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⁎, Rosalinda Posadas-Sánchez, Julian Ramírez-Bello

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

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
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 azymogen 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/ACE2 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.
| Variant ID | Minor allele | AMR (%) | AFR (%) | EUR (%) | EAS (%) | Amino acid position and change | Potential functional effect |
|------------|--------------|---------|---------|---------|---------|-------------------------------|----------------------------|
| rs6175794C/T | T            | 0.8     | 0       | 2.8     | 0       |                               |                            |
| rs2298659G/A | A            | 31.2    | 16.2    | 21.4    | 29.6    | Gly250Gly                      |                            |
| rs17854725A/G | G            | 47.7    | 36.1    | 56.4    | 17.5    | Ile256Ile                      |                            |
| rs61175796G/A | A            | 1.6     | 0       | 0.5     | 0       | Tyr140Iyr                      |                            |
| rs1232976C/T | T            | 18.0    | 25.9    | 20.9    | 41.3    |                               |                            |
| rs3787950T/C | C            | 1.6     | 30.1    | 7.1     | 11.7    | Thr27Thr                       |                            |
| rs6175793G/A | A            | 0       | 0       | 1.1     | 0       | Pro63Pro                       |                            |
| rs4303794A/C | C            | 28.1*   | 41.2**  | 41.8†   | 1‡      |                               | The C allele creates binding sites for AP2, and SP1, and WT1 |
| rs11088551A/G | G            | 28.1*   | 41.2**  | 41.8†   | 1‡      |                               | The A allele creates binding sites for BRCA, MYB, NF1, and RFX |
| rs6678213GCGCGCGGCGC/C (INDEL) | C | 28.1 | 41.2** | 41.8† | 1‡ |                               |                            |
| rs4303795A/G | G            | 28.1*   | 41.2**  | 41.8†   | 1‡      |                               |                            |
| rs5844077G/A (INDEL) | G | 29.6 | 25.3 | 25.8 | 8.3 |                               |                            |
| rs7683541G/A | A            | 31.3    | 0       | 10.4    | 0       |                               |                            |
| rs2483504G/T | T            | 16.4    | 37.0    | 12.6    | 23.3    |                               |                            |
| rs1249198T/C | C            | 27.3*   | 39.4**  | 40.7†   | 1‡      |                               | The T allele creates binding sites for DBP, HSF1, and NXX25 |
| rs2870750G/A | A            | 25.8*   | 38.4**  | 40.1†   | 1‡      |                               | The C allele creates a binding site for HNF31/47 |
| rs5525742AGCTG (INDEL) | CT | 26.6 | 38.4 | 39.0 | 1.5 |                               | The G allele creates binding sites for HNF3, ALPHALPHA, and TBP |
| rs1262635R/C | G            | 26.8    | 26.4    | 9.3     | 56.8    |                               | The G allele creates a binding site for KASO |
| rs8128074C/T | T            | 16.4    | 2.8     | 12.6    | 23.8    |                               | The C allele creates binding sites for ETF, KROX, LRF, and SPZ1 |
| rs5621846A/G | A            | 25.8*   | 40.3**  | 40.1†   | 1‡      |                               | The A allele creates a binding site for PPARG |
| rs11281229TCTCAGG (INDEL) | TCCAGG | 25.8 | 40.3 | 40.1 | 0.9 |                               |                            |
| rs8127674G/A | 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       |                               |                            |
| rs453727A/T | A            | 26.6    | 4.2     | 46.2    | 0.5     |                               |                            |
| rs462471G/A | A            | 36.7*** | 34.3**  | 13.7**  | 53.4‡‡  |                               |                            |
| rs7600363G/A | A            | 12.5**  | 5.6     | 12.1‡‡  | 6.3**   |                               |                            |
| 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 |
| rs456142C/T | T            | 36.7*** | 34.3**  | 13.7**  | 53.4‡‡  |                               | The T allele can create a binding site for hsa-miR-943 |
| rs12863489C/T | T            | 0.8     | 7.9     | 0       | 6.3     |                               | The A allele can create a binding site for hsa-miR-1324 |
| rs462574G/A | A            | 24.2    | 14.8**  | 1.7     | 47.1‡‡  |                               | The A allele can create a binding site for hsa-miR-450b-5p |
| rs546298A/T | T            | 37.5*** | 34.3**  | 13.7**  | 53.4‡‡  |                               | The A alleles can create a binding site for hsa-miR-220b |
| rs1700104G/A | A            | 0.8     | 13.9    | 0       | 0       |                               |                            |
| rs11910678T/C | C            | 1.6     | 13.9**  | 6.3     | 6.3‡‡   |                               |                            |
| rs77975406G/A | A            | 12.5**  | 4.6     | 12.1‡‡  | 6.3‡‡   |                               |                            |
| rs1262374C/T | T            | 0       | 0       | 0       | 13.6‡‡  |                               | The C allele can create a binding site for hsa-miR-345 |
| rs62221752C/T | T            | 3.9     | 0       | 6.0     | 0       |                               | The C allele can create a binding site for hsa-miR-1226 |
| rs77975406G/A | A            | 4.6     | 0.8     | 0       | 0       |                               |                            |
| rs149695191GT/T (DELETION) | T | 22.7 | 0.8 | 0 | 0 |                               |                            |

(continued on next page)
Among all the genes analyzed, we included polymorphisms with frequencies greater than 1.5% in at least one of the four populations from the American, African, European, and Asian continents.

### 3.1. **ACE2 polymorphisms**

Table 1 shows thirteen ACE2 polymorphisms with a frequency greater than 1.5% in at least one of the populations reviewed. As can be seen, two of these polymorphisms (rs35803318 and rs4646179) were located in the coding sequence without a change of amino acid. On the
thirty-nine polymorphisms had a frequency higher than 1.5% in at least three populations with different ancestry. Ten polymorphisms located in the promoter and the 5′ region of the 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.

### 3.2. **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).

### 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, rs977728-Met11le). According to PolyPhen-2 results, rs353163 was possibly 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.
3.4. 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 0.015, sensitivity 0.96, specificity 0.79) and tolerated (SIFT score 0.57, low confidence). Alternatively, we did not identify any effect of Met1 or 1IIe (rs977728).

Both five 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 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).

3.5. 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 4

| Variant ID | MAF (%) in populations with different ancestry |
|------------|-----------------------------------------------|
| Coding sequence | Minor allele AMER (MXL) AFR (YRI) EUR (GBR) EAS (CHB) |
| rs17223045/T | T | 0.8 | 11.6 | 1.1 | 0 |
| rs17216663/T | T | 1.6 | 0 | 0.6 | 0 |

#### Potential functional effect

- Amn130Amn Pro257Leu |

---

**Table 3**

| Coding sequence | Minor allele AMER (MXL) AFR (YRI) EUR (GBR) EAS (CHB) |
|-----------------|----------------------------------------------------------|
| rs17223045/T    | T 0.8 | 11.6 | 1.1 | 0 |
| rs17216663/T    | T 1.6 | 0   | 0.6 | 0 |

#### Potential functional effect

- Amn130Amn Pro257Leu |

---

**Table 4**

| ELANE polymorphisms. | MAF (%) in populations with different ancestry | Potential functional effect |
|----------------------|-----------------------------------------------|-----------------------------|
| Variant ID | Minor allele AMER (MXL) AFR (YRI) EUR (GBR) EAS (CHB) |
| Coding sequence | | |
| rs17223045/T | T | 0.8 | 11.6 | 1.1 | 0 |
| rs17216663/T | T | 1.6 | 0 | 0.6 | 0 |

#### Potential functional effect

- Amn130Amn Pro257Leu |

---

**Table 5**

| Variant ID | Minor allele AMER (MXL) AFR (YRI) EUR (GBR) EAS (CHB) |
|-----------------|----------------------------------------------------------|
| rs17223045/T    | T 0.8 | 11.6 | 1.1 | 0 |
| rs17216663/T    | T 1.6 | 0 | 0.6 | 0 |

#### Potential functional effect

- Amn130Amn Pro257Leu |

---

**Table 6**

| Variant ID | Minor allele AMER (MXL) AFR (YRI) EUR (GBR) EAS (CHB) |
|-----------------|----------------------------------------------------------|
| rs17223045/T    | T 0.8 | 11.6 | 1.1 | 0 |
| rs17216663/T    | T 1.6 | 0 | 0.6 | 0 |

#### Potential functional effect

- Amn130Amn Pro257Leu |

---

**Table 7**

| Variant ID | Minor allele AMER (MXL) AFR (YRI) EUR (GBR) EAS (CHB) |
|-----------------|----------------------------------------------------------|
| rs17223045/T    | T 0.8 | 11.6 | 1.1 | 0 |
| rs17216663/T    | T 1.6 | 0 | 0.6 | 0 |

#### Potential functional effect

- Amn130Amn Pro257Leu |

---

**Table 8**

| Variant ID | Minor allele AMER (MXL) AFR (YRI) EUR (GBR) EAS (CHB) |
|-----------------|----------------------------------------------------------|
| rs17223045/T    | T 0.8 | 11.6 | 1.1 | 0 |
| rs17216663/T    | T 1.6 | 0 | 0.6 | 0 |

#### Potential functional effect

- Amn130Amn Pro257Leu |

---

**Table 9**

| Variant ID | Minor allele AMER (MXL) AFR (YRI) EUR (GBR) EAS (CHB) |
|-----------------|----------------------------------------------------------|
| rs17223045/T    | T 0.8 | 11.6 | 1.1 | 0 |
| rs17216663/T    | T 1.6 | 0 | 0.6 | 0 |

#### Potential functional effect

- Amn130Amn Pro257Leu |


| Variant ID | Minor allele | AMER (MXL) | AFR (YRI) | EUR (GBR) | EAS (CHB) | Amino acid position and change | Potential functional effect |
|------------|--------------|------------|-----------|-----------|-----------|-------------------------------|-----------------------------|
| Coding sequence |              |            |           |           |           |                               |                             |
| rs11541204G/A | A            | 0          | 0         | 5.0       | 0         |                               | Gln134Gln                   |
| rs14242183G/T | T            | 4.7*       | 0         | 0         | 15.5*     |                               | X                           |
| rs14242183C/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                           |
| rs11786311G/T | C            | 1.6        | 16.2      | 2.8†       | 0         |                               | X                           |
| rs1389221C/GAA (INDEL) | C/GAA | 43.8       | 15.3      | 51.7      | 76.7      | Both alleles can create binding sites for BRCA, DBP, LRF, MYB, and STAT4 | X                           |
| rs5695235AA/T | T            | 2.3        | 4.6       | 0         | 0         | Both alleles can create binding sites for GATA, GFI1, and TEL2 | X                           |
| rs75567766G/C | C            | 2.3        | 7.9       | 0         | 0         |                               | X                           |
| rs3118869G/A | A            | 39.8       | 46.8      | 47.3      | 32.5      | The C allele creates binding sites for SREBP, AHR, and AHRHF | X                           |
| rs4130745G/C | A            | 3.1        | 23.6      | 2.8†       | 0         | Both alleles can create binding sites for BRCA, DBP, LRF, MYB, and STAT4 | X                           |
| rs5969930G/A | A            | 0.8        | 3.7       | 2.8†       | 0         | Both alleles can create binding sites for STAT, and RFX. The C allele can create a binding site for SF2ASF1 | X                           |

* Variants in high LD or are tagSNPs between them in an American, European, and Asian population, respectively.
†, ‡, ±  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, CDS; consensus coding sequence.

CTSL is located on chromosome 9q21.33. Six transcripts have been reported for CTSL, 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.
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, EFF, 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. 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.
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 $ACE1$ 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
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 *TMPRSS2* 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 *TMPRSS11A* 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 *TMPRSS11A* 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
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 \textit{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 Pro257 (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 \textit{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
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] World Health Organization, 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) Situation Report – 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.

[4] Coronavirus 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.1407067111.

[9] S. Bertram, R. Dijkman, M. Habjan, A. Heinrich, 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.02372-12.
X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, et al., Characterization of spike glycoprotein of SARS-CoV-2, J. Virol. 87 (2013) 5502–5511, https://doi.org/10.1128/JVI.01218-13.

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) e74669, https://doi.org/10.1371/journal.pone.0074669.

K. Shirato, M. Kawase, S. Matsuuya, Middle East respiratory syndrome coronavirus infection mediated by the transmembrane serine protease TMEM234, J. Virol. 87 (2013) 12552–12561, https://doi.org/10.1128/JVI.01890-13.

M. Shigourie, M. Takeda, M. Ishikawa, N. Ishiguro, H. Takeuchi, Y. Nakatani, et al., Efficient multiplication of human metapneumovirus in Vero cells expressing the transmembrane serine protease TMEM234, J. Virol. 82 (2008) 8942–8946, https://doi.org/10.1128/JVI.00676-08.

J.E. Park, K. Li, A. Barlan, A.R. Fehr, S. Perlman, P.B. McGary, 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.1512977113.

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.

C. Liu, Y. Li, T. Guan, Y. Lai, Y. Shen, A. Zeyaweiding, et al., ACE2 polymorphisms (SNPs) in protein-coding and non-coding RNA genes in multiple populations, Life Sciences 260 (2020) 118313, https://doi.org/10.1016/j.lfs.2020.118313.

L.T. McCollum, P.E. Gallagher, E.A. Tallant, Angiotensin-(1-7) attenuates angio- and vaso-constriction in the hamster, Curr. Med. Chem. Cardiovasc. Hematol. Agent 3 (2005) 383–391, https://doi.org/10.2174/1567201053715894.

Y. Shirogane, M. Takeda, M. Iwasaki, N. Ishiguro, H. Takeuchi, Y. Nakatani, et al., Mutations in ELANE and COH1 (VPS13B) genes associated with neutrophil elastase activity in acute lung injury and respiratory distress syndrome, Respirology (2020), https://doi.org/10.1111/res.13769.

K. Shirato, M. Kawase, S. Matsuuya, 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.

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, Cardiovasc. Diabetol. 17 (2018) 127, https://doi.org/10.1186/s12933-018-0771-3.

F. Qi, S. Qian, S. Zhang, Z. Zhang, Single cell RNA sequencing of 13 human tissues and organs, Front. Bioeng. Biotechnol. 8 (2020) 1–10, https://doi.org/10.3389/fbioe.2020.00012.

T.M. Antalis, T.H. Bugge, Q. Wu, Membrane-anchored serine proteases in health and disease, Prog. Mol. Biol. Transl. Sci. 99 (2011) 1–37, https://doi.org/10.1016/B978-0-12-385504-6.00004-4.

P. Zmora, M. Hoffmann, H. Kollum, A.S. Moldenhauer, O. Danov, A. Braun, et al., TMPRSS11A activates the influenza 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. https://doi.org/10.1074/jbc.RA118.001273.

A.J. Barrett, N.D. Rawlings, J.F. Woensm, Handbook of Proteolytic Enzymes, Elsevier Academic Press, London, 2004.

K. Kawabata, Y. 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.

H. Xia, S. Sriramula, K.H. Chhabra, E. Lazartigues, Brain angiotensin-converting enzyme (ACE) conversion of angiotensin I to angiotensin 1–9, Circ. Res. 87 (2000) E1–E9, https://doi.org/10.1161/01.RES.87.5.e1.

C. Bhattacharyya, C. Das, A. Ghosh, A.K. Singh, S. Mukherjee, P.P. Majumder, 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.02232-14.

D. Han, Y. Han, Y. Huang, Y. Yao, Z. Wang, D. Li, et al., Evaluation of the clinical significance and role of neutrophil elastase in acute lung injury, J. Clin. Lab. Anal. 21 (2007) 87–92, https://doi.org/10.1002/jcla.20141.

Z. Cheng, J. Zhou, K. Kai-Wang To, H. Cho, 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/jiu426.

B. Zou, 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. 119 (2020) 11888–11893, https://doi.org/10.1073/pnas.1902310116.

H. Arikawa, Y. Kuwabara, K. Kuramochi, T. Horii, S. Kondo, Y. Ueda, et al., The role of neutrophil elastase in the pathogenesis of sepsis, Shock 36 (2011) 254–260, https://doi.org/10.1097/SHK.0b013e318210a83b.

C. Kimchi-Sarfaty, J.M. Oh, I.W. Kim, Z.E. Sauna, A.M. Calcagno, S.V. Ambudkar, et al., 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.