ACE2, TMPRSS2, TYK2, SLC6A20, and IFNAR2 human genes variants influence SARS-CoV-2 infection susceptibility

Seyedeh Sepideh Aghamirli[1] and Nasrollah Saleh-Gohari[1]

[1]Kerman University of Medical Sciences Department of Medical Genetics

July 26, 2023

Abstract

The Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) causes a global pandemic named COVID-19. That virus causes a range of human respiratory tract infections; that severity varies from relatively mild to severe respiratory injury syndrome. It has indicated that some individuals might offer susceptibility to SARS-CoV-2 infection due to genetic factors. ACE2, TMPRSS2, TYK2, SLC6A20, and IFNAR2 human genes are involved in the pathogenesis of coronavirus in various populations and geographic territories; Therefore, examining the genetic variants of these genes can determine their association with the severity of the COVID-19 disease. In this study, the Whole-Exome Sequencing (WES) technique was used to identify variants of the mentioned human genes concerning the presence or absence of SARS-CoV-2 infection in the cohort of 100 individuals from Iran; that may modulate viral infectivity and make some individuals more vulnerable than others. Next, the frequency of variants found in the Iranian population was compared with those belonging to reference individuals from the 1000 Genomes Project, genomAD, and ExAC. In addition, due to the extraordinary importance of the protein's three-dimensional structure in maintaining the optimal function of the protein, also protein modeling was performed for the essential found variants. The ACE2 gene showed a high level of polymorphism. While TMPRSS2 is less polymorphic. The variants rs759499720/ACE2, rs776459296/ACE2, rs386818798/TMPRSS2, rs771922681/TYK2, rs753470142/TYK2, c.675G>T/TYK2, rs147760034/SLC6A20, rs139008024/SLC6A20, and rs759744926/IFNAR2 showed a significant association with SARS-CoV-2 infection and COVID-19. These variants have previously been detected in studies.

ACE2, TMPRSS2, TYK2, SLC6A20, and IFNAR2 human genes variants influence SARS-CoV-2 infection susceptibility

Seyedeh Sepideh Aghamirli and Nasrollah Saleh-Gohari

Department of Medical Genetics, Faculty of Medicine, Kerman University of Medical Sciences, Kerman, Iran

Correspondence

Nasrollah Saleh-Gohari, Department of Medical Genetics, Faculty of Medicine, Kerman University of Medical Sciences, Kerman, Iran.

Email: n_salehgohari@kmu.ac.ir and salehgohari@yahoo.co.uk

Postal address: Kerman - End of 22 Bahman Boulevard - Bahonar University Complex - Afzalipour Medical School

Telephone: (+98) 34 32532451

Fax number: (+98) 34-33257671

Abstract

The Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) causes a global pandemic named COVID-19. That virus causes a range of human respiratory tract infections; that severity varies from relatively mild to severe respiratory injury syndrome. It has indicated that some individuals might offer susceptibility to SARS-CoV-2 infection due to genetic factors. ACE2, TMPRSS2, TYK2, SLC6A20, and IFNAR2 human genes are involved in the pathogenesis of coronavirus in various populations and geographic territories; Therefore, examining the genetic variants of these genes can determine their association with the severity of the COVID-19 disease. In this study, the Whole-Exome Sequencing (WES) technique was used to identify variants of the mentioned human genes concerning the presence or absence of SARS-CoV-2 infection in the cohort of 100 individuals from Iran; that may modulate viral infectivity and make some individuals more vulnerable than others. Next, the frequency of variants found in the Iranian population was compared with those belonging to reference individuals from the 1000 Genomes Project, genomAD, and ExAC. In addition, due to the extraordinary importance of the protein's three-dimensional structure in maintaining the optimal function of the protein, also protein modeling was performed for the essential found variants. The ACE2 gene showed a high level of polymorphism. While TMPRSS2 is less polymorphic. The variants rs759499720/ACE2, rs776459296/ACE2, rs386818798/TMPRSS2, rs771922681/TYK2, rs753470142/TYK2, c.675G>T/TYK2, rs147760034/SLC6A20, rs139008024/SLC6A20, and rs759744926/IFNAR2 showed a significant association with SARS-CoV-2 infection and COVID-19. These variants have previously been detected in studies.

Keywords

COVID-19, SARS-CoV-2 infection, whole-exome sequencing, ACE2, TMPRSS2, TYK2, SLC6A20, IFNAR2

Introduction

Since December 2019, a new form of severe acute respiratory syndrome (SARS) from a new strain of coronavirus (SARS coronavirus 2 [SARS-CoV-2]) has been spreading worldwide [1]. The disease caused by SARS-CoV-2 was named COVID-19 and was declared a pandemic by the World Health Organization (WHO) in March 2020[2]. The signs and symptoms of COVID-19 disease are diverse in different patients [3]. A combination of genetic and non-genetic factors between the genome of the invading virus and the host may interact in the process of virus infection and determine the severity of the disease outcome [4]. Potential genetic and epigenetic factors of the host are associated with increased severity of COVID-19 infection [5,6]. So far, genetic variants of human genes related to the entry mechanism of SARS-CoV-2 and genes related to the host's innate immune response have been identified as essential host determinants in the severity of COVID-19 [7]. The advent of the Next-Generation Sequencing (NGS) technique has opened a new path for variant screening, also Whole-Exome Sequencing (WES) offers a promising approach to identify causative variants in known genes as well as in novel disease-related candidate genes [8,9]. Through the technique of WES, the frequency of exonic variants of genes concerning the severity of SARS-CoV-2 infectivity in patients can be tracked[10]. It is expected that this technique will be the first choice of clinical services in the near future because it provides useful information for genetic counseling [11], treatment [12], and disease management[13]. Host genetic factors contributing to severe COVID-19 have been investigated mainly in adults by Genome-Wide Association Studies (GWAS) and Whole-Exome Sequencing (WES)/Whole-Genome Sequencing (WGS) [14,15]. The first detailed insight into host genetics associated with severe COVID-19 infection was provided by GWAS [16]. GWAS is ideal for identifying variants with a relatively high allelic frequency (>5% minor allelic frequency) associated with specific Single Nucleotide Polymorphisms (SNPs) [17]. While the overall relative risk of these SNPs is mostly small [18], they can highlight specific genes that may have key pathogenic roles in the disease [19]. Several GWAS have investigated host genetic variants in clinical phenotypes of severity/susceptibility to COVID-19 [20]. The COVID-19 Host Genetics Initiative (HGI) is constantly updating its GWAS meta-analysis (latest HGI version 6 available at https://www.covid19hg.org/results/r6/)[21]. GWAS have identified various loci associated with susceptibility to infection or severity of COVID-19[22], which were later confirmed by colocalization analysis [23] or Mendelian Randomization (MR)[24]. The most consistent region identified among several GWAS associated with severe COVID-19 is at 3p21.31 chromosomal region, which includes several genes, including SLC6A20[25]. Several studies that also considered gene expression data have shown alleles

with low expression of the IFNAR2 gene are associated with an increased risk of severe COVID-19[26,27]. Finally, there is evidence that genetic polymorphisms may influence the risk of SARS-CoV-2 infection through viral entry [28], including ACE2 and TMPRSS2 genes[29,30]. Several studies have used a gene-based approach to identify functional variants of these proteins associated with the disease and severity of COVID-19 [31]. Overall, GWAS in adults with severe COVID-19 has identified genes associated with antiviral functions such as TYK2 [32]. The present study was designed to identify variants of human genes that are the cause of severe COVID-19 and investigate the effect of variants on the three-dimensional structure of the relevant proteins. An advanced understanding of host genetic and epigenetic factors and SARS-CoV-2 virus interactions is critical for improved prognostic tools and innovative therapies [33]. This research not only contributes to our better understanding of the disease and accurate identification of genetic differences; Rather, it facilitates the process of predicting the severity of the disease and, as a result, prompt and timely treatment of patients. In this study, Whole-Exome Sequencing is used to analyze variants in ACE2, TMPRSS2, TYK2, SLC6A20, and IFNAR2 in a cohort of 100 patients with a history of COVID-19 from Iran.

Material and Methods

Sample selection

In this study, 100 Iranian subjects were selected from among the patients referred to the genetics laboratory to carry out investigations to determine the etiology of various non-infectious genetic diseases. The participants have been genotyped and followed up for COVID-19 genetic risk factors for 2 years (2021-2023). Individuals under the age of 18 are one of the exclusion criteria. Information about SARS-CoV-2 infection was collected from a group of 100 selected individuals via a questionnaire administered to patients [34].

Genotype analysis

Blood samples taken from patients were washed with lysis buffer so that RBCs were separated. Then, genomic DNA was extracted from the WBCs through the salting-out method and the extracted DNA samples were stored at -20°C until analysis. To assess the purity of the extracted DNA, the Optical Density (OD) of the samples was measured by spectrophotometry in a nanodrop device. The whole exome of 100 participants in this project was sequenced using the Illumina HiSeq 2500 platform with an average coverage of 50X [35]. Informed written consent had been acquired from all participants. The study was approved by the ethics committee of the Kerman University of Medical Science. To have a brief review, after checking the read quality by measuring Quality Control (QC) score through the Phred scale, the raw reads aligned to the human reference genome assembly (GRCh38) using BWA. As well, VCF files of multi-sample were generated utilizing the GATK tool. All ACE2, TMPRSS2, TYK2, SLC6A20, and IFNAR2 variants were extracted for further analysis. Variants were annotated using the dbSNP, ClinVar, Varsome, and Franklin databases, which include population-oriented data on nucleotide and amino acid sequence changes. The highest population minor allele frequency (MAF) of all variants was checked to be less than 1%.

In-Silico analysis

Various bioinformatics software for molecular dynamics simulation has been used to determine the effect of the genetic variant on the amino acid sequence, including determining the effect of the variant on the primary transcripts of gene and alternative transcripts, as well as the potential effect of the variant on the function and tertiary structure of the protein. Data are also provided on Polyphen-2, SIFT, MutationTaster, FATHMM-MKL, and CADD scores.

Protein modeling

Most methods generate models interactively based on the user requests; For example, I-TASSER. Here, homology modeling was applied by the I-TASSER server to create the 3D structure of the trimeric studied protein that can calculate the effect of genetic variants on protein structure and stability. All files in PDB (Protein Data Bank) format were obtained from the I-TASSER server. UCSF ChimeraX was applied for the graphical visualization of molecules. The matchmaker tool was chosen to superimpose related structures without worrying about numbering or missing residues. This tool superimposes proteins by creating an

alignment and then matches the aligned residues to the 3D structure. Also, all figures of 3D structures and alignments have been assembled with the UCSF ChimeraX software.

Statistical analysis

Statistical analysis was performed using SPSS version 26.0 software and GraphPad Prism 9.4 software was used to draw graphs. The Chi-Square statistical test was used to evaluate the association between different variants of human genes ACE2, TMPRSS2, TYK2, SLC6A20, and IFNAR2 with the severity and incidence rate of COVID-19. It should be noted that in all analyses the group without variants was considered as the reference group for calculating the Odds Ratio (OR). P-Value < 0.05 was considered statistically significant.

Results

Demographic characteristics of the study participants

A total of 100 individuals underwent sequencing, 48 were male (48.0% of the study subjects), and 52 were female (52.0% of the study subjects) with an average age of 25.89 years. Of the total of 100 subjects analyzed, forty cases of SARS-CoV-2 infection (40.0%) were identified, with the remaining sixty (60.0%) not having been infected. The analysis of demographic variables shows that the gender of the participants in the study does not have a significant difference between subjects with a history of contracting COVID-19 and those without a history of contracting COVID-19 (P-Value = 0.438). Vaccinated subjects had a significantly higher history of contracting COVID-19, which may have been because these people were more willing to get vaccinated after contracting COVID-19 (P-Value < 0.001). Direct or indirect contact of subjects participating in the study with patients suffering from COVID-19 has significantly caused more people to contract COVID-19; In other words, the frequency of people with a history of COVID-19 is significantly higher in people who had contact with COVID-19 patients (P-Value > 0.0001) (Table 1). The age of the participants in the study is also significantly higher in those with a history of COVID-19, which is consistent with previous studies (P-Value = 0.001) (Table 2).

Table 1: Demographic characteristics of study participants

Table 2: Average age of study participants

The patient's VCF file analysis

Among the 100 analyzed patients, 40 patients had at least one variant in the five investigated human genes ACE2, TMPRSS2, TYK2, SLC6A20, and IFNAR2, and subsequently 60 patients did not have any variants in the mentioned genes. The result of this analysis shows a total of 140 variants. 35 variants for ACE2 gene, 21 variants for the TMPRSS2 gene, 29 variants for TYK2 gene, 30 variants for SLC6A20 gene, and 25 variants for IFNAR2 gene have been identified. The frequency graph of variants shows that ACE2 gene variants are more common among the study subjects (Diagram 1. a). Not considering duplicate variants, ACE2 gene has 7 unique variants, TMPRSS2 gene has 3 unique variants, TYK2 gene has 4 unique variants, SLC6A20 gene has 4 unique variants, and also IFNAR2 gene has 3 unique variants. Therefore, ACE2 gene is more polymorphic than other genes in this study (Diagram 1. b). It should be noted that in this study, 4 variants were found that were identified for the first time. All variants except one variant are in the category of uncertain significance (VUS) (Table 3).

Diagram 1

Table 3: Variants found in patients

The frequency of variants

The highest population MAF for all variants found in patients is less than 0.01. The highest minor allele frequency in the population was obtained from the 1000 Genomes Project Phase 3, ExAC, and gnomAD databases. A "rare" variant has a minor allele frequency of less than 1% in all projects reporting frequencies. Table 4 shows the frequency of each variant in patients with and without SARS-CoV-2 infection.

Table 4: ACE2, TMPRSS2, TYK2, SLC6A20, and IFNAR2 variants identified in a study of 100 individuals

The ACE2 gene analysis

ACE2 polymorphisms were prevalent in the cohort. In the group of individuals with SARS-CoV-2 infection, 28 (51.85%) presented no variant. Seven exonic ACE2 variants were detected in the cohort. The Chi-square test was used to analyze the effect of different ACE2 gene variants on contracting COVID-19. The result of this test shows that there is a significant difference in the frequency of the rs759499720 and rs776459296 variants of the ACE2 gene between people with and without a history of COVID-19 (OR = 7.857, 95% CI: 0.947-94.31; p = 0.034 and OR = 4.714, 95% CI: 1.318-16.971; p = 0.019, respectively). In other words, the rs759499720 and rs776459296 variants of the ACE2 gene has a greater risk of contracting COVID-19.

The TMPRSS2 gene analysis

This gene showed a lower level of polymorphism than ACE2, with 7 patients without SARS-CoV-2 infection (11.87%) and 14 with the infection (32.55%) presenting variants. No difference was observed in the distribution of variants between men and women. Three variants were detected. The Chi-square test was used to analyze the effect of different TMPRSS2 gene variants on contracting COVID-19. The result of this test shows that there is a significant difference in the frequency of the rs386818798 variant of the TMPRSS2 gene between people with and without a history of COVID-19 (OR = 3.587, 95% CI: 10.11-1.135; p = 0.025). In other words, the rs386818798 variant of the TMPRSS2 gene has a higher risk of contracting COVID-19. No significant differences were found between individuals with and without SARS-CoV-2 infection for the remaining TMPRSS2 variants.

The TYK2 gene analysis

The Chi-square test was used to analyze the effect of different TYK2 gene variants on contracting COVID-19. The result of this test shows that there is a significant difference in the frequency of the rs771922681 and rs753470142 variants of the TYK2 gene between people with and without a history of COVID-19 (OR = 5.500, 95% CI: 1.083-26.95; p = 0.024 and OR = 7.857, 95% CI: 0.975-93.97; p = 0.032, respectively). In other words, the rs771922681 and rs753470142 variants of the TYK2 gene has a higher risk of contracting COVID-19. In addition, there is a significant difference in the frequency of the new diagnosed variant (chr19:10365853:C:A) of the TYK2 gene between people with and without a history of COVID-19 (OR = 6.286, 95% CI: 1.364-30.26; p = 0.012). It means the new variant (chr19:10365853:C:A) of the TYK2 gene has a higher risk of contracting COVID-19.

The SLC6A20 gene analysis

The Chi-square test was used to analyze the effect of different SLC6A20 gene variants on contracting COVID-19. The result of this test shows that the frequency of the rs147760034 and rs139008024variants of the SLC6A20 gene is significantly different between people with and without a history of COVID-19 (OR = 6.875, 95% CI: 1.597-32.63; p = 0.007 and OR = 5.347, 95% CI: 1.055-26.19; p = 0.027, respectively). In other words, the rs147760034 and rs139008024 variants of the SLC6A20 gene has a greater risk of contracting COVID-19.

The IFNAR2 gene analysis

The variant with the least P-Value between other was diagnosed in IFNAR2 gene. The Chi-square test was used to analyze the effect of different IFNAR2 gene variants on contracting COVID-19. The result of this test shows that the frequency of the rs759744926 variant of the IFNAR2 gene is significantly different between people with and without a history of COVID-19 (OR = 6.171, 95% CI: 1.713-21.31; p = 0.003). In other words, the rs759744926 variant of the IFNAR2 gene has a higher risk of contracting COVID-19.

In-silico analysis

PolyPhen-2 and SIFT bioinformatics tools determined that the missense variants rs147760034, rs753470142, TYK2: c.675G>T, and rs759744926 lead to damage in the three-dimensional structure of proteins and also

disrupt protein function. And the same goes for continuing bioinformatics tools (Table 5).

Table 5: Results of In-silico analysis of variants

The final prediction of the functional effect of human variants using CADD

Because multiple variant interpretations and scoring tools are available, a widely applicable criterion that accurately and unbiasedly measures and integrates diverse information is needed. A C-score greater than or equal to 10 indicates that these are predicted to be the 10% most deleterious substitutions that can be considered for the human genome, a score greater than or equal to 20 indicates the 1% most deleterious substitutions, and a score greater than or equal to 30 It represents 0.1% of the most destructive substitutions. Table 6 shows that c.884C>T genetic variant of the human IFNAR2 gene is more than 10% of most deleterious substitutions as the most damaging variant for the 3D structure and function of the corresponding protein.

Table 6: The final effect of the variants on the structure and function of the corresponding proteins by CADD

Homology modeling and finding variants of TYK2 residues

The first missense variant of the TYK2 gene is Q225H. This variant changes the amino acid glutamine, which is an uncharged polar amino acid, at position 225 (a total of 1188 amino acids) to the amino acid histidine, which is a positively charged amino acid. Since the amino acid histidine contains an imidazole ring in its side chain; It occupies more three-dimensional space and turns the alpha helix into a beta loop (Figure 1. a). The second missense variant of the TYK2 gene is R465Q. This variant changes the amino acid arginine, which is a positively charged amino acid, at position 465 (a total of 1188 amino acids) to the amino acid glutamine, which is a polar uncharged amino acid. Since the amino acid glutamine contains a shorter side chain than arginine; It occupies less three-dimensional space and converts the beta loop into an alpha helix. In this way, it leads to the extension of the alpha helix and the change of the three-dimensional structure of the protein (Figure 1. b). The third missense variant of the TYK2 gene is R1159S. This variant changes the amino acid arginine, which is a positively charged amino acid, at position 1159 (a total of 1188 amino acids) to the amino acid serine, which is a polar uncharged amino acid. Since the amino acid serine contains a short hydroxymethyl side chain; It occupies less 3D space, resulting in an alpha helix formation in a 3D position close to the remaining variant. Serine is common in many proteins, as seen in the figure below it is present in significant concentrations in the outer regions of soluble proteins due to its hydrophilic nature (Figure 1. c).

Figure 1

Homology modeling and finding variants of SLC6A20 residues

The first missense variant of the SLC6A20 gene is V104I. This variant changes the amino acid valine at position 104 (a total of 593 amino acids) to the amino acid isoleucine, both of them are branched hydrophobic amino acids. This amino acid substitution has shortened both the N-terminal and C-terminal sides of the alpha helix. The isoleucine prefers to be buried in the hydrophobic cores of proteins due to its hydrophobicity. Perhaps the most obvious effect of this is that this amino acid is rarely placed in an alpha helix, although it is easier and even preferred to place in beta sheets. For this reason, the amino end of the alpha helix becomes β turn and the carboxyl end of the alpha helix becomes β loop (Figure 1. d). The second missense variant of the SLC6A20 gene is F249S. This variant changes the amino acid phenylalanine, a hydrophobic amino acid containing an aromatic ring, at position 249 (a total of 593 amino acids) to the amino acid serine, an uncharged polar amino acid. This amino acid substitution has made the C-terminal of the alpha helix longer. Since phenylalanine is hydrophobic, it prefers to be placed in the hydrophobic cores of proteins. The presence of an aromatic side chain can also mean that phenylalanine is involved in interactions with other aromatic side chains. For this reason, replacing phenylalanine 249 with serine has disturbed the interaction of this amino acid with phenylalanine 250, and on the other hand, the relatively reactive hydroxyl group of

serine amino acid has enabled this amino acid to form hydrogen bonds with various polar substrates; As a result, it leads to the transformation of the beta loop into an alpha helix (Figure 1. e).

Homology modeling and finding variants of IFNAR2 residues

The missense variant of the IFNAR2 gene is P295L. This variant replaces the amino acid proline at position 295 (a total of 516 amino acids) with the amino acid leucine, which is a branched hydrophobic amino acid. Proline is unique in that it is the only amino acid whose side chain is attached twice to the protein skeleton, forming a pentagonal nitrogen-containing ring. More precisely, this property makes proline an imino acid. This difference is very important; because proline cannot participate in many of the main chain connections that are easily established by other amino acids. For this reason, proline can often be found in very tight β-turns and β-loops in protein structures (i.e., where the polypeptide chain must change direction). Functionally, proline plays an important role in molecular recognition, especially in intracellular signaling. Similar to what is observed in the IFNAR2 signaling pathway. Domains such as SH3 bind to specific proline-containing peptides that are key parts of many signaling cascades. Leucine contains a very unreactive side chain, so it is rarely directly involved in protein function. For this reason, by replacing proline with an active role in signaling cascades with leucine, the role of IFNAR2 as part of the signaling pathway of the immune system is disrupted (Figure 1. f).

Discussion

Previous studies have explored genetic risk factors associated with susceptibility to SARS-CoV-2 infection and the severity of COVID-19. It has been indicated that the disease is associated with several variants of different genes such as the OAS1, DPP9, CCR2, and ASHG genes[36]. The association between susceptibility to COVID-19 and ACE2 gene variants should be analyzed in various populations. The ACE2 gene is the main functional receptor for SARS-CoV-2 invasion to host cells. The other candidate genes related to COVID-19 susceptibility are TMPRSS2, TYK2, SLC6A20, and IFNAR2. The TMPRSS2 gene codes its protein that mediates the SARS-CoV-2 invasion to cells so that the TMPRSS2 protein functions as a coreceptor for the virus. The TYK2 and IFNAR2 genes are involved in the anti-viral immune system defense. Also, the SLC6A20 gene has a protective role against cytokine storms caused by SARS-CoV-2 infection. Therefore, the variants of mentioned genes may modulate viral infectivity in humans, making some individuals more sensitive to SARS-Cov-2 infection than others. In this study, we have investigated a cohort of 100 Iranian patients with or without a history of COVID-19, to evaluate the association between genetic variants in specific genes and SARS-CoV-2 infection and COVID-19. We have divided this section according to the different genes examined. In this study, using the Whole-Exome Sequencing technique, approximately 145,000 variants were identified in each patient. Finally, among this number, 9 variants predisposed to contracting COVID-19 were reported in 40 patients with a history of contracting the disease.

ACE2 gene

2 variants out of the total of 9 variants predisposing to COVID-19 were located in the ACE2 gene, which lead to the facilitation of SARS-CoV-2 virus infection. Angiotensin-converting enzyme 1 (ACE) and angiotensin-converting enzyme 2 (ACE2) are homologous genes that regulate the physiological homeostasis of the renin-angiotensin system (RAS) [37]. Since SARS-CoV-2 uses the ACE2 receptor to attack alveolar epithelial cells, it is conceivable that the expression level of ACE2 in different tissues can reveal potential genetic susceptibility to COVID-19 infection. The age-dependent expression of the ACE2 gene in the nasal epithelium was investigated in a group of 305 people and it was found that children (less than 10 years old) have the lowest level of ACE2 gene expression [38]. Therefore, it was suggested that the lower risk of developing COVID-19 among children may be due to the lower expression of the ACE2 gene related to age. Two polypyrimidine tract splice variants of the ACE2 gene (rs759499720 delT and rs776459296 insT) were identified, which were significantly more common among SARS-CoV-2 patients than the control population, which indicates that they play a role in the infectivity of the SARS-CoV-2 and the severity of the COVID-19. The genetic variants c.584-8delT and c.584-4dupT are a type of sequence variant located in the polypyrimidine fragment at the 3' end of intron 4. These variants increase ACE2 gene expression, possibly through alternative mRNA

splicing mechanisms[39]. Furthermore, statistically higher allelic diversity has been reported in the group of patients with a history of COVID-19, suggesting that a predisposing genetic background may determine inter-individual differences associated with COVID-19.

TMPRSS2 gene

1 variant out of the total of 9 variants predisposing to COVID-19 was located in the TMPRSS2 gene, which leads to an increase in the penetration of the SARS-CoV-2 virus into the host cell. The TMPRSS2 gene encodes a serine protease enzyme involved in cleavage and activation of the SARS-CoV-2 spike protein during membrane fusion[40]. Current study showed that rs386818798 (GGA vs. AGT) of the TMPRSS2 genotype is a polynucleotide substitution associated with a more aggressive pattern of the disorder. The polymorphism associated with COVID-19 in the present study was rs386818798. which could affect the level of mRNA transcripts and influence the splicing and possibly the generation of one of the 20 TMPRSS2 isoforms reported so far [41]. SRp40 protein is a pre-mRNA splicing factor. It is expressed in all cells, tissues, and organs, including the lung, where airway epithelial cells are found. The GGA allele of rs386818798 can provide a new binding site for SRp40, a splicing protein involved in the generation of protein isoforms (20 mRNA isoforms have been identified). This allele can increase the expression level of TMPRSS2 protein with the subsequent increase of spike protein priming[42]. This fact could facilitate the entry of SARS-CoV-2 into the cell and increase the risk of contracting COVID-19.

TYK2 gene

3 variants out of a total of 9 variants predisposing to COVID-19 were located in the TYK2 gene, which lead to the weakening of anti-viral immunity. The TYK2 gene encodes a member of the intracellular non-receptor tyrosine kinases of the Janus Kinase family, which plays a key role in the immune response against viral infections[43]. This protein is functionally related to the IFNAR1 receptor subunit. This connection has a positive effect on the binding of the ligand to the receptor complex. Therefore, proper TYK2 activity is an important step to initiate the type I IFN response. In this study, SNVs located within the protein-coding region of the tyrosine kinase 2 gene and clinically important include rs771922681, rs753470142, and a new variant that has not been reported so far. The rs771922681 variant, within exon 10 of the total 25 exons of the TYK2 gene, affects the expression level of this gene, which is associated with the virulence of COVID-19 infection. It is a missense variant and leads to a decrease in the expression of the TYK2 gene by the destruction of the mRNA transcription molecule using the nonsense mutation-mediated decay (NMD) mechanism [44]. Another TYK2 variant, rs753470142, is also significantly associated with contracting COVID-19. A known likely pathogenic missense variant of TYK2 reduces its activity by creating another codon at nucleotide 46 of the 436 nucleotides belonging to exon 25. The rs753470142 variant disrupts the function of the leader sequence, thereby causing intracellular accumulation of non-functional TYK2. Another variant, Q225H, which is a novel variant, can disrupt the inclusion of exon 8, which is essential for TYK2 binding to co-receptors involved in signaling pathways [45].

SLC6A20 gene

2 variants out of the total of 9 variants predisposing to COVID-19 were located in the SLC6A20 gene, which lead to the weakening of its protective role against cytokine storms. The protein product of the SLC6A20 gene is sodium-dependent imino transporter 1 (System IMINO transporter (SIT1)), which is involved in the transport of amino acids such as glycine and proline [46]. The main link between SLC6A20 and COVID-19 comes from the functional interaction between SLC6A20 and ACE2 in the membrane of small intestinal enterocytes. The expression levels of SLC6A20 and ACE2 in the lungs are positively correlated, and both genes are mainly expressed in alveolar epithelial type 2 (AT2) cells [47]. SIT1 heterodimerizes with ACE2, which appears to be required to form quaternary structures that can serve as binding sites for SARS-CoV-2 spike glycoproteins. The expression of ACE2 increases the expression level of SIT1, its location in the plasma membrane, and its function in amino acid transport. SLC6A20 is a novel regulator of glycine levels. ACE2 activation increases plasma glycine levels. Glycine has anti-inflammatory, anti-oxidative, neurological, and metabolic regulation functions. By binding to its receptor, GlyR, glycine induces chloride entry into

the cell, which causes hyperpolarization of the cell membrane, which protects cells from pyroptosis and a proinflammatory cytokine storm induced by SARS-CoV-2 infection. Both variants found in this study, rs147760034 and rs139008024, lead to impaired function of SLC6A20 protein and as a result increased plasma concentration of glycine amino acid. Subsequently, the intensity of the cytokine storm does not subside and the severity of COVID-19 is high [48]. In particular, this molecule prevents neurological damage due to its effect on the central nervous system (high expression of SIT1 protein in the brain and spinal cord) and due to its beneficial protective effect against the release of pro-inflammatory cytokines caused by SARS-CoV-2 infection.

IFNAR2 gene

1 variant out of the total of 9 variants predisposing to COVID-19 was located in the IFNAR2 gene, which leads to a defect in the interferon signaling pathway to stimulate T helper cells in creating an immune system response. Type I IFN binds to a receptor complex consisting of interferon alpha and beta receptors 1 and 2 (IFNAR1 and IFNAR2, respectively) [49], which are associated with Janus kinases, TYK2 and JAK1, respectively. The activation of these kinases causes the tyrosine phosphorylation of STAT1 and STAT2, which leads to the formation of a heterotrimer with IRF-family member 9 (IRF-9) and thus organizes the interferon-stimulated gene transcription factor (ISGF3) complex. This signaling pathway activates a self-reinforcing positive feedback loop that leads to the induction of the production of high and protective amounts of IFN-α. The P295L mutant may destabilize critical binding to the IFN molecule in response to COVID-19[50]. Therefore, P295L is likely to have a major effect in disrupting the structure of the IFNAR2 protein. The rs759744926 variant is a missense variant that affects receptor structure and function and limits the antiviral effects of IFN-α/β.

Strengths and limitations of the study

This study has been conducted for the first time worldwide with 5 various genes related to the infectivity of SARS-CoV-2. This study, in addition to investigating the effect of variants on contracting COVID-19, the effect of found variants on the amino acid sequence of the related proteins, and the structure and function of the related proteins have been measured. This research confirms previous studies on the importance of genetic factors predisposing to COVID-19 in ACE2, TMPRSS2, TYK2, SLC6A20, and IFNAR2 genes. Some limitations of the study are that first of all this study only focused on genetic factors, so all other factors such as epigenetic factors, comorbidities, and lifestyle were not considered. Second, for some polymorphisms that did not show any association with contracting COVID-19, we cannot rule out the possibility that they could be associated with disease severity or death, perhaps larger sample sizes are needed to demonstrate this association.

Conclusion

This study aims to find potential genetic factors that may be associated with severe outcomes of COVID-19; It was conducted on individuals with a history of COVID-19 in the Iranian population. The present study shows that the rs759499720/ACE2, rs776459296/ACE2, rs386818798/TMPRSS2, rs771922681/TYK2, rs753470142/TYK2, c.675G>T/TYK2, rs147760034/SLC6A20, rs139008024/SLC6A20 and rs759744926/IFNAR2 variants are genetic risk factors for severe forms of COVID-19. However, further research with a larger sample size and different human populations is needed to confirm the findings of this study. Also, more studies are necessary to find the association between these found variants and gender and different ethnicities.

Acknowledgments

The authors would like to acknowledge the participating patients for the information and the DNA samples provided.

Conflict of interests

The authors declare no conflicts of interest in this work.

Author contributions

All authors listed on the title page contributed equally to the full process of designing, planning, and conducting the study and are fully aware of the submission. All authors read and approved the final manuscript.

Ethics approval

All processes were under the ethical standards of the ethics committee on human subject research at Kerman University of Medical Sciences (code of "IR.KMU.AH.REC.1401.215").

Consent to participate

Informed written consent had been received from all participants. The ethics committee of the Kerman University of Medical Sciences approved the study.

Consent for publication

As the corresponding author, I verify that the manuscript has been read and authorized for submission by all the named authors. We affirm that this manuscript is original and has not been published ago in another journal, and is not presently being regarded for publication elsewhere.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Seyedeh Sepideh Aghamirli https://orcid.org/0009-0009-6532-1319

Nasrollah Saleh-Gohari https://orcid.org/0000-0003-1256-0323

References

1. Yang X, Yu Y, Xu J, et al. 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* . 2020;8(5):475-481. doi:10.1016/S2213-2600(20)30079-5

2. Rello J, Belliato M, Dimopoulos M-A, et al. Update in COVID-19 in the intensive care unit from the 2020 HELLENIC Athens International symposium. *Anaesthesia, Crit Care Pain Med* . 2020;39(6):723. doi:10.1016/J.ACCPM.2020.10.008

3. Weiss SR, Navas-Martin S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol Mol Biol Rev* . 2005;69(4):635-664. doi:10.1128/MMBR.69.4.635-664.2005

4. S U, P S, AV R, et al. Origin, transmission, diagnosis and management of coronavirus disease 2019 (COVID-19). *Postgrad Med J* . 2020;96(1142):753-758. doi:10.1136/POSTGRADMEDJ-2020-138234

5. Heymann DL, Shindo N. COVID-19: what is next for public health?*Lancet (London, England)* . 2020;395(10224):542-545. doi:10.1016/S0140-6736(20)30374-3

6. Ge H, Wang X, Yuan X, et al. The epidemiology and clinical information about COVID-19. *Eur J Clin Microbiol Infect Dis* . 2020;39(6):1011-1019. doi:10.1007/S10096-020-03874-Z

7. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet (London, England)* . 2020;395(10223):497-506. doi:10.1016/S0140-6736(20)30183-5

8. Fisher D, Heymann D. Q&A: The novel coronavirus outbreak causing COVID-19. *BMC Med* . 2020;18(1). doi:10.1186/S12916-020-01533-W

9. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nat 2020 5797798* . 2020;579(7798):265-269. doi:10.1038/s41586-020-2008-3

10. Sahu P. Closure of Universities Due to Coronavirus Disease 2019 (COVID-19): Impact on Education and Mental Health of Students and Academic Staff. *Cureus* . 2020;12(4). doi:10.7759/CUREUS.7541

11. Sanyaolu A, Okorie C, Hosein Z, et al. Global Pandemicity of COVID-19: Situation Report as of June 9, 2020. *Infect Dis (Auckl)* . 2021;14:1178633721991260. doi:10.1177/1178633721991260

12. Hua W, Xiaofeng L, Zhenqiang B, Jun R, Ban W, Liming L. [The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China]. *Zhonghua Liu Xing Bing Xue Za Zhi* . 2020;41(2):297-300. doi:10.3760/CMA.J.ISSN.0254-6450.2020.02.003

13. C W, X C, Y C, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med* . 2020;180(7):934-943. doi:10.1001/JAMAINTERNMED.2020.0994

14. MC P, S S, P D, et al. COVID-19-related Genes in Sputum Cells in Asthma. Relationship to Demographic Features and Corticosteroids.*Am J Respir Crit Care Med* . 2020;202(1):83-90. doi:10.1164/RCCM.202003-0821OC

15. Chong WP, Ip WE, Tso GHW, et al. The interferon gamma gene polymorphism +874 A/T is associated with severe acute respiratory syndrome. *BMC Infect Dis 2006 61* . 2006;6(1):1-4. doi:10.1186/1471-2334-6-82

16. Yuan FF, Boehm I, Chan PKS, et al. High Prevalence of the CD14-159CC Genotype in Patients Infected with Severe Acute Respiratory Syndrome-Associated Coronavirus. *Clin Vaccine Immunol* . 2007;14(12):1644. doi:10.1128/CVI.00100-07

17. Torre-Fuentes L, Matías-Guiu J, Hernández-Lorenzo L, et al. ACE2, TMPRSS2, and Furin variants and SARS-CoV-2 infection in Madrid, Spain.*J Med Virol* . 2021;93(2):863-869. doi:10.1002/JMV.26319

18. R Y, Y Z, Y L, L X, Y G, Q Z. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* . 2020;367(6485):1444-1448. doi:10.1126/SCIENCE.ABB2762

19. RAS S, WO S, AC A, et al. The ACE2/Angiotensin-(1-7)/MAS Axis of the Renin-Angiotensin System: Focus on Angiotensin-(1-7). *Physiol Rev* . 2018;98(1):505-553. doi:10.1152/PHYSREV.00023.2016

20. K K, Y I, JM P. Angiotensin-converting enzyme 2 in lung diseases.*Curr Opin Pharmacol* . 2006;6(3):271-276. doi:10.1016/J.COPH.2006.03.001

21. K K, B S, B Z, et al. Functional prediction and comparative population analysis of variants in genes for proteases and innate immunity related to SARS-CoV-2 infection. *Infect Genet Evol* . 2020;84. doi:10.1016/J.MEEGID.2020.104498

22. Ortiz-Fernández L, López-Mejias R, Carmona FD, et al. The role of a functional variant of TYK2 in vasculitides and infections. *Clin Exp Rheumatol* . 2020;38(5):949-955.

23. Semiz S. SIT1 transporter as a potential novel target in treatment of COVID-19. *Biomol Concepts* . 2021;12(1):156-163. doi:10.1515/BMC-2021-0017

24. Hurtado-Guerrero I, Hernáez B, Pinto-Medel MJ, et al. Antiviral, Immunomodulatory and Antiproliferative Activities of Recombinant Soluble IFNAR2 without IFN-ß Mediation. *J Clin Med* . 2020;9(4). doi:10.3390/JCM9040959

25. L L, D X, G Y, et al. Positive RT-PCR Test Results in Patients Recovered From COVID-19. *JAMA* . 2020;323(15):1502-1503. doi:10.1001/JAMA.2020.2783

26. Zhang JC, Wang S Bin, Xue YD. Fecal specimen diagnosis 2019 novel coronavirus-infected pneumonia. *J Med Virol* . 2020;92(6):680-682. doi:10.1002/JMV.25742

27. Shi H, Han X, Jiang N, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study.*Lancet Infect Dis* . 2020;20(4):425-434. doi:10.1016/S1473-3099(20)30086-4

28. K X, H C, Y S, et al. [Management of corona virus disease-19 (COVID-19): the Zhejiang experience]. *Zhejiang Da Xue Xue Bao Yi Xue Ban* . 2020;49(1):147-157. doi:10.3785/J.ISSN.1008-9292.2020.02.02

29. O M, B C. Use of antiviral drugs to reduce COVID-19 transmission.*Lancet Glob Heal* . 2020;8(5):e639-e640. doi:10.1016/S2214-109X(20)30114-5

30. EJ G-B, MG N, N R, et al. Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. *Cell Host Microbe* . 2020;27(6):992-1000.e3. doi:10.1016/J.CHOM.2020.04.009

31. van Doremalen N, Miazgowicz KL, Milne-Price S, et al. Host species restriction of Middle East respiratory syndrome coronavirus through its receptor, dipeptidyl peptidase 4. *J Virol* . 2014;88(16):9220-9232. doi:10.1128/JVI.00676-14

32. SF A, AA Q, MR M. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses* . 2020;12(3). doi:10.3390/V12030254

33. Alekseyev YO, Fazeli R, Yang S, et al. A Next-Generation Sequencing Primer-How Does It Work and What Can It Do? *Acad Pathol* . 2018;5. doi:10.1177/2374289518766521

34. Schulert GS, Blum SA, Cron RQ. Host genetics of pediatric SARS-CoV-2 COVID-19 and multisystem inflammatory syndrome in children. *Curr Opin Pediatr* . 2021;33(6):549-555. doi:10.1097/MOP.0000000000001061

35. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* . 2013;369(16):1502-1511. doi:10.1056/NEJMOA1306555

36. Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* . 2009;19(5):596-604. doi:10.1016/J.SBI.2009.08.003

37. Donoghue M, Hsieh F, Baronas E, et al. A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circ Res* . 2000;87(5). doi:10.1161/01.RES.87.5.E1

38. Verdecchia P, Cavallini C, Spanevello A, Angeli F. The pivotal link between ACE2 deficiency and SARS-CoV-2 infection. *Eur J Intern Med* . 2020;76:14-20. doi:10.1016/J.EJIM.2020.04.037

39. Chen YY, Zhang P, Zhou XM, et al. Relationship between genetic variants of ACE2 gene and circulating levels of ACE2 and its metabolites. *J Clin Pharm Ther* . 2018;43(2):189-195. doi:10.1111/JCPT.12625

40. Stopsack KH, Mucci LA, Antonarakis ES, Nelson PS, Kantoff PW. TMPRSS2 and COVID-19: Serendipity or Opportunity for Intervention?*Cancer Discov* . 2020;10(6):779-782. doi:10.1158/2159-8290.CD-20-0451

41. Strope JD, PharmD CHC, Figg WD. TMPRSS2: Potential Biomarker for COVID-19 Outcomes. *J Clin Pharmacol* . 2020;60(7):801-807. doi:10.1002/JCPH.1641

42. Paniri A, Hosseini MM, Akhavan-Niaki H. First comprehensive computational analysis of functional consequences of TMPRSS2 SNPs in susceptibility to SARS-CoV-2 among different populations. *J Biomol Struct Dyn* . 2021;39(10):1-18. doi:10.1080/07391102.2020.1767690

43. Strobl B, Stoiber D, Sexl V, Mueller M. Tyrosine kinase 2 (TYK2) in cytokine signalling and host immunity. *Front Biosci (Landmark Ed* . 2011;16(9):3224-3232. doi:10.2741/3908

44. Akbari M, Akhavan-Bahabadi M, Shafigh N, et al. Expression analysis of IFNAR1 and TYK2 transcripts in COVID-19 patients. *Cytokine* . 2022;153. doi:10.1016/J.CYTO.2022.155849

45. Dendrou CA, Cortes A, Shipman L, et al. Resolving TYK2 locus genotype-to-phenotype differences in autoimmunity. *Sci Transl Med* . 2016;8(363). doi:10.1126/SCITRANSLMED.AAG1974

46. Yao Y, Ye F, Li K, et al. Genome and epigenome editing identify CCR9 and SLC6A20 as target genes at the 3p21.31 locus associated with severe COVID-19. *Signal Transduct Target Ther* . 2021;6(1). doi:10.1038/S41392-021-00519-1

47. Bae M, Roh JD, Kim Y, et al. SLC6A20 transporter: a novel regulator of brain glycine homeostasis and NMDAR function. *EMBO Mol Med* . 2021;13(2). doi:10.15252/EMMM.202012632

48. Takanaga H, Mackenzie B, Suzuki Y, Hediger MA. Identification of mammalian proline transporter SIT1 (SLC6A20) with characteristics of classical system imino. *J Biol Chem* . 2005;280(10):8974-8984. doi:10.1074/JBC.M413027200

49. Fricke-Galindo I, Martínez-Morales A, Chávez-Galán L, et al. IFNAR2 relevance in the clinical outcome of individuals with severe COVID-19.*Front Immunol* . 2022;13. doi:10.3389/FIMMU.2022.949413

50. Smieszek SP, Polymeropoulos VM, Xiao C, Polymeropoulos CM, Polymeropoulos MH. Loss-of-function mutations in IFNAR2 in COVID-19 severe infection susceptibility. *J Glob Antimicrob Resist* . 2021;26:239-240. doi:10.1016/J.JGAR.2021.06.005

Supporting information

Additional supporting information may be found online in the Supporting Information section.

Hosted file

`Figure.docx` available at https://authorea.com/users/643337/articles/656751-ace2-tmprss2-tyk2-slc6a20-and-ifnar2-human-genes-variants-influence-sars-cov-2-infection-susceptibility

Hosted file

`Table.docx` available at https://authorea.com/users/643337/articles/656751-ace2-tmprss2-tyk2-slc6a20-and-ifnar2-human-genes-variants-influence-sars-cov-2-infection-susceptibility