DOI: 10.26650/BB/CH22.2020.008.07

7. BÖLÜM / CHAPTER 7

Inter-Populations Comparative Analysis of the *ACE2*, *TMPRSS2*, *CTSB* and *CTSL* Gene Variants Identified in Turkish Individuals

Türk Bireylerde Tanımlanan *ACE2, TMPRSS2, CTSB* ve *CTSL* Gen Varyantlarının Populasyonlar Arası Karşılaştırmalı Analizi

Evrim Kömürcü-Bayrak[1], Güven Toksoy[1], Oya Uyguner[1], Birsen Karaman[1,2], Seher Başaran[1]

[1]Department of Medical Genetics, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey
[2]Departments of Basic Pediatric Sciences, Child Health Institute, Istanbul University, Istanbul, Turkey
e-mail: ebayrak@istanbul.edu.tr
ORCID: E.KB. 0000-0003-1271-1208; G.T. 0000-0002-8103-9980; O.U. 0000-0002-2035-4338; B.K. 0000-0001-8640-0176; S.B. 0000-0003-1271-120

ABSTRACT

Objective: Recent research studies have suggested that susceptibility to SARS-CoV-2 infection might be associated with genetic variants of host machinery for viral entry. The allele frequencies of *ACE2, TMPRSS2, CTSB,* and *CTSL* gene variants defined as viral entry machinery have significant differences among population groups, but there is no information in the Turkish population. To provide a basis for future studies on susceptibility to SARS-CoV-2 infection, these gene variants were screened in a group of Turkish population.
Material and Method: We analysed 138 unrelated non-anonymous individuals for single nucleotide variants (SNVs) of the exonic and flanking intronic regions in *ACE2, TMPRSS2, CTSB,* and *CTSL* genes using in-house whole exome sequencing data. These defined variants were also compared with other population data sets (gnomAD).
Results: For the first time in Turkish individuals, the frequencies of the alternate alleles in these genes were identified in this preliminary analysis study. As a result, three novel SNVs were identified in these genes and it was found that the allele frequencies of some variants were different among the population groups.
Conclusion: Whether these variants, if any, contribute to the transmissibility of viral infection, they require to be screened in affected/unaffected study groups and evaluated with functional studies for significant variants to elucidate the roles of these variants in SARS-CoV-2 infection.
Keywords: SARS-CoV-2, viral entry machinery, genetic variants

ÖZ

Amaç: Son araştırmalar, SARS-CoV-2 enfeksiyonuna yatkınlıkta viral girişin konakçı sistemlerin genetik varyantları ile ilişkili olabileceğini düşündürmektedir. Viral giriş sistemleri (makineleri) olarak tanımlanmış *ACE2, TMPRSS2, CTSB* ve *CTSL* genlerindeki varyantların allel frekansları, populasyon grupları arasında önemli farklılıklara sahiptir, ancak Türk popülasyonunda bu konuda herhangi bir bilgi mevcut değildir. SARS-CoV-2 enfeksiyonuna duyarlılık konusunda gelecekteki çalışmalar için bir temel sağlamak amacıyla, Türk bireylerde bu gen varyantları araştırılmıştır.
Gereç ve Yöntem: *ACE2, TMPRSS2, CTSB* ve *CTSL* genlerindeki ekzonik ve komşu intronik bölgelerin tek nükleotid varyantları (SNV'ler) için kurum içi tüm ekzom sekanslama verileri kullanılarak ilişkisiz 138 kişi analiz edilmiştir. Ayrıca tanımlanmış varyantlar, diğer populasyon veri kümeleriyle (gnomAD) karşılaştırılmıştır.
Bulgular: Bu ön çalışmada, ilk kez Türk bireylerin varyant analizleri yapılarak, bu genlerdeki alternatif alellerin frekansları belirlenmiştir. Sonuç olarak, bu genlerde üç yeni SNV tanımlanmış ve bazı varyantların allel frekanslarının populasyon grupları arasında farklı olduğu bulunmuştur.
Sonuç: Gelecekte SARS-CoV-2 enfeksiyonunda fonksiyonel çalışmalarla bu varyantların rollerinin belirlenmesi, ayrıca bu varyantların viral enfeksiyonun bulaşabilirliğine katkıda bulunup bulunamayacağı, etkilenmiş/etkilenmemiş çalışma gruplarında araştırılması önem taşımaktadır.
Anahtar Kelimeler: SARS-CoV-2, viral giriş makineleri, genetik varyantlar

INTRODUCTION

Since the beginning of the 21st century, public health has been combating contagious diseases such as Severe Acute Respiratory Syndrome (1,2), Middle East Respiratory Syndrome (3) and the latest COVID-19 (4) caused by coronaviruses (SARS-CoV, MERS-CoV and SARS-CoV-2, respectively). Susceptibility to SARS-CoV has been shown to be associated with genetic variants in angiotensin converting enzyme 2 receptors, the main host machine for viral entry among different animals (5-7).

In the human SARS-CoV/CoV-2 infection, angiotensin converting enzyme 2 (ACE2) receptor (8,9) for entry into cells as well as endosomal cysteine protease cathepsin B (CTSB) and cathepsin L (CTSL) (9,10) and transmembrane serine protease 2 (TMPRSS2) (9,11,12) for viral spike (S) protein priming (9) were determined as host machineries of the viral entry. In vitro studies, it has been shown that endosomal cysteine proteases cathepsin B/L activity with proteases inhibitors blockades the entry of SARS-CoV into the cell (13) but does not completely prevent the entry of SARS-CoV-2 (9). This comparative in vitro study reveals important commonalities between SARS-CoV-2 and SARS-CoV infections and states that especially TMPRSS2 inhibitors might be a potential target for antiviral intervention (13). Recent research studies have suggested that susceptibility to the SARS-CoV-2 infection might be associated with genetic variants of viral entry machineries, and some specific *ACE2* gene variants are important for viral infection (13-16). Also, the allele frequencies of these variants have been shown to differ significantly among population groups (14-17). On the other hand, it has been suggested that some genetic variants may have effects on *ACE2* and proteases (*TMPRSS2, CTSB and CTSL*) gene expression levels and change the severity of infection (14-17).

In this preliminary study, it was aimed to identify the alternate allele frequencies of 138 non-anonymous unrelated individuals from Turkey for which may provide important variants of *ACE2, TMPRSS2, CTSB,* and *CTSL* genes for future studies related to human coronaviruses infections.

MATERIAL AND METHODS

The participants of the study

The study included 138 unrelated non-anonymous individuals with whole exome sequencing data. The informed consent was obtained from all of the participants. The study was conducted in accordance with the ethical standards of the Ethics Committee of the University of Istanbul, Istanbul Faculty of Medicine and with the Helsinki Declaration (1964). The variants were screened in our in-house exome sequencing data set called MGexome (Medical Genetics Exome).

In-House Whole-Exome Sequencing Data Sets

Whole-exome sequencing was performed on peripheral blood DNA from 138 unrelated non-anonymous individuals, in MGexome. The variant analyses were performed on files in variant call format (vcf) obtained from sequencing data sets using the Illumina HiSeq and S5 Ion Torrent platforms. Illumina and IonTorrent sequence analysis was performed using the DNAscan and IonReporter analysis pipelines, respectively. With a sequencing run yield of approximately 10 Gb, the samples achieved 97% of the targeted exome bases covering a depth of at least 20x.

Variant Detection and Analysis

The single nucleotide variants (SNVs) of the exonic and flanking intronic regions in *ACE2, TMPRSS2, CTSB,* and *CTSL* genes were analysed in MGexome data. The variants with known pathogenicity

in *ACE2* (NM_021804), *TMPRSS2* (NM_001135099), *CTSB* (NM_001908) and *CTSL* (NM_001257972) genes of viral entry machineries were listed using VarSome (UniProt, ClinVar, and PubMed), dbSNP (https://www.ncbi.nlm.nih.gov/snp/), HGMD Professional (Human Genome Mutation Database; BIOBASE, https://portal.biobase-international.com/cgi-bin/portal/login.cgi), and classified according to American College of Medical Genetics and Genomics-ACMG Standards (19) for interpretation of their status. The pathogenicity and conservation scores of variants were checked in VarSome (https://varsome.com/) using *in silico* tools such as MutationTaster, DANN, SIFT, PROVEAN and GERP (Genomic Evolutionary Rate Profiling). The alternate allele frequencies of defined variants in our in-house data set were compared with African/African American, Ashkenazi Jewish, East Asian (Koreans and Japanese), European (Finnish, Bulgarian, Estonian, North-Western European, Southern European and Swedish), Latino/American Admixed, and South Asian populations in public Genome Aggregation Database (gnomAD, https://gnomad.broadinstitute.org/). The data set of gnomAD provides 123,136 exome sequences and 15,496 whole-genome sequences from unrelated individuals sequenced as part of various disease-specific and population genetic studies.

RESULTS

Chromosomal location of genes required for viral entry, number of SNPs (single nucleotide polymorphisms) and clinical importance of SNVs in *ACE2, TMPRSS2, CTSB,* and *CTSL* genes are summarized in Table 1. Although there are thousands of single nucleotide polymorphisms (SNPs) in these genes in the dbSNP database, there are few SNVs with known pathogenicity classifications according to UniProt, ClinVar, VarSome and PubMed data sets. SNVs associated with various phenotypes in HGMD Professional are shown in Table 1.

The allele frequencies compared to seven different populations and ACMG classifications of *ACE2* variants detected in the Turkish population were listed in Table 2. We determined nine SNVs in our in-house whole exome sequencing data set. For the first time in this study, the novel Leu120Pro variant, one of these nine variants, was identified in one of 138 individuals. This missense variant was conserved among genomes of 35 mammals (GERP score; 5.8) and was determined as disease-causing *in silico* tools (MutationTaster, high DANN score, PROVEAN and SIFT). This missense variant was defined as an uncertain significance variant according to the ACMG classification, such as the His505Arg variant (rs1016409802), a rare variant previously described in dbSNP databank.

We determined 28 different SNVs in *TMPRSS2* gene, 28 different SNVs in *CTSB* gene and 8 different SNVs in *CTSL* gene in our in-house whole-exome sequencing data set. The allele frequencies compared to other populations and ACMG classifications of *TMPRSS2, CTSB* and *CTSL* gene variants were listed in Table 3, Table 4 and Table 5, respectively. In this study, it was found that there were differences in allele frequencies of Gly8Arg, Gly8Val, Thr112Ile, Asp158Asn and Val197Met variants of *TMPRSS2* gene compared to other population groups (Table 3). In addition, two novel variants called c.111G>T (p.Arg37=) in *CTSB* gene and c.126+19G>C in *CTSL* gene were identified in two different individuals. This novel synonymous variant in *CTSB* gene was not conserved among genomes of 35 mammals (GERP score; 5.16), and had a low DANN score (0.73), but it was determined as disease-causing according to the Mutation Taster tool. In addition, the novel intronic variant determined in *CTSL* gene had a lower pathogenicity score (DANN score; 0.26) and also conservation score (GERP score; 3.66). This variant was also predicted as polymorphism according to the Mutation Taster tool.

Table 1. The general information of *ACE2*, *TMPRSS2*, *CTSB*, and *CTSL* genes and single nucleotide variants (SNVs)

Gene	Description	Cytobands	Numbers of RefSNPs in dbSNP	Number of SNVs with known pathogenicity*						HGMD	Reported Phenotype
				Coding impact	Pathogenic	Likely Pathogenic	Uncertain Significance	Likely Benign	Benign		
ACE2	angiotensin I converting enzyme 2, 805 amino acids, 19 exons	Xp22.2	7480	Synonymous	0	0	0	0	1	0	-
				Missense	0	0	2	1	1	2 (Ala412Glu, Ala627Val)	Autism spectrum disorder, West Syndrome
				Non-coding	0	0	0	1	0	1 (c.1896+64A>T)	Blood pressure salt sensitivity
TMPRSS2	transmembrane protease, serine 2, 529 amino acids, 14 exons	21q22.3	14145	Synonymous	0	0	1	0	4	0	-
				Missense	0	0	0	0	3	2 (Val294Met, Ala423Thr)	Left ventricular obstruction, Autism spectrum disorder
				Frameshift	0	0	1	0	0	1 (c.13827C>T)	Reduced androgen receptor binding
CTSB	cathepsin B, 339 amino acids, 10 exons	8p23.1	14120	Synonymous	0	0	0	0	4	0	-
				Missense	0	0	0	1	3	3 (Trp2Arg, Leu26Val, Gln334Pro)	Developmental disorder, Pancreatitis
				Non-coding	0	0	0	1	0	0	-
CTSL	cathepsin L, 333 amino acids, 8 exons	9q21.33	1709	Synonymous	0	0	0	1	0	0	-
				Missense	0	0	0	0	1	0	-
				Non-coding	0	0	0	0	0	c.-461C>A	Hypertension

*Total classified variants (UniProt, ClinVar, VarSome and PubMed); HGMD; Human Genome Mutation Database.

Table 2. The alternate allele frequencies of *ACE2* gene variants in Turkish and other populations and their clinical definitions according to ACMG classification (18)

	Allele Frequencies in gnomAD data set, n (allele number)									In-House data set				
Variant Description	**African, n=712**	**Ashkenazi Jewish, n=4158**	**East Asian, n=1746**	**European (Finnish), n=3116**	**European (Non-Finnish), n=17568**	**Latino, n=11012**	**South Asian, n=10288**	**Other, n=1552**	**Total, n=50152**	**MGexome, n=276**	**dbSNP ID**	**Position of the variant on the protein**	**ACMG Definition**	**ACMG Verdict Rules**
c.359T>C	-	-	-	-	-	-	-	-	-	0.00362	novel	p.Leu120Pro	VUS	PM2, PP3
c.439+4G>A	**0.221**	**0.258**	**0.547**	**0.191**	**0.2**	**0.394**	**0.453**	**0.254**	**0.28**	**0.174**	rs2285666		Benign	BA1, BP4
c.584-71A>G	**0.767**	**0.634**	**0.998**	**0.67**	**0.628**	**0.788**	-	**0.709**	**0.695**	**0.725**	rs971249		VUS	BP4
c.802+24G>A	0.124	-	0.0246	0.000395	0.000442	0.00753	**0.0619**	0.0106	**0.0179**	0.00362	rs4646140		Benign	BA1, BP4
c.1164A>G	-	-	-	-	**0.0000245**	-	0.000524	-	**0.0000656**	0.00362	rs773676270	p.Gln388=	Likely Benign	BS1, BP4, BP7
c.1514A>G	-	-	**0.0000956**	-	-	-	-	-	**0.00000655**	0.00362	rs1016409802	p.His505Arg	VUS	PM2, PP3
c.2070T>C	**0.0649**	0.00576	-	-	0.000355	0.00319	0.000158	0.00245	**0.00561**	0.00362	rs4646179	p.Asn690=	Benign	BA1, BP4, BP6, BP7
c.2114+4A>G	-	-	-	-	**0.0000248**	-	-	-	**0.0000111**	0.00725	rs371381538		VUS	PM2
c.2247G>A	0.00705	0.103	-	0.0165	0.0454	**0.0774**	0.00348	0.0491	**0.0396**	0.0254	rs35803318	p.Val749=	Benign	BA1, BP4, BP7

Green: AF>0.05; red: AF<0.0001; bold: AF of total gnomAD populations; VUS: Uncertain Significance; PM2: absent from control databanks; BA1: allele frequency is greater than 5%; PP3: multiple lines of computational evidence support a deleterious effect; BP4: impact on gene in computational analysis; BP6: in recently reports variant as benign; BP7: a silent variant. The abbreviations is according to ACMG guideline (18).

Table 3. The alternate allele frequencies of *TMPRSS2* gene variants in Turkish and other populations and their clinical definitions according to ACMG classification (18).

Variant Description	Allele Frequencies in gnomAD data set, n (allele number)									In-House data set	dbSNP ID	Position of the variant on the protein	ACMG Definiation	ACMG Verdict Rules
	African, n=712	Ashkenazi Jewish, n=4158	East Asian, n=1746	European (Finnish), n=3116	European (Non-Finnish), n=17568	Latino, n=11012	South Asian, n=10288	Other, n=1552	Total, n=50152	MGexome, n=276				
c.-50G>A	-	-	-	-	-	-	-	-	-	0.00362	rs1601595829		VUS	PM2, BP4
c.22G>C	0.0014	0.00577	-	0.00193	0.0109	0.00309	0.000389	0.00644	**0.00542**	0.00362	rs200291871	p.Gly8Arg	Likely Benign	BS1, BP4
c.23G>T	**0.334**	**0.383**	0.0119	**0.419**	**0.431**	**0.275**	**0.269**	**0.379**	**0.342**	**0.152**	rs75603675	p.Gly8Val	Benign	BA1, BP4
c.55+153C>T	0.0402	**0.234**	**0.0533**	**0.129**	**0.177**	**0.182**	-	**0.163**	**0.128**	0.0326	rs8126497		Benign	BA1, BP4
c.300C>T	0.00291	0.00734	**0.0000551**	0.0181	0.0141	0.0106	0.00294	0.0155	**0.0106**	0.0145	rs61735792	p.Pro100=	Likely Benign	BS1, BP4, BP7
c.335C>T	0.00142	0.00812	-	0.0108	0.0101	0.00301	0.00406	0.00761	**0.00704**	0.0217	rs61735793	p.Thr112Ile	Likely Benign	BS1, BP4, BP7
c.336A>G	**0.202**	**0.152**	**0.147**	0.0296	**0.0775**	0.0478	**0.191**	**0.102**	**0.0997**	**0.167**	rs3787950	p.Thr112=	Benign	BA1, BP4, BP7
c.350-25C>T	0.00174	0.00625	-	0.0193	0.0142	0.00375	0.00242	0.0109	**0.00952**	0.0109	rs144948620		Likely Benign	BS1, BP4
c.437-48T>C	0.000787	0.014	-	**0.0000583**	0.00125	0.00229	0.00134	0.00379	**0.00183**	0.00362	rs187290362		Likely Benign	BS1, BP4
c.472G>A	-	-	-	-	**0.0000265**	-	**0.0000654**	-	**0.0000199**	0.00362	rs199865751	p.Asp158Asn	VUS	BP4
c.556+14G>A	**0.549**	**0.811**	**0.287**	**0.546**	**0.709**	**0.489**	**0.645**	**0.664**	**0.619**	**0.710**	rs422471		Benign	BA1, BP4
c.589G>A	**0.294**	**0.136**	**0.381**	**0.371**	**0.229**	**0.153**	**0.248**	**0.224**	**0.245**	**0.203**	rs12329760	p.Val197Met	Benign	PP3, BA1
c.651C>T	0.00229	0.0161	-	0.00591	0.0139	0.00899	0.000653	0.0136	**0.00926**	0.0181	rs61735789	p.Tyr217=	Likely Benign	BS1, BP4, BP6, BP7
c.683+83G>T	**0.241**	**0.552**	0.00513	**0.355**	**0.491**	**0.309**	-	**0.387**	**0.374**	**0.315**	rs734056		Benign	BA1, BP4
c.683+92C>T	0.00758	0.0276	-	**0.105**	**0.0584**	0.0307	-	**0.0763**	**0.0462**	0.0181	rs34561135		Benign	BA1, BP4
c.759C>T	0.00209	0.0115	0.00277	0.0125	0.0031	0.00228	0.000489	0.00619	**0.0038**	0.00725	rs141788162	p.Ala253=	Likely Benign	BS1, BP4, BP6, BP7
c.795-32C>T	0.00445	0.00788	**0.0000544**	0.0368	0.0285	0.00622	0.00729	0.0172	**0.0188**	0.0109	rs74423429		Likely Benign	BS1, BP4
c.838+47G>A	0.000743	0.00189	**0.0000544**	0.00515	0.00636	0.00182	0.00654	0.00492	**0.00461**	0.0109	rs75756279		Likely Benign	BS1, BP4
c.879T>C	**0.379**	**0.583**	**0.129**	**0.522**	**0.55**	**0.389**	**0.471**	**0.489**	**0.473**	**0.457**	rs17854725	p.Ile293=	Benign	BA1, BP4, BP7
c.888C>T	**0.192**	**0.103**	**0.264**	**0.363**	**0.231**	**0.257**	**0.211**	**0.221**	**0.238**	**0.156**	rs2298659	p.Gly296=	Benign	BA1, BP4, BP7
c.1010+36A>C	**0.116**	0.00463	**0.0000758**	0.000127	0.00261	0.0103	0.00621	0.00919	**0.0103**	0.0109	rs113288437		Benign	BA1, BP4
c.1010+45C>T	**0.108**	0.00422	0.000163	0.000136	0.00261	0.00947	0.00611	0.00913	**0.00982**	0.0109	rs75168613		Benign	BA1, BP4
c.1010+85C>G	0.00472	**0.0621**	-	0.0449	0.0337	0.0259	-	0.0404	**0.0255**	0.00725	rs73230068		Likely Benign	BS1, BP4
c.1187-101G>C	**0.31**	**0.4**	**0.219**	**0.252**	**0.278**	**0.222**	-	**0.301**	**0.282**	0.0362	rs28524972		Benign	BA1, BP4
c.1187-47A>T	0.00129	-	0.00109	0.0217	-	0.000441	-	-	**0.000784**	0.00362	rs777860329		VUS	PM2, BP4
c.1266G>A	0.00477	0.0213	**0.0000544**	0.0289	0.0291	0.0128	0.0111	0.0245	**0.0205**	0.0254	rs61735794	p.Gly422=	Likely Benign	BS1, BP4, BP7
c.1579-58T>A	**0.0691**	0.0414	**0.0571**	**0.236**	**0.158**	**0.0932**	-	**0.159**	**0.134**	0.00362	rs73905370		Benign	BA1, BP4
c.1579-24C>G	0.000186	-	**0.0916**	-	0.000108	0.000291	0.00122	0.00344	**0.00708**	0.00362	rs118028230		Benign	BA1, BP4

Green: AF>0.05; red: AF<0.0001; bold: AF of total gnomAD populations; VUS: Uncertain Significance; PM2: absent from control databanks; BA1: allele frequency is greater than 5%; PP3: multiple lines of computational evidence support a deleterious effect; BP4: impact on gene in computational analysis, BP6: in recently reports variant as benign; BP7: a silent variant; BS1: allele frequency is greater than expected for disorder. The abbreviations is according to ACMG guideline (18).

Table 4. The alternate allele frequencies of *CTSB* gene variants in Turkish and other populations and their clinical definitions according to ACMG classification (18)

Variant Description	Allele Frequencies in gnomAD data set, n (allele number)									In-House data set	dbSNP ID	Position of the variant on the protein	ACMG Definiation	ACMG Verdict Rules
	African, n=712	Ashkenazi Jewish, n=4158	East Asian, n=1746	European (Finnish), n=3116	European (Non-Finnish), n=17568	Latino, n=11012	South Asian, n=10288	Other, n=1552	Total, n=50152	MGexome, n=276				
c.-22G>A	0.0229	0.00192	-	**0.0000493**	0.00164	0.0046	0.000756	0.00547	**0.00318**	0.00362	rs144735294		Benign	BS1, BS2, BP4
c.-11C>T	0.0122	**0.0711**	0.000382	**0.0832**	**0.0894**	0.0445	0.0214	**0.0786**	**0.0617**	**0.0833**	rs1137063		Benign	BA1, BP4
c.76C>G	**0.329**	**0.284**	**0.503**	**0.402**	**0.383**	**0.351**	**0.357**	**0.382**	**0.378**	**0.399**	rs12338	p.Leu26Val	Benign	BA1, BP4
c.111G>T	-	-	-	-	-	-	-	-	-	0.00362	novel	p.Arg37=	Likely Benign	PM2, BP4, BP7
c.157A>G	0.0459	**0.094**	0.000109	**0.0934**	**0.134**	**0.0633**	0.0469	**0.115**	**0.0926**	**0.0942**	rs1803250	p.Ser53Gly	Benign	BA1, BP6
c.223A>G	**0.0000621**	0.000898	0.0275	-	0.000133	0.000697	0.000891	0.00163	**0.00236**	0.00362	rs74996838	p.Thr75Ala	Benign	BS1, BS2, BP4
c.327+20C>T	**0.316**	**0.245**	**0.492**	**0.347**	**0.365**	**0.343**	**0.339**	**0.364**	**0.359**	**0.127**	rs2272766		Benign	BA1, BP4
c.328-38G>T	**0.0000653**	0.000101	-	-	0.000155	-	-	-	**0.0000783**	0.00362	rs754014623		Benign	BS1, BS2, BP4
c.328-32G>C	-	-	-	-	-	**0.0000291**	-	-	**0.00000408**	0.00362	rs1321186406		VUS	PM2, BP4
c.420A>C	**0.598**	**0.572**	**0.492**	**0.504**	**0.612**	**0.454**	**0.546**	**0.589**	**0.561**	**0.594**	rs13332	p.Thr140=	Benign	BS1, BP4, BP7
c.420A>G	**0.194**	**0.0592**	0.000164	0.0164	0.0263	0.0249	0.0232	0.038	**0.0353**	0.0109	rs13332	p.Thr140=	Benign	BA1, BP4, BP7
c.446+9G>C	0.000631	0.00129	**0.0000546**	0.000795	0.00504	0.00119	0.00319	0.00346	**0.00308**	0.00725	rs199582927		Benign	BS1, BS2, BP4
c.446+42G>C	**0.145**	**0.167**	0.000493	**0.106**	**0.129**	0.0467	**0.0983**	**0.109**	**0.104**	0.0254	rs28592650		Benign	BA1, BP4
c.447-16G>C	**0.47**	**0.349**	**0.479**	**0.346**	**0.386**	**0.368**	**0.435**	**0.407**	**0.397**	**0.188**	rs2294140		Benign	BA1, BP4
c.533-117C>T	**0.0604**	**0.279**	0.00128	**0.176**	**0.262**	**0.125**	-	**0.234**	**0.179**	0.0435	rs1692819		Benign	BA1, BP4
c.533-111C>A	**0.263**	**0.302**	**0.421**	**0.315**	**0.345**	**0.335**	-	**0.328**	**0.321**	**0.0507**	rs2294139		Benign	BA1, BP4
c.533-11C>G	**0.542**	**0.59**	**0.473**	**0.503**	**0.619**	**0.461**	**0.511**	**0.589**	**0.559**	**0.272**	rs1736088		Benign	BA1, BP4
c.584G>A	-	-	-	-	**0.00000915**	-	**0.0000705**	-	**0.0000127**	0.00362	rs774859450	p.Arg195Gln	VUS	PM2, PP3
c.676+32G>A	0.00126	-	0.000141	-	**0.0000475**	-	**0.000071**	0.000262	**0.000153**	0.00362	rs376744805		Benign	BS1, BS2, BP4
c.676+49G>A	**0.247**	**0.269**	**0.463**	**0.327**	**0.354**	**0.328**	**0.376**	**0.356**	**0.348**	**0.362**	rs2294138		Benign	BA1, BP4
c.677-42G>T	**0.224**	**0.32**	0.000655	**0.178**	**0.262**	**0.12**	**0.143**	**0.228**	**0.2**	**0.112**	rs1692817		Benign	BA1, BP4
c.677-30C>T	0.00499	0.0243	-	0.00685	0.0329	0.0163	0.00263	0.0266	**0.0199**	0.0145	rs147685738		Benign	BS1, BS2, BP4
c.704G>A	**0.136**	0.00268	**0.0000544**	-	0.000941	0.00762	0.00049	0.00636	**0.0106**	0.00362	rs17573	p.Ser235Asn	Benign	BA1, BP4
c.737A>C	0.00861	0.000298	-	**0.0000924**	0.00174	0.00347	**0.0000653**	0.00293	**0.00192**	0.00362	rs114308907	p.Asn246Thr	Benign	PP3, BS1, BS2, BP6
c.923-36G>T	0.0039	**0.0561**	0.000113	0.0164	0.0266	0.0167	0.0232	0.0319	**0.0218**	0.0435	rs142249463		Benign	BA1, BP4
c.949G>A	0.000123	0.000893	0.0238	**0.0000924**	0.000123	0.000694	0.000784	0.00147	**0.00207**	0.00362	rs79487342	p.Asp317Asn	Benign	BS1, BS2, BP4
c.1001A>C	-	-	0.0327	0.0114	0.000501	**0.0000578**	0.0015	0.00147	**0.00383**	0.00362	rs117613666	p.Gln334Pro	Benign	BS1, BS2, BP4, BP6
c.*40C>G	**0.111**	**0.315**	0.0006	**0.177**	**0.261**	**0.114**	**0.146**	**0.223**	**0.192**	**0.199**	rs709821		Benign	BA1, BP4

Green: AF>0.05; red: AF<0.0001; bold: AF of total gnomAD populations; VUS: Uncertain Significance; PM2: absent from control databanks; BA1: allele frequency is greater than 5%; PP3: multiple lines of computational evidence support a deleterious effect; BP4: impact on gene in computational analysis, BP6: in recently reports variant as benign; BP7: a silent variant; BS1: allele frequency is greater than expected for disorder. The abbreviations is according to ACMG guideline (18).

Table 5. The alternate allele frequencies of *CTSL* gene variants in Turkish and other populations and their clinical definitions according to ACMG classification (18).

	Allele Frequencies in gnomAD data set, n (allele number)									In-House Data Set				
Variant Description	**African, n=712**	**Ashkenazi Jewish, n=4158**	**East Asian, n=1746**	**European (Finnish), n=3116**	**European (Non-Finnish), n=17568**	**Latino, n=11012**	**South Asian, n=10288**	**Other, n=1552**	**Total, n=50152**	**MGexome, n=276**	**dbSNP ID**	**Position of the variant on the protein**	**ACMG Definiation**	**ACMG Verdict Rules**
c.-10-33C>G	**0.0629**	**0.075**	**0.0000544**	0.0265	0.0414	0.0302	0.0225	**0.0509**	**0.0362**	0.00362	rs10123692		Benign	BA1, BP4
c.5A>C	0.00123	0.000695	-	0.00115	0.00702	0.00543	0.000719	0.00391	**0.00431**	0.00362	rs112682750		Likely Benign	BS1, BP4
c.126+11C>T	0.000246	0.00179	-	0.00129	0.00376	0.00139	-	0.00179	**0.00213**	0.00362	rs184253842		Likely Benign	BS1, BP4
c.126+19G>C	-	-	-	-	-	-	-	-	**-**	0.00362	novel		VUS	PM2, BP4
c.126+31C>T	**0.552**	**0.508**	**0.345**	**0.431**	**0.451**	**0.412**	**0.455**	**0.436**	**0.445**	**0.435**	rs2274611		Benign	BA1, BP4
c.402G>A	0.00796	0.0207	0.000109	**0.0594**	**0.0509**	0.0278	0.0337	0.0369	**0.0383**	0.0290	rs11541204	p.Gln134=	Benign	BA1, BP4, BP7
c.933C>T	-	-	0.000598	-	**0.0000264**	**0.0000289**	**0.0000327**	-	**0.0000636**	0.00362	rs138353727	p.Tyr311=	Likely Benign	BS1, BP4, BP7
c.*11G>C	0.000124	-	-	**0.0000463**	0.00103	**0.0000871**	**0.0000656**	0.000328	**0.000507**	0.00362	rs370032344		Likely Benign	BS1, BP4

Green: AF>0.05; red: AF<0.0001; bold: AF of total gnomAD populations; VUS: Uncertain Significance; PM2: absent from control databanks; BA1: allele frequency is greater than 5%; PP3: multiple lines of computational evidence support a deleterious effect; BP4: impact on gene in computational analysis, BP6: in recently reports variant as benign; BP7: a silent variant; BS1: allele frequency is greater than expected for disorder. The abbreviations is according to ACMG guideline (18).

DISCUSSION

For the first time, in-house whole-exome data set of 138 unrelated individuals were analysed in this preliminary study to reflect the allele frequencies of exonic and flanking intronic variants in *ACE2, TMPRSS2, CTSB,* and *CTSL* gene in a group of unrelated non-anonymous individuals from Turkey. SARS-CoV/ CoV-2 coronaviruses utilize the proteins encoded by these genes as entry machinery into the host cell. The tissues and organs such as lung, small intestine, liver, kidney, minor salivary glands, and heart in which these genes are expressed are the regions most affected by the coronaviruses infection and the gene variants were claimed to be associated with gene expression levels (14-16).

To systematically investigate the candidate functional coding variants in *ACE2* and allele frequency differences among populations, Cao and colleagues have analysed the 1700 variants in ChinaMAP (China Metabolic Analytics Project) and public 1KGP (1000 Genomes Project), gnomAD and TopMed databases (15). In this study 32 of 62 variants of *ACE2* were identified to be potentially affecting the amino acid sequences. However, their results demonstrated that there was no natural variant for coronavirus S-protein binding in different populations except the Gln300X variant of ACE2 identified in the ChinaMAP (15). In previous studies, the lysine 31, tyrosine 41 and 82-84 residues in alpha-helix 1 and 353-357 residues in beta5 region of ACE2 were shown to be important for binding to the spike protein of SARS-CoV-1 (19) and several variants, including Met383Thr, Pro389His, Asp427Tyr and His505Arg variants of ACE2 were reported to slightly inhibit the interaction with the spike protein (14,19). Among these variants only His505Arg was identified in our study. The allele frequency of His505Arg variant as an interaction-inhibitor variant in different populations was very low or absent compared to our population (allele frequency, 0.00362 in MGexome; Table 2) (14). In addition, the novel missense variant (Leu120Pro) was identified in one individual. On the other hand, Lys26Arg, Ile468Val, Ala627Val, Asn638Ser, Ser692Pro, Asn720Asp, and Leu731Ile/Leu731Phe variants that were termed as hotspot regions in different population data sets (16) were not shown in our group of the Turkish population (Table 2). Interestingly, the c.439+4G>A variant (rs2285666), which has the highest allele frequency among 62 variants identified in the study of Cao et al. (China-MAP and Han China South populations, 0.556 and 0.557, respectively) (15) compared to others (African, 0.211; Ashkenazi Jewish, 0.258; East Asian, 0.547; European, 0.191; Finnish, 0.2; Latino, 0.394; South Asian; 0.453) has been shown to exhibit much higher allele frequency when compared to the Turkish population group (0.174) (Table 2). The comparative genetic analyses in previous studies (14-16) suggest that *ACE2* variants may play important roles in susceptibilities to COVID-19 and its associated cardiovascular conditions by altering the Renin-Angiotensin pathway.

In conclusion, we have explored possible genetic variants impacting the severity of COVID-19, focusing on four genes involved in viral entry machinery in a group of the Turkish population. Finally, these variants may contribute to the transmissibility of SARS-CoV-2 infection. In future studies, the variant analysis needs to be searched in expanded study groups and the functional studies are required to elucidate the roles of these variants and their interactions in individual treatments in the SARS-CoV-2 infection.

REFERENCES / KAYNAKLAR

1. Lee N, Hui D, Wu A, Chan P, Cameron P, Joynt GM, et al. A major outbreak of severe acute respiratory syndrome in Hong Kong. N Engl J Med. 2003; 348(20): 1986-94.
2. Peiris JS, Guan Y, Yuen KY. Severe acute respiratory syndrome. Nat Med. 2004; 10(12 Suppl): S88-97.
3. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med. 2012; 367(19): 1814-20.

4. Wang Z, Yang B, Li Q, Wen L, Zhang R. Clinical Features of 69 Cases with Coronavirus Disease 2019 in Wuhan, China. Clin Infect Dis. 2020; 71(15): 769-77.
5. Li W, Greenough TC, Moore MJ, Vasilieva N, Somasundaran M, Sullivan JL, et al. Efficient replication of severe acute respiratory syndrome coronavirus in mouse cells is limited by murine angiotensin-converting enzyme 2. J Virol. 2004; 78(20): 11429-33.
6. Li KK, Yip CW, Hon CC, Lam CY, Zeng F, Leung FC. Characterisation of animal angiotensin-converting enzyme 2 receptors and use of pseudotyped virus to correlate receptor binding with susceptibility of SARS-CoV infection. Hong Kong Med J. 2012; 18(3): 35-8.
7. Hou Y, Peng C, Yu M, Li Y, Han Z, Li F, et al. Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry. Arch Virol. 2010; 155(10): 1563-9.
8. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature. 2003; 426(6965): 450-4.
9. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020; 181(2): 271-80.
10. Simmons G, Gosalia DN, Rennekamp AJ, Reeves JD, Diamond SL, Bates P. Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. Proc Natl Acad Sci USA. 2005; 102(33): 11876-81.
11. Matsuyama S, Nagata N, Shirato K, Kawase M, Takeda M, Taguchi F. Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. J Virol. 2010; 84(24): 12658-64.
12. Shulla A, Heald-Sargent T, Subramanya G, Zhao J, Perlman S, Gallagher T. A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. J Virol. 2011; 85(2): 873-82.
13. Kawase M, Shirato K, van der Hoek L, Taguchi F, Matsuyama S. Simultaneous treatment of human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute respiratory syndrome coronavirus entry. J Virol. 2012; 86(12): 6537-45.
14. Darbani B. The Expression and Polymorphism of Entry Machinery for COVID-19 in Human: Juxtaposing Population Groups, Gender, and Different Tissues. Int J Environ Res Public Health. 2020; 17(10): 3433. doi:10.3390/ijerph17103433
15. Cao Y, Li L, Feng Z, Wan S, Huang P, Sun X, et al. Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. Cell Discov. 2020; 6: 11. doi: 10.1038/s41421-020-0147-1
16. Hou Y, Zhao J, Martin W, Kallianpur A, Chung MK, Jehi L, et al. New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis. Version 2. BMC Med. 2020; 18(1): 216. doi: 10.1186/s12916-020-01673-z
17. Asselta R, Paraboschi 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 (Albany NY). 2020; 12(11): 10087-98.
18. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015; 17(5): 405-24.
19. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J. 2005; 24(8): 1634-43.