


# Transcriptome-wide summary data-based Mendelian randomization analysis reveals 38 novel genes associated with severe COVID-19

Suhas Krishnamoorthy<sup>1</sup> | Gloria H.-Y. Li<sup>2</sup> | Ching-Lung Cheung<sup>1,3</sup> 

<sup>1</sup>Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong

<sup>2</sup>Department of Health Technology and Informatics, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

<sup>3</sup>Laboratory of Data Discovery for Health (D24H), Pak Shek Kok, Hong Kong

## Correspondence

Ching-Lung Cheung, Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, 21 Sassoon Rd, Pokfulam, Hong Kong.  
Email: [lung1212@hku.hk](mailto:lung1212@hku.hk)

## Abstract

Severe COVID-19 has a poor prognosis, while the genetic mechanism underlying severe COVID-19 remains largely unknown. We aimed to identify genes that are potentially causally associated with severe COVID-19. We conducted a summary data-based Mendelian randomization (SMR) analysis using expression quantitative trait loci (eQTL) data from 49 different tissues as the exposure and three COVID-19-phenotypes (very severe respiratory confirmed COVID-19 [severe COVID-19], hospitalized COVID-19, and SARS-CoV-2 infection) as the outcomes. SMR using multiple SNPs was used as a sensitivity analysis to reduce false positive rate. Multiple testing was corrected using the false discovery rate (FDR)  $q$ -value. We identified 309 significant gene-trait associations (FDR  $q$  value < 0.05) across 46 tissues for severe COVID-19, which mapped to 64 genes, of which 38 are novel. The top five most associated protein-coding genes were Interferon Alpha and Beta Receptor Subunit 2 (*IFNAR2*), 2'-5'-Oligoadenylate Synthetase 3 (*OAS3*), mucin 1 (*MUC1*), Interleukin 10 Receptor Subunit Beta (*IL10RB*), and Napsin A Aspartic Peptidase (*NAPSA*). The potential causal genes were enriched in biological processes related to type I interferons, interferon-gamma inducible protein 10 production, and chemokine (C-X-C motif) ligand 2 production. In addition, we further identified 23 genes and 5 biological processes which are unique to hospitalized COVID-19, as well as 13 genes that are unique to SARS-CoV-2 infection. We identified several genes that are potentially causally associated with severe COVID-19. These findings improve our limited understanding of the mechanism of COVID-19 and shed light on the development of therapeutic agents for treating severe COVID-19.

## KEYWORDS

COVID-19, eQTL, Mendelian randomization, transcriptome

## 1 | INTRODUCTION

The coronavirus disease 2019 (COVID-19) is a highly contagious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which has resulted in a global pandemic and public health crisis.<sup>1</sup> As of 12 March 2022, there have been over 452 million

confirmed cases of COVID-19, as well as over 6 million deaths.<sup>2</sup> COVID-19 infections display extensive variability in symptoms, prognosis, and severity among individuals, with some infections being asymptomatic and others being lethal.<sup>3</sup> Like other infectious diseases, host genetics affect susceptibility to COVID-19 infection, severity, and prognosis.<sup>4,5</sup> The COVID-19 Host Genetic Initiative

(HGI) is a collaborative effort, which generates and shares data regarding the genetic determinants of COVID-19 outcomes, and has previously identified several risk loci associated with the disease.<sup>4</sup> These findings provide important insights into how host genetics may influence COVID-19 infection and progression. However, no causality can be inferred by the genome-wide association study (GWAS). Thus, there is an urgent need to better understand the causal factors of COVID-19 infection and progression,<sup>5</sup> especially to explain the mechanism underlying severe COVID-19.

Although GWAS per se does not infer causality, combining summary statistics from multiple GWAS could infer causality via the Mendelian randomization (MR) framework. At the same time, MR has been shown to be a valid approach for drug repurposing, drug identification, and clinical management. Expression quantitative trait loci (eQTL) are the genetic variants that affect the expression level of a gene. Combining summary statistics from both transcriptome-wide eQTL and GWAS of COVID-19 using MR approach is powerful in identifying genes with their expression levels being causally associated with COVID-19 outcomes. Previous studies have identified 2<sup>6</sup> and 14<sup>7</sup> putatively causal genes of COVID-19. These studies, which were conducted using earlier releases of HGI data, provide valuable insights into the pathological mechanism of the disease. However, in these studies, only one genetic instrument was used without sensitivity analysis, leading to potentially biased findings and inflated false positive rates.

To better understand whether host genes, and hence its expression, could affect the susceptibility to severe COVID-19, we performed a summary data-based Mendelian Randomization (SMR) analysis using the top single-nucleotide polymorphism (SNP) in eQTL study as the instrument and the severe COVID-19, hospitalized COVID-19 and SARS-CoV-2 infection in release 6 of the COVID-19 HGI data as the outcomes.<sup>8,9</sup> SMR analysis using the summary statistics from multiple SNPs (multi-SNPs SMR) was performed as the sensitivity analysis to reduce the false positive rate.

## 2 | MATERIALS AND METHODS

### 2.1 | Study design and data sources

In the current study, we evaluated the causality of transcriptome-wide gene expression (exposure) on COVID-19 outcomes using the SMR analysis. Summary statistics from eQTL studies provide information on the effect of genetic variants on gene expression levels. As expression levels vary by tissue-type, eQTL studies are generally conducted for a specific tissue. In each tissue, probes are used to measure a gene's expression and the cis-eQTLs for that gene are any SNPs located within 1 Mb of the gene probe that are significantly associated with the gene's expression as defined by the  $P_{eQTL} < 5E-8$ . The cis-eQTLs were obtained from the 48 different tissues from v7 of the Genotype-Tissue Expression (GTEx) project with a sample size ranging from 80 to 491<sup>10</sup> as well as the cis-eQTL data from blood provided by the eQTLGen Consortium with a sample

size of 31 684.<sup>11</sup> Details of the sample size and tissue used are provided in Table 1. We utilized the expression data set generated from all the aforementioned 49 tissues because gene expression in various tissues could have a systematic effect, especially when the gene codes for a secreted protein. Previously, COVID-19 has been found to affect virtually all organs, while the virus binds to ACE2, which is present in nearly all tissues.<sup>12</sup> Moreover, some genes are tissue-specific, with the expression only detected in a few tissues. To obtain the broadest coverage of genes, we, therefore, studied 49 tissues in total.

Three COVID-19 outcomes were used in the SMR analysis: severe COVID-19 ( $n = 8779$ , controls = 1 001 875), hospitalized COVID-19 ( $n = 24 274$ , controls = 2 061 529) and SARS-CoV-2 infection ( $n = 112 612$ , controls = 2 474 079). The control groups were subjects from the general population who did not show the respective phenotype. Severe COVID-19 was considered the primary outcome since it was often associated with a poor prognosis. Hospitalized COVID-19 and SARS-CoV-2 infection were the secondary outcomes. The phenotypes (Supporting Information: Table S1) were defined by the COVID-19 HGI<sup>13</sup> and have been used previously in several studies.<sup>6,7,14-16</sup> GWAS summary data of these phenotypes were obtained from the meta-analyses round 6 data released by HGI (<https://www.covid19hg.org/results/r6/>). Each contributing study conducted its GWAS independently but following the HGI consortium guidelines which suggested accounting for the following covariates: age, sex, age<sup>2</sup>, age\*sex, and the first 20 principal components.<sup>13</sup>

### 2.2 | SMR and HEIDI analysis

SMR is a method that integrates summary statistics from GWAS and eQTL studies under the MR framework to prioritize genes whose expression levels are potentially causally associated with an outcome trait. MR analysis uses genetic variants as instrumental variables to infer a causal relationship between an exposure and an outcome. While conventional two-sample MR utilizes summary statistics from two independent GWAS to estimate the effect of one phenotype on another, SMR utilizes summary statistics from independent eQTL study and GWAS to estimate the effect of a gene's expression level on a phenotype. We adopted the SMR method (version 1.03) for our primary analysis (Figure 1A). For each gene, the cis-eQTL having the strongest association signal was used as the single genetic instrument in the primary analysis. This method has been adopted in previous SMR studies.<sup>6,7</sup> However, as more than one cis-eQTL could be implicated in the expression of one gene, using a single eQTL as the instrument could lead to biased results and potentially inflated false positive rate. In addition, SMR uses a single variant as the genetic instrument disallowing the distinguishment between associations that arise due to causality or due to horizontal pleiotropy (i.e., the association between the genetic variant and the exposure being independent of that between the genetic variant and the outcome). Therefore, we used the multi-SNPs SMR (Figure 1B) as the sensitivity

**TABLE 1** List of tissues used as exposures and their respective sample size

Tissue	Number of samples	Tissue	Number of samples
Brain – Substantia nigra	80	Colon – Sigmoid	203
Brain – Spinal cord (cervical c-1)	83	Esophagus – Gastroesophageal Junction	213
Minor Salivary Gland	85	Pancreas	220
Brain – Amygdala	88	Testis	225
Uterus	101	Stomach	237
Vagina	106	Colon – Transverse	246
Brain – Hypothalamus	108	Breast – Mammary Tissue	251
Brain – Anterior cingulate cortex (BA24)	109	Heart – Atrial Appendage	264
Brain – Hippocampus	111	Artery – Aorta	267
Brain – Putamen (basal ganglia)	111	Heart – Left Ventricle	272
Cells – EBV-transformed lymphocytes	117	Cells – Transformed fibroblasts	300
Brain – Frontal Cortex (BA9)	118	Adipose – Visceral (Omentum)	313
Ovary	122	Esophagus – Muscularis	335
Small Intestine – Terminal Ileum	122	Skin – Not Sun Exposed (Suprapubic)	335
Brain – Cerebellar Hemisphere	125	Esophagus – Mucosa	358
Brain – Nucleus accumbens (basal ganglia)	130	Nerve – Tibial	361
Prostate	132	Whole Blood	369
Brain – Cortex	136	Lung	383
Brain – Caudate (basal ganglia)	144	Adipose – Subcutaneous	385
Spleen	146	Artery – Tibial	388
Artery – Coronary	152	Thyroid	399
Liver	153	Skin – Sun Exposed (Lower leg)	414
Brain – Cerebellum	154	Muscle – Skeletal	491
Pituitary	157	Blood (eQTLgen)	31 684
Adrenal Gland	175		

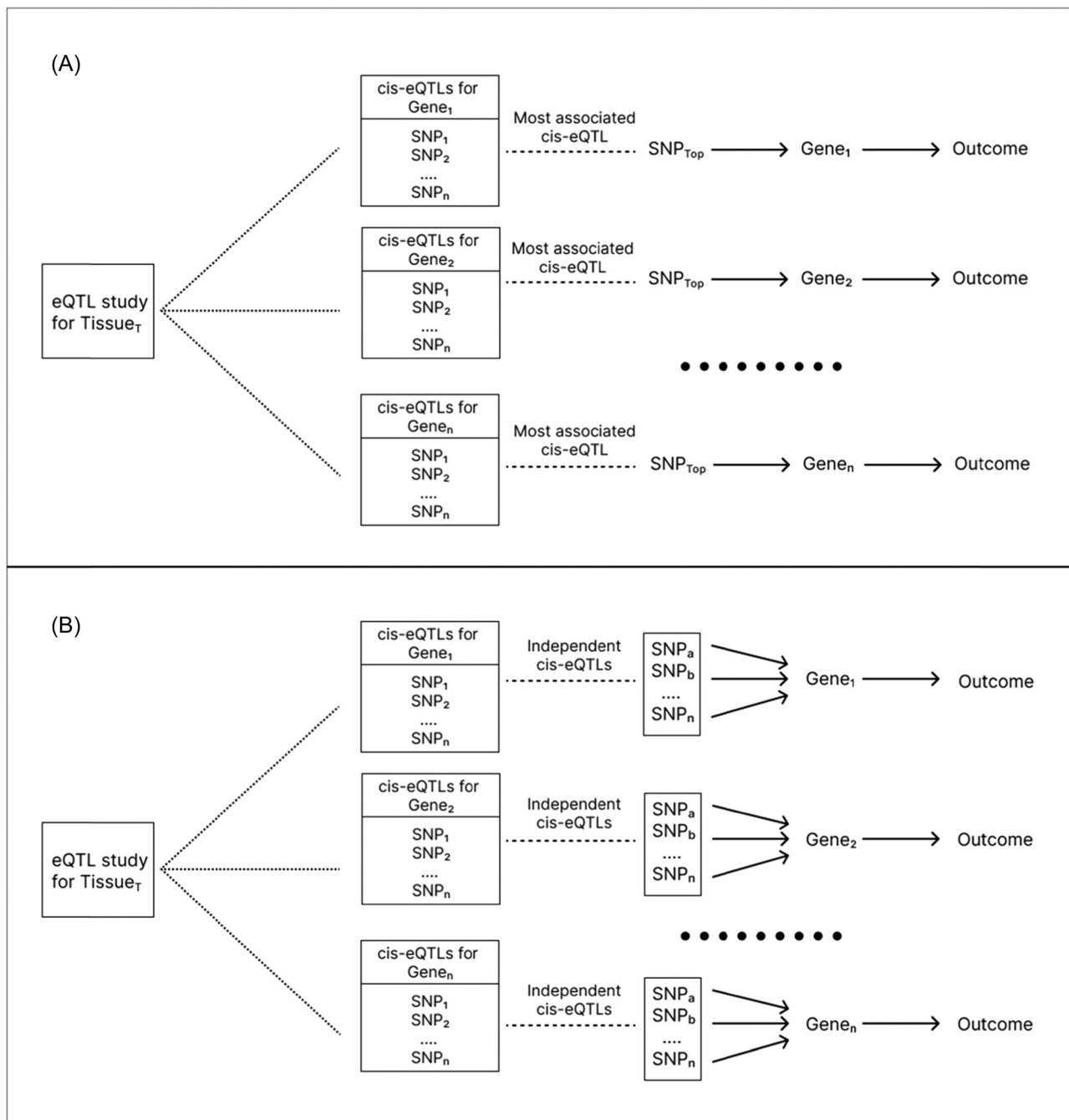
analysis to reduce such bias. By including multiple instruments, especially within the cis region of a probe, the likelihood of horizontal pleiotropy is diminished, and the statistical power of the analysis increases.<sup>17,18</sup> Detailed information of the SMR and multi-SNPs SMR has been described previously.<sup>8,9</sup>

For SMR, we followed the standard approach.<sup>8,9</sup> The analysis was conducted for each tissue independently. The cis-eQTL within the cis-region of a probe having the strongest association with the gene's expression was selected as the instrumental variable for the SMR analysis to evaluate the association between a gene's expression level in the specific tissue and the outcome, and only genes with at least one cis-eQTL with  $P_{eQTL} < 5E-8$  were included.<sup>8</sup> SNPs with an allele frequency  $< 0.01$  were removed, along with any SNPs having an allele frequency difference  $> 0.2$  between any of the three data sets (GWAS summary statistics, eQTL summary statistics, and linkage disequilibrium [LD] reference panel data).<sup>8</sup> The analysis was repeated

in all the 49 tissues with available eQTL data for all 3 COVID-19 outcomes

The heterogeneity in dependent instruments (HEIDI) test was used to distinguish pleiotropy from linkage.<sup>8</sup> HEIDI tests against the null hypothesis that the association detected by the SMR test is due to pleiotropy (i.e., due to a single causal variant as opposed to two variants in high LD with each other). For the HEIDI test, cis-eQTLs that were in strong LD ( $r^2 > 0.9$ ) or in weak LD ( $r^2 < 0.05$ ) were excluded. The HEIDI test was only conducted if the number of cis-eQTLs for the test was  $\geq 3$ .<sup>8</sup> Associations with a  $P_{HEIDI} < 0.01$  were considered to show significant heterogeneity and hence excluded.<sup>9</sup> To correct for multiple testing, false discovery rate (FDR)  $q$  value was used. FDR  $q$  value  $< 0.05$  was considered statistically significant.

In the sensitivity analysis using multi-SNPs SMR, we used all independent cis-eQTLs (SNPs with  $r^2 < 0.1$  with the top cis-eQTL)



**FIGURE 1** Illustration of SMR and multi-SNPs SMR. (A) Illustration of the primary analysis, SMR. For each tissue T, the most associated cis-eQTL for each gene (SNP<sub>top</sub>) was used as the genetic instrument to estimate the effect of the expression level of the gene on the outcome. (B) Illustration of the sensitivity analysis, multi-SNPs SMR. For each tissue T, all independent cis-eQTLs for a gene ( $r^2 < 0.1$  with the top cis-eQTL) which were significant ( $P_{eQTL} < 5E-8$ ) were used as the genetic instruments to estimate the effect of the expression level of the gene on the outcome. Both analyses were repeated for each of the 49 tissues and all 3 COVID-19 outcomes.

with  $p < 5E-8$  within the cis-region as instrumental variables.<sup>9</sup> The originally proposed threshold for defining independent cis-eQTL was  $r^2 < 0.9$ . However, this threshold may include SNPs with high LD (as defined as  $r^2 \geq 0.1$ ). We, therefore, adopted the  $r^2 < 0.1$  as the threshold for defining independent SNPs.<sup>19</sup> The reference panel of 503 European individuals from the 1000 genomes project (phase 3) was used to compute LD estimates for the analysis.<sup>20</sup> Only gene-trait

associations that show significant associations in both primary and sensitivity analyses were considered potential causal genes. FDR  $q$  value  $< 0.05$  was considered statistically significant. Data analyses were done using SMR v1.03 (<https://cnsngenomics.com/software/smr/>) and R version 4.1.0 (<https://www.r-project.org/>). LD reference panel was generated using BCFtools 1.11,<sup>21</sup> VCFtools 0.1.15,<sup>22</sup> and PLINK 1.90 (<https://www.cog-genomics.org/plink/1.9/>).<sup>23</sup>

To gain insight into the biological function of the potential causal genes, gene ontology (GO) enrichment analysis was conducted using g:Profiler (<https://biit.cs.ut.ee/gprofiler>)<sup>24</sup> to annotate the results. Correction for multiple testing was done using g:SCS (Set Counts and Sizes),<sup>25</sup> a method developed to estimate thresholds in complex functional profiling data such as GO.

## 3 | RESULTS

### 3.1 | Severe COVID-19

In the transcriptome-wide analysis of 49 tissues with severe COVID-19, 826 gene-trait associations were identified by univariable SMR analysis across 46 tissues (including blood) with FDR  $q$  value  $< 0.05$ . Thirty-nine of these associations were excluded as they were insignificant in the multi-SNPs SMR analysis. Of the remaining 787 associations, 478 did not pass the HEIDI test, resulting in 309 significant associations. The 309 associations were spread across 46 tissues (Supporting Information: Table S2), while the associations were mapped to 64 genes (Table 2), of which 38 were protein coding. Thirty-nine of the genes were identified in two or more tissues (Supporting Information: Table S3). The top five most associated protein-coding genes were Interferon Alpha and Beta Receptor Subunit 2 (*IFNAR2*), 2'-5'-Oligoadenylate Synthetase 3 (*OAS3*), mucin 1 (*MUC1*), Interleukin 10 Receptor Subunit Beta (*IL10RB*), and Napsin A Aspartic Peptidase (*NAPSA*). The tissue with the highest number of significant genes was esophagus mucosa ( $n = 17$ ) (Supporting Information: Table S4).

GO enrichment analysis (Table 3) for biological processes showed the genes were significantly enriched in seven GO terms relating to immune system regulation. The associated biological processes ( $p_{\text{adjusted}} < 0.05$ ) were driven by four genes (2'-5'-Oligoadenylate Synthetase 1 [*OAS1*], Tyrosine Kinase 2 [*TYK2*], *IFNAR2*, and *OAS3*) and related to type I interferon, interferon-gamma inducible protein 10 (IP-10) production, and chemokine (C-X-C motif) ligand 2 production.

### 3.2 | Hospitalized COVID-19

A total of 816 gene-trait associations were identified by univariable SMR analysis across eQTL in 49 tissues on hospitalized COVID-19 (FDR  $q$  value  $< 0.05$ ). Fifty-nine of these associations were excluded as they were insignificant in the multi-SNPs SMR analysis. Of the remaining 757 associations, 462 did not pass the HEIDI test, resulting in 295 significant associations. The 295 associations were spread across 49 tissues (Supporting Information: Table S5), while the associations were mapped to 63 genes (Table 4), of which 36 were protein coding. Thirty-eight genes were identified in two or more tissues (Supporting Information: Table S6). The top five most associated protein-coding genes were *IFNAR2*, *NAPSA*, *IL10RB*, alpha 1-3-N-acetylgalactosaminyltransferase and alpha 1-3-galactosyltransferase (*ABO*), and *OAS3*. Twenty-three genes

(14 protein coding and 9 non-protein coding) were unique to hospitalized COVID-19 (Supporting Information: Table S7). The tissue with the highest number of significant genes was blood ( $n = 11$ ; Supporting Information: Table S4).

GO enrichment analysis (Table 5) for biological processes showed the genes were significantly enriched in 12 GO terms relating to immune system regulation and homeostasis. The associated biological processes were driven by seven genes (*OAS1*, 2'-5'-Oligoadenylate Synthetase 2 [*OAS2*], *OAS3*, *TYK2*, *IFNAR2*, *NAPSA*, and Surfactant Protein D [*SFTPD*]). The associated biological processes ( $p_{\text{adjusted}} < 0.05$ ) were similar to those identified in the GO enrichment analysis of severe COVID-19, with additional broad biological processes related to nuclease regulation, surfactant and chemical homeostasis, and interferon-beta production.

### 3.3 | SARS-CoV-2 infection

We identified a total of 86 gene-trait associations which were significant in both SMR and multi-SNPs SMR while not being rejected by the HEIDI test. These were mapped to 20 genes across 38 tissues (Supporting Information: Table S8). Among the 20 genes, 14 had a significant association in more than one tissue, 16 were protein-coding and 13 genes were unique to SARS-CoV-2 infection (Supporting Information: Table S7). The top 5 protein-coding genes were Neurexophilin And PC-Esterase Domain Family Member 3 (*NXPE3*), SUMO Specific Peptidase 7 (*SEN7*), Centrosomal Protein 97 (*CEP97*), TUB Like Protein 2 (*TULP2*), and Netrin 5 (*NTN5*). No GO terms were found to be significantly enriched in the GO analysis of the significant causal genes for SARS-CoV-2 infection.

## 4 | DISCUSSION

In this study, we conducted the SMR analysis between 49 transcriptome-wide eQTL datasets as exposures and severe COVID-19 as the primary outcome. We identified a total of 309 gene-tissue associations and 64 potentially causal genes for severe COVID-19. Using hospitalized COVID-19 as the secondary outcome, similar genes and GO enrichment were observed, suggesting the findings observed in the severe COVID-19 analysis were robust. In addition, we identified 23 genes and 5 pathways that may be specific to hospitalized COVID-19, as well as 13 genes that may be specific to SARS-CoV-2 infection. These findings potentially explain the underlying mechanisms of severe COVID-19 and serve as potential therapeutic targets for the treatment of COVID-19.

A previous SMR study<sup>6</sup> found no genes to be significantly associated with severe COVID-19 but found one gene, *IFNAR2*, to be associated with hospitalized COVID-19 in the eQTL data set from the blood tissue. In the current study, *IFNAR2* was identified as a causal gene in the analysis of both severe and hospitalized COVID-19. The lack of associations in the previous study could be due to the use of the GWAS summary statistics from an earlier release of HGI data (release 3), and the

TABLE 2 Genes that were associated with severe COVID-19.

Gene	No. of tissues	Results in the most associated tissue			Tissue	Protein coding
		$\beta$	SE	p Value		
IFNAR2	6	-0.375	0.068	2.9E-08	Blood	Yes
OAS3	1	0.129	0.024	4.7E-08	Cells Transformed fibroblasts	Yes
MUC1	1	0.786	0.146	7.8E-08	Blood	Yes
IL10RB	3	-0.48	0.091	1.2E-07	Cells Transformed fibroblasts	Yes
NAPSA	2	-0.479	0.092	2.0E-07	Blood	Yes
KCNC3	4	1.151	0.222	2.2E-07	Blood	Yes
PLEKHM1	2	0.155	0.032	1.9E-06	Brain Cerebellar Hemisphere	Yes
OAS1	4	-0.384	0.081	2.1E-06	Skin Sun Exposed Lower leg	Yes
ARL17A	10	-0.102	0.022	2.3E-06	Brain Cortex	Yes
TYK2	4	0.362	0.077	2.5E-06	Adrenal Gland	Yes
NUTM2B	28	-0.133	0.03	8.0E-06	Skin Not Sun Exposed Suprapubic	Yes
WNT3	11	-0.178	0.04	9.5E-06	Pancreas	Yes
KANSL1	1	-0.383	0.087	1.1E-05	Cells Transformed fibroblasts	Yes
MUC5B	1	-0.324	0.075	1.6E-05	Lung	Yes
TRIM23	3	0.286	0.068	2.8E-05	Skin Sun Exposed Lower leg	Yes
CENPK	6	-0.343	0.084	4.4E-05	Cells Transformed fibroblasts	Yes
ARHGAP27	1	-0.234	0.057	4.4E-05	Brain Nucleus accumbens basal ganglia	Yes
RMI2	2	-0.189	0.047	6.1E-05	Esophagus Mucosa	Yes
ICAM5	2	-0.23	0.057	6.4E-05	Cells Transformed fibroblasts	Yes
ZNF528	30	-0.14	0.035	6.7E-05	Colon Sigmoid	Yes
HLA-DQA1	1	0.098	0.025	7.6E-05	Blood	Yes
ADAMTS6	1	0.132	0.033	8.1E-05	Testis	Yes
LRRC37A	7	-0.165	0.042	8.5E-05	Artery Coronary	Yes
ELF5	1	0.427	0.11	1.0E-04	Lung	Yes
CRHR1	1	-0.274	0.071	1.1E-04	Esophagus Muscularis	Yes
TOMM7	8	0.182	0.047	1.3E-04	Adipose Subcutaneous	Yes
SLC22A31	1	0.13	0.034	1.4E-04	Spleen	Yes
CCHCR1	1	0.253	0.066	1.4E-04	Stomach	Yes
ZGLP1	1	-0.353	0.094	1.7E-04	Brain Cerebellum	Yes
ICAM3	2	0.1	0.027	1.8E-04	Brain Amygdala	Yes
RASIP1	1	-0.096	0.026	1.9E-04	Cells Transformed fibroblasts	Yes
NSF	1	0.619	0.168	2.2E-04	Esophagus Mucosa	Yes
PPWD1	2	0.153	0.042	2.6E-04	Colon Transverse	Yes
HLA-DQB2	3	-0.077	0.021	2.8E-04	Brain Cortex	Yes
NOTCH4	1	-0.087	0.025	3.9E-04	Cells EBV-transformed lymphocytes	Yes
SNX31	3	0.084	0.024	4.8E-04	Brain Hypothalamus	Yes
SELE	1	-0.14	0.04	4.9E-04	Liver	Yes
HLA-DQB1	3	0.081	0.023	5.3E-04	Brain Hypothalamus	Yes

TABLE 2 (Continued)

Gene	No. of tissues	Results in the most associated tissue				Protein coding
		$\beta$	SE	p Value	Tissue	
RP11-119F19.2	22	-0.22	0.041	7.4E-08	Blood	No
RP11-259G18.3	10	-0.216	0.042	3.0E-07	Blood	No
NAPSB	31	-0.12	0.024	5.3E-07	Whole Blood	No
FAM22B	1	-0.175	0.036	1.2E-06	Blood	No
FAM215B	6	-0.177	0.037	1.3E-06	Thyroid	No
RP11-798G7.5	2	0.208	0.043	1.6E-06	Blood	No
DND1P1	5	-0.103	0.022	2.9E-06	Blood	No
RP11-259G18.2	10	-0.345	0.075	4.0E-06	Blood	No
KANSL1-AS1	5	-0.106	0.023	5.2E-06	Brain Cortex	No
BEND3P3	3	-0.307	0.068	5.7E-06	Blood	No
RP11-259G18.1	8	-0.17	0.038	7.1E-06	Esophagus Mucosa	No
RP11-506M13.3	23	-0.155	0.036	1.6E-05	Nerve Tibial	No
RP11-182L21.5	4	-0.208	0.05	2.9E-05	Nerve Tibial	No
LRRC37A17P	1	0.26	0.062	3.1E-05	Esophagus Mucosa	No
AC091132.1	2	0.437	0.106	3.7E-05	Esophagus Mucosa	No
RP11-707O23.5	2	-0.088	0.021	4.2E-05	Brain Substantia nigra	No
CTD-2020K17.1	1	0.234	0.059	7.0E-05	Brain Cerebellum	No
RPS26P8	2	-0.191	0.049	8.2E-05	Adipose Subcutaneous	No
CTD-2116N20.1	1	0.152	0.039	9.1E-05	Testis	No
CTC-534A2.2	1	-0.548	0.141	1.0E-04	Blood	No
MAPT-AS1	2	0.169	0.044	1.0E-04	Brain Nucleus accumbens basal ganglia	No
RP11-798G7.8	1	-0.135	0.035	1.3E-04	Brain Cerebellar Hemisphere	No
XXbac-BPG299F13.17	2	0.197	0.052	1.4E-04	Esophagus Gastroesophageal Junction	No
AC005682.5	1	-0.201	0.053	1.6E-04	Thyroid	No
RP11-46C24.3	1	-0.117	0.032	2.5E-04	Brain Cerebellum	No
RP11-707M1.1	1	-0.143	0.039	2.8E-04	Brain Caudate basal ganglia	No

TABLE 3 GO biological process enrichment analysis of the significant genes for severe COVID-19

Term name	Term id	Adjusted p value	Intersections
Type I interferon signaling pathway	GO:0060337	0.004	OAS1, TYK2, IFNAR2, OAS3
Cellular response to type I interferon	GO:0071357	0.005	OAS1, TYK2, IFNAR2, OAS3
Response to type I interferon	GO:0034340	0.008	OAS1, TYK2, IFNAR2, OAS3
Negative regulation of IP-10 production	GO:0071659	0.015	OAS1, OAS3
IP-10 production	GO:0071612	0.030	OAS1, OAS3
Regulation of IP-10 production	GO:0071658	0.030	OAS1, OAS3
Negative regulation of chemokine (C-X-C motif) ligand 2 production	GO:2000342	0.030	OAS1, OAS3



TABLE 4 Genes that were associated with hospitalized COVID-19

Gene	No. of tissues	Results in the most associated tissue			Tissue	Protein coding
		$\beta$	SE	p Value		
IFNAR2	5	-0.272	0.044	5.3E-10	Skin Sun Exposed Lower leg	Yes
NAPSA	3	-0.317	0.053	2.5E-09	Blood	Yes
IL10RB	3	0.329	0.058	1.9E-08	Muscle Skeletal	Yes
ABO	1	0.298	0.053	1.9E-08	Artery Tibial	Yes
OAS3	3	0.078	0.014	2.8E-08	Cells Transformed fibroblasts	Yes
ARL17A	10	-0.072	0.013	4.6E-08	Brain Cortex	Yes
MUC1	1	0.41	0.076	5.8E-08	Blood	Yes
PLEKHM1	4	0.095	0.019	5.2E-07	Brain Cerebellar Hemisphere	Yes
ELF5	1	0.333	0.067	5.8E-07	Lung	Yes
KANSL1	6	-0.268	0.054	7.0E-07	Cells Transformed fibroblasts	Yes
KCNC3	2	0.069	0.014	1.4E-06	Spleen	Yes
WNT3	2	-0.108	0.022	1.4E-06	Pancreas	Yes
OAS1	4	-0.302	0.063	1.5E-06	Adipose Subcutaneous	Yes
LRRC37A2	5	-0.091	0.019	1.9E-06	Vagina	Yes
ARHGAP27	1	-0.167	0.037	6.1E-06	Brain Nucleus accumbens basal ganglia	Yes
MUC5B	1	-0.191	0.042	6.2E-06	Lung	Yes
SPPL2C	1	-0.141	0.032	8.5E-06	Brain Cerebellar Hemisphere	Yes
LRRC37A	24	-0.052	0.012	1.4E-05	Brain Cerebellar Hemisphere	Yes
CCHCR1	5	0.111	0.026	1.5E-05	Blood	Yes
NXPE3	2	0.08	0.019	2.0E-05	Esophagus Muscularis	Yes
CRHR1	1	-0.199	0.047	2.3E-05	Esophagus Muscularis	Yes
ICAM3	11	0.187	0.044	2.6E-05	Adipose Visceral Omentum	Yes
MAPT	1	-0.181	0.044	3.2E-05	Skin Not Sun Exposed Suprapubic	Yes
ARL17B	2	0.091	0.022	3.5E-05	Whole Blood	Yes
TCF19	1	0.085	0.021	3.8E-05	Blood	Yes
RAB2A	1	0.252	0.062	4.5E-05	Artery Tibial	Yes
CYP4B1	5	0.22	0.054	4.6E-05	Nerve Tibial	Yes
TYK2	2	0.492	0.121	4.8E-05	Skin Sun Exposed Lower leg	Yes
OAS2	1	0.385	0.096	5.5E-05	Blood	Yes
ZNF778	2	0.054	0.013	5.6E-05	Pancreas	Yes
NSF	2	0.392	0.098	5.9E-05	Esophagus Mucosa	Yes
SLC22A31	10	0.066	0.017	9.0E-05	Esophagus Mucosa	Yes
BMP1	1	0.23	0.06	1.1E-04	Artery Tibial	Yes
AVEN	1	0.104	0.027	1.3E-04	Heart Atrial Appendage	Yes
SCAMP5	2	0.098	0.026	1.5E-04	Skin Sun Exposed Lower leg	Yes
SFTPD	1	0.076	0.021	3.3E-04	Pituitary	Yes
RP11-259G18.3	18	-0.072	0.012	4.0E-10	Whole Blood	No
KANSL1-AS1	12	-0.076	0.013	1.5E-09	Colon Transverse	No
RP11-259G18.2	4	-0.243	0.042	8.5E-09	Blood	No



TABLE 4 (Continued)

Gene	No. of tissues	Results in the most associated tissue				Protein coding
		$\beta$	SE	p Value	Tissue	
LRRC37A4P	4	0.082	0.015	4.1E-08	Liver	No
CRHR1-IT1	1	-0.115	0.021	6.6E-08	Brain Putamen basal ganglia	No
DND1P1	5	-0.061	0.011	1.0E-07	Artery Coronary	No
NAPSB	28	-0.074	0.014	1.5E-07	Whole Blood	No
RP11-259G18.1	19	-0.123	0.024	2.7E-07	Esophagus Mucosa	No
FAM215B	7	-0.108	0.021	4.5E-07	Thyroid	No
ATP5O	1	0.154	0.032	1.4E-06	Lung	No
MAPT-AS1	2	0.195	0.042	3.9E-06	Muscle Skeletal	No
RP11-707O23.5	1	-0.059	0.013	3.9E-06	Brain Substantia nigra	No
AC091132.1	3	0.318	0.069	4.4E-06	Esophagus Mucosa	No
RP11-798G7.5	14	0.106	0.024	1.0E-05	Brain Cerebellar Hemisphere	No
RP11-798G7.6	20	0.07	0.016	1.2E-05	Thyroid	No
CTD-2020K17.1	1	0.161	0.037	1.8E-05	Brain Cerebellum	No
RPS26P8	2	-0.131	0.031	2.1E-05	Breast Mammary Tissue	No
BEND3P3	1	-0.153	0.036	2.2E-05	Blood	No
RP11-798G7.8	1	-0.09	0.021	2.9E-05	Brain Cerebellar Hemisphere	No
RP11-119F19.2	1	-0.087	0.022	6.2E-05	Blood	No
PDCL3P4	13	-0.109	0.027	6.3E-05	Heart Left Ventricle	No
HLA-J	3	-0.094	0.024	8.7E-05	Colon Transverse	No
ZBTB11-AS1	1	-0.166	0.043	9.9E-05	Muscle Skeletal	No
CTD-3092A11.2	1	-0.094	0.024	1.1E-04	Lung	No
RP11-322D14.1	1	0.082	0.021	1.5E-04	Whole Blood	No
AL133481.1	2	0.041	0.011	1.8E-04	Brain Cerebellum	No
IL10RB-AS1	1	0.06	0.016	2.6E-04	Brain Cerebellum	No

TABLE 5 GO biological process enrichment analysis of the significant genes for hospitalized COVID-19

Term name	Term id	Adjusted p value	Intersections
Type I interferon signaling pathway	GO:0060337	3.2E-05	OAS1, OAS2, OAS3, IFNAR2, TYK2
Cellular response to type I interferon	GO:0071357	3.8E-05	OAS1, OAS2, OAS3, IFNAR2, TYK2
Response to type I interferon	GO:0034340	6.7E-05	OAS1, OAS2, OAS3, IFNAR2, TYK2
Regulation of ribonuclease activity	GO:0060700	2.6E-04	OAS1, OAS2, OAS3
Surfactant homeostasis	GO:0043129	0.003	OAS1, NAPSA, SFTPD
Chemical homeostasis within a tissue	GO:0048875	0.004	OAS1, NAPSA, SFTPD
Regulation of nuclease activity	GO:0032069	0.006	OAS1, OAS2, OAS3
Negative regulation of IP-10 production	GO:0071659	0.010	OAS1, OAS3
IP-10 production	GO:0071612	0.020	OAS1, OAS3
Regulation of IP-10 production	GO:0071658	0.020	OAS1, OAS3
Negative regulation of chemokine (C-X-C motif) ligand 2 production	GO:2000342	0.020	OAS1, OAS3
Positive regulation of interferon-beta production	GO:0032728	0.048	OAS1, OAS2, OAS3

analysis was conducted using only eQTL summary statistics from blood and lung, hence leading to limited statistical power. Another transcriptome-wide SMR study<sup>7</sup> was conducted using eQTL and mQTL data from lung and whole blood with the hospitalized COVID-19 GWAS from HGI release 5, which identified associations involving seven protein-coding genes *TYK2*, *IFNAR2*, *OAS1*, *OAS3*, *XCR1*, *CCR5*, and *MAPT*. The current study found all but two (*XCR1* and *CCR5*) of these genes to be significantly associated with hospitalized COVID-19. Compared to these two studies, the current study had the highest power and identified the largest number of potentially causal genes for severe and hospitalized COVID-19 using the most recent release of HGI data and 49 eQTL data sets. In total, we identified 64 protein-coding genes (Supporting Information: Table S7). Of these, several have previously been found to be associated with severe COVID-19, hospitalized COVID-19 and SARS-CoV-2 infection through various experimental techniques and definitions of the COVID-19 outcomes: these include *ABO*, *OAS1*, *OAS2*, *OAS3*, *TYK2*, *IFNAR2*, *IL10RB*, *MAPT*, *ARL17A*, *ARL17B*, *CCHCR1*, *CEP97*, *LRRC37A*, *LRRC37A2*, *NXPE3*, *ICAM5*, *MUC1*, *NTN5*, *ELF5*, *FUT2*, *KANSL1*, *MUC5B*, *SFTPD*, *SLC22A31*, *SELE*, *SENP7*, *NAPSA*, *NOTCH4*, *PLEKHA4*, *WNT3*, *NSF*, *TOMM7*, *HLA-DQB2*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1*, *TULP2*, and *TCF19*.<sup>4,6,7,26-37</sup> On top of these, our analysis has identified several novel genes which are potentially causally associated with severe COVID-19, hospitalized COVID-19 and SARS-CoV-2 infection; 15 genes were associated with severe COVID-19 (*ADAMTS6*, *CENPK*, *NUTM2B*, *PPWD1*, *RM12*, *RASIP1*, *SNX31*, *TRIM23*, *ARHGAP27*, *CRHR1*, *ICAM3*, *KCNC3*, *PLEKHM1*, *ZGLP1*, and *ZNF528*), 12 with hospitalized COVID-19 (*AVEN*, *BMP1*, *CYP4B1*, *RAB2A*, *SCAMP5*, *SPPL2C*, *ARHGAP27*, *CRHR1*, *ICAM3*, *KCNC3*, *PLEKHM1*, and *ZNF778*), and 5 were associated with SARS-CoV-2 infection (*HSD17B14*, *MED24*, *CLK2*, *RASIP1*, and *MAMSTR*). We also identified 39 non-protein-coding genes (Supporting Information: Table S7). Of these, only *DND1P1*, *KANSL1-AS1*, *LRRC37A4P*, *LCN1P1*, *MAPT-AS1*, *IL10RB-AS1*, *PDCL3P4*, *ZBTB11-AS1*, and *CRHR1-IT1* have been reported previously.<sup>31,36,38</sup> The remaining noncoding genes are novel.

The top 10 most associated protein-coding genes with severe COVID-19 were also found to be associated with hospitalized COVID-19 (*IFNAR2*, *OAS3*, *MUC1*, *IL10RB*, *NAPSA*, *KCNC3*, *PLEKHM1*, *OAS1*, *ARL17A*, and *TYK2*). Four of these, *OAS1*, *IFNAR2*, *OAS3*, and *TYK2* were the genes that drove the enrichment of biological processes in the GO enrichment analysis. The enriched biological processes were related to type I interferon signaling, IP-10 production, and CXCL2 production. *IFNAR2* and *TYK2* are involved in the signaling pathway of type I interferons,<sup>39</sup> while *OAS1* and *OAS3* are known to regulate type I interferons and chemokines.<sup>40</sup> Type I and III interferons have previously been implicated in COVID-19 and may have a detrimental and beneficial effect on SARS-CoV-2 replication.<sup>41</sup> These interferons are a central part of the innate antiviral response. However, the response of these interferons was found to be diminished and delayed in patients with COVID-19 and observed only in a small number of patients as their infection reached critical stages.<sup>42,43</sup> It has been suggested that type I and III interferons confer early protection but late amplification of the disease.<sup>43</sup> *IL10RB* belongs to the cytokine receptor family and has previously been implicated in COVID-19.<sup>36</sup> Two of the top 10 genes,

*KCNC3* and *PLEKHM1*, are novel genes associated with severe COVID-19 and may play a causal role in the disease through the immune system. *KCNC3* is a gated potassium voltage channel that may affect the immune response by inhibiting T cell activation.<sup>44</sup> *PLEKHM1* regulates autophagosome-lysosome fusion, and depletion of *PLEKHM1* enhances the presentation of MHC class 1 molecules, thereby affecting the immune response.<sup>45</sup> Taken together, these genes highlight the role of immune system dysregulation in severe COVID-19 and provide further insight into the potentially causal role of certain genes in this process.

In the supplementary analysis using SARS-CoV-2 infection as an outcome, we identified four genes in common with severe COVID-19 (*OAS1*, *IFNAR2*, *RASIP1*, and *CCHCR1*) and seven in common with hospitalized COVID-19 (*OAS1*, *IFNAR2*, *CCHCR1*, *PDCL3P4*, *NXPE3*, and *ZBTB11-AS1*). The remaining 13 unique SARS-CoV-2 infection genes may be associated with SARS-CoV-2 infection, but not a severe response as a consequence. In the GO enrichment analysis, no significant GO enrichment was found. This may be due to a lower statistical power of the SARS-CoV-2 infection GWAS. The asymptomatic nature of COVID-19 and self-reported phenotype may lead to misclassification of the SARS-CoV-2 cases in the GWAS of SARS-CoV-2 infection. As a result, a higher negative predictive value and lower sensitivity of covid infection diagnosis could be observed, resulting in a lower statistical power of the GWAS.<sup>46</sup>

The present study also identified several novel genes associated with severe COVID-19 that may underlie the pathogenesis of severe COVID-19. *ARHGAP27*, *RASIP1*, and *CENPK* are involved in signaling by the Rho GTPase pathway and may increase COVID-19 infection through dysregulation of the pathway, affecting cytoskeletal dynamics.<sup>47,48</sup> *TRIM23* is known to induce autophagy in response to viral infections.<sup>49</sup> *ZNF528* and *ZGLP1* are involved in transcription and may function as regulators in pathways affecting the severity of COVID-19 infections.<sup>50</sup> *ICAM3* expression in several brain tissues was associated with an increased risk of severe and hospitalized COVID-19. Previously, *ICAM3* was found to increase lactate dehydrogenase levels, and lower total white blood cell counts in patients infected with severe acute respiratory syndrome (SARS)<sup>51</sup> which may similarly affect those infected by SARS-CoV-2.

Our study has important clinical implications. The ongoing COVID-19 pandemic is unprecedented in modern times, while severe COVID-19 is particularly associated with poor prognosis. Our findings provide further insight into the mechanism of severe COVID-19. The current study has identified several novel genes which have a potentially causal association with severe COVID-19. Notably, some of the genes identified in the current study have been implicated in COVID-19 infection and progression in previous studies with different study designs and definitions, which serve as positive controls and provide evidence of the robustness of our findings. Thus, the genes identified in this study provide direction for further functional validation studies and shed light on the development of therapeutic agents for treating severe COVID-19.

Our study also has several strengths. First, we utilized 49 different tissue types, allowing a comprehensive evaluation of gene expression with severe COVID-19 across tissues. Second, we utilized the multi-SNPs SMR as the sensitivity analysis, providing further robust evidence of association and reducing the false positive rates. Finally, we used GWAS from release 6 of COVID-19 HGI data in the analysis providing larger statistical power compared to previous studies. Nevertheless, there are limitations. First, the GWAS did not control for all confounders, such as diabetes, obesity, and other risk factors of severe COVID-19, which may affect the outcome. However, the nature of the study using eQTL is less likely to be affected by confounders. Second, although several associations only appeared in one tissue, it may not be a true tissue-specific association since many eQTLs are underpowered due to the low sample size. Thus, cautious interpretation is required. Third, although we conducted additional sensitivity analysis to reduce bias, we cannot eliminate the possibility of the association being due to horizontal pleiotropy. For this reason, we consider the identified genes as “potentially causal.” Further study is required to validate the role of these genes on covid outcomes.

In conclusion, we identified 64 genes that are potentially causally associated with severe COVID-19, of which 38 genes are novel. These results lead to a better understanding of the mechanism of severe COVID-19 and show potential therapeutic targets for the treatment and/or reduction of symptoms and mortality after SARS-CoV-2 infection. Further research into the relationship between the genes put forward here, and severe COVID-19 is warranted.

#### AUTHOR CONTRIBUTIONS

Ching-Lung Cheung designed the study. Suhas Krishnamoorthy gathered data, conducted the analysis, and drafted the manuscript. Ching-Lung Cheung and Gloria H.-Y. Li revised the manuscript for intellectual content. All authors read and approved the final manuscript.

#### ACKNOWLEDGMENT

This study was supported by AIR@InnoHK administered by the Innovation and Technology Commission.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### DATA AVAILABILITY STATEMENT

All data that support the findings of this paper are available online. GWAS data are available from release 6 of COVID-19 HGI data (<https://www.covid19hg.org/results/r6/>). eQTL data from version 7 of the GTEx Project were downloaded in SMR binary (BESD) format (<https://yanglab.westlake.edu.cn/software/smr/#DataResource>), eQTL data from eQTL-Gen were downloaded in BESD format (<https://www.eqtngen.org/cis-eqtls.html>) and LD reference panel data was downloaded from 1000 genomes project phase 3 (<https://www.internationalgenome.org/data-portal/sample>). The data generated from this paper are available in the tables and Supporting Information: tables.

#### ORCID

Ching-Lung Cheung  <http://orcid.org/0000-0002-6233-9144>

#### REFERENCES

1. Thomson B. The COVID-19 pandemic: A global natural experiment. *Circulation*. 2020;142(1):14-16.
2. World Health Organization. *WHO COVID-19 Dashboard*. Geneva: World Health Organization; 2022.
3. Casanova JL, Su HC. A global effort to define the human genetics of protective immunity to SARS-CoV-2 infection. *Cell*. 2020;181(6):1194-1199.
4. COVID-19 Host Genetics Initiative. 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 Human Genet*. 2020;28(6):715-718.
5. Chapman SJ, Hill AV. Human genetic susceptibility to infectious disease. *Nat Rev Genet*. 2012;13(3):175-188.
6. Liu D, Yang J, Feng B, Lu W, Zhao C, Li L. Mendelian randomization analysis identified genes pleiotropically associated with the risk and prognosis of COVID-19. *J Infect*. 2021;82(1):126-132.
7. Baranova A, Cao H, Zhang F. Unraveling risk genes of COVID-19 by multi-omics integrative analyses. *Front Med*. 2021;8:738687.
8. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48(5):481-487.
9. Wu Y, Zeng J, Zhang F, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun*. 2018;9(1):918.
10. GTEx Consortium; Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204-213.
11. Vösa U, Claringbould A, Westra H-J, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv*. 2018:447367.
12. Jain U. Effect of COVID-19 on the organs. *Cureus*. 2020;12(8):e9540.
13. Niemi MEK, Karjalainen J, Liao RG, et al. Mapping the human genetic architecture of COVID-19. *Nature*. 2021;600(7889):472-477.
14. Sobczyk MK, Gaunt TR. The effect of circulating zinc, selenium, copper and vitamin K(1) on COVID-19 outcomes: a Mendelian randomization study. *Nutrients*. 2022;14(2):233.
15. Huffman JE, Butler-Laporte G, Khan A, et al. Multi-ancestry fine mapping implicates OAS1 splicing in risk of severe COVID-19. *Nature Genet*. 2022;54(2):125-127.
16. Downes DJ, Cross AR, Hua P, et al. Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. *Nature Genet*. 2021;53(11):1606-1615.
17. Swerdlow DI, Kuchenbaecker KB, Shah S, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int J Epidemiol*. 2016;45(5):1600-1616.
18. Palmer TM, Lawlor DA, Harbord RM, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res*. 2012;21(3):223-242.
19. Liu Y, Shen H, Greenbaum J, et al. Gene expression and RNA splicing imputation identifies novel candidate genes associated with osteoporosis. *J Clin Endocrinol Metab*. 2020;105(12):e4742-e4757.
20. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
21. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.
22. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156-2158.

23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575.
24. Raudvere U, Kolberg L, Kuzmin I, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 2019;47(W1):W191-W198.
25. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* 2007;35:W193-W200.
26. Pairo-Castineira E, Clohisey S, Klaric L, et al. Genetic mechanisms of critical illness in COVID-19. *Nature.* 2021;591(7848):92-98.
27. Ellinghaus D, Degenhardt F, Bujanda L, et al. Genomewide association study of severe Covid-19 with respiratory failure. *N Engl J Med.* 2020;383(16):1522-1534.
28. Wu L, Zhu J, Liu D, Sun Y, Wu C. An integrative multiomics analysis identifies putative causal genes for COVID-19 severity. *Genet Med Off J Am Coll Med Genet.* 2021;23(11):2076-2086.
29. van Moersel CHM, van der Vis JJ, Duckworth A, et al. The MUC5B promoter polymorphism associates with severe COVID-19 in the European population. *Front Med.* 2021;8:668024.
30. Kousathanas A, Pairo-Castineira E, Rawlik K, et al. Whole genome sequencing reveals host factors underlying critical Covid-19. *Nature.* 2022;607:97-103.
31. D'antonio M, Nguyen JP, Arthur TD, et al. SARS-CoV-2 susceptibility and COVID-19 disease severity are associated with genetic variants affecting gene expression in a variety of tissues. *Cell Rep.* 2021;37(7):110020.
32. Castelli EC, de Castro MV, Naslavsky MS, et al. MHC variants associated with symptomatic versus asymptomatic SARS-CoV-2 infection in highly exposed individuals. *Front Immunol.* 2021;12:742881.
33. Hu J, Li C, Wang S, Li T, Zhang H. Genetic variants are identified to increase risk of COVID-19 related mortality from UK Biobank data. *Hum Genomics.* 2021;15(1):10.
34. de Sousa E, Ligeiro D, Lérias JR, et al. Mortality in COVID-19 disease patients: Correlating the association of major histocompatibility complex (MHC) with severe acute respiratory syndrome 2 (SARS-CoV-2) variants. *Int J Infect Dis Off Publ Int Soc Infect Dis.* 2020;98:454-459.
35. Harb H, Benamar M, Lai PS, et al. Notch4 signaling limits regulatory T-cell-mediated tissue repair and promotes severe lung inflammation in viral infections. *Immunity.* 2021;54(6):1186-1199.e7.
36. Schmiedel BJ, Rocha J, Gonzalez-Colin C, et al. COVID-19 genetic risk variants are associated with expression of multiple genes in diverse immune cell types. *Nat Commun.* 2021;12(1):6760.
37. Pamos AB, Millischer V, Menon DK, et al. Proteome-wide Mendelian randomization identifies causal links between blood proteins and severe COVID-19. *PLoS Genet.* 2022;18(3):e1010042.
38. Fullard JF, Lee HC, Voloudakis G, et al. Single-nucleus transcriptome analysis of human brain immune response in patients with severe COVID-19. *Genome Med.* 2021;13(1):118.
39. Duncan CJ, Mohamad SM, Young DF, et al. Human IFNAR2 deficiency: Lessons for antiviral immunity. *Sci Transl Med.* 2015;7(307):307ra154.
40. Lee WB, Choi WY, Lee DH, Shim H, Kim-Ha J, Kim YJ. OAS1 and OAS3 negatively regulate the expression of chemokines and interferon-responsive genes in human macrophages. *BMB Rep.* 2019;52(2):133-138.
41. Sa Ribero M, Jouvenet N, Dreux M, Nisole S. Interplay between SARS-CoV-2 and the type I interferon response. *PLoS Pathog.* 2020;16(7):e1008737.
42. Galani IE, Rovina N, Lampropoulou V, et al. Untuned antiviral immunity in COVID-19 revealed by temporal type I/III interferon patterns and flu comparison. *Nat Immunol.* 2021;22(1):32-40.
43. Paludan SR, Mogensen TH. Innate immunological pathways in COVID-19 pathogenesis. *Sci Immunol.* 2022;7(67):eabm5505.
44. Koo GC, Blake JT, Talento A, et al. Blockade of the voltage-gated potassium channel Kv1.3 inhibits immune responses in vivo. *J Immunol.* 1997;158(11):5120-5128.
45. McEwan DG, Popovic D, Gubas A, et al. PLEKHM1 regulates autophagosome-lysosome fusion through HOPS complex and LC3/GABARAP proteins. *Mol Cell.* 2015;57(1):39-54.
46. Duan R, Cao M, Wu Y, et al. An empirical study for impacts of measurement errors on EHR based association studies. *AMIA Annu Symp Proc.* 2016;2016:1764-1773.
47. Katoh Y, Katoh M. Identification and characterization of ARHGAP27 gene in silico. *Int J Mol Med.* 2004;14:943-947.
48. Post A, Pannekoek WJ, Ross SH, Verlaan I, Brouwer PM, Bos JL. Rasip1 mediates Rap1 regulation of Rho in endothelial barrier function through ArhGAP29. *Proc Natl Acad Sci.* 2013;110(28):11427-11432.
49. Sparrer KMJ, Gableske S, Zurenski MA, et al. TRIM23 mediates virus-induced autophagy via activation of TBK1. *Nat Microbiol.* 2017;2(11):1543-1557.
50. Cassandri M, Smirnov A, Novelli F, et al. Zinc-finger proteins in health and disease. *Cell Death Discov.* 2017;3(1):17071.
51. Chan KY, Ching JC, Xu MS, et al. Association of ICAM3 genetic variant with severe acute respiratory syndrome. *J Infect Dis.* 2007;196(2):271-280.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Krishnamoorthy S, Li GH-Y, Cheung C-L. Transcriptome-wide summary data-based Mendelian randomization analysis reveals 38 novel genes associated with severe COVID-19. *J Med Virol.* 2022;1-12.  
doi:10.1002/jmv.28162