



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Variability in genes related to SARS-CoV-2 entry into host cells (*ACE2*, *TMPRSS2*, *TMPRSS11A*, *ELANE*, and *CTSL*) and its potential use in association studies

Gilberto Vargas-Alarcón^{a,*}, Rosalinda Posadas-Sánchez^b, Julian Ramírez-Bello^c

^a Department of Molecular Biology, Instituto Nacional de Cardiología Ignacio Chávez, Mexico City, Mexico

^b Department of Endocrinology, Instituto Nacional de Cardiología Ignacio Chávez, Mexico City, Mexico

^c Research Unit, Hospital Juárez de México, Mexico City, Mexico

ARTICLE INFO

Keywords:

SARS-CoV2
COVID19
ACE2
TMPRSS2
TMPRSS11A
Cathepsin
Elastase
Polymorphisms

ABSTRACT

Background: The prevalence and mortality of the outbreak of the COVID-19 pandemic show marked geographic variation. The presence of several subtypes of the coronavirus and the genetic differences in the populations could condition that variation. Thus, the objective of this study was to propose variants in genes that encode proteins related to the SARS-CoV-2 entry into the host cells as possible targets for genetic associations studies. **Methods:** The allelic frequencies of the polymorphisms in the *ACE2*, *TMPRSS2*, *TMPRSS11A*, cathepsin L (*CTSL*), and elastase (*ELANE*) genes were obtained in four populations from the American, African, European, and Asian continents reported in the 1000 Genome Project. Moreover, we evaluated the potential biological effect of these variants using different web-based tools.

Results: In the coding sequences of these genes, we detected one probably-damaging polymorphism located in the *TMPRSS2* gene (rs12329760) that produces a change of amino acid. Furthermore, forty-eight polymorphisms with possible functional consequences were detected in the non-coding sequences of the following genes: three in *ACE2*, seventeen in *TMPRSS2*, ten in *TMPRSS11A*, twelve in *ELANE*, and six in *CTSL*. These polymorphisms produce binding sites for transcription factors and microRNAs. The minor allele frequencies of these polymorphisms vary in each community; indeed, some of them are high in specific populations.

Conclusion: In summary, using data of the 1000 Genome Project and web-based tools, we propose some polymorphisms, which, depending on the population, could be used for genetic association studies.

1. Introduction

The coronavirus disease 2019 (COVID-19) pandemic, as declared by the World Health Organization on 11 March 2020 [1], has accumulated 6057.853 confirmed cases globally until June 1 [2]. The etiologic agent of COVID-19 is a novel beta coronavirus [3], which was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses [4]. Zhou et al. established that the SARS-CoV-2 is 96% identical at the whole-genome level to a bat SARS-like coronavirus and 79.5% identical to SARS-CoV [5].

Coronaviruses possess an enveloped, single, positive-stranded RNA genome that encodes for four membrane polypeptides, namely spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins [6]. The spike glycoprotein (S) present in the coronavirus envelope is used to bind and penetrate the host cells. The S protein is composed of two subunits: S1 and S2; the S1 subunit allows the virus to bind the host cell

receptors, while S2 enables the fusion of viral and cellular membranes. The SARS-CoV-2 entry into target cells requires S protein priming by cellular proteases, which entails S protein cleavage at the S1/S2 and S2' sites [7]. Depending on virus strains and cell types, coronavirus (CoV) S proteins may be cleaved by one or several host proteases, including furin, cathepsins, transmembrane protease serine protease-2 (TMPRSS-2), neutrophil elastase (ELANE), and probably TMPRSS11A [8–15]. The availability of these proteases on target cells largely determines whether CoV particles enter cells through plasma membrane or endocytosis. Hoffmann et al. demonstrated that SARS-CoV-2 uses the SARS-CoV receptor angiotensin-converting enzyme 2 (ACE2) for entry into target cells and the transmembrane protease 2 (TMPRSS2) for S protein priming [16]. In the same way, Ou et al. found that cathepsin L (CTSL) is critical for virus entry [17]. It has also been reported that the S protein of the A2a subtype has an additional elastase-specific proteolytic cleavage site that endows the virus with an increased ability to

* Corresponding author.

E-mail address: gvargas63@yahoo.com (G. Vargas-Alarcón).

<https://doi.org/10.1016/j.lfs.2020.118313>

Received 7 June 2020; Received in revised form 7 August 2020; Accepted 19 August 2020

Available online 21 August 2020

0024-3205/ © 2020 Published by Elsevier Inc.

Table 1
ACE2 polymorphisms.

ACE2	MAF (%) in populations with different ancestry					Potential functional effect			
	Variant ID	Minor allele	AMER (MXL)	AFR (YRI)	EUR (GBR)	EAS (CHB)	Amino acid position and change	Y	N
Coding sequence									
rs35803318C/T	T	8.3	0	4.4	0	Val749Val			X
rs4646179A/G	G	0	12.2	0	0	Asn690Asn			X
Promoter and 5' near the gene									
rs113009615AAAAAA/AAAAAAA (INDEL)	A	2.1	17.7	0.7	0				X
rs7885856G/A	A	2.1 ^a	7.3	0	0		Both alleles can create binding sites for AP2ALPHA, BCL6, CEBP, and ETS.		
rs112621533T/C	C	0	8.5	0.7 ^c	0				X
rs11336754ATTT/ATT (INDEL)	ATT	4.2	14.6	0.7	2.5				X
rs760084155G/A	A	18.7	0	0	0				X
rs765471058A/T	T	17.7	0	0	0				X
rs9698134C/T	T	2.1 ^a	17.7 ^b	0.7 ^c	0		Allele C can create a binding site for HIC1		
rs9698150G/C	C	2.1 ^a	17.7 ^b	0.7 ^c	0		Both alleles can create binding sites for BRCA, DBP, ETF, MYB, RFX, and WT1		
rs112593415A/G	G	2.1 ^a	17.7 ^b	0.7 ^c	0				X
rs184697926A/C	C	0	0	0	11.9				X
rs142049267A/G	G	0	4.3	0	0				X

ACE2; Angiotensin I Converting Enzyme 2, MAF; Minor allele frequency, AMER; Americans, AFR; Africans, EUR; Europeans, EAS; East Asia, MXL; Mexicans from Los Angeles, YRI; Yoruba in Ibadan, Nigeria, CHB; Han Chinese in Beijing, China, GBR; British in England and Scotland, Y; Yes, N; No, INDEL; Insertion/Deletion, LD; Linkage disequilibrium. CCDS; consensus coding sequence.

ACE2 is located on chromosome Xp22.2. Five transcripts have been reported for ACE2, two of them synthesize the CCDS of 805 amino acids. The first transcript consists of 18 exons and 17 introns, 18 exons encode this protein, transcript length; 3339 bps. The second transcript consists of 19 exons and 18 introns, the CCDS consist of 18 exons, transcript length; 3507 bps.

^a Variants in high LD or tagSNPs between them in an American population.

^b Variants in high LD or tagSNPs between them in an African population.

^c Variants in high LD or tagSNPs between them in a European population.

penetrate host cells [10]. This virus subtype was reported in China and spread rapidly in Europe and North America [18,19].

ACE2 is a carboxypeptidase that converts angiotensin II to angiotensin-(1-7) [Ang-(1-7)], which evokes anti-fibrosis, anti-hypertrophy, vasodilatation, and other beneficial effects [20–22]. Tissue-bound or membrane-bound ACE2 is a kind of transmembrane protein with a single metalloprotease active site and a transmembrane domain [23,24]. The ACE2 receptor is expressed at high levels not only in alveolar type-2 cells in the lung, but also in liver cholangiocytes, myocardial cells, esophagus keratinocytes, kidney proximal tubules, bladder urothelial cells, and gastrointestinal epithelial cells [25,26]. In the lung, ACE2 is abundantly expressed in Clara cells, type I and II alveolar epithelial cells, macrophages, endothelium, vascular smooth muscle cells, and bronchial epithelia [27]. ACE2 is encoded in the chromosome Xp22 and spans 39.98 kb of genomic DNA. This gene generates two transcripts that originate the same 805-amino-acid-residue protein; one transcript consists of 18 exons and 17 introns (transcript length: 3339 bps) and the other is composed of 19 exons and 18 introns (transcript length: 3507 bps). The ACE2 gene exhibits a high level of polymorphism; in fact, some single nucleotide polymorphisms (SNPs) have been associated with susceptibility to diseases, such as type 2 diabetes and hypertension [28,29].

The transmembrane serine protease TMPRSS2 is an essential enzyme that can cleave hemagglutinin of many subtypes of the influenza virus and the coronavirus S protein [30,31]. It has been reported that TMPRSS2 deficiency protects mice against H1N1, mouse-adapted H1N1, and H7N9 influenza A virus infections [30,32]. Recently, it has been shown that TMPRSS2 can help SARS-CoV-2 enter host cells by cleaving the S protein [16]. Matsuyama et al. demonstrated that TMPRSS2-expressing cell lines are highly susceptible to SARS-CoV, MERS-CoV, and SARS-CoV-2 [33]. The gene that encodes TMPRSS2 is polymorphic and is considered a susceptibility gene for H1N1 and H7N9 influenza [34]. Similarly, TMPRSS11A is another member of the subfamily of type II transmembrane serine proteases. This enzyme is

synthesized as a zymogen and can be activated upon auto-proteolytic cleavage at a site located between the protease domain and the stem region [35]. Zmora et al. demonstrated that TMPRSS11A cleaves and activates the MERS-CoV spike protein and the influenza A virus hemagglutinin [36].

Elastase is known to be secreted by neutrophils as part of an inflammatory response to a viral infection and is also produced by opportunistic bacteria that can colonize virally infected respiratory tissue [37]. The increase of elastase activity as a result of an extreme inflammatory process produces an important pulmonary injury contributing significantly to the pathogenesis of chronic obstructive pulmonary disease, cystic fibrosis, acute respiratory distress syndrome, and pulmonary fibrosis [38,39].

Cathepsin L is a peptidase that preferentially cleaves peptide bonds with aromatic residues in the P2 position and hydrophobic residues in the P3 position [40]. It has been previously reported that cathepsin L participates in the viral glycoprotein processing of Ebola and SARS-CoV. It is well established that this viral process is important for cell membrane fusion and host cell entry [41]. Using inhibitors of cathepsin B and L in HEK 293/hACE2 cells, Ou et al. [17] demonstrated that the treatment with cathepsin L inhibitor decreases the entry of SARS-CoV-2 into the cells. This result suggests that cathepsin L could be very important for S protein priming in lysosome for viral entry.

The outbreak of the COVID-19 pandemic shows marked geographic variation in its prevalence and mortality. This variability could be due to both the presence of several subtypes of the virus and the genetic differences in the human populations [18,19,42–44]. Considering this fact and the important role of the ACE2, TMPRSS2, TMPRSS11A, cathepsin L, and elastase in the process of virus entry into the host cell, the present study aims to propose possible variants in these *loci* for genetic association studies in patients with SARS-CoV-2 infection.

Table 2
TMPRSS2 polymorphisms.

TMPRSS2		MAF (%) in populations with different ancestry					Potential functional effect	
Variant ID	Minor allele	AMER (MXL)	AFR (YRU)	EUR (GBR)	EAS (CHB)	Amino acid position and change	Y	N
Coding sequence								
rs61735794C/T	T	0.8	0	2.8	0	Gly385Gly		X
rs2298659G/A	A	31.2	16.2	21.4	29.6	Gly290Gly		X
rs17854725A/G	G	47.7	36.1	56.4	17.5	Ile256Ile		X
rs61735789G/A	A	1.6	0	0.5	0	Tyr180Tyr	Probably Damaging (by PolyPhen-2)	X
rs12329760C/T	T	18.0	25.5	20.9	41.3	Val160Met	Deleterious (by SIFT)	
rs3787950T/C	C	1.6	30.1	7.1	11.7	Thr75Thr	Without post-translational modification (by ModPred server)	X
rs61735792G/A	A	0	0	1.1	0	Pro63Pro		X
Promoter and 5' near the gene								
rs4303794A/C	C	28.1*	41.2 ^{***}	41.8*	1 ±		The C allele creates binding sites for AP2, and SP1, and WTI	
rs11088551A/G	G	28.1*	41.2 ^{***}	41.8*	1 ±		The A allele creates binding sites for BRCA, MYB, NF1, and RFX	
rs66492316GGCGGAGGC/C (INDEL)	C	28.1	41.2 ^{***}	41.8	1			X
rs4303795A/G	G	28.1*	41.2	41.2*	1 ±		The G allele creates binding sites for HNF4 and KID3	
rs5844077G/GA (INDEL)	G	29.6	25.3	25.8	8.3			X
rs76833541G/A	A	31.3	0	10.4	0			X
rs4283504G/T	T	16.4	37.0	12.6	23.3		The T allele creates binding sites for DBP, HSF1, and NKX25	
rs12481984T/C	C	27.3*	39.4 [†]	40.7*	1 ±		The C allele creates a binding site for HAND1E47	
rs28707508G/A	A	25.8*	38.4 [†]	40.1*	1 ±		The A allele creates binding sites for HNF3, ALPHA, and TBP	
rs552257429C/CT (INDEL)	CT	26.6	38.4	39.0	1.5		The G allele creates a binding site for KAISO	
rs12626358C/G	G	26.8	26.4	9.3	56.8		The C allele creates binding sites for ETF, KROX, LRF, and SPZ1	
rs128074C/T	T	16.4	2.8	12.6	23.8		The A allele creates a binding site for PPAR_DRI	
rs56218846G/A	A	25.8*	40.3 [†]	40.1*	1			X
rs11281229T/TCCAGG (INDEL)	TCCAGG	25.8	40.3	40.1	0.9			
rs8127674A/G	G	25.8*	40.3 [†]	40.1*	1		The G allele creates binding sites for AP2ALPHA, ETF, and SPZ1	
3' near the gene								
rs11088550G/A	A	12.5**	0	9.3	0			X
rs463727T/A	A	26.6	4.2	46.2	0.5			X
rs462471G/A	A	36.7***	34.3 ^{***}	13.7**	53.4 ±			X
rs76000363G/A	A	12.5**	5.6	12.1**	6.3**			X
3' UTR								
rs143680939GA/G (INDEL)	G	12.5	5.6	12.1	6.8		The C allele creates a binding site for hsa-miR-548c-3p	X
rs456142C/T	T	36.7***	34.3 ^{***}	13.7**	53.4 ±			
rs112657409C/T	T	0.8	7.9	0	6.3 ± ± ±			X
rs2838038A/T	T	12.5**	4.6	12.1**	6.3 ± ±		The T allele can create a binding site for hsa-miR-943	
rs462574G/A	A	24.2	14.8 ^{***}	1.7	47.1 ± ±		The A allele can create a binding site for hsa-miR-1324	
rs456298A/T	T	37.5***	34.3 ^{***}	13.7**	53.4 ±		The A allele can create a binding site for hsa-miR-450b-5p	
rs17001042G/A	A	0.8	13.9	0	0		The A alleles can create a binding site for hsa-miR-220b	X
rs11910678T/C	C	1.6	13.9 ^{***}	0	6.3 ± ± ±			X
rs77675406G/A	A	12.5**	4.6	12.1**	6.3 ± ±			
rs12627374C/T	T	0	0	0	13.6		The C allele can create a binding site for hsa-miR-345	
rs62217525C/T	T	3.9	0	6.0	0		The C allele can create a binding site for hsa-miR-1226	X
rs77996454G/A	A	4.6	0.8	0	0			X
rs149695119TG/T (DELETION)	T	22.7	0.8	0	0			

(continued on next page)

Table 3
TMPRSS11A polymorphisms.

TMPRSS11A		MAF (%) in populations with different ancestry				Potential functional effect		
Variant ID	Minor allele	AMER (MXL)	AFR (YRI)	EUR (GBR)	EAS (CHB)	Amino acid position and change	Y	N
Coding sequence								
rs1371932A/G	G	46.1	32.4	49.5	31.1 ±	Asp334Asp		X
rs353163C/T	T	46.1	13.4	40.1	15.5	Arg290Gln	Benign (by PolyPhen-2) Tolerated (by SIFT) Arg290 originates a proteolytic cleavage (by ModPred server)	
rs1370840G/A	A	41.4*	55.6	20.3 [‡]	10.7 ± ±	Thr81Thr		X
rs139010197T/C	C	0.8	0	4.4	0	Lys48Arg	Benign (by PolyPhen-2) Tolerated (by SIFT) 48Arg originates a proteolytic cleavage (by ModPred server)	
rs11930532T/C	C	41.4*	78.7	20.3 [‡]	10.7 ± ±	Val6Val		X
rs97728C/T	T	39.1*	10.6	20.9 [‡]	9.7 ± ±	Met1Ile	Benign (by PolyPhen-2) Tolerated (by SIFT) Without post-translational modification (by ModPred server)	
Promoter and 5' near the gene								
rs17088849T/C	C	15.6	13.4	22.0	53.4		The C allele creates binding sites for BRCA and MYB	
rs200058897TA/T (INDEL)	T	5.5	13.9	5.0	0			X
rs536791104C/G	G	5.5**	13.9 [‡]	5.0 ^{‡‡}	0			X
rs6552135A/G	G	37.5	0.5	52.2	36.9		The C allele creates binding sites for CEBPA, CEBPDELTA, and CEBP	
rs17088850A/G	G	5.5**	20.4	5.0 ^{‡‡}	0		The C allele creates binding sites for BRCA and MYB	
rs17088851T/C	C	5.5**	11.6 [‡]	5.0 ^{‡‡}	0		The C allele creates binding sites for AREB6, ETF, KID3, and SPZ1	
rs720009T/A	A	1.6	24.1	0	0		The C allele creates a binding site for GATA6	
5' UTR								
rs6552134A/G	G	46.9	79.2	25.8	9.7 ± ±		The G allele creates a binding site for AP2ALPHA	
rs11947613G/A	A	2.3	47.7	0	0		The G allele creates a binding site for TBP	
3' UTR								
rs4860265A/G	G	47.7	37.4	34.1	30.1 ±		The G allele creates a binding site for hsa-miR-658	
rs9998258T/C	C	2.3***	1.8	6.6 ^{‡‡‡}	0		The T allele creates binding sites for hsa-miR-1, hsa-miR-613, and hsa-miR-148b	
rs33929303C/T	T	20.3	25.9	31.3	9.2			X
rs28648375T/A	A	2.3***	0	6.6 ^{‡‡‡}	0		The T allele creates a binding site for hsa-miR-1244	
rs12646286C/T	T	25.8	5.1	18.1	60.7			X

*, **, *** Variants in high LD or tagSNPs between them in an American population. [‡], Variants in high LD or tagSNPs between them in an African population. [‡], ^{‡‡}, ^{‡‡‡} Variants in high LD or tagSNPs between them in a European population. ±, ± ± Variants in high LD or tagSNPs between them in an Asian population. TMPRSS11A; Transmembrane Serine Protease 11A, MAF; Minor allele frequency, AMER; Americans, AFR; Africans, EUR; Europeans, EAS; East Asia, MXL; Mexicans from Los Angeles, YRI; Yoruba in Ibadan, Nigeria, CHB; Han Chinese in Beijing, China, GBR; British in England and Scotland, Y; Yes, N; No, INDEL; Insertion/Deletion, UTR; untranslated region, LD; Linkage disequilibrium. CCDS; consensus coding sequence.

TMPRSS11A is located on chromosome 4q13.2. Three transcripts have been reported for TMPRSS11A, two produce CCDS. The first transcript consists of 10 exons and 9 introns, 10 exons encode this 421 amino-acid protein, transcript length; 3054 bps. The second transcript consists of 10 exons and 9 introns, 10 exons encode this 418 amino-acid protein, transcript length; 3247.

other hand, three SNPs (located in the promoter or in the 5' region near the gene) had a possible functional effect: rs7885856 (both alleles can create binding sites for AP2alpha, BCL6, CEBP, and ETS transcription factors), rs9698134 (the C allele can produce a binding site for HIC1 transcription factor), and rs9698150 (both alleles produced binding sites for BRCA, DBP, ETF, MYB, RFX, and WT1).

3.2. TMPRSS2 polymorphisms

The TMPRSS2 polymorphisms are shown in Table 2. In this case, thirty-nine polymorphisms had a frequency higher than 1.5% in at least one of the populations; four of them were located in the coding sequence and only one (rs12329760) produces an amino acid change (Val160Met). This change was probably damaging (PolyPhen-2, score 0.989, sensitivity 0.72, specificity 0.97). Indeed, using SIFT, we identified that this variant was deleterious (SIFT score 0.009). Our analysis,

based in the ModPred server, showed that neither Val160 nor 160Met undergoes any possible post-translational modification (e.g., acetylation, proteolytic cleavage, glycosylation, phosphorylation). As for this gene, seventeen polymorphisms located in the promoter, in the 5' region near the gene, and the 3' UTR region had a possible functional effect. Ten polymorphisms located in the promoter and the 5' region near the gene produced binding sites for several transcription factors, whereas seven located in the 3' UTR region created potential binding sites for several microRNAs.

3.3. TMPRSS11A polymorphisms

Out of twenty polymorphisms in the TMPRSS11A gene, six were in the coding sequence; three of these generated a nonsynonymous substitution (rs353163-Arg290Gln, rs139010197-Lys48Arg, rs97728-Met1Ile). According to PolyPhen-2 results, rs353163 was possibly

Table 4
ELANE polymorphisms.

Variant ID	Minor allele	MAF (%) in populations with different ancestry				Amino acid position and change	Potential functional effect	
		AMER (MXL)	AFR (YRI)	EUR (GBR)	EAS (CHB)		Y	N
Coding sequence								
rs17223045C/T	T	0.8	11.6	1.1	0	Asn130Asn		X
rs17216663C/T	T	1.6	0	0.6	0	Pro257Leu	Benign (by PolyPhen-2) Tolerated (by SIFT) Pro257 undergoes a hydroxylation (by ModPred server)	
Promoter and 5' near the gene								
rs74876755C/T	T	0	5.6	0	0			X
rs10413889G/A	A	4.7	18.1	12.6	0.5		The A allele creates a binding site for SPZ1	
rs3761007G/A	A	4.7	0	7.7	25.7		The G allele creates a binding site for DR4	
rs3761006G/A	A	5.5	0	0.5	18.0		The A allele creates binding sites for OCT and P53	
rs10409474C/G	G	10.2	28.2	12.6	28.6		The G allele creates a binding site for YY1	
rs3761005T/A	A	44.5	68.5	31.3	59.7		The A allele creates binding sites for CEBPDELTA and YY1	
rs351107T/G	G	0.8	9.3	1.7	0		The G allele creates a binding site for DBP	
rs3761001G/A	A	14.8	56.0	25.3*	29.6 [‡]		The G allele creates binding sites for USF and LRF	
rs2007647G/A	A	7.0	9.7	24.2*	1.0		The A allele creates binding sites for ETS, HMG1Y, NFAT, and OCT1	
rs17216593C/T	T	0.8	7.4	0.6	0		The T allele creates binding sites for PAX8 and SREBP	
rs740021C/A	A	7.0	27.3	1.1	28.2 [‡]		The A allele creates a binding site for CEBPGAMMA	
3' near the gene								
rs187713106T/A	A	0.8	11.1	3.3	0			X
rs113311784T/TA (INDEL)	TA	12.5	6.5	16.5	38.8			X
rs6510983C/T	T	15.6	38.0	25.8	1.9		The A allele creates a binding site for CEBPA	
rs17223066G/A	A	54.7	23.1	44.5	33.5		The G allele creates a binding site for CREB	

* , [‡], Variants in high LD or tagSNPs between them in a European, and Asian population, respectively.

ELANE; Elastase, neutrophil expressed, MAF; Minor allele frequency, AMER; Americans, AFR; Africans, EUR; Europeans, EAS; East Asia, MXL; Mexicans from Los Angeles, YRI; Yoruba in Ibadan, Nigeria, CHB; Han Chinese in Beijing, China, GBR; British in England and Scotland, Y; Yes, N; No, INDEL; Insertion/Deletion, LD; Linkage disequilibrium. CCDS; consensus coding sequence.

ELANE is located on chromosome 19p13.3. Two transcripts have been reported for this gene, which produce CCDS. The first transcript consists of 5 exons and 4 introns, 5 exons encode this 267 amino-acid protein, transcript length; 909 bps. The second transcript consists of 6 exons and 5 introns, 5 exons encode this 267 amino-acid protein, transcript length; 1028.

benign (Polyphen-2 score 0.015, sensitivity 0.96, specificity 0.79) and tolerated (SIFT score 1). Using ModPred, we found that 290Gln was not affected; in contrast, variant Arg290 was predicted to undergo a translational modification: a proteolytic cleavage (score 0.71 and medium confidence). A similar result was observed with the 48Arg variant (rs139010197); this variant was predicted to be benign (PolyPhen-2 score 0.02, sensitive 0.95, specificity 0.8) and tolerated (SIFT score 0.53). Using ModPred, we identified that the 48Arg variant might undergo a proteolytic cleavage (score 0.57, low confidence). Alternatively, we did not identify any effect of Met1 or 1Ile (rs977728). Both five polymorphisms located in the promoter and the 5' region near the gene and the two located in the 5'UTR region had a possible functional effect; these SNPs produced binding sites for several transcription factors. Finally, three out of five polymorphisms located in the 3'-UTR region created binding sites for some microRNAs (Table 3).

3.4. *ELANE* polymorphism

The polymorphisms of the *ELANE* gene are shown in Table 4. As can be seen, two of them (rs17223045 and rs17216663) were located in the coding region. In fact, according to the bioinformatic analysis, rs17216663 provoked a change of amino acid (Pro257Leu), which is benign (PolyPhen-2 score 0.01, sensitivity 0.96, specificity 0.77) and tolerated (SIFT score 0.197). Using ModPred, we found that Pro257

undergoes hydroxylation (score 0.80, medium confidence), while 257Leu is not predicted to be post-translationally modified. In this gene, 12 polymorphisms had a possible functional effect. These polymorphisms were located in several regions of the gene and spawned binding sites for some transcription factors.

3.5. *CTSL* polymorphisms

In this gene, one polymorphism (rs11541204) was in the coding region, without a change of amino acid. Four out of nine polymorphisms located in the promoter region and the 5' region near the gene, presented a possible functional effect: a binding site for some transcription factors. Both the polymorphism located in the 5' UTR region and the one located in the 3' region near the gene generated binding sites for transcriptional factors (Table 5).

3.6. LD between SNPs of *ACE2*, *TMPRSS2*, *TMPRSS11A*, *ELANE*, and *CTSL*

We conducted an LD analysis between the SNPs in *ACE2* (Fig. 1), *TMPRSS2* (Fig. 2), *TMPRSS11A* (Fig. 3), *ELANE* (Fig. 4), and *CTSL* (Fig. 5) proposed here. Thus, our analysis of LD included SNPs in these 5 genes but not INDELS. We observed several non-informative SNPs (minor allele frequency = 0%) in the 4 populations included in our

Table 5
Catepsin L polymorphisms.

Variant ID	Minor allele	MAF (%) in populations with different ancestry						Amino acid position and change		Potential functional effect
		AMER (MXL)	AFR (YRI)	EUR (GBR)	EAS (CHB)	Y	N			
Coding sequence										
rs11541204G/A	A	0	0	5.0	0		Gln134Gln			X
Promoter and 5' near the gene										
rs78985072G/A	A	4.7*	0	0	15.5 [±]					X
rs142421833C/T	T	4.7*	0	0	15.5 [±]					X
rs3128509G/A	A	45.3	13.0	41.8	3.4			Both alleles can create binding sites for BRCA, GATA4, MYB, and RFX		X
rs111786311T/G	G	1.6	16.2	2.8*	0					X
rs11389221C/CAAA (INDEL)	CAAA	43.8	15.3	51.7	76.7			Both alleles can create binding sites for GATA, GFLI, and TEL2		X
rs56952354A/T	T	2.3	4.6	0	0					X
rs75567776G/C	C	2.3	7.9	0	0			The C allele creates binding sites for SREBP, AHR, and AHRHIF		X
rs3118869C/A	A	39.8	46.8	47.3	32.5			Both alleles can create binding sites for BRCA, DBP, LRF, MYB, and STAT4		
rs41307457C/A	A	3.1	23.6	2.8*	0					
5' UTR										
rs41312184C/T	T	1.6	0.5	11.0	0			Both alleles can create binding sites for STAT, and RFX. The C allele can create a binding site for SF2ASF1		
3' near the gene										
rs59063901G/A	A	0.8	3.7	2.8*	0			Both alleles can create binding sites for STAT, SPZ1, and GAPP		

*, *, ± Variants in high LD or are tagSNPs between them in an American, European, and Asian population, respectively.

MAF; Minor allele frequency, AMER; American, AFR; Africans, EUR; Europeans, EAS; East Asia, MXL; Mexicans from Los Angeles, YRI; Yoruba in Ibadan, Nigeria, CHB; Han Chinese in Beijing, China, GBR; British in England and Scotland, Y; Yes, N; No, INDEL; Insertion/Deletion, UTR; untranslated region, LD; Linkage disequilibrium. CCDS; consensus coding sequence.

CTS1 is located on chromosome 9q21.33. Six transcripts have been reported for CTS1, three produce CCDS, and two of them synthesize the 333 amino acid protein. The first transcript consists of 8 exons and 7 introns, 7 exons encode this protein, transcript length; 1436 bps. The second transcript consists of 8 exons and 7 introns, 7 exons encode this protein, transcript length; 1654 pb.

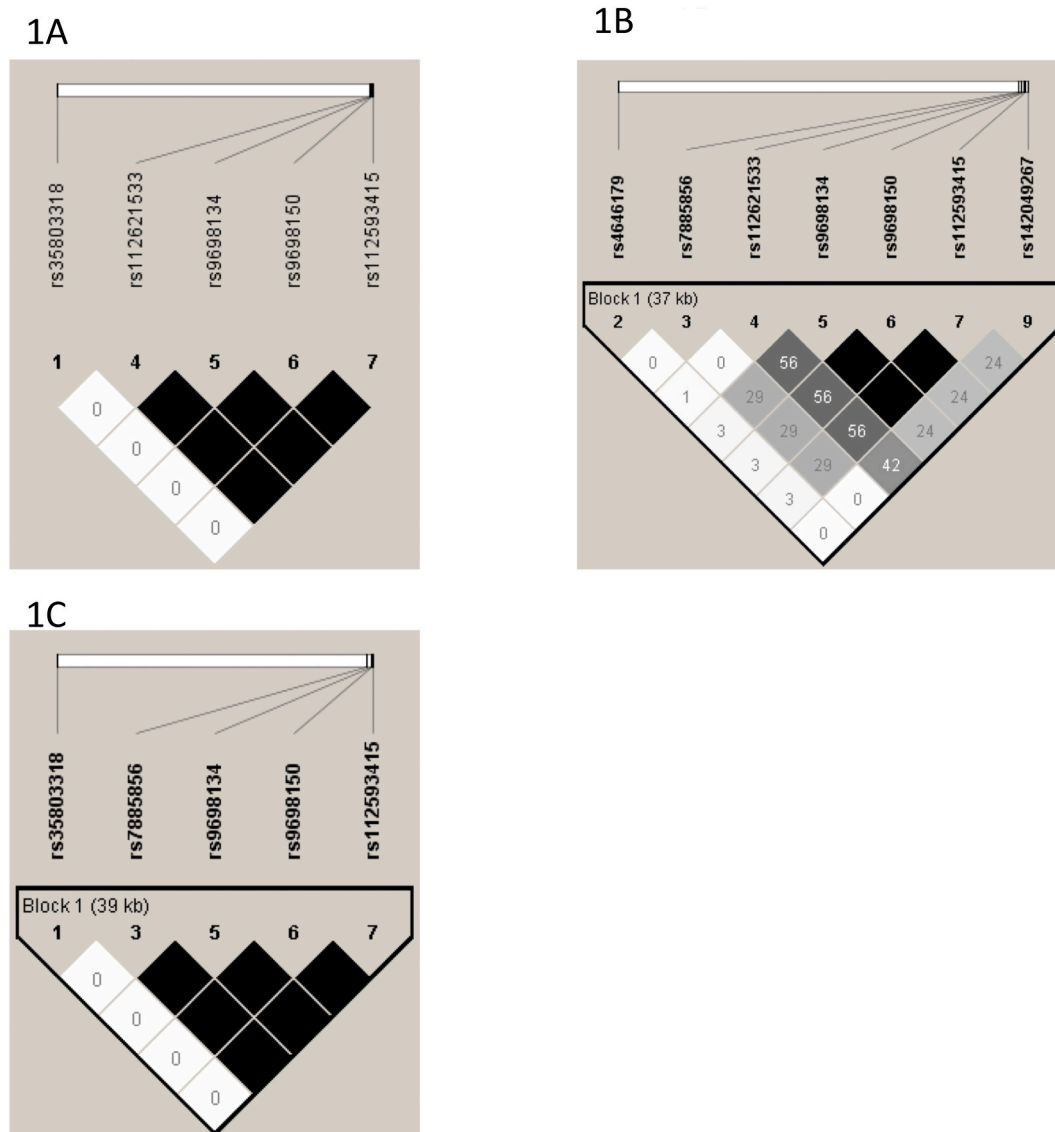


Fig. 1. Linkage disequilibrium (r^2) in the *ACE2* gene in the included populations. Of the 13 variants shown in Table 1, two were INDELS and in two of them no information was found, so they were not added to the Haploview program. Of the remaining 9, some were not polymorphic in the different populations. Linkage disequilibrium (LD) between variants is shown in the figures, 5 in Americans (Fig. 1A), 7 in Africans (Fig. 1B), 5 in Europeans (Fig. 1C). In Asians, none of the variants were in LD.

analysis. For example, for *TMPRSS2*, 2, 7, 5, and 8 SNPs were eliminated in the American, African, European, and Asian populations included in our study, respectively. This reflects the heterogeneity between the populations. Similar results can be observed for *ACE2*, *TMPRSS11A*, *ELANE*, and *CTSL*.

4. Discussion

Using the information about allelic frequencies obtained from dbSNPs, Ensembl Genome Browser, and the 1000 Genome Project, as well as different web-based tools, we defined some polymorphic variants in the *ACE2*, *TMPRSS2*, *TMPRSS11A*, *ELANE*, and *CTSL* genes that could be important for association studies in the SARS-CoV-2 infection. SARS-CoV-2 enters the cell by binding its S protein with cellular receptors (e.g., *ACE2* membrane-bound protein) [16]. Some proteases, such as *TMPRSS2*, cathepsin L, neutrophil elastase, and probably *TMPRSS11A* participate in this process [8–15]; in fact, polymorphisms

in their encoding genes could not only have an impact in the expression and/or structure of these proteases but also be associated with SARS-CoV-2 infection susceptibility.

Even though most of the *ACE2* variants occur at low frequencies in human populations, we detected three polymorphisms with a possible functional effect: binding site generation for some transcription factors. AP2alpha, BCL6, CEBP, ETS (rs7885856), HIC1 (rs9698134), BRCA, DBP, ETF, MYB, RFX, and WT1 (rs9698150) are some of these factors, which could have a role in the virus infection. It has been reported that BCL6 modulates tissue neutrophil survival and exacerbates pulmonary inflammation following influenza virus infection [45]. Han et al. [46] demonstrated that the CEBP alpha participates in the activation of hfg12 prothrombinase during SARS-CoV infection, thus having an important role in the development of thrombosis in SARS. The three *ACE2* polymorphisms with possible functional effects have a high frequency of its minor allele only in the African population. Thus, these polymorphisms could be genetic targets for association studies in this

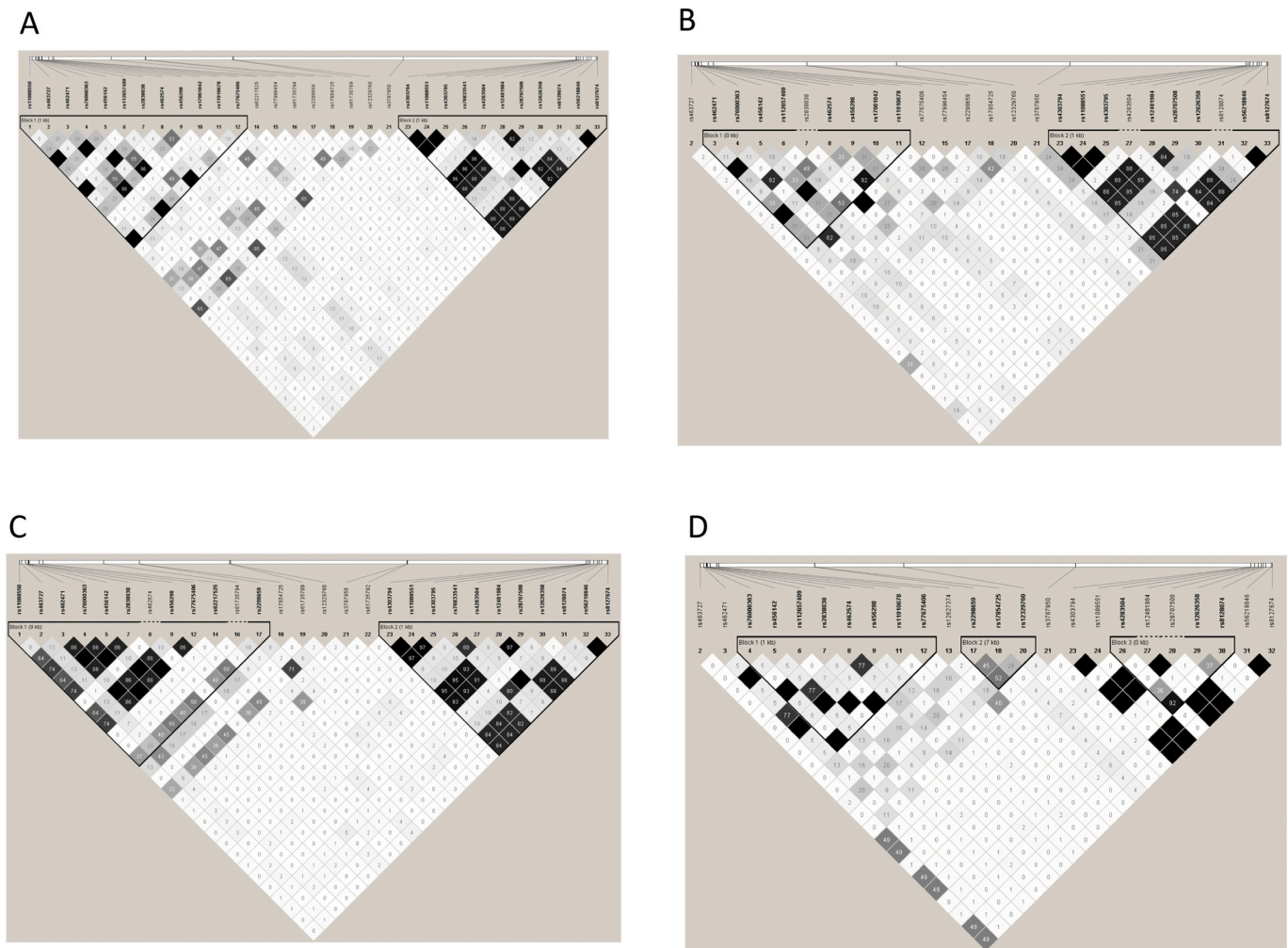


Fig. 2. Linkage disequilibrium (r^2) in the *TMPRSS2* gene in the included populations. Of the 39 variants shown in Table 2, five were INDELs and one a deletion, so they were not added to the Haploview program. Of the remaining 33, some were not polymorphic in the different populations. Linkage disequilibrium between variants is shown in the figures, 31 in Americans (Fig. 2A), 26 in Africans (Fig. 2B), 28 in Europeans (Fig. 2C), and 24 in Asians (Fig. 2D).

population. Two recent studies have analyzed the association of *ACE2* polymorphisms with susceptibility to SARS-CoV-2 infection [42,43]; however, the evidence stating that low-frequency variants can participate in SARS-CoV-2 infection is not convincing. In the same way, Cao et al. [47] systematically investigated the candidate functional-coding variants in *ACE2* and the allele frequency differences between several populations. The results of this analysis suggested that there are no variants in the *ACE2* gene resistant to coronavirus S-protein binding in the study populations.

It was recently suggested that a renin-angiotensin system (RAS) imbalance impacts all stages of SARS-CoV-2 infection and clinical findings thereof, placing RAS molecules at the center of COVID-19 pathophysiology. The imbalance between the ACE/Ang II/AT1R and ACE2/Ang-(1-7)/MasR axes results in multiple organ dysfunction and uncontrolled inflammatory response [48]. The insertion/deletion (*I/D*) polymorphism of the *ACE1* gene is associated with plasma and tissue levels of ACE. In this context, Delanghe et al. [49] analyzed not only the prevalence and mortality data (per 1,000,000 inhabitants) of the COVID-19 infection of several countries but also the frequency of several polymorphisms in genes of some human plasma proteins, including the *ACE I/D* polymorphism. The results of this study suggest that the prevalence of COVID-19 is significantly correlated with the *ACE1* polymorphism.

Contrary to the *ACE2* gene, the polymorphisms in the *TMPRSS2* gene had a considerable variation in its frequencies between human populations. In this gene, we detected one polymorphism (rs12329760) located in the coding sequence that created a nonsynonymous substitution (Val160Met). Our in silico analysis using ModPred did not show a possible effect of the *TMPRSS2* rs12329760 polymorphism on any post-translational modification (e.g., proteolytic cleavage, acetylation, glycosylation, phosphorylation, and sulfation). However, this variant was predicted to be damaging by PolyPhen-2 and deleterious by SIFT. It has been recently reported that the *TMPRSS2* Val160Met variant decreases the stability of the protein, which might impede viral entry [50]. In a previous in silico analysis of the *TMPRSS2* gene, it was found that this polymorphism creates a de novo pocket protein [44]. The frequency of the minor allele of this polymorphism was high in the four study populations. Seventeen *TMPRSS2* polymorphisms (located in the promoter, in the 5' region near the gene, and the 3' UTR region) generated a possible functional effect: the binding of different transcription factors and microRNAs. Two of them had a high frequency of its minor allele in the four populations (rs4283504 and rs12626358) and created binding sites for the DBP, HSF1, NKX25, and KAISO factors. It has been reported that heat shock factor 1 (HSF1) is an innate repressor of HIV-induced inflammation [51]. The frequency of the minor allele of 7 of these polymorphisms was high in populations from the

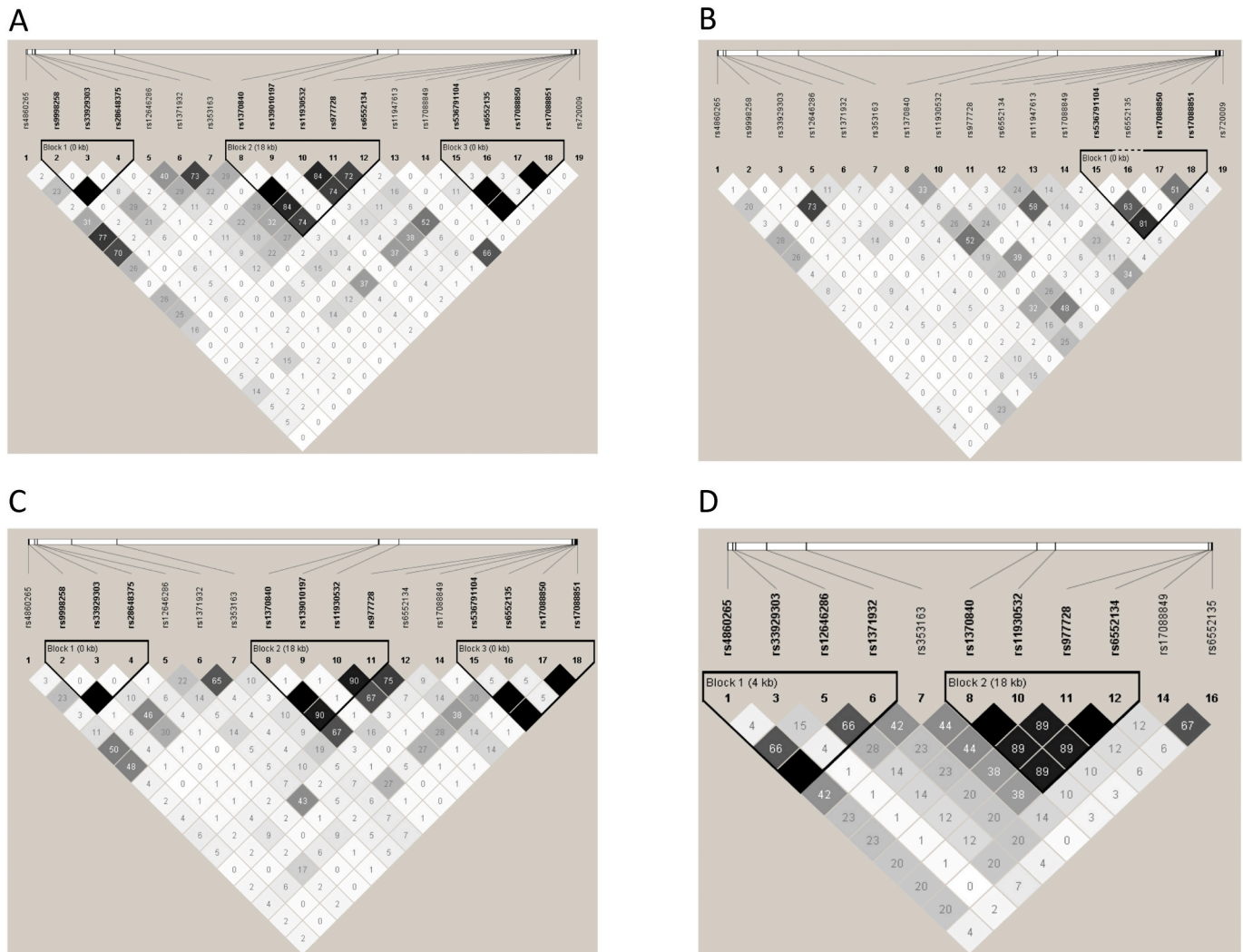


Fig. 3. Linkage disequilibrium (r^2) in the *TMPPSS11A* gene in the included populations. Of the 20 variants shown in Table 3, one was INDEL and was not added to the Haploview program. Of the remaining 19, some were not polymorphic in the different populations. Linkage disequilibrium between variants is shown in the figures, 19 in Americans (Fig. 3A), 17 in Africans (Fig. 3B), 17 in Europeans (Fig. 3C), and 11 in Asians (Fig. 3D).

American, African, and European continents. However, in the Asian population, only 3 (rs4283504, rs12626358, and rs8128074) out of the ten polymorphisms were observed with a minor allele frequency higher than 10%. In the same *TMPPSS2* gene, we detected 7 polymorphisms with a functional effect: all of them producing binding sites for microRNAs and two of them (rs456142 and rs456298) with high frequencies of its minor allele in the four study populations. The rs12627374 polymorphism produces a binding site for the microRNA-345. This polymorphism only was present in the Han Chinese population. Using computational analysis, we observed that it can affect a wide spectrum of microRNAs profile [44].

Although we identified that the three non-synonymous variants of *TMPPSS11A* are benign (according to PolyPhen-2) or tolerated (using SIFT), two of them (rs353163-Arg290Gln and rs13901019-Lys48Arg) possibly undergo a post-translational modification: proteolytic cleavage (according to ModPred server). Moreover, since this variant is located in the catalytic domain, it has been suggested that its activity is reduced because of the impact on the protein three-dimensional structure [52]. On the other hand, the rs353163 (Arg290Gln) polymorphism has been associated with the risk of esophageal squamous cell carcinoma [52]. Therefore, it is possible that these variants could affect viral entry;

however, future functional studies should be carried out to establish its role on SARS-CoV-2 susceptibility. The minor allele of the rs353163 polymorphism was present in a high frequency in the four study populations. Another 10 polymorphisms in this gene evoke a possible functional effect. Five located in the promoter and the 5' region near the gene and two in the 5' UTR region produced binding sites for several transcription factors. On the other hand, three polymorphisms located in the 3' UTR region created microRNAs binding sites. Out of these polymorphisms, the minor alleles of only 3 (rs17088849, rs6552134, and rs4860265) were present in high frequencies in the four populations. The minor alleles of three (rs17088850, rs17088851, rs720009) out of ten *TMPPSS11A* gene polymorphisms with a possible functional effect were seen in a high frequency only in the African population. It could be interesting to study the association of these polymorphisms with SARS-CoV-2 infection in African populations to define if they are related to the low infection rate in the continent.

In the *ELANE* gene that encodes the neutrophil elastase, 12 polymorphisms with possible functional effects were detected: ten in the promoter and the 5' region near the gene, and two in the 3' region near the gene. These twelve polymorphisms produced binding sites for several transcription factors and microRNAs. The minor allele of four of

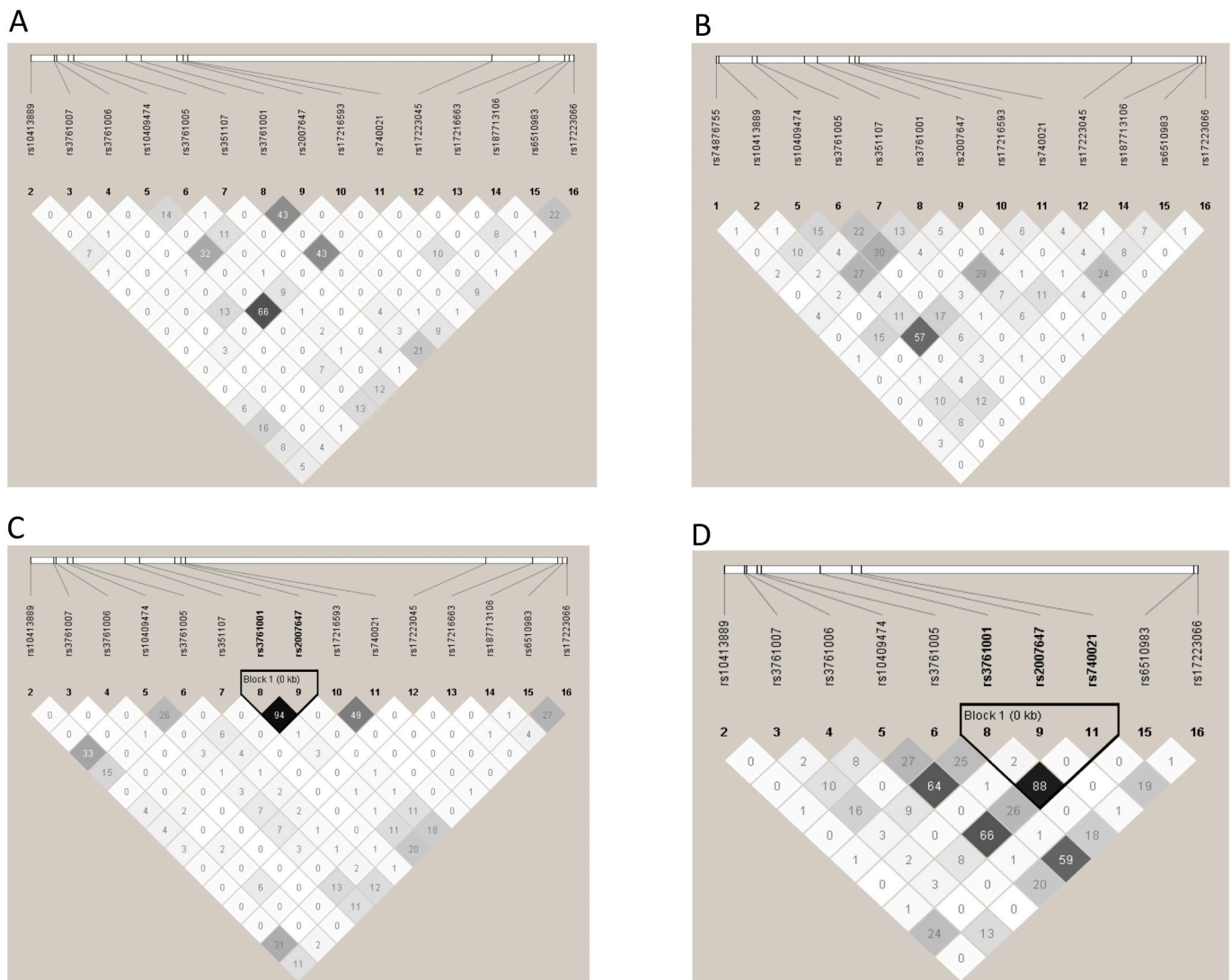


Fig. 4. Linkage disequilibrium (r^2) in the *ELANE* gene in the included populations. Of the 17 variants shown in Table 4, one was INDEL and was not added to the Haploview program. Of the remaining 16, some were not polymorphic in the different populations. Linkage disequilibrium between variants is shown in the figures, 15 in Americans (Fig. 4A), 13 in Africans (Fig. 4B), 15 in Europeans (Fig. 4C), and 10 in Asians (Fig. 4D).

these polymorphisms (rs10409474, rs3761005, rs3761001, rs17223066) was present in high frequency in the four populations. The minor allele of two polymorphisms (rs3761007 and rs3761006) had a high frequency only in the Han Chinese population. In a like manner, the minor allele frequency of rs2007647 was high only in the European (British) population. As for the SARS-CoV-2 infection, these polymorphisms could be relevant in the Asian and European populations. The nonsynonymous *ELANE* Pro257Leu (rs17216663) variant is predicted to be benign (PolyPhen-2) and tolerated (SIFT). Nevertheless, we found with the ModPred server that Pro257 may undergo hydrolyzation, which could affect the function of the *ELANE* protein. Previously, one study reported that Pro257Leu (located in the *ELANE* carboxyl terminus) is a risk factor for severe congenital neutropenia; however, the biological significance of this variant remains uncertain [53]. Therefore, future functional studies are essential to determine its effect.

In the *CTSL* gene, six polymorphisms with possible functional effects were detected. These polymorphisms were located in several regions of the gene and created binding sites for transcription factors. The minor allele of one of these polymorphisms (rs41307457) showed a high frequency only in the African population. Similarly, the minor allele of

rs41312184 was present in high frequency only in the European population. The association of these polymorphisms with the SARS-CoV-2 infection should be analyzed in these populations.

It is important to note the high heterogeneity in the different populations, which is evident in the linkage disequilibrium analysis that we carried out. Different linkage disequilibrium patterns were observed for each gene in each population. The above requires an adequate selection of the SNPs to be studied in each of the populations.

Of note, we did not evaluate potentially important variants located in the gene introns. Admittedly, some of these variants could have a role in producing different mRNAs and protein isoforms on these 5 genes. Even though the synonymous variants (substitutions that do not lead to an amino acid change) seem to have no functional effect on proteins, some authors have published an effect on the structure and function of them [54,55]. In our study, we included only information from the dbSNPs, Ensembl Genome Browser, and 1000 Genome Project databases. Discrete sequence databases of individuals infected with SARS-CoV2 were not analyzed. The phenotypic classification was not linked with the allelic patterns.

In summary, using web-based tools, we identified herein some

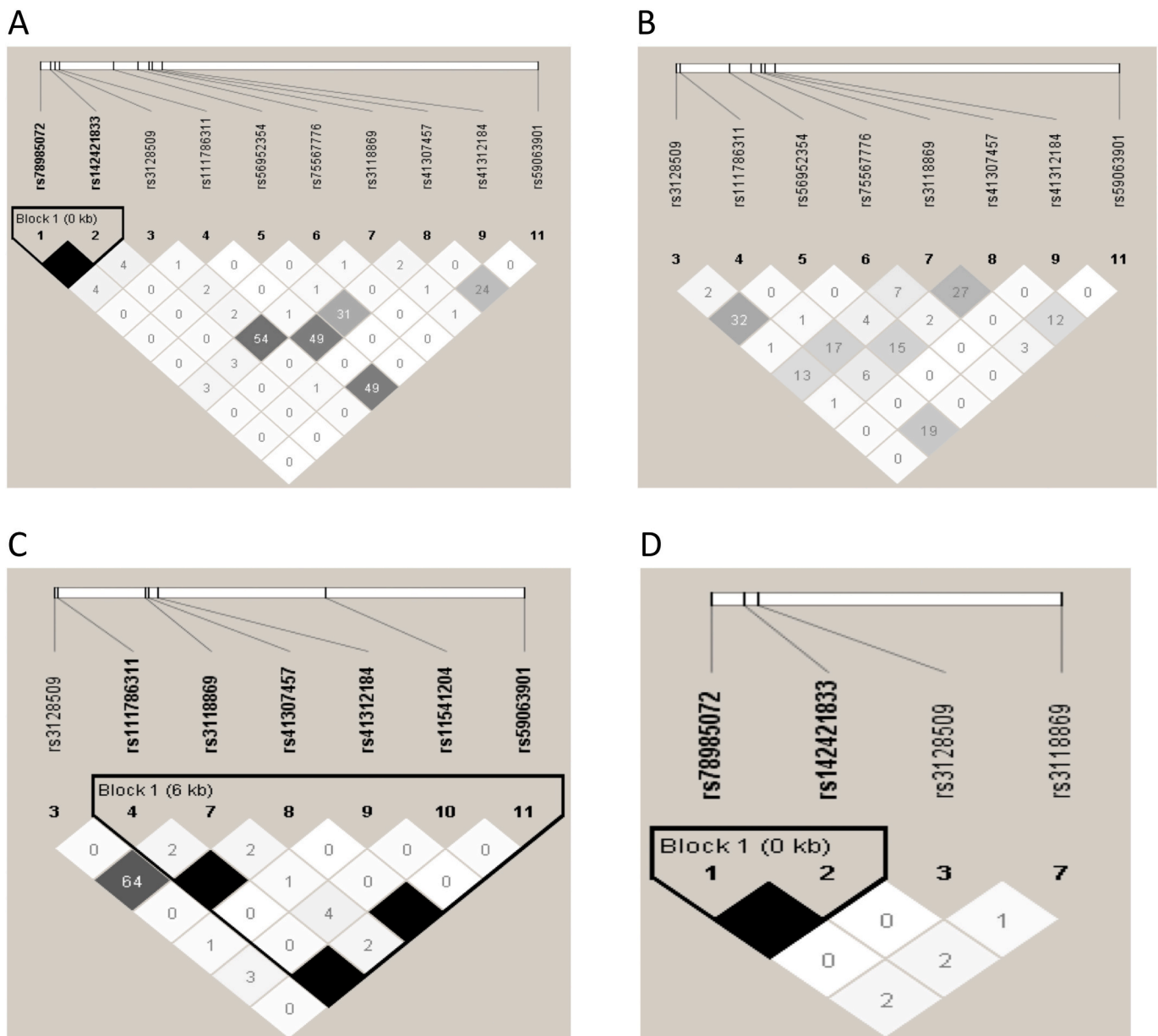


Fig. 5. Linkage disequilibrium (r^2) in the *CTSL* gene in the included populations. Of the 12 variants shown in Table 5, one was INDEL and was not added to the Haploview program. Of the remaining 11, some were not polymorphic in the different populations. Linkage disequilibrium between variants is shown in the figures, 10 in Americans (Fig. 5A), 8 in Africans (Fig. 5B), 7 in Europeans (Fig. 5C), and 4 in Asians (Fig. 5D).

polymorphisms in the genes that encode proteins related to the SARS-CoV-2 entry into the host cells that could be used for genetic association studies.

Declaration of competing interest

The authors declare no competing interests.

References

- [1] WorldHealthOrganization, Coronavirus disease 2019 (COVID-19) situation report – 51, https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10, (2020).
- [2] World Health Organization, Coronavirus disease 2019 (COVID-19) SituationReport – 133, https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200601-covid-19-sitrep-133.pdf?sfvrsn=9a56f2ac_4, (2020).
- [3] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, et al., Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, *Lancet* 395 (2020) 565–574, [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8).
- [4] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2, *Nat. Microbiol.* 5 (2020) 536–544, <https://doi.org/10.1038/s41564-020-0695-z>.
- [5] P. Zhou, X.L. Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature* 579 (2020) 270–273.
- [6] A.R. Fehr, S. Perlman, Coronaviruses: an overview of their replication and pathogenesis, *Methods Mol. Biol.* 1282 (2015) 1–23, https://doi.org/10.1007/978-1-4939-2438-7_1.
- [7] W. Li, M.J. Moore, N. Vasilieva, J. Sui, S.K. Wong, M.A. Berne, et al., Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus, *Nature* 426 (2003) 450–454.
- [8] J.K. Millet, G.R. Whittaker, Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 15214–15219, <https://doi.org/10.1073/pnas.1407087111>.
- [9] S. Bertram, R. Dijkman, M. Habjan, A. Heurich, S. Gierer, I. Glowacka, et al., TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium, *J. Virol.* 87 (2013) 6150–6160, <https://doi.org/10.1128/JVI.03372-12>.

- [10] C. Bhattacharyya, C. Das, A. Ghosh, A.K. Singh, S. Mukherjee, P.P. Majumder, et al., Global Spread of SARS-CoV-2 Subtype with Spike Protein Mutation D614G Is Shaped by Human Genomic Variations that Regulate Expression of *TMPRSS2* and *MX1* Genes, (2020), <https://doi.org/10.1101/2020.05.04.075911>.
- [11] S. Gierer, S. Bertram, F. Kaup, F. Wrensch, A. Heurich, A. Kramer-Kuhl, et al., The spike protein of the emerging betacoronavirus EMC uses a novel coronavirus receptor for entry, can be activated by *TMPRSS2*, and is targeted by neutralizing antibodies, *J. Virol.* 87 (2013) 5502–5511, <https://doi.org/10.1128/JVI.00128-13>.
- [12] Z. Qian, S.R. Dominguez, K.V. Holmes, Role of the spike glycoprotein of human Middle East respiratory syndrome coronavirus (MERS-CoV) in virus entry and syncytia formation, *PLoS One* 8 (2013) e76469, <https://doi.org/10.1371/journal.pone.0076469>.
- [13] K. Shirato, M. Kawase, S. Matsuyama, Middle East respiratory syndrome coronavirus infection mediated by the transmembrane serine protease *TMPRSS2*, *J. Virol.* 87 (2013) 12552–12561, <https://doi.org/10.1128/JVI.01890-13>.
- [14] Y. Shirogane, M. Takeda, M. Iwasaki, N. Ishiguro, H. Takeuchi, Y. Nakatsu, et al., Efficient multiplication of human metapneumovirus in Vero cells expressing the transmembrane serine protease *TMPRSS2*, *J. Virol.* 82 (2008) 8942–8946, <https://doi.org/10.1128/JVI.00676-08>.
- [15] J.E. Park, K. Li, A. Barlan, A.R. Fehr, S. Perlman, P.B. McCray, et al., Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 12262–12267, <https://doi.org/10.1073/pnas.1608147113>.
- [16] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, et al., SARS-CoV-2 cell entry depends on ACE2 and *TMPRSS2* and is blocked by a clinically proven protease inhibitor, *Cell* 181 (2020) 271–280, <https://doi.org/10.1016/j.cell.2020.02.052>.
- [17] X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, et al., Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV, *Nat. Commun.* 11 (2020) 1620, <https://doi.org/10.1038/s41467-020-15562-9>.
- [18] N.K. Biswas, P.P. Majumder, Analysis of RNA sequences of 3636 SARS-CoV-2 collected from 55 countries reveals selective sweep of one virus type, *Indian J. Med. Res.* (2020), https://doi.org/10.4103/ijmr.IJMR_1125_20 In press.
- [19] D.F. Gudbjartsson, A. Helgason, H. Jonsson, O.T. Magnusson, P. Melsted, G.L. Norddahl, et al., Spread of SARS-CoV-2 in the Icelandic population, *N. Engl. J. Med.* (2020), <https://doi.org/10.1056/nejmoa2006100> In press.
- [20] M. Donoghue, F. Hsieh, E. Baronas, K. Godbout, M. Gosselin, N. Stagliano, et al., A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1–9, *Circ. Res.* 87 (2000) E1–E9, <https://doi.org/10.1161/01.res.87.5.e1>.
- [21] V.B. Patel, J.C. Zhong, M.B. Grant, G.Y. Oudit, Role of the ACE2/angiotensin 1–7 axis of the renin-angiotensin system in heart failure, *Circ. Res.* 118 (2016) 1313–1326, <https://doi.org/10.1161/CIRCRESAHA.116.307708>.
- [22] L.T. McCollum, P.E. Gallagher, E.A. Tallant, Angiotensin-(1–7) attenuates angiotensin II-induced cardiac remodeling associated with upregulation of dual-specificity phosphatase 1, *Am. J. Physiol. Heart Circ. Physiol.* 302 (2012) H801–H810, <https://doi.org/10.1152/ajpheart.00908.2011>.
- [23] D.W. Lambert, M. Yarski, F.J. Warner, P. Thornhill, E.T. Parkin, A.I. Smith, et al., Tumor necrosis factor- α convertase (ADAM17) mediates regulated ectodomain shedding of the severe-acute respiratory syndrome-coronavirus (SARS-CoV) receptor, angiotensin-converting enzyme-2 (ACE2), *J. Biol. Chem.* 280 (2005) 30113–30119, <https://doi.org/10.1074/jbc.M505111200>.
- [24] H. Xia, S. Sriramula, K.H. Chhabra, E. Lazartigues, Brain angiotensin-converting enzyme type 2 shedding contributes to the development of neurogenic hypertension, *Circ. Res.* 113 (2013) 1087–1096, <https://doi.org/10.1161/CIRCRESAHA.113.301811>.
- [25] X. Zou, K. Chen, J. Zou, P. Han, J. Hao, Z. Han, Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection, *Front. Med.* 14 (2020) 185–192, <https://doi.org/10.1007/s11684-020-0754-0>.
- [26] F. Qi, S. Qian, S. Zhang, Z. Zhang, Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses, *Biochem. Biophys. Res. Commun.* 526 (2020) 135–140, <https://doi.org/10.1016/j.bbrc.2020.03.044>.
- [27] R.A. Santos, F. Frezard, A.J. Ferreira, Angiotensin-(1–7): blood, heart, and blood vessels, *Curr. Med. Chem. Cardiovasc. Hematol. Agent* 3 (2005) 383–391, <https://doi.org/10.2174/156801605774322373>.
- [28] Q. Zhang, M. Cong, N. Wang, X. Li, H. Zhang, K. Zhang, et al., Association of angiotensin-converting enzyme 2 gene polymorphism and enzymatic activity with essential hypertension in different gender a case-control study, *Medicine* 97 (2018) e12917, <https://doi.org/10.1097/MD.00000000000012917>.
- [29] C. Liu, Y. Li, T. Guan, Y. Lai, Y. Shen, A. Zeyaweidung, et al., ACE2 polymorphisms associated with cardiovascular risk in Uyghurs with type 2 diabetes mellitus, *Cardiovasc. Diabetol.* 17 (2018) 127, <https://doi.org/10.1186/s12933-018-0771-3>.
- [30] K. Sakai, Y. Ami, M. Tahara, T. Kubota, M. Anraku, M. Abe, et al., The host protease *TMPRSS2* plays a major role *in vivo* replication of emerging H7N9 and seasonal influenza viruses, *J. Virol.* 88 (2014) 5608–5616, <https://doi.org/10.1128/JVI.03677-13>.
- [31] I. Glowacka, S. Bertram, M.A. Muller, P. Allen, E. Soilleux, S. Pfeifferle, et al., Evidence that *TMPRSS2* activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response, *J. Virol.* 85 (2011) 4122–4134, <https://doi.org/10.1128/JVI.02232-10>.
- [32] C. Tarnow, G. Engels, A. Arendt, F. Schwalm, H. Sediri, A. Preuss, et al., *TMPRSS2* is a host factor that is essential for pneumotropism and pathogenicity of H7N9 influenza virus in mice, *J. Virol.* 88 (2014) 4744–4751, <https://doi.org/10.1128/JVI.03799-13>.
- [33] S. Matsuyama, N. Nao, K. Shirato, M. Kawase, S. Saito, I. Takayama, et al., Enhanced isolation of SARS-CoV-2 by *TMPRSS2*-expressing cells, *Proc. Natl. Acad. Sci. U. S. A.* 117 (2020) 7001–7003, <https://doi.org/10.1073/pnas.2002589117>.
- [34] Z. Cheng, J. Zhou, K. Kai-Wang To, H. Chu, C. Li, D. Wang, et al., Identification of *TMPRSS2* as a susceptibility gene for severe 2009 pandemic A(H1N1) influenza and A(H7N9) influenza, *J. Infect. Dis.* 212 (2015) 1214–1221, <https://doi.org/10.1093/infdis/jiv246>.
- [35] T.M. Antalis, T.H. Bugge, Q. Wu, Membrane-anchored serine proteases in health and disease, *Prog. Mol. Biol. Transl. Sci.* 99 (2011) 1–50, <https://doi.org/10.1016/B978-0-12-385504-6.00001-4>.
- [36] P. Zmora, M. Hoffmann, H. Kollmus, A.S. Moldenhauer, O. Danov, A. Braun, et al., *TMPRSS11A* activates the influenza A virus hemagglutinin and the MERS coronavirus spike protein and is insensitive against blockade by HAI-1, *J. Biol. Chem.* 293 (2018) 13863–13873, <https://doi.org/10.1074/jbc.RA118.001273>.
- [37] A.J. Barrett, N.D. Rawlings, J.F. Woessner, *Handbook of Proteolytic Enzymes*, Elsevier Academic Press, London, 2004.
- [38] K. Kawabata, T. Hagio, S. Matsuoka, The role of neutrophil elastase in acute lung injury, *Eur. J. Pharmacol.* 451 (2002) 1–10, [https://doi.org/10.1016/S0014-2999\(02\)02182-9](https://doi.org/10.1016/S0014-2999(02)02182-9).
- [39] S. Hashimoto, Y. Okayama, N. Shime, A. Kimura, Y. Funakoshi, K. Kawabata, et al., Neutrophil elastase activity in acute lung injury and respiratory distress syndrome, *Respirology* 13 (2008) 581–584, <https://doi.org/10.1111/j.1440-1843.2008.01283.x>.
- [40] H. Kirschke, L. Cathepsin, *Handb. Proteolytic Enzym* (2013) 1808–1817, <https://doi.org/10.1016/B978-0-12-382219-2.00410-5>.
- [41] H.A. Elshabrawy, J. Fan, C.S. Haddad, K. Ratia, C.C. Broder, M. Caffrey, et al., Identification of a broad-spectrum antiviral small molecule against severe acute respiratory syndrome coronavirus and Ebola, Hendra, and Nipah viruses by using a novel high-throughput screening assay, *J. Virol.* 88 (2014) 4353–4365, <https://doi.org/10.1128/JVI.03050-13>.
- [42] Stawiski EW, Diwanji D, Suryamohan K, Gupta R, Fellouse FA, Sathirapongsasuti JF, et al., Human ACE2 receptor polymorphisms predict SARS-CoV-2 susceptibility, bioRxiv, in press, doi:<https://doi.org/10.1101/2020.04.07.024752>.
- [43] Renieri A, Benetti E, Tita R, Spiga O, Ciolfi A, Birolo G, et al., ACE2 variants underlie interindividual variability and susceptibility to COVID-19 in Italian population, medRxiv, in press, doi:<https://doi.org/10.1101/2020.04.03.20047977>.
- [44] A. Paniri, M.M. Hosseini, H. Akhavan-Niaki, First comprehensive computational analysis of functional consequences of *TMPRSS2* SNPs in susceptibility to SARS-CoV-2 among different populations, *Biomol Struct Dyn* 15 (2020) 1–18, <https://doi.org/10.1080/07391102.2020.1767690>.
- [45] B. Zhu, R. Zhang, C. Li, L. Jiang, M. Xiang, Z. Ye, et al., BCL6 modulates tissue neutrophil survival and exacerbates pulmonary inflammation following influenza virus infection, *Proc. Natl. Acad. Sci. U. S. A.* 116 (2019) 11888–11893, <https://doi.org/10.1073/pnas.1902310116>.
- [46] M. Han, W. Yan, Y. Huang, H. Yao, Z. Wang, D. Xi, et al., The nucleocapsid protein of SARS-CoV induces transcription of hfg12 prothrombinase gene dependent on C/EBP alpha, *J. Biochem.* 144 (2008) 51–62, <https://doi.org/10.1093/jb/mvn042>.
- [47] Y. Cao, L. Li, Z. Feng, S. Wan, P. Huang, X. Sun, et al., Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations, *Cell Discov* 6 (2020) 11, <https://doi.org/10.1038/s41421-020-0147-1>.
- [48] Lanza K, Perez LG, Costa LB, Cordeiro TM, Palmeira VA, Ribeiro VT, et al. Covid-19: the renin-angiotensin system imbalance hypothesis. *Clin Sci (Lond)*. 2020; 134: 1259–1264. <https://doi.org/10.1042/CS20200492>.
- [49] J.E. Delanghe, M.M. Speeckaert, M.L. De Buyzere, COVID-19 infections are also affected by human ACE1 D/I polymorphism, *Clin. Chem. Lab. Med.* 58 (2020) 1125–1126, <https://doi.org/10.1515/cclm-2020-0425>.
- [50] Vishnubhotla R, Vankadari N, Ketavarapu V, Amanchy R, Avanthi S, Bale G, et al., Genetic variants in *TMPRSS2* and structure of SARS-CoV-2 spike glycoprotein and *TMPRSS2* complex. bioRxiv preprint doi: <https://doi.org/10.1101/2020.06.30.179663>.
- [51] X. Pan, J. Lin, X. Zeng, W. Li, W. Wu, W.Z. Lu, et al., Heat shock factor 1 suppresses the HIV-induced inflammatory response by inhibiting nuclear factor- κ B, *Cell. Immunol.* 327 (2018) 26–35, <https://doi.org/10.1080/07391102.2020.1767690>.
- [52] Li Y, Zhang X, Huang G, Miao X, Guo L, Lin D, et al. Identification of a novel polymorphism Arg290Gln of esophageal cancer related gene 1 (ECRG1) and its related risk to esophageal squamous cell carcinoma. *Carcinogenesis*. 2006; 27: 798–802. <https://doi.org/10.1093/carcin/bgi258>.
- [53] L. Beene, B. Xin, C. Lukas, H. Wang, Mutations in ELANE and COH1 (VPS13B) genes cause severe neutropenia in a patient with cohen syndrome, *J Clin Cell Immunol* 6 (2015) 6, <https://doi.org/10.4172/2155-9899.1000378>.
- [54] C. Kimchi-Sarfaty, J.M. Oh, I.W. Kim, Z.E. Sauna, A.M. Calcagno, S.V. Ambudkar, et al., A “silent” polymorphism in the *MDR1* gene changes substrate specificity, *Science* 315 (2007) 525–528, <https://doi.org/10.1126/science.1135308>.
- [55] J. Ramírez-Bello, M. Jiménez-Morales, Functional implications of single nucleotide polymorphisms (SNPs) in protein-coding and non-coding RNA genes in multifactorial diseases, *Gac. Med. Mex.* 153 (2017) 238–250.