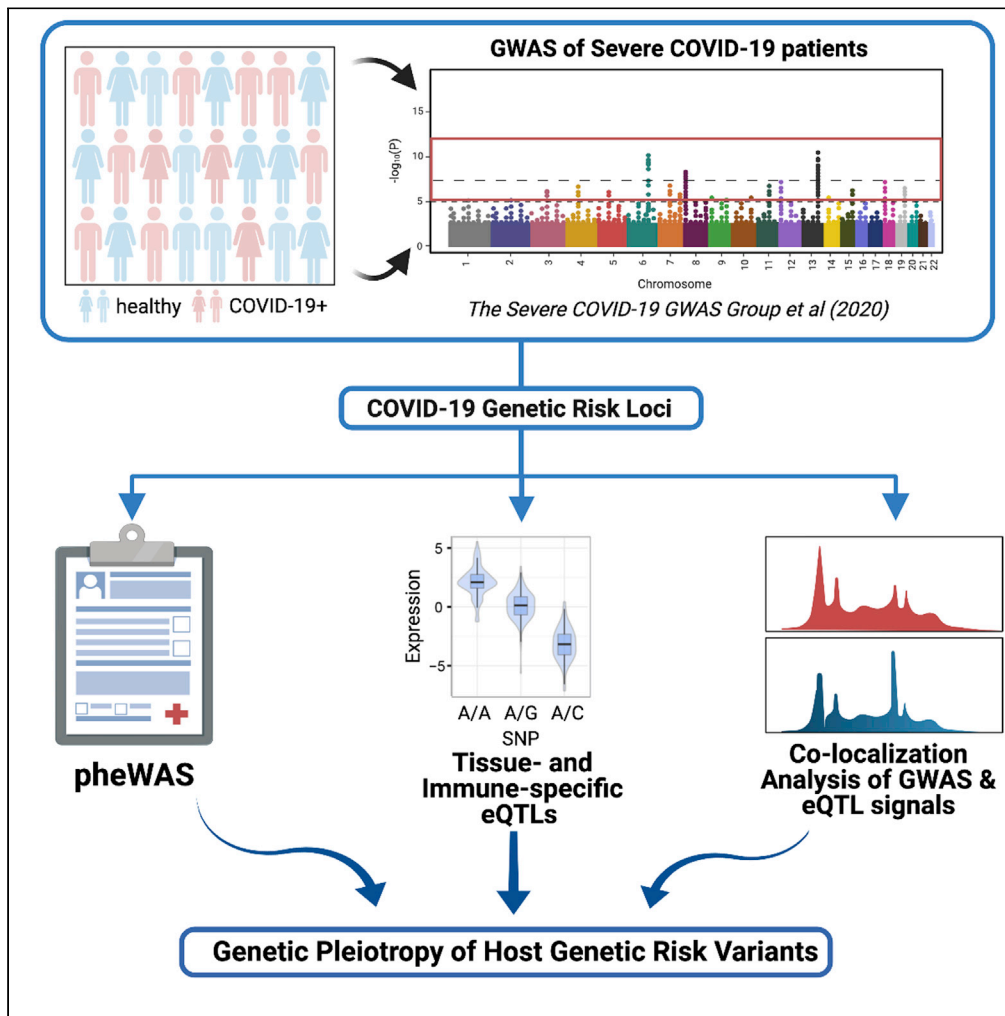


Article

Phenome-wide and expression quantitative trait locus associations of coronavirus disease 2019 genetic risk loci



Chang Yoon Moon, Brian M. Schilder, Towfique Raj, Kuan-lin Huang

kuan-lin.huang@mssm.edu

**Highlights**

Immune, metabolic, and respiratory traits are associated with severe COVID-19 loci

Tissue- and immune-subtype-specific eQTL analyses uncover associated immune genes

Immune gene eQTLs colocalize with COVID-19 GWAS associations

Macrophages of patients with severe COVID-19 differentially express eQTL-linked genes

Moon et al., iScience 24, 102550  
June 25, 2021 © 2021 The Authors.  
<https://doi.org/10.1016/j.isci.2021.102550>



## Article

## Phenome-wide and expression quantitative trait locus associations of coronavirus disease 2019 genetic risk loci

Chang Yoon Moon,<sup>1,2,3,4</sup> Brian M. Schilder,<sup>1,4,5,6,7</sup> Towfique Raj,<sup>1,4,5,6,7</sup> and Kuan-lin Huang<sup>1,2,3,4,8,\*</sup>

## SUMMARY

While several genes and clinical traits have been associated with higher risk of severe coronavirus disease 2019 (COVID-19), how host genetic variants may interact with these parameters and contribute to severe disease is still unclear. Herein, we performed phenome-wide association study, tissue and immune-cell-specific expression quantitative trait locus (eQTL)/splicing quantitative trait locus, and colocalization analyses for genetic risk loci suggestively associated with severe COVID-19 with respiratory failure. Thirteen phenotypes/traits were associated with the severe COVID-19-associated loci at the genome-wide significance threshold, including monocyte counts, fat metabolism traits, and fibrotic idiopathic interstitial pneumonia. In addition, we identified tissue and immune subtype-specific eQTL associations affecting 48 genes, including several ones that may directly impact host immune responses, colocalized with the severe COVID-19 genome-wide association study associations, and showed altered expression in single-cell transcriptomes. Collectively, our work demonstrates that host genetic variations associated with multiple genes and traits show genetic pleiotropy with severe COVID-19 and may inform disease etiology.

## INTRODUCTION

Upon emerging from Wuhan, China, in December 2019, the outbreak of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), has become a global pandemic, infecting, as of August 2020, more than 24 million people worldwide (Dong et al., 2020). While most infected individuals either are asymptomatic or present with mild flu-like symptoms (i.e., fever, cough, sore throat, malaise, and so on), few experience severe manifestations of COVID-19 that span multiple organ systems (Chen et al., 2020; Guan et al., 2020; Huang et al., 2020). Although several clinical observational studies have identified several demographic and clinical factors associated with increased disease severity and mortality such as sex, age, and other comorbidities such as cancer, chronic kidney disease, and obesity, which and how host genetic factors may contribute to the pathogenesis and pathophysiology of severe COVID-19 are poorly understood (Williamson et al., 2020). While a recent effort by the severe COVID-19 genome-wide association study (GWAS) group identified two significant loci (3p21.31 and 9q34.2) and multiple suggestive host genetic loci associated with increased risk of suffering from severe COVID-19 with respiratory failure, the functional effects of these loci and how they may influence disease severity remain unclear (Ellinghaus et al., 2020).

A phenome-wide association study (PheWAS) is a genotype-to-phenotype approach to identify associations between a particular genetic variant and a wide range of human traits (Hebbring, 2014). Since the first published PheWAS in 2010 (Denny et al., 2010), the PheWAS has been used to interrogate genetic pleiotropy, facilitate drug discovery and repositioning and discover novel genetic etiologies of complex and polygenic disorders (Diogo et al., 2018; Hebbring, 2014; Semmes et al., 2020; Zhao et al., 2018) and systematic compilations of GWAS results, such as that recently undertaken by GWAS Atlas, have facilitated the PheWAS (Watanabe et al., 2019). Considering the multiorgan involvement and manifold individual variations of severe COVID-19, it is likely that host susceptibility to severe COVID-19 is determined by multiple genetic factors. By connecting variants associated with both severe COVID-19 and other diseases or phenotypes, PheWAS analysis can use genetic pleiotropy to validate known comorbidities and unveil novel biological insights. Furthermore, PheWAS may offer new insights into the genetic etiology of severe COVID-19 by leveraging known biology behind seemingly disparate phenotypes.

<sup>1</sup>Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>2</sup>Center for Transformative Disease Modeling, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>3</sup>Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>4</sup>Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>5</sup>Nash Family Department of Neuroscience & Friedman Brain Institute, New York, NY, USA

<sup>6</sup>Ronald M. Loeb Center for Alzheimer's disease, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>7</sup>Estelle and Daniel Maggin Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>8</sup>Lead contact

\*Correspondence: kuan-lin.huang@mssm.edu  
<https://doi.org/10.1016/j.isci.2021.102550>



In addition to PheWAS that can link categorical phenotypes, tissue-specific quantitative trait locus (QTL) analysis can help to prioritize gene-disease connections for mechanistic investigations. With the advent of immense functional genomic databases such as GTEx (GTEx Consortium et al., 2020), the transcriptional effects of genetic variants can be investigated at tissue-specific resolution. The findings gathered from tissue-specific QTL analysis to severe COVID-19 can help to (1) identify candidate genes that may contribute to the pathophysiology of severe COVID-19, (2) focus the range of mechanistic hypotheses regarding the effect of specific gene expressions to organs most affected by severe COVID-19, and (3) cross-reference with known biology of linked phenotypes uncovered by the PheWAS analysis to validate and refine hypotheses regarding the pathophysiological mechanisms of severe COVID-19.

In this present study, we conducted both PheWAS and QTL analyses to identify potential mechanisms through which host genomic factors contribute to the pathogenesis and pathophysiology of severe COVID-19. We extracted the lead single-nucleotide polymorphism (SNP) with the smallest p value from each locus showing suggestive associations with severe COVID-19 cases with respiratory failure using the recent GWAS (Ellinghaus et al., 2020) and conducted the PheWAS using these lead SNPs by leveraging the GWAS Atlas database and identified a list of linked phenotypes. We systematically performed colocalization analyses between each locus identified by the PheWAS and 112 QTL data sets from GTEx (GTEx Consortium et al., 2020) and the eQTL Catalogue (Kerimov et al., 2020), spanning 69 unique tissue types. This provided a robust list of candidate genes with their tissue-specific expression alterations. Together, the results presented here have generated multiple focused hypotheses regarding the pathophysiology of severe COVID-19 for future validation and investigation.

## RESULTS

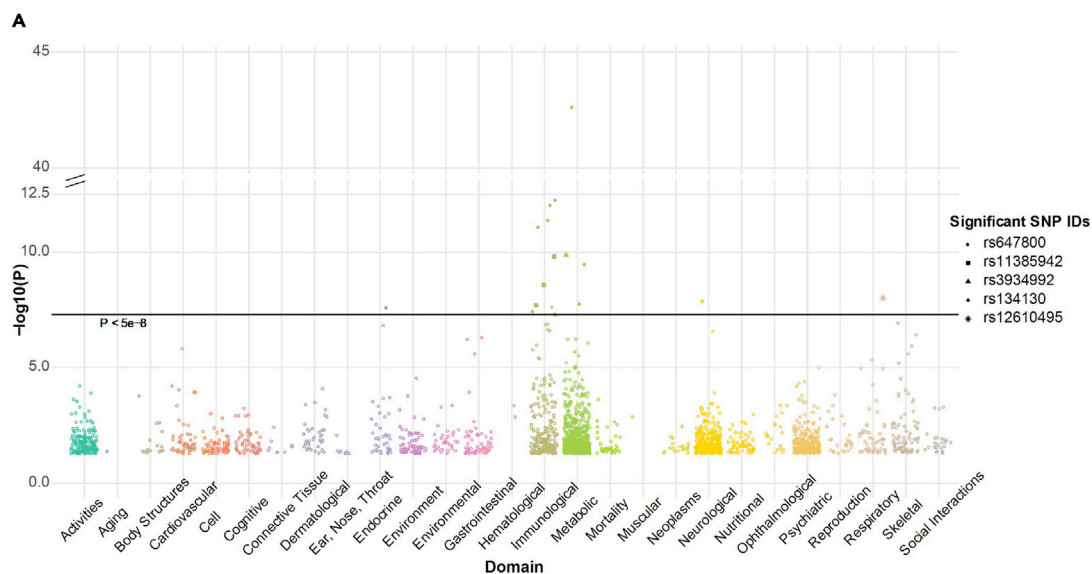
### PheWAS of genetic risk loci associated with severe COVID-19

To identify phenotypes sharing genetic etiologies with severe COVID-19, we conducted PheWAS of 22 pruned genetic risk loci significantly or suggestively associated ( $p < 1e-5$ ) with severe COVID-19 based on GWAS results published by The severe Covid-19 GWAS group (accession numbers: GCST90000256) (Ellinghaus et al., 2020) (Table S1). We first sought to replicate the 22 SNPs using GWAS performed by the COVID-19 Host Genetics Initiative (COVID-19 hg). Five SNPs, including rs2440652, rs3934992, rs12610495, rs11385942, and rs657152, showed an association passing the Bonferroni threshold ( $p < 0.05/22$ ) and were further considered as “COVID-19 hg-replicated SNPs” in the following texts. To gauge the independence of the signal for each lead SNP, linkage disequilibrium (LD) analysis was performed using the LDmatrix tool (Machiela and Chanock, 2015). LD analysis demonstrated that the 22 lead SNPs were independent from each other ( $R^2 < 0.01$ ; Table S2) and each replicated SNP was found within an LD block ( $R^2 > 0.8$ ) consisting of its neighboring variants.

On conducting the PheWAS of the 22 lead SNPs using GWAS Atlas (Watanabe et al., 2019) that has aggregated summary statistics of more than 4,756 GWAS studies, we identified 5 SNPs, including the 2 reported COVID-19 GWAS SNPs (rs647800, rs11385942) (Ellinghaus et al., 2020) and two COVID-19 hg-replicated SNPs (rs3934992, rs12610495), that showed associated traits reaching the genome-wide significance threshold ( $p < 5e-8$ ) (a full list of PheWAS results was listed in Table S3). A total of 13 traits were significant (Figure 1A)—8 of which were associated with rs647800, an intronic variant of the *ABO* gene (Yeung et al., 2017). Significant associated traits were predominantly immunological or metabolic (Figure 1B). The immunological traits were mostly relevant to population counts/proportions of monocytes and erythrocytes. Top metabolic traits were biochemical and physiological metrics concerning coagulation (“activated partial thromboplastin time”) and fat metabolism (“total cholesterol,” “leggs-leg fat ratio [male]”). Of all the significantly associated traits, fibrotic idiopathic interstitial pneumonia was the only disease linked through the COVID-19 GWAS-lead SNPs. Taken together, the PheWAS analysis of 22 lead SNPs identified several physiological metrics as well as a disease associated with severe COVID-19.

### eQTL/sQTL analyses

To evaluate the functional relevance of the lead GWAS loci, we used 22 lead SNPs to query GTEx and compiled e/sQTLs associated with the lead SNPs. After filtering associations based on QTL FDR  $< 0.05$ , 11 lead GWAS SNPs were identified as significant e/sQTLs (Tables S4 and S5) with the vast majority of eQTLs ( $n = 227$  of 258) and sQTLs ( $n = 140$  of 143) being associated with 1 SNP (rs3934992; Figure 2A). No one particular tissue type dominated the eQTL and sQTL associations; the top tissue types with



**B**  
**Significant PheWAS traits associated with lead SNPs**

SNP	Domain	Trait	P
rs647800	Endocrine	Thyroid-stimulating hormone	2.46e-08
rs11385942	Immunological	Monocyte percentage of white cells*	1.56e-10
rs11385942	Immunological	Monocyte count	2.59e-09
rs647800	Immunological	Hematocrit*	9.20e-13
rs647800	Immunological	Monocyte count*	4.12e-12
rs647800	Immunological	Hemoglobin concentration*	8.13e-12
rs647800	Immunological	Red blood cell count	3.60e-08
rs3934992	Metabolic	Waist-hip ratio (adjusted for BMI)	1.34e-10
rs647800	Metabolic	Activated partial thromboplastin time	2.29e-43
rs647800	Metabolic	Total cholesterol	3.29e-10
rs647800	Metabolic	Legs-leg fat ratio (male)	1.68e-08
rs134130	Neurological	Cerebellar vermal lobules VI VII	1.34e-08
rs12610495	Respiratory	Fibrotic idiopathic interstitial pneumonias	9.57e-09

**Figure 1. PheWAS of the lead SNPs showing suggestive associations ( $p < 1e-5$ ) with severe COVID-19**

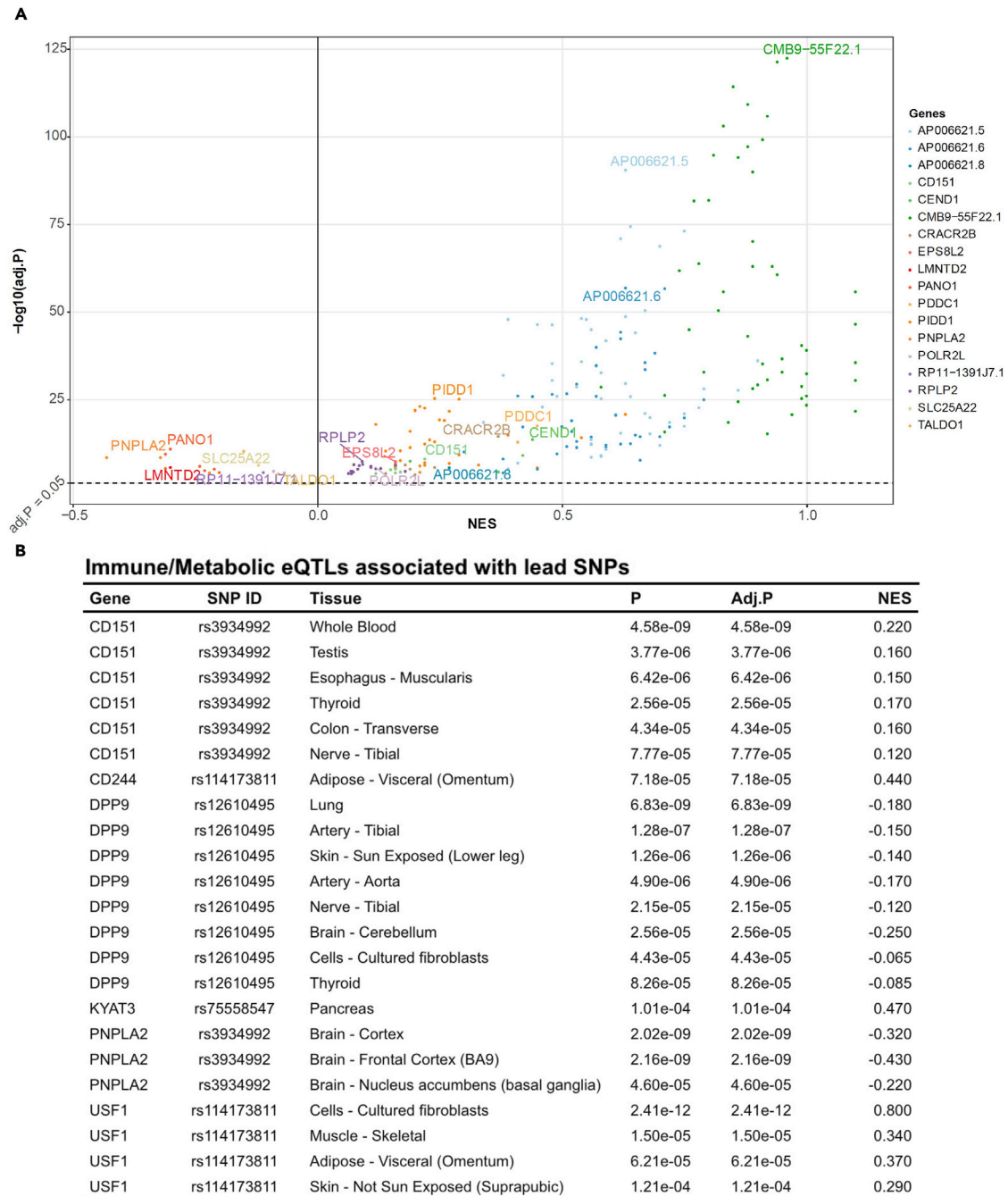
(A) A Manhattan plot of the phenotypes associated with the lead SNPs. Only phenotypes that had p-values reaching genome-wide significance ( $p < 5e-8$ , marked by the horizontal line) were marked with their associated SNPs (See “Significant SNP IDs”).

(B) A table of significant associated traits. (\*) indicates that there were duplicate entries resulted from two- and three-way meta-analyses for the same trait, where the three-way meta-analysis results were used for all these associations.

See also [Figure S1](#) and [Tables S1, S2, and S3](#).

associated eQTLs include the muscularis layer of the esophagus ( $n = 10$ ), tibial nerve ( $n = 10$ ), sun-exposed skin at the lower leg ( $n = 10$ ), and lungs ( $n=6$ ). The eQTL analysis encompassing 22 lead SNPs identified 29 novel genes in addition to 4 genes (i.e., *SLC6A20*, *FYCO1*, *CXCR6*, *CCR1*) previously reported by the severe COVID-19 GWAS group to be associated with rs11385942 at locus 3p21.31 ([Ellinghaus et al., 2020](#)). Two sets of three genes (*CMB9-55F22.1*, *PIDD1*, *RPLP2*), all of which are associated with the predominant lead COVID-19 hg-replicated SNP (rs3934992), constituted top 10% of statistically significant eGenes and sGenes, respectively.

In addition to these top genes, we also identified several genes with known immune and metabolic functions—paralleling our findings in the PheWAS analysis ([Figure 2B](#)). Corresponding to the top metabolic PheWAS traits related to fat metabolism, one distinct COVID-19 hg-replicated SNP was a genome-wide significant e/sQTL of *PNPLA2*, a critical enzyme involved in the first step of triglyceride hydrolysis



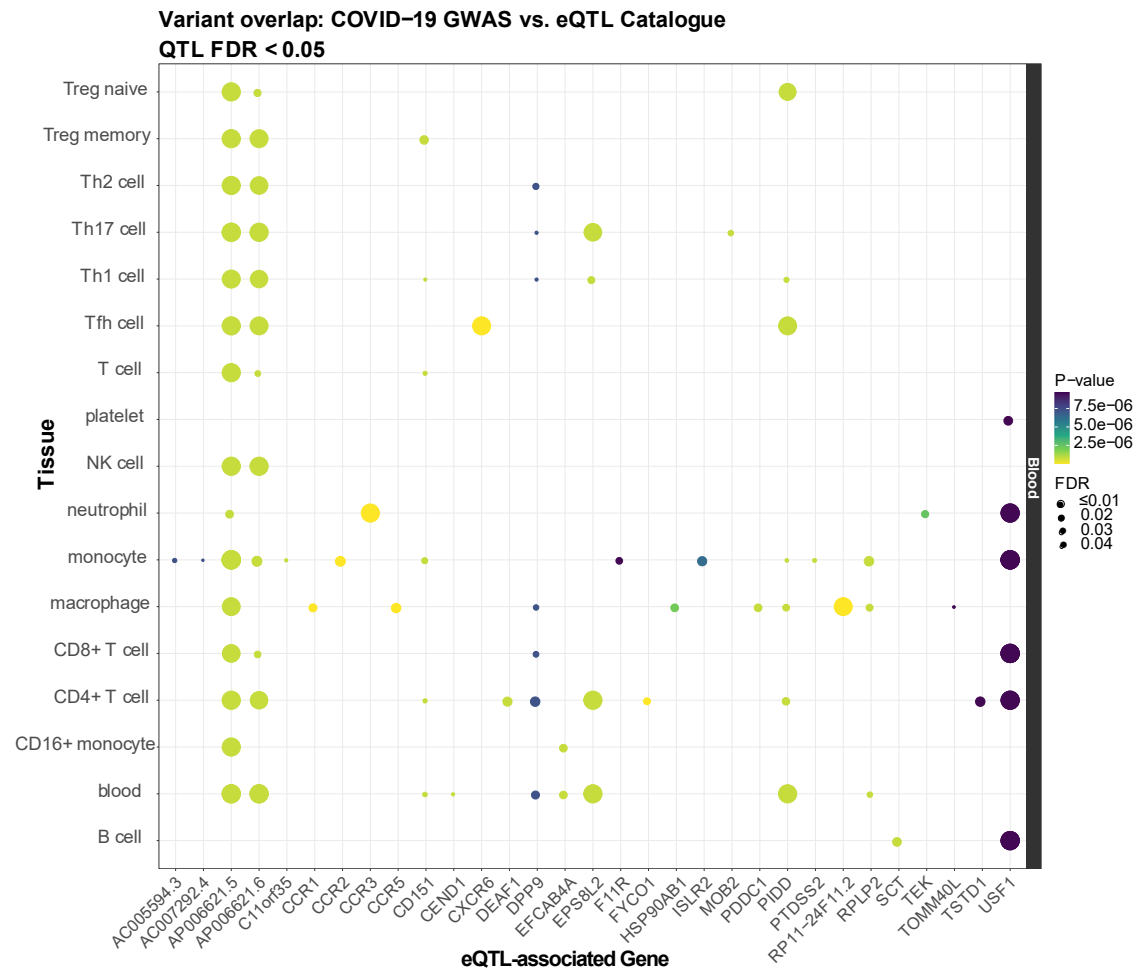
**Figure 2. Significant eQTLs associated with lead SNPs**

(A) A volcano plot of eQTLs associated with rs3934992. Dashed line (-) represents adjusted p-value of 0.05. NES: Normalized Effect Size.

(B) A table of eQTLs affecting immune and metabolic genes.

See also [Tables S4](#) and [S5](#)

([Taxiarchis et al., 2019](#)). In terms of immunity, expression changes in *CD151*, *PIDD1*, and *DPP9*—genes involved in immune cell activation and NF- $\kappa$ B-mediated immune response ([Okondo et al., 2017](#); [Rocha-Pe-rugini et al., 2014](#); [Tinel et al., 2007](#))—were associated with two distinct COVID-19 hg-replicated SNPs (rs3934992, rs12610495). Overall, QTL analyses of the lead SNPs identified genes of which expression changes may be functionally associated with severe COVID-19 at a tissue-specific resolution.



**Figure 3. Significant immune subtype-specific eQTLs associated with lead SNPs**

The eQTLs were compiled from the eQTL Catalogue (Kerimov et al., 2020) database and filtered at a Benjamini-Hochberg FDR < 0.05. Point color reflects the significance of that SNP in the COVID-19 GWAS (Ellinghaus et al., 2020) (more yellow = more significant p-values), while point size reflects the significance of the eQTL SNP (larger = more significant FDR). When multiple SNPs overlap within a given gene-tissue combination, the SNP with the lowest p-value/FDR is plotted. See also Figures S2 and S3 and Table S6

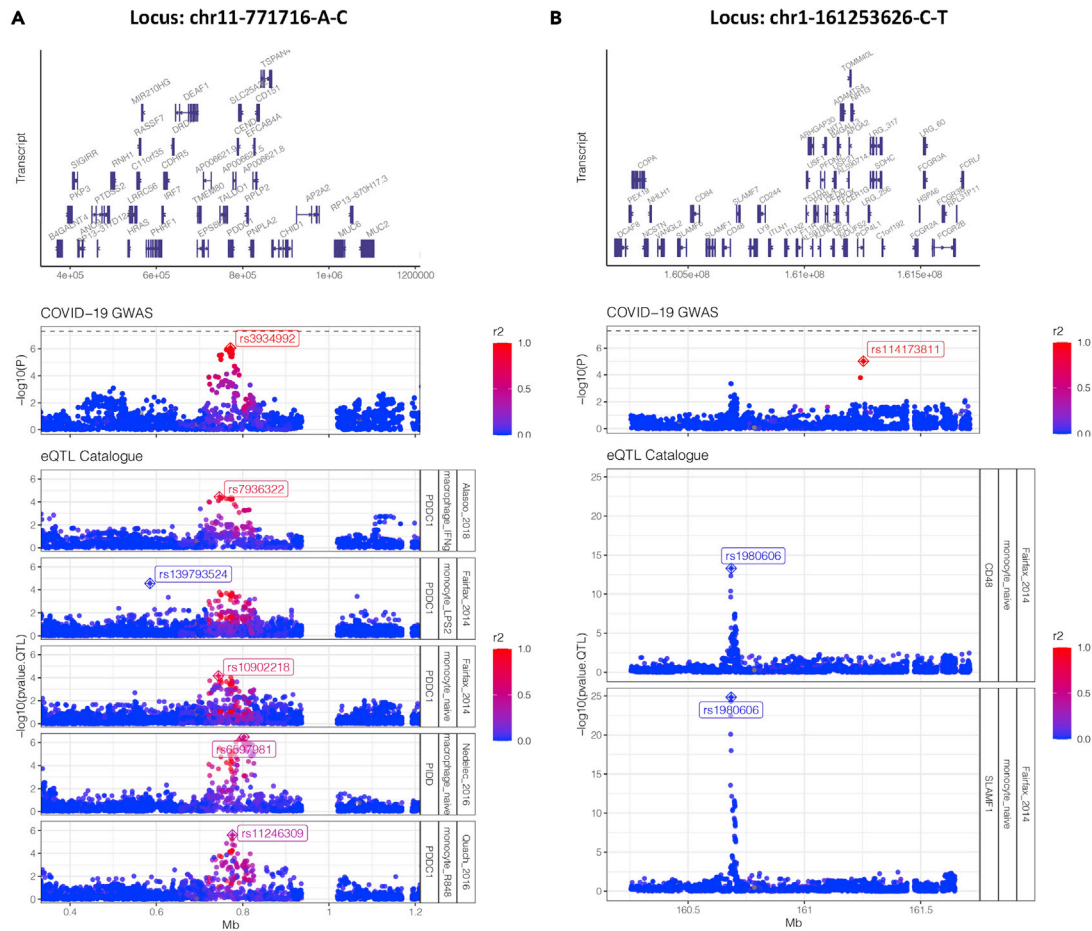
### Immune-specific eQTL analysis

Based on the identification of genes with known immune function in the eQTL analysis, as well as the reported involvement of the dysregulated immune response in severe COVID-19 (Lucas et al., 2020), we further queried the 22 lead SNPs in the eQTL Catalogue to determine whether they affected gene expression in specific immune cell subtypes. We identified 134 immune-cell-subtype-specific eQTL associations (FDR < 0.05) that are distributed across 7 GWAS SNPs and 31 eGenes. Several genes that were the most statistically significant in the GTEx analysis (*RPLP2*, *PIDD1*) reemerged in the immune-subtype-specific eQTL analysis and were primarily seen in monocytes and macrophages (Mo/M $\phi$ ) and several T cell subsets. Moreover, expression changes of a few immune genes (*DPP9*, *CD151*, *PIDD1*) that were previously highlighted were predominantly seen in both CD4<sup>+</sup> and CD8<sup>+</sup> T cells as well as Mo/M $\phi$ . Notably, 14 of these genes were not nominated by the GTEx eQTL analysis but were associated in various T, B, natural killer (NK), and myeloid cell subsets (Figure 3), including several genes involved in leukocyte recruitment and migration such as *CCR1*, *CCR2*, *CCR3*, *CCR5*, and *CXCR6*. These results highlight the resolution gained by conducting cell-type-specific analyses into immune-cell-expressed genes that may be linked to the genetic factors underlying severe COVID-19.

### GWAS/eQTL colocalization analysis

Owing to LD, overlapping significant SNPs between GWAS and QTL do not necessarily mean they share the same underlying causal signal. Or does a lack of overlapping lead variants necessarily mean that





**Figure 4. Colocalization analysis**

Manhattan plots of two severe COVID-19 GWAS loci (A) chr11-771716-A-C (B) chr1-161253626-C-T and their respective colocalized immune eQTLs from the eQTL Catalogue (colocalization probability >0.8). Each point represents a SNP across a genomic range, and color indicates the degree of LD ( $r^2$ ) with the proxy GWAS SNPs in each locus (indicated by crossed diamonds). In the eQTL rows, crossed diamonds indicate the lead QTL SNP (with the smallest association p value for a given gene).

they do not share the same signal being tagged by different variants. To robustly test for GWAS-eQTL signal sharing, we performed a series of colocalization analyses (Wakefield, 2009) in the 9 loci in which the GWAS-lead SNPs overlapped with significant eVariants (FDR <0.05) in the eQTL analysis. We identified 7 colocalizations that met the posterior probability threshold (see STAR Methods), spread across 2 GWAS loci, 5 eQTL data sets (all in macrophages and monocytes), and affecting the expression of 4 genes (*PDDC1*, *PIDD1*, *CD48*, and *SLAMF1*) (Figure 4)

In locus chr11-771716-A-C (Figure 4A), the colocalized genetic signal both increased severe COVID-19 risk and modulated expression of *PDDC1* in monocytes under three conditions (naive, LP2, and R848) and in IFN- $\gamma$ -stimulated macrophages. The expression of P53-induced death domain protein 1 (*PIDD1*) was also affected in macrophages (naive), suggesting that both *PDDC1* and *PIDD1* are perhaps controlled by a common regulatory element that is disrupted by the severe COVID-19 risk signal in this region.

While locus chr1-161253626-C-T was also colocalized with eQTLs in monocytes (naive), the signal modulated expression of *CD48* and signaling lymphocytic activation molecule family member 1, both of which are immunoglobulin-like receptors of the CD2 subfamily (Figure 4B). Interestingly, a closer inspection revealed that the colocalizing GWAS signal was not the same one tagged by the previously identified proxy SNP showing the most significant association with severe COVID-19 (rs114173811) but instead a signal ~0.5Mb upstream tagged by the SNP rs1980606. LD analysis confirmed that these signals are independent

(no SNPs in the upstream signal had  $r^2 > 0$  with the lead SNP in the downstream signal). This highlights the complexity of genetic association signals as well as the additional insights that colocalization analyses can reveal.

### Differential gene expression in bronchoalveolar lavage fluid using single-cell analyses

To determine if expression of any of these eQTL-nominated immune genes as well as genes involved in implicated pathways (i.e., inflammasome activation) was altered in patients with severe COVID-19 compared with patients with moderate COVID-19 or healthy controls, we performed a DGE analysis on a recently published scRNA-seq data set of bronchoalveolar lavage fluid of patients with COVID-19 of varying disease severity and healthy controls (Liao et al., 2020b). Because the functions of these immune genes are either most well-characterized in Mo/M $\phi$ s (*DPP9*, *NLRP1*, *GSDMD*) (Okondo et al., 2017) or known to be most relevant for Mo/M $\phi$ s (*CCR1*, *CCR2*, *CCR5*) (Dyer et al., 2019), we specifically focused our DGE analysis on Mo/M $\phi$ s subsets within the data set and evaluated the differential expression of six eQTL-nominated immune genes. Average expression of genes involved in inflammasome activation (*DPP9*, *NLRP1*, *GSDMD*) was significantly altered (adj.  $p < 0.001$ ) between Mo/M $\phi$ s of patients with severe and moderate COVID-19. Moreover, average expression of several genes involved in leukocyte recruitment and migration (*CCR1*, *CCR2*, *CCR5*) was significantly higher (Bonferroni adj.  $p < 0.001$ ) in Mo/M $\phi$ s of patients with severe COVID-19 than in both patients with moderate COVID-19 as well as healthy controls (Figure S3). These results demonstrate, in at least a subset of patients with COVID-19, the differential expression of eQTL-nominated immune genes between Mo/M $\phi$ s of patients with severe COVID-19 and of patients with moderate COVID-19 or healthy controls.

### DISCUSSION

In this study, we characterized the phenome- and transcriptome-wide associations of genetic loci implicated in severe COVID-19 with respiratory failure. We leveraged GWAS ATLAS database (Watanabe et al., 2019) containing summary statistics from 4,756 GWASs to perform a PheWAS of the lead SNPs and identified 13 significant traits linked to genetic risk loci of severe COVID-19. Using both eQTL databases, we performed eQTL analysis of the lead SNPs to characterize their effects on gene expression in tissue- and immune-cell-subset-specific resolution. Eighteen lead SNPs were identified as eQTLs/sQTLs of 48 genes. Taken together, these results reveal insights into potential mechanisms that could drive the formation of focused hypotheses about the pathophysiology of severe COVID-19.

Immunological and metabolic traits were among significant phenotypes uncovered by the PheWAS analysis. Severe cases of COVID-19 are often characterized by dysregulated and hyperinflammatory immune system (Lucas et al., 2020; Merad and Martin, 2020; Vabret et al., 2020) and metabolic dysfunction (Ayes, 2020)—suggesting that these SNPs may either play a role in multisystem manifestations of severe COVID-19 or contribute to the development of comorbidities known to exacerbate the severity of disease. For instance, that top immunological traits were hematologic parameters such as monocyte and erythrocyte populations is concordant with the hematological findings in severe COVID-19 cases such as monocyte expansion (Merad and Martin, 2020; Wen et al., 2020) and decrease in hemoglobin levels (Chowdhury and Anwar, 2020; Li et al., 2020). Recent reports have suggested that cytokine profiles of patients with severe COVID-19 show similarities to cytokine release syndromes such as macrophage activation syndromes (Giamarellos-Bourboulis et al., 2020; Mehta et al., 2020; Moore and June, 2020) and that monocytes/macrophages extensively accumulate in the lungs (Sánchez-Cerrillo et al., 2020). Further investigations are required to disseminate how these lead SNPs (rs647800 and rs11385492) may similarly or differentially contribute to monocyte count and monocyte-driven inflammation seen in COVID-19.

Coupled with hyperinflammation driven partly by these mononuclear cells, another major prong of severe COVID-19 pathology is systemic coagulopathy. Patients suffering from severe COVID-19 often have lower levels of platelet counts and increased levels of D-dimers (Becker, 2020; Chan and Weitz, 2020; Iba et al., 2020; Liao et al., 2020a). Numerous cases of widespread pulmonary thrombosis and disseminated intravascular coagulation or sepsis-induced coagulopathy have also been described in severe COVID-19 patients (Becker, 2020; Tang et al., 2020). Possibly reflective of the predominant role of coagulopathies in the pathophysiology of severe COVID-19, the most statistically significant trait identified in our PheWAS analysis was activated partial thrombin time ( $p$  value =  $2.29 \times 10^{-43}$ ; associated SNP: rs647800) which has been reported to be prolonged in patients with severe COVID-19. Notably, a recent review by Merad et al. (Merad and Martin, 2020) has suggested a link between hyperinflammatory monocytes and coagulopathies—as



even in the absence of any vascular injury, coagulation pathway can still be activated through the recruitment of tissue-factor-expressing inflammatory monocytes. Taken together, it would be interesting to interrogate the genetic pleiotropy underlying rs647800 and how it affects monocyte count, activated partial thrombin time, and severe COVID-19.

In addition to coagulation, another metabolic traits highlighted by the PheWAS results were regarding cholesterol levels (“total cholesterol”;  $p$ -value =  $3.29e-10$ ) and physiological measures associated with obesity (“waist-hip ratio” and “legs-leg fat ratio”;  $p$ -values =  $1.34e-10$  and  $1.68e-8$ , respectively). Both obesity and high cholesterol levels have been associated with increased risk of severe COVID-19 (Wang et al., 2020; Zhu et al., 2020), thereby suggesting that the PheWAS analysis can capture some of the comorbidities found by epidemiological studies.

The PheWAS also identified rs12610495 to be associated with fibrotic idiopathic interstitial pneumonias (FIIP;  $p$ -value:  $9.57e-9$ ), a disease directly connected to respiratory failure. FIIP shares a similar risk profile with severe COVID-19 disease and thus is a potential comorbidity of severe COVID-19 (George et al., 2020). Notably, in addition to sharing a similar risk profile with FIIP, some patients with severe COVID-19 suffer from lung function abnormalities that strongly suggest the presence of pulmonary fibrosis (Raghu and Wilson, 2020), implicating similar etiologies underlying COVID-19 and FIIP.

In line with the immunological and metabolic traits that emerged from the PheWAS analysis, our QTL analyses identified several genes with related functions that contribute to those traits. For instance, we found that rs12610495, an intron variant of *DPP9*, is within a susceptibility locus for FIIP (Fingerlin et al., 2013) and modulates both the expression and splicing of *DPP9*. *DPP9* can inhibit *NLRP1* and repress downstream inflammasome activation (Zhong et al., 2018). Conversely, pharmacological inhibition of *DPP9* has been shown to induce *NLRP1* inflammasome-mediated pyroptosis in monocytes and macrophages (Gai et al., 2019; Okondo et al., 2017; Vasconcelos et al., 2019; Zhong et al., 2018). *NLRP1*-dependent pyroptosis in the lung alveolar macrophages has been shown to cause progressive lung damage that is similar to the one seen in acute respiratory distress syndrome (Kovarova et al., 2012). Furthermore, *NLRP1*-dependent pyroptosis is sufficient to cause progressive lung injury, independent of *NLRP3* activation or *IL-1b* secretion, which suggests a relevant role of *DPP9* in mediating lung injury seen in severe COVID-19 (Kovarova et al., 2012). Finally, it was recently shown that human *NLRP1* senses double-stranded RNA (dsRNA) and activates an inflammasome complex upon dsRNA sensing (Bauernfried et al., 2021)—thereby raising a distinct possibility that dsRNA replication intermediates of SARS-CoV-2 (V'kovski et al., 2021) can be sensed by *NLRP1* and set off an inflammatory cascade mediated by inflammasome activation. Considering the evidence together, it is possible that decreased *DPP9* expression in the lungs can predispose patients with COVID-19 to develop lung injury mediated by increased *NLRP1*-dependent pyroptosis. Concordant with this hypothesis, the lead SNP rs12610495 is associated not only with decreased expression of *DPP9* specifically in the lung and in macrophages but also with increased risk of severe COVID-19 with respiratory failure. Furthermore, our DEG analysis of scRNA-seq of bronchoalveolar fluid of patients with COVID-19 showed that expression of *DPP9*, *NLRP1*, and *GSDMD*, an effector protein for pyroptosis (Mathur et al., 2018), was significantly altered in Mo/Mφs of patients with severe COVID-19 compared to their moderate counterparts. Overall, our results suggest the potential role of inflammasome activation and pyroptosis in exacerbating respiratory distress in severe COVID-19—a point that was recently echoed in a recent review of inflammasomes in COVID-19 (Yap et al., 2020).

Beyond *DPP9*, our QTL analyses identified several immune genes that may play a role in potential mechanisms through which host immune response against SARS-CoV-2 can be affected. These potential mechanisms include NF-κB-mediated immune response (i.e. *PIDD1*) (Bock et al., 2013; Tinel et al., 2007), immune synaptic interaction involved in T cell activation (i.e. *CD151*) (Rocha-Perugini et al., 2014) and T cell exhaustion in viral infection (i.e. *CD244*) (Agresta et al., 2018; Pacheco et al., 2013). In our immune subset-specific QTL analysis, we were able to resolve changes in expression of *PIDD1* and *CD151* in diverse T cell subsets as well as monocytes. Moreover, the QTL overlap and colocalization analyses showed that the genetic signal associated with increased severe COVID-19 risk also modulated immune genes (i.e., *PDDC1*, *CD150*, *CD48*) in monocytes. For instance, one of the colocalized genes, *CD48*, is known to interact with *CD244* to modulate NK cell and CD8<sup>+</sup> T cell effector function (Agresta et al., 2018; McArdel et al., 2016). Taken together, both QTL overlap and colocalization analyses highlight strong associations between risk of severe COVID-19 and transcriptional changes that underpin monocyte-lymphocyte interactions.

Furthermore, we also identified several leukocyte adhesion and chemokine receptors (*CCR3*, *CCR5*) in addition to the chemokine receptors (*CCR1*, *CCR2*, *CXCR6*) previously reported by the severe COVID-19 GWAS group (Ellinghaus et al., 2020)—all of which can mediate the increased leukocyte recruitment seen in the lungs of patients with severe COVID-19 (Sánchez-Cerrillo et al., 2020). In line with our immune-subset-specific QTL analysis and the aforementioned histological findings, we observed in our DEG analysis a significant increase in the expression of these chemokine receptors (*CCR1*, *CCR2*, *CCR5*) in Mo/Mφs of patients with severe COVID-19 compared with those of both patients with moderate COVID-19 and healthy controls. While current scRNA-seq data set did not include genotype information, future single-cell data sets with paired DNA/RNA data can help confirm the link between genetic risk allele and their eQTL effects at a single-cell resolution.

Concordant with several metabolic traits identified in the PheWAS such as obesity and cholesterol levels (Kanai et al., 2018; Pulit et al., 2019; Rask-Andersen et al., 2019), the QTL analyses identified two genes, *PNPLA2* and *USF1*, whose functions are directly connected to those traits. Furthermore, we were able to resolve changes in expression of *USF1* to the level of various immune cell compartments and found that *USF1* expression was altered predominantly in Mo/Mφs and T cell subtypes. *USF1* is a transcription factor that modulates several genes involved in lipid and glucose metabolism such as apolipoproteins and lipases including *PNPLA2*, a triglyceride lipase that catalyzes the initial step in triglyceride hydrolysis (Aguilar-Salinas et al., 2014; Taghizadeh et al., 2019). Knockout studies of *USF1* in mice have been shown to enhance cholesterol efflux in macrophages and downregulate secretion of pro-inflammatory cytokines such as MCP-1 and IL-1β (Ruuth et al., 2018) and *USF1* variants in humans have been associated with familial combined hyperlipidemia and coronary artery disease (Fan et al., 2014; Laurila et al., 2016; Ruuth et al., 2018). Interestingly, a recent study by Dias et al. showed that monocytes from COVID-19 have increased lipid droplet accumulation and that pharmacological inhibition of lipid droplet formation led to a significant inhibition of replication of SARS-CoV-2 (Dias et al., 2020). In addition to lipid metabolism, kynurenine metabolism has also been implicated by one of our nominated genes, *KYAT3*, an aminotransferase that transaminates kynurenine to kynurenic acid. Considering that kynurenine pathway plays an integral role in immunosuppression (Routy et al., 2016; Wirthgen et al., 2018; Zhai et al., 2020), it would be interesting to investigate how expression changes in *KYAT3* can affect the immune response against SARS-CoV2. Notably, alterations in both metabolic pathways implicated in the QTL analyses—fatty acid and kynurenine metabolism—have also been observed in patients with COVID-19 (Thomas et al., 2020), further supporting that these genes may play a role in immune dysfunction seen in severe COVID-19.

In summary, we conducted PheWAS and QTL analyses to investigate how host genetic factors may contribute to pathophysiology of severe COVID-19 and uncovered numerous gene-disease-tissue connections. Our results provide plausible consequences of the host genetic risk factors underlying severe disease, informing future investigations into immune, metabolic, and other aspects of COVID-19 pathophysiology.

### Limitations of the study

We acknowledge the limitations of our approaches, such as the restriction of discovering traits studied by available GWAS or QTL data sets and their correlational nature, which may be addressed by techniques such as Mendelian randomization as the COVID-19 GWAS expands in size (T.C.-19 H.G., 2020). Widespread LD also poses challenges when trying to distinguish causal variants from their close correlates—an issue which fine-mapping techniques aim to resolve (Broekema et al., 2020; Hutchinson et al., 2020; Schaid et al., 2018). Furthermore, the suggestive genetic risk loci of severe COVID-19 used for analyses herein as well as the identified PheWAS and QTL associations also require validation.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS

- Extraction of severe COVID-19 associated lead SNPs
- PheWAS analysis
- Tissue- and immune-specific eQTL/sQTL analysis
- GWAS/eQTL colocalization analyses
- scRNA-seq analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.102550>.

## ACKNOWLEDGMENTS

The authors wish to acknowledge D. Ellinghaus, The Severe Covid-19 GWAS Group and the COVID-19 Host Genetics Initiative for contributing GWAS summary statistics data and M. Liao et al. for contributing scRNA-seq data of bronchoalveolar lavage fluid collected from both COVID-19 patients and healthy controls. We would also like to acknowledge each study's participating patients and families that generously contributed the data. This work was supported by NIH R35GM138113 to K.H. and by NIH 2T32GM007280 to C.Y.M.

## AUTHOR CONTRIBUTIONS

K.H. and C.Y.M. conceived the research and designed the analyses. C.Y.M. conducted the PheWAS, GTEx QTL, and scRNA-seq analyses. B.M.S. and T.R. conducted the immune QTL analyses. C.Y.M. wrote the manuscript, while K.H., B.M.S., and T.R. edited the manuscript. All authors read and approved the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 10, 2020

Revised: March 24, 2021

Accepted: May 13, 2021

Published: June 25, 2021

## REFERENCES

- Agresta, L., Hoebe, K.H.N., and Janssen, E.M. (2018). The emerging role of CD244 signaling in immune cells of the tumor microenvironment. *Front. Immunol.* *9*, 2809.
- Aguilar-Salinas, C.A., Tusie-Luna, T., and Pajukanta, P. (2014). Genetic and environmental determinants of the susceptibility of Amerindian derived populations for having hypertriglyceridemia. *Metabolism* *63*, 887–894.
- Althuler, D., and Donnelly, P. (2005). A haplotype map of the human genome. *Nature* *437*, 1299–1320.
- Ayres, J.S. (2020). A metabolic handbook for the COVID-19 pandemic. *Nat. Metab.* *2*, 572–585.
- Bauernfried, S., Scherr, M.J., Pichlmair, A., Duderstadt, K.E., and Hornung, V. (2021). Human NLRP1 is a sensor for double-stranded RNA. *Science* *371*, eabd0811.
- Becker, R.C. (2020). COVID-19 update: Covid-19-associated coagulopathy. *J. Thromb. Thrombolysis* *50*, 54–67.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* *57*, 289–300.
- Bock, F.J., Krumschnabel, G., Manzi, C., Peintner, L., Tanzer, M.C., Hermann-Kleiter, N., Baier, G., Llacuna, L., Yelamos, J., and Villunger, A. (2013). Loss of PIDD limits NF- $\kappa$ B activation and cytokine production but not cell survival or transformation after DNA damage. *Cell Death Differ.* *20*, 546–557.
- Broekema, R.V., Bakker, O.B., and Jonkers, I.H. (2020). A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open Biol.* *10*, 190221.
- Chan, N.C., and Weitz, J.I. (2020). COVID-19 coagulopathy, thrombosis, and bleeding. *Blood* *136*, 381–383.
- Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* *395*, 507–513.
- Chowdhury, S.F., and Anwar, S. (2020). Management of hemoglobin disorders during the COVID-19 pandemic. *Front. Med.* *7*, 306.
- 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.
- Denny, J.C., Ritchie, M.D., Basford, M.A., Pulley, J.M., Bastarache, L., Brown-Gentry, K., Wang, D., Masys, D.R., Roden, D.M., and Crawford, D.C. (2010). PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. *Bioinformatics* *26*, 1205–1210.
- Denny, J.C., Bastarache, L., Ritchie, M.D., Carroll, R.J., Zink, R., Mosley, J.D., Field, J.R., Pulley, J.M., Ramirez, A.H., Bowton, E., et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* *31*, 1102–1111.
- Dias, S.S.G., Soares, V.C., Ferreira, A.C., Sacramento, C.Q., Fintelman-Rodrigues, N., Temezo, J.R., Teixeira, L., Barreto, E., Mattos, M., de Freitas, C.S., et al. (2020). Lipid droplets fuel SARS-CoV-2 replication and production of inflammatory mediators. *Biorxiv*.

- Diogo, D., Tian, C., Franklin, C.S., Alanne-Kinnunen, M., March, M., Spencer, C.C.A., Vangjeli, C., Weale, M.E., Mattsson, H., Kilpeläinen, E., et al. (2018). Phenome-wide association studies across large population cohorts support drug target validation. *Nat. Commun.* **9**, 4285.
- 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.
- Dyer, D.P., Medina-Ruiz, L., Bartolini, R., Schuette, F., Hughes, C.E., Pallas, K., Vidler, F., Macleod, M.K.L., Kelly, C.J., Lee, K.M., et al. (2019). Chemokine receptor redundancy and specificity are context dependent. *Immunity* **50**, 378–389.e5.
- Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., Asselta, R., et al. (2020). Genomewide association study of severe covid-19 with respiratory failure. *New Engl. J. Med.* **383**, 1522–1534.
- Fan, Y.-M., Hernesniemi, J., Oksala, N., Levula, M., Raitoharju, E., Collings, A., Hutri-Kähönen, N., Juonala, M., Marniemi, J., Lyytikäinen, L.-P., et al. (2014). Upstream Transcription Factor 1 (USF1) allelic variants regulate lipoprotein metabolism in women and USF1 expression in atherosclerotic plaque. *Sci. Rep.* **4**, 4650.
- Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278.
- Fingerlin, T.E., Murphy, E., Zhang, W., Peljto, A.L., Brown, K.K., Steele, M.P., Loyd, J.E., Cosgrove, G.P., Lynch, D., Groshong, S., et al. (2013). Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat. Genet.* **45**, 613–620.
- Gai, K., Okondo, M.C., Rao, S.D., Chui, A.J., Ball, D.P., Johnson, D.C., and Bachovchin, D.A. (2019). DPP8/9 inhibitors are universal activators of functional NLRP1 alleles. *Cell Death Dis.* **10**, 587.
- George, P.M., Wells, A.U., and Jenkins, R.G. (2020). Pulmonary fibrosis and COVID-19: the potential role for antifibrotic therapy. *Lancet Respir. Med.* **8**, 807–815.
- Giamarellos-Bourboulis, E.J., Netea, M.G., Rovina, N., Akinosoglou, K., Antoniadou, A., Antonakos, N., Damoraki, G., Gkavogianni, T., Adami, M.-E., Katsaounou, P., et al. (2020). Complex immune dysregulation in COVID-19 patients with severe respiratory failure. *Cell Host Microbe* **27**, 992–1000.e3.
- GTEX Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330.
- Guan, W.-J., Ni, Z.-Y., Hu, Y., Liang, W.-H., Ou, C.-Q., He, J.-X., Liu, L., Shan, H., Lei, C.-L., Hui, D.S.C., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New Engl. J. Med.* **382**, 1708–1720.
- Hebbring, S.J. (2014). The challenges, advantages and future of phenome-wide association studies. *Immunology* **141**, 157–165.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506.
- Hutchinson, A., Asimit, J., and Wallace, C. (2020). Fine-mapping genetic associations. *Hum. Mol. Genet.* **29**, ddaa148.
- Iba, T., Levy, J.H., Levi, M., and Thachil, J. (2020). Coagulopathy in COVID-19. *J. Thromb. Haemost.* **18**, 2103–2109.
- Kanai, M., Akiyama, M., Takahashi, A., Matoba, N., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., et al. (2018). Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400.
- Kerimov, N., Hayhurst, J.D., Manning, J.R., Walter, P., Kolberg, L., Peikova, K., Samoviča, M., Burdett, T., Jupp, S., Parkinson, H., et al. (2020). eQTL catalogue: a compendium of uniformly processed human gene expression and splicing QTLs. *Biorxiv*.
- Kovarova, M., Hesker, P.R., Jania, L., Nguyen, M., Snouwaert, J.N., Xiang, Z., Lommatzsch, S.E., Huang, M.T., Ting, J.P.-Y., and Koller, B.H. (2012). NLRP1-Dependent pyroptosis leads to acute lung injury and morbidity in mice. *J. Immunol.* **189**, 2006–2016.
- Laurila, P.-P., Soronen, J., Kooijman, S., Forsström, S., Boon, M.R., Surakka, I., Kaiharju, E., Coomans, C.P., Berg, S.A.A.V.D., Autio, A., et al. (2016). USF1 deficiency activates brown adipose tissue and improves cardiometabolic health. *Sci. Transl. Med.* **8**, 323ra13.
- Li, Q., Cao, Y., Chen, L., Wu, D., Yu, J., Wang, H., He, W., Chen, L., Dong, F., Chen, W., et al. (2020). Hematological features of persons with COVID-19. *Leukemia* **34**, 2163–2172.
- Liao, D., Zhou, F., Luo, L., Xu, M., Wang, H., Xia, J., Gao, Y., Cai, L., Wang, Z., Yin, P., et al. (2020a). Haematological characteristics and risk factors in the classification and prognosis evaluation of COVID-19: a retrospective cohort study. *Lancet Haematol.* **7**, e671–e678.
- Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., Cheng, L., Li, J., Wang, X., Wang, F., et al. (2020b). Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* **26**, 842–844.
- Lucas, C., Wong, P., Klein, J., Castro, T.B.R., Silva, J., Sundaram, M., Ellingson, M.K., Mao, T., Oh, J.E., Israelow, B., et al. (2020). Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* **584**, 463–469.
- 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.
- Mathur, A., Hayward, J.A., and Man, S.M. (2018). Molecular mechanisms of inflammasome signaling. *J. Leukoc. Biol.* **103**, 233–257.
- McArdel, S.L., Terhorst, C., and Sharpe, A.H. (2016). Roles of CD48 in regulating immunity and tolerance. *Clin. Immunol.* **164**, 10–20.
- Mehta, P., McAuley, D.F., Brown, M., Sanchez, E., Tattersall, R.S., Manson, J.J., and UK, H.A.S.C. (2020). COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet* **395**, 1033–1034.
- Merad, M., and Martin, J.C. (2020). Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. *Nat. Rev. Immunol.* **20**, 355–362.
- Moore, B.J.B., and June, C.H. (2020). Cytokine release syndrome in severe COVID-19. *Science* **368**, 473–474.
- Okondo, M.C., Johnson, D.C., Sridharan, R., Go, E.B., Chui, A.J., Wang, M.S., Poplawski, S.E., Wu, W., Liu, Y., Lai, J.H., et al. (2017). DPP8 and DPP9 inhibition induces pro-caspase-1-dependent monocyte and macrophage pyroptosis. *Nat. Chem. Biol.* **13**, 46–53.
- Pacheco, Y., McLean, A.P., Rohrbach, J., Porichis, F., Kaufmann, D.E., and Kavanagh, D.G. (2013). Simultaneous TCR and CD244 signals induce dynamic downmodulation of CD244 on human antiviral T cells. *J. Immunol.* **191**, 2072–2081.
- Pe'er, I., Yelensky, R., Altshuler, D., and Daly, M.J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385.
- Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., et al. (2019). Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174.
- Raghu, G., and Wilson, K.C. (2020). COVID-19 interstitial pneumonia: monitoring the clinical course in survivors. *Lancet Respir. Med.* **8**, 839–842.
- Rask-Andersen, M., Karlsson, T., Ek, W.E., and Johansson, Å. (2019). Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nat. Commun.* **10**, 339.
- Rocha-Perugini, V., González-Granado, J.M., Tejera, E., López-Martin, S., Yañez-Mó, M., and Sánchez-Madrid, F. (2014). Tetraspanins CD9 and CD151 at the immune synapse support T-cell integrin signaling. *Eur. J. Immunol.* **44**, 1967–1975.
- Routy, J.-P., Routy, B., Graziannini, G.M., and Mehraj, V. (2016). The kynurenine pathway is a double-edged sword in immune-privileged sites and in cancer: implications for immunotherapy. *Int. J. Tryptophan. Res.* **9**, IJTR.S38355.
- Ruuth, M., Soronen, J., Kaiharju, E., Merikanto, K., Perttilä, J., Metso, J., Lee-Rueckert, M., Taskinen, M.-R., Kovanen, P.T., Öörni, K., et al. (2018). USF1 deficiency alleviates inflammation, enhances cholesterol efflux and prevents cholesterol accumulation in macrophages. *Lipids Health Dis.* **17**, 285.
- Sánchez-Cerrillo, I., Landete, P., Aldave, B., Sánchez-Alonso, S., Sánchez-Azofra, A., Marcos-Jiménez, A., Ávalos, E., Alcaraz-Serna, A., de los

- Santos, I., Mateu-Albero, T., et al. (2020). COVID-19 severity associates with pulmonary redistribution of CD1c+ DC and inflammatory transitional and nonclassical monocytes. *J. Clin. Invest.* **130**, 6290–6300.
- Schaid, D.J., Chen, W., and Larson, N.B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504.
- Schilder, B.M., Humphrey, J., and Raj, T. (2020). echolocatoR: an automated end-to-end statistical and functional genomic fine-mapping pipeline. *Biorxiv*.
- Semmes, E.C., Vijaykrishnan, J., Zhang, C., Hurst, J.H., Houlston, R.S., and Walsh, K.M. (2020). Leveraging genome and phenome-wide association studies to investigate genetic risk of acute lymphoblastic leukemia. *Cancer Epidemiol. Prev. Biomarkers* **29**, 1606–1614.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoekius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21.
- Taghizadeh, E., Esfehiani, R.J., Sahebkar, A., Parizadeh, S.M., Rostami, D., Mirinezhad, M., Poursheikhani, A., Mobarhan, M.G., and Pasdar, A. (2019). Familial combined hyperlipidemia: an overview of the underlying molecular mechanisms and therapeutic strategies. *Iubmb Life* **71**, 1221–1229.
- Tang, N., Bai, H., Chen, X., Gong, J., Li, D., and Sun, Z. (2020). Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy. *J. Thromb. Haemost.* **18**, 1094–1099.
- Taxiarchis, A., Mahdessian, H., Silveira, A., Fisher, R.M., and van't Hooft, F.M. (2019). PNPLA2 influences secretion of triglyceride-rich lipoproteins by human hepatoma cells. *J. Lipid Res.* **60**, 1069–1077.
- Thomas, T., Stefanoni, D., Reisz, J.A., Nemkov, T., Bertolone, L., Francis, R.O., Hudson, K.E., Zimring, J.C., Hansen, K.C., Hod, E.A., et al. (2020). COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. *JCI Insight* **5**, e140327.
- Tinel, A., Janssens, S., Lippens, S., Cuenin, S., Logette, E., Jaccard, B., Quadroni, M., and Tschopp, J. (2007). Autoproteolysis of PIDD marks the bifurcation between pro-death caspase-2 and pro-survival NF- $\kappa$ B pathway. *EMBO J.* **26**, 197–208.
- Vabret, N., Britton, G.J., Gruber, C., Hegde, S., Kim, J., Kuksin, M., Levantovsky, R., Malle, L., Moreira, A., Park, M.D., et al. (2020). Immunology of COVID-19: current state of the science. *Immunity* **52**, 910–941.
- de Vasconcelos, N.M., Vliegen, G., Gonçalves, A., Hert, E.D., Martín-Pérez, R., Opdenbosch, N.V., Jallapally, A., Geiss-Friedlander, R., Lambeir, A.-M., Augustyns, K., et al. (2019). DPP8/DPP9 inhibition elicits canonical Nlrp1b inflammasome hallmarks in murine macrophages. *Life Sci. Alliance* **2**, e201900313.
- V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **19**, 155–170.
- Wakefield, J. (2009). Bayes factors for genome-wide association studies: comparison with P-values. *Genet. Epidemiol.* **33**, 79–86.
- Wang, H., Yuan, Z., Pavel, M.A., Hobson, R., and Hansen, S.B. (2020). The role of high cholesterol in age-related COVID19 lethality. *Biorxiv*.
- Watanabe, K., Stringer, S., Frei, O., Mirkov, M.U., Leeuw, C.de, Polderman, T.J.C., Sluis, S.van der, Andreassen, O.A., Neale, B.M., and Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348.
- Wen, W., Su, W., Tang, H., Le, W., Zhang, X., Zheng, Y., Liu, X., Xie, L., Li, J., Ye, J., et al. (2020). Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. *Cell Discov.* **6**, 31.
- Williamson, E.J., Walker, A.J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C.E., Curtis, H.J., Mehrkar, A., Evans, D., Inglesby, P., et al. (2020). Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **584**, 430–436.
- Wirthgen, E., Hoefflich, A., Rebl, A., and Günther, J. (2018). Kynurenine acid: the janus-faced role of an immunomodulatory tryptophan metabolite and its link to pathological conditions. *Front. Immunol.* **8**, 1957.
- Yap, J.K.Y., Moriyama, M., and Iwasaki, A. (2020). Inflammasomes and pyroptosis as therapeutic targets for COVID-19. *J. Immunol.* **205**, ji2000513.
- Yeung, E.W., Craggs, J.G., and Gizer, I.R. (2017). Comorbidity of alcohol use disorder and chronic pain: genetic influences on brain reward and stress systems. *Alcohol. Clin. Exp. Res.* **41**, 1831–1848.
- Zhai, L., Bell, A., Ladomersky, E., Lauing, K.L., Bollu, L., Sosman, J.A., Zhang, B., Wu, J.D., Miller, S.D., Meeks, J.J., et al. (2020). Immunosuppressive IDO in cancer: mechanisms of action, animal models, and targeting strategies. *Front. Immunol.* **11**, 1185.
- Zhao, J., Cheng, F., Jia, P., Cox, N., Denny, J.C., and Zhao, Z. (2018). An integrative functional genomics framework for effective identification of novel regulatory variants in genome-phenome studies. *Genome Med.* **10**, 7.
- Zhong, F.L., Robinson, K., Teo, D.E.T., Tan, K.-Y., Lim, C., Harapas, C.R., Yu, C.-H., Xie, W.H., Sobota, R.M., Au, V.B., et al. (2018). Human DPP9 represses NLRP1 inflammasome and protects against autoinflammatory diseases via both peptidase activity and FIIND domain binding. *J. Biol. Chem.* **293**, 18864–18878.
- Zhu, Z., Hasegawa, K., Ma, B., Fujiogi, M., Camargo, C.A., and Liang, L. (2020). Obesity & genetic predisposition with COVID-19. *Metabolism* **112**, 154345.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Genome-wide association study of patients with severe COVID-19	The severe COVID-19 GWAS group et al., PMID: 32558485	Accession number: GCST90000256
scRNA-seq of bronchoalveolar lavage fluid of COVID-19 patients	Liao et al. 2020a, 2020b PMID: 32398875	Accession Number: GSE145926
GWAS Atlas	Watanabe et al., 2019 PMID: 31427789	N/A
eQTL Catalogue	Kerimov et al., 2020 ( <a href="https://www.biorxiv.org/content/10.1101/2020.01.29.924266v2.full">https://www.biorxiv.org/content/10.1101/2020.01.29.924266v2.full</a> )	N/A
GTEx database	GTEx Consortium et al., 2020 PMID: 32913098	N/A
Software and algorithms		
R v3.6.3	The R Project for Statistical Computing	<a href="https://www.r-project.org">https://www.r-project.org</a>
BioRender	©BioRender	<a href="https://biorender.com">biorender.com</a>

## RESOURCE AVAILABILITY

## Lead contact

Information and requests for resources should be directed to and will be fulfilled by the lead contact, Kuan-lin Huang ([kuan-lin.huang@mssm.edu](mailto:kuan-lin.huang@mssm.edu)).

## Materials availability

This study did not generate new unique reagents.

## Data and code availability

All relevant data are available from the lead contact upon request. All code use in the eQTL Catalogue and colocalization analyses can be found in the following GitHub repository: [https://github.com/RajLabMSSM/COVID19\\_PheWAS](https://github.com/RajLabMSSM/COVID19_PheWAS).

Additional Supplemental Items are available from Mendeley Data at <http://doi.org/10.17632/nyn84yyms7.1> and <https://data.mendeley.com/datasets/nyn84yyms7/1>.

## METHOD DETAILS

## Extraction of severe COVID-19 associated lead SNPs

We downloaded the summary statistics of the genome-wide meta-analysis that was corrected for age and sex as reported by The Severe Covid-19 GWAS group (Accession numbers: GCST90000256) (Ellinghaus et al., 2020). Aside from the reported SNPs, namely rs11385942 at locus 3p21.31 and with rs657152 at locus 9q34.2, we extracted the SNPs with suggestive associations ( $p < 1e-5$ ), resulting in 319 SNPs from GCST90000256. We then pruned the associated SNPs based on the most significant p value per locus ( $\pm 100$ Kb regions surrounding the lead SNP) recursively within the study, resulting in 22 non-adjacent loci, including 2 genome-wide significant SNPs ( $p < 5e-8$ ) (Altshuler and Donnelly, 2005; Pe'er et al., 2008) and 20 suggestive SNPs associated with severe COVID-19 with respiratory failure (Table S1). We then cross-referenced 22 lead SNPs against the meta-analysis of GWAS studies between hospitalized COVID-19 patients and the population controls (B2\_ALL, dated 09/30) performed by the COVID-19 Host Genetics Initiative (COVID-19 hg) (COVID-19 Host Genetics Initiative, 2020). We considered these SNPs' association as "replicated" in the independent GWAS if they surpassed the Bonferroni-corrected p-value for the 22 queried SNPs ( $P < 0.05/22$ ), validating associations for 3 additional SNPs (rs2440652, rs3934992, rs12610495) in addition to the two previously reported COVID-19 GWAS SNPs. In order to evaluate the level of independence of each SNP, linkage disequilibrium (LD) analysis was performed using the LDmatrix



tool (Machiela and Chanock, 2015) (1) between any pair of the 22 lead SNPs that are located on the same chromosome ((Table S2) and (2) between each replicated SNP and other variants showing suggestive associations ( $p < 1e-5$ ) from GCST90000256 within  $\pm 100\text{Kb}$  surrounding each replicated SNP (Figure S1). Spanish and Italian population were used as reference population for the LDmatrix analysis to correspond with the population used in the original COVID-19 GWAS study.

### PheWAS analysis

Phenotype associations of the 22 lead SNPs were queried using the PheWAS database available at the GWAS ATLAS (Watanabe et al., 2019) and the PheWAS results of each SNP was compiled into the candidate list of PheWAS hits (Table S3). Upon compilation, the list of associations was filtered using the genome-wide significance threshold ( $p < 5e-8$ ) (Altshuler and Donnelly, 2005; Pe'er et al., 2008), resulting in 17 significant phenotypes. Considering the varying statistical thresholds used in the PheWAS literature (Denny et al., 2013; Diogo et al., 2018; Hebring, 2014), we focused on a concise description of the associations passing the genome-wide significance threshold in the main texts, while noting that the full list of PheWAS results are available in Table S3 enabling examination of other associations.

### Tissue- and immune-specific eQTL/sQTL analysis

Similar to the PheWAS analysis, we used our list of 22 lead SNPs to query tissue-specific expression QTL (eQTL) and splicing QTL (sQTL) datasets from the GTEx database (GTEx Consortium et al., 2020). In order to correct for multiple testing, e/sQTL associations were then filtered using the Benjamini-Hochberg (Benjamini and Hochberg, 1995), resulting in 258 eQTLs and 143 sQTLs with  $FDR < 0.05$ . In addition to the GTEx analysis, we used the R package *catalogueR* (<https://github.com/RajLabMSSM/catalogueR>) (Schilder et al., 2020) to query these same 22 SNPs in the eQTL Catalogue (Kerimov et al., 2020), which contains 63 eQTL datasets across 20 studies (excluding GTEx). GWAS hits within sex chromosomes were excluded as the eQTL Catalogue only contains results for autosomes. eQTL Catalogue contains many immune subtype-specific and stimulation condition-specific datasets, permitting greater resolution of our biological interpretations. 190 significant eVariants associated with 9 lead GWAS SNPs (Table S6) remained after filtering by  $FDR < 0.05$  (Figure S2)—132 of which were immune subtype- or state- specific.

### GWAS/eQTL colocalization analyses

Using the summary statistics of the original GWAS study (GCST90000256) and the eQTL Catalogue datasets, colocalization analysis was performed on 9 lead SNPs whose eVariants identified in the eQTL Catalogue were statistically significant ( $FDR < 0.05$ ). The locus for each lead SNP was defined as  $\pm 1\text{Mb}$  from the lead SNP (2Mb windows). Since the original GWAS summary statistics do not provide both minor allele frequency (MAF) or the sample size for each SNP, MAF of each SNP was assumed to be the same as the one provided by the eQTL Catalogue. Each SNP was assumed to have the same effective sample size calculated from the total number of participants in the GWAS while accounting for the proportion of cases/controls using the formula:

$$\frac{4}{\left(\frac{1}{N_{\text{cases}}} + \frac{1}{N_{\text{controls}}}\right)}$$

A series of Bayesian colocalization tests of were run to evaluate the probability of shared genetic signal between each GWAS locus and each eQTL locus. Specifically, we used the *run\_coloc()* function within *catalogueR* (Schilder et al., 2020), which runs Approximate Bayes Factor colocalization (Wakefield, 2009) to first infer the probability that each SNP is causal for the phenotype within each respective dataset and then calculate the probability that the same underlying genetic signal is shared in both datasets. eQTL pseudogenes were identified via Ensembl and removed from downstream analysis. For each GWAS-QTL locus comparison, the signals were considered to be colocalized if they met the following criterion (where PP.H4 is the posterior probability that the signals are associated with their respective traits and shared, and PP.H3 is the posterior probability that the signals are associated with their respective traits but not shared):

$$(PP.H3 + PP.H4 > 0.8) \& \left(\frac{PP.H4}{PP.H3}\right) \geq 2$$

### scRNA-seq analysis

Previously published single-cell RNA-seq (scRNA-seq) dataset (accession number: GSE145926) of bronchoalveolar lavage fluid (BALF) from COVID-19 patients and healthy controls was downloaded and analyzed for differential expression of six immune genes nominated from the immune-specific eQTL analyses above (*DPP9*, *NLRP1*, *GSDMD*, *CCR1*, *CCR2*, *CCR5*) in monocyte/macrophage subsets using the analytical pipeline described previously (Liao et al., 2020b). Briefly, using the cluster annotations described in the original publication to isolate monocyte/macrophage subsets, the macrophage clusters were then integrated using *IntegrateData()* function with default parameters from the Seurat v3 package (Stuart et al., 2019) and differential gene expression analysis between disease severity groups was performed using MAST (Finak et al., 2015) in Seurat v3 (Stuart et al., 2019) to generate a list of differentially expressed genes between two groups (i.e. Severe vs. Moderate, Moderate vs. Healthy, Severe vs. Healthy). The list was then queried for specific immune genes of interest for their statistics. In accordance with Seurat's best practices, Bonferroni adjusted p values were computed taking into account the full number of genes tested in the dataset (Stuart et al., 2019). Violin plots comparing their expression levels as well as the UMAP plot of the integrated macrophage subset are shown in Figure S3.

### QUANTIFICATION AND STATISTICAL ANALYSIS

All quantification and statistical analyses were performed as described in the [method details](#) section of the [STAR Methods](#).