

# **Integrative Genomics Analysis Reveals a Novel 21q22.11 Locus Contributing to Susceptibility of COVID-19**

Yunlong Ma<sup>1,#</sup>, Yukuan Huang<sup>1,#</sup>, Sen Zhao<sup>2</sup>, Yinghao Yao<sup>3</sup>, Yaru Zhang<sup>1</sup>, Jia Qu<sup>1</sup>, Nan Wu<sup>2\*</sup>,  
Jianzhong Su<sup>1,3\*</sup>

<sup>1</sup>Institute of Biomedical Big Data, School of Ophthalmology & Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou, 325027, China.

<sup>2</sup> Beijing Key Laboratory for Genetic Research of Skeletal Deformity, Key laboratory of big data for spinal deformities, Department of Orthopedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China.

<sup>3</sup> Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou 325011, China.

<sup>#</sup> These authors contributed equally to this work.

<sup>\*</sup> To whom correspondence should be addressed: Prof. Jianzhong Su, e-mail: [sujz@wmu.edu.cn](mailto:sujz@wmu.edu.cn); Prof. Nan Wu, e-mail: [dr.wunan@pumch.cn](mailto:dr.wunan@pumch.cn).

**Keywords:** GWAS, meta-analysis, COVID-19, susceptibility, genetic variants

## Abstract

The systematic identification of host genetic risk factors is essential for the understanding and treatment of COVID-19. By performing a meta-analysis of two independent genome-wide association (GWAS) summary datasets (N = 680,128), a novel locus at 21q22.11 was identified to be associated with COVID-19 infection (rs9976829 in *IFNAR2* and upstream of *IL10RB*, OR = 1.16, 95% CI = 1.09 - 1.23,  $P = 2.57 \times 10^{-6}$ ). The rs9976829 represents a strong splicing quantitative trait locus (sQTL) for both *IFNAR2* and *IL10RB* genes, especially in lung tissue ( $P = 1.8 \times 10^{-24}$ ). Gene-based association analysis also found *IFNAR2* was significantly associated with COVID-19 infection ( $P = 2.58 \times 10^{-7}$ ). Integrative genomics analysis of combining GWAS with eQTL data showed the expression variations of *IFNAR2* and *IL10RB* have prominent effects on COVID-19 in various types of tissues, especially in lung tissue. The majority of *IFNAR2*-expressing cells were dendritic cells (40%) and plasmacytoid dendritic cells (38.5%), and *IL10RB*-expressing cells were mainly nonclassical monocytes (29.6%). *IFNAR2* and *IL10RB* are targeted by several interferons-related drugs. Together, our results uncover 21q22.11 as a novel susceptibility locus for COVID-19, in which individuals with G alleles of rs9976829 have a higher probability of COVID-19 susceptibility than those with non-G alleles.

## 1. Introduction

Coronavirus disease 2019 (COVID-19) has rapidly evolved into a global pandemic <sup>[1]</sup>. The health and economy systems of most nations worldwide are suffering from severe disruptions <sup>[2]</sup>. As of July 13<sup>th</sup>, 2020, there were more than 12.9 million confirmed patients worldwide with more than

550,000 deaths<sup>[3]</sup>. The clinical manifestations of COVID-19 range from asymptomatic to severe respiratory failure<sup>[4]</sup>. Early studies on COVID-19 infection have concentrated on epidemiology<sup>[5]</sup>, clinical characteristics<sup>[6]</sup>, and genomic features of virus<sup>[7]</sup>. Understanding host genetic factors contributing to COVID-19 susceptibility is essential for the precise management in the community.

Recently, a growing number of researchers have concentrated on the involvement of host genetic factors in COVID-19. Through performing a genome-wide association study (GWAS) with 1,610 severe COVID-19 patients and 2,205 controls, Ellinghaus et al.<sup>[8]</sup> reported two important gene clusters of 3p21.31 and 9q34.2 as genetic susceptibility loci for severe COVID-19, and confirmed a potential involvement of the ABO blood-group system. From a population perspective, the COVID-19 Host Genetic Consortium launched the “COVID-19 Host Genetics Initiative” to collect data from the genetics community to uncover the genetic determinants of COVID-19 susceptibility, severity, and outcomes<sup>[2]</sup>. However, identification of more host genetic risk factors is limited by the sample size of a single study.

Here, we performed a meta-analysis by combining two independent GWAS summary statistics with a large-scale sample size to identify novel genes for COVID-19 susceptibility. The systematic bioinformatics analyses were performed, including MAGMA gene-based association analysis, S-PrediXcan and S-MultiXcan analysis, Sherlock-based inference analysis, gene-property analysis, single-cell RNA analysis, and functional enrichment analysis. We uncover the genes and pathways convey risk of COVID-19 infection and give a clue of the potential effective drugs for treating COVID-19.

## 2. Results

### 2.1. SNP-level association analysis reveals a novel susceptible locus 21q22.11 for COVID-19

By conducting a meta-analysis of GWAS summary data from Ellinghaus et al.<sup>[8]</sup> (COVID\_I: 1,610 COVID-19 patients and 2,205 controls) and the COVID-19 Host Genetic Consortium<sup>[2]</sup> (ANA5: 1,678 COVID-19 patients and 674,635 controls), we confirmed two reported loci of 3p21.31 and 9q34.2 to be associated with COVID-19 infection (rs11385942 in *SLC6A20*,  $P = 2.87 \times 10^{-16}$ , and rs8176719 in *ABO*,  $P = 4 \times 10^{-7}$ ; Figure 1, Table 1, and Supplemental Figures S1-S2). The reported rs657152 in *ABO*, which is high linkage disequilibrium with rs8176719, remains suggestively significant ( $P = 5.53 \times 10^{-6}$ ). Notably, we identified a novel locus at 21q22.11 to be associated with COVID-19 infection (rs9976829 in *IFNAR2-IL10RB*, OR = 1.16, 95% CI = 1.09 - 1.23,  $P = 2.57 \times 10^{-6}$ ; Table 1 and Figure 1). The rs9976829 represents a splicing quantitative trait locus (sQTL) for both *IFNAR2* and *IL10RB* genes across multiple tissues with the strongest significance in the lung tissue ( $P = 1.8 \times 10^{-24}$ ; Supplemental Figure S3).

### 2.2. Gene-based association analysis identifies nine risk genes for COVID-19

We performed a MAGMA gene-based association analysis by using the meta-GWAS results and found that nine genes in three loci of 3p21.31 (*LZTFL1*, *XCRI*, *CCR9*, *FYCO1*, *SLC6A20*, and *CXCR6*), 9q21.32 (*HNRNPK* and *RMII*), and 21q22.11 (*IFNAR2*) were significantly associated with COVID-19 infection (FDR < 0.05; Figure 2 and Table 2). As expected, the gene of *ABO* showed a nominally significant association with COVID-19 ( $P = 6.55 \times 10^{-4}$ ) consistent with the previous report<sup>[8]</sup>. Of note, *IFNAR2* ( $P = 2.58 \times 10^{-7}$ ), *HNRNPK* ( $P = 1.46 \times 10^{-5}$ ), and *RMII* ( $P = 1.86 \times 10^{-5}$ ) were

identified for the first time to be associated with COVID-19 infection. There were other 30 genes showing suggestive associations with COVID-19 ( $P < 1 \times 10^{-3}$ , Supplemental Table S1). Meanwhile, we performed two MAGMA gene-based association analyses for COVID\_I GWAS summary data from Ellinghaus et al. [8] and the COVID-19 Host Genetic Consortium (ANA5) [2], respectively. We found that these nine identified genes showed significant or suggestive associations with COVID-19 in both COVID-I and ANA5 datasets (Table 2), indicating that our meta-analysis based on larger samples enhance the statistical power to uncover risk genes for COVID-19.

### 2.3. Cytokine-related pathways enriched by risk genes for COVID-19

As mentioned above, there were 41 genes showing significant or suggestive associations with COVID-19 ( $P < 1 \times 10^{-3}$ , Supplemental Table S1). We performed a pathway enrichment analysis of these 41 identified genes for COVID-19 susceptibility (Methods), and found they were significantly enriched in two pathways of cytokine-cytokine receptor interaction (FDR = 0.009) and chemokine signaling pathway (FDR = 0.009) (Figure 3a). Additionally, there were seven pathways showing suggestive associations ( $P < 0.05$ ; Supplemental Table S2), including Kaposi sarcoma-associated herpesvirus infection ( $P = 0.0086$ ), human cytomegalovirus infection ( $P = 0.014$ ), and human papilloma virus infection ( $P = 0.027$ ). Meanwhile, we conducted a GO enrichment analysis and found 3 GO-terms were significantly overrepresented (FDR < 0.05; Figure 3b and Supplemental Table S3). That is, cytokine receptor activity (FDR =  $1.01 \times 10^{-5}$ ), cytokine binding (FDR = 0.022), and peptide receptor activity (FDR = 0.027). Previous studies demonstrated that soluble cytokines activate an anti-viral and anti-proliferate state by inducing the expression of many

interferon-stimulated genes to prevent viral replication <sup>[9]</sup>. Network-based enrichment analysis showed that these 41 genes are at least partially biologically connected ( $P = 0.013$ , Supplemental Figure S8). These results suggest cytokine-related pathways or functional terms may play important roles in the process of COVID-19 infection.

#### **2.4. Expression of *IFNAR2* and *IL10RB* associated with COVID-19 across multiple tissues**

To highlight the functional association of these 11 identified genes with COVID-19, we conducted two independent integrative genomics analyses by incorporating meta-GWAS summary data with eQTL data across 49 GTEx tissues. Using S-PrediXcan analysis, we found that the expression variations of these 11 genes have prominent effects on COVID-19 in various types of tissues (Supplemental Table S5). Consistently, Sherlock-based integrative genomics analysis showed seven genes including *IFNAR2* and *IL10RB* whose genetically regulated expression were significantly associated with COVID-19 across multiple tissues (Supplemental Table S6). For example, the *IFNAR2* gene was associated with COVID-19 across six tissues including lung ( $P = 2.44 \times 10^{-4}$ ). The *IL10RB* gene showed associations with COVID-19 across 24 tissues with the most significant tissue of cells-transformed fibroblasts ( $P = 2.80 \times 10^{-5}$ ).

We further used S-MultiXcan to meta-analyze the tissue-specific associations from S-PrediXcan across 49 GTEx tissues, and found these 11 genes whose expression were significantly associated with COVID-19, which support our results from the MAGMA analysis (Figure 3c and Supplemental Table S7). Additionally, our *in silico* permutation analysis also showed that genes identified from MAGMA analysis had a significantly higher overlap with S-MultiXcan-identified genes than random

events ( $P < 1 \times 10^{-5}$ ; Figure 3d).

## 2.5. *IFNAR2* and *IL10RB* specially expressed in immunity associated cell types in lung tissue

To examine the links between tissue-specific gene expression profiles and COVID-19 gene associations, we conducted a MAGMA gene-property analysis in 53 specific tissue types and 30 general tissue types. We found that the association signals were enriched in lung, thyroid, and esophagus tissue in 30 general tissue types (Supplemental Figure S4a). In the analysis of 53 specific tissues, COVID-19 gene associations were also enriched in lung, cultured fibroblasts, and thyroid (Supplemental Figure S4b). These results suggest that these identified genes for COVID-19 may have important functions in lung tissue.

Using 50 cell populations across four compartments (epithelial, endothelial, stromal, and immune) of lung tissue, we identified the primarily expressed cells of these 11 risk genes associated with COVID-19 (Figure 4, Supplemental Table S4, and Supplemental Figure S5). The *HNRNPK* gene was widely expressed in various cell types across all four compartments (Supplemental Figure S6). However, the majority of *IFNAR2*-expressing cells were dendritic cells (40%) and plasmacytoid dendritic cells (38.5%) (Figure 4). The *IFNAR2* was also expressed within lipofibroblast (20%), ciliated (19.1%) and nonclassical monocyte (18.5%), albeit at diminished abundance compared with dendritic cells; *IL10RB* was primarily expressed within nonclassical monocyte (29.6%). Plasmacytoid dendritic cells produce large amounts of type I interferons-proteins that are important for immunity to viruses<sup>[10]</sup>. Our data indicated that *IFNAR2* and *IL10RB* could play regulatory roles in the pulmonary immune response.

## 2.6. Potential drugs targeted with *IFNAR2* and *IL10RB*

By performing a drug-gene interaction analysis and literature mining, we found that seven of 11 COVID-19-associated genes (63.6%) were enriched in five potential “druggable” gene categories (Supplemental Table S8 and Supplemental Figure S7). The gene of *CCR9* is targeted by two FDA-approved drugs including hydroxyurea and hydralazine (Supplemental Table S9). There are nine FDA-approved drugs showing agonist-receptor interactions with *IFNAR2* (Supplemental Table S9), including interferon alfa-2a, interferon alfacon-1, and interferon beta-1a, which could be useful alone or in combination with other antiviral drugs for treating SARS-CoV infection <sup>[11]</sup>. Loutfy et al. <sup>[12]</sup> reported that interferon alfacon-1 plus corticosteroids showed association with improved oxygen saturation and more rapid resolution of radiographic lung opacities than systemic corticosteroid alone in severe acute respiratory syndrome (SARS). Treatment with interferon has shown preliminary benefits for patients with COVID-19 <sup>[13]</sup> and is being evaluated by numerous ongoing clinical trials. The IFN-beta treatment could effectively block SARS-CoV-2 replication <sup>[14]</sup>. The *IL10RB* gene is targeted by peginterferon lambda-1a (Supplemental Table S9), which has been designed to treat mild COVID-19 in a clinical trial (identifier: NCT04331899). These results suggest that interferons associated with *IFNAR2* and *IL10RB* exert potential effects on the treatment of COVID-19.

## 3. Discussion

Using a meta-analytic method to combine two existing GWAS summary datasets with a large-scale sample size, we validated two reported genetic loci on chromosome 3p21.31 and 9q34.2



to be significantly associated with COVID-19 infection, and found a novel locus at a chromosome 21q22.11 gene cluster conveying susceptibility of COVID-19.

On chromosome 21q22.11, the peak association signal covered two genes of *IFNAR2* and *IL10RB*, which have biological functions that probably related to COVID-19. *IFNAR2* encodes a type I membrane protein, which forms one of the two chains of a receptor for interferons alpha and beta. The deficiency of *IFNAR2* supports an essential role for interferons alpha and beta in human antiviral immunity <sup>[15]</sup>. Notably, dysregulation of type I interferon response has been observed in COVID-19 patients <sup>[16]</sup>. Impaired type I interferon activity in the blood could be a hallmark of severe COVID-19 <sup>[17]</sup>. Furthermore, *IFNAR2* is required for anti-influenza immunity and related to the risk of post-influenza bacterial superinfections <sup>[18]</sup>. As for *IL10RB*, its encoded protein belongs to the cytokine receptor family and is an accessory chain essential for the active interleukin 10 receptor complex. Variants in *IFNAR2* and *IL10RB* gene were associated with the susceptibility to hepatitis B virus (HBV) infection <sup>[19]</sup>. Most recently, the interferon pathway is identified to be targeted by the COVID-19 viral protein of Nsp13 <sup>[20]</sup>.

Functional enrichment analysis showed that cytokine-related pathways including cytokine receptor activity, cytokine-cytokine receptor interaction, and chemokine signaling pathway were significantly enriched by genes associated with COVID-19. Seven risk genes of *IFNAR2*, *IL10RB*, *XCR1*, *CXCR6*, *CCR9*, *CCR1*, and *GNG12* have implicated in these pathways. There were four genes encoding chemokine receptors, including the X-C motif chemokine receptor 1 (*XCR1*), the C-X-C motif chemokine receptor 6 (*CXCR6*), the CC motif chemokine receptor 9 (*CCR9*), and the CC motif chemokine receptor 1 (*CCR1*). Vaccine molecules targeting *XCR1* on cross-presenting dendritic cells enhance a protective CD8<sup>+</sup> T-cell responses against influenza virus <sup>[21]</sup>. *CXCR6*

modulates the localization of lung tissue-resident memory CD8<sup>+</sup> T-cells throughout the sustained immune response to respiratory pathogens, including influenza viruses <sup>[22]</sup>. Both CCR9 and CCR1 also have related functions on immune response to respiratory influenza infection <sup>[23]</sup>. Varied manifestations in COVID-19 infection may result from different host genetic factors, which are probably related to immune response <sup>[24]</sup>. Through inducing the expression of many interferon-stimulated genes, soluble cytokines have anti-viral, anti-proliferate, and immunomodulatory effects on obstructing viral replication <sup>[9]</sup>. Together, cytokine-related pathways potentially play important roles in the pathogenesis of COVID-19 infection, and more relevant studies should be performed to explore the underlying biological mechanisms.

With regard to the novel gene of *HNRNPK*, it belongs to the subfamily of ubiquitously expressed heterogeneous nuclear ribonucleoproteins (hnRNP), of which proteins have important roles in cell cycle progression. *HNRNPK* acts as a central hub in the replication cycle of multiple viruses including HCV <sup>[25]</sup>. An interaction of *HNRNPK* and *HNRNPA2B1* with hepatitis E virus (HEV) promoters has important roles in HEV replication <sup>[26]</sup>. Two cellular RNA binding proteins of hnRNP and NS1-BP have important roles in regulating influenza A virus RNA splicing <sup>[27]</sup>. As for the *RMI1* gene, its protein is an essential component of the RMI complex, which has a crucial role in DNA repair and maintaining genome stability <sup>[28]</sup>. Since there existed important functions of these identified genes, molecular studies are warrant to illustrate the functional consequences of detected association signals.

The power of our study is limited by the difference in study design between the two datasets we analyzed: the COVID-19 Host Genetic Consortium included patients with mild or severe COVID-19 but the Ellinghaus et al. <sup>[8]</sup> study only included severe COVID-19. Because most of the COVID-19

infected individuals are asymptomatic, population-based controls used in the original studies may contain a substantial proportion of asymptomatic patients, which further reduced the power of our meta-analysis. Due to data from the two included GWASs were based on summary statistics, the population stratification was not assessed in the current meta-analysis. For the Ellinghaus et al.<sup>[8]</sup> study, to examine for population stratification within and across Italian and Spanish panels, a principal component analysis (PCA) was performed by using the FlashPCA<sup>[29]</sup>. Covariates from 10 PCA were conducted to control for potential population stratification (COVID\_I). As for the ANA5 GWAS summary data from the COVID-19 Host Genetic Consortium constructed by 10 contributing studies, the strategy of population stratification is unknown.

#### **4. Conclusions**

In summary, our findings uncover 21q22.11 as a novel risk locus for COVID-19 susceptibility and implicate the potential role of interferons targeting *IFNAR2* and *IL10RB* in the treatment of COVID-19. Individuals with the G alleles of rs9976829 have a 16% greater chance of COVID-19 infection compared with those carrying no such allele. The efficacy and the safety of interferon products are still being evaluated by numerous ongoing clinical trials, which may be strengthened by subgrouping the patients according to their genotypes of the *IFNAR2* and the *IL10RB* loci. Further studies are needed to delineate current findings and understand the underlying pathophysiology of COVID-19.

#### **5. Methods**

### ***GWAS summary data from Ellinghaus et al. (Dataset #1)***

For this GWAS recently reported by Ellinghaus et al. <sup>[8]</sup>, there were 1,980 patients with severe COVID-19 enrolled from seven hospitals in the Italian and Spanish epicenters of the SARS-CoV-2 pandemic in the Europe. A total of 2,381 control participants were enrolled from Italy and Spain. After stringent quality control and excluding population outliers, 1,610 patients with COVID-19 with respiratory failure (835 Italian and 775 Spanish COVID-19 cases) and 2,205 control participants (1,255 Italian and 950 Spanish controls) were included in the final GWAS. In total, 8,965,091 high-quality SNPs (post imputation  $R^2 \geq 0.6$  and minor allele frequency (MAF)  $\geq 1\%$ ) were included in the Italian cohort and 9,140,716 high-quality SNPs in the Spanish cohort. The GWAS summary statistics (COVID\_I) are publicly available in the website ([www.c19-genetics.eu](http://www.c19-genetics.eu)). For more detailed information, please refer to the original article <sup>[8]</sup>.

### ***GWAS summary data from the COVID-19 Host Genetic Consortium (Dataset #2)***

This GWAS summary statistics of the publicly available COVID-19 HGI GWAS meta-analyses round 2 (ANA5, susceptibility [affected vs. population]) was downloaded from the official website of the COVID-19 Host Genetic Consortium <sup>[2]</sup> ([www.covid19hg.org/results](http://www.covid19hg.org/results); analysis named “20200508-results-ANA5\_ALL\_inv\_var\_meta”; file named “COVID19\_HGI\_ANA5\_20200513.txt.gz”; release date of May 15 2020). There were 1,678 COVID-19 patients and 674,635 control participants from 10 contributing studies. For the GWAS summary statistics, there were a total of 34,010,457 genetic variants included with a MAF threshold of 0.0001 and an imputation score filter of 0.6. For more detailed information, please refer to the

original article <sup>[2]</sup>.

### ***Meta-analysis of GWAS summary data***

By using the meta-analysis tool of METAL <sup>[30]</sup>, a fixed-effects meta-analysis was performed to identify risk genes for COVID-19 across two GWAS summary datasets (Dataset #1: COVID\_I GWAS summary statistics, and Dataset #2: ANA5 GWAS summary statistics). After removing low-quality and non-matched SNPs, there were 8,424,883 high-quality SNPs with a MAF  $\geq 1\%$  and imputation  $R^2 \geq 0.6$  that were common to both datasets with the use of effect-size estimates (BETA) and their standard errors (SE) for the meta-analysis. With regard to the genome-wide meta-analysis, we adopted the widely-used threshold of  $5 \times 10^{-8}$  for combined P values to determine statistical significance. As reported in Ellinghaus et al. <sup>[8]</sup>, we used the combined P value and combined effect (E) with its SE generated by the METAL to compute the odds ratio (OR) and its 95% confidence interval (CI): 1)  $OR = \exp(E)$ ; 2) the upper confidence limit ( $OR_{95\ U} = \exp(E + 1.96 \cdot SE)$ ); 3) the lower confidence limit ( $OR_{95\ L} = \exp(E - 1.96 \cdot SE)$ ). The *qqman* package in R platform was used to generate Manhattan plot and quantile-quantile (QQ) plot. The web-access tool of *LocusZoom* <sup>[31]</sup> was used to visualize regional association plots (<http://locuszoom.sph.umich.edu/>).

### ***Gene-based association analysis***

We conducted a gene-based association analysis of our meta-GWAS summary data for COVID-19 by using the Multi-marker Analysis of GenoMic Annotation (MAGMA) <sup>[32]</sup>, which

utilizes a multiple regression method to identify multi-marker aggregated effects that account for SNP P values and linkage disequilibrium (LD) between SNPs. The analyzed SNP set of each gene was based on whether the SNP located in the gene body region or within extended +/- 20 kb downstream or upstream of the gene. The LDs among SNPs were calculated based on the 1,000 Genomes Phases 3 European Panel <sup>[33]</sup>. The Benjamini-Hochberg false discovery rate (FDR) method was used to correct the association results for multiple testing. The P value threshold of  $1.86 \times 10^{-5}$  was applied.

### ***Functional enrichment analysis***

To annotate the molecular functions and biological pathways of these COVID-19-associated genes (Supplemental Table S2), we performed a functional enrichment analysis by using the WebGestalt tool <sup>[34]</sup> based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and gene ontology (GO) terms. Using the overrepresentation analysis, the WebGestalt could identify functional association between COVID-19-associated genes and KEGG pathways. Furthermore, GO enrichment analysis was performed by using 3 categories of GO terms: biological process, cellular component, and molecular function. The redundancies of GO-terms were removed. The hypergeometric test was applied to assess the statistical significance. P values were corrected for multiple testing using the FDR, and a P value threshold of  $2.9 \times 10^{-4}$  was applied.

### ***Gene-property analysis***

We performed a MAGMA gene-property analysis<sup>[32]</sup>, which is implemented in FUMA<sup>[35]</sup>. The gene expression data from 83 tissues GTEx RNA-seq data (version 8) were used to parse the gene expression profiles. Expression values (TPM) were log2 transformed and average expression values were adopted per tissue. The gene-property analysis was conducted for 53 specific tissue types and 30 general tissue types, respectively. Bonferroni correction for multiple testing was used for the examined tissue types.

### ***Single-cell RNA-seq analysis for lung tissue***

We performed single-cell data analysis from normal lung tissue sequenced by using Smart-Seq2. This data set is a part of the Human Lung Atlas<sup>[36]</sup>, consisting of 50 cell populations across 4 compartments (epithelial, endothelial, stromal, and immune) of lung tissue. All 9,404 cells with distinct cellular identities were download from the Human Lung Atlas. Cells from 3 donors with FACS-sorted strategy were available at the Synapse (accession numbers: syn22168639, syn22168625, and syn22168622).

### ***S-PrediXcan and S-MultiXcan analysis***

We applied S-PrediXcan<sup>[37]</sup> to integrate expression quantitative trait loci (eQTL) data with genetic associations from GWAS summary statistics to identify genes, which genetically predicted expression levels are associated with COVID-19. S-PrediXcan firstly estimates gene expression weights by training a linear prediction model (MASHR model) in samples with both SNP genotype

and gene expression data. These estimated weights are processed with beta values and standard errors from meta-GWAS summary data on COVID-19 to predict gene expression from GWAS summary statistics, while combining the variance and co-variance of SNPs from an LD reference panel based on the 1000 Genomes Project Phase 3 genotypes<sup>[33]</sup>. The eQTL data for 49 tissues from the GTEx Project (version 8) were used in the current analysis. S-PrediXcan was performed for each of 49 tissues for a total of 659,158 gene-tissue pairs. To increase power to identify genes whose expression is similarly differentially regulated across tissues, we meta-analyzed the S-PrediXcan results with the use of the S-MultiXcan method<sup>[38]</sup>, which employs multivariate regression to integrate evidence across 49 GTEx tissues with a total of 22,327 genes. Significant associations were determined by using Bonferroni correction.

### ***Sherlock-based integrative genomics analysis***

To further validate these identified host genes for COVID-19, we also applied an independent approach of Sherlock-based integrative genomics analysis based on a Bayesian inference algorithm<sup>[39]</sup>. The Sherlock-based integrative analysis incorporates genetic information from meta-GWAS summary statistics on COVID-19 with eQTL data across 49 tissues from the GTEx Project (version 7) to prioritize important risk genes. It should be noted that Sherlock employs different algorithm and strategies to perform the statistical inference compared with S-PrediXcan. Briefly, Sherlock first searches expression-associated SNPs (named as eSNPs) across different GTEx tissues. Then, Sherlock estimates the possible association of eSNPs with COVID-19 using our meta-GWAS summary data. Sherlock computes individual Logarithm of the Bayes Factor (LBF) for each SNP



pair, and the sum of these constitutes the final LBF score for each gene. There are three potential scenarios: 1) if an eSNP in a given gene showed a significant association with COVID-19, a positive score would be assigned; 2) if an eSNP in a given gene showed a non-significant association with COVID-19, a negative score would be assigned; 3) no score was assigned if an SNP was not eSNP but showed a significant association with COVID-19. The Sherlock applied the simulation analysis to compute the p value of the Bayes factor for each gene, as reference of a method of Bayes/non-Bayes compromise <sup>[40]</sup>. We used the Bonferroni correction method to correct the significance for multiple testing. Due to we tested the Sherlock analysis in 49 GTEx tissues, there existed different thresholds across tissues. For example, the P value threshold was  $6.64 \times 10^{-6}$  ( $0.05/7,529$ ) for lung tissue.

### ***In silico permutation analysis***

To determine whether there exist a higher overlapped rate of genes identified from MAGMA analysis (Gene set #1:  $N = 1,005$ ,  $P < 0.05$ ) with genes from S-MultiXcan analysis (Gene set #2:  $N = 1,141$ ,  $P < 0.05$ ) than that from random selections, we conducted a computer-based permutation analysis of 100,000 times of random selections <sup>[41]</sup>. First, we counted the overlapped genes between Gen sets #1 and #2 ( $N_{\text{observation}}$ ). Secondly, we used all tested genes from S-MultiXcan analysis as background genes ( $N_{\text{background}} = 22,326$  genes). Then, through randomly selecting the same number of genes as Gene set #2 from background genes to compare with Gene set #1 for 100,000 times, we counted the number of overlapped genes in each time ( $N_{\text{random}}$ ). In the third step, we summed the counts of  $N_{\text{observation}} \leq N_{\text{random}}$  and divided by 100,000 to calculate empirically permuted P value.  $P <$

0.05 is considered to be significant.

### ***Drug-gene interaction analysis***

We submitted these 11 COVID-19-associated genes into the widely-used Drug Gene Interaction Database (DGIdb v.3.0.2; <http://www.dgiddb.org/>) to identify drug-gene interactions with Food and Drug Administration (FDA)-approved pharmaceutical compounds as well as antineoplastic and immunotherapies drugs depended on 20 databases with 51 known interaction types, and search 10 databases to find genes with potential drug abilities.

### **Supporting Information**

Supporting Information is available from the Wiley Online Library or from the author.

### **Acknowledgements**

We thanks to the helpful suggestions of Dr. Guijun Zhang. We appreciate all the authors from the COVID-19 Host Genetic Consortium, as well as Ellinghaus and his colleagues who have deposited and shared GWAS summary data on public databases. This study was funded by the National Natural Science Foundation of China (61871294 to J.S., 81501852 to N.W.), and Science Foundation of Zhejiang Province (LR19C060001 to J.S). All analysis codes in the Methods are available in a publicly available GitHub repository at

[https://github.com/YukuanHuang/Host\\_genetics\\_for\\_COVID-19](https://github.com/YukuanHuang/Host_genetics_for_COVID-19).

## Authors' contributions

J.S., N.W., Y.M. and J.Q. conceived and designed the study. Y.M., Y.H., S.Z., G.Q. and J.Z. contributed to management of data collection. Y.M., Y.H., G.Q., J.Z., J.Q., Z.M., J.Y., J.S., N.W. and S.Z. conducted bioinformatics analysis and data interpretation. Y.M., J.S., N.W. and S.Z. wrote the manuscripts. All authors reviewed and approved the final manuscript.

## Conflict of Interest

The authors declare no conflict of interest.

## Figure Legend

### **Figure 1. Meta-analysis of GWAS summary data highlighting susceptibility loci for COVID-19.**

a) Manhattan plot of the meta-analysis GWAS summary statistics highlighting three susceptibility loci for COVID-19. The Manhattan plot is shown of the meta-GWAS summary statistics of meta-analyzing the COVID\_I GWAS data (controlled for potential population stratification) with ANA5 GWAS data. The red horizontal line marks the genome-wide significance threshold of a P value less than  $5 \times 10^{-8}$ . b) Quantile-quantile (QQ) plot of the meta-analysis GWAS summary statistics. All 8,424,883 high-quality SNPs with a MAF  $\geq 1\%$  and imputation  $R^2 \geq 0.6$  were used for

plotting. In QQ plot the 2.5<sup>th</sup> and 97.5<sup>th</sup> centiles of the distribution under random sampling and the null hypothesis form the 95% concentration band. The genomic inflation factor lambda ( $\lambda$ ) is 1.0075.

c) Regional association plot for 21q22.11 locus of meta-GWAS summary statistics. Regional association plot is shown for 21q22.11 locus of the meta-GWAS summary statistics of meta-analyzing the COVID\_I GWAS data (controlled for potential population stratification) with ANA5 GWAS data. The purple diamond marks the most strongly associated SNP of rs9976829 with COVID-19. The color illustrates LD information with rs9976829, as shown in the color legend.

**Figure 2. Circus plot showed the results of gene-based association analysis.** Note: The inner ring shows the 22 autosomal human chromosomes (Chr1-22) and X chromosome (Chr23). A circular symbol in the outer ring represents a gene. Color indicates the statistical significance of genes (red marks genes significantly associated with COVID-19 with  $FDR < 0.05$ , yellow indicates genes with  $1.86 \times 10^{-5} < P \leq 1 \times 10^{-3}$ , green marks genes with  $1 \times 10^{-3} < P \leq 0.05$ , and gray represents genes with  $P > 0.05$ ).

**Figure 3. Functional enrichment analysis of genes associated with COVID-19.** a) Pathway enrichment analysis identified 9 significant or suggestive KEGG pathways enriched by COVID-19-associated genes. b) GO enrichment analysis identified 10 significant or suggestive GO-terms enriched by COVID-19-associated genes. a) - b) The green bar represents a suggestive enrichment ( $P < 0.05$ ), and the orange bar represents a significant enrichment ( $FDR < 0.05$ ). c) Scatter plot show the consistency of 11 risk genes identified from both MAGMA and S-MultiXcan

analysis. The vertical and horizontal dotted lines represent  $-\log_{10}(P = 0.05)$ . d) *In silico* permutation analysis of 100,000 times of random selections. This permutation analysis was used to compare the overlapped genes between MAGMA and S-MultiXcan (see Methods). The empirical P value is less than  $1 \times 10^{-5}$ .

#### **Figure 4. Expression of *IFNAR2* and *IL10RB* among 50 cellular populations from lung tissue.**

This plot is based on a data set as a part of the Human Lung Atlas, consisting of 50 cell populations across 4 compartments (epithelial, endothelial, stromal, and immune) of lung tissue (x axis). y axis represents the expression level with log transformed count.

#### **References**

- [1] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, E. H. Y. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Y. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, Z. Feng, *N Engl J Med* 2020, 382, 1199.
- [2] European journal of human genetics : EJHG 2020, 28, 715.
- [3] E. Dong, H. Du, L. Gardner, *The Lancet. Infectious diseases* 2020, 20, 533.
- [4] Z. Wu, J. M. McGoogan, *JAMA* 2020.
- [5] J. F. Chan, S. Yuan, K. H. Kok, K. K. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C. Yip, R. W. Poon, H. W. Tsoi, S. K. Lo, K. H. Chan, V. K. Poon, W. M. Chan, J. D. Ip, J. P. Cai, V. C. Cheng, H. Chen, C. K. Hui, K. Y. Yuen, *Lancet (London, England)* 2020, 395, 514; G. Onder, G. Rezza, S. Brusaferro, *JAMA* 2020.
- [6] F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, L. Guan, Y. Wei, H. Li, X. Wu, J. Xu, S. Tu, Y. Zhang, H. Chen, B. Cao, *Lancet (London, England)* 2020, 395, 1054; C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, *Lancet (London, England)* 2020, 395, 497.

- [7] P. Zhou, X. L. Yang, X. G. Wang, B. Hu, L. Zhang, W. Zhang, H. R. Si, Y. Zhu, B. Li, C. L. Huang, H. D. Chen, J. Chen, Y. Luo, H. Guo, R. D. Jiang, M. Q. Liu, Y. Chen, X. R. Shen, X. Wang, X. S. Zheng, K. Zhao, Q. J. Chen, F. Deng, L. L. Liu, B. Yan, F. X. Zhan, Y. Y. Wang, G. F. Xiao, Z. L. Shi, *Nature* 2020, 579, 270; R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W. J. Liu, D. Wang, W. Xu, E. C. Holmes, G. F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, *Lancet* (London, England) 2020, 395, 565.
- [8] D. Ellinghaus, F. Degenhardt, L. Bujanda, M. Buti, A. Albillos, P. Invernizzi, J. Fernández, D. Prati, G. Baselli, R. Asselta, M. M. Grimsrud, C. Milani, F. Aziz, J. Kässens, S. May, M. Wendorff, L. Wienbrandt, F. Uellendahl-Werth, T. Zheng, X. Yi, R. de Pablo, A. G. Chercoles, A. Palom, A. E. Garcia-Fernandez, F. Rodriguez-Frias, A. Zanella, A. Bandera, A. Protti, A. Aghemo, A. Lleo, A. Biondi, A. Caballero-Garralda, A. Gori, A. Tanck, A. Carreras Nolla, A. Latiano, A. L. Fracanzani, A. Peschuck, A. Julià, A. Pesenti, A. Voza, D. Jiménez, B. Mateos, B. Nafria Jimenez, C. Quereda, C. Paccapelo, C. Gassner, C. Angelini, C. Cea, A. Solier, D. Pestaña, E. Muñoz-Diaz, E. Sandoval, E. M. Paraboschi, E. Navas, F. García Sánchez, F. Ceriotti, F. Martinelli-Boneschi, F. Peyvandi, F. Blasi, L. Téllez, A. Blanco-Grau, G. Hemmrich-Stanisak, G. Grasselli, G. Costantino, G. Cardamone, G. Foti, S. Aneli, H. Kurihara, H. ElAbd, I. My, I. Galván-Femenia, J. Martín, J. Erdmann, J. Ferrusquía-Acosta, K. Garcia-Etxebarria, L. Izquierdo-Sanchez, L. R. Bettini, L. Sumoy, L. Terranova, L. Moreira, L. Santoro, L. Scudeller, F. Mesonero, L. Roade, M. C. Rühlemann, M. Schaefer, M. Carrabba, M. Riveiro-Barciela, M. E. Figuera Basso, M. G. Valsecchi, M. Hernandez-Tejero, M. Acosta-Herrera, M. D'Angiò, M. Baldini, M. Cazzaniga, M. Schulzky, M. Cecconi, M. Wittig, M. Ciccarelli, M. Rodríguez-Gandía, M. Bocciolone, M. Miozzo, N. Montano, N. Braun, N. Sacchi, N. Martínez, O. Özer, O. Palmieri, P. Faverio, P. Preatoni, P. Bonfanti, P. Omodei, P. Tentorio, P. Castro, P. M. Rodrigues, A. Blandino Ortiz, R. de Cid, R. Ferrer, R. Gualtierotti, R. Nieto, S. Goerg, S. Badalamenti, S. Marsal, G. Matullo, S. Pelusi, S. Juzenas, S. Aliberti, V. Monzani, V. Moreno, T. Wesse, T. L. Lenz, T. Pumarola, V. Rimoldi, S. Bosari, W. Albrecht, W. Peter, M. Romero-Gómez, M. D'Amato, S. Duga, J. M. Banales, J. R. Hov, T. Folseraas, L. Valenti, A. Franke, T. H. Karlsen, *N Engl J Med* 2020.
- [9] L. C. Platanias, *Nature reviews. Immunology* 2005, 5, 375.
- [10] F. J. Barrat, L. Su, *The Journal of experimental medicine* 2019, 216, 1974; M. Macal, Y. Jo, S. Dallari, A. Y. Chang, J. Dai, S. Swaminathan, E. J. Wehrens, P. Fitzgerald-Bocarsly, E. I. Zúñiga, *Immunity* 2018, 48, 730.
- [11] J. Cinatl, B. Morgenstern, G. Bauer, P. Chandra, H. Rabenau, H. W. Doerr, *Lancet* (London, England) 2003, 362, 293.
- [12] M. R. Loutfy, L. M. Blatt, K. A. Siminovitch, S. Ward, B. Wolff, H. Lho, D. H. Pham, H. Deif, E. A. LaMere, M. Chang, K. C. Kain, G. A. Farcas, P. Ferguson, M. Latchford, G. Levy, J. W. Dennis, E. K. Lai, E. N. Fish, *JAMA* 2003, 290, 3222.
- [13] I. F. Hung, K. C. Lung, E. Y. Tso, R. Liu, T. W. Chung, M. Y. Chu, Y. Y. Ng, J. Lo, J. Chan, A. R. Tam, H. P. Shum, V. Chan, A. K. Wu, K. M. Sin, W. S. Leung, W. L. Law, D. C. Lung, S. Sin, P. Yeung, C. C. Yip, R. R. Zhang, A. Y. Fung, E. Y. Yan, K. H. Leung, J. D. Ip, A. W. Chu, W. M. Chan, A. C. Ng, R. Lee, K. Fung, A. Yeung, T. C. Wu, J. W. Chan, W. W. Yan, W. M. Chan, J. F. Chan, A. K. Lie, O. T. Tsang, V. C. Cheng, T. L. Que, C. S. Lau, K. H. Chan, K. K. To, K. Y. Yuen, *Lancet* 2020, 395, 1695.
- [14] X. Lei, X. Dong, R. Ma, W. Wang, X. Xiao, Z. Tian, C. Wang, Y. Wang, L. Li, L. Ren, F. Guo, Z. Zhao, Z. Zhou, Z. Xiang, J. Wang, *Nat Commun* 2020, 11, 3810.

- [15] C. J. Duncan, S. M. Mohamad, D. F. Young, A. J. Skelton, T. R. Leahy, D. C. Munday, K. M. Butler, S. Morfopoulou, J. R. Brown, M. Hubank, J. Connell, P. J. Gavin, C. McMahon, E. Dempsey, N. E. Lynch, T. S. Jacques, M. Valappil, A. J. Cant, J. Breuer, K. R. Engelhardt, R. E. Randall, S. Hambleton, *Sci Transl Med* 2015, 7, 307ra154.
- [16] D. Blanco-Melo, B. E. Nilsson-Payant, W. C. Liu, S. Uhl, D. Hoagland, R. Moller, T. X. Jordan, K. Oishi, M. Panis, D. Sachs, T. T. Wang, R. E. Schwartz, J. K. Lim, R. A. Albrecht, B. R. tenOever, *Cell* 2020, 181, 1036.
- [17] J. Hadjadj, N. Yatim, L. Barnabei, A. Corneau, J. Boussier, N. Smith, H. Péré, B. Charbit, V. Bondet, C. Chenevier-Gobeaux, P. Breillat, N. Carlier, R. Gauzit, C. Morbieu, F. Pène, N. Marin, N. Roche, T. A. Szwed, S. H. Merkl, J. M. Treluyer, D. Veyer, L. Mouthon, C. Blanc, P. L. Tharaux, F. Rozenberg, A. Fischer, D. Duffy, F. Rieux-Laucat, S. Kernéis, B. Terrier, *Science* 2020, 369, 718.
- [18] K. M. Shephardson, K. Larson, L. L. Johns, K. Stanek, H. Cho, J. Wellham, H. Henderson, A. Rynda-Appl, *Frontiers in immunology* 2018, 9, 2589.
- [19] Q. M. Gong, X. F. Kong, Z. T. Yang, J. Xu, L. Wang, X. H. Li, G. D. Jin, J. Gao, D. H. Zhang, J. H. Jiang, Z. M. Lu, X. X. Zhang, *Journal of viral hepatitis* 2009, 16, 674.
- [20] D. E. Gordon, G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K. M. White, M. J. O'Meara, V. V. Rezell, J. Z. Guo, D. L. Swaney, T. A. Tummino, R. Hüttenhain, R. M. Kaake, A. L. Richards, B. Tutuncuoglu, H. Foussard, J. Batra, K. Haas, M. Modak, M. Kim, P. Haas, B. J. Polacco, H. Braberg, J. M. Fabius, M. Eckhardt, M. Southeray, M. J. Bennett, M. Cakir, M. J. McGregor, Q. Li, B. Meyer, F. Roesch, T. Vallet, A. Mac Kain, L. Miorin, E. Moreno, Z. Z. C. Naing, Y. Zhou, S. Peng, Y. Shi, Z. Zhang, W. Shen, I. T. Kirby, J. E. Melnyk, J. S. Chorb, K. Lou, S. A. Dai, I. Barrio-Hernandez, D. Memon, C. Hernandez-Armenta, J. Lyu, C. J. P. Mathy, T. Perica, K. B. Pilla, S. J. Ganesan, D. J. Saltzberg, R. Rakesh, X. Liu, S. B. Rosenthal, L. Calviello, S. Venkataramanan, J. Liboy-Lugo, Y. Lin, X. P. Huang, Y. Liu, S. A. Wankowicz, M. Bohn, M. Safari, F. S. Ugur, C. Koh, N. S. Savar, Q. D. Tran, D. Shengjuler, S. J. Fletcher, M. C. O'Neal, Y. Cai, J. C. J. Chang, D. J. Broadhurst, S. Klippsten, P. P. Sharp, N. A. Wenzell, D. Kuzuoglu-Ozturk, H. Y. Wang, R. Trenker, J. M. Young, D. A. Cavero, J. Hiatt, T. L. Roth, U. Rathore, A. Subramanian, J. Noack, M. Hubert, R. M. Stroud, A. D. Frankel, O. S. Rosenberg, K. A. Verba, D. A. Agard, M. Ott, M. Emerman, N. Jura, M. von Zastrow, E. Verdin, A. Ashworth, O. Schwartz, C. d'Enfert, S. Mukherjee, M. Jacobson, H. S. Malik, D. G. Fujimori, T. Ideker, C. S. Craik, S. N. Floor, J. S. Fraser, J. D. Gross, A. Sali, B. L. Roth, D. Ruggero, J. Taunton, T. Kortemme, P. Beltrao, M. Vignuzzi, A. García-Sastre, K. M. Shokat, B. K. Shoichet, N. J. Krogan, *Nature* 2020.
- [21] E. Fossum, G. Grødeland, D. Terhorst, A. A. Tveita, E. Vikse, S. Mjaaland, S. Henri, B. Malissen, B. Bogen, *European journal of immunology* 2015, 45, 624.
- [22] A. N. Wein, S. R. McMaster, S. Takamura, P. R. Dunbar, E. K. Cartwright, S. L. Hayward, D. T. McManus, T. Shimaoka, S. Ueha, T. Tsukui, T. Masumoto, M. Kurachi, K. Matsushima, J. E. Kohlmeier, *The Journal of experimental medicine* 2019, 216, 2748.
- [23] J. Wang, F. Li, H. Wei, Z. X. Lian, R. Sun, Z. Tian, *The Journal of experimental medicine* 2014, 211, 2397; S. M. Henson, R. Snelgrove, T. Hussell, D. J. Wells, R. Aspinall, *Journal of immunology (Baltimore, Md. : 1950)* 2005, 175, 4112; J. Zhou, H. K. Law, C. Y. Cheung, I. H. Ng, J. S. Peiris, Y. L. Lau, *The Journal of infectious diseases* 2006, 194, 61.

- [24] W. Wang, Y. Xu, R. Gao, R. Lu, K. Han, G. Wu, W. Tan, JAMA 2020, 323, 1843.
- [25] M. Poenisch, P. Metz, H. Blankenburg, A. Ruggieri, J. Y. Lee, D. Rupp, I. Rebhan, K. Diederich, L. Kaderali, F. S. Domingues, M. Albrecht, V. Lohmann, H. Erfle, R. Bartenschlager, PLoS pathogens 2015, 11, e1004573.
- [26] G. D. Kanade, K. D. Pingale, Y. A. Karpe, Frontiers in microbiology 2019, 10, 2501.
- [27] P. L. Tsai, N. T. Chiou, S. Kuss, A. García-Sastre, K. W. Lynch, B. M. Fontoura, PLoS pathogens 2013, 9, e1003460.
- [28] L. Fang, X. Sun, Y. Wang, L. Du, K. Ji, J. Wang, N. He, Y. Liu, Q. Wang, H. Zhai, J. Hao, C. Xu, Q. Liu, FASEB journal : official publication of the Federation of American Societies for Experimental Biology 2019, 33, 5561.
- [29] G. Abraham, Y. Qiu, M. Inouye, Bioinformatics 2017, 33, 2776.
- [30] C. J. Willer, Y. Li, G. R. Abecasis, Bioinformatics 2010, 26, 2190.
- [31] R. J. Pruim, R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, T. P. Gliedt, M. Boehnke, G. R. Abecasis, C. J. Willer, Bioinformatics 2010, 26, 2336.
- [32] C. A. de Leeuw, J. M. Mooij, T. Heskes, D. Posthuma, PLoS computational biology 2015, 11, e1004219.
- [33] A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, Nature 2015, 526, 68.
- [34] B. Zhang, S. Kirov, J. Snoddy, Nucleic acids research 2005, 33, W741.
- [35] K. Watanabe, E. Taskesen, A. van Bochoven, D. Posthuma, Nat Commun 2017, 8, 1826.
- [36] K. J. Travaglini, A. N. Nabhan, L. Penland, R. Sinha, A. Gillich, R. V. Sit, S. Chang, S. D. Conley, Y. Mori, J. Seita, G. J. Berry, J. B. Shrager, R. J. Metzger, C. S. Kuo, N. Neff, I. L. Weissman, S. R. Quake, M. A. Krasnow, bioRxiv 2020, 742320.
- [37] A. N. Barbeira, S. P. Dickinson, R. Bonazzola, J. Zheng, H. E. Wheeler, J. M. Torres, E. S. Torstenson, K. P. Shah, T. Garcia, T. L. Edwards, E. A. Stahl, L. M. Huckins, G. T. Consortium, D. L. Nicolae, N. J. Cox, H. K. Im, Nat Commun 2018, 9, 1825.
- [38] A. N. Barbeira, M. Pividori, J. Zheng, H. E. Wheeler, D. L. Nicolae, H. K. Im, PLoS genetics 2019, 15, e1007889.
- [39] X. He, C. K. Fuller, Y. Song, Q. Meng, B. Zhang, X. Yang, H. Li, American journal of human genetics 2013, 92, 667.
- [40] B. Servin, M. Stephens, PLoS genetics 2007, 3, e114.
- [41] N. Akula, J. R. Wendland, K. H. Choi, F. J. McMahon, Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology 2016, 41, 886.





**Table 1. Susceptibility loci associated with COVID-19 identified by meta-analysis of GWAS summary data**

SNP	CHR	Position	Loci	ALT	REF	COVID_I			ANA5			Meta-analysis		
						OR	95% CI	P-value	OR	95% CI	P-value	OR	95% CI	P-value
rs11385942	3	45876459	3p21.31	GA	G	1.77	1.49-2.11	$1.15 \times 10^{-10}$	1.50	1.29-1.74	$1.10 \times 10^{-7}$	1.61	1.43-1.80	$2.87 \times 10^{-16}$
rs8176719	9	136132908	9q34.2	TC	T	1.32	1.19-1.46	$9.93 \times 10^{-8}$	1.10	1.01-1.19	$2.10 \times 10^{-2}$	1.17	1.10-1.25	$4.76 \times 10^{-7}$
rs657152	9	136139265	9q34.2	A	C	1.33	1.20-1.47	$4.95 \times 10^{-8}$	1.07	0.99-1.15	$7.86 \times 10^{-2}$	1.15	1.08-1.21	$5.53 \times 10^{-6}$
rs9976829	21	34614834	21q22.11	G	A	1.18	1.07-1.32	$1.77 \times 10^{-3}$	1.15	1.06-1.24	$3.58 \times 10^{-4}$	1.16	1.09-1.23	$2.57 \times 10^{-6}$

Note: CHR = chromosome, OR = odds ratio, 95% CI = 95% confidence interval, ALT = Altered allele, REF = Reference allele, COVID\_I = COVID\_I GWAS summary statistics (Dataset #1), ANA5 = ANA5 GWAS summary statistics (Dataset #2). The meta-analysis data were based on the combination of COVID\_I GWAS summary data (controlled for potential population stratification) with ANA5 GWAS summary data.

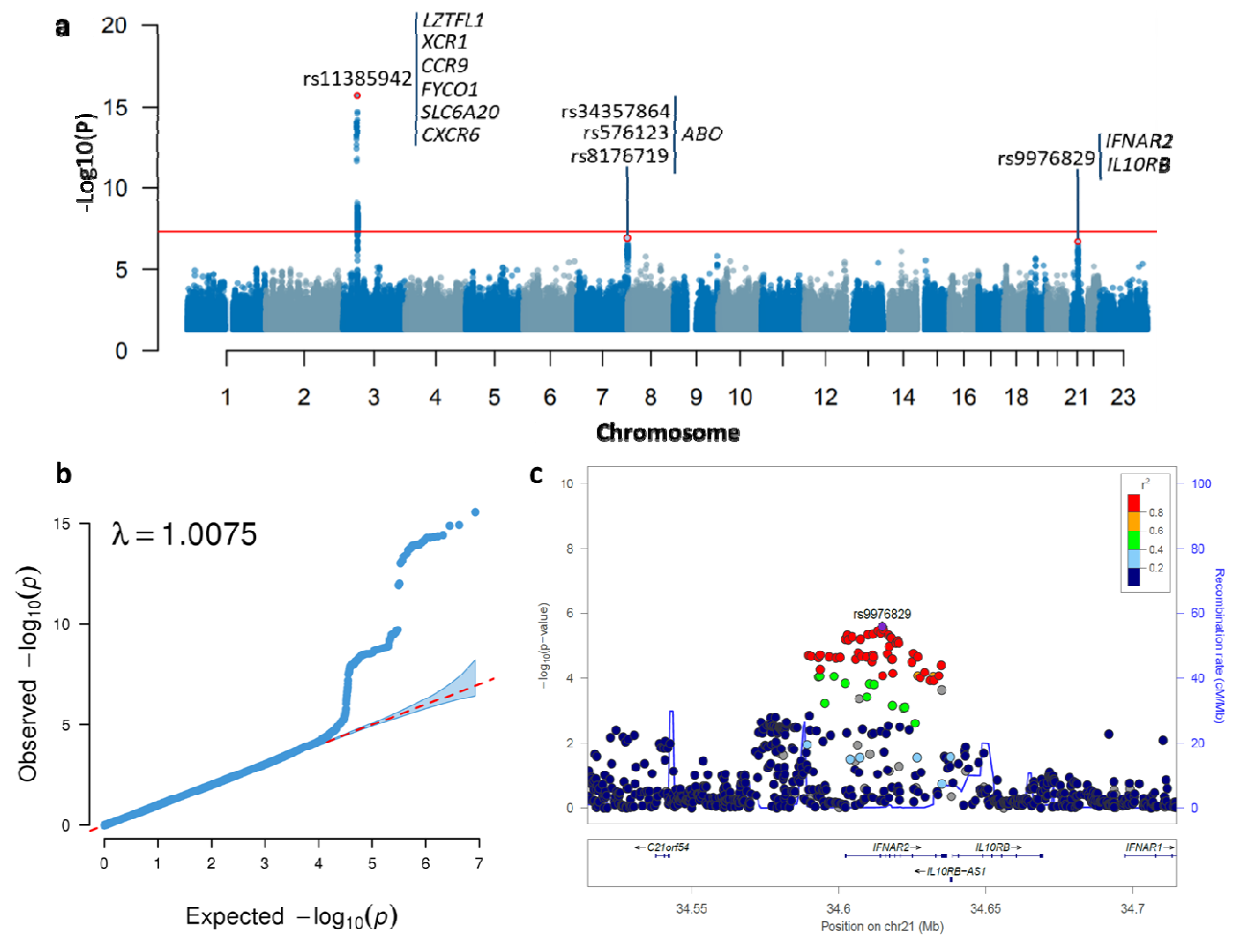
**Table 2. Significant genes associated with Covid-19 identified by MAGMA gene-based association analysis**

Gene	CHR	Start position	Stop position	Loci	MAGMA on COVID_I		MAGMA on ANA5		MAGMA on meta-analysis data		
					Z score	P-value	Z score	P-value	Z score	P-value	FDR
<i>LZTFL1</i>	3	45844808	45977216	3p21.31	5.71	$5.61 \times 10^{-9}$	4.51	$3.29 \times 10^{-6}$	7.05	$8.83 \times 10^{-13}$	$1.69 \times 10^{-8}$
<i>XCRI</i>	3	46042291	46088979	3p21.31	4.82	$7.13 \times 10^{-7}$	4.14	$1.71 \times 10^{-5}$	6.03	$8.41 \times 10^{-10}$	$8.03 \times 10^{-6}$
<i>CCR9</i>	3	45907996	45964667	3p21.31	4.81	$7.70 \times 10^{-7}$	3.61	$1.54 \times 10^{-4}$	5.96	$1.28 \times 10^{-9}$	$8.14 \times 10^{-6}$
<i>FYCO1</i>	3	45939391	46057316	3p21.31	4.28	$9.55 \times 10^{-6}$	3.32	$4.49 \times 10^{-4}$	5.50	$1.95 \times 10^{-8}$	$9.31 \times 10^{-5}$
<i>SLC6A20</i>	3	45776941	45858039	3p21.31	4.05	$2.60 \times 10^{-5}$	2.72	$3.26 \times 10^{-3}$	5.11	$1.58 \times 10^{-7}$	$6.03 \times 10^{-4}$
<i>IFNAR2</i>	21	34582231	34656831	21q22.11	3.53	$2.10 \times 10^{-4}$	3.50	$2.31 \times 10^{-4}$	5.02	$2.58 \times 10^{-7}$	$7.58 \times 10^{-4}$

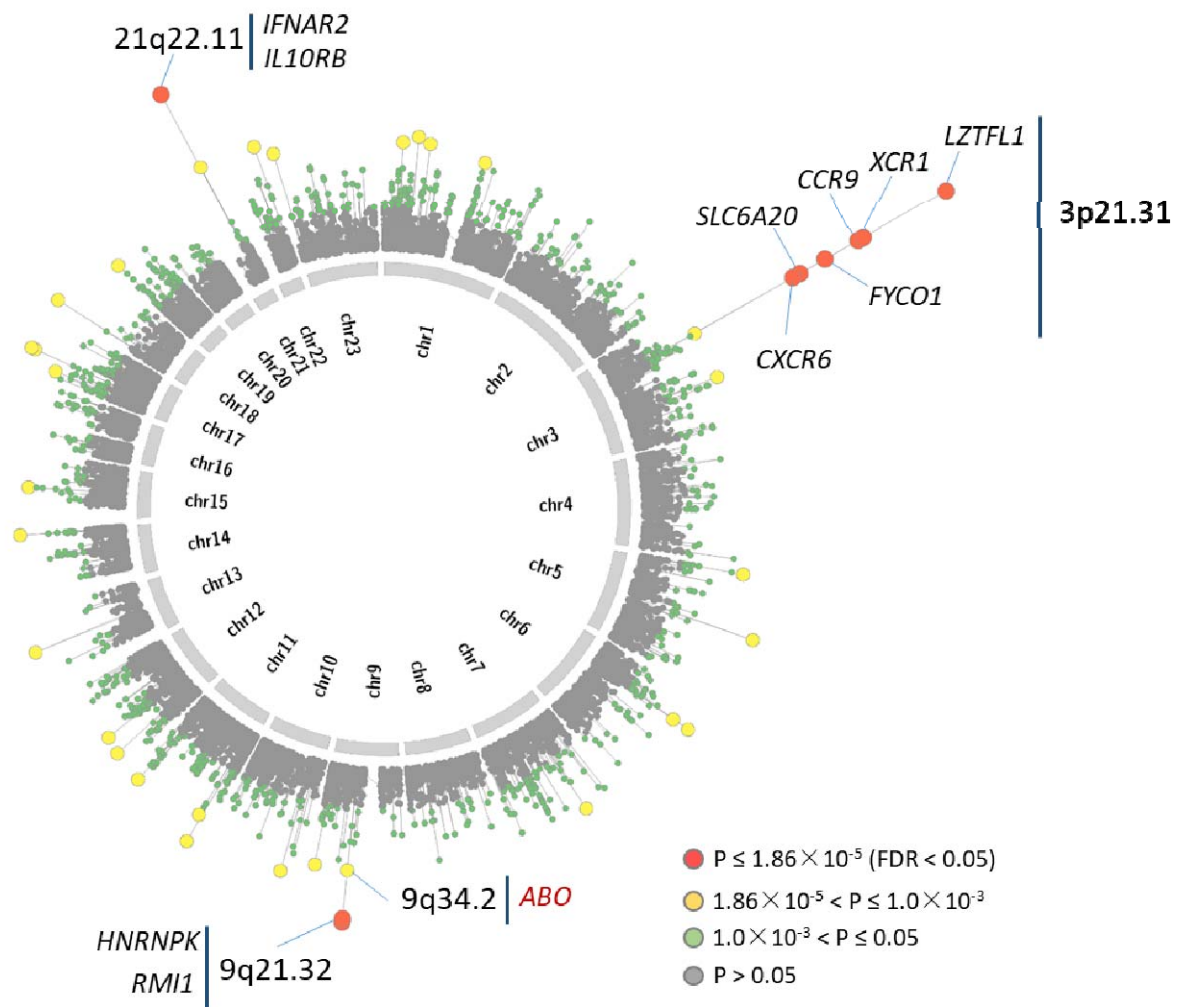
<i>CXCR6</i>	3	45964973	46009845	3p21.31	3.85	$5.99 \times 10^{-5}$	3.13	$8.83 \times 10^{-4}$	5.01	$2.78 \times 10^{-7}$	$7.58 \times 10^{-4}$
<i>HNRNPK</i>	9	86562998	86615692	9q21.32	2.02	$2.16 \times 10^{-2}$	3.18	$7.42 \times 10^{-4}$	4.18	$1.46 \times 10^{-5}$	$3.48 \times 10^{-2}$
<i>RMI1</i>	9	86575321	86638989	9q21.32	2.09	$1.84 \times 10^{-2}$	3.15	$8.29 \times 10^{-4}$	4.12	$1.86 \times 10^{-5}$	$3.94 \times 10^{-2}$
<i>ABO</i>	9	136110563	136170630	9q34.2	4.85	$6.28 \times 10^{-7}$	0.10	0.46	3.21	$6.55 \times 10^{-4}$	0.43

Note: CHR = chromosome, FDR = False discovery rate, COVID\_I = COVID\_I GWAS summary statistics (Dataset #1), ANA5 = ANA5 GWAS summary statistics (Dataset #2). The meta-analysis data were based on the combination of COVID\_I GWAS summary data (controlled for potential population stratification) with ANA5 GWAS summary data.

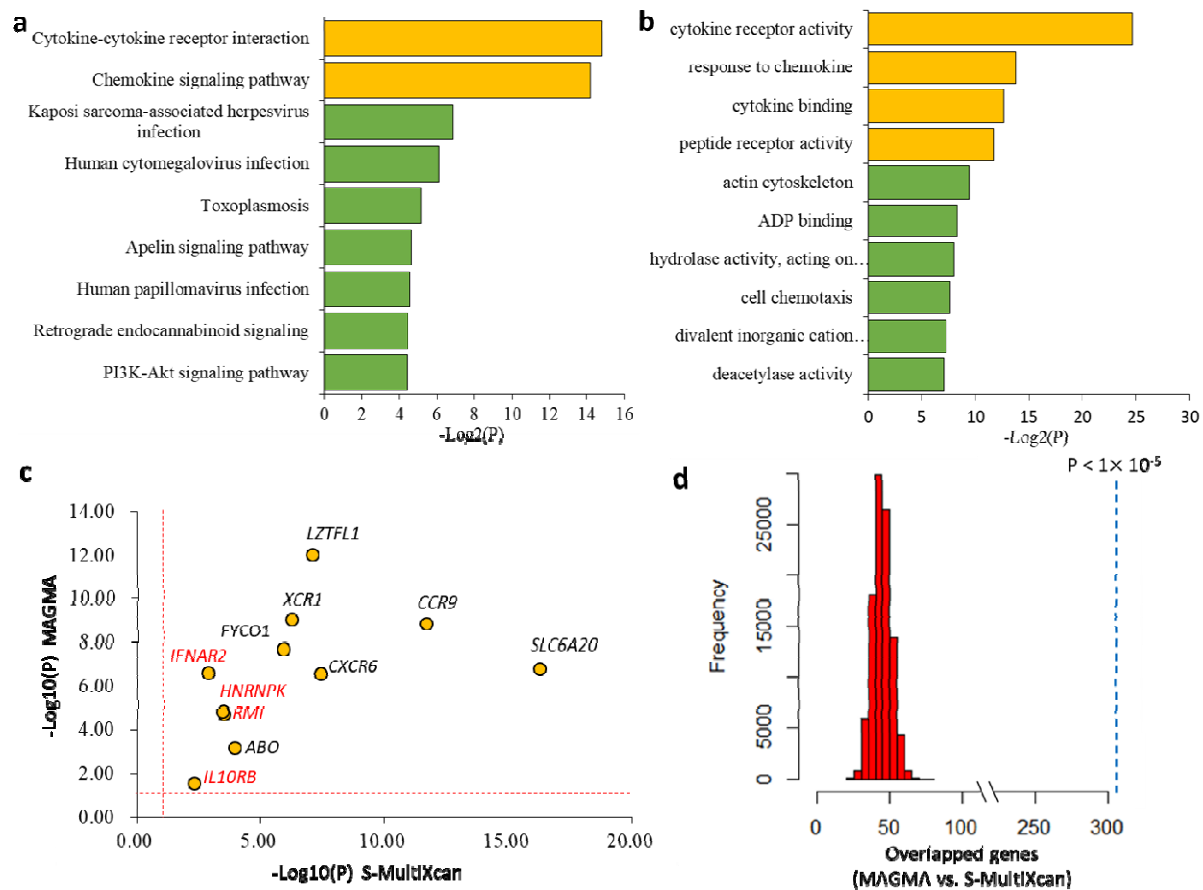
**Figure 1.**



**Figure 2.**



**Figure 3.**



**Figure 4.**

