

RESEARCH

Open Access



# Exploring the genetic landscape of COVID-19 susceptibility and severity among patients in Türkiye

Ahmet Gorkem Er<sup>1,2,3†</sup>, Yavuzhan Çakır<sup>2†</sup>, Berrin Er<sup>4</sup>, Hüseyin Cahit Burduroğlu<sup>2</sup>, Mehmet Çakmak<sup>5</sup>, Lale Özışık<sup>5</sup>, Mine Durusu Tanrıöver<sup>6</sup>, Arzu Topeli<sup>4</sup>, Yeşim Aydın Son<sup>2\*</sup> and Serhat Ünal<sup>1\*</sup>

## Abstract

**Background** One of the most challenging factors for clinicians in managing COVID-19 has been differences in the clinical course. To investigate the parameters associated with severe disease in detail, along with examining known risk factors such as advanced age and comorbidities, understanding personal genetic factors is necessary, as the clinical course may change due to differences in the host genome.

**Methods** Human genetic variants reported to be associated with severe disease were genotyped in 68 patients in COVID-19 medical wards and 52 in COVID-19 intensive care units at Hacettepe University Adult Hospital.

**Results** The rs17860115 variant was significantly more prevalent in our cohort than in the European (non-Finish) population, whereas the rs2298659, rs2298661, rs4290734, and rs9271609 variants were significantly less common, which may reflect genetic differentiation, selective pressures, or protective factors within this population. While no significant association was found between variants and disease severity, notably, the ACE2 rs1548474 allele frequency was 38.0% in the ICU group and 22.9% in the non-ICU group (OR = 2.06; 95% CI 1.10–3.90;  $p = 0.02$ ).

**Conclusion** These findings emphasize the importance of examining genetic differences both within and across populations when developing new strategies for disease control and public health policies, particularly for infectious diseases such as COVID-19. They also point to the necessity for further research involving larger and more varied populations to validate these associations and to investigate the genetic factors that may drive them.

**Keywords** COVID-19, Severe disease, Genetics, Human variants

<sup>†</sup>Ahmet Gorkem Er and Yavuzhan Çakır contributed equally to this work.

\*Correspondence:

Yeşim Aydın Son  
yesim@metu.edu.tr  
Serhat Ünal  
sunal@hacettepe.edu.tr

<sup>1</sup>Department of Infectious Diseases and Clinical Microbiology, Hacettepe University Faculty of Medicine, Ankara, Türkiye

<sup>2</sup>Department of Health Informatics, Graduate School of Informatics, Middle East Technical University, Ankara, Türkiye

<sup>3</sup>Department of Medicine, Stanford University, Stanford, USA

<sup>4</sup>Division of Intensive Care Medicine, Department of Internal Medicine, Hacettepe University Faculty of Medicine, Ankara, Türkiye

<sup>5</sup>Department of Internal Medicine, Hacettepe University Faculty of Medicine, Ankara, Türkiye

<sup>6</sup>Vaccine Institute, Hacettepe University, Ankara, Türkiye



## Background

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), emerged in 2019, rapidly spreading worldwide and leading to a pandemic. The virus primarily affects the upper and lower respiratory tract, and patients present with respiratory symptoms ranging from mild to severe disease [1]. In addition to being a respiratory tract infection, COVID-19 is also considered a systemic disease due to lymphocytic endothelitis and widespread micro- and macrovascular involvement [2]. During the pandemic, we have observed how advancements in molecular biology and biotechnology, particularly next-generation sequencing, significantly influence genetics and virology [3, 4]. As the virus evolves, identifying emerging viral variants has become important, as some may spread more rapidly or cause changes in disease severity.

Single nucleotide variations (SNVs) in the human genome contribute to genetic diversity and can affect gene function. They are pivotal in understanding genetic disorders, drug responses, and disease susceptibility. The variability in COVID-19 symptoms, even among those living in the same household, highlights the complex interplay between host genetic makeup and disease outcomes. During the pandemic, several studies have focused on host genome sequencing to reveal SNVs associated with the susceptibility to and severity of COVID-19. This has led to the need to identify genetic markers that can predict an individual's susceptibility to infection with SARS-CoV-2 or even severe disease [5–7].

Large consortia, such as the severe COVID-19 Genome-Wide Association Study (GWAS) group, the GenOMICC and ISARIC groups, and the COVID-19 Host Genetics Initiative, have conducted GWAS studies on an unprecedented scale, with international data sharing to investigate the associations of host genetic variants with SARS-CoV-2 infection [2, 8]. The largest GWAS from the COVID-19 Host Genetics Initiative utilized genomic data from multiple clinical trials, existing biobank and cohort studies, and data from consumer genetic companies to investigate genetic loci associated with disease susceptibility or severity [9]. Their primary analysis included over 2.5 million population controls and over 50,000 COVID-19 patients categorized by disease severity and hospital setting; several genetic loci were associated with disease susceptibility or severity. The first signals reported to be associated with severe disease included genetic loci on chromosomes 3p21.31 and 9q34.2 [2, 10]. The 3p21.31 cluster is the most robustly transcribed signal in many studies, is associated with a twofold increased risk of respiratory failure from COVID-19, and is suggested to be inherited from Neanderthals [9, 11]. With increasing sample sizes, various other genetic loci have gained genome-wide significance

[8, 12, 13]. Many of these loci contain genes close to SNPs encoding proteins that regulate antiviral or proinflammatory host responses. For example, rs2298661, identified near the *TMPRSS2* and *MX1* genes, has been linked to severe outcomes [13]. Variants in *ACE2* and other genes have also been associated with disease progression [14, 15].

In this study, through a data-driven approach, we compile COVID-19-associated SNVs published during the COVID-19 pandemic to provide deeper insights into the genetic factors contributing to COVID-19 outcomes. The selected SNV profiles are genotyped by targeted multiplex sequencing to assess the differences in the genotypes of hospitalized COVID-19 patients in medical wards and intensive care units (ICUs) to explore the variants associated with susceptibility and severity [16–19].

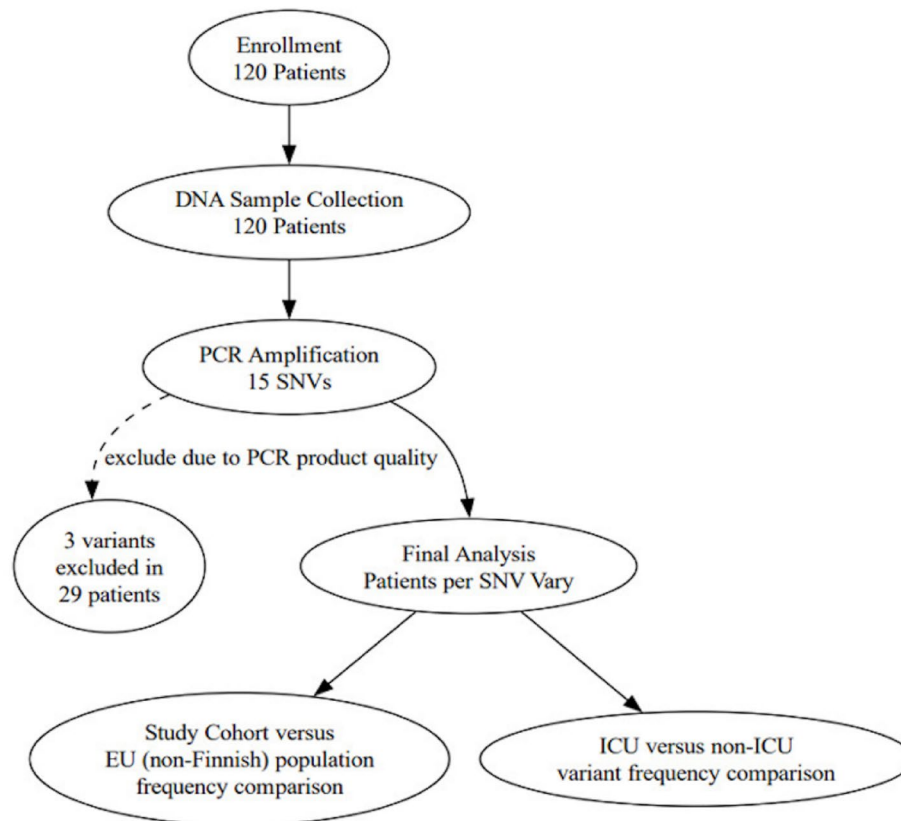
We hope that this study will enhance our understanding of genetic factors in the Turkish population and contribute to the broader scientific community, informing future genomic research, personalized medicine, and public health strategies against pandemic threats.

## Methods

### Patient data

This prospective cohort study was conducted at Hacettepe University Adult Hospital in Ankara, Turkey, between December 22, 2020, and May 5, 2021. After approval from the Institutional Ethics Committee with the number GO 2021/02–22, patients aged 18 years and older with a positive COVID-19 PCR test and who provided written informed consent were included in the study.

The current study cohort is a subset of participants originally recruited for our previous research in which we used sparse canonical correlation analysis and collaborative learning to identify and quantify associations between different data modalities, including viral genome sequencing, imaging, clinical data, and laboratory results [20]. DNA was collected from 120 COVID-19 patients hospitalized in medical wards and ICUs (Fig. 1). Sixty-eight patients were followed up in COVID-19 medical wards (non-ICU group), and 52 patients were followed up in COVID-19 intensive care units (ICU group). The ICU admission criteria were as follows: respiratory distress, tachypnea, increased oxygen demand, hypoxemia, symptoms related to sepsis, and acute organ dysfunction. Saliva samples were obtained via NORGEN BIOTEK CORP Saliva DNA kits, a noninvasive and user-friendly method frequently used in genotyping studies [21–23]. Patients provide saliva by spitting directly into sterile collection tubes containing a preservative solution, and for intubated patients, suction methods are employed to collect respiratory fluids. The samples were promptly sent to the laboratory for processing or stored at the



**Fig. 1** Overview of the study workflow

recommended temperature. DNA isolation was performed via a standardized protocol adapted for saliva samples [24]. All isolation procedures were conducted in a Class II biosafety cabinet in laboratories because of the airborne nature of SARS-CoV-2. After DNA isolation, the DNA products were controlled for the requirements of the DNA material by nanodrop measurements.

### Selected variants

The present study is grounded in a comprehensive literature search to identify SNVs associated with COVID-19 susceptibility and severity. This search was methodically structured to cover a broad spectrum of peer-reviewed research to identify key genetic markers that provide valuable insights into the genetic determinants influencing COVID-19 outcomes.

However, owing to the unprecedented number of COVID-19-related publications in which many were not peer reviewed and the time constraints imposed during the COVID-19 pandemic, it was not feasible to conduct a fully exhaustive review of the literature. Therefore, we relied on a substantial body of high-quality,

credible sources to build a short list of SNVs to be genotyped in patients (Table 1). Prominent studies, such as those investigating TMPRSS2 variants (e.g., rs2298661 and rs2298659), have elucidated their potential role in modulating disease severity. Additionally, other significant genetic loci (e.g., rs1123573 and rs8178521) associated with critical COVID-19 outcomes and long-term COVID-19 symptoms have been identified, further contributing to the genetic variability underlying susceptibility and disease progression. Linkage disequilibrium between SNVs on the same chromosome was assessed via LDlink's LDmatrix tool [25].

### Genotyping

The primers were designed via PyroMark Q24 software to amplify regions of interest, as listed in Table 2 [29, 30]. PCR amplification was conducted for each patient and each target region, totaling 1800 reactions with conventional PCR methods. The PCR amplification process was controlled and confirmed via gel electrophoresis. PCR products from individual patients were pooled for targeted multiplex sequencing. Sequencing was performed

**Table 1** List of polymorphic positions genotyped

Polymorphism	Gene	Location (hg38)	Allele	Population Frequency*	References
rs11088551	TMPRSS2	chr21:41508389	A>G	0,42	[26]
rs1123573	BCL11A	chr2:60480453	A>G	0,38	[7]
rs12610495	DPP9	chr19:4717660	A>G	0,28	[7]
rs1548474	ACE2	chrX:15,621,438	T>G	0,28	[15]
rs17574	DPP4	chr2:162073469	A>G	0,34	[27]
rs17860115	IFNAR2	chr21:33230000	C>A	0,32	[7]
rs2298659	TMPRSS2	chr21:41473447	G>A	0,23	[28]
rs2298661	TMPRSS2	chr21:41473715	C>A	0,16	[13]
rs35899679	TMPRSS2	chr21:41491393	C>A	0,34	[28]
rs4290734	TMPRSS2	chr21:41481156	A>G	0,41	[28]
rs4303794	TMPRSS2	chr21:41508379	A>C	0,42	[26]
rs463727	TMPRSS2	chr21:41464259	T>A	0,45	[7]
rs61882275	ELF5	chr11:34482745	G>A	0,37	[7]
rs8178521	IL10RB	chr21:33287378	C>T	0,26	[7]
rs9271609	HLA-DRB1	chr6:32623820	T>C	0,28	[7]

\*Allele frequencies are based on European(non-Finnish) population data

**Table 2** Forward (F) and reverse (R) primer pairs for amplifying variant regions

Variant	Primer Pairs
rs11088551	F - GTTCCTGCCTGGCTCAAC R - GGCTCACCCAGGACTCCA
rs1123573	F-GATTCAAGCGCTGTGCACT R-CCTTCCAGCCAGGTCATTTAGAAT
rs12610495	F-ATTATCTTGCTCTGATCACCCT R-AATCCATCTCTGCAGCTGTGTA
rs1548474	F-AAAATCTGGATTTGTGGCAGAAG R-ATCAGGCCAAATCACAGTAACAT
rs17574	F-CGCGGTCTCCCTCTCTAAC R-CACGGTGATGATGGTGACAAG
rs17860115	F-CACCCGCACTAAAGACGCTTC R-GCTCGGGGAAGTCTTCCG
rs2298659	F-CCCCTGGCATACTTTCC R-AACTCAAGCCGCCAGAGC
rs2298661	F-ACTGCTGGGGATGGACTTAGG R-GAGTCCCAGGGCCTTGTA
rs35899679	F-TTTTTTGAGACAGGGTCTTGC R-GCAGCCGAGTTATGAGAATC
rs4290734	F-AAATGTTCACTGCAACCCTCTTA R-TTGCAGCCTGTGTGAATTT
rs4303794	F-GCTGCGGAGGGACCCATA R-CGCCTACAGGAGCTCGTGA
rs463727	F-TGGGTCTTCTGTGCTTTTT R-GGTCCCTCAAATGACTCCTCTTA
rs61882275	F-CACCCACCCTGATGAGAATAA R-CGTCTTGCTTTCTACCCCTCATA
rs8178521	F-GGCAGCCTTGGGTTTTTCC R-CGAGGCTGTGATAGTGAGCTATGA
rs9271609	F-TTTTCATCACCTCCAAGGAGACC R-GGGCCTAAGCTCCTTCTGCATA

via Illumina multiplex sequencing protocols [31]. The sequencing data were stored in FASTQ format for downstream analysis.

Variant calling was conducted via standard bioinformatics workflows [32, 33]: Quality assessment and trimming of the raw sequencing reads were performed using Cutadapt version 2.6. The trimmed reads were then aligned to the human reference genome (hg38) using BWA version 0.7.17. The resulting SAM files were converted into sorted BAM files using SAMtools version 1.19. Variant calling was conducted with FreeBayes to generate VCF files. Variants with read depths less than 10 were filtered out to ensure data quality. Finally, variant information was extracted using Python and the pandas library, and the processed results were saved in CSV format. Individual observations of each allele were subsequently aggregated to ascertain the total allele frequency for the cohort for each SNV. Moreover, to provide further insight into the data, the numbers of heterozygous and homozygous patients were also recorded.

Allele frequencies were calculated for each SNV, considering only regions with sufficient read depth. The minor allele frequencies (MAFs) were determined as follows:

$$MAF = \frac{\text{Number of minor alleles}}{\text{Total number of alleles}}$$

#### Population studies

The genetic makeup of the modern Turkish population exhibits notable similarities with, particularly those in the Mediterranean region, as evidenced by clustering with Iberians from Spain and Tuscans from Italy. This clustering suggests a shared genetic heritage, further supported

by the similar frequency distributions of GWAS SNPs between Turkish and southern European populations, which indicate a greater proportion of ancestry sharing. This study assumes that the population of Turkey shares a genetic predisposition similar to that of the European (non-Finnish) population under normal conditions [34, 35].

### Statistical analysis

Statistical comparisons between the study cohort and the European (non-Finnish) population for host variants were performed via z tests. To account for multiple comparisons, a Bonferroni correction was applied, adjusting the significance threshold to  $p < 0.0033$ . The power for all variant comparisons is calculated on the basis of the number of patient alleles genotyped as the sample size.

Additionally, we simulated our data with 100 alleles for 1000 cohorts via the bootstrap approach [36]. After 1000 cohorts were simulated with 100 alleles via the bootstrap algorithm, 95% confidence interval limits were determined. We suggest that the frequencies of these variants, which differ in 1000 simulated cohorts containing 100 alleles, significantly differ between the study cohort and the population.

If the allele frequency seen in the European population was not within this confidence interval, it was defined as having a significant difference. The overall workflow of the bootstrapping algorithm was as follows:

1. Simulated allele frequencies were generated via NumPy on the basis of the observed frequencies.
2. Multiple cohorts (k) were created through resampling, and allele frequencies were recorded in each cohort.
3. The 95% and 99% confidence intervals were derived from the resampled data.
4. The observed allele frequencies in the European population were compared against these intervals to assess statistical significance.

Fisher's chi-square test and odds ratios were used to assess and measure the association of the MAF between the ICU and non-ICU patient groups. Bonferroni correction was also applied to this analysis.

### Pathway analysis

Significant SNVs were mapped to genes and analyzed for pathway enrichment via the KEGG database and functional annotation via Gene Ontology (GO) enrichment [37].

## Results

A total of 120 patients were included in the study. Fifty-two patients (43.3%) were admitted to the ICU. The mean age was  $59.0 \pm 16.1$  years, and 49 patients (40.9%) were female.

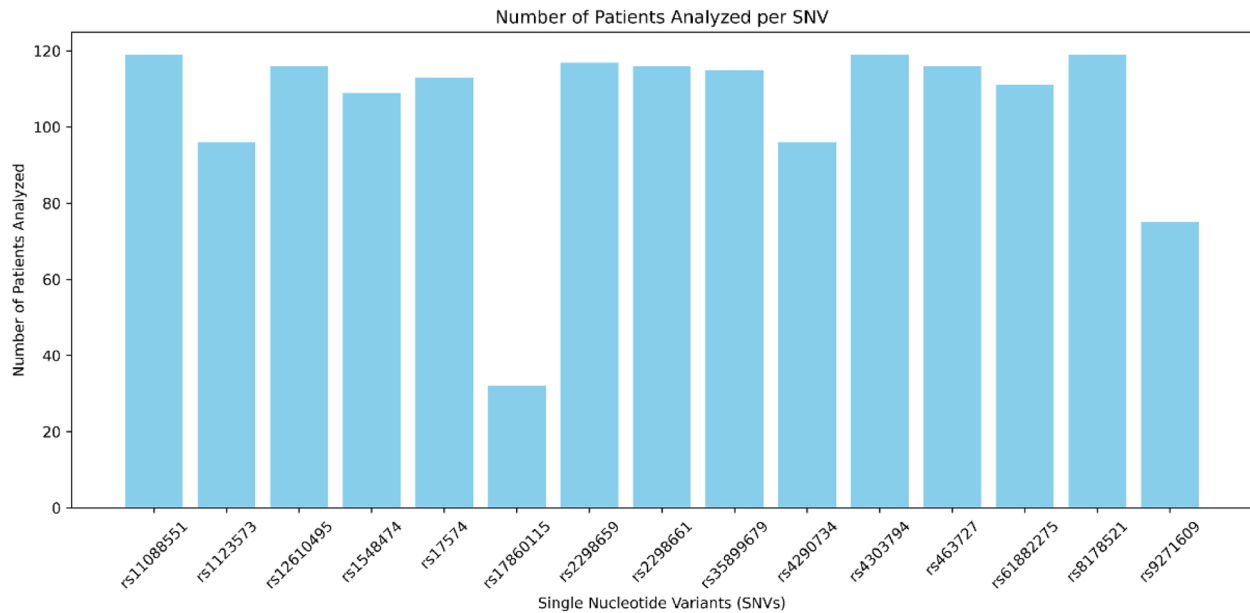
For host genome analyses, PCR was performed on DNA samples from 120 patients to amplify 15 regions of interest. Each PCR was designed to target specific SNVs associated with COVID-19 susceptibility. To ensure robust amplification, the PCR conditions were carefully optimized according to the steps mentioned in the methods. To ensure consistency when conducting multiple experiments simultaneously, we designed primers for different targets to run under similar PCR conditions. In 29 patient samples, for rs1123573, rs9271609, and rs17860115, either there was not enough material for PCR experiments or correct PCR products were not observed. Therefore, these 29 PCRs were not included in the analyses of these SNVs. On average, each SNV was represented in  $105.6 \pm 23.5$  patients, whereas 3 SNVs were genotyped in 119 patient samples (Fig. 2; Table 3).

Multiplex sequencing of the PCR-amplified target variant regions yielded predominantly satisfactory results, ensuring comprehensive coverage and accurate genotyping of SNVs in the patient cohort. The majority of the samples presented excellent read depths (up to 162334, with a mean read depth of 16522.5). In a few samples in which regions with read depths below the threshold of ten were identified, these low-quality reads were omitted from further analysis to maintain the integrity and accuracy of the dataset.

Overall, the sequencing process demonstrated robust performance, with the majority of target regions being accurately sequenced and analyzed. The quality control measures effectively mitigated the impact of sequencing errors, ensuring the reliability of the genetic insights derived from this study.

The patients were categorized on the basis of the presence of homozygous wild-type alleles, heterozygous alleles, and homozygous variant alleles, with their respective frequencies reported in Table 3.

Z tests were conducted to compare the variant allele frequencies between our cohort and the European (non-Finnish) population. Z test analysis results for the comparisons of SNV and European (non-Finnish) allele frequencies are presented in Table 4. After multiple testing corrections with the Bonferroni method, 5 SNVs demonstrated statistically significant differences in variant allele frequencies between our cohort and the European (non-Finnish) population. The rs17860115 variant allele was significantly more prevalent in COVID-19 patients, suggesting a potential cohort-specific variant or association with specific traits or conditions prevalent in this group. The rs2298659, rs2298661, rs4290734,



**Fig. 2** Patient counts per SNV

**Table 3** Patient allele frequencies for homozygous and heterozygous observations

Region	Patient Count	Wild type patient count	Wild type homozygous patient frequency	Heterozygous patient count	Heterozygous patient frequency	Homozygous variant patient count	Variant homozygous patient frequency	Variant allele frequency
rs11088551	119	40	0.33	61	0.51	18	0.15	0.404167
rs1123573	96	41	0.42	44	0.45	11	0.11	0.340206
rs12610495	116	74	0.63	39	0.33	3	0.03	0.192308
rs1548474	109	67	0.61	19	0.17	23	0.21	0.295455
rs17574	113	56	0.49	49	0.43	8	0.07	0.285088
rs17860115	32	9	0.27	4	0.12	19	0.58	0.636364
rs2298659	117	85	0.72	30	0.25	2	0.02	0.144068
rs2298661	116	84	0.72	30	0.26	2	0.02	0.145299
rs35899679	115	46	0.40	50	0.43	19	0.16	0.37931
rs4290734	96	93	0.96	0	0.00	3	0.03	0.030928
rs4303794	119	40	0.33	61	0.51	18	0.15	0.404167
rs463727	116	42	0.36	48	0.41	26	0.22	0.42735
rs61882275	111	35	0.31	59	0.53	17	0.15	0.415179
rs8178521	119	59	0.49	49	0.41	11	0.09	0.295833
rs9271609	75	46	0.61	29	0.38	0	0.00	0.190789

and rs9271609 variants were significantly less common in our cohort, which may reflect genetic differentiation, selective pressures, or protective factors within this population.

The remaining SNVs did not exhibit significant differences in allele frequency after Bonferroni correction, suggesting that their MAF distribution is within a similar range between our cohort and the European (non-Finnish) population. These SNVs may serve as genetic markers with similar prevalence across these populations. The statistical power for the significant SNVs ranged from 81 to 100%, indicating a high probability of detecting true

differences. However, some SNVs with higher p values had lower power, so they were classified as nonsignificant results.

#### Linkage results

The SNVs rs17860115, rs8178521, rs463727, rs2298659, rs2298661, rs4290734, rs35899679, rs4303794, and rs11088551 are located on chromosome 21. The heatmap for the linkage analysis was created as presented in Fig. 3. The linkage analysis for rs4303794 and rs1108855, with a score of 1, revealed that they were perfectly linked and

**Table 4** Frequency comparisons of the study cohort and the European (non-Finnish) population

SNV	Allele count	Variant allele frequency	Euro-pean (non-Finnish) frequency	z test p value	Power
rs11088551	238	0.40	0.42	0.472	9%
rs1123573	192	0.34	0.39	0.174	99%
rs12610495	232	0.19	0.27	0.004	84%
rs1548474	218	0.30	0.29	0.792	5%
rs17574	226	0.28	0.34	0.043	52%
rs17860115	64	0.63	0.32	<0.001*	100%
rs2298659	234	0.15	0.23	<0.001*	89%
rs2298661	232	0.15	0.22	0.001*	81%
rs35899679	230	0.38	0.47	0.016	80%
rs4290734	192	0.03	0.49	<0.001*	100%
rs4303794	238	0.40	0.42	0.472	9%
rs463727	232	0.43	0.46	0.305	15%
rs61882275	222	0.41	0.37	0.222	22%
rs8178521	238	0.29	0.27	0.377	10%
rs9271609	150	0.19	0.3	0.0002*	90%

According to the p values obtained from individual Z tests conducted with the frequencies and the sample sizes of patients' alleles, the following SNVs showed statistically significant differences, with p values < 0.0033: rs17860115, rs2298659, rs2298661, rs4290734, and rs9271609. While rs17860115 had a higher variant rate in the study cohort, other SNVs had lower MAFs

were expected to have the same MAF, which was 0.40, supporting the genotyping results.

#### Bootstrap algorithm results

As mentioned, a bootstrap-based analysis was developed and applied with the existing variant frequency. With variant frequencies obtained from the frequency analyses, 1000 cohorts of 100 alleles were created, and 95% confidence intervals of frequency means of cohorts were compared with European (non-Finnish) allele frequencies. Table 5 shows that the European (non-Finnish) allele frequencies are within the limits of the 95% confidence intervals of the simulated data.

#### Comparison of ICU and non-ICU frequencies

The comparison of allele frequencies between the ICU and non-ICU groups is summarized in Table 6. No statistically significant difference was found between these groups according to the MAF values in post hoc analyses considering the Bonferroni correction.

#### Kyoto encyclopedia of genes and genomes pathway analysis

For the SNVs that presented significant differences in MAF between the European (non-Finnish) population and the study cohort, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was conducted. The subnetwork revealed significant associations related to COVID-19 and other viral diseases (Fig. 4). Gene

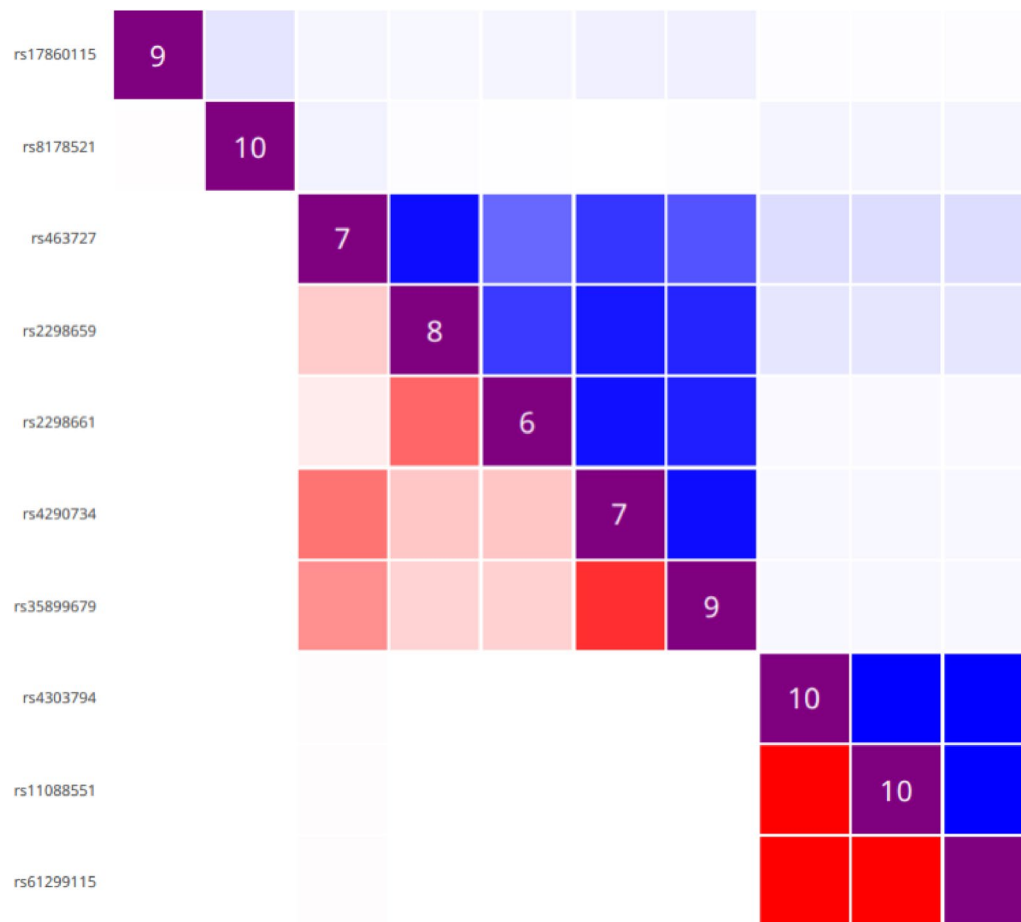
Ontology data revealed that HLA-DRB1 and TMPRSS2 are involved in the biological processes regulating viral entry into host cells and positively regulating viral entry into host cells. Both genes also play roles in the positive regulation of the viral life cycle and positive regulation by symbionts of entry into host cells. Additionally, HLA-DRB1 and ACE2 are implicated in the regulation of cytokine production, a critical response during viral infections.

#### Discussion

This study presents a detailed genetic analysis of SNVs and their associations with COVID-19 susceptibility and severity among patients at Hacettepe University Adult Hospital in Türkiye. Through a robust methodological approach encompassing sample collection, DNA isolation, PCR, sequencing, and advanced statistical analyses, our findings reveal significant insights into the genetic predispositions that may influence COVID-19 outcomes within the Turkish population.

While a case-control study design would be optimal, as the study was conducted during the pandemic, healthy individuals were not accessible due to COVID-19 precautions. Therefore, a cohort-population comparison study was conducted. The Turkish population frequencies from TÜSEB published in 2024 were only based on 100 whole-genome sequencing data [38]. To overcome these limitations, the European (non-Finnish) population was used to represent the reference MAF. When the frequencies were compared, significant MAF differences were observed for the rs17860115, rs2298659, rs2298661, rs4290734, and rs9271609 variants between the study cohort and the European (non-Finnish) populations. All five variants showed statistically significant differences according to the z test analysis, with powers ranging between 81% and 100%. The variant rs17860115 was more common in the COVID-19 patient cohort, indicating a possible association with an increased risk of COVID-19 infection. Conversely, the rs2298659, rs2298661, rs4290734, and rs9271609 variants were observed at significantly lower frequencies in the patient cohort compared to the NFE population MAF, suggesting a potential protective role against COVID-19 infection for those genotypes.

A bootstrapping approach was employed for the first time to validate the z test results and construct confidence intervals, allowing us to evaluate the MAF distributions of the variants in the study cohort compared with the population frequencies. The bootstrap analysis confirmed the z test results that rs17860115, rs2298659, rs4290734, and rs9271609 were significantly different from the population frequency. Only for rs2298661, which was the variant with the lowest power in the z test among the significant ones, bootstrap analysis was not



**Fig. 3** Linkage matrix heatmap for Chr 21 variants. rs11088551 and rs4303794 show perfect linkage ( $R^2 = 1.00$ ) and the same variant allele frequencies both in the study cohort and the European populations (0.40 in the study cohort, 0.42 in the European (non-Finnish) population)

confirmative, filtering out the variant with the lowest power (81%) after the z test analysis.

On the basis of our observations with this dataset, we suggest the use of the Bonferroni-adjusted p values to determine the significance of the variants in studies where cohort–population MAF comparisons are needed. Additionally, further validation of the significant variants with the bootstrapping approach is valuable for testing the power thresholds suggested with the z test. With bootstrapping, setting the power threshold according to the power of the last variant that satisfies the Bonferroni-adjusted p value yields a power threshold of 84%, which will filter out rs2298661 with a power of 81%. However, it should also be noted that, on the other hand, a more relaxed Holm–Bonferroni correction would filter rs12610495 with a power of 84%.

Here, the final list of variants was located in the HLA-DRB1 (rs9271609), IFNAR2 (rs17860115), and TMPRSS2 (rs2298659, rs4290734) genes. Studies have shown potential relationships between variants in the HLA-DRB1 gene and COVID-19 severity [39]. In this study, the rs9271609 variant in the HLA-DRB1 gene may be

potentially associated with decreased disease risk. The IFNAR2 gene has been identified in several studies as a significant factor in COVID-19 susceptibility and severity [40, 41]. Similarly, in this study, the rs17860115 variant was potentially related to COVID-19 susceptibility. According to the data from dbSNP, the single-nucleotide polymorphism rs17860115 is located within the 5' untranslated region (UTR) of the gene. The positioning of this variant in the UTR suggests that it may play a significant role in the regulation of gene expression or translation [42]. The TMPRSS2 gene has been found to play a significant role in COVID-19 susceptibility and severity. Rokni et al. reported an increased risk of COVID-19 in carriers of certain TMPRSS2 polymorphisms, whereas Wulandari et al. reported a possible association between the p.Val160Met polymorphism and SARS-CoV-2 infectivity and disease outcome [43, 44]. The rs2298659, rs2298661, and rs4290734 variants identified in this study showed potentially protective effects against COVID-19. According to the study of Uslu. et al., rs34536443 variant in TYK2, rs7515509 variant in AK5 gene were found

**Table 5** Significance analysis of the study cohort and European (non-Finnish) MAF with the bootstrapping approach

index	95% CI Lower bound	95% CI Upper bound	European (non- Finnish) allele frequency	Signif- icant Differ- ence
rs11088551	0.30975	0.5	0.42	No
rs1123573	0.25	0.43	0.39	No
rs12610495	0.12	0.27	0.27	No
rs1548474	0.21	0.38	0.30	No
rs17574	0.2	0.37	0.34	No
<b>rs17860115</b>	<b>0.54</b>	<b>0.73</b>	<b>0.32</b>	<b>Yes</b>
<b>rs2298659</b>	<b>0.08</b>	<b>0.21</b>	<b>0.23</b>	<b>Yes</b>
rs2298661	0.08	0.22	0.22	No
rs35899679	0.29	0.47	0.47	No
<b>rs4290734</b>	<b>0</b>	<b>0.07</b>	<b>0.49</b>	<b>Yes</b>
rs4303794	0.31	0.5	0.42	No
rs463727	0.34	0.52	0.46	No
rs61882275	0.33	0.51	0.37	No
rs8178521	0.22	0.39	0.27	No
<b>rs9271609</b>	<b>0.12</b>	<b>0.27</b>	<b>0.30</b>	<b>Yes</b>

MAFs for SNVs rs17860115, rs2298659, rs4290734, and rs9271609 are significantly different than the exact European (non-Finnish) MAF based on the bootstrap confidence intervals

statistically different between different severe, moderate, and mild/asymptomatic COVID-19 patient groups [45].

The comparison of allele frequencies between ICU patients and non-ICU patients reveals important insights that could have implications for understanding the role of genetic variants in determining COVID-19 severity. However, although no statistical significance was found for ACE2, rs1548474 attracted our attention, since the allele frequency of rs1548474 of ACE2 was 38.0% in the ICU group and 22.9% in the non-ICU group ( $p=0.02$ ;  $\alpha_{\text{Bonferroni}}=0.0036$ ). In studies examining the relationship of the ACE2 gene with COVID-19, direct or inversely proportional relationships were observed between ACE2 gene polymorphisms and COVID-19 severity [14, 46]. A recent meta-analysis of 84 different studies evaluating the associations of 130 polymorphisms in 61 candidate genes in more than 6,000 patients with severe symptoms and 8,000 patients with mild symptoms of COVID-19 revealed that ACE2 was significantly associated with the severity of COVID-19 [47]. In the preprint of the study conducted by Wooster et al. in the United States, six variants of ACE2 were associated with disease severity, and only rs1548474 was found to be significant for mild disease. The allele frequency of the variant was 19% in hospitalized patients and 42% in nonhospitalized patients. However, while the polymorphism in the dbSNP database was T > G, the major allele was determined to be G in this study [15]. On the other hand, Celik, et al.'s study population, ACE gene I/D, ACE2 receptor gene rs2106809, and rs2285666 polymorphisms were not associated with the severity of COVID-19 infection [48]. In a study from our

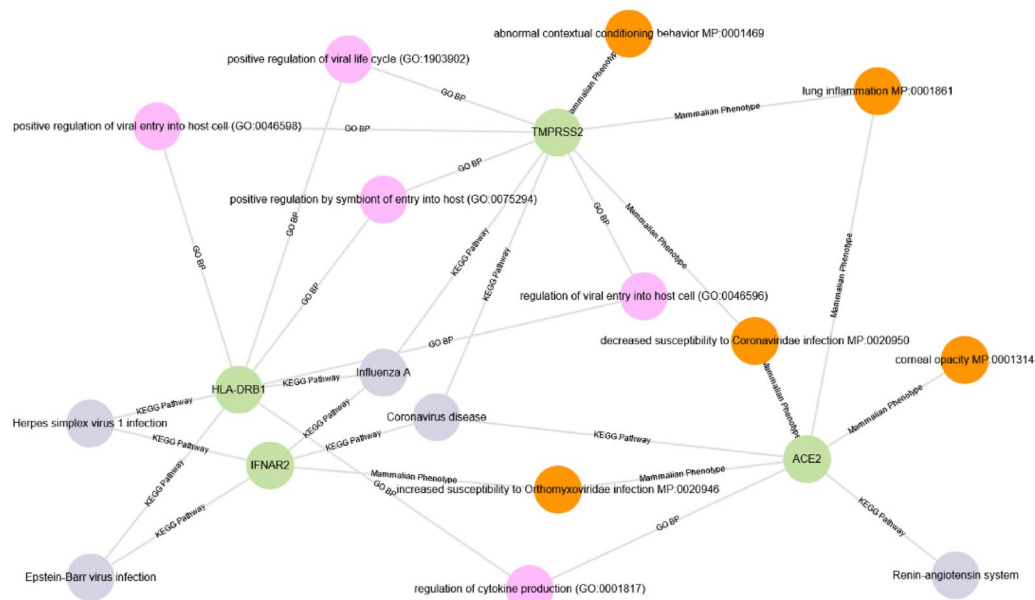
**Table 6** Comparison of allele frequencies between the ICU and non-ICU groups.

Polymorphism	ICU alleles	Non-ICU alleles	p value	OR (%95 CI)
rs11088551	n = 104 44 (42.3%)	n = 134 53 (39.6%)	p = 0.77	1.12 (0.64–1.95)
rs1123573	n = 92 33 (35.9%)	n = 100 33 (33.0%)	p = 0.79	1.14 (0.60–2.15)
rs12610495	n = 104 22 (21.2%)	n = 128 23 (18.0%)	p = 0.66	1.22 (0.60–2.48)
rs1548474	n = 100 38 (38.0%)	n = 118 27 (22.9%)	p = 0.02	2.06 (1.10–3.90)
rs17574	n = 98 28 (28.9%)	n = 128 37 (28.9%)	p = 1.00	0.98 (0.53–1.83)
rs17860115	n = 18 9 (50.0%)	n = 46 33 (71.7%)	p = 0.18	0.40 (0.11–1.42)
rs2298659	n = 104 17 (16.3%)	n = 130 17 (13.1%)	p = 0.60	1.30 (0.59–2.88)
rs2298661	n = 104 17 (16.3%)	n = 128 17 (13.3%)	p = 0.82	1.16 (0.56–2.41)
rs35899679	n = 104 36 (34.6%)	n = 126 52 (41.3%)	p = 0.37	0.75 (0.44–1.29)
rs4290734	n = 90 4 (4.4%)	n = 102 1 (1.0%)	p = 0.17	4.70 (0.52– 42.83)
rs4303794	n = 104 44 (42.3%)	n = 134 67 (50.0%)	p = 0.29	0.73 (0.44–1.23)
rs463727	n = 102 41 (40.2%)	n = 130 59 (45.4%)	p = 0.51	0.81 (0.48–1.37)
rs61882275	n = 100 43 (43.0%)	n = 122 50 (41.0%)	p = 0.87	1.09 (0.64–1.87)
rs8178521	n = 104 34 (32.7%)	n = 134 37 (27.6%)	p = 0.48	1.27 (0.73–2.26)
rs9271609	n = 72 15 (20.8%)	n = 78 14 (18.0%)	p = 0.81	1.20 (0.53–2.71)

$\alpha_{\text{Bonferroni}}=0.0033$

region; enrichment of the rs4646116 ACE2 functional allele in the Turkish population was suggested, and also The two TMPRSS2 missense variants, rs12329760 and rs75603675, that were detected in the Turkish population and have differential frequency distributions in dbSNP may have a role in population-specific outcomes in COVID-19 severity [49].

KEGG pathway analysis provided insights into pathway enrichment, focusing on the genes HLA-DRB1 (rs9271609), IFNAR2 (rs17860115), and TMPRSS2 (rs2298659, rs4290734) because of their significant frequency differences between the European (non-Finnish) population and our study cohort. Additionally, ACE2 was included in the analysis because of its notable frequency variation between the ICU and non-ICU patient groups. According to data from the MGI Mammalian Phenotypic database, knockout (KO) models for IFNAR2 and ACE2 in mice demonstrated increased susceptibility to Orthomyxoviridae infections. TMPRSS2- and ACE2-KO mice



**Fig. 4** Pathway enrichment analysis of susceptibility and severity variants revealed in the study with a KEGG pathway (enrichr). In this figure, green icons indicate genes, pink icons correspond to GO biological processes, gray icons represent items related to KEGG 2021 human pathways, and orange icons are designated MGI mammalian phenotype associations

presented with lung inflammation, whereas ACE2-KO mice presented with corneal opacity. Notably, TMPRSS2- and ACE2-KO mice presented reduced susceptibility to Coronaviridae infections, and TMPRSS2-KO mice presented abnormal contextual conditioning behavior [37]. The KEGG pathway database indicated that HLA-DRB1 and IFNAR2 gene products are components of the Epstein–Barr virus infection pathway. Furthermore, HLA-DRB1, IFNAR2, and TMPRSS2 are linked to the influenza A pathway. HLA-DRB1 and IFNAR2 are also involved in the herpes simplex virus 1 infection pathway. IFNAR2, TMPRSS2, and ACE2 were identified as key elements in the coronavirus disease pathway. Additionally, the involvement of ACE2 in the renin–angiotensin system pathway underscores its potential impact on COVID-19 severity and progression [37].

There are important limitations in our study. It can be assumed that our cohort size is limited, which makes the study prone to type 2 error. This study was conducted during the pandemic which limited our ability to expand the cohort. Furthermore, the study dates to the early days of the pandemic, and few patients were vaccinated for COVID-19. As it is difficult to appropriately quantify the impact of vaccination and more patients at risk for severe disease were vaccinated during the study period, the impact of vaccination on the development of severe disease cannot be determined. Our study is aiming to focus on previously reported variants, but also a genome-wide association study (GWAS) would provide a broader analysis of genetic factors associated with COVID-19 susceptibility and severity.

The frequency results for some variants could not be obtained for every patient. As shown in Fig. 2; Table 3 in the results section, although we were able to obtain information about most variants in most patients, we observed a high level of misses, especially in three variants. Although the inadequacy of the reading depths at this point is the final result, some situations may cause this. For example, the fact that some patients (especially in the ICU group) were intubated with saliva samples could not be obtained very efficiently and that there was not enough concentration in the isolated DNA due to errors that may have occurred during DNA isolation may have caused this situation. Furthermore, the fact that many patients experienced problems with variants for which we could not obtain sufficient results may indicate that this situation may be related to primer design or that there may have been problems during sequencing, especially in these regions. However, it was not possible to renew the tests after these failed sequencings because the PCR products or isolated DNAs had lost their freshness or because, owing to storing conditions, some samples degraded. Because of this, we did not sequence the same products since we were suspicious of the adequacy of the degraded samples. Additionally, since it would not be possible to obtain DNA samples from the same patients again, we continued the frequency scans by excluding these regions.

Additionally, while our study focused on genetic predispositions, the interaction between genetic, environmental, and social factors is crucial in determining disease outcomes. Comprehensive models integrating

these factors are needed to provide a more complete picture of COVID-19 susceptibility and severity.

## Conclusion

This study provides significant insights into the clinical and genetic factors associated with COVID-19 susceptibility and severity. As the global community continues to combat COVID-19, integrating genetic research into public health strategies remains a priority, promising more effective responses to this and future pandemics. Further research in this area will enhance our understanding of the genetic basis of infectious diseases and improve our ability to predict and mitigate their impacts on diverse populations.

## Abbreviations

COVID-19	Coronavirus disease
SNVs	Single nucleotide variations
ICU	intensive care unit
MAFs	minor allele frequencies
GWAS	Genome-Wide Association Study
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes

## Acknowledgements

We gratefully acknowledge Gen Era Diagnostics for providing support for DNA isolation, PCR, and genome sequencing. AGE also thanks to The Scientific and Technological Research Council of Türkiye (TUBITAK).

## Author contributions

A.G.E., Y.Ç. contributed equally to this work. Conceptualization, A.G.E., Y.Ç., M.D.T., Y.A.S., S.U.; methodology, A.G.E., Y.Ç., Y.A.S., M.D.T.; formal analysis, A.G.E., Y.Ç.; investigation, A.G.E., Y.Ç.; resources, M.D.T., Y.A.S., S.U.; data curation, A.G.E., B.E., Y.Ç., H.C.B., M.Ç.; writing—original draft preparation, A.G.E., Y.Ç.; writing—review and editing, A.G.E., Y.Ç., B.E., H.C.B., M.Ç., L.O., M.D.T., A.T., Y.A.S., S.U.; visualization, Y.Ç.; supervision, M.D.T., Y.A.S., S.U.; project administration, M.D.T., S.U.; funding acquisition, M.D.T., S.U. All authors have read and agreed to the published version of the manuscript.

## Funding

This study was funded by the Foundation for Influenza Epidemiology and the Turkish Society of Internal Medicine.

## Data availability

The datasets generated and analysed during the current study are available in the "https://dataview.ncbi.nlm.nih.gov/object/PRJNA1210972?reviewer=a95j109sku7k9644rkos5sek2r".

## Declarations

### Ethical approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Ethical Committee of Hacettepe University (protocol code: GO 2021/02–22). Informed consent was obtained from all the subjects involved in the study.

### Consent for publication

Informed consent was obtained from all the subjects involved in the study.

### Competing interests

The authors declare no competing interests.

Received: 4 January 2025 / Accepted: 24 July 2025

Published online: 30 December 2025

## References

1. Mehta OP, Bhandari P, Raut A, Kacimi SEO, Huy NT. Coronavirus disease (COVID-19): comprehensive review of clinical presentation. *Front Public Health*. 2020;8:582932. <https://doi.org/10.3389/fpubh.2020.582932>.
2. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, et al. Genomewide association study of severe Covid-19 with respiratory failure. *N Engl J Med*. 2020;383(16):1522–34. <https://doi.org/10.1056/NEJMoa2020283>.
3. Louten J. Chapter 7 - Virology research and diagnosis of viral infections. In: Louten J, editor. *Essential human virology (Second Edition)*. Second Ed ed. Boston: Academic; 2023. pp. 119–44.
4. Phalke S, Sawant SA, Samudra P, Yadav P, Chakraborty C, Jadhav A et al. Chapter 5.1 - Viral Genome Sequencing and Its Significance in Latest Clinical and Research Findings. In: Das S, Dash HR, editors. *Microbial Diversity in the Genomic Era (Second Edition)*. Second Ed ed: Academic Press; 2024. pp. 517–39.
5. Azzarà A, Cassano I, Paccagnella E, Tirindelli MC, Nobile C, Schittone V, et al. Genetic variants determine intrafamilial variability of SARS-CoV-2 clinical outcomes in 19 Italian families. *PLoS ONE*. 2022;17(10):e0275988. <https://doi.org/10.1371/journal.pone.0275988>.
6. Dasgupta I, Saini S, Khan MA, Chaudhary K. 1 - Population-level differences in COVID-19 prevalence, severity, and clinical outcome. In: Pandey R, editor. *Genomic surveillance and pandemic. Preparedness: Academic; 2023*. pp. 3–25.
7. Kousathanas A, Pairo-Castineira E, Rawlik K, Stuckey A, Odhams CA, Walker S, et al. Whole-genome sequencing reveals host factors underlying critical COVID-19. *Nature*. 2022;607(7917):97–103. <https://doi.org/10.1038/s41586-022-04576-6>.
8. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, et al. Genetic mechanisms of critical illness in COVID-19. *Nature*. 2021;591(7848):92–8. <https://doi.org/10.1038/s41586-020-03065-y>.
9. Niemi MEK, Karjalainen J, Liao RG, Neale BM, Daly M, Ganna A, et al. Mapping the human genetic architecture of COVID-19. *Nature*. 2021;600(7889):472–7. <https://doi.org/10.1038/s41586-021-03767-x>.
10. Shelton JF, Shastri AJ, Ye C, Weldon CH, Filshtein-Sonmez T, Coker D, et al. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat Genet*. 2021;53(6):801–8. <https://doi.org/10.1038/s41588-021-00854-7>.
11. Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from neanderthals. *Nature*. 2020;587(7835):610–2. <https://doi.org/10.1038/s41586-020-2818-3>.
12. Mapping the human genetic architecture of COVID-19. *Nature*. 2021;600(7889):472–7. <https://doi.org/10.1038/s41586-021-03767-x>.
13. Andolfo I, Russo R, Lasorsa VA, Cantalupo S, Rosato BE, Bonfiglio F, et al. Common variants at 21q22.3 locus influence MX1 and TMPRSS2 gene expression and susceptibility to severe COVID-19. *iScience*. 2021;24(4):102322. <https://doi.org/10.1016/j.isci.2021.102322>.
14. Sabater Molina M, Nicolás Rocamora E, Bendicho AI, Vázquez EG, Zorio E, Rodriguez FD, et al. Polymorphisms in ACE, ACE2, AGTR1 genes and severity of COVID-19 disease. *PLoS ONE*. 2022;17(2):e0263140. <https://doi.org/10.1371/journal.pone.0263140>.
15. Wooster L, Nicholson CJ, Sigurðslid HH, Cardenas CLL, Malhotra R. Polymorphisms in the ACE2 locus associate with severity of COVID-19 infection. *MedRxiv*. 2020. <https://doi.org/10.1101/2020.06.18.20135152>.
16. Brody T. Chapter 19 - Biomarkers. In: Brody T, editor. *Clinical trials (Second Edition)*. Second Ed ed. Boston: Academic; 2016. pp. 377–419.
17. Børsting C, Morling N. In: Siegel JA, Saukko PJ, Houck MMBTEFS, editors. *Single-Nucleotide polymorphisms*. Waltham: Academic; 2013. pp. 233–8.
18. Gunderson K, Barker DL, Bibikova M, Fan J-B. CHAPTER 7 - Genotype and epigenotype by single nucleotide polymorphism (SNP) analysis. In: Loring JF, Wesselschmidt RL, Schwartz PH, editors. *Human stem cell manual*. Oxford: Academic; 2007. pp. 85–95.
19. Nelson MR, Marnellos G, Kammerer S, Hoyal CR, Shi MM, Cantor CR, et al. Large-scale validation of single nucleotide polymorphisms in gene regions. *Genome Res*. 2004;14(8):1664–8. <https://doi.org/10.1101/gr.2421604>.
20. Er AG, Ding DY, Er B, Uzun M, Cakmak M, Sadee C, et al. Multimodal data fusion using sparse canonical correlation analysis and cooperative learning: a COVID-19 cohort study. *NPJ Digit Med*. 2024;7(1):117. <https://doi.org/10.1038/s41746-024-01128-2>.
21. Gudiseva HV, Hansen M, Gutierrez L, Collins DW, He J, Verkuil LD, et al. Saliva DNA quality and genotyping efficiency in a predominantly elderly population. *BMC Med Genom*. 2016;9:17. <https://doi.org/10.1186/s12920-016-0172-y>.

22. Rogers NL, Cole SA, Lan H-C, Crossa A, Demerath EW. New saliva DNA collection method compared to buccal cell collection techniques for epidemiological studies. *Am J Hum Biol*: Official J Hum Biology Council. 2007;19(3):319–26. <https://doi.org/10.1002/ajhb.20586>.
23. van Oorschot RAH, Verdon TJ, Ballantyne KN. Collection of samples for DNA analysis. *Methods in molecular biology*. (Clifton NJ). 2016;1420:1–12. [https://doi.org/10.1007/978-1-4939-3597-0\\_1](https://doi.org/10.1007/978-1-4939-3597-0_1).
24. Garbieri TF, Brozowski DT, Dionisio TJ, Santos CF, das Neves LT. Human DNA extraction from whole saliva that was fresh or stored for 3, 6 or 12 months using five different protocols. *J Appl Oral Science: Revista FOB*. 2017;25(2):147–58. <https://doi.org/10.1590/1678-77572016-0046>.
25. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, et al. Linkage disequilibrium in the human genome. *Nature*. 2001;411(6834):199–204. <https://doi.org/10.1038/35075590>.
26. Barash A, Machluf Y, Ariel I, Dekel Y. The pursuit of COVID-19 biomarkers: putting the spotlight on ACE2 and TMPRSS2 regulatory sequences. *Front Med*. 2020;7. <https://doi.org/10.3389/fmed.2020.582793>.
27. Alkharsah KR, Aljaroodi SA, Rahman JU, Alnafie AN, Al Dossary R, Aljindan RY, et al. Low levels of soluble DPP4 among Saudis May have constituted a risk factor for MERS endemicity. *PLoS ONE*. 2022;17(4):e0266603. <https://doi.org/10.1371/journal.pone.0266603>.
28. Asselta R, Paroschi EM, Mantovani A, Duga S. ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. *Aging*. 2020;12(11):10087–98. <https://doi.org/10.18632/aging.103415>.
29. Chuang L-Y, Cheng Y-H, Yang C-H. Specific primer design for the polymerase chain reaction. *Biotechnol Lett*. 2013;35(10):1541–9. <https://doi.org/10.1007/s10529-013-1249-8>.
30. Dieffenbach CW, Lowe TM, Dveksler GS. General concepts for PCR primer design. *PCR Methods Appl*. 1993;3(3):S30–7. <https://doi.org/10.1101/gr.3.3.s30>.
31. Heather JM, Chain B. The sequence of sequencers: the history of sequencing DNA. *Genomics*. 2016;107(1):1–8. <https://doi.org/10.1016/j.jygeno.2015.11.003>.
32. Koboldt DC. Best practices for variant calling in clinical sequencing. *Genome Med*. 2020;12(1):91. <https://doi.org/10.1186/s13073-020-00791-w>.
33. Zverinova S, Guryev V. Variant calling: considerations, practices, and developments. *Hum Mutat*. 2022;43(8):976–85. <https://doi.org/10.1002/humu.24311>.
34. Alkan C, Kavak P, Somel M, Gokcumen O, Ugurlu S, Saygi C, et al. Whole genome sequencing of Turkish genomes reveals functional private alleles and impact of genetic interactions with Europe, Asia and Africa. *BMC Genomics*. 2014;15(1):963. <https://doi.org/10.1186/1471-2164-15-963>.
35. Özçelik T, Kanaan M, Avraham KB, Yannoukakos D, Mégarbané A, Tadmouri GO, et al. Collaborative genomics for human health and Cooperation in the Mediterranean region. *Nat Genet*. 2010;42(8):641–5. <https://doi.org/10.1038/ng0810-641>.
36. Pattengale ND, Allipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A, editors. How many bootstrap replicates are necessary?? BT - Research in computational molecular Biology 2009 2009. Berlin, Heidelberg: Springer Berlin Heidelberg.
37. Evangelista JE, Xie Z, Marino GB, Nguyen N, Clarke Daniel JB, Ma'ayan A. Enrichr-KG: bridging enrichment analysis across multiple libraries. *Nucleic Acids Res*. 2023;51(W1):W168–79. <https://doi.org/10.1093/nar/gkad393>.
38. Turkish Genome Project. Data Sharing Portal [Available from: <https://tgd.tuseb.gov.tr/en/about>].
39. Anzures A, Naka I, Miki S, Nakayama-Hosoya K, Isshiki M, Watanabe Y, et al. Association of HLA-DRB1\*09:01 with severe COVID-19. *Hla*. 2021;98(1):37–42. <https://doi.org/10.1111/tan.14256>.
40. Ma Y, Huang Y, Zhao S, Yao Y, Zhang Y, Qu J, et al. Integrative Genomics Analysis Reveals a Novel 21q22.11 Locus Contributing to Susceptibility of COVID-19. *medRxiv*. 2020.
41. Ma Y, Huang Y, Zhao S, Yao Y, Zhang Y, Qu J, et al. Integrative genomics analysis reveals a 21q22.11 locus contributing risk to COVID-19. *Hum Mol Genet*. 2021;30(13):1247–58. <https://doi.org/10.1093/hmg/ddab125>.
42. TA B. *Genomes*. 2nd edition ed. Oxford: Wiley-Liss; 2002.
43. Rokni M, Heidari Nia M, Sarhadi M, Mirinejad S, Sargazi S, Moudi M, et al. Association of TMPRSS2 gene polymorphisms with COVID-19 severity and mortality: a Case-Control study with computational analyses. *Appl Biochem Biotechnol*. 2022;194(8):3507–26. <https://doi.org/10.1007/s12010-022-03885-w>.
44. Wulandari L, Hamidah B, Pakpahan C, Damayanti NS, Kurniati ND, Adiatmaja CO, et al. Initial study on TMPRSS2 p.Val160Met genetic variant in COVID-19 patients. *Hum Genomics*. 2021;15(1):29. <https://doi.org/10.1186/s40246-021-00330-7>.
45. Uslu K, Özcelik F, Zararsiz G, Eldem V, Cephe A, Sahin IO, et al. Deciphering the host genetic factors conferring susceptibility to severe COVID-19 using exome sequencing. *Genes Immun*. 2024;25(1):14–42. <https://doi.org/10.1038/s41435-023-00232-9>.
46. Rodrigues R, Costa de Oliveira S. The impact of Angiotensin-Converting enzyme 2 (ACE2) expression levels in patients with comorbidities on COVID-19 severity: A comprehensive review. *Microorganisms*. 2021;9(8). <https://doi.org/10.3390/microorganisms9081692>.
47. Gupta K, Kaur G, Pathak T, Banerjee I. Systematic review and meta-analysis of human genetic variants contributing to COVID-19 susceptibility and severity. *Gene*. 2022;844:146790. <https://doi.org/10.1016/j.gene.2022.146790>.
48. Karakas Celik S, Cakmak Genc G, Piskin N, Acikgoz B, Altinsoy B, Kurucu Issiz B, et al. Polymorphisms of ACE (I/D) and ACE2 receptor gene (Rs2106809, Rs2285666) are not related to the clinical course of COVID-19: A case study. *J Med Virol*. 2021;93(10):5947–52. <https://doi.org/10.1002/jmv.27160>.
49. Duman N, Tuncel G, Bisgin A, Bozdogan ST, Sag SO, Gul S, et al. Analysis of ACE2 and TMPRSS2 coding variants as a risk factor for SARS-CoV-2 from 946 whole-exome sequencing data in the Turkish population. *J Med Virol*. 2022;94(11):5225–43. <https://doi.org/10.1002/jmv.27976>.

## Publisher's note

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