431-322801 | *IFITM3* | rs79196191 | 11:316475 | G/T | 0.08/0.03 | 2.52E-03 | 2.94 (1.44–6.01) | Severe covid vs. population | 0.99633 |
| 11:304431-322801 | *IFITM3* | rs28655829 | 11:320166 | C/T | 0.05/0.01 | 3.44E-02 | 4.82 (1.08–21.45) | Severe covid vs. mild covid | 0.8658 |
| 11:304431-322801 | *IFITM3* | rs1248936100 | 11:321074 | T/TG | 0.03/0.08 | 1.52E-02 | 0.33 (0.14–0.81) | Severe covid vs. mild covid | 0.9975 |
| 11:324001-331788 | *IFITM3* | rs10902125 | 11:329951 | A/G | 0.04/0.01 | 3.42E-02 | 4.46 (0.99–19.98) | Severe covid vs. mild covid | 0.96307 |
| 21:41496486-41511100 | *TMPRSS2* | rs1440999733 | 21:41501641 | C/A | 0/0.02 | 2.47E-02 | 0.00 (0.00-NaN) | Severe covid vs. mild covid | 0.85643 |

outcomes when comparing mild and severe patients ($p$ = 2.38E-2, OR = 0.39, 95% CI = 0.17–0.89), was associated with increased expression of *IFITM3* in whole blood ($p$ = 1.72E-10), while T allele of rs28655829 in the enhancer of *IFITM3*, which increased the risk of severity when comparing mild and severe patients ($p$ = 2.38E-2, OR = 0.39, 95% CI = 0.17–0.89), was associated with decreased expression of *IFITM3* in cultured fibroblasts ($p$ = 7.55E-8). This indicated that these variants might confer genetic risk or protection by affecting the gene expression and highlighted the importance of *IFITM3* gene in the defense of SARS-CoV-2 infection.

## Discussion

In this study, out of 1238 variant previously reported to be association with susceptibility and severity of COVID-19, we confirmed 49 variants, corresponding to 18 independent loci, including 3p21.31 locus, *ABO*, *IFNAR2*, *TNF*, *APOE*, and *FOXP4-AS1* gene.

3p21.31 locus has been identified as a risk factor by GWAS studies of Italian and Spanish (Ellinghaus et al., 2020), British (Wang et al., 2008), Americans (Shelton et al., 2021), and meta-GWAS study of HGI (Initiative, 2021). However, most participants of the studies are Europeans and the major genetic risk factor in 3p21.31 for severe COVID-19 is proposed to be inherited from Neanderthals (Zeberg and Pääbo, 2020), which is almost absent in East Asians. In accordance, previous COVID-19 GWAS study of Chinese population was unable to replicate the locus (Wang et al., 2020; Wu et al., 2021). In our study, consistent with the low frequency of the previously reported lead variant rs11385942 in 3p21.31, only one individual was identified to harbor this variant. This individual had severe COVID-19, which might be due to the risk variant rs11385942. On the other hand, we confirmed another variant in 3p21.31, rs71327056, which is an intergenic variant between *UQCRC2P1* and *CCR2* genes and is not linkage disequilibrium (LD) with rs11385942 (Ellinghaus et al., 2020) (1000 Genomes Project CEU, $r^2$ = 0.19), suggesting there might be more than one independent variant in 3p21.31 contribution to COVID-19 susceptibility and severity. Notably, the minor allele G of rs71317056 was found to increase severity in HGI release4 where most of individuals were Europeans while our study revealed that the minor allele of rs71317056 effected oppositely in Chinese population (Tables 3, 4). We speculated this might due to the interaction with another risk haplotype which Europeans inherited from Neanderthals but was absent in Asians. In addition, rs71317056 is in LD with rs35943069 (1000 Genomes Project CEU, $r^2$ = 1; CHB, $r^2$ = 1), which resides in a potentially enhancer region that is annotated by GeneHancer (Fishilevich et al., 2017). The minor allele is associated with increased *CCR1* gene expression in cultured fibroblasts (GTEx V8, Supplementary Table S10) (Consortium, 2020), suggesting that it might function through expression regulation of *CCR1*, receptor for a C-C type chemokine which play an important role in immune system against viral infection (Zlotnik and Yoshie, 2012).

The *ABO* locus has also been revealed to be associated with COVID-19 severity (Ellinghaus et al., 2020). However, the previously reported lead variant rs657152 had no significant difference in allele frequency between cases and controls of our study. Instead, we confirmed several other variants in the *ABO* locus when comparing all COVID-19 cases with general Chinese population. Specifically, one of the lead variants,

rs34357864 was also been confirmed when comparing severe COVID-19 cases with general Chinese population. Similar to our study, HGI data revealed that this variant was significant when either all COVID-19 patients or hospitalized COVID-19 patients compared with general population, suggesting this is a variant associated with susceptibility.

The T allele of rs9975538 in the intron of gene *IFNAR2* had a lower frequency in severe COVID-19 patients compared with general Chinese population, consistent with results of HGI release 4 comparing either very severe respiratory confirmed COVID-19 patients or hospitalized COVID-19 patients with general population. rs9975538 was also in LD with rs2236757 (1000 Genomes Project CEU, $r^2$ = 0.75; CHB, $r^2$ = 1), the previously reported variant that was associated with critical illness of COVID-19 (Pairo-Castineira et al., 2021). Notably, the T allele of rs9975538 increased the expression of *IFNAR2* and *IL10RB* in lung (GTEx V8, $p$ = 1.09E-5, $p$ = 1.99E-5 respectively). As *IFNAR2* and *IL10RB* are type I and III IFN receptor respectively, this variant might confer protection by increasing IFN receptor expression and thus upregulating the antiviral activity of IFN pathway. In addition, *IFNAR2* play a vital role in multiple sclerosis, a chronic autoimmune disorder characterized by inflammation of the central nervous system, demyelination and axonal damage (Gilli et al., 2008; Órpez-Zafra et al., 2017). This leads to a hypothesis that COVID-19 patients harboring *IFNAR2* rs9975538 variant might be more likely to develop neurological disorders (Douaud et al., 2022; Lee et al., 2022), possibly though neuroinflammatory pathways.

Rs1799964, which is located upstream of *TNF* gene, was found to be associated with disease severity, with C allele being associated with mild phenotype. TNF is multifunctional proinflammatory cytokine and plays a key role in regulating the immunological response to infections (Waters et al., 2013). The C allele of rs1799964 was associated with increased expression of *TNF* than T allele (Nourian et al., 2017) and increased lymphocyte counts (Chen et al., 2020), which might protect the body from severe disease.

*APOE* gene polymorphisms have been reported to be associated with susceptibility or severity of COVID-19 in British, Czech, Spanish, Finnish and Kurdish population (Kuo et al., 2020; Al-Jaf et al., 2021; Del Ser et al., 2021; Hubacek et al., 2021; Kurki et al., 2021). Consistent with above findings, we revealed an association of *APOE* variant rs429358 with susceptibility to COVID-19 in Chinese population. Given that *APOE* is associated with Alzheimer's and cardiovascular diseases and type 2 diabetes (Liu et al., 2013; Mahley, 2016; Liu et al., 2019), comorbidities that are related to COVID-19 susceptibility and severity, the effect of the *APOE* variant on COVID-19 could be indirect. Meanwhile, recent researches indicated that APOE might also affect SARS-CoV-2 infection directly by interacting with ACE2 inhibiting SARS-CoV-2 cellular entry (Zhang et al., 2022), regulating cellular cholesterol homeostasis (Gao et al., 2022) and modulating antiviral immunity (Ostendorf et al., 2022). Notably, another variant that determines APOE isoforms, rs7412, did not pass the significant threshold, probably because this variant mainly contributes to APOE ε2 isoform while COVID-19 is more associated with APOE ε4 isoform (Kuo et al., 2020; Al-Jaf et al., 2021; Del Ser et al., 2021; Hubacek et al., 2021; Kurki et al., 2021).

Rs1886814 in the intron of the lncRNA *FOXP4-AS1* was found to be associated with disease susceptibility when comparing all COVID-19 patients or severe COVID-19 patients with general Chinese population. In HGI release 4 datasets, it is associated with COVID-19 hospitalization when comparing with general population. Recent trans-ethnic genome-wide association study of severe COVID-19 that incorporated Chinese population and HGI results revealed another significant variant in the intron of *FOXP4-AS1*, rs1853837, which is LD with rs1886814 in Chinese population (1000 Genomes Project CHB, $r^2$ = 0.68) (Wu et al., 2021). The risk allele C of rs1886814 is an eQTL in positive association with the expression of FOXP4 in lung (GETx V8, $p$ = 3.28E-6) (Consortium, 2020). FOXP4 is a transcription factor expressed in both thymocytes and peripheral $CD4^+$ and $CD8^+$ T cells, and is necessary for normal T cell cytokine recall responses to antigen following pathogenic infection (Wiehagen et al., 2012).

We noted that the overall confirmation rate was not high, possibly due to different population structure and limited sample size of our study. The statistic power was provided in Supplementary Figure S3. On the other hand, it is worth noting that for the validated variants, the phenotypes of cases and controls were highly coordinated in our study and original study. All variants validated in our mild and severe group were specific to be identified in previous association study of severity, that is, when comparing hospitalized or very severe respiratory confirmed COVID-19 patients with not hospitalized ones. Nearly all variants validated in the COVID-19 patients and general Chinese population group were specific to be identified in previous association study of susceptibility, that is, comparing COVID-19 patients with general population. The remarkable specificity suggested that susceptibility and severity might have different genetic basis and also indicated the accuracy of our study.

In addition, we also identified 67 variants in the 12 regulatory regions of 11 candidate genes associated with susceptibility or severity of COVID-19, which have not been reported before. Further annotation by RegulomeDB and GTEx database revealed two variants affected the expression of *IFITM3* and conferred genetic risk and protection respectively. *IFITM3* gene encodes a transmembrane protein that could be induced by interferons and function as a broad-spectrum antiviral effector molecule by directly limiting cellular entry of a number of pathogenic viruses, including influenza A virus, West Nile virus, dengue virus, SARS-CoV and SARS-CoV-2 (Diamond and Farzan, 2013; Shi et al., 2021). Moreover, rs12252 variant in the gene has been found to be associated with COVID-19 severity (Zhang et al., 2020b; Alghamdi et al., 2021; Gómez et al., 2021). Our results

indicated that, in addition to genetic variants, enhancer variants of *IFITM3* might confer genetic risk or protection by affecting gene expression as well. Though larger cohort studies are needed to confirm these genetic associations, our data presented here also highlighted the important role of *IFITM3* in host defense against SARS-CoV-2.

In conclusion, we confirmed a list of previously reported variants associated with susceptibility and severity of COVID-19, and identified several enhancer variants potentially regulating expression of genes associated with COVID-19. Though larger cohort studies and further experiments are needed to confirm these genetic associations and explore the molecular mechanism, elucidation of host genetic factors contributing to susceptibility to severe infection will provide the opportunity for clinical risk profiling of patients with COVID-19, mechanistic understanding of the underlying pathophysiology and further identification of potential therapeutic targets.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ Supplementary Material.

## Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Huoshenshan Hospital (HSSLL036). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## Author contributions

Conceptualization: ZZ, YZ, and CW; Methodology: YZ and PL; Formal analysis: PL and WS; Resources: CW, YK; Data Curation: SS, YW, KL, and XG; Writing—original draft: PL; Writing—review and editing: PL and all; Supervision: ZZ, YZ, and CW; Project administration: YZ and YK; Funding acquisition: PL and YZ.

## Funding

This work was supported by grants from the National Key Research and Development Program of China (2018YFC1200704, 2018YFA0900801) and the National Natural Science Foundation of China (31801087).

## Acknowledgments

We thank all the patients participating in this study. We thank the COVID-19 Host Genetic Initiative for sharing the HGI release4 summary statistics, and the 1000 Genomes Project for sharing the human genetic variation data.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1073880/full#supplementary-material

## References

Al-Jaf, S. M. A., Niranji, S. S., Ali, H. N., and Mohammed, O. A. (2021). Association of Apolipoprotein e polymorphism with SARS-CoV-2 infection. *Infect. Genet. Evol.* 95, 105043. doi:10.1016/j.meegid.2021.105043

Alghamdi, J., Alaamery, M., Barhoumi, T., Rashid, M., Alajmi, H., Aljasser, N., et al. (2021). Interferon-induced transmembrane protein-3 genetic variant rs12252 is associated with COVID-19 mortality. *Genomics* 113 (4), 1733–1741. doi:10.1016/j.ygeno.2021.04.002

Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., et al. (2015). A global reference for human genetic variation. *Nature* 526 (7571), 68–74. doi:10.1038/nature15393

Bastard, P., Rosen, L. B., Zhang, Q., Michailidis, E., Hoffmann, H-H., Zhang, Y., et al. (2020). Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* 370 (6515), eabd4585. doi:10.1126/science.abd4585

Benetti, E., Tita, R., Spiga, O., Ciolfi, A., Birolo, G., Bruselles, A., et al. (2020). ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. *Eur. J. Hum. Genet.* 28 (11), 1602–1614. doi:10.1038/s41431-020-0691-z

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi:10.1093/bioinformatics/btu170

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. doi:10.1186/s13742-015-0047-8

Chen, M-H., Raffield, L. M., Mousas, A., Sakaue, S., Huffman, J. E., Moscati, A., et al. (2020). Trans-ethnic and ancestry-specific blood-cell genetics in 746, 667 individuals from 5 global populations. *Cell* 182 (5), 1198–1213. doi:10.1016/j.cell.2020.06.045

Ching, J. C-Y., Chan, K. Y. K., Lee, E. H. L., Xu, M-S., Ting, C. K. P., So, T. M. K., et al. (2010). Significance of the myxovirus resistance A (MxA) gene -123C>a single-nucleotide polymorphism in suppressed interferon beta induction of severe acute respiratory syndrome coronavirus infection. *J. Infect. Dis.* 201 (12), 1899–1908. doi:10.1086/652799

Claringbould, A., and Zaugg, J. B. (2021). Enhancers in disease: Molecular basis and emerging treatment strategies. *Trends Mol. Med.* 27 (11), 1060–1073. doi:10.1016/j.molmed.2021.07.012

Consortium, G. T. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369 (6509), 1318–1330. doi:10.1126/science.aaz1776

Del Ser, T., Fernández-Blázquez, M. A., Valentí, M., Zea-Sevilla, M. A., Frades, B., Alfayate, E., et al. (2021). Residence, clinical features, and genetic risk factors associated with symptoms of COVID-19 in a cohort of older people in madrid. *Gerontology* 67 (3), 281–289. doi:10.1159/000513182

Di Maria, E., Latini, A., Borgiani, P., and Novelli, G. (2020). Genetic variants of the human host influencing the coronavirus-associated phenotypes (SARS, MERS and COVID-19): Rapid systematic review and field synopsis. *Hum. Genomics* 14 (1), 30. doi:10.1186/s40246-020-00280-6

Diamond, M. S., and Farzan, M. (2013). The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nat. Rev. Immunol.* 13 (1), 46–57. doi:10.1038/nri3344

Douaud, G., Lee, S., Alfaro-Almagro, F., Arthofer, C., Wang, C., McCarthy, P., et al. (2022). SARS-CoV-2 is associated with changes in brain structure in UK Biobank. *Nature* 604 (7907), 697–707. doi:10.1038/s41586-022-04569-5

Downes, D. J., Cross, A. R., Hua, P., Roberts, N., Schwessinger, R., Cutler, A. J., et al. (2021). Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. *Nat. Genet.* 53 (11), 1606–1615. doi:10.1038/s41588-021-00955-3

Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., et al. (2020). Genomewide association study of severe covid-19 with respiratory failure. *N. Engl. J. Med.* 383 (16), 1522–1534. doi:10.1056/NEJMoa2020283

Faul, F., Erdfelder, E., Lang, A-G., and Buchner, A. G. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39 (2), 175–191. doi:10.3758/bf03193146

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., et al. (2017). *GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards*. Oxford): Database.

Gao, P., Ji, M., Liu, X., Chen, X., Liu, H., Li, S., et al. (2022). Apolipoprotein E mediates cell resistance to influenza virus infection. *Sci. Adv.* 8 (38), eabm6668. doi:10.1126/sciadv.abm6668

Gilli, F., Valentino, P., Caldano, M., Granieri, L., Capobianco, M., Malucchi, S., et al. (2008). Expression and regulation of IFNalpha/beta receptor in IFNbeta-treated patients with multiple sclerosis. *Neurology* 71 (24), 1940–1947. doi:10.1212/01.wnl.0000327340.50284.8d

Gómez, J., Albaiceta, G. M., Cuesta-Llavona, E., García-Clemente, M., López-Larrea, C., Amado-Rodríguez, L., et al. (2021). The Interferon-induced transmembrane protein 3 gene (IFITM3) rs12252 C variant is associated with COVID-19. *Cytokine* 137, 155354. doi:10.1016/j.cyto.2020.155354

Hamano, E., Hijikata, M., Itoyama, S., Quy, T., Phi, N. C., Long, H. T., et al. (2005). Polymorphisms of interferon-inducible genes OAS-1 and MxA associated with SARS in the Vietnamese population. *Biochem. Biophys. Res. Commun.* 329 (4), 1234–1239. doi:10.1016/j.bbrc.2005.02.101

He, J., Feng, D., de Vlas, S. J., Wang, H., Fontanet, A., Zhang, P., et al. (2006). Association of SARS susceptibility with single nucleic acid polymorphisms of OAS1 and MxA genes: A case-control study. *BMC Infect. Dis.* 6, 106. doi:10.1186/1471-2334-6-106

Hoffmann, M., Kleine-Weber, H., and Pöhlmann, S. (2020). A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol. Cell* 78 (4), 779–784. doi:10.1016/j.molcel.2020.04.022

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181 (2), 271–280. doi:10.1016/j.cell.2020.02.052

Hubacek, J. A., Dlouha, L., Dusek, L., Majek, O., and Adamkova, V. (2021). Apolipoprotein E4 allele in subjects with COVID-19. *Gerontology* 67 (3), 320–322. doi:10.1159/000516200

Initiative, C-H. G. (2021). Mapping the human genetic architecture of COVID-19. *Nature* 600 (7889), 472–477.

Initiative, C-H. G. (2020). The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet.* 28 (6), 715–718. doi:10.1038/s41431-020-0636-6

Kuo, C-L., Pilling, L. C., Atkins, J. L., Masoli, J. A. H., Delgado, J., Kuchel, G. A., et al. (2020). APOE e4 Genotype Predicts Severe COVID-19 in the UK Biobank Community Cohort. *J. Gerontol. A Biol. Sci. Med. Sci.* 75 (11), 2231–2232. doi:10.1093/gerona/glaa131

Kurki, S. N., Kantonen, J., Kaivola, K., Hokkanen, L., Mäyränpää, M. I., Puttonen, H., et al. (2021). APOE ε4 associates with increased risk of severe COVID-19, cerebral microhaemorrhages and post-COVID mental fatigue: A Finnish biobank, autopsy and clinical study. *Acta Neuropathol. Commun.* 9 (1), 199. doi:10.1186/s40478-021-01302-7

Latini, A., Agolini, E., Novelli, A., Borgiani, P., Giannini, R., Gravina, P., et al. (2020). COVID-19 and genetic variants of protein involved in the SARS-CoV-2 entry into the host cells. *Genes* 11 (9), E1010. doi:10.3390/genes11091010

Lee, M. H., Perl, D. P., Steiner, J., Pasternack, N., Li, W., Maric, D., et al. (2022). Neurovascular injury with complement activation and inflammation in COVID-19. *Brain.* 145 (7), 2555–2568. doi:10.1093/brain/awac151

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Li, P., Shi, M-L., Shen, W-L., Zhang, Z., Xie, D-J., Zhang, X-Y., et al. (2017). Coordinated regulation of IFITM1, 2 and 3 genes by an IFN-responsive enhancer through long-range chromatin interactions. *Biochim. Biophys. Acta. Gene Regul. Mech.* 1860 (8), 885–893. doi:10.1016/j.bbagrm.2017.05.003

Liu, C-C., Liu, C-C., Kanekiyo, T., Xu, H., and Bu, G. (2013). Apolipoprotein E and alzheimer disease: Risk, mechanisms and therapy. *Nat. Rev. Neurol.* 9 (2), 106–118. doi:10.1038/nrneurol.2012.263

Liu, S., Liu, J., Weng, R., Gu, X., and Zhong, Z. (2019). Apolipoprotein E gene polymorphism and the risk of cardiovascular disease and type 2 diabetes. *BMC Cardiovasc. Disord.* 19 (1), 213. doi:10.1186/s12872-019-1194-0

Mahley, R. W. (2016). Apolipoprotein E: From cardiovascular disease to neurodegenerative disorders. *J. Mol. Med.* 94 (7), 739–746. doi:10.1007/s00109-016-1427-y

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20 (9), 1297–1303. doi:10.1101/gr.107524.110

Nourian, M., Chaleshi, V., Pishkar, L., Azimzadeh, P., Baradaran Ghavami, S., Balaii, H., et al. (2017). Evaluation of tumor necrosis factor (TNF)-α mRNA expression level and the rs1799964 polymorphism of the TNF-α gene in peripheral mononuclear cells of patients with inflammatory bowel diseases. *Biomed. Rep.* 6 (6), 698–702. doi:10.3892/br.2017.908

Novelli, A., Biancolella, M., Borgiani, P., Cocciadiferro, D., Colona, V. L., D'Apice, M. R., et al. (2020). Analysis of ACE2 genetic variants in 131 Italian SARS-CoV-2-positive patients. *Hum. Genomics* 14 (1), 29. doi:10.1186/s40246-020-00279-z

Órpez-Zafra, T., Pavía, J., Hurtado-Guerrero, I., Pinto-Medel, M. J., Rodriguez Bada, J. L., Urbaneja, P., et al. (2017). Decreased soluble IFN-β receptor (sIFNAR2) in multiple sclerosis patients: A potential serum diagnostic biomarker. *Mult. Scler.* 23 (7), 937–945. doi:10.1177/1352458516667564

Ostendorf, B. N., Patel, M. A., Bilanovic, J., Hoffmann, H. H., Carrasco, S. E., Rice, C. M., et al. (2022). Common human genetic variants of APOE impact murine COVID-19 mortality. *Nature* 611, 346–351. doi:10.1038/s41586-022-05344-2

Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., Pasko, D., et al. (2021). Genetic mechanisms of critical illness in COVID-19. *Nature* 591 (7848), 92–98. doi:10.1038/s41586-020-03065-y

Richardson, S., Hirsch, J. S., Narasimhan, M., Crawford, J. M., McGinn, T., Davidson, K. W., et al. (2020). Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York city area. *JAMA* 323 (20), 2052–2059. doi:10.1001/jama.2020.6775

Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., et al. (2020). Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* 117 (21), 11727–11734. doi:10.1073/pnas.2003138117

Shelton, J. F., Shastri, A. J., Ye, C., Weldon, C. H., Filshtein-Sonmez, T., Coker, D., et al. (2021). Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat. Genet.* 53 (6), 801–808. doi:10.1038/s41588-021-00854-7

Shi, G., Kenney, A. D., Kudryashova, E., Zani, A., Zhang, L., Lai, K. K., et al. (2021). Opposing activities of IFITM proteins in SARS-CoV-2 infection. *EMBO J.* 40 (3), e106501. doi:10.15252/embj.2020106501

Wang, F., Huang, S., Gao, R., Zhou, Y., Lai, C., Li, Z., et al. (2020). Initial whole-genome sequencing and analysis of the host genetic contribution to COVID-19 severity and susceptibility. *Cell Discov.* 6 (1), 83. doi:10.1038/s41421-020-00231-4

Wang, S., Wei, M., Han, Y., Zhang, K., He, L., Yang, Z., et al. (2008). Roles of TNF-alpha gene polymorphisms in the occurrence and progress of SARS-cov infection: A case-control study. *BMC Infect. Dis.* 8, 27. doi:10.1186/1471-2334-8-27

Waters, J. P., Pober, J. S., and Bradley, J. R. (2013). Tumour necrosis factor in infectious disease. *J. Pathol.* 230 (2), 132–147. doi:10.1002/path.4187

Wiehagen, K. R., Corbo-Rodgers, E., Li, S., Staub, E. S., Hunter, C. A., Morrisey, E. E., et al. (2012). Foxp4 is dispensable for T cell development, but required for robust recall responses. *PloS one* 7 (8), e42273. doi:10.1371/journal.pone.0042273

Wu, P., Ding, L., Li, X., Liu, S., Cheng, F., He, Q., et al. (2021). Trans-ethnic genome-wide association study of severe COVID-19. *Commun. Biol.* 4 (1), 1034. doi:10.1038/s42003-021-02549-5

Wu, Z., and McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72 314 cases from the Chinese center for disease control and prevention. *JAMA* 323 (13), 1239–1242. doi:10.1001/jama.2020.2648

Zeberg, H., and Pääbo, S. (2020). The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* 587 (7835), 610–612. doi:10.1038/s41586-020-2818-3

Zhang, H., Shao, L., Lin, Z., Long, Q-X., Yuan, H., Cai, L., et al. (2022). APOE interacts with ACE2 inhibiting SARS-CoV-2 cellular entry and inflammation in COVID-19 patients. *Signal Transduct. Target. Ther.* 7 (1), 261. doi:10.1038/s41392-022-01118-4

Zhang, Q., Bastard, P., Liu, Z., Le Pen, J., Moncada-Velez, M., Chen, J., et al. (2020)., 370. New York, NY), eabd4570. doi:10.1126/science.abd4570Inborn errors of type I IFN immunity in patients with life-threatening COVID-19*Science*6515

Zhang, Y., Qin, L., Zhao, Y., Zhang, P., Xu, B., Li, K., et al. (2020). Interferon-induced transmembrane protein 3 genetic variant rs12252-C associated with disease severity in coronavirus disease 2019. *J. Infect. Dis.* 222 (1), 34–37. doi:10.1093/infdis/jiaa224

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in wuhan, China: A retrospective cohort study. *Lancet* 395 (10229), 1054–1062. doi:10.1016/S0140-6736(20)30566-3

Zhou, P., Yang, X-L., Wang, X-G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579 (7798), 270–273. doi:10.1038/s41586-020-2012-7

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382 (8), 727–733. doi:10.1056/NEJMoa2001017

Zlotnik, A., and Yoshie, O. (2012). The chemokine superfamily revisited. *Immunity* 36 (5), 705–716. doi:10.1016/j.immuni.2012.05.008

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1041470



OPEN ACCESS

EDITED BY
Maritha J. Kotze,
Tygerberg Hospital, South Africa

REVIEWED BY
Yawei Li,
Northwestern University, United States
Bin Guo,
Merck, United States
Kelebogile Moremi,
Stellenbosch University, South Africa

*CORRESPONDENCE
Ke Zhang,
kezhang@gmc.edu.cn
Zhong-Shan Cheng,
cheng.zhong.shan@gmail.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 10 September 2022
ACCEPTED 01 December 2022
PUBLISHED 04 January 2023

CITATION
Luo Y-S, Shen X-C, Li W, Wu G-F, Yang X-M, Guo M-Y, Chen F, Shen H-Y, Zhang P-P, Gao H, Nie Y, Wu J-H, Mou R, Zhang K and Cheng Z-S (2023), Genetic screening for hypertension and COVID-19 reveals functional variation of *SPEG* potentially associated with severe COVID-19 in women. *Front. Genet.* 13:1041470. doi: 10.3389/fgene.2022.1041470



# Genetic screening for hypertension and COVID-19 reveals functional variation of *SPEG* potentially associated with severe COVID-19 in women

Yu-Si Luo[1,2†], Xiang-Chun Shen[3,4†], Wei Li[5], Guo-Feng Wu[1], Xiao-Meng Yang[2], Ming-Yang Guo[2], Fang Chen[2], Hu-Yan Shen[2], Ping-Ping Zhang[2], Han Gao[1], Ying Nie[2], Jia-Hong Wu[2], Rong Mou[2], Ke Zhang[2,3]* and Zhong-Shan Cheng[6]*

[1]Department of Emergency, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, [2]The Key and Characteristic Laboratory of Modern Pathogenicity Biology, School of Basic Medical Sciences, Guizhou Medical University, Guiyang, China, [3]The High Efficacy Application of Natural Medicinal Resources Engineering Center of Guizhou Province, School of Pharmaceutical Sciences, Guizhou Medical University, Guiyang, China, [4]State Key Laboratory for Functions and Applications of Medicinal Plants, Guizhou Medical University, Guiyang, China, [5]Department of Cardiovascular, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, [6]Center for Applied Bioinformatics, St. Jude Children's Research Hospital, Memphis, TN, United States

The coronavirus disease 2019 (COVID-19) pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has led to more than 6.4 million deaths worldwide. The prevalent comorbidity between hypertension and severe COVID-19 suggests common genetic factors may affect the outcome of both diseases. As both hypertension and severe COVID-19 demonstrate sex-biased prevalence, common genetic factors between the two diseases may display sex-biased differential associations. By evaluating COVID-19 association signals of 172-candidate hypertension single nucleotide polymorphisms (SNPs) derived from more than 1 million European individuals in two sex-stratified severe COVID-19 genome-wide association studies from UK BioBank with European ancestry, we revealed one functional cis expression quantitative trait locus of *SPEG* (rs12474050) showing sex-biased association with severe COVID-19 in women. The risk allele rs12474050*T associates with higher blood pressure. In our study, we found it is significantly correlated with lower *SPEG* expression in muscle-skeletal but with higher expression in both brain cerebellum and cerebellar hemisphere. Additionally, nominal significances were detected for the association between rs12474050*T and lower *SPEG* expression in both heart left ventricle and atrial appendage; among these tissues, the *SPEG* expression is nominally significantly higher in females than in males. Further analysis revealed *SPEG* is mainly expressed in cardiomyocytes in heart and is upregulated upon SARS-CoV-2 infection, with significantly higher upregulation of *SPEG* only observed in female but not in male COVID-19 patients compared to both normal female and male individuals, suggesting upregulation of *SPEG* is a female-specific protective

mechanism against COVID-19 induced heart damage. Taken together, our analyses suggest the involvement of *SPEG* in both hypertension and severe COVID-19 in women, which provides new insights for sex-biased effect of severe COVID-19 in women.

KEYWORDS

**hypertension, severe COVID-19, GWAS, *SPEG*, women, cardiomyocyte**

## Introduction

The coronavirus disease 2019 (COVID-19) pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has been quickly spreading in more than 200 countries and overwhelmingly challenging the global population. Since November 2019, there are 591,683,619 cases have been reported and 6,443,306 deaths have been documented (WHO, 2022). The overall mortality rate of COVID-19 is approximately 1.09% at the time of writing.

Numerous studies, including observational and retrospective research, had demonstrated that hypertension was the most frequent comorbidity of COVID-19 patients and was reported as a common and independent risk factor for the severity and mortality of patients suffered from COVID-19 (Chen et al., 2021; Dai et al., 2022). Typically, hypertension is measured based on diastolic blood pressure (DBP) and systolic blood pressure (SBP), and patients with DBP ≥ 90 mmHg or SBP ≥ 140 mmHg are diagnosed with hypertension (Chen et al., 2021). A recent study suggested hypertension contributed 2.5-fold increased risk of disease severity and mortality in COVID-19 patients (Lippi et al., 2020). Nevertheless, SARS-CoV-2 infection is also able to induce hypertension post COVID-19, named as Post-COVID-19 hypertension and highlighted by two studies from Turkey research groups (Akpek, 2022; Uysal et al., 2022). According to the data from the retrospective cohort study presented by Akpek et al., the DBP and SBP of 153 eligible COVID-19 patients (mean age 46.5 ± 12.7 years) were statistically higher in the post COVID-19 period than on admission (Akpek, 2022). The other prospective multicenter study reported the DBP and SBP of 100 eligible COVID-19 young patients (mean age 189.45 ± 23.78 months) were remarkably higher in the post COVID-19 period than on admission (Uysal et al., 2022). The potential cause of Post-COVID-19 hypertension by SARS-CoV-2 infection may be due to the shift of renin-angiotensin system (RAS) related balance from Mas [angiotensin converting enzyme 2 (ACE2)/angiotensin (Ang) 1-7/Mas] axis to RAS [Ang converting enzyme (ACE)/Ang II/Ang II type I receptor (AT1R)] axis (Uysal et al., 2022) upon SARS-CoV-2 infection. A new study further revealed that severe COVID-19 induced autoantibodies against Ang II is correlated with blood pressure dysregulation and COVID-19 severity (Briquez et al., 2022). Additionally, infection of SARS-CoV-2 might also lead to aggravation of pre-existing hypertension and damages of other organs, such as heart, which consequently contributed to more severe clinical consequences (Wang et al., 2020).

In order to explore potentially sex-biased genetic factors influencing both COVID-19 severity and hypertension, we performed an integrative screening of known hypertension association single nucleotide polymorphisms (SNPs) that are also cis-expression quantitative trait loci (cis-eQTLs) in severe COVID-19 GWASs stratified by sex. Our analyses reveal that only one cis-eQTL, rs12474050, mapped to the gene striated preferentially expressed protein kinase (*SPEG*) encoding the protein SPEG, shown significant, sex-biased association with severe COVID-19 in women. Further scientific literature reviewing underpinned that *SPEG* played critical roles in the development, maintenance, and function of cardiac and skeletal muscles (Campbell et al., 2021; Luo et al., 2021), and the expression of *SPEG* is suggestively higher in muscle skeletal, heart atrial appendage, and heart left ventricle in females than in males. We thus investigated the potential link between sex-biased expression of *SPEG* and COVID-19 severity in females.

## Materials and methods

### Selection of hypertension SNPs that are cis-eQTLs in GTEx

According to Evangelou et al. (Evangelou et al., 2018), hypertension association SNPs ($n$ = 172 lead SNPs with independent genome-wide significant association signals) that were also cis-eQTLs in GTEx database (GTEx, 2020) were obtained. These SNPs were derived from a hypertension GWAS conducted over 1 million European individuals. Two hypertension traits, including DBP and SBP were studied by the hypertension GWAS. Among these 172 cis-eQTLs, the top five tissues harboring the largest number of these cis-eQTLs include adipose subcutaneous ($n$ = 43), artery tibial ($n$ = 41), artery aorta ($n$ = 39), nerve tibial (n = 39), muscle skeletal ($n$ = 38). We referred to these 172-candidate hypertension association SNPs as hypertension cis-eQTLs.

### Evaluation of hypertension cis-eQTLs in severe COVID-19 GWASs in European populations

To screen potential functional SNPs associated with both hypertension and severe COVID-19, we evaluated the

association signals of these candidate hypertension cis-eQTLs that were reported as independent genome-wide significant SNPs in the hypertension GWAS in the current severe COVID-19 GWASs of European ancestry that were performed separately by sex, the summary statistics of which were freely available from GRASP database (Thibord et al., 2022). Details for the two GWASs are as follows: the first GWAS, severe COVID-19 positive vs. non-severe COVID-19 positive in European (EUR) Female UK BioBank (UKBB) tested positive samples, has 283 cases and 7,113 controls; the second GWAS, severe COVID-19 positive vs. non-severe COVID-19 positive in EUR Male UKBB tested positive samples, has 565 cases and 6,093 controls. The rationale to focus on severe COVID-19 by sex is due to the observation that sex is a risk factor for both hypertension and severe COVID-19. Furthermore, due to the nature that these 172 cis-eQTLs associated with hypertension were derived from European samples, we therefore only focused on severe COVID-19 GWASs conducted on European samples.

We conducted sex-biased differential association analysis for these 172 hypertension cis-eQTLs extracted from the sex-stratified severe COVID-19 GWASs. Sex-biased differential association analysis was performed with Z-test according to Thibord et al. (Thibord et al., 2022). See the below formula for calculating delta Z-score (ΔZ-score) and its corresponding $p$-value. Bonferroni correction was used to adjust the significance of differential association between sex, with the threshold set at $p < 0.05/172 = 3e\text{-}4$. Only one cis-eQTL, rs12474050, mapped to *SPEG*, passed the statistical significance threshold, with its ΔZ-score equal to 3.75 and differential Z-score $p$-value equal to 1.8e-4. rs12474050 was also a top SNP showing suggestive association with severe COVID-19 in females ($p$ = 1.78e-4; beta = 0.36; se = 0.10). In the corresponding severe COVID-19 GWAS of males, the cis-eQTL was not significant ($p$ = 0.23; beta = −0.08; se = 0.07).

$$\Delta Z - score = \frac{female.snp.\beta - male.snp.\beta}{\sqrt{(female.snp.se^2 + male.snp.se^2)}}$$

$$p = pnorm(-|\Delta Z - score|)*2$$

## Functional annotation for *SPEG* cis-eQTL rs12474050

As rs12474050 was reported as a cis-eQTL of *SPEG* in GTEx database V7 by the hypertension GWAS, we further evaluated the association between rs12474050 and *SPEG* expression in GTEx database V8 (GTEx, 2020). In addition, we also investigated *SPEG* expression among 49 GTEx tissues by sex and inferred tissue specific biological functions of *SPEG* by determining whether *SPEG* was highly expressed in a specific tissue.

## Expression analysis of *SPEG* in multiple GTEx tissues

Similar to our previous analysis (Luo et al., 2022), bulk RNAseq transcript per million (TPM) matrix data of 49 tissues, as well as sex information of these samples from GTEx database V8 (GTEx, 2020), were downloaded from GTEx Portal. By matching samples with sex information, as well as tissue information, we determined differential expression of *SPEG* between males and females across multiple GTEx tissues with ANOVA test implemented in the procedure proc GLM using SAS OnDemand for Academics.

We also evaluated *SPEG* expression among different single cells from eight tissues, including breast mammary tissue, esophagus mucosa, esophagus muscularis, heart left ventricle, lung, muscle skeletal, prostate, skin sun exposed lower leg, using the visualization tool 'GTEx Multi-Gene Single Cell Query' from GTEx Portal.

## Evaluation of *SPEG* expression upon SARS-CoV-2 infection

To determine the expression of *SPEG* upon SARS-CoV-2 infection, we firstly re-analyzed publicly available single cell expression data that were related to COVID-19 and stored in the UCSC Cell Browser (Speir et al., 2021). By querying the keyword "COVID" for the option of disease type, we obtained 13 COVID-19 related single cell data sets from UCSC Cell Browser (accessed on 8 August 2022; see Supplementary Material for these 13 COVID-19 single cell data sets). After evaluating *SPEG* expression across the above 13 COVID-19 single cell expression data sets in UCSC Cell Browser, we revealed that *SPEG* is highly expressed in cardiomyocyte cells in heart tissue based on single cell data from the dataset "Cellular Targets of SARS-CoV-2" (Melms et al., 2021). Thus, we selected the heart single cell data set "Cellular Targets of SARS-CoV-2" for deep investigation of *SPEG* expression. The expression data matrix for the heart tissue single cell data set, as well as its sample meta data and Uniform Manifold Approximation and Projection (UMAP) for all single cells, are provided by UCSC Cell Browser.

Since our aim is to evaluate whether *SPEG* expression shows sex-biased differential expression between females and males for both COVID-19 patients and healthy controls, we first performed sample level differential expression analysis for the data set according to Trump et al. (Trump et al., 2021). We linked the normalized read count matrix and clinical information of these COVID-19 samples, such as sex and COVID-19 status, as well as single cell UMAP coordinates, for further analyses using SAS OnDemand for Academics. Differential expression of *SPEG* between healthy controls and COVID-19 patients by sex were conducted for each single cell type at sample level. In total, there were four groups, including COVID-19 females ($n$ = 6), COVID-

19 males ($n$ = 13), healthy females ($n$ = 12), and healthy males ($n$ = 16), subjected to analysis. Samples without specific single cell types were not included in the final differential expression or other comparisons, such as comparisons of cell frequency and cell percentage. We used median log2 (expression of *SPEG* + 1) in each sample for different single cell types and subsequently performed differential gene expression analysis for *SPEG*. We applied the option lsmeans from SAS proc GLM procedure to perform multiple comparisons and the $p$ values for multiple comparisons among different groups in each single cell type were adjusted with Tukey adjustment. The significance threshold was set at adjusted $p \leq 0.05$. In terms of cell counts and cell percentage for each single cell type from each sample, as well as cell counts and cell percentage of *SPEG* expressed cells in each sample, we compared these parameters across the above four groups using the same SAS procedure and used the same adjusted $p$-value threshold to determine statistical significance.

Alternatively, we also conducted single cell level differential expression analysis for *SPEG* for each single cell type by aggregating all cells of specific single cell type together among the four patient groups, including COVID-19 females, COVID-19 males, healthy females, and healthy males. This analysis was performed using R package Seurat (Hao et al., 2021) (see details of analysis procedures in Supplementary Material). Briefly, normalization was performed to obtain relative gene expression abundances between cells by scaling count data with default setting in Seurat. Differential gene expression analysis of *SPEG* was conducted among each cell cluster defined by (Melms et al., 2021) using the default Wilcoxon Rank Sum test in Seurat. As only two cell types display *SPEG* expression, including cardiomyocyte and vascular smooth muscle cell (VSMC), differential expression analysis for *SPEG* was only conducted among the two cell types, with multiple adjusted $p$-value threshold set at $p < 0.05/(2 \times 6) = 4e\text{-}3$ to determine significance. Additionally, the average log2 (fold change) of *SPEG* expression and the percent of cells expressing *SPEG* between two compared patient groups were generated by Seurat.

We also evaluated *SPEG* expression in cardiomyocytes upon SARS-CoV-2 infection in a bulk RNAseq dataset downloaded from NCBI gene expression omnibus (GEO) (accession number: GSE156754) (Perez-Bermejo et al., 2021). This published data set evaluated genes responding to SARS-CoV-2 infection among multiple cardiac cells, including human induced pluripotent stem cell-derived cardiomyocytes, cardiac fibroblasts, and endothelial cells. We obtained the normalized read counts of *SPEG* and compared *SPEG* expression between SARS-CoV-2 infected cardiac cells and its corresponding mock controls, differential gene expression analyses were performed using Student's t-test with the software Prism GraphPad. Multiple adjustment $p$-value threshold was set at $p \leq 0.008$ (0.05/6).

## Results

The frequent observation of comorbidity between hypertension and severe COVID-19 promoted us to search for genetic factors associated with both diseases. Based on a previously published large-scale hypertension GWAS conducted over 1 million European individuals (Evangelou et al., 2018), we selected 172 independent hypertension association SNPs that were also cis-eQTLs for evaluation in two severe COVID-19 GWASs conducted separately by sex from UK BioBank with European ancestry. In the sex-biased differential association analysis, only one SNP out of 172-candidate SNPs passed the multiple adjusted association $p$-value threshold ($p < 3e\text{-}4$). The SNP rs12474050 (sex-biased differential association $p$ = 1.8e-4) is a cis-eQTL of *SPEG* and emerged with suggestive association with severe COVID-19 in women ($p$ = 1.8e-4; beta = 0.36; se = 0.10) but not in men ($p$ = 0.23; beta = −0.08; se = 0.07) (Figure 1). We thus evaluated local COVID-19 association signals in a 1-Mbp window where rs12474050 located at the center in the two severe COVID-19 GWASs by sex. The local Manhattan plot by sex in Figure 1 demonstrated that rs12474050 was one of the top COVID-19 association signals in the female COVID-19 GWAS, with more SNPs showing suggestive associations (p < 1e-3) in females than in males, suggesting rs12474050 is a sex-biased genetic factor showing association with COVID-19 in females. Collectively, our integrative genetic screening of both hypertension and severe COVID-19 among European samples revealed rs12474050 is a sex-biased marker associated with severe COVID-19 in women.

As rs12474050 was reported as a cis-eQTL of *SPEG* by the previously published hypertension GWAS (Evangelou et al., 2018), in which the GTEx database V7 was used to annotate these cis-eQTLs, we further evaluated its association with *SPEG* expression across 49 tissues in GTEx database V8 (GTEx, 2020). We confirmed that rs12474050 is indeed a cis-eQTL of *SPEG* (Figure 2) and the COVID-19 risk allele rs12474050*T negatively associates with *SPEG* expression among muscle skeletal (nominal $p$ = 1.5e-8; beta = −0.106), heart atrial appendage (nominal p = 9e-3; beta = −0.074), and heart left ventricle (nominal $p$ = 3.7e-3; beta = −0.069), although only in the first tissue the SNP passed the multiple adjusted $p$-value threshold of $p \leq 0.001$ (0.05/49). In addition, rs12474050*T demonstrates positive association with *SPEG* expression in two brain tissues, including brain cerebellar hemisphere (nominal $p$ = 3.7e-5; beta = 0.28) and brain cerebellum (nominal $p$ = 4.3e-5; beta = 0.25), the association $p$ values in which all passed the multiple adjusted $p$-value threshold. Nevertheless, in the cis-eQTL meta-analysis, the posterior probability (m-value in Figure 2) of rs12474050 in each tissue revealed that muscle skeletal, heart atrial appendage, and heart left ventricle all display m-value >0.5, indicating the association pattern of rs12474050 with *SPEG* is more constant between these three tissues along with the majority of other tissues. In short, the risk allele rs12474050*T is a cis-eQTL of

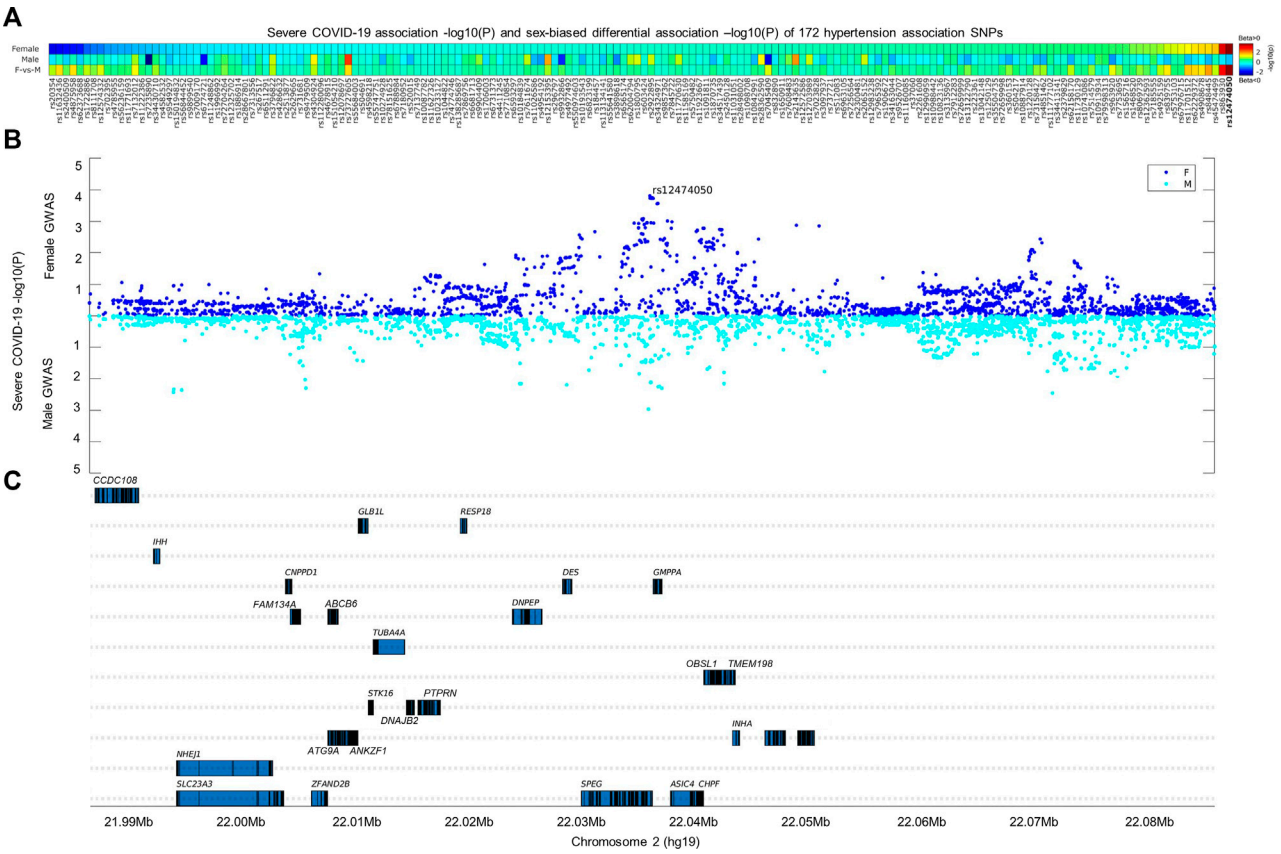



FIGURE 1
Integrative analysis of 172-hypertension association SNPs with severe COVID-19 genome-wide association signals by sex. The heatmap in the **(A,B)** display the COVID-19 association signals of these 172-hypertension association SNPs separated by sex, as well as sex-biased differential associations signals at the **(C)** of the heatmap (see method for calculation of sex-biased differential association *p* values); these SNPs are genome-wide significantly associated with hypertension (p < 5e-8) and are also cis expression quantitative trait loci (cis-eQTLs) according to Evangelou et al. (Evangelou et al., 2018). Only one SNP, rs12474050 (highlighted in red on the end-right side of the heatmap), mapped to *SPEG*, demonstrates sex-biased association (differential association *p*-value = 1.8e-4; passed the multiple adjusted *p*-value threshold of p < 3e-4) with severe COVID-19 in females. The SNP demonstrates suggestive association with severe COVID-19 females (*p* = 1.8e-4) and but not with severe COVID-19 males (*p* = 0.23). The scatterplots in the middle panel illustrate the severe COVID-19 association signals by sex for common SNPs (minor allele frequency >0.05) within a window of 1-Mbp where rs12474050 is located at the center, and genes covered in the window are located in the lower panel. Note: the COVID-19 GWAS association signals for females and males are colored in green and blue, respectively; the sex-biased multiple adjusted association *p*-value was set as $p \leq$ 3e-4, i.e., 0.05/172.

*SPEG* and associated with lower *SPEG* expression among the majority of GTEx tissues, particularly with significant association in muscle skeletal and suggestive associations in heart atrial appendage and heart left ventricle.

Since rs12474050*T was only found suggestively associated with severe COVID-19 in females but not in males, we questioned whether *SPEG* was differentially expressed between sex in muscle skeletal, heart atrial appendage, and heart left ventricle, as well as breast mammary tissue. We revealed that among the first three tissues *SPEG* expression was indeed suggestively higher in females than in males (Figure 2; nominally differential expression *p* values: 0.02, 0.03, and 0.004, respectively). Additionally, we also evaluated *SPEG* expression in the breast mammary tissue, which is biologically different between females and males, it revealed that females demonstrated statistically higher expression of *SPEG* (nominal *p* = 7.1e-7). Nevertheless, when tested the eQTL association in two brain tissues, including brain cerebellar hemisphere and brain cerebellum, there were no significant differences between females and males for *SPEG* expression. In conclusion, apart from the genetic effect of rs12474050, sex is another factor affecting *SPEG* expression in breast mammary tissue, muscle skeletal, heart atrial appendage, and heart left ventricle.

To investigate the potential involvement of *SPEG* in severe COVID-19, we determined *SPEG* expression among 13 single cell RNAseq data sets collected by UCSC Cell Browser and one standard along bulk RNAseq data related to cardiac cells upon SARS-CoV-2 infection from GEO (see method section for detail

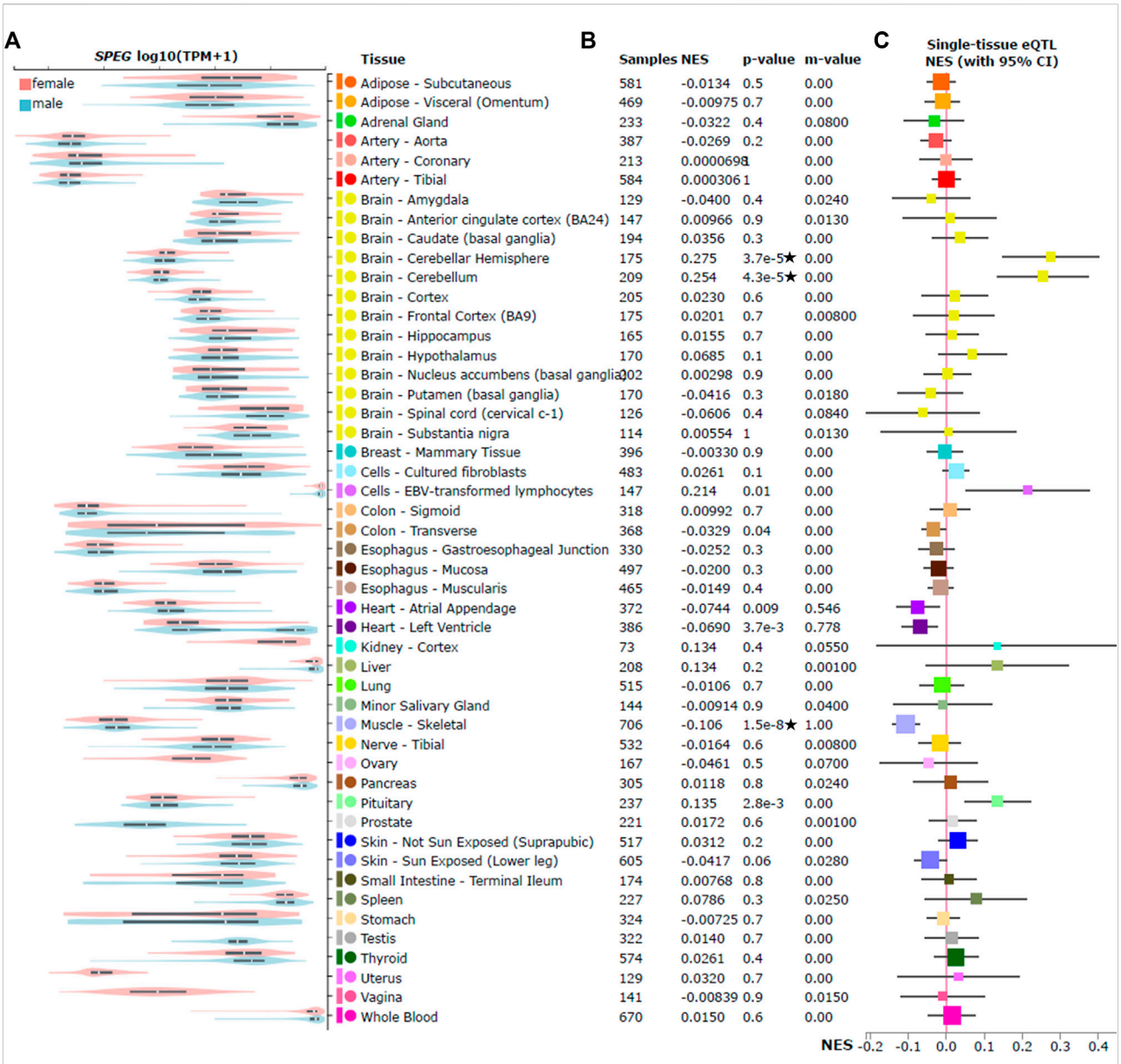



FIGURE 2
*SPEG* expression and cis expression quantitative trait locus (cis-eQTL) analysis of rs12474050 across 49 GTEx tissues. Violin plots on the **(A)** illustrates the expression of *SPEG* among 49 GTEx tissues by sex; among these tissues, only breast mammary tissue show statistically (nominal $p$ = 7.1e-7) higher expression of *SPEG* in females, although in heart left ventricle the SNP is nominally associated with *SPEG* expression (nominal $p$ = 0.004) but not passed the strict multiple adjustment $p$-value threshold $p \leq 0.001$ (0.05/49). Meanwhile, no significant differences between females and males for *SPEG* expression in the two brain tissues (brain cerebellar hemisphere and brain cerebellum that display significant eQTL signals) where rs12474050 is positively correlated with *SPEG* expression. **(B)** displays the multi-tissue cis-eQTL plot for rs12474050, with the COVID-19 risk allele rs12474050*T mostly, but negatively associated with *SPEG* expression in muscle skeletal that passes the multiple adjusted $p$-value threshold of $p \leq 0.001$. Note: star labeled along with the nominal cis-eQTL $p$-value indicates that the $p$-value passes the multiple adjustment $p$-value threshold. Nevertheless, the risk allele is positively associated with *SPEG* expression in two brain tissues, including cerebellar hemisphere and cerebellum. Details of cis-eQTL results are also provided, including tissue name, sample size, normalized effect size (NES), the posterior probability of rs12474050 for each tissue in cis-eQTL meta-analysis (m-value), and unadjusted cis-eQTL association $p$-value. **(C)** is the NES forest plot of rs12474050, with the mean of NES represented by square that are weighted and colored according to size of NES and tissue type, respectively.

of these data sets). Most of these single cell data sets related to COVID-19 from UCSC Cell Browser are linked to epithelium cells and immune cells responding to SARS-CoV-2 infection. However, the expression of *SPEG* was rare among different epithelium and immune cells based on evaluation of these single cell data sets in UCSC Cell Browser (data not shown).


FIGURE 3
Single cell analysis of *SPEG* expression among 8 GTEx tissues reveals *SPEG* expression mainly detected in myocyte related single cell types. Different single cell types from 8 tissues, including breast mammary tissue, esophagus mucosa, esophagus muscularis, heart left ventricle, lung, muscle skeletal, prostate, skin sun exposed lower leg, are determined for the % of cells expression *SPEG*. Higher % of cells expressing *SPEG* are found among myocyte related cell types. Low level of *SPEG* expression is detected among immune cells and epithelial cells.

This is in line with the observation of *SPEG* in GTEx single cell data is mainly expressed in myofibroblast and myocyte and <50% cells expression *SPEG* among all different single cell types (Figure 3); for cardiomyocytes from heart left ventricle, only 21% of myocytes (cardiac) and 16% of myocytes (cardiac, cytoplasmic) express *SPEG*. For single cell data from UCSC Cell Browser, only in heart tissue, *SPEG* was observed to be mainly expressed in the cardiac cell cardiomyocytes (average cardiomyocytes in each sample = 1,586; Figures 4A, B), with almost absence of *SPEG* expression in other single cell types,

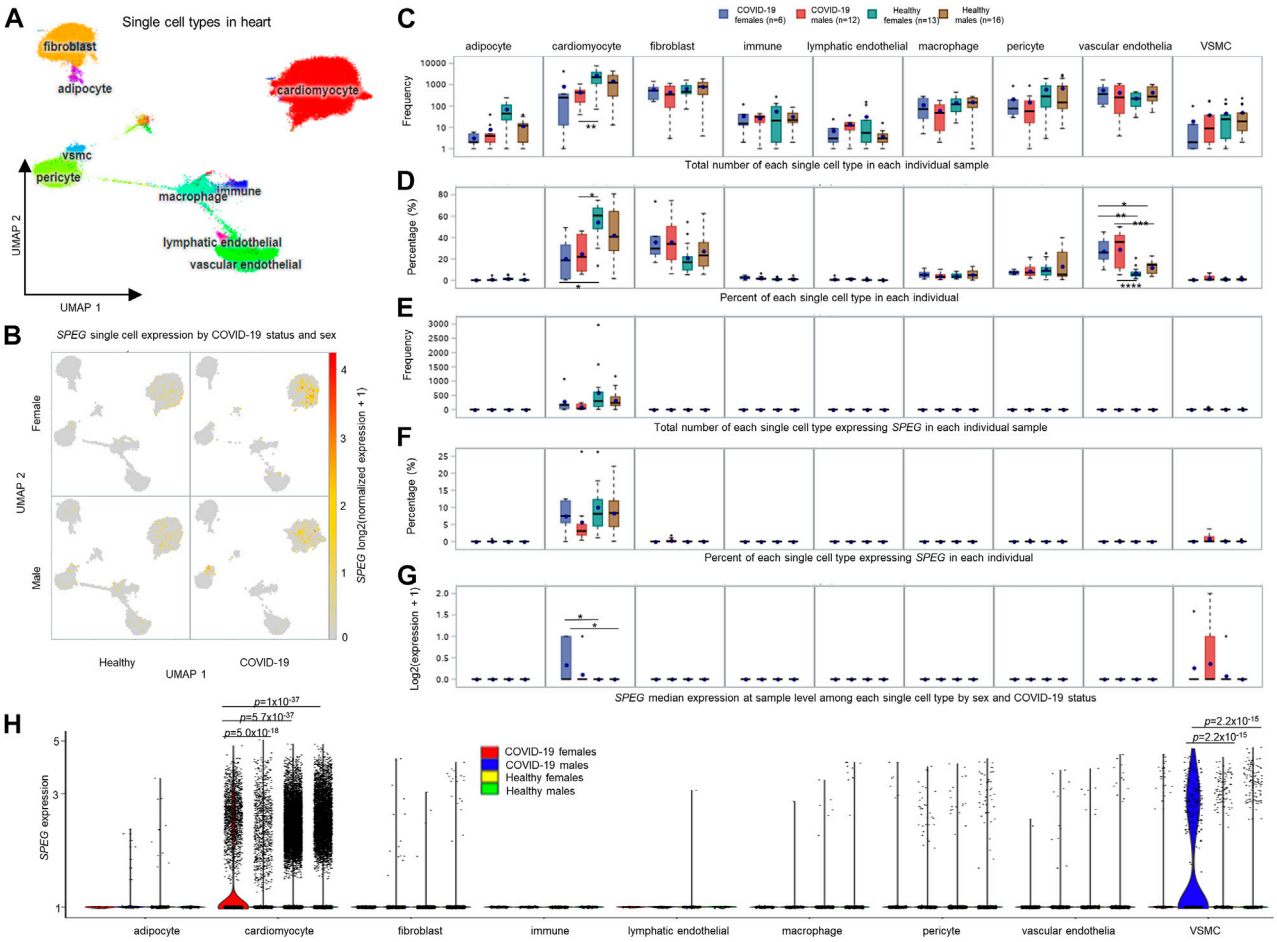



FIGURE 4
Single cell analysis of heart tissue reveals that *SPEG* expression is up-regulated in cardiomyocytes upon SARS-CoV-2 infection. **(A)** Uniform Manifold Approximation and Projection (UMAP) analysis of heart single cell data set derived from healthy controls (13 females and 16 males) and COVID-19 patients (6 females and 12 males) published by Melms et al. (Melms et al., 2021). **(B)** UMAPs divided by COVID-19 status and sex illustrating the cardiomyocytes is the major cell type in heart to express *SPEG*. **(C)** Total number of each single cell type in each individual sample compared among different groups, including COVID-19 females, COVID-19 males, healthy females, and healthy males. **(D)** Percent of each single cell type in each individual measured and evaluated across different sample groups. **(E)** Total number of each single cell type expressing *SPEG* in each individual sample determined across different groups. **(F)** Percent of each single cell type expressing *SPEG* in each individual examined among different sample groups. **(G)** Differential gene expression analysis of *SPEG* median expression at sample level among each single cell type by sex and COVID-19 status. Note: For each single cell type, multiple comparison *p* values were subjected to Tukey adjustment using the lsmeans statement in SAS proc GLM procedure, with the adjusted statistical significance represented by * ($p \leq 0.05$), ** ($p \leq 0.01$), *** ($p \leq 0.001$), and **** ($p \leq 0.0001$). The box-and-whisker plots display the mean (dot within box), median (line inside the box), inter-quantile interval (box), minimum (lowest value of whisker) and maximum (maximum value of whisker), with outliers represented by dots up or down the whiskers. **(H)** Differential gene expression analysis of *SPEG* expression at single cell level by aggregating all cells of each single cell type by sex and COVID-19 status. Nominal adjusted *p* values were determined using R package Seurat (Hao et al., 2021). All these nominal *p* values pass the multiple adjustment *p*-value threshold of $p < 0.05/(2 \times 6) = 4e\text{-}3$, as there are only 2 cell types, including cardiomyocyte and vascular smooth muscle cell (VSMC) displaying *SPEG* expression.

including adipocytes, fibroblast cells, immune cells, lymphatic endothelial cells, macrophages, pericytes, and vascular endothelia cells. Of note, VSMC was an exception, which was found to express *SPEG* but the average number of VSMC in each sample was only ~10 (Figure 4C).

To determine sex-biased differences among COVID-19 and healthy samples by sex, we conducted sample level analysis by separating samples into four groups, including COVID-19 females, COVID-19 males, healthy females, and healthy males, and evaluated cell counts and cell percentage of each single cell type across these four groups by individual. We revealed that both COVID-19 females and males had lower counts of cardiomyocytes compared to healthy individuals, but only between COVID-19 males and healthy females, the number of cardiomyocyte counts was significantly different (Figure 4C). Further comparisons of cell percentage for each single cell type

across these four groups revealed that both COVID-19 females and males show lower cardiomyocyte percentage compared to healthy females (adjusted $p$ values <0.05; Figure 4D). Surprisingly, we found the percentages of vascular endothelial cells are significantly higher in both COVID-19 females and males than both healthy females and males (all adjusted $p$ values <0.05; Figure 4D). However, no differences for the percent of vascular endothelial cells were observed between females and males of COVID-19 or healthy individuals. We further dissected the cell count and percentage of each single cell type expressing *SPEG* [log2 (normalized expression +1) > 0] in the data set and observed no significant variations for the two parameters in cardiomyocytes and other single cell types (Figures 4E, F). Furthermore, we determined the upregulation of *SPEG* expression in cardiomyocytes by comparing these four groups (Figure 4G). We confirmed that the increased fold change was higher in female COVID-19 patients compared to both healthy females and males (multiple adjusted $p$ values <0.05). When compared COVID-19 females with COVID-19 males, the *SPEG* expression is higher in females (fold change = 1.55), although the multiple adjusted $p$-value is not significant ($p$ = 0.3).

Furthermore, we performed differential gene expression analysis for *SPEG* at single cell level by collecting all cells of a specific cell type from a group of individuals across four patient groups, including COVID-19 females, COVID-19 males, healthy females, and healthy males. It turned out that the strategy is more robust, as more significant upregulation of *SPEG* in cardiomyocyte was observed in COVID females (0.29 percent of cardiomyocytes expression *SPEG*) than in COVID-19 males (0.22 percent of cardiomyocytes expression *SPEG*) (differential expression $p$-value = 5.0e-18 and fold change = 1.2; see Figure 4H; Supplementary Figure S1). Detailed evaluation of *SPEG* expression in each individual sample (Supplementary Figure S1) revealed that three out of six COVID-19 females shown higher upregulation of *SPEG* in cardiomyocytes, with another 2 COVID-19 females contained <100 cardiomyocytes. Meanwhile, among 12 COVID-19 males, there are two COVID-19 males with no cardiomyocytes detected in the single cell data, seven COVID-19 males demonstrated lower expression of *SPEG*, and two male COVID-19 individuals displayed higher *SPEG* expression comparable to COVID-19 females. The lack of enough cardiomyocytes in the two female COVID-19 individuals and the higher expression of *SPEG* in two COVID-19 males are the reason why the insignificant difference at sample level was observed between COVID-19 females and males in terms of *SPEG* expression. In addition, COVID-19 females but not COVID-19 males demonstrated significantly higher expression of *SPEG* compared to these healthy females and males (Figure 4H; all $p$ values <1e-36). Individual level single cell expression of *SPEG* in cardiomyocyte confirmed that majority of healthy females and males demonstrated lower expression of *SPEG* (Supplementary Figure S1). In terms of VSMCs, although aggregating all single cells by patient group increased the differences of *SPEG* expression observed between COVID-19 males and either of healthy males and healthy females (both $p$ values = 2.2e-15), there are few VSMC cells expressing *SPEG* among all these individual samples and no significant difference in terms of *SPEG* expression in VSMCs found between females and males of COVID-19. The significant difference for VSMCs may be driven by two COVID-19 males with more VSMCs showing relatively higher expression of *SPEG* compared to all other healthy individuals. Thus, the significant up-regulation of VSMCs in COVID-19 males is not conclusive.

Finally, we evaluated *SPEG* expression in a bulk RNAseq data of cardiac cells upon SARS-CoV-2 infection. The bulk RNAseq dataset consists of multiple cardiac cells, including human induced pluripotent stem cell-derived cardiomyocytes, cardiac fibroblasts, and endothelial cells. We only observed that the significant up-regulation of *SPEG* upon SARS-CoV-2 infection with high multiplicity of infection equal to 0.1 in cardiomyocytes (Supplementary Figure S2).

To this end, both in single cell and bulk RNAseq analysis, *SPEG* is specifically expressed in heart cardiomyocytes, and upon SARS-CoV-2 infection, *SPEG* is up-regulated in cardiomyocytes, with only statistically significant upregulation of *SPEG* in COVID-19 females but not in COVID-19 males compared to either female or male healthy controls at both sample level and single cell level analyses.

## Discussion

Growing evidence suggests that hypertension is the frequently observed comorbidity with severe COVID-19 (Wang et al., 2020; Yang et al., 2020; Gallo et al., 2022). Through integrative screening of genetic factors associated with both hypertension and severe COVID-19, we revealed that the cis-eQTL rs12474050 of *SPEG* is a potential host factor predisposing to higher DBP and severe COVID-19 in women. The risk allele rs12474050*T correlates with lower expression of *SPEG* in multiple tissues, including muscle skeletal, heart atrial appendage, and heart left ventricle. Among 49 GTEx tissues, higher expression of *SPEG* was detected in muscle skeletal, heart atrial appendage, and heart left ventricle of females compared to males, potentially suggesting *SPEG* expression is under sex-biased regulation. Coincidently, Fagerberg et al. and Singh et al. demonstrated *SPEG* was highest decoded in endometrium (Fagerberg et al., 2014; Singh et al., 2020). The close relationship between endometrium and cardiomyocyte was presented by Fan et al. that the endometrium-derived stem cells could repair myocardial ischemia injury (Fan et al., 2021). In our study, we further confirmed that higher *SPEG* expression in breast mammary tissue between females and males. Such observations rationalize our hypothesis that *SPEG* is highly sex-biased with

severe COVID-19 in women. Compared to males, females tended to have higher risk for heart related diseases, such as atrial fibrillation (Chugh et al., 2013; Emdin et al., 2016). Females preferred to suffer from long COVID than males since immune system of females is more sensitive upon viral infection (Ganesh et al., 2022). Our analyses implicated the interplay among sex, hypertension, cardiovascular diseases, and severe COVID-19. This might be due to the important functions of *SPEG* in heart, especially for its potential protective roles in cardiomyocytes in COVID-19. Notably, the post-acute cardiovascular manifestations of COVID-19 were widely reported (Xie et al., 2022) and the putative mechanisms interestingly involved fibrosis and scarring of cardiac tissue caused by activated TGF-β signaling (Mustroph et al., 2021). Coincidently, we revealed that upon SARS-CoV-2 infection the expression of *SPEG* was upregulated in cardiomyocytes, implicating the potential damage in heart in COVID-19. As SARS-CoV-2 infection could damage the cardiac tissue both directly and indirectly (Topol, 2020; Saeed et al., 2021), the upregulation of *SPEG*, partially increased in higher magnitude in female cardiomyocytes, would be a potentially compensatory response and self-cardioprotective action due to the infection of SARS-CoV-2 in heart. Taken together, the expression of *SPEG* is influenced by sex and SARS-CoV-2 infection, and *SPEG* is a critical host factor involved in severe COVID-19 in women.

Previous studies advocated that *SPEG* played pivotal roles in the development, maintenance, and function of cardiac and skeletal muscles (Campbell et al., 2021; Luo et al., 2021). The *SPEG* gene encoded SPEG belongs to the Unc89 subfamily, myosin light chain kinase (MLCK) protein family (Geisler et al., 2007). The Unc89 subfamily members could induce phosphorylation of junctophilin 2 (JPH2), ryanodine receptor (RyR2), sarcoplasmic/endoplasmic reticulum $Ca^{2+}$ ATPase 2a (SERCA2a), and α-tropomyosin (TPM1) in cardiac muscle (Campbell et al., 2021). As a result, *via* the triggering of phosphorylation by SPEG, the JPH2, RyR2, SERCA2a, and TPM1 were tightly involved in excitation–contraction (E-C) coupling (Bers, 2008; Landstrom et al., 2017). The E-C coupling is a key physiological process of conversion of electrical stimuli into a mechanical action in skeletal muscle contraction and is a core phenotype as one of the biological functions of SPEG. *SPEG* mutations have been found in patients with centronuclear myopathy which is an inherited neuromuscular disorder characterized by clinical features of a congenital myopathy and centrally placed nuclei on muscle biopsy (Agrawal et al., 2014). In addition, the lower expression of *SPEG* was observed in human end-stage HF (Quick et al., 2017). Further *in vivo* studies suggests that SPEG is essential for proper myocyte formation and maturation, and for cardiac development and function (Ono et al., 2005). Consequently, Quick et al. reported that specific tamoxifen-inducible acute down-regulation of *SPEG* in cardiomyocytes of over 8 weeks mice resulted in disruption of transverse tubule integrity, impaired calcium handling, altered E-C coupling, and HF (Quick et al., 2017). Based on the abundant pieces of information for critical role of *SPEG* in cardiovascular system, we confirmed the enhanced expression of *SPEG* in female but not male would be protective against severe COVID-19.

Our integrative genetic screening of SNPs associated with both hypertension and severe COVID-19 revealed that the cis-eQTL rs12474050 of *SPEG* was potentially associated with severe COVID-19 in women. Both single cell RNAseq and bulk RNAseq analyses confirmed the upregulation of *SPEG* in cardiac cell type cardiomyocytes upon SARS-CoV-2 infection. Furthermore, *SPEG* expression was higher in normal heart tissues of females than that of males. In addition, *SPEG* was upregulated in cardiomyocytes of female COVID-19 patients compared to both healthy females and males. Our finding suggested that *SPEG* plays a protective role in heart damage of female COVID-19 patients. This has even broad implication regarding to the substantially increased risk of heart disease in COVID-19 survivors according to two recent large-scale studies, including a 2021 study based on 13,638 health records (Mainous et al., 2021) from Florida and a 2022 study considering 153,760 COVID-19 survivors and thousands of controls (Xie et al., 2022). COVID-19 might induce cardiac injury through systemic inflammation and ischemic pathways, also including stress cardiomyopathy, acute and fulminant myocarditis (Abbasi, 2021). In our study, the percentage of vascular endothelial cells in heart tissue from COVID-19 patients of both sex is dramatically increased, which is consistent with previous report that vascular endothelia dysfunction is commonly associated with COVID-19 induced heart failure (Otifi and Adiga, 2022). Nevertheless, female-gender differences in cardiovascular diseases from SARS-CoV-2 were not fully understood. Some reports demonstrated these may be due to COVID-19 induced microvascular coagulopathy and worsening consequent thrombocytopenia (Muir et al., 2021; Bechmann et al., 2022) thereafter to cause heart damage. Additionally, stress cardiomyopathy was frequently found in COVID-19 female patients (Magadum and Kishore, 2020). Regarding to the crucial function of *SPEG* in heart-left ventricular E-C coupling, we deduce that *SPEG* is involved in severe COVID-19 or long COVID induced heart damage particular in woman by providing protective roles. Further investigations using cardiomyocytes with *SPEG* knockout, the tamoxifen treatment mouse model, or an inducible heart-specific *SPEG* knockout/overexpression mouse model will be warranted in order to systematically delineate the exact mechanisms for genetic functional variants in female severe COVID-19 patients.

Our study has limitations. One limitation is that we are not able to infer the potential causal relationship between hypertension and sex-biased severe COVID-19 in women. Performing colocalization analysis to infer causal relationship between hypertension and severe COVID-19 requires to have GWAS summary statistics from both hypertension and sex-

stratified severe COVID-19 GWASs (Foley et al., 2021). Unfortunately, the hypertension GWAS summary statistics is not freely available, and thus we are not able to conduct the analysis in current study. However, we found that rs12474050 is an independent genome-wide significant SNPs in the hypertension GWAS, and in the sex-stratified severe COVID-19 GWAS the SNP is also one of the top signals associated with severe COVID-19 in female patients (Figure 1B). Therefore, we trust that rs12474050 is a potential SNP associated with both hypertension and COVID-19 in females. Nevertheless, given the small sample size of severe female COVID-19 patients in the current severe COVID-19 GWAS, further replication of the severe COVID-19 association in women needs to be conducted first before evaluating the causal relationship between hypertension and severe COVID-19 in females, as well as between hypertension or severe COVID-19 and *SPEG* gene expression. In addition, other important hypertension signals may be excluded in our analysis. A more comprehensive analysis would be to search for all hypertension signals from all published GWAS studies in the sex-stratified severe COVID-19 GWASs and subsequently perform sex-biased analysis for these curated hypertension SNPs in COVID-19. Alternatively, it is also feasible to carry out genome-wide cross-trait meta-analysis between sex-biased COVID-19 GWASs and hypertension GWAS according to cai et al. (Cai et al., 2022). The other limitation is that rs12474050 is also a missense variant of *SPEGNB*, which is also named as *SPEG* neighbor. However, the expression data of *SPEGNB* is not included in the GTEx portal due to it is expression level is extremely low across 49 tissues. Further evaluation of its expression in the human protein expression atlas (Uhlén et al., 2015) confirmed that it is only lowly and specifically expressed in the tongue and muscle skeletal tissue, among which the TPM of *SPEGNB* is less than 3. In addition, the function of *SPEGNB* is still undetermined. The potential involvement of *SPEGNB* in both hypertension and severe COVID-19 as well as it is relationship with *SPEG* warrants for further study. The third limitation is about the exact reason of the low percentage of cardiomyocytes expressing *SPEG* in both the heart samples from COVID-19 patients and healthy control is not clear. Although potential sub-clusters of cardiomyocytes may specifically express *SPEG*, we could not find evidences to support this. In details, in the heart single cell data set, we find the cardiomyocytes that express *SPEG* tend to be evenly distributed in the cardiomyocyte cluster (Figure 4B), and all cardiomyocytes are clustered tightly together in the UMAP, suggesting no potential sub-clusters exist among cardiomyocytes. The above conclusion is further confirmed after re-analyzing the single cardiac cell data of 'Cardiac Differentiation' published by Kathiriya et al. (Kathiriya et al., 2021) *via* UCSC Cell Browser. Kathiriya et al. studied the knockout of an important cardiac cell development transcription factor *TBX5* by comparing single cell expression data of wild type of human induced pluripotent stem cell (iPSC) lines, Crispr-cas9-exposed control, with two Crispr-cas9-enginered *TBX5* knockout iPSC lines, including one single copy knockout and two-copy knockout of *TBX5* iPSC lines. They generated single cell data sets of iPSC lines developed to cardiomyocytes in a range of time points, including 6 days, 11 days, and 23 days. By performing trajectory inference analysis on the single cell data (Supplementary Figure S3), we confirm that *SPEG* tend to be highly expressed at the time point 11 days, with lower expression of *SPEG* detected at 3 days and no expression of *SPEG* observed at 23 days. We also find that *SPEG* is evenly expressed among cardiomyocytes at the time point 11 days and no obvious sub-clusters specifically expressing *SPEG*. In conclusion, it would be difficult to determine cardiomyocyte sub-clusters in the current heart single cell data sets, mainly due to the expression characteristics of *SPEG* during cardiomyocyte development. Future research is warranted to address the above problem.

## Data availability statement

Publicly available datasets were analyzed in this study. Used data can be found in the text.

## Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## Author contributions

Z-SC, Y-SL, X-CS, and KZ conceived the study. Y-SL, X-CS, WL, X-MY, M-YG, FC, H-YS, and HG collected the raw data. Y-SL, G-FW, P-PZ, YN, J-HW, and RM proceeded the data statistics. Z-SC, X-CS, and KZ analyzed and interpreted the data. Z-SC, Y-SL, X-CS, and KZ drafted and revised the manuscript.

## Funding

This project was partly supported by Science and Technology Supporting Plan of Scientific and Technological Department of Guizhou Province ([2020]4Y163), Science and Technology Program of Guiyang Science and Technology Bureau [(2020)-10-2], Cultivation Project of National Natural Science Foundation of China of the Affiliated Hospital of Guizhou Medical University and Guizhou Medical University (I-2020-06, 20NSP010).

## Acknowledgments

We are thankful to all researchers who participated and contributed to the GRASP project, as without their generosities in freely sharing the two severe COVID-19 GWASs to public we might not be able to make current finding. We also need to thank the reviewer who provides constructive suggestions to improve the quality of the manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1041470/full#supplementary-material

**SUPPLEMENTARY FIGURE S1**
Violin plots and Uniform Manifold Approximation and Projection (UMAP) plots for SPEG expression at single sample level. **(A,B)** illustrate SPEG expression in each individual sample for two cell types, including cardiomyocyte and vascular smooth muscle cell (VSMC), respectively. These individuals are categorized into four patient groups, including COVID-19 females, COVID-19 males, healthy females, and healthy males, which are indicated by different color bars at the bottom of these violin plots. SPEG expression is further evaluated at single simple level using UMAP plots for the four groups, including COVID-19 females **(C)**, COVID-19 males **(D)**, healthy females **(E)** and healthy males **(F)**. Note: R package Seurat (Hao et al., 2021) was used to perform single cell analysis at single sample level.

**SUPPLEMENTARY FIGURE S2**
Bulk RNAseq analysis of multiple cardiac cell lines, including human induced pluripotent stem cell-derived cardiomyocytes, cardiac fibroblasts, and endothelial cells, reveal that SPEG expression is up-regulated in cardiomyocytes upon SARS-CoV-2 infection. Among multiple cardiac cell lines, including human induced pluripotent stem cell (iPSC), iPSC-derived cardiomyocytes (Cardiomyocytes), Cardiac fibroblasts, and Endothelial cells, *SPEG* expression is increased in cardiomyocytes upon SARS-CoV-2 infection, compared to its corresponding mock controls. The data set was published by Perez-Bermejo et al., cells were treated with either vehicle or SARS-CoV-2 virus for 48 hours at a MOI of either 0.0006 (low), 0.01 (middle), or 0.1 (high) (Perez-Bermejo et al., 2021). Note: Pairwise differential expression analysis were performed with the Student's t-test, with the unadjusted statistical significance represented by * ($p \leq 0.05$), ** ($p \leq 0.01$), *** ($p \leq 0.001$), and **** ($p \leq 0.0001$); MOI: multiplicity of infection. The multiple adjusted p value threshold is set as $p < 0.008$ (0.05 / 6), thus, only the comparison of *SPEG* expression between high MOI SASR-CoV-2 infection and its corresponding mock control passes the multiple adjusted p value threshold.

**SUPPLEMENTARY FIGURE S3**
Single cell trajectory inference analysis for SPEG expression in cardiac cells using UCSC Cell Browser. Three panels, including **(A–C)**, illustrating the UMAP for all single cells (upper part), UMAP of SPEG expressed cells (middle part) , and pseudotime analyses for SPEG expressed cells and all cells, with cells are colored by SPEG expression level or by different sample source, respectively (see lower two consecutive parts). The single cell data were derived from the following 4 groups: wild type of human induced pluripotent stem cell (iPSC) lines, Crispr-cas9-exposed control lines, and two Crispr-cas9-enginered TBX5 knockout iPSC lines, including one single copy knockout and two-copy knockout of TBX5 iPSC lines, for 3 time points of 6 days, 11 days, and 23 days. The data set was published by Kathiriya et al. (Kathiriya et al., 2021) and is named as "Cardiac Differentiation" at UCSC Cell Browser. Note: on day 23, SPEG expression is not detectable.

## References

Abbasi, J. (2021). Researchers investigate what COVID-19 does to the heart. *Jama* 325 (9), 808–811. doi:10.1001/jama.2021.0107

Agrawal, P., Pierson, C. R., Joshi, M., Liu, X., Ravenscroft, G., Moghadaszadeh, B., et al. (2014). SPEG interacts with myotubularin, and its deficiency causes centronuclear myopathy with dilated cardiomyopathy. *Am. J. Hum. Genet.* 95 (2), 218–226. doi:10.1016/j.ajhg.2014.07.004

Akpek, M. (2022). Does COVID-19 cause hypertension? *Angiology* 73 (7), 682–687. doi:10.1177/00033197211053903

Bechmann, N., Barthel, A., Schedl, A., Herzig, S., Varga, Z., Gebhard, C., et al. (2022). Sexual dimorphism in COVID-19: Potential clinical and public health implications. *Lancet. Diabetes Endocrinol.* 10 (3), 221–230. doi:10.1016/S2213-8587(21)00346-6

Bers, D. M. (2008). Calcium cycling and signaling in cardiac myocytes. *Annu. Rev. Physiol.* 70, 23–49. doi:10.1146/annurev.physiol.70.113006.100455

Briquez, P. S., Rouhani, S. J., Yu, J., Pyzer, A. R., Trujillo, J., Dugan, H. L., et al. (2022). Severe COVID-19 induces autoantibodies against angiotensin II that correlate with blood pressure dysregulation and disease severity. *Sci. Adv.* 8 (40), eabn3777. doi:10.1126/sciadv.abn3777

Cai, L., He, C., Liu, Y., Sun, Y., He, L., and Baranova, A. (2022). Inflammation and immunity connect hypertension with adverse COVID-19 outcomes. *Front. Genet.* 13, 933148. doi:10.3389/fgene.2022.933148

Campbell, H., Aguilar-Sanchez, Y., Quick, A. P., Dobrev, D., and Wehrens, X. H. T. (2021). Speg: A key regulator of cardiac calcium homeostasis. *Cardiovasc. Res.* 117 (10), 2175–2185. doi:10.1093/cvr/cvaa290

Chen, J., Liu, Y., Qin, J., Ruan, C., Zeng, X., Xu, A., et al. (2021). Hypertension as an independent risk factor for severity and mortality in patients with COVID-19: A retrospective study. *Postgrad. Med. J.* 98 (1161), 515–522. doi:10.1136/postgradmedj-2021-140674

Chugh, S. S., Havmoeller, R., Narayanan, K., Singh, D., Rienstra, M., Benjamin, E. J., et al. (2013). Worldwide epidemiology of atrial fibrillation: A global burden of disease 2010 study. *Circulation* 129 (8), 837–847. doi:10.1161/CIRCULATIONAHA.113.005119

Dai, L. S., Zhu, M. P., Li, Y. M., Zhou, H. M., Liao, H. L., Cheng, P. P., et al. (2022). Hypertension exacerbates severity and outcomes of COVID-19 in elderly patients: A retrospective observational study. *Curr. Med. Sci.* 42 (3), 561–568. doi:10.1007/s11596-022-2539-y

Emdin, C. A., Wong, C. X., Hsiao, A. J., Altman, D. G., Peters, S. A., Woodward, M., et al. (2016). Atrial fibrillation as risk factor for cardiovascular disease and death in women compared with men: Systematic review and meta-analysis of cohort studies. *BMJ* 532, h7013. doi:10.1136/bmj.h7013

Evangelou, E., Warren, H. R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H., et al. (2018). Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* 50 (10), 1412–1425. doi:10.1038/s41588-018-0205-x

Fagerberg, L., Hallstrom, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., et al. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics.* 13 (2), 397–406. doi:10.1074/mcp.M113.035600

Fan, X., Song, H., Yin, W., Zhang, J., Peng, Z., Yang, K., et al. (2021). Human endometrium-derived stem cell improves cardiac function after myocardial ischemic injury by enhancing angiogenesis and myocardial metabolism. *Stem Cell Res. Ther.* 12 (1), 344. doi:10.1186/s13287-021-02423-5

Foley, C. N., Staley, J. R., Breen, P. G., Sun, B. B., Kirk, P. D. W., Burgess, S., et al. (2021). A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat. Commun.* 12 (1), 764. doi:10.1038/s41467-020-20885-8

Gallo, G., Calvez, V., and Savoia, C. (2022). Hypertension and COVID-19: Current evidence and perspectives. *High. Blood Press. Cardiovasc. Prev.* 29 (2), 115–123. doi:10.1007/s40292-022-00506-9

Ganesh, R., Grach, S. L., Ghosh, A. K., Bierle, D. M., Salonen, B. R., Collins, N. M., et al. (2022). The female-predominant persistent immune dysregulation of the post-COVID syndrome. *Mayo Clin. Proc.* 97 (3), 454–464. doi:10.1016/j.mayocp.2021.11.033

Geisler, S. B., Robinson, D., Hauringa, M., Raeker, M. O., Borisov, A. B., Westfall, M. V., et al. (2007). Obscurin-like 1, OBSL1, is a novel cytoskeletal protein related to obscurin. *Genomics* 89 (4), 521–531. doi:10.1016/j.ygeno.2006.12.004

Gtex, C. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369 (6509), 1318–1330. doi:10.1126/science.aaz1776

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., 3rd, Zheng, S., Butler, A., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184 (13), 3573–3587.e29. e29. doi:10.1016/j.cell.2021.04.048

Kathiriya, I. S., Rao, K. S., Iacono, G., Devine, W. P., Blair, A. P., Hota, S. K., et al. (2021). Modeling human TBX5 haploinsufficiency predicts regulatory networks for congenital heart disease. *Dev. Cell* 56 (3), 292–309. e9. doi:10.1016/j.devcel.2020.11.020

Landstrom, A. P., Dobrev, D., and Wehrens, X. H. T. (2017). Calcium signaling and cardiac arrhythmias. *Circ. Res.* 120 (12), 1969–1993. doi:10.1161/CIRCRESAHA.117.310083

Lippi, G., Wong, J., and Henry, B. (2020). Hypertension in patients with coronavirus disease 2019 (COVID-19): A pooled analysis. *Pol. Arch. Intern. Med.* 130 (4), 304–309. doi:10.20452/pamw.15272

Luo, S., Rosen, S. M., Li, Q., and Agrawal, P. B. (2021). Striated Preferentially Expressed Protein kinase (SPEG) in muscle development, function, and disease. *Int. J. Mol. Sci.* 22 (11), 5732. doi:10.3390/ijms22115732

Luo, Y.-S., Luo, L., Li, W., Chen, Y., Wu, G. F., Chen, F., et al. (2022). Evaluation of a functional single nucleotide polymorphism of the SARS-CoV-2 receptor ACE2 that is potentially involved in long COVID. *Front. Genet.* 13, 931562. doi:10.3389/fgene.2022.931562

Magadum, A., and Kishore, R. (2020). Cardiovascular manifestations of COVID-19 infection. *Cells* 9 (11), 2508. doi:10.3390/cells9112508

Mainous, A. G., 3rd, Rooks, B. J., Wu, V., and Orlando, F. A. (2021). COVID-19 post-acute sequelae among adults: 12 month mortality risk. *Front. Med.* 8, 778434. doi:10.3389/fmed.2021.778434

Melms, J. C., Biermann, J., Huang, H., Wang, Y., Nair, A., Tagore, S., et al. (2021). A molecular single-cell lung atlas of lethal COVID-19. *Nature* 595 (7865), 114–119. doi:10.1038/s41586-021-03569-1

Muir, K. L., Kallam, A., Koepsell, S. A., and Gundabolu, K. (2021). Thrombotic thrombocytopenia after Ad26.COV2.S vaccination. *N. Engl. J. Med.* 384 (20), 1964–1965. doi:10.1056/NEJMc2105869

Mustroph, J., Hupf, J., Baier, M. J., Evert, K., Brochhausen, C., Broeker, K., et al. (2021). Cardiac fibrosis is a risk factor for severe COVID-19. *Front. Immunol.* 12, 740260. doi:10.3389/fimmu.2021.740260

Ono, Y., Schwach, C., Antin, P. B., and Gregorio, C. C. (2005). Disruption in the tropomodulin1 (Tmod1) gene compromises cardiomyocyte development in murine embryonic stem cells by arresting myofibril maturation. *Dev. Biol.* 282 (2), 336–348. doi:10.1016/j.ydbio.2005.03.015

Otifi, H. M., and Adiga, B. K. (2022). Endothelial dysfunction in covid-19 infection. *Am. J. Med. Sci.* 363 (4), 281–287. doi:10.1016/j.amjms.2021.12.010

Perez-Bermejo, J. A., Kang, S., Rockwood, S. J., Simoneau, C. R., Joy, D. A., Silva, A. C., et al. (2021). SARS-CoV-2 infection of human iPSC–derived cardiac cells reflects cytopathic features in hearts of patients with COVID-19. *Sci. Transl. Med.* 13 (590), eabf7872. doi:10.1126/scitranslmed.abf7872

Quick, A. P., Wang, Q., Philippen, L. E., Barreto-Torres, G., Chiang, D. Y., Beavers, D., et al. (2017). SPEG (Striated Muscle Preferentially Expressed Protein Kinase) is essential for cardiac function by regulating junctional membrane complex activity. *Circ. Res.* 120 (1), 110–119. doi:10.1161/CIRCRESAHA.116.309977

Saeed, S., Tadic, M., Larsen, T. H., Grassi, G., and Mancia, G. (2021). Coronavirus disease 2019 and cardiovascular complications: Focused clinical review. *J. Hypertens.* 39 (7), 1282–1292. doi:10.1097/HJH.0000000000002819

Singh, M., Bansal, V., and Feschotte, C. (2020). A single-cell RNA expression map of human coronavirus entry factors. *Cell Rep.* 32 (12), 108175. doi:10.1016/j.celrep.2020.108175

Speir, M. L., Bhaduri, A., Markov, N. S., Moreno, P., Nowakowski, T. J., Papatheodorou, I., et al. (2021). UCSC cell browser: Visualize your single-cell data. *Bioinformatics* 37 (23), 4578–4580. doi:10.1093/bioinformatics/btab503

Thibord, F., Chan, M. V., Chen, M. H., and Johnson, A. D. (2022). A year of COVID-19 GWAS results from the GRASP portal reveals potential genetic risk factors. *HGG Adv.* 3 (2), 100095. doi:10.1016/j.xhgg.2022.100095

Topol, E. J. (2020). COVID-19 can affect the heart. *Science* 370 (6515), 408–409. doi:10.1126/science.abe2813

Trump, S., Lukassen, S., Anker, M. S., Chua, R. L., Liebig, J., Thurmann, L., et al. (2021). Hypertension delays viral clearance and exacerbates airway hyperinflammation in patients with COVID-19. *Nat. Biotechnol.* 39 (6), 705–716. doi:10.1038/s41587-020-00796-1

Uhlén, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347 (6220), 1260419. doi:10.1126/science.1260419

Uysal, B., Akca, T., Akacı, O., and Uysal, F. (2022). The prevalence of Post-COVID-19 hypertension in children. *Clin. Pediatr.* 61 (7), 453–460. doi:10.1177/00099228221085346

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., et al. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 323 (11), 1061–1069. doi:10.1001/jama.2020.1585

WHO (2022). Numbers at a glance for global COVID-19 cases. [Internet] 2022 [cited 2022 19 August]; Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (Accessed 08 19, 2022).

Xie, Y., Xu, E., Bowe, B., and Al-Aly, Z. (2022). Long-term cardiovascular outcomes of COVID-19. *Nat. Med.* 28 (3), 583–590. doi:10.1038/s41591-022-01689-3

Yang, J., Zheng, Y., Gou, X., Pu, K., Chen, Z., Guo, Q., et al. (2020). Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: A systematic review and meta-analysis. *Int. J. Infect. Dis.* 94, 91–95. doi:10.1016/j.ijid.2020.03.017

# Frontiers in Genetics

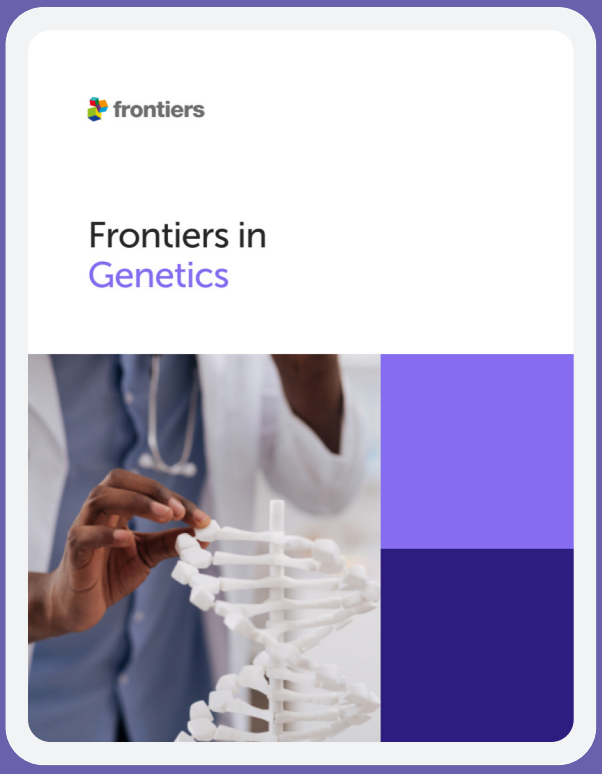
**Highlights genetic and genomic inquiry relating to all domains of life**

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

---

## Discover the latest Research Topics

See more →




**Frontiers**

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

**Contact us**

+41 (0)21 510 17 00
frontiersin.org/about/contact