



Genome-wide association study of SARS-CoV-2 infection in Chinese population

Jie Fan^{1,2} · Quan-Xin Long³ · Ji-Hua Ren³ · Hao Chen¹ · Meng-Meng Li¹ · Zheng Cheng¹ · Juan Chen^{3,4} · Li Zhou¹ · Ai-Long Huang³

Received: 10 May 2022 / Accepted: 18 July 2022 / Published online: 4 August 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Coronavirus disease 2019 (COVID-19) is a global public health concern. The purpose of this study was to investigate the association between genetic variants and SARS-CoV-2 infection and the COVID-19 severity in Chinese population. A total of 256 individuals including 87 symptomatic patients (tested positive for SARS-CoV-2), 84 asymptomatic cases, and 85 close contacts of confirmed patients (tested negative for SARS-CoV-2) were recruited from February 2020 to May 2020. We carried out the whole exome genome sequencing between the individuals and conducted a genetic association study for SARS-CoV-2 infection and the COVID-19 severity. In total, we analyzed more than 100,000 single-nucleotide polymorphisms. The genome-wide association study suggested potential correlation between genetic variability in *POLR2A*, *ANKRD27*, *MAN1A2*, and *ERAP1* genes and SARS-CoV-2 infection susceptibility. The most significant gene locus associated with SARS-CoV-2 infection was located in *POLR2A* ($p = 5.71 \times 10^{-6}$). Furthermore, genetic variants in *PCNX2*, *CD200R1L*, *ZMAT3*, *PLCL2*, *NEIL3*, and *LINC00700* genes ($p < 1 \times 10^{-5}$) were closely associated with the COVID-19 severity in Chinese population. Our study confirmed that new genetic variant loci had significant association with SARS-CoV-2 infection and the COVID-19 severity in Chinese population, which provided new clues for the studies on the susceptibility of SARS-CoV-2 infection and the COVID-19 severity. These findings may give a better understanding on the molecular pathogenesis of COVID-19 and genetic basis of heterogeneous susceptibility, with potential impact on new therapeutic options.

Keywords COVID-19 · SARS-CoV2 · GWAS · Severity · Susceptibility

✉ Juan Chen
chenjuan2014@cqmu.edu.cn

✉ Li Zhou
zhouli@cqmu.edu.cn

✉ Ai-Long Huang
ahuang@cqmu.edu.cn

¹ Department of Epidemiology, School of Public Health and Management, Research Center for Medicine and Social Development, Innovation Center for Social Risk Governance in Health, Chongqing Medical University, Chongqing 400016, China

² Nanan District Center for Disease Control and Prevention, Chongqing, China

³ Key Laboratory of Molecular Biology On Infectious Diseases, Chinese Ministry of Education, Chongqing Medical University, Room 617, College of Life Sciences Building, 1 YixueYuan Road, YuZhong District, Chongqing 400016, China

⁴ Key Laboratory of Laboratory Medical Diagnostics, Chinese Ministry of Education, Chongqing Medical University, Chongqing, China

Introduction

In December 2019, a pneumonia outbreak occurred in Wuhan, China. This disease caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) rapidly spread worldwide and the WHO declared the pandemic of COVID-19 on March 11, 2020 [1, 2]. Globally, as of 2 May 2022, there have been 511,479,320 confirmed cases of COVID-19, including 6,238,832 deaths (data from WHO, <https://covid19.who.int/>).

Although it is still unclear about the early transmission at the initial of SARS-CoV-2 infection, accumulating information verified that this virus can be transmitted between humans and many subclinical cases exit after intimate contact. Like most viruses, there is a huge variation on the severity of symptoms in patients with SARS-CoV-2 infection [3]. Most patients with SARS-CoV-2 infection have reported mild to severe respiratory symptoms. Adults infected with SARS-CoV-2 usually present fever, cough, dyspnea, and

pneumonia. Elder patients with underlying disease or immunocompromised conditions are prone to severe situation such as acute respiratory distress syndrome [4]. However, there are some patients who are diagnosed positive by RT-PCR but are either asymptomatic or minimally symptomatic [5]. In addition, not all individuals exposed to SARS-CoV-2 are infected according to the epidemiological observation of the patients' close contacts. The sources of these variations are undoubtedly multifactorial, and one possibility lies with genetic variants and variable gene expression of host cells.

Genetic variants play a role in susceptibility to infectious diseases; the global genetic community has been actively investigating on genetic contribution to COVID-19 [6, 7]. New research showed that genetics also plays a role in the severity of COVID-19 [8, 9]. David et al. identified a 3p21.31 gene cluster as a genetic susceptibility locus in COVID-19 patients with respiratory failure and confirmed potential involvement of the *ABO* blood-group system [10]. A genetic variant in the factor 3 (*F3*) gene has been identified as a risk factor in severe COVID-19 patients [11]. Another mutation (rs150892504) found in the *ERAP2* gene codes for a zinc metalloaminopeptidase protein [12], which is important for the way that antigens are presented via the *HLA* class 1 binding peptides, thus activating T cells to respond and kill the infected cells [13]. However, whether genetic variants influence the outcome of SARS-CoV-2 infection (asymptomatic or symptomatic) in Chinese population still remains unknown.

In this study, to identify new susceptibility genes that might predispose patients to COVID-19, we conducted the whole exome genome sequencing (GWAS) for COVID-19 in a Han Chinese population, and identified new COVID-19 susceptibility loci that were involved in the immune response and associated with increased risk of SARS-CoV-2 infection.

Materials and methods

Study subjects

From February 1 to May 30, 2020, a total of 256 Han Chinese subjects were consecutively recruited in three hospitals of Chongqing (Chongqing Three Gorges Central Hospital, University-Town Hospital of Chongqing Medical University, and Yongchuan Hospital Affiliated to Chongqing Medical University). Among these individuals, 171 were confirmed cases of COVID-19 (tested positive for SARS-CoV-2) and 85 were close contacts of confirmed patients (tested negative for SARS-CoV-2). Written informed consent was obtained from all individuals involved in the study. This study was approved by the Ethics in Research Committee of Chongqing Medical University in Chongqing, China.

Laboratory confirmation

To identify SARS-CoV-2 infection, nasopharyngeal swabs were collected at least twice and tested by RT-PCR. RNA from all samples was isolated within 24 h. Viral RNA samples were extracted according to the manufacturer's instructions using the Nucleotide Acid Extraction Kit (DAAN Gene, Registration No. 20170583), based on an automated magnetic bead purification procedure. A commercial RT-PCR kit (DAAN Gene, Registration No. 20203400063) was used to test samples for SARS-CoV-2. Briefly, two target genes, namely open reading frame1ab (ORF1ab) and nucleocapsid protein (N), were simultaneously amplified and tested by RT-PCR. Primers of RT-PCR testing for SARS-CoV-2 were adopted according to the recommendation by the Chinese Center for Disease Control and Prevention. PCR cycling: 50 °C for 15 min, 95 °C for 15 min, 45 cycles containing 94 °C for 15 s, 55 °C for 45 s (fluorescence collection) [14].

Ct values less than 37 and greater than 40 were defined as positive and negative, respectively, for both genes. Samples with Ct values from 37 to 40 were defined as inconclusive, and a second test was needed. Starting 1 week after admission, nasopharyngeal samples were tested by RT-PCR every 2–3 days for the remainder of the hospitalization period. Patients with one positive RT-PCR result were defined as patients with SARS-CoV-2 infection. Patients with two consecutive negative RT-PCR results were defined as SARS-CoV-2 negative [14].

Definitions

A confirmed case of COVID-19 was defined as an individual with consecutive positive nucleic acid tests for SARS-CoV-2 (twice in every 24 h), using laboratory-based RT-PCR.

The symptomatic patients were defined as the patients who were laboratory-confirmed with COVID-19 and presented symptoms such as fever, cough, sore throat, and sputum.

An asymptomatic case was defined as an individual with a positive nucleic acid test result but without any relevant clinical symptoms in the preceding 14 days and during hospitalization.

A close contact was defined as anyone who had direct contact with infectious secretions of a COVID-19 patient. Close contact can occur while caring for, living with, visiting, or sharing a healthcare waiting area or room with patients with COVID-19. The duration of shedding was calculated as the number of days from the first positive nasopharyngeal sample to the last positive sample based

on RT-PCR testing. The last positive sample was followed by a negative RT-PCR result on two sequential tests [14].

Genetic analysis

Human genomic DNA samples were extracted by the Magnetic Beads Genomic DNA Extraction Kit (Nanjing ZhongkeBio Medical Co., Ltd.) according to the manufacturer's instruction, and the concentration was measured by a Nanodrop2000cspectrophotometer (Thermo Scientific, DE). A total amount of 0.3 µg DNA per sample was required for library generation. Standard instructions from the manufacturer were used for the SureSelectXT Target Enrichment System for Illumina Multiplexed Sequencing (G7530-90000, Qiagen), target capture with the Agilent SureSelect Human All Exon V6 (Qiagen), and 150 bp paired-end sequencing reads on the Illumina NovaSeq platform with bioinformatic processing and variant annotation. Briefly, genomic DNA sample was fragmented by sonication to a size of 350 bp. Then, DNA fragments were end-polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing, followed by further PCR amplification. After PCR products were purified (AMPure XP system), libraries were analyzed for size distribution by Agilent 2100 Bioanalyzer and quantified by real-time PCR (3 nM). At last, DNA library was sequenced on Illumina for paired-end 150 bp reads.

All WES data of individuals included in the initial discovery cohort were analyzed according to standardized GATK4 pipeline. The raw data were aligned to the hg19 human reference genome with the Burrows Wheeler Alignment (v 0.7.17) MEM algorithm. Duplication was marked by the Picard Mark duplicates (v 2.18.0) tool. Base Quality Score Recalibration (BQSR) was performed with GATK tools (v4.1.2.0) before SNP and indel calling was done with Haplotype Caller with interval lists specific to the exome enrichment kit for each sample. gVCF files were combined with Combine GVCFs, and then, genotyping was performed with the genotype GVCF tool implemented in GATK. The cohort call set was filtered with GATK Variant Recalibrator and Apply VQSR. Annotation of the VCF file was performed by a customized version

of ANNOVAR. The parameters were based on the GATK Best Practice.

Statistical analysis

Statistical analysis was performed with SPSS (version 22.0, SPSS Inc., Chicago, IL, USA). Descriptive statistics (mean, standard deviation, and percentage) were conducted to reflect the background characteristics of the participants. The χ^2 test or Fisher's exact test was used to analyze categorical variables. A *p* value of less than 0.05 was considered significant. Plink-1.9 was used in the GWAS analysis performed on the VQSR-passed germline SNPs. For quality control, SNPs with missing rate over 90%, MAF less than 1%, and hwe *p* value less than 0.000001 were filtered. Samples with DST over 0.85 were considered closely related and filtered. Then, chi-squared association statistics were calculated for each SNP based on default parameters. The Manhattan and QQ plots were drawn using R qqman packages.

Results

Demographic characteristics

The demographic characteristics are described in Table 1. The study included 87 symptomatic patients (G1), 84 asymptomatic cases (G2), and 85 close contacts of confirmed patients (G3), with the mean age of 49.45 years (SD 19.31), 43.55 years (SD 17.03), and 49.76 years (SD 19.08) in three groups respectively. No statistically significant difference was found in age or sex between groups.

Genome-wide association analysis

Symptomatic patients (G1) vs. close contacts (G3)

We performed a GWA scan in 87 symptomatic patients and 85 close contacts using the Sure Select XT Target Enrichment System for Illumina Multiplexed Sequencing. After the standard quality control (QC) filtering, 194,883 SNPs were included in the follow-up analyses. There was little evidence of inflation of the test statistic (λ 1000 for all invasive analysis = 1.015). We identified eleven risk associations at

Table 1 Characteristics of symptomatic and asymptomatic patients and close contacts

Variables	G1 (<i>n</i> = 87)	G2 (<i>n</i> = 84)	G3 (<i>n</i> = 85)	<i>p</i> value
Sex				
Male	49 (56.32)	40 (45.45)	48 (56.47)	<i>p</i> > 0.05
Female	38 (43.68)	44 (54.55)	37 (43.53)	
Age (year), mean ± SD	49.45 ± 19.31	43.55 ± 17.03	49.76 ± 19.08	<i>p</i> > 0.05

p value < 0.05 was considered significant

Table 2 Association evidence for 11 SNPs at 9 loci in GWAS (G1 vs. G3)

CHR	SNP	GENE	Allele1	Allele2	OR	SE	Raw <i>p</i>	Location
1	rs12728046	MAN1A2	G	A	0.05065	1.038	7.66E−05	Intronic
1	rs12119249	TARS2	T	C	0.3972	0.2365	8.16E−05	Intronic
1	rs4912132	IFFO2	A	G	2.547	0.2405	8.74E−05	Upstream
5	rs27042	ERAP1	A	G	3.136	0.2725	1.78E−05	Intronic
5	rs469876	ERAP1	G	A	2.969	0.2765	5.76E−05	Intronic
9	rs944513	PSAT1	G	T	0.3793	0.2468	7.01E−05	Intronic
15	rs62022788	PARP6	G	T	2.775	0.2451	2.46E−05	Intronic
17	rs34151785	POLR2A	T	C	6.476	0.458	5.71E−06	Intronic
18	rs28701630	C18orf63	A	G	0.3462	0.2735	7.57E−05	Intronic
19	rs3760943	ANKRD27	A	G	3.406	0.3235	8.64E−05	Intronic
19	rs74776662	ANKRD27	T	G	3.586	0.3411	9.54E−05	Intronic

nine different loci reaching genome-wide significance (effect allele frequency, $EAF \geq 0.05$, p value $< 1 \times 10^{-5}$) (Table 2; Fig. 1a). There was no evidence of heterogeneity for these associations. The SNP rs34151785 was the most significant SNP in COVID-19 GWAS (p value = 5.71×10^{-6}).

Symptomatic patients (G1) vs. asymptomatic cases (G2)

Figure 1b is the Manhattan plots of the COVID-19 GWAS, demonstrating the genome-wide association results for COVID-19. The association between 194,805 genotyped SNPs and COVID-19 severity (87 symptomatic patients and 84 asymptomatic patients) was analyzed, and the green horizontal line represented a p value of 1.0×10^{-4} . As shown in Table 3, rs1033323, with a p value of 7.37×10^{-5} in the joint analysis, is located in chromosome region 1q42.2 within the intronic region of the *PCNX2* gene.

Asymptomatic cases (G2) vs. close contacts (G3)

We also identified sixteen loci that showed marginal evidence of risk associations (p value of 1×10^{-5} to 8×10^{-6}) in 84 asymptomatic cases and 85 close contacts (Table 4; Fig. 1c). The SNPs, rs17032820, rs2289273, rs2289274, rs4684686, and rs2278554 were located in chromosome region 3p25.3 and associated with COVID-19 infection in joint analysis. These SNPs were located in a linkage disequilibrium (LD) block *ATP2B2* gene.

Pairwise LD patterns in three SNPs of the ERAP1 gene and expression quantitative trait locus (eQTL) analysis

The pair wise LD of three SNPs was generated from the Haploview4.2 software (Fig. 2). The LD block built up by rs27042, rs469876, and rs26618 suggested that they were significantly associated with each other ($D' > 0.90$).

Moreover, they were significantly associated with the mRNA levels of *ERAP1* in eQTL analysis (Table 5).

Molecular analyses

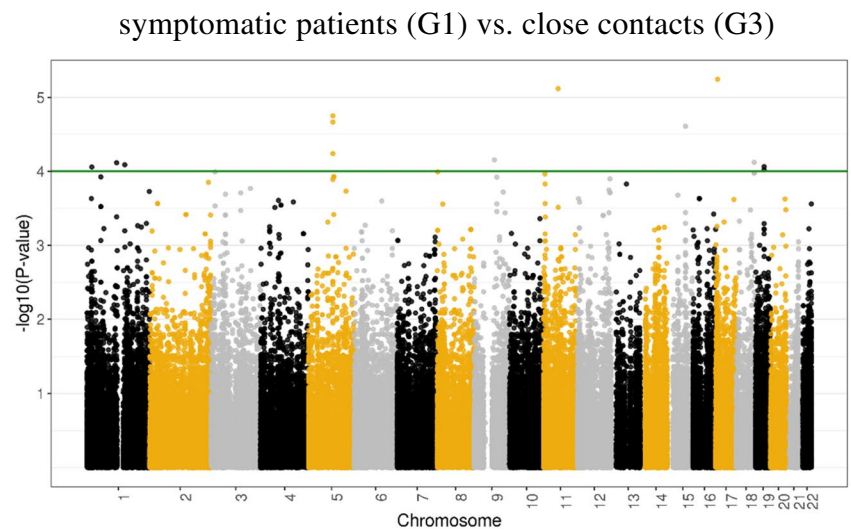
The SNP may exert a long-range effect on the expression of upstream and downstream genes of the loci. A search of genes that were present within 1 Mb of the SNPs (rs27042, rs469876) revealed some potentially interesting genes (Fig. 3a, b).

Discussion

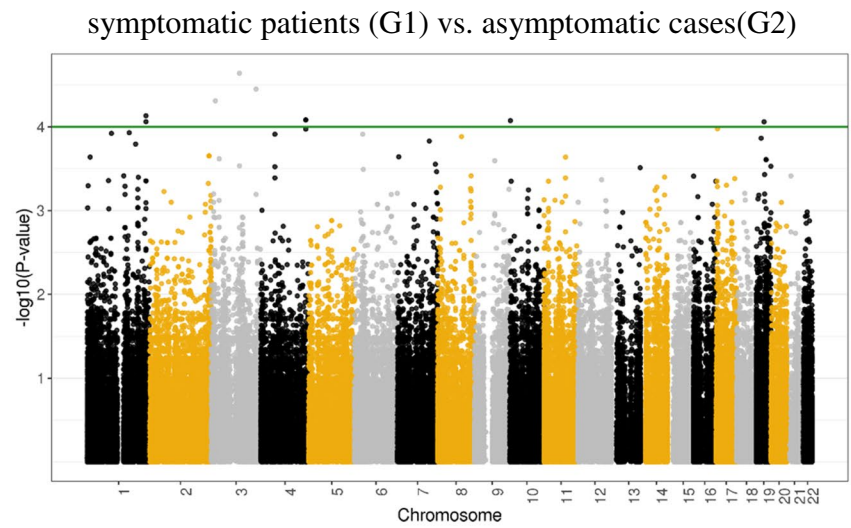
Individuals carrying specific variants of genes directly involved in viral infection (e.g., *ACE2*, *TMPRSS2*) or exhibiting differential expression of those genes may have inherently different susceptibility to SARS-CoV-2, which may explain the broad spectrum of symptoms and disease severity associated with COVID-19. We have conducted the genetic association study for the susceptibility to SARS-CoV-2 and COVID-19 severity.

Recently, genetic variants in *ABO*, *LZTFL1*, *ABCD3*, *C4BPA*, *TMEM181*, *BRF2*, *ERAP2*, *ALOXE3*, and *IFNAR2* genes have been reported to be associated with SARS-CoV-2 infection or COVID-19 severity [10, 15–22]. In the genome-wide association analysis, our study showed potential correlation between genetic variability in *POLR2A*, *ANKRD27*, *MANIA2*, and *ERAP1* genes (in Table 2) and the disease susceptibility. We compared our results with previously identified candidate regions. However, these loci except *ERAP1* have not been reported in European and other populations. D'Amico et al. [23] reported that the dysfunctional status of *ERAP1* and *ERAP2* enzymes may exacerbate the effect of SARS-CoV-2 infection. Surprisingly, we noted that rs27042 and rs469876 were closely associated with susceptibility to COVID-19. SNP rs27042 and rs469876 at chromosome5q15 (linked with $r^2 = 0.83$) of the *ERAP1* gene were significantly

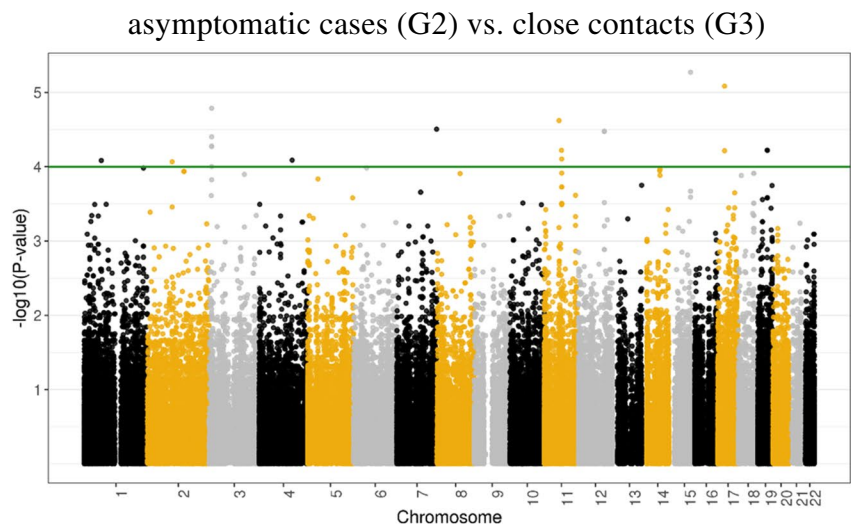
Fig. 1 Manhattan plot. Manhattan plot showing raw *p* value results from GWAS analysis. Each chromosome is depicted in a different color; the green horizontal line corresponds to the commonly adopted genome-wide significant level at 1×10^{-4} . **a** Summary of genome-wide association results for 87 symptomatic patients and 85 close contacts individuals. **b** Summary of genome-wide association results for 87 symptomatic patients and 84 asymptomatic cases. **c** Summary of genome-wide association results for 84 asymptomatic cases and 85 close contacts



(a)



(b)



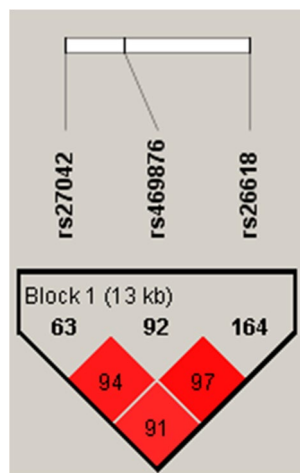
(c)

Table 3 Association evidence for 6 SNPs at 6 loci in GWAS (G1 vs. G2)

CHR	SNP	GENE	Allele1	Allele2	OR	SE	Raw <i>p</i>	Location
1	rs1033323	PCNX2	A	C	3.706	0.3437	7.37E−05	Intronic
3	rs6770923	CD200R1L	G	A	0.2379	0.3555	2.29E−05	Intronic
3	rs4955810	ZMAT3	C	T	9.175	0.6309	3.54E−05	Intronic
3	rs7653834	PLCL2	T	C	2.784	0.2551	4.90E−05	Exonic
4	rs3792606	NEIL3	G	A	0.3168	0.2992	8.25E−05	Intronic
10	rs1414125	LINC00700	C	A	0.3482	0.2727	8.43E−05	ncRNA_exonic

Table 4 Association evidence for 16 SNPs at 9 loci in GWAS (G2 vs. G3)

CHR	SNP	GENE	Allele1	Allele2	OR	SE	Raw <i>p</i>	Location
1	rs79498460	WDR78	A	G	0.2059	0.4324	8.25E−05	Intronic
2	rs11888101	FAM178B	G	A	3.412	0.3234	8.59E−05	Intronic
3	rs17032820	ATP2B2	G	A	0.2833	0.3021	1.63E−05	Intronic
3	rs2289273	ATP2B2	A	G	0.2989	0.3028	3.96E−05	Exonic
3	rs2289274	ATP2B2	A	G	0.31	0.2977	5.28E−05	Exonic
3	rs4684686	ATP2B2	G	A	0.3336	0.2773	5.34E−05	Intronic
3	rs2278554	ATP2B2	G	T	0.3859	0.2473	9.98E−05	Intronic
4	rs11737495	SCLT1	T	G	19.95	1.046	8.17E−05	Intronic
7	rs887607	MNX1	A	G	2.809	0.2512	3.13E−05	Intronic
11	rs4930642	TPCN2	A	G	0.2991	0.3156	7.87E−05	UTR5
12	rs2002895	UTP20	A	G	21.76	1.04	3.35E−05	Intronic
12	rs7963896	UTP20	T	C	21.76	1.04	3.35E−05	Intronic
12	rs7977402	UTP20	C	T	21.76	1.04	3.35E−05	Intronic
15	rs2305366	SLC28A1	A	G	0.2851	0.2834	5.34E−06	Intronic
17	rs634065	FOXN1	G	T	0.2767	0.2973	8.23E−06	Intronic
17	rs634061	FOXN1	A	G	0.3223	0.2892	6.10E−05	Intronic

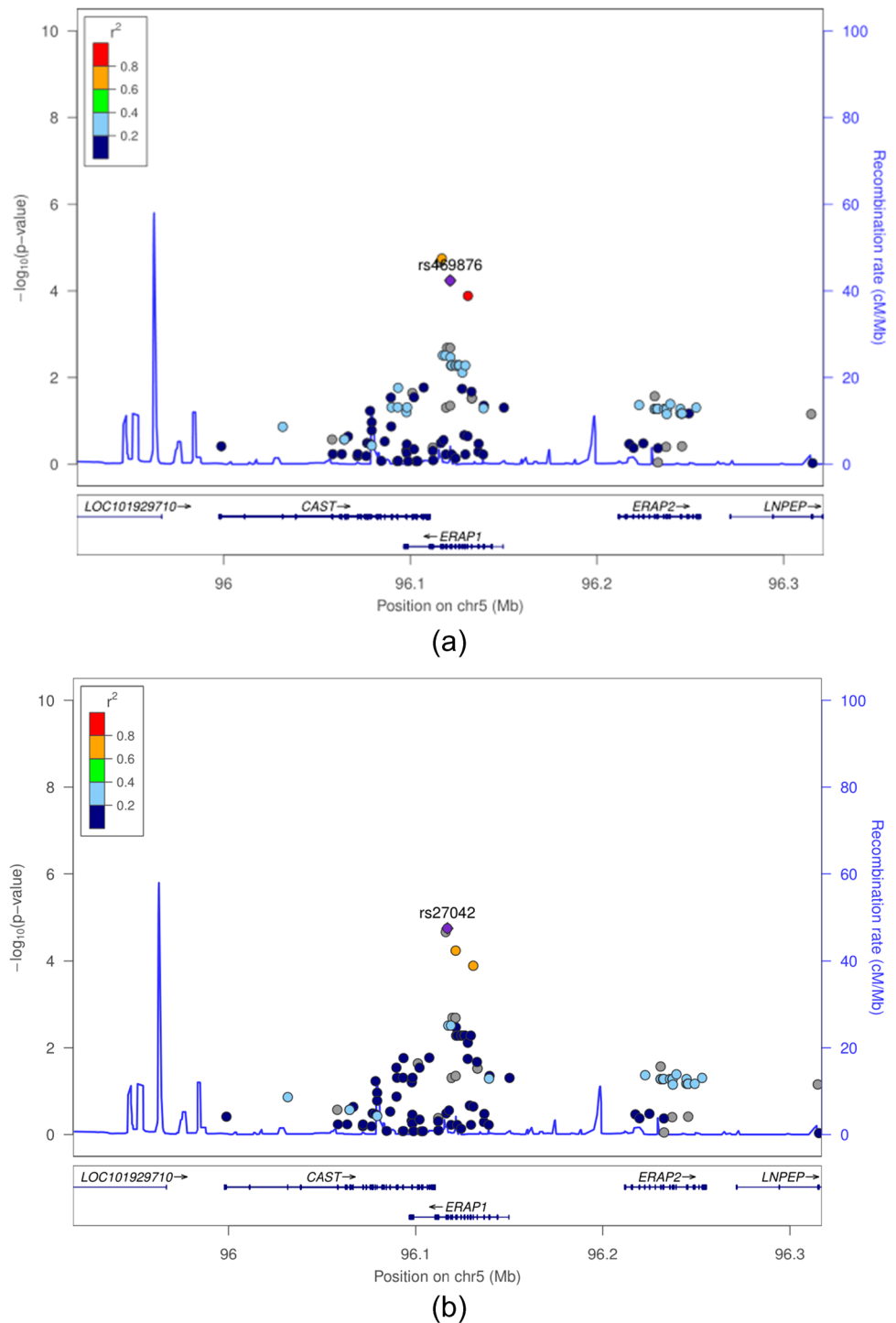
Fig. 2 LD patterns (*D'* plots) of the 3 SNPs in the *ERAP1* gene, as generated by Haploview v4.2. The LD block built up by the rs27042, rs469876, and rs26618**Table 5** Functional annotation for SNPs with strong linkage disequilibrium with the marker SNP rs rs469876

SNP	<i>r</i> ²	OR	Raw <i>p</i>	eQTL	Location
rs27042	0.83	3.136	1.78E−05	eQTL	Intronic
rs469876	1	2.969	5.76E−05	eQTL	Intronic
Rs26118	0.91	2.766	1.29 E−04	eQTL	Exonic

associated with the mRNA levels of *ERAP1* in eQTL analysis. *ERAP1* is an M1 zinc metalloprotease family member, which contains a transmembrane domain and an active site with GAMEN and Zn-binding HEXXH(X) 18 E motifs. *ERAP1* plays a role in peptide trimming in the generation of most *HLA* (human leukocyte antigen) class I-binding peptides, and is also involved in regulating proinflammatory cytokine signaling through cleavage of cytokine cell surface receptors [24]. Studies have highlighted the role of *ERAP1* in innate immune-mediated pathways involved in inflammatory responses, and indicated that SNP in *ERAP1* is associated with a number of autoimmune/inflammatory conditions [25–27].

We further explored the association between genetic variants and COVID-19 severity. GWAS using symptomatic patients of COVID-19 as the case and asymptomatic patients as the control suggested potential correlation between genetic variability in *PCNX2*, *CD200R1L*, *ZMAT3*, *PLCL2*, *NEIL3*, and *LINC00700* genes (in Table 3) and COVID-19 severity. None of these genes had been previously reported in the literature. Among the genetic loci associated with severe COVID-19, the 3p21.31 gene cluster has been reported to be robustly associated with COVID-19 severity. CCR9 and

Fig. 3 Regional plots of the two SARS-CoV-2 infection susceptibility loci



CXCR6 have been identified as putative causal genes of the 3p21.31 locus [10]. We found little evidence to suggest that allele frequency differences at this locus could account for the higher rate of severe outcomes from COVID-19. Janie F. Shelton and Anjali J. Shastri et al. also found no correlation between gene cluster 3P21.31 and severe COVID-19 [28]. In fact, the primary risk allele at the chromosome 3p21.31 locus is most common in European populations [29]. It is

well known that individuals of diverse racial and ethnic backgrounds harbor different allelic variants.

There are several limitations in our study. Firstly, the sample size is relatively small, and the power analysis indicates that sample size of around 250 is barely sufficient to identify genome-wide significant genetic variants with MAF greater than 0.2 and odds ratio greater than 1.8 given type I error rate 0.05. Secondly, the patients with COVID-19 were only

divided into asymptomatic and symptomatic group, which were not subdivided based on their severity. Thirdly, the extent to which individual genetic variant affects the susceptibility to SARS-CoV-2 and COVID-19 severity has not been assessed yet.

Conclusion

This study shows the genetic variability of SARS-CoV-2 genomes in Chinese population using multiple sequence alignment techniques. We have discovered new and highly plausible genetic association with the susceptibility to SARS-CoV-2 infection and COVID-19 severity. Importantly, our results provide preliminary insights that necessitate functional validation in future studies.

Acknowledgements We are sincerely grateful to all volunteering participants and the healthcare staff in our study.

Author contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Quan-Xin Long, Ji-Hua Ren, Hao Chen, Meng-Meng Li, and Zheng Cheng. The first draft of the manuscript was written by Jie Fan and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Ethics approval This study was approved by the Ethics in Research Committee of Chongqing Medical University in Chongqing, China.

Consent to participate Written informed consent was obtained from all individuals involved in the study.

Consent for publication All authors consent to publish.

Conflict of interest The authors declare no competing interests.

References

1. Tsankov BK, Allaire JM, Irvine MA, Lopez AA, Sauvé LJ, Valance BA et al (2021) Severe COVID-19 infection and pediatric comorbidities: a systematic review and meta-analysis. *Int J Infect Dis* 103:246–256. <https://doi.org/10.1016/j.ijid.2020.11.163>
2. Zu ZY, Jiang MD, Xu PP et al (2020) Coronavirus disease 2019 (COVID-19): a perspective from China. *Radiology* 296(2):E15–E25
3. Yu X, Yang R (2020) COVID-19 transmission through asymptomatic carriers is a challenge to containment. *Influenza Other Respir Viruses* 14(4):474–475
4. Zhang Y, Yang H, Li S, Li WD, Wang J, Wang Y (2021) Association analysis framework of genetic and exposure risks for COVID-19 in middle-aged and elderly adults. *Mech Ageing Dev* 194:111433
5. Chen C, Zhu C, Yan D, Liu H, Li D, Zhou Y et al (2021) The epidemiological and radiographical characteristics of asymptomatic infections with the novel coronavirus (COVID-19): a systematic review and meta-analysis. *Int J Infect Dis* 104:458–464. <https://doi.org/10.1016/j.ijid.2021.01.017>
6. Armstrong J, Rudkin JK, Allen N et al (2020) Dynamic linkage of COVID-19 test results between Public Health England's Second Generation Surveillance System and UK Biobank. *Microb Genom* 6(7):mgen00937
7. Grigorescu F, Lautier C (2020) How geneticists contribute to understanding of COVID-19 disease pathogenicity. *Acta Endocrinol (Bucharest)* 16(3):346–352
8. Clayton PT (2020) Is susceptibility to severe COVID-19 disease an inborn error of metabolism. *J Inher Metab Dis* 43(5):906–907
9. Kaser A (2020) Genetic risk of severe COVID-19. *N Engl J Med* 383(16):1590–1591
10. Ellinghaus D, Degenhardt F, Bujanda L et al (2020) Genomewide association study of severe COVID-19 with respiratory failure. *N Engl J Med* 383(16):1522–1534
11. Ramlall V, Thangaraj PM, Meydan C et al (2020) Immune complement and coagulation dysfunction in adverse outcomes of SARS-CoV-2 infection. *Nat Med* 26(10):1609–1615
12. Saulle I, Vanetti C, Goglia S et al (2020) A new ERAP2/Iso3 isoform expression is triggered by different microbial stimuli in human cells. Could it play a role in the modulation of SARS-CoV-2 infection. *Cells* 9(9):1951
13. Hu J, Li C, Wang S, Li T, Zhang H (2020) Genetic variants are identified to increase risk of COVID-19 related mortality from UK Biobank data. *medRxiv* 15(1):10
14. Long QX, Tang XJ, Shi QL et al (2020) Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat Med* 26(8):1200–1204
15. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur J Hum Genet* (2020) 28(6):715–718
16. Elhabyan A, Elyaacoub S, Sanad E, Abukhadra A, Elhabyan A, 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
17. Nelson CP, Sama IE, Codd V et al (2020) Genetic associations with plasma angiotensin converting enzyme 2 concentration: potential relevance to COVID-19 risk. *Circulation* 142(11):1117–1119
18. Smatti MK, Al-Sarraj YA, Albagha O, Yassine HM (2020) Host genetic variants potentially associated with SARS-CoV-2: a multi-population analysis. *Front Genet* 11:578523
19. Wallentin L, Lindbäck J, Eriksson N et al (2020) Angiotensin-converting enzyme 2 (ACE2) levels in relation to risk factors for COVID-19 in two large cohorts of patients with atrial fibrillation. *Eur Heart J* 41(41):4037–4046
20. Wang F, Huang S, Gao R 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
21. Zeberg H, Pääbo S (2020) The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* 587(7835):610–612
22. Beyerstedt S, Casaro EB, Rangel ÉB (2021) COVID-19: angiotensin-converting enzyme 2 (ACE2) expression and tissue susceptibility to SARS-CoV-2 infection. *Eur J Clin Microbiol Infect Dis* 40(5):905–919
23. D'Amico S, Tempora P, Lucarini V et al (2021) ERAP1 and ERAP2 enzymes: a protective shield for RAS against COVID-19. *Int J Mol Sci* 22(4):1705

24. López de Castro JA (2018) How ERAP1 and ERAP2 shape the peptidomes of disease-associated MHC-I proteins. *Front Immunol* 9:2463
25. Admon A (2019) ERAP1 shapes just part of the immunopeptidome. *Hum Immunol* 80(5):296–301
26. Pepelyayeva Y, Amalfitano A (2019) The role of ERAP1 in autoinflammation and autoimmunity. *Hum Immunol* 80(5):302–309
27. Reeves E, James E (2018) The role of polymorphic ERAP1 in autoinflammatory disease. *Biosci Rep* 38(4):BSR20171503
28. Shelton JF, Shastri AJ, Ye C et al (2021) Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat Genet* 53(6):801–808
29. Mapping the human genetic architecture of COVID-19. *Nature* (2021) 600(7889):472–477

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.