Rare Variants in Inborn Errors of Immunity Genes Associated With Covid-19 Severity

Panhong Liu[1,2], Mingyan Fang[2,3], Yuxue Luo[2], Fang Zheng[4], Yan Jin[4], Fanjun Cheng[5], Huanhuan Zhu[2*] and Xin Jin[2,3,6*]

[1] College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, [2] Beijing Genomeics Institute At Shenzhen, BGI-Shenzhen, Shenzhen, China, [3] Beijing Genomeics Institute In Singapore, BGI-Singapore ,Singapore, Singapore, [4] Department of Pediatrics, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [5] Department of Hematology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [6] School of Medicine, South China University of Technology, Guangzhou, China

Host genetic factors have been shown to play an important role in SARS-CoV-2 infection and the course of Covid-19 disease. The genetic contributions of common variants influencing Covid-19 susceptibility and severity have been extensively studied in diverse populations. However, the studies of rare genetic defects arising from inborn errors of immunity (IEI) are relatively few, especially in the Chinese population. To fill this gap, we used a deeply sequenced dataset of nearly 500 patients, all of Chinese descent, to investigate putative functional rare variants. Specifically, we annotated rare variants in our call set and selected likely deleterious missense (LDM) and high-confidence predicted loss-of-function (HC-pLoF) variants. Further, we analyzed LDM and HC-pLoF variants between *non-severe* and *severe* Covid-19 patients by (a) performing gene- and pathway-level association analyses, (b) testing the number of mutations in previously reported genes mapped from LDM and HC-pLoF variants, and (c) uncovering candidate genes *via* protein-protein interaction (PPI) network analysis of Covid-19-related genes and genes defined from LDM and HC-pLoF variants. From our analyses, we found that (a) pathways Tuberculosis (hsa:05152), Primary Immunodeficiency (hsa:05340), and Influenza A (hsa:05164) showed significant enrichment in *severe* patients compared to the *non-severe* ones, (b) HC-pLoF mutations were enriched in Covid-19-related genes in *severe* patients, and (c) several candidate genes, such as *IL12RB1*, *TBK1*, *TLR3*, and *IFNGR2*, are uncovered by PPI network analysis and worth further investigation. These regions generally play an essential role in regulating antiviral innate immunity responses to foreign pathogens and in responding to many inflammatory diseases. We believe that our identified candidate genes/pathways can be potentially used as Covid-19 diagnostic markers and help distinguish patients at higher risk.

Keywords: Covid-19, inborn errors of immunity, gene-level tests, pathway-based analysis, PPI network analysis

INTRODUCTION

Since December 2019, the coronavirus diseases 2019 (Covid-19) (Gorbalenya et al., 2020) caused by the SARS-CoV-2 virus (Severe Acute Respiratory Syndrome Coronavirus 2) (Zhu et al., 2020) has spread rapidly across the world. By January 2022, the ongoing SARS-CoV-2 pandemic has caused more than 360 million confirmed cases and more than 5 million deaths. Host genetic factors have been shown to play critical roles in the disease susceptibility and severity (Ellinghaus et al., 2020; COVID-19 Host Genetics Initiative, 2021; Pairo-Castineira et al., 2021; Shelton et al., 2021; Kousathanas et al., 2022). The Covid-19 Host Genetics Initiative (Covid-19 HGI, https://www.covid19hg.org/) is an international initiative to share the results of host genome-wide associations study (GWAS) meta-analysis of Covid-19 disease. The most recent Covid-19 HGI release 6 has reported 24 lead SNPs (P < 5e-8) mapped to nearly 136 genes, such as *LZTFL1*, *ABO*, *OAS1*, *DPP9*, and *IFNAR2* (COVID-19 Host Genetics Initiative, 2021). The estimated heritability of Covid-19 symptoms explained by these common variants was 6.5% (Pairo-Castineira et al., 2021). A twin study with participants from the TwinsUK cohort reported that 31% of phenotypic variance of predicted Covid-19 is due to host genetic factors (Williams et al., 2020). This leads to a large proportion of unexplained heritability (nearly 25%), commonly referred to as "missing heritability". There is increasing evidence that rare variants also make a major contribution to missing heritability of many complex diseases and traits (Hunt et al., 2013; Zuk et al., 2014; Misawa et al., 2020).

Recently, the rare variants attracted researchers' attention in explaining the missing heritability of Covid-19 disease. For example, Zhang et al. found that the rare predicted loss-of-function (pLoF) variants in the IRF7- and TLR3-dependent type I interferon (IFN) pathway were enriched in patients who developed risky Covid-19 (Zhang Q. et al., 2020). Smieszek et al. reported that pLoF variant in gene *IFNAR2* (c.966C>A/ p.Y322X) might play a role not only in clinical manifestation of Covid-19 but also in the response to vaccination (Smieszek et al., 2021). In addition, multiple studies found that the pLoF variants in *TLR7* gene enriched in severely affected male patients, and the deficiency of TLR7 would impair innate immunity and increase severity of COVID-19 (van der Made et al., 2020; Asano et al., 2021; Fallerini et al., 2021; Mantovani et al., 2021). As previously reported, the rare variants were more likely to be functional and tended to have stronger effects on complex diseases (Gorlov et al., 2011). The study of genetic effects of rare variants is necessary to elucidate the severity of Covid-19.

To explore the genetic contributions of rare variants in Covid-19 patients with inborn errors of immunity (IEI) genes, we recruited and investigated nearly 500 hospitalized patients from Union Hospital of Tongji Medical College of Huazhong University of Science and Technology (abbr. Union Hospital) (Zhu et al., 2021). Based on patients' genomic data and clinical information, we carried out three major analyses to investigate the effects of putative functional rare variants (LDM and HC-pLoF): (a) gene- and pathway-level tests of these rare variants between *severe* and *non-severe* patients; (b) examination of the significance of previously reported rare variants and genes in our dataset; and c) rare mutation accumulation analysis and PPI network analysis in only *severe* patients. From these analyses, we (a) identified candidate functional pathways that are responsible for innate immune disorders and respiratory diseases, such as Tuberculosis (hsa:05152), Primary Immunodeficiency (hsa:05340), and Influenza A (hsa:05164); (b) successfully replicated two Covid-19 associated SNPs (rs780744847 and rs541048548) mapped on genes *TLR3* and *ICAM3*, respectively; and (c) suggested several candidate genes, including *IL12RB1, TBK1, TLR3,* and *IFNGR2*, which might be involved in SARS-CoV-2 cell entry, host immune responses, and Covid-19 disease severity.

Until now, literature based on the Chinese population has replicated and discovered some Covid-19-associated common variants (Wang et al., 2020; Wu et al., 2021a; Wu et al., 2021b; Zhu et al., 2021), but genetic background of rare variants is currently insufficiently understood in the Chinese population. Our work is an effort to fill this gap. We hope that it will serve as a useful scientific reference to assess the genetic mechanism of rare variants in Covid-19 and advance our understanding of disease etiology.

MATERIALS AND METHODS

Patient Recruitment and Quality Control

All subjects in this study were collected from the Union Hospital. We used PLINK 2.0 (Chang et al., 2015) to infer sex of individuals from SNP genotypes and VerifyBamID (Zhang F. et al., 2020) to assess DNA contamination level. Individuals were excluded if their inferred sex was inconsistent with that of clinically recorded. We also removed individuals with estimated contamination rates greater than 0.05. After sample quality control, there were 451 unrelated individuals with 159 mild and moderate patients, and 292 severe and critical patients. The severity classification criteria were made by the National Health Commission of P.R. China (Wu and McGoogan, 2020). We reclassified the mild/moderate patients as *non-severe* patients and the severe/critical patients as *severe* patients.

Whole Genome Sequencing

Cell-free DNA (cf-DNA) was extracted from 200 ul plasma using MagPure Circulating DNA Kit following the manufacturer's instructions. The 40 ul cf-DNA was extracted from each sample and used to create a library using MGIEasy Cell-free DNA Library Prep Kit according to the library preparation pipeline. Sequencing was conducted by the DNBSEQ platform (MGI, Shenzhen, China) to generate 100 bp paired-end reads. The mean sequencing depth was 17.8× for all samples.

Genotype Calling

The blood samples of some patients were collected at different time points during hospitalization. To increase the average depth of study, sequence fastq files of each patient were merged together to generate one GVCF file by BWA (Li and Durbin, 2009) and Sentieon Genomics software (Freed et al., 2017).

Joint variant calling was then performed on GVCF files of all participants using the Sentieon GVCFtyper algorithm. The resulting VCF file was used for subsequent genomic analyses.

After the application of excessHet (<54.69) filter, Variant Quality Score Recalibration (VQSR) was completed by using the Genome Analysis Toolkit (GATK version 4.1.2) (DePristo et al., 2011). Known variant files were downloaded from the GATK bundle. For SNP sets, we used SNPs of HapMap, 1000G_omin and 1000G_phase1 database as training sets, the SNPs of HapMap as true sets, and the SNPs of dbSNP as known sets. For indel sets, we used indel of Mills_and_1000G_gold_standard database as training and true sets. We used the annotations "DP", "QD", "MQRankSum", "ReadPosRankSum", "FS", and "SOR" to train VQSR. Finally, we used a sensitivity threshold of 99.7% and 99% for the SNPs and INDELs, respectively, to define genotyped sites that passed VQSR filtration.

To improve the genotyping accuracy, we used the Beagle 4.0 software (Browning and Browning, 2016) to perform LD-based genotype refinement and imputation by taking genotype likelihoods as inputs. Low-quality variants with dosage imputation score DR2 < 0.3 were filtered out.

Principal Component Analysis

Principal component analysis (PCA) (Pearson, 1901) was performed using a subset of autosomal bi-allelic SNPs by applying PLINK 2.0 (Chang et al., 2015). Several restrictions were applied to select SNPs for PCA analysis, including keeping SNPs with minor allele frequency (MAF) ≥ 5%, Hardy–Weinberg Equilibrium P ≥ 1e-6, and removing one of a pair of SNPs if the LD was greater than 0.5 (in a window of 50 SNPs with a step of 5 SNPs). To select an adequate number of significant PCs in the association analysis, we performed hypothesis testing for each eigenvalue by using software Eigenstrat (Price et al., 2006).

Functional Annotation

We annotated rare variants (MAF < 0.5%) in our final call set by using the Ensembl Variant Effect Predictor (VEP, build 103, GRCh38) (McLaren et al., 2016) with default parameters. The databases for annotation included dbSNP (Sherry et al., 2001), gnomAD (Karczewski et al., 2020), and 1000 Genomes Project (Clarke et al., 2012). In addition, we used Combined Annotation Dependent Depletion (CADD) score to predict missense variants that had potential effects on protein function. The CADD score was annotated by CADD plug-in (Kircher et al., 2014). Missense variants with CADD score > MSC (Mutation Significance Cutoff) score (95% confidence interval) (Itan et al., 2016) were predicted as likely deleterious missense (LDM) variants. We also used LOFTEE (Karczewski et al., 2020) plug-in to identify high-confidence pLoF (HC-pLoF) for stop-gained, frameshift, and splice site disrupting variants. Finally, we focused on the LDM and HC-pLoF variants in the subsequent analyses.

Rare Variants Analyses

To investigate the cumulative effects of multiple rare variants, we performed gene-based association analysis using KGGSeq 1.0

(Li et al., 2017) with the sequence kernel association test (SKAT) (Wu et al., 2010), the Optimized SKAT (SKAT-O) (Lee et al., 2012), and Burden test. We used the binary collapsing method (burden test) implemented in KGGseq. We further carried out pathway-based analysis by testing the Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets (Kanehisa and Goto, 2000). The adjusted covariates included age, sex, and the top six principal components. We defined the suggestive significance threshold for gene-based association test as 1e–6 and for pathway-based association test as 0.05.

We also focused on 24 type I IFN genes (denoted as IFN-genes) that were found as an enrichment in a life-threatening Covid-19 study (Zhang Q. et al., 2020; Kosmicki et al., 2021) and 136 genes located in 50 kb of lead SNPs reported by the Covid-19 Host Genetics Initiative (release 6, denoted as HGI-genes) (COVID-19 Host Genetics Initiative, 2021) (**Supplementary Table S1**). The mutation accuracy of variants in these 159 known candidate genes (one overlap between 24 IFN-genes and 136 HGI-genes) was manually checked by using Samtools 1.10 (Danecek et al., 2021).

Finally, we performed an analysis of rare variant accumulation in genes identified by two approaches. The first approach detected genes if one variant met the following two conditions: (a) the mutations occurred in only *severe* patients, and (b) the variant harbored no less than three effect allele counts. We denoted these genes as "individual variant-driven" genes. The second approach determined genes if (a) all mutations in the gene occurred in only *severe* patients, and (b) the total number of mutations in the gene is at least three. We denoted these genes as "all variant-driven" genes. We note that genes identified by the two methods may have some overlap. Each of the two gene sets was then used for PPI network analysis with the above 159 known candidate genes. We used the STRING version 10.5 (Search Tool for the Retrieval of Interacting Genes/Proteins) (Szklarczyk et al., 2019) to build the PPI network. The minimum required interaction score to the highest confidence was set to 0.900.

RESULTS

Participant Characteristics

In this study, participants included 451 Covid-19 patients aged 23 to 97 years old and all declared Han Chinese population. In **Table 1**, we provide participant characteristics for *non-severe* and *severe* patients, respectively. A total of 159 (35.25%) and 292 (64.75%) patients were grouped as *non-severe* and *severe*, respectively. The same as previously reported (Cummings et al., 2020; Yang et al., 2020), patients with older age (*severe*: an average of 64 years old vs. *non-severe*: an average of 58 years old, t-test p = 4.6e-05) and men (*severe* 52.74% vs. *non-severe* 42.14%, Fisher's exact test p = 0.04) were at a higher risk of developing severe symptoms.

Data Quality

After quality control, the dataset consisted of 22,107,585 and 680,522 variants from autosomes and X chromosome,
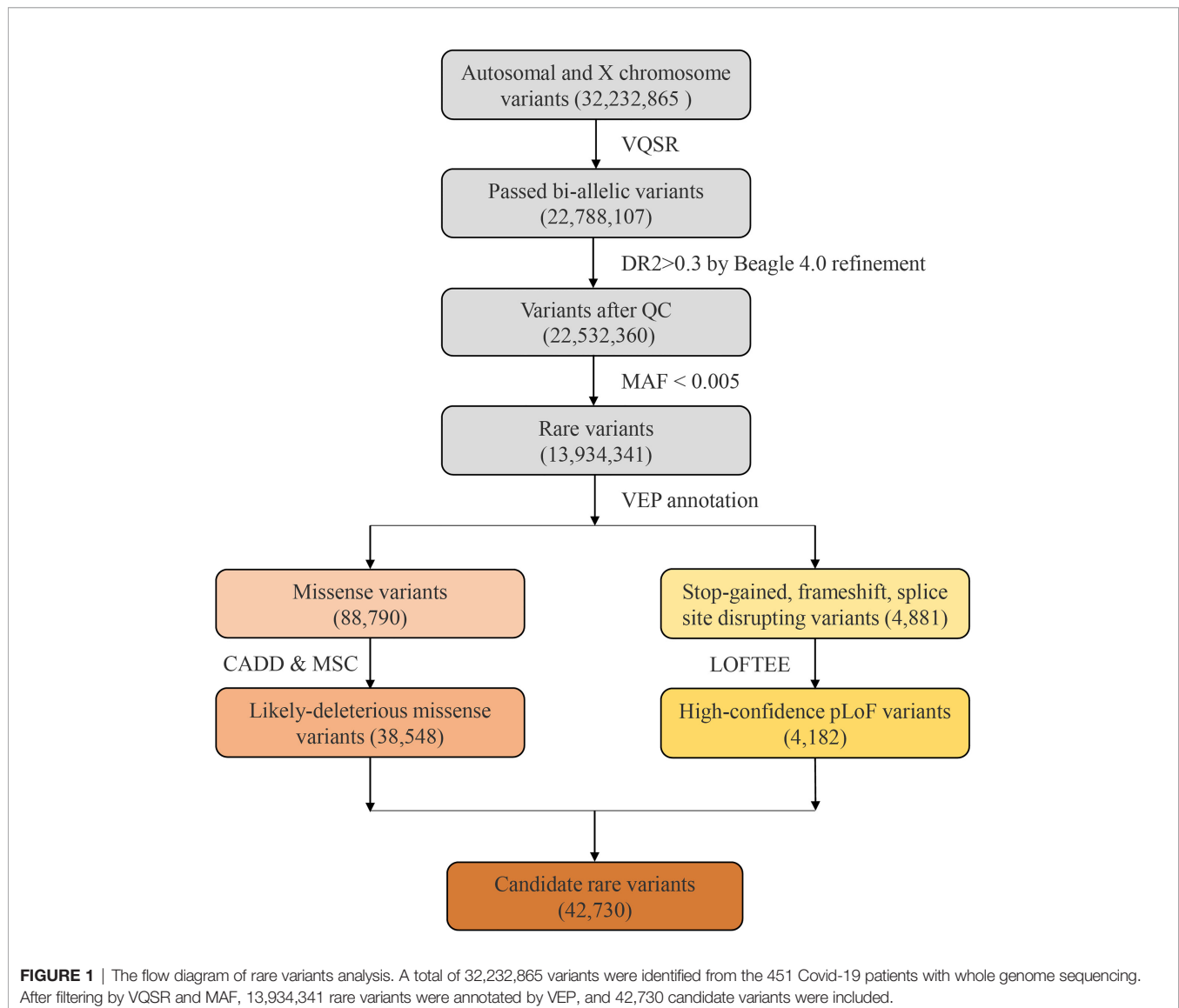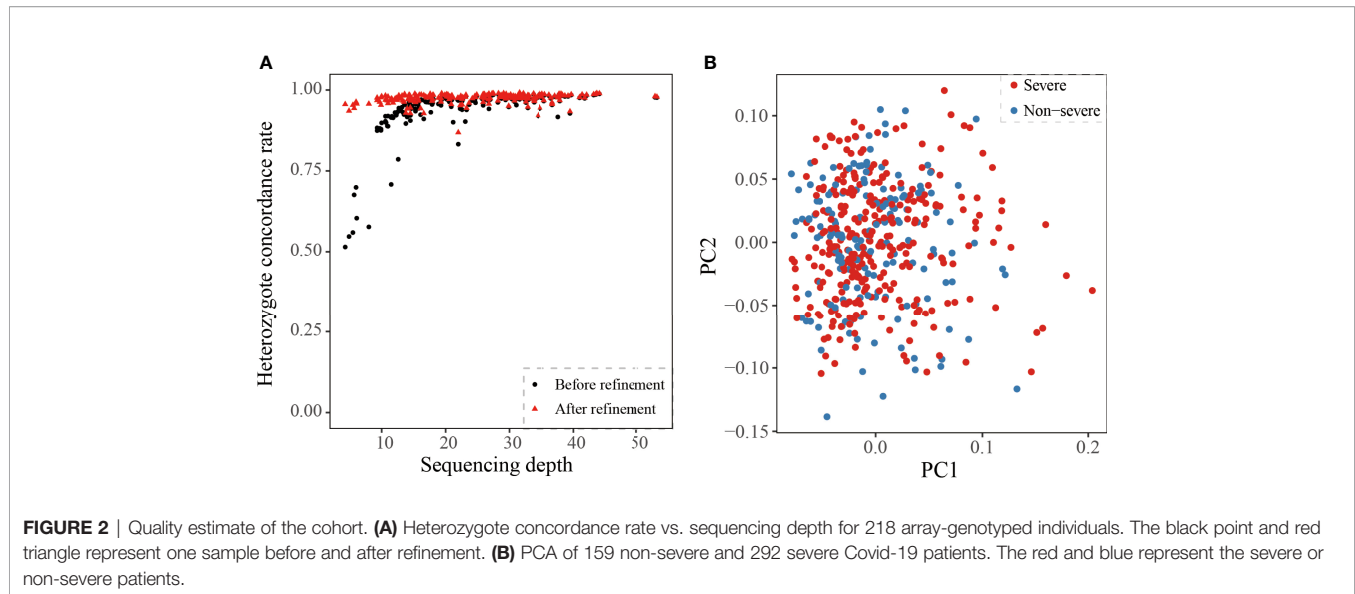
TABLE 1 | Participant characteristics.

	Number, n (%)	Men, n (%)	Age, average (sd)	Depth, average (sd)
All patients	451			21.67
Non-severe	159 (35.25%)	67 (42.14%)	58.33 (14.62)	19.04 (8.94)
Severe	292 (64.75%)	154 (52.74%)	64.11 (13.31)	23.1 (10.43)

respectively (**Figure 1**). Then we compared chip array sequencing results with genotype after LD-based refinement by Beagle 4.0 on 218 individuals. The heterozygote concordance rate increased from an average of 94.4% to 97.4%, and the improvement is more dramatic for samples with lower sequencing depth (**Figure 2A**). After filtering in variants by imputation score DR2 > 0.3, the final dataset for further analyses had a total of 22,532,360 variants, and the PCA on 575,888 autosome SNPs detected no outlier samples (**Figure 2B**).

Rare Variants Statistics

After filtering by MAF, we obtained a total of 13,934,341 rare variants for VEP annotation. Among the resulting annotations, there were 88,790 missense variants and 4881 pLoF variation (including stop-gained, frameshift, and splice site disrupting variants). Damaging effects of these missense and pLoF variants were then predicted by CADD and LOFTEE plug-in, respectively. About 43.41% missense variants were predicted as LDM variants (38,548) and 85.68% pLoF variants were predicted as HC-pLoF variants (4182). Thus, in total, 42,730 predicted

FIGURE 1 | The flow diagram of rare variants analysis. A total of 32,232,865 variants were identified from the 451 Covid-19 patients with whole genome sequencing. After filtering by VQSR and MAF, 13,934,341 rare variants were annotated by VEP, and 42,730 candidate variants were included.

FIGURE 2 | Quality estimate of the cohort. **(A)** Heterozygote concordance rate vs. sequencing depth for 218 array-genotyped individuals. The black point and red triangle represent one sample before and after refinement. **(B)** PCA of 159 non-severe and 292 severe Covid-19 patients. The red and blue represent the severe or non-severe patients.

likely damaging variants were applied for further analysis. For both LDM variants and HC-pLoF variants, we first tested the difference of their numbers between *non-severe* and *severe* patients and found no significant difference (**Supplementary Table S2**).

Gene- and Pathway-Level Analysis of Rare Variants

The gene-level analysis of rare variants was performed between *severe* and *non-severe* patients *via* KGGSeq. We performed the gene-based tests for genes mapped by all 42,730 rare variants, 38,548 LDM variants, and 4182 HC-pLoF variants, respectively. The gene-based analyses did not identify genes that passed the significance threshold of 1e-6 (**Supplementary Figures S1A–C**).

Furthermore, we leveraged the biological knowledge that sets of genes acting together in pathways. In total, we tested 307 KEGG pathways and detected Tuberculosis (hsa:05152, P-burden = 0.036) between *severe* and *non-severe* patients on LDM and HC-pLoF variants (**Supplementary Table S3**). Tuberculosis (TB) is an airborne infectious disease caused by Mycobacterium tuberculosis (Mtb). It first attacks the lungs, then other parts of the body through the circulatory system. This transmission characteristic is remarkably similar to that of Covid-19. As previously reported, Tuberculosis pathway was significant with acute respiratory distress syndrome and lung injury in mice and humans (Sweeney et al., 2017). The TB/Covid-19 Global Study Group observed a phenomenon that TB and SARS-CoV-2 might be co-infected, that is, TB was often diagnosed concurrently or after Covid-19 infection and the co-infection might account for increased case fatality rate (The TB/COVID-19 Global Study Group, 2021). Our finding brought up a possible explanation that patients with rare mutations enriched in Tuberculosis pathway were more likely to develop severe Covid-19 symptoms. More details about this candidate pathway, including the contributed genes, the corresponding gene-based results, and the number of carriers in cases and controls, were

provided in **Supplementary Table S4**. When focusing on only pLOF variants enriched on KEGG, two significant pathways highlighted: Primary immunodeficiency (hsa:05340, P-burden = 0.014) and Influenza A (hsa:05164, P-burden = 0.021) (**Supplementary Table S2**). Primary immunodeficiencies (PID) are a group of potentially serious disorders that can cause increased susceptibility to severe infections, autoimmune diseases, and malignancy. Several studies revealed that patients with PID displayed higher morbidity and mortality from Covid-19 (Hunt et al., 2013; Zuk et al., 2014; Trinder et al., 2020; Kousathanas et al., 2022). Influenza is an infectious respiratory disease caused by influenza virus. Bibert et al. observed that gene pathways involved in the detection of Influenza A overlapped with those involved in the detection of SARS-CoV-2 virus (Bibert et al., 2021). In these two biological pathways, three functional genes, *IKBKG*, *IRF7*, and *IFNAR1*, were previously identified to have an effect on Covid-19 severity (Zhang Q. et al., 2020). More details about these two potentially functional pathways are provided in **Supplementary Tables S5 and S6**.

Tested on 159 Candidate Genes

In addition to uncovering unknown possibly associated genes or pathways, we also tested 159 previously reported candidate genes, with 24 in the IFN pathway (Zhang Q. et al., 2020; Kosmicki et al., 2021) and 136 located within 50 kb of significant common variants in the Covid-19 HGI (COVID-19 Host Genetics Initiative, 2021). Specifically, we focused on LDM and HC-pLOF variants to aggregate potential effects of rare variants. For missense variants in both IFN- and HGI-genes, we did not detect significant difference between *severe* and *non-severe* patients.

In the 24 IFN-genes, we found one HC-pLoF variant rs780744847 (c.1180C>T/p.R394*) on *TLR3* mutated in only *severe* patients but no mutations in *non-severe* patients. It was reported that the *TLR3* deficiency may lead to increased incidences of viral infections and impair the production of type

I IFN throughout SARS-CoV-2 infection (Zhang Q. et al., 2020). Moreover, mutations of inborn errors of TLR3-dependent type I IFN immunity more often occurred in highly critical patients than in mild patients and healthy controls.

For the 136 HGI-genes, we found that the number of HC-pLoF mutations occurred in the *severe* group was more than that of the *non-severe* groups (16 in *severe* and 2 in *non-severe* patients, Fisher's exact test p = 0.043) (**Table 2**). We also detected a HC-pLoF variant rs541048548 (c.1053del/p.A352fs) on *ICAM3* only mutated in *severe* patients. The gene *ICAM3* played an important role in the immunopathogenesis of SARS virus (Chan et al., 2007) and had been reported that its expression was downregulated in asymptomatic Covid-19 cases compared with symptomatic patients (Masood et al., 2021).

Mutation Accumulation Analyses

In the mutation accumulation analysis, we first investigated whether there were potentially functional mutations unique to *severe* patients. We filtered in rare variants mutated in only *severe* patients and with minor allele count (MAC) greater than or equal to three. This resulted in 756 rare variants mapped to 700 genes. Among these variants, we observed a very rare mutation rs777044791 in gene *CCR3* at locus 3p21.31 (**Table 3**). The physical distance between rs777044791 and rs11385942 is 0.43 MB (GRCh38), a distance typically flanked into the same genomic region (Casto and Feldman, 2011). The variant rs11385942 is a common variant located at locus 3p21.31 in European populations and was first identified to be associated with respiratory failure due to Covid-19 from GWAS analysis in Italian and Spanish populations (Ellinghaus et al., 2020). This finding was repeated in other studies based on European populations (COVID-19 Host Genetics Initiative, 2021; Pairo-Castineira et al., 2021; Shelton et al., 2021), verifying its effects on Covid-19 disease. In the Chinese population, common variant studies at this locus did not replicate significance (Wang et al., 2020; Wu et al., 2021b;

TABLE 2 | The pLoF variants identified in Covid-19 patients in 159 candidate genes.

Gene	SNP	Variant annotation	HGVSc/HGVSp	Genotype	Sample ID	Sex	Age range	Phenotype	Category
TLR3	rs780744847	Stop gained	c.1180C>T/p.Arg394Ter	Het	U312	F	70-79	Severe	IFN-genes
THBS3	rs748584696	Stop gained	c.853C>T/p.Arg285Ter	Het	U088	F	80-89	Severe	HGI-genes
THBS3	chr1_155206198_A_C	Splice donor variant	c.286+2T>G	Het	U359	F	70-79	Severe	HGI-genes
TAC4	rs372635644	Splice acceptor variant	c.124-1G>A	Het	U429	F	60-69	Severe	HGI-genes
TYK2	rs770927552	Frameshift variant	c.209_212del/p.Cys70SerfsTer21	Het	U422	F	80-89	Severe	HGI-genes
C6orf15	chr6_31112292_C_T	Splice acceptor variant	c.68-1G>A	Het	U107	M	60-69	Severe	HGI-genes
CAT	rs777641795	Splice donor variant	c.1195+1G>A	Het	U012	M	40-49	Non-severe	HGI-genes
CDH15	chr16_89179469_C_G	Stop gained	c.96C>G/p.Tyr32Ter	Het	U174	F	50-59	Severe	HGI-genes
CDSN	chr6_31116133_G_GA	Frameshift variant	c.1481dup/p.Cys496LeufsTer20	Het	U021	F	50-59	Non-severe	HGI-genes
ICAM3	rs541048548	Frameshift variant	c.1053delC/p.Ala352ArgfsTer11	Het	U225	M	70-79	Severe	HGI-genes
ICAM3	rs541048548	Frameshift variant	c.1053delC/p.Ala352ArgfsTer11	Het	U047	F	70-79	Severe	HGI-genes
MYDGF	rs745851558	Splice donor variant	c.174+1G>T	Het	U071	M	70-79	Severe	HGI-genes
PLEKHA4	chr19_48853718_CA_C	Frameshift variant	c.1289delT/p.Leu430ArgfsTer4	Het	U261	F	60-69	Severe	HGI-genes
PLEKHA4	chr19_48853720_GCCGGT_G	Frameshift variant	c.1283_1287delACCGG/ p.Asp428AlafsTer76	Het	U261	F	60-69	Severe	HGI-genes
PPP1R15A	rs768756506	Frameshift variant	c.1535_1536delAT/ p.Tyr512CysfsTer14	Het	U309	M	60-69	Severe	HGI-genes
PSORS1C2	rs79153019	Frameshift variant	c.281delC/p.Pro94LeufsTer35	Het	U075	M	60-69	Severe	HGI-genes
PSORS1C2	rs79153019	Frameshift variant	c.281delC/p.Pro94LeufsTer35	Het	U150	M	60-69	Severe	HGI-genes
PSORS1C2	rs79153019	Frameshift variant	c.281delC/p.Pro94LeufsTer35	Het	U144	M	60-69	Severe	HGI-genes
TULP2	chr19_48881045_T_TC	Frameshift variant	c.1528_1529insG/ p.Gln510ArgfsTer17	Het	U176	F	70-79	Severe	HGI-genes

	rs777044791	rs11385942
CHROM	chr3	chr3
POS (hg38)	46,266,186	45,834,967
ALT	T	GA
REF	C	G
Variant annotation	Missense variant	Intron variant
Allele frequency		
Severe (N = 292)	0.005	0
Non-severe (N = 159)	0	0
ChinaMAP	0.002	0.004
1000G_EAS	0	0.005
1000G_EUR	0	0.0805
1000G_SAS	0	0.296
1000G_AFR	0	0.053
gnomAD_EAS	0.0005	0.0006

Zhu et al., 2021), and no rare variant studies had been conducted. Our work raised a possibility that SNPs in locus 3p21.31 might also play an important role in Covid-19 severity in the Chinese population.

Then, we performed PPI network analysis for the 700 "individual variant-driven" genes with the 159 known genes (**Figure 3A**). From the results, we found two candidate genes *IL12RB1* and *TRAF3IP3* that had extensive interactions with IFN- and HGI-genes. Gene *IL12RB1* (Interleukin 12 Receptor Subunit Beta 1) encodes a type I transmembrane protein that binds to interleukin-12 (IL12) and is involved in IL12 transduction. Mutations in *IL12RB1* damage the development of IL17-producing T lymphocytes and increase the susceptibility to Salmonella and mycobacterial infections (van de Vosse et al., 2013). Our PPI network analysis indicated that *IL12RB1* and *TYK2* had experimentally determined interactions, which were compiled from a set of public databases and were more likely to be credible (Zhang et al., 2013). Gene *TYK2* had been previously identified to be associated with Covid-19 critical illness (Pairo-Castineira et al., 2021), implying the potential effects of *IL12RB1* to the aggravation of Covid-19. The gene *TRAF3IP3* (TRAF3 Interacting Protein 3) encodes a protein that plays essential roles in both innate and adaptive immunity. Knockout mouse experiments of this gene observed a decrease in white blood cell count in males and an increased susceptibility to bacterial infection (Gardin and White, 2011). In our results, TRAF3IP3 was experimentally determined with protein TRAF3 encoded by gene *TRAF3*, which was included in a newly created pathway "Activation of NLRP3 inflammasome by SARS-CoV-2" (WP4876) (Siu et al., 2019). In response to viral infection, *TRAF3IP3* bridges *TRAF3* and *MAVS* leading to interferon production, indicating it's probably strong relationship with Covid-19 disease.

We also performed PPI network analysis for the 778 "all variant-driven" genes with the 159 known genes (**Figure 3B**), from which three genes, *TBK1*, *TLR3*, and *IFNGR2* were highlighted. Specifically, *TBK1* (TANK Binding Kinase 1) encodes a protein that plays important roles in antiviral innate immune response and in regulating inflammatory response to foreign agents (Fitzgerald et al., 2003; Mori et al., 2004).

A previous study observed colocalization of *TBK1* with the M protein of SARS-CoV-2, which might hinder the dsRNA-induced IFN production at the step or upstream of *TBK1* (Zheng et al., 2020). The gene *TLR3* (Toll Like Receptor 3) encodes a member of the TLR family that plays a primary role in recognition of pathogen and innate immunity activation. It recognizes dsRNA participated in multiple viral infections and induces type I IFNs production (Kawai and Akira, 2007). The gene *IFNGR2* (Interferon Gamma Receptor 2) encodes the non-ligand-binding beta chain of the gamma interferon receptor. A recent study revealed a new set of genes that upregulated in severe Covid-19 patients compared to mild ones, probably triggered by *IFNGR1* and *IFNGR2* (Henriques-Pons et al., 2021).

In summary, our mutation accumulation analyses and PPI network analyses suggested that *IL12RB1*, *TRAF3IP3*, *TBK1*, and *TLR3* and *IFNGR2* are key regions in *severe* Covid-19 patients compared with *non-severe*, implying their functions and associations with Covid-19 severity.

CONCLUSION

In our study, we have uncovered several functional pathways associated with Covid-19 severity, for example, Tuberculosis (hsa:05152), Primary Immunodeficiency (hsa:05340), and Influenza A (hsa:05164). These pathways are all responsible for innate immune disorders and respiratory diseases, highlighting the importance of host innate immune system against Covid-19. Our mutation accumulation analysis and PPI network analysis suggested several novel candidate genes in the Chinese population, including *IL12RB1*, *TBK1*, *TLR3*, and *IFNGR2*. These genes are potentially involved in SARS-CoV-2 cell entry, host immune responses, and finally influencing Covid-19 severity.

On one hand, our work filled the gap of IEI analysis in Covid-19 patients in the Chinese population; on the other hand, we replicated several Covid-19-associated genes first identified from a European population and also discovered some candidate genes specific to the Chinese population.
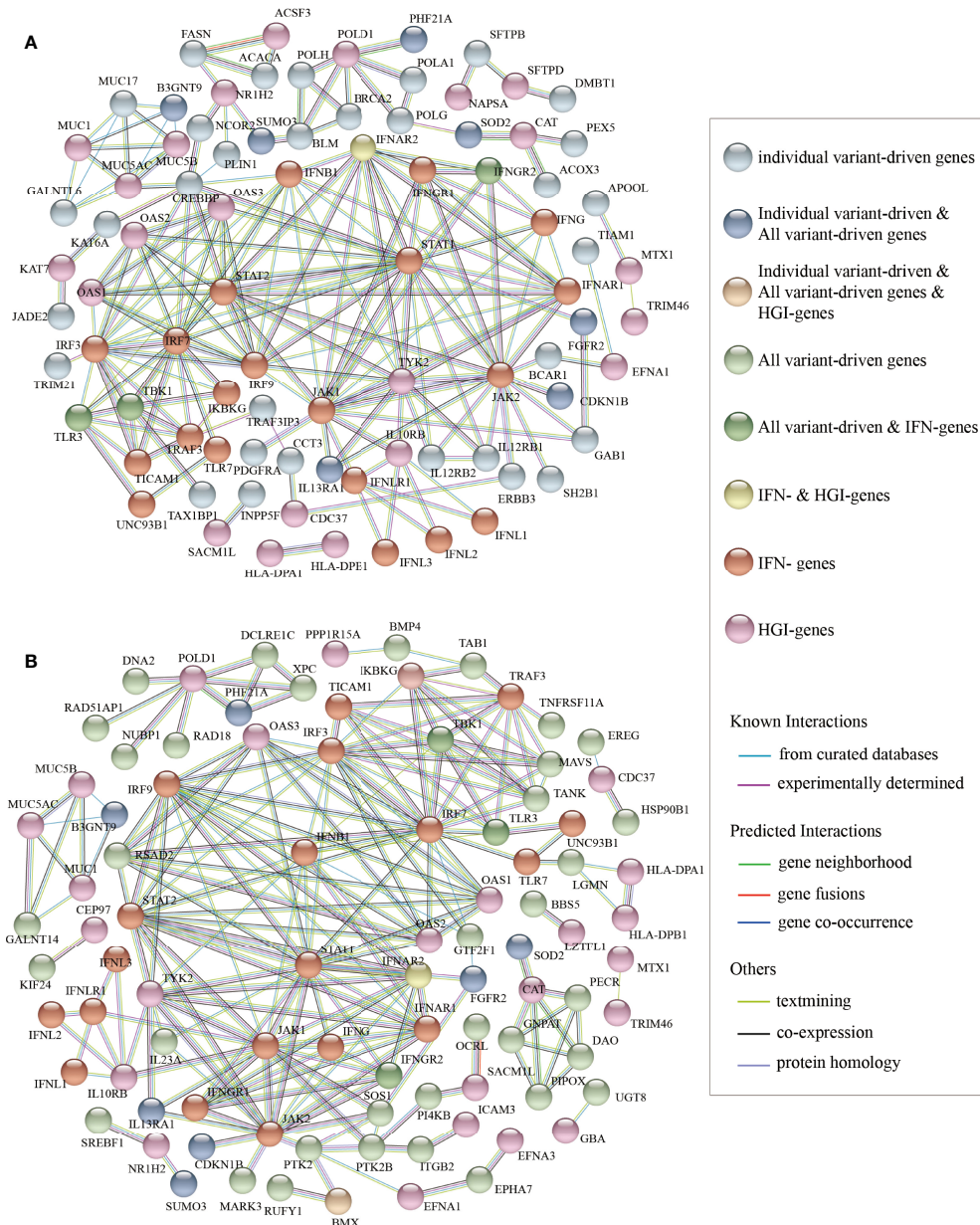
FIGURE 3 | The results of protein-protein interaction network analysis. The plot of the PPI network **(A)** between the "individual variant-driven" genes with the candidate genes, and **(B)** between the "all variant-driven" genes with the candidate genes.

DISCUSSION

SARS-CoV-2 is a strain of coronavirus and is highly pathogenic and transmissible. After exposure to the virus, ordinary people may not develop noticeable symptoms or develop mild to moderate symptoms, while people with IEI tend to suffer severe and critical symptoms, or even death (Zhang S.-Y. et al., 2020). There is increasing evidence that the host genetic variants in genes related to immunodeficiency or inflammasomes might attribute to Covid-19 clinical manifestations (Elhabyan et al., 2020). Many clinical drug treatments for Covid-19 were

cultivated from this finding, including type I IFNs (e.g., IFN-α1b), TNF inhibitors, anti-IFN-γ antibodies, JAK1 inhibitors, and STAT1 inhibitors (Ku et al., 2021).

In this work, we carried out the first study of rare variants in IEI genes associated with Covid-19 severity in the Chinese population. The identified functional candidate pathways Tuberculosis, Primary Immunodeficiency, and Influenza A were previously known to be part of antiviral immune responses and viral eradication, and we discovered their potential influences in Covid-19. We also suggested several putative genetic regions probably involved in susceptibility and severity of Covid-19, including genes *IL12RB1*, *TRAF3IP3*,

TBK1, *TLR3*, and *IFNGR2*. Our work highlighted the importance of rare IEI in Covid-19 patients and people with IEI defects are more likely to be infected with SARS-CoV-2 and to develop severe symptoms. Thus, we appeal more studies on IEI in both the Chinese population and other populations to pinpoint causal genes of Covid-19 severity and finally help identify those patients at higher risk.

Despite the many compelling and significant findings of our work, there are still a few limitations to be noted. First, the sample size we used is relatively small, and the limited sample size limits the statistical power for identifying rare variants. More studies with large sample sizes are demanded to validate our results and uncover more candidate variants. Second, our work has suggested several candidate genes and pathways potentially related to Covid-19 severity, yet unfortunately, due to resources limitations, we are unable to perform web-lab experiments and verify gene functions at this stage. More persuasive experimental designs are needed to investigate how these candidate genes/pathways affect disease progression.

Covid-19 is assessed as a complex infectious disease and affected many risk factors. Symptoms of Covid-19 are highly variable, ranging from unnoticeable to severe and even death. The host genetic background is only partly responsible for the phenotypic heterogeneity. In recent years, multi-omics studies have proven a powerful and successful strategy to provide a broader perspective in understanding disease development and biological phenomena. Several multi-omics analyses of Covid-19 have been proposed to integrate multiple "omes" data to unravel disease mechanisms at multiple omics levels (Su et al., 2020; Montaldo et al., 2021; Overmyer et al., 2021; Stephenson et al., 2021; Wu et al., 2021a). The integrative analyses of rare genome and other "omes" data (e.g., proteome, transcriptome, epigenome, metabolome, and microbiome) may inspire us to discover new risk factors for severe Covid-19 disease.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study have been deposited into CNGB Sequence Archive (CNSA) (Guo et al., 2020) of China National GeneBank DataBase (CNGBdb) (Guo et al., 2020) with accession number CNP0002853.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Medical Ethics Committee of Union Hospital, Tongji Medical College, Huazhong University of Science and Technology. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XJ and HZ conceived the study, designed the research program, and managed the project. PL and HZ performed the statistical analyses. MF advised on statistical methods. PL and HZ wrote the manuscript. FC, FZ, and YJ collected the samples. YL finished the laboratory processing and data acquisition. All authors participated in revising the manuscript. All authors read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2022.888582/full#supplementary-material

Supplementary Figure 1 | Results of gene-based association tests of rare variants. **(A)**. The QQ plots of gene-based association analyses between severe and non-severe patients for **(A)** 42,730 candidate rare variants; **(B)** 38,548 rare likely-deleterious missense variants, and **(C)** 4182 high-confidence pLoF variants. The color represents different association methods integrated in KGGseq, including SKAT test (in blue), SKAT-O (in red), and burden test (in orange).

REFERENCES

Asano, T., Boisson, B., Onodi, F., Matuozzo, D., Moncada-Velez, M., Luxman Maglorius Renkilaraj, M. R, et al. (2021). "X-Linked Recessive TLR7 Deficiency in ~1% of Men Under 60 Years Old With Life-Threatening COVID-19". *Sci. Immunol.* 6 (62), eabl4348. doi: 10.1126/sciimmunol.abl4348

Bibert, S., Guex, N., Lourenco, J., Brahier, T., Papadimitriou-Olivgeris, M., Damonti, L., et al. (2021). Transcriptomic Signature Differences Between SARS-CoV-2 and Influenza Virus Infected Patients. *Front. Immunol.* 12, 1990. doi: 10.3389/fimmu.2021.666163

Browning, B. L., and Browning, S. R. (2016). Genotype Imputation With Millions of Reference Samples. *Am. J. Hum. Genet.* 98 (1), 116–126. doi: 10.1016/j.ajhg.2015.11.020

Casto, A. M., and Feldman, M. W. (2011). Genome-Wide Association Study SNPs in the Human Genome Diversity Project Populations: Does Selection Affect Unlinked SNPs With Shared Trait Associations? *PLoS Genet.* 7 (1), e1001266. doi: 10.1371/journal.pgen.1001266

Chan, K. Y. K., Ching, J. C. Y., Xu, M. S., Cheung, A. N. Y., Yip, S.-P., Yam, L. Y. C., et al. (2007). Association of ICAM3 Genetic Variant With Severe Acute Respiratory Syndrome. *J. Infect. Dis.* 196 (2), 271–280. doi: 10.1086/518892

Chang, ,. C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., . Purcell, S. M, and Lee, J. J (2015). Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *GigaScience* 4, 7. doi: 10.1186/s13742-015-0047-8

Clarke, L., Zheng-Bradley, X., Smith, R., Kulesha, E., Xiao, C., Toneva, I., et al. (2012). The 1000 Genomes Project: Data Management and Community Access. *Nat. Methods* 9 (5), 459–462. doi: 10.1038/nmeth.1974

COVID-19 Host Genetics Initiative. (2021). Mapping the Human Genetic Architecture of COVID-19. *Nature* 600 (7889), 472–477. doi: 10.1101/2021.11.08.21265944

Cummings, M. J., Baldwin, M. R., Abrams, D., Jacobson, S. D., Meyer, B. J., Balough, E. M., et al. (2020). Epidemiology, Clinical Course, and Outcomes of Critically Ill Adults With COVID-19 In New York City: A Prospective Cohort Study. *Lancet (Lond. Engl.)* 395 (10239), 1763–1770. doi: 10.1016/S0140-6736(20)31189-2

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve Years of SAMtools and BCFtools. *GigaScience* 10 (2), giab008. doi: 10.1093/gigascience/giab008

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A Framework for Variation Discovery and Genotyping Using Next-Generation DNA Sequencing Data. *Nat. Genet.* 43 (5), 491–498. doi: 10.1038/ng.806

Elhabyan, A., Elyaacoub, S., Sanad, E., Abukhadra, A., Elhabyan, A., and Dinu, V. (2020). The Role of Host Genetics in Susceptibility to Severe Viral Infections in Humans and Insights Into Host Genetics of Severe COVID-19: A Systematic Review. *Virus Res.* 289, 198163. doi: 10.1016/j.virusres.2020.198163

Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., et al. (2020). Genomewide Association Study of Severe Covid-19 With Respiratory Failure. *N. Engl. J. Med* 383 (16), 1522–1534. doi: 10.1056/NEJMoa2020283

Fallerini, C., Daga, S., Mantovani, S., Benetti, E., Picchiotti, N., Francisci, D., et al. (2021). Association of Toll-Like Receptor 7 Variants With Life-Threatening COVID-19 Disease in Males: Findings From a Nested Case-Control Study. *ELife* 10, 1–2. doi: 10.7554/eLife.67569

Fitzgerald, K. A., McWhirter, S. M., Faia, K. L., Rowe, D. C., Latz, E., Golenbock, D. T., et al. (2003). IKKepsilon and TBK1 Are Essential Components of the IRF3 Signaling Pathway. *Nat. Immunol.* 4 (5), 491–496. doi: 10.1038/ni921

Freed, D.N., Aldana, R., Weber, J.A., and Edwards, J. (2017). The Sentieon Genomics Tools - A Fast and Accurate Solution to Variant Calling From Next-Generation Sequence Data. *bioRxiv*. Available at: http://biorxiv.org/content/early/2017/05/12/115717.abstract.

Gardin, A., and White, J. (2011). "The Sanger Mouse Genetics Programme: High Throughput Characterisation of Knockout Mice". *Acta Ophthalmolog.* 89 (s248). doi: 10.1111/j.1755-3768.2011.4451.x

Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., et al. (2020). The Species Severe Acute Respiratory Syndrome-Related Coronavirus: Classifying 2019-NCoV and Naming It SARS-CoV-2. *Nat. Microbiol.* 5 (4), 536–544. doi: 10.1038/s41564-020-0695-z

Gorlov, I. P., Gorlova, O. Y., Frazier, M. L., Spitz, M. R., and Amos, C. I. (2011). Evolutionary Evidence of the Effect of Rare Variants on Disease Etiology. *Clin. Genet.* 79 (3), 199–206. doi: 10.1111/j.1399-0004.2010.01535.x

Guo, X., Chen, F., Gao, F., Li, L., Liu, K., You, L., et al. (2020). CNSA: A Data Repository for Archiving Omics Data. *Database*, 2020 1–6. doi: 10.1093/database/baaa055

Henriques-Pons, A., Beghini, D. G., Santos Silva, V. D., Horita, S. I., and da Silva, F. A. B. (2021). Pulmonary Mesenchymal Stem Cells in Mild Cases of COVID-19 Are Dedicated to Proliferation; In Severe Cases, They Control Inflammation, Make Cell Dispersion, and Tissue Regeneration. *Front. Immunol.* 12. doi: 10.3389/fimmu.2021.780900

Hunt, K. A., Mistry, V., Bockett, N. A., Ahmad, T., Ban, M., Barker, J. N., et al. (2013). Negligible Impact of Rare Autoimmune-Locus Coding-Region Variants on Missing Heritability. *Nature* 498 (7453), 232–235. doi: 10.1038/nature12170

Itan, Y., Shang, L., Boisson, B., Ciancanelli, M. J., Markle, J. G., Martinez-Barricarte, R., et al. (2016). The Mutation Significance Cutoff: Gene-Level Thresholds for Variant Predictions. *Nat. Methods* 13 (2), 109–110. doi: 10.1038/nmeth.3739

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi: 10.1093/nar/28.1.27

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., et al. (2020). The Mutational Constraint Spectrum Quantified From Variation in 141,456 Humans. *Nature* 581 (7809), 434–443. doi: 10.1038/s41586-020-2308-7

Kawai, T., and Akira, S. (2007). Signaling to NF-KappaB by Toll-Like Receptors. *Trends Mol. Med.* 13 (11), 460–469. doi: 10.1016/j.molmed.2007.09.002

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A General Framework for Estimating the Relative Pathogenicity of Human Genetic Variants. *Nat. Genet.* 46 (3), 310–315. doi: 10.1038/ng.2892

Kosmicki, J. A., Horowitz, J. E., Banerjee, N., Lanche, R., Marcketta, A., Maxwell, E., et al. (2021). Pan-Ancestry Exome-Wide Association Analyses of COVID-19 Outcomes in 586,157 Individuals. *Am. J. Hum. Genet.* 108 (7), 1350–1355. doi: 10.1016/j.ajhg.2021.05.017

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

Ku, C.-L., Chen, I. T., and Lai, M.-Z. (2021). Infection-Induced Inflammation From Specific Inborn Errors of Immunity to COVID-19. *FEBS J.* 288 (17), 5021–5041. doi: 10.1111/febs.15961

Lee, S., Wu, M. C., and Lin, X. (2012). Optimal Tests for Rare Variant Effects in Sequencing Association Studies. *Biostat (Oxf. Engl.)* 13 (4), 762–775. doi: 10.1093/biostatistics/kxs014

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment With Burrows-Wheeler Transform. *Bioinf. (Oxf. Engl.)* 25 (14), 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, M., Li, J., Li, M. J., Pan, Z., Hsu, J. S., Liu, D. J., et al. (2017). Robust and Rapid Algorithms Facilitate Large-Scale Whole Genome Sequencing Downstream Analysis in an Integrative Framework. *Nucleic Acids Res.* 45 (9), e75. doi: 10.1093/nar/gkx019

Mantovani, S., Daga, S., Fallerini, C., Baldassarri, M., Benetti, E., Picchiotti, N., et al. (2022). Rare Variants in Toll-Like Receptor 7 Results in Functional Impairment and Downregulation of Cytokine-Mediated Signaling in COVID-19 Patients. *Genes Immun.* 23, 51–56. doi: 10.1038/s41435-021-00157-1

Masood, K. I., Yameen, M., Ashraf, J., Shahid, S., Mahmood, S. F., Nasir, A., et al. (2021). Upregulated Type I Interferon Responses in Asymptomatic COVID-19 Infection Are Associated With Improved Clinical Outcome. *Sci. Rep.* 11 (1), 22958. doi: 10.1038/s41598-021-02489-4

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.* 17 (1), 122. doi: 10.1186/s13059-016-0974-4

Misawa, H., Hasegawa, T., Mishima, E., Jutabha, P., Ouchi, M., and Kojima, K. (2020). Contribution of Rare Variants of the SLC22A12 Gene to the Missing Heritability of Serum Urate Levels. *Genetics* 214 (4), 1079–1090. doi: 10.1534/genetics.119.303006

Montaldo, C., Messina, F., Abbate, I., Antonioli, M., Bordoni, V., Aiello, A., et al. (2021). Multi-Omics Approach to COVID-19: A Domain-Based Literature Review. *J. Trans. Med.* 19 (1), 501. doi: 10.1186/s12967-021-03168-8

Mori, M., Yoneyama, M., Ito, T., Takahashi, K., Inagaki, F., and Fujita, T. (2004). Identification of Ser-386 of Interferon Regulatory Factor 3 as Critical Target for Inducible Phosphorylation That Determines Activation. *J. Biol. Chem.* 279 (11), 9698–9702. doi: 10.1074/jbc.M310616200

Overmyer, K. A., Shishkova, E., Miller, I. J., Balnis, J., Bernstein, M. N., Peters-Clarke, T. M., et al. (2021). Large-Scale Multi-Omic Analysis of COVID-19 Severity. *Cell Syst.* 12 (1), 23–40.e7. doi: 10.1016/j.cels.2020.10.003

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

Pearson, K. (1901). "LIII. On Lines and Planes of Closest Fit to Systems of Points in Space". *London Edinburgh Dublin Philos. Magazine J. Sci.* 2 (11), 559–572. doi: 10.1080/14786440109462720

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies. *Nat. Genet.* 38 (8), 904–909. doi: 10.1038/ng1847

Shelton, J. F., Shastri, A. J., Ye, C., Weldon, C. H., Filshtein-Sonmez, T., Coker, D., et al. (2021). Trans-Ancestry Analysis Reveals Genetic and Nongenetic

Associations With COVID-19 Susceptibility and Severity. *Nat. Genet.* 53 (6), 801–808. doi: 10.1038/s41588-021-00854-7

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M, et al. (2001). "DbSNP: The NCBI Database of Genetic Variation". *Nucleic Acids Res.* 29 (1), 308–311. doi: 10.1093/nar/29.1.308

Siu, K.-L., Yuen, K.-S., Castaño-Rodriguez, C., Ye, Z.-W., Yeung, M.-L., Fung, S.-Y., et al. (2019). Severe Acute Respiratory Syndrome Coronavirus ORF3a Protein Activates the NLRP3 Inflammasome by Promoting TRAF3-Dependent Ubiquitination of ASC. *FASEB J.* 33 (8), 8865–8877. doi: 10.1096/fj.201802418R

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

Stephenson, E., Reynolds, G., Botting, R. A., Calero-Nieto, F. J., Morgan, M. D., Tuong, Z. K., et al. (2021). Single-Cell Multi-Omics Analysis of the Immune Response in COVID-19. *Nat. Med.* 27 (5), 904–916. doi: 10.1038/s41591-021-01329-2

Su, Y., Chen, D., Yuan, D., Lausted, C., Choi, J., Dai, C. L., et al. (2020). Multi-Omics Resolves a Sharp Disease-State Shift Between Mild and Moderate COVID-19. *Cell* 183 (6), 1479–1495.e20. doi: 10.1016/j.cell.2020.10.037

Sweeney, T. E., Lofgren, S., Khatri, P., and Rogers, A. J. (2017). Gene Expression Analysis to Assess the Relevance of Rodent Models to Human Lung Injury. *Am. J. Respir. Cell Mol. Biol.* 57 (2), 184–192. doi: 10.1165/rcmb.2016-0395OC

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING V11: Protein–Protein Association Networks With Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res.* 47 (D1), D607–D613. doi: 10.1093/nar/gky1131

The TB/COVID-19 Global Study Group. (2021). Tuberculosis and COVID-19 Co-Infection: Description of the Global Cohort. *Eur. Respir. J.* 59, 2102538. doi: 10.1183/13993003.02538-2021

Trinder, M., Walley, K. R., Boyd, J. H., and Brunham, L. R. (2020). Causal Inference for Genetically Determined Levels of High-Density Lipoprotein Cholesterol and Risk of Infectious Disease. *Arteriosclerosis Thrombosis Vasc. Biol.* 40 (1), 267–278. doi: 10.1161/ATVBAHA.119.313381

van der Made, C. I., Simons, A., Schuurs-Hoeijmakers, J., van den Heuvel, G., Mantere, T., Kersten, S., et al. (2020). Presence of Genetic Variants Among Young Men With Severe COVID-19. *JAMA* 324 (7), 663–673. doi: 10.1001/jama.2020.13719

van de Vosse, E., Haverkamp, M. H., Ramirez-Alejo, N., Martinez-Gallo, M., Blancas-Galicia, L., Metin, A., et al. (2013). IL-12rβ1 Deficiency: Mutation Update and Description of the IL12RB1 Variation Database. *Hum. Mutat.* 34 (10), 1329–1339. doi: 10.1002/humu.22380

Wang, F., Huang, S., Gao, H., Zhou, Y., Lai, C., Li, Z., et al. (2020). Initial Whole Genome Sequencing and Analysis of the Host Genetic Contribution to COVID-19 Severity and Susceptibility. *Cell Discov.* 6 (1), 83. doi: 10.1101/2020.06.09.20126607

Williams, F. M. K., Freydin, M., Mangino, M., Couvreur, S., Visconti, A., Bowyer, R. C. E., et al. (2020). Self-Reported Symptoms of COVID-19, including Symptoms Most Predictive of SARS-CoV-2 Infection, Are Heritable. *Twin Res. Hum. Genet.* 23, 316–321. doi: 10.1017/thg.2020.85

Wu, P., Chen, D., Ding, W., Wu, P., Hou, H., Bai, Y., et al. (2021a). The Trans-Omics Landscape of COVID-19. *Nat. Commun.* 12 (1), 4543. doi: 10.1038/s41467-021-24482-1

Wu, P., Ding, L., Li, X., Liu, S., Cheng, F., He, Q., et al. (2021b). Trans-Ethnic Genome-Wide Association Study of Severe COVID-19. *Commun. Biol.* 4 (1), 1034. doi: 10.1038/s42003-021-02549-5

Wu, M. C., Kraft, P., Epstein, M. P., Taylor, D. M., Chanock, S. J., Hunter, D. J., et al. (2010). Powerful SNP-Set Analysis for Case-Control Genome-Wide Association Studies. *Am. J. Hum. Genet.* 86 (6), 929–942. doi: 10.1016/j.ajhg.2010.05.002

Wu, Z., and McGoogan, J. M. (2020). Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 323 (13), 1239–1242. doi: 10.1001/jama.2020.2648

Yang, X., Yu, Y., Xu, J., Shu, H., Xia, J., Liu, H., et al. (2020). Clinical Course and Outcomes of Critically Ill Patients With SARS-CoV-2 Pneumonia in Wuhan, China: A Single-Centered, Retrospective, Observational Study. *Lancet Respir. Med.* 8 (5), 475–481. doi: 10.1016/S2213-2600(20)30079-5

Zhang, F., Flickinger, M., Gagliano Taliun, S. A., Abecasis, G. R., Scott, L. J., McCaroll, S. A., et al. (2020). Ancestry-Agnostic Estimation of DNA Sample Contamination From Sequence Reads. *Genome Res.* 30 (2), 185–194. doi: 10.1101/gr.246934.118

Zhang, Q., Liu, Z., Moncada-Velez, M., Chen, J., Ogishi, M., Bigio, B., et al. (2020). Inborn Errors of Type I IFN Immunity in Patients With Life-Threatening COVID-19. *Science* 370 (6515), eabd4570. doi: 10.1126/science.abd4570

Zhang, Q. C., Petrey, D., Garzón, J. I., Deng, L., and Honig, B. (2013). PrePPI: A Structure-Informed Database of Protein-Protein Interactions. *Nucleic Acids Res.* 41 (Database issue), D828–D833. doi: 10.1093/nar/gks1231

Zhang, S.-Y., Zhang, Q., Casanova, J.-L., Su, H. C.COVID Team (2020). Severe COVID-19 in the Young and Healthy: Monogenic Inborn Errors of Immunity? *Nat. Rev. Immunol.* 20 (8), 455–456. doi: 10.1038/s41577-020-0373-7

Zheng, Y., Zhuang, M.-W., Han, L., Zhang, J., Nan, M.-L., Zhan, P., et al. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Membrane (M) Protein Inhibits Type I and III Interferon Production by Targeting RIG-I/MDA-5 Signaling. *Signal Transduct. Target. Ther.* 5 (1), 299. doi: 10.1038/s41392-020-00438-7

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A Novel Coronavirus From Patients With Pneumonia in China 2019. *N. Engl. J. Med.* 382 (8), 727–733. doi: 10.1056/NEJMoa2001017

Zhu, H., Zheng, F., Li, L., Jin, Y., Luo, Y., Li, Z., et al. (2021). A Chinese Host Genetic Study Discovered Type I Interferons and Causality of Cholesterol Levels and WBC Counts on COVID-19 Severity. *iScience* 24, 103186. doi: 10.1016/j.isci.2021.103186

Zuk, O., Schaffner, S. F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., et al. (2014). Searching for Missing Heritability: Designing Rare Variant Association Studies. *Proc. Natl. Acad. Sci.* 111 (4), E455 LP–E464. doi: 10.1073/pnas.1322563111