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.
Disparities in COVID-19 severities and casualties across ethnic groups around the globe and patterns of ACE2 and PIR variants

Sabrina Samad Shoily a,1, Tamim Ahsan b,1, Kaniz Fatema a, Abu Ashfaqur Sajib a,*

a Department of Genetic Engineering & Biotechnology, University of Dhaka, Dhaka 1000, Bangladesh
b Department of Mathematics and Natural Sciences, Ibrac University, Dhaka 1212, Bangladesh

ARTICLE INFO

Keywords:
COVID-19
SARS-CoV-2
ACE2
PIR
Genetic variants
Linkage disequilibrium
Haplotype
Geographic distribution

ABSTRACT

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) mediated Coronavirus disease-19 (COVID-19) has affected millions of individuals around all corners of the globe. Symptoms and severities of infection with this highly contagious virus vary among individuals and there is disparity in the number of COVID-19 related casualties across different ethnic groups. The primary receptor for SARS-CoV-2 entry into the host cells is angiotensin-converting enzyme 2 (ACE2). Certain variants of ACE2 are known to be associated with COVID-19 comorbidities such as hypertension, cardiovascular complications, diabetes, chronic lung disease, etc. In this study, we looked into the geographic distribution of disease-associated variants of ACE2 as well as closely located PIR gene to explore any possible correlation with the disparities in COVID-19 severities and casualties across ethnic groups. Frequencies of the ACE2 variants associated with COVID-19 comorbidities are higher in the European and the admixed American populations. These variants are also present with stronger pairwise linkage disequilibrium (LD) in the European and the admixed American populations. On the other hand, the variants with protective role are more prevalent in the East and the South Asian populations. Strong pairwise LD exists among the activity modifying (modifier) variants of the ACE2 and PIR genes only in the European super-population. Absence of these PIR variants in the South Asian population may contribute to the overall lower COVID-19 case fatality rates (CFR) despite the dense population in this region.

INTRODUCTION

SARS-CoV-2 mediated COVID-19 has emerged as one of the most widespread and frightful pandemics in human history (Devaux et al., 2020; Esakandari et al., 2020). Symptoms and severities of infection with SARS-CoV-2 vary among individuals. While some of the infected individuals may be asymptomatic carriers or exhibit mild flu-like symptoms, the others may develop severe symptoms and progress to pneumonia, respiratory failure as well as death (Alimohamadi et al., 2020; Feng et al., 2020b; Wu and McGoogan, 2020). Till April 12, 2021 COVID-19 has spread in 215 countries and regions on earth with over 135,646,500 confirmed cases of infection and more than 2,930,500 deaths. There is disparity in the number of COVID-19 caused mortalities across ethnic groups in different geographic regions around the globe (World Health Organization, 2020; Yamamoto and Bauer, 2020). In addition to environmental, socioeconomic and regulatory factors, genetic variants may be responsible for the observed interpopulation and interethnic variability.

The primary receptor for SARS-CoV-2 entry into the host cell is the angiotensin-converting enzyme 2 (ACE2) (Davidson Anne et al., 2020; Lan et al., 2020). As SARS-CoV-2 occupies the ACE2 molecules on cell surface, it prevents ACE2-mediated degradation of angiotensin II (Ang II), which consequently raises Ang II level (Hirano and Murakami, 2020). Increased levels of tissue and circulating Ang II exert a number of direct pro-atherosclerotic effects and promotes monocyte and endothelial cell activation causing progression to atherosclerotic plaque formation (Thomas Merlin et al., 2010). Ang II induces production of proinflammatory cytokines and chemokines via Toll-like receptor 4 (TLR4) (Feng et al., 2020a). Alongside the lungs, SARS-CoV-2 invades other organs that express high level of ACE2. SARS-CoV-2 mediated damage occurs directly by viral propagation and indirectly by inflammatory and immune factors as well as imbalance between the renin angiotensin system (RAS) and ACE2/angiotensin-(1–7) axis caused by ACE2 deficiency (Ni et al., 2020). Genetic factors significantly affect the
circuiting level of ACE2 in individuals (Rice et al., 2006). The expression level of ACE2 correlates positively with COVID-19 susceptibility, whereas down-regulation of ACE2 activity exacerbate Ang II mediated inflammatory events (Behl et al., 2020).

The highly prevalent comorbidities in COVID-19 patients are hypertension, diabetes, cardiovascular complication and chronic lung disease, among which hypertension is the most common (Sanyaolu et al., 2020). COVID-19 patients with diabetes and hypertension often exhibit severe symptoms, and percentage of non-survivors is high in this group of patients (Richardson et al., 2020; Singh et al., 2020; Zhou et al., 2020). Hypertension per se is strongly associated with diabetes and other complications (Adler et al., 2000; de Boer et al., 2017). On the other hand, ACE2 variants are associated with hypertension, hypertension-related atrial fibrillation and left atrial remodeling (Lozano-Gonzalez et al., 2020; Luo et al., 2019; Patel et al., 2012).

This study explored the geographic distribution of disease associated variants of ACE2 as well as other genes in its vicinity to understand the interethnic variability in the susceptibility and severity of SARS-CoV-2 mediated COVID-19.

2. Materials and methods

2.1. Retrieval of disease-associated variants

ACE2 gene variants, that are associated with any disease, were retrieved from the DisGeNET database (version 7) (Pinero et al., 2020). The search was conducted using “ACE2” gene (Entrez Id: 59272) as the keyword. Redundant variants were removed from the list. Regulatory roles of the variants were retrieved from rsSNPBase 3.1 (Guo and Wang, 2018). All filtering options (i.e. TF binding regions, chromatin interactive regions, topologically associated domains, IncRNAs coding regions, mature miRNAs regions, target sites of miRNAs and circRNA regions) were selected to identify any type of regulatory element.

2.2. Haplotype and linkage disequilibrium analysis

Frequencies of disease-associated ACE2 gene variants in 26 populations (Supplementary Table 1) across 5 super-populations (African, admixed American, East Asian, European and South Asian) were collected from the 1000 Genomes Database (1000 Genomes Project Consortium et al., 2015) using the Ensembl Allele Frequency Calculator based on the phase 3 data (Yates et al., 2020). Pairwise linkage disequilibrium (LD) among these variants and frequencies of haplotypes comprising these variants in the five super-populations were calculated using LDmatrix and LDhap modules of LDlink (version 5.0), respectively (Machiela and Chanock, 2015).

2.3. Identification of other modifier variants

LDproxy module of LDlink (Machiela and Chanock, 2015) was used to identify the genetic variants in absolute LD ($r^2 = 1$) with any of the disease-associated variants in ACE2 within ±0.5 Mb (total 1mb) region in any of the super-populations. These SNPs were mapped to several genes. SNPs that are supposed to modify the activity (modifier) of the corresponding gene products were identified and selected using Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016). Frequencies of these modifier variants were retrieved from the 1000 Genomes Database using the Ensembl Allele Frequency Calculator based on the phase 3 data.

2.4. Identification of candidate genes associated with inflammation

UniProt database (Pundir et al., 2017) was searched to identify whether any of these genes is directly or indirectly associated with inflammation. Among the identified genes, only PIR was found to be associated with inflammation. LDmatrix was used to identify the LD patterns among the modifier variants of ACE2 and PIR genes in 5 super-populations.

2.5. Statistical analyses

The statistical tool available at Metaboanalyst (version 5) (Xia et al., 2009) was used for multivariate principal component analysis (PCA) based on the variant allele frequencies (VAFs) of the above mentioned disease-associated and modifier variants. In total five principal components were considered.

3. Result

3.1. Disease associated variants of ACE2

ACE2 harbors 15 variants that are associated with various diseases (Table 1). Except one (rs1514280), the other variants are associated with at least one of the following conditions- hypertension, dyslipidemia, type 2 diabetes mellitus, atrial fibrillation, and/or left ventricular hypertrophy. These are the major comorbidities of COVID-19 (Apicella et al., 2020; Das et al., 2020; Gawalko et al., 2020; Ma et al., 2020; Sanyaolu et al., 2020; Shibata et al., 2020; Stone et al., 2020). Among the identified variants, thirteen are intron variants, one (rs4830542) is a downstream gene variant, and one (rs2285666) is a splice region modifier. Intron variants can influence expression of the genes with introns through altering transcriptional activity or affecting splicing efficiency (Cooper, 2010; Lin et al., 2019). Splice site variants that cause creation of a cryptic splice site, incorrect recognition of splice site or disruption in activities of regulatory elements involved in splicing, can lead to

| Variant ID | Consequence | Alleles | Disease |
|------------|-------------|---------|---------|
| rs4830542  | Downstream gene variant | C/G,T | Essential hypertension, atrial fibrillation, dyslipidemias |
| rs2074192  | Intron variant | C/T | Essential hypertension, Atrial fibrillation, dyslipidemias, left ventricular hypertrophy, hypertensive disease, pregnancy associated hypertension, metabolic syndrome X, cardiovascular diseases, carotid atherosclerosis |
| rs233575  | Intron variant | G/A | Dyslipidemias |
| rs1514280  | Intron variant | G/G | Negative affectivity, temporal pain, psychological symptom |
| rs4240157  | Intron variant | C/T | Essential hypertension, atrial fibrillation, dyslipidemias, hypertensive disease |
| rs879222  | Intron variant | C/G | Essential hypertension, dyslipidemias, carotid atherosclerosis, Non-insulin-dependent diabetes mellitus |
| rs4646156  | Intron variant | A/T | Angina pectoris, dyslipidemias |
| rs4646155  | Intron variant | A/C | Atrial fibrillation, essential hypertension, dyslipidemias |
| rs4646188  | Intron variant | A/G | Non-insulin-dependent diabetes mellitus, dyslipidemias, essential hypertension, hypertensive disease |
| rs4646142  | Intron variant | G/A,C | Dyslipidemias |
| rs2048663  | Intron variant | T/G | Essential hypertension |
| rs2285666  | Splice region variant | C/T | Dyslipidemias, non-insulin-dependent diabetes mellitus, hypertensive disease, orthostatic hypertension |
| rs6632677  | Intron variant | G/C | Atrial fibrillation, essential hypertension, ischmic stroke, hypertrophic cardiomyopathy |
| rs2106809  | Intron variant | A/G | Essential hypertension, hypertensive disease, cardiac event, hypertrophic cardiomyopathy, dyslipidemias, left ventricular hypertrophy, ischmic stroke |
| rs1978124  | Intron variant | T/A,C | Dyslipidemias, hypertensive disease |
3.2. ACE2 variant allele frequencies in different populations

Variant allele frequencies (VAFs) of the retrieved disease associated variants of ACE2 in different populations are given in Table 2. Variant alleles with frequencies ≥ 0.25 are shown as bold italic. The variant alleles at rs4646155, rs4646188 and rs6632677 loci are present with low frequencies in all populations. At the other loci, the distribution patterns of the variant alleles are apparently different among the super-populations. The variant alleles at rs4830542, rs2074192, rs233575, rs1514280, rs4240157, rs4646157, rs2048683, and rs1978124 are relatively more prevalent in the European (EUR) populations. A similar VAF pattern is followed by two admixed American (AMR) populations—Puerto Rican (PUR) and Colombian (CLM). VAFs at majority of these loci are higher in these two populations compared to the other two admixed American populations (Mexican (MXL) and Peruvians (PEL)). The global minor (variant) alleles at rs4646142, rs2285666, and rs2106809 loci are present as the major alleles in the East Asian (EAS) populations as well as three South Asian (SAS) populations (Bangladeshi (BEB), Indian Telegu (ITU) and Sri Lankan Tamil (STU)), but are relatively less prevalent (frequency < 0.3) in the European populations. The variant allele at rs2074192 is present with high frequencies (> 0.3) in all, except the South Asian populations. In general, the variant alleles are present at much lower frequencies in the East Asian populations followed by the African (AFR) populations.

Drastic interpopulation variation exists for the VAFs at three loci—rs4830542, rs4240157 and rs879922 (Table 2). In the African populations, the VAFs at these loci range from 0.42 to 0.63. At the same loci, the VAFs in the East Asian populations range from 0.0 to 0.15. The other populations (in AMR, EUR and SAS) have intermediate VAFs at these loci. In the East Asian populations, the VAFs at rs1514280, rs4240157, rs879922, rs4646156, and rs4830542 are very low compared to the other populations.

3.3. LD among the ACE2 variants in different super-populations

LD among the 15 disease associated variants in five super-populations is shown in Fig. 1. In the EUR super-population, there is strong LD among rs4830542, rs233575, rs1514280, rs4240157, rs879922, rs4646156 and rs2048683. Although similar LD patterns for these variants are present in the admixed American and the South Asian (SAS) populations, these associations are not as strong as those observed in the European populations. The patterns of VAFs at these loci are similar in the EUR populations and the two admixed American populations (PUR and CLM) (Table 2), but quite different from the East Asian (EAS) populations. These variants, except rs1514280 and rs2048683, are associated with dyslipidemia (Table 1). Contrary to the other populations, relatively strong LD exists among rs2074192, rs4646142, rs2285666 and rs2106809. In fact, majority of the LDs that are present in the East Asians are quite strong and distinct. rs4646188 variant allele is non-existent among the East Asians. Apparently, LD among the disease associated SNPs is the lowest in AFR compared to the other super-populations.
In all super-populations, rs4646142 is in high LD with rs2285666. Except in AFR, these two variants are in high LD with rs2106809. rs2048683 has high pairwise LD with rs4646156 in all super-populations. The pairwise LD of rs2048683 with the other SNPs vary among super-populations.

### 3.4. Haplotype frequencies in different super-populations

The haplotypes with frequencies ≥ 0.01 in different super-populations are shown in Table 3 (variant alleles are shown as bold italic). Among these, the C_{C\_G\_A\_C\_A\_G\_T\_G\_A\_T} haplotype (#8 in Table 3) with variant alleles at rs4830542, rs233575, rs1514280, rs4240157, rs879922, rs466156, rs2048683, and rs1978124 is present at >0.01 frequencies in EUR (0.277), AMR (0.179), SAS (0.164) and AFR (0.011) super-populations. The C_{C\_G\_A\_C\_A\_G\_T\_C\_G\_C\_G\_C\_G\_C\_G\_C} haplotype (#9 in Table 3) harboring variant alleles at rs4830542, rs233575, rs1514280, rs4240157, rs879922, rs466156, rs2048683 and rs2106809 is present only in the EUR (0.029) and SAS (0.429) super-populations. The T\_C\_A\_G\_T\_G\_C\_A\_C\_G\_T\_G\_C\_C (#14 in Table 3) haplotype is also highly prevalent in the SAS populations (0.429). The haplotype with none of the variant alleles (#15 in Table 3) is more frequent in the South Asian super-population (0.022).

### 3.5. Variants in the regulatory elements and their haplotype frequencies

Among the disease associated ACE2 variants, five (rs1514280, rs4240157, rs4646155, rs4646156, and rs6632677) were identified to be associated with regulatory elements as shown in Table 4. rs1514280 and rs4240157 are associated with circular RNA (circRNA) coding region. circRNAs are produced from both protein-coding and non-coding regions of the genome and lack poly(A) tails and 5′ caps (Szabo and Salzman, 2016). These RNAs generated by backsplicing play key role in controlling the levels of endogenous miRNAs and act as protective factor against development of cardiac hypertrophy and atherosclerosis (Holdt et al., 2016; Stępień et al., 2018; Tan et al., 2017; Wang et al., 2016).
gene targeting chromatin interactive region. Through formation of chromatin loop, distal regulatory elements are connected to their target genes (Peng et al., 2019; Sobhy et al., 2019). The frequencies of haplotypes containing these 5 regulatory SNPs in different super-genes (Peng et al., 2019; Sobhy et al., 2019). The frequencies of haplotypes in different super-populations (Peng et al., 2019; Sobhy et al., 2019). The frequencies of haplotypes in different super-populations. Variants in PIR in the European super-population appear to have significantly stronger LD and form multiple small blocks compared to the other super-populations. Variants in PIR in the European super-population also have relatively more LDs with the variants of ACE2. PIR intrinsic modifier variant rs191989275 is present only in the European super-population (Fig. 2).

There is a striking difference in the frequencies of PIR modifier variants between the South Asian and other super-populations (Supplementary Table 2). PIR modifier variants at rs112433960, rs113923657, rs141501969, rs142269326, rs147768156, rs149822007, and rs191989275 loci are non-existent in the SAS super-population, although these are mostly present in the AFR, AMR, EAS and/or EUR super-populations. In all super-populations, except SAS, the rs138854040 and rs6632677 variants of ACE2 form strong pairwise LD with several variants of the PIR gene. These differences in distribution of the above-mentioned disease-associated and modifier variants are reflected in the un-supervised multivariate clustering of the populations based on VAFs (Fig. 3). As shown, populations of similar ethnic origins cluster together in the PCA plot.

### 4. Discussion

COVID-19 patients with history of chronic illnesses such as hypertension, diabetes, cardiovascular disease, chronic lung disease and obesity are at higher risk of developing severe illness and deteriorating outcomes like acute respiratory distress syndrome (ARDS) and pneumonia (Chen et al., 2020b; Sanyaolu et al., 2020; Singh et al., 2020). ACE2 is a strong hypertensive quantitative trait locus (QTL) (Crackower et al., 2002). Several variants of ACE2 are also associated with

### 3.6. Modifier variants in genes adjacent to ACE2 and their LD patterns

There are several genes on the X chromosome within ±0.5 Mbps of the disease associated variants of ACE2. Among these PIR is the only gene which is associated with inflammation. Frequencies of the modifier variants of ACE2 and PIR genes are given in supplementary Table 2 and pairwise LD between these variants is shown in Fig. 2. Modifier variants of PIR in the European super-population appear to have significantly stronger LD and form multiple small blocks compared to the other super-populations. Variants in PIR in the European super-population also have relatively more LDs with the variants of ACE2. PIR intrinsic modifier variant rs191989275 is present only in the European super-population (Fig. 2).

#### Table 3

| Haplotypes | Frequencies in super-populations |
|------------|----------------------------------|
| # | C_C_A_C_T_T_C_T_G_A_T | AFR (0.072), AMR (0.021), EUR (0.014), SAS (0.013) |
| 1 | C_C_A_C_T_T_C_T_G_A_T | AFR (0.026) |
| 2 | C_C_A_C_T_T_C_T_G_A_T | AFR (0.091) |
| 3 | C_C_A_C_T_T_C_T_G_A_T | AFR (0.109), AMR (0.012) |
| 4 | C_C_A_C_T_T_C_T_G_A_T | AFR (0.016) |
| 5 | C_C_A_C_T_T_C_T_G_A_T | AFR (0.069) |
| 6 | C_C_A_C_T_T_C_T_G_A_T | AFR (0.129), AMR (0.015), EAS (0.030), SAS (0.075) |
| 7 | C_C_A_C_T_T_C_T_G_A_T | AFR (0.179), EUR (0.277), SAS (0.164) |
| 8 | C_C_A_C_T_T_C_T_G_A_T | EUR (0.020) |
| 9 | C_C_A_C_T_T_C_T_G_A_T | AMR (0.017), EUR (0.022) |
| 10 | C_C_A_C_T_T_C_T_G_A_T | AMR (0.036), EAS (0.064) |
| 11 | C_C_A_C_T_T_C_T_G_A_T | AFR (0.032), EAS (0.029), SAS (0.011) |
| 12 | C_C_A_C_T_T_C_T_G_A_T | AMR (0.013), EUR (0.022), SAS (0.014) |
| 13 | C_C_A_C_T_T_C_T_G_A_T | AFR (0.059), AMR (0.246), EAS (0.429), EUR (0.165), SAS (0.429) |
| 14 | C_C_A_C_T_T_C_T_G_A_T | AFR (0.012), AMR (0.019), EUR (0.016), SAS (0.022) |
| 15 | C_C_A_C_T_T_C_T_G_A_T | EUR (0.025), EUR (0.026) |
| 16 | C_C_A_C_T_T_C_T_G_A_T | AFR (0.325), AMR (0.288), EAS (0.405), EUR (0.144), SAS (0.124) |
| 17 | C_C_A_C_T_T_C_T_G_A_T | EUR (0.119) |
| 18 | C_C_A_C_T_T_C_T_G_A_T | EUR (0.012) |
| 19 | C_C_A_C_T_T_C_T_G_A_T | AFR (0.012), AMR (0.048), SAS (0.020) |
| 20 | C_C_A_C_T_T_C_T_G_A_T | EUR (0.032), EUR (0.107), SAS (0.061) |

* order of the variants in the haplotype: C_T_G_A_C_T_G_C_T_G_A

#### Table 4

| rs ID | Chromosome | Allele | Related regulatory elements | Target genes | rSNP or LD-proxies | eQTL |
|-------|------------|--------|----------------------------|--------------|-------------------|------|
| rs1514280 | chrX:15586448 | C/T | circRNA region | N/A | rsSNP | N |
| rs4240157 | chrX:15586964 | A/G | circRNA region | N/A | rsSNP | N |
| rs4646155 | chrX:15597509 | G/A | Chromatin interactive region | 1 gene | rsSNP | N |
| rs4646156 | chrX:15597043 | A/T | Chromatin interactive region | 1 gene | rsSNP | Y |
| rs6632677 | chrX:15614872 | C/G | Chromatin interactive region | 2 genes | rsSNP | N |

* Regulatory SNP.

#### Table 5

| # | Haplotype | ALL | AFR | AMR | EAS | EUR | SAS |
|---|----------|-----|-----|-----|-----|-----|-----|
| 1 | G_T_T_C_G | 0.6477 | 0.4526 | 0.6565 | 0.8901 | 0.6084 | 0.6978 |
| 2 | A_C_A_C_G | 0.1528 | 0.0847 | 0.2099 | 0.0026 | 0.3146 | 0.1936 |
| 3 | G_C_T_T_G | 0.0607 | 0.1326 | 0.0191 | 0.0314 | 0.0013 | 0.085 |
| 4 | A_C_T_C_G | 0.045 | 0.1236 | 0.0363 | 0.0013 | 0.03 | 0.0042 |
| 5 | C_G_C_A_G | 0.0315 | 0.1107 | 0.0115 | – | 0.0026 | – |
| 6 | G_C_T_C_G | 0.0278 | 0.0917 | 0.0038 | 0.0013 | 0.0104 | 0.0028 |
| 7 | G_T_T_C_G | 0.0207 | 0.0001 | 0.0363 | 0.0081 | 0.0013 | 0.007 |
| 8 | G_T_A_C_G | 0.013 | 0.0001 | 0.0267 | 0.0039 | 0.0013 | 0.0097 |
| 9 | G_T_T_G_G | – | – | – | – | – | – |

* Order of variants: rs1514280, rs4240157, rs4646156, rs4646155, rs6632677.

All variant haplotype: A_C_A_T_C_G.
dyslipidemia, type 2 diabetes mellitus, atrial fibrillation and left ventricular hypertrophy, among others. Based on the global distribution of COVID-19, the cases and mortality rates have been higher in the European and the American populations (World Health Organization, 2020; Yamamoto and Bauer, 2020). In this study, we have explored the interethnic and interpopulation variability in the distribution of ACE2 genetic variants that are associated with COVID-19 comorbidities.

Generally, ACE2 confers a protective effect on the epithelial cells of the lungs through the ACE2-Angiotensin (1–7)-MAS receptor axis that reduces inflammation, dilates blood vessels and reduces thrombosis. When ACE2 level is reduced on cell surface, the protective effect through ACE2-Angiotensin (1–7)-MAS receptor axis is minimized and amplification of the ACE-Angiotensin II-AT1 receptor axis occurs. ACE-Angiotensin II-AT1 receptor axis is responsible for various effects (such as vasoconstriction, increased inflammation, fibrosis and pulmonary damage) mediated by the Ang II (Verdecchia et al., 2020b). Individuals with higher ACE activity (thus, a more active ACE-Angiotensin II-AT1 receptor axis) in conjunction with reduced ACE2 activity (thus, a less active ACE2-Angiotensin (1–7)-MAS receptor axis to counterregulate the ACE-Angiotensin II-AT1 receptor axis effects) are more susceptible to hypertension in association with classical cardiovascular risk factors such as old age, dyslipidemia, and diabetes (Ghafouri-Fard et al., 2020; Pinheiro et al., 2019).

When SARS-CoV-2 enters the lung epithelial cells, the ACE2 receptor is internalized and, thus, removed from the functional site, compromising the balance between ACE and ACE2 activity. Decreased level of ACE2 contributes to vasoconstriction, increased inflammation, fibrosis and pulmonary damage and ultimately severe consequences of SARS-CoV-2 infection (Samavati and Uhal, 2020; Verdecchia et al., 2020a). In addition, ACE2 deficiency leads to Ang II accumulation. This elevated level of Ang II induces the production of pro-inflammatory cytokines and soluble IL-6 receptor, followed by STAT3 and NF-κB pathway-mediated activation of IL-6, which in turn causes overproduction of pro-inflammatory cytokines and chemokines including IL-6, and thus sets off a positive feedback loop (Hirano and Murakami, 2020). It is also known that ACE2 deficiency comes with old age, more in male than in females, and also diabetes mellitus (possibly due to excessive glycosylation) (Verdecchia et al., 2020b). ACE2 deficiency is also associated with Ang II induced exacerbation of in the lungs through the ACE2-Angiotensin (1–7)-MAS receptor axis that reduces inflammation, dilates blood vessels and reduces thrombosis. When ACE2 level is reduced on cell surface, the protective effect through ACE2-Angiotensin (1–7)-MAS receptor axis is minimized and amplification of the ACE-Angiotensin II-AT1 receptor axis occurs. ACE-Angiotensin II-AT1 receptor axis is responsible for various effects (such as vasoconstriction, increased inflammation, fibrosis and pulmonary damage) mediated by the Ang II (Verdecchia et al., 2020b). Individuals with higher ACE activity (thus, a more active ACE-Angiotensin II-AT1 receptor axis) in conjunction with reduced ACE2 activity (thus, a less active ACE2-Angiotensin (1–7)-MAS receptor axis to counterregulate the ACE-Angiotensin II-AT1 receptor axis effects) are more susceptible to hypertension in association with classical cardiovascular risk factors such as old age, dyslipidemia, and diabetes (Ghafouri-Fard et al., 2020; Pinheiro et al., 2019).

When SARS-CoV-2 enters the lung epithelial cells, the ACE2 receptor is internalized and, thus, removed from the functional site, compromising the balance between ACE and ACE2 activity. Decreased level of ACE2 contributes to vasoconstriction, increased inflammation, fibrosis and pulmonary damage and ultimately severe consequences of SARS-CoV-2 infection (Samavati and Uhal, 2020; Verdecchia et al., 2020a). In addition, ACE2 deficiency leads to Ang II accumulation. This elevated level of Ang II induces the production of pro-inflammatory cytokines and soluble IL-6 receptor, followed by STAT3 and NF-κB pathway-mediated activation of IL-6, which in turn causes overproduction of pro-inflammatory cytokines and chemokines including IL-6, and thus sets off a positive feedback loop (Hirano and Murakami, 2020). It is also known that ACE2 deficiency comes with old age, more in male than in females, and also diabetes mellitus (possibly due to excessive glycosylation) (Verdecchia et al., 2020b). ACE2 deficiency is also associated with Ang II induced exacerbation of

Fig. 2. Pairwise LD among the modifier variants of ACE2 and PIR. LD measures are given as $r^2$. For the missing alleles in any population LD could not be calculated (NA). AFR = African, AMR = Admixed American, EAS = East Asian, EUR = European, SAS = South Asian.

Fig. 3. Multivariate clustering of populations based on disease-associated and modifier VAFs of ACE2 and PIR. Clustering was performed using unsupervised principal component analysis (PCA) method. Only the first three principal components (PCs) are shown in this fig. (5 PCs were considered). As shown in the Fig. 3, the first three components could explain most part of the variance (cumulative variance for the PC1, PC2 and PC3 is 98.1%).
Infection, Genetics and Evolution 92 (2021) 104888

The variant allele at rs233575 is associated with left atrial remodeling after myocardial infarction (Kassiri et al., 2009; Pal and Bhansali, 2020; Xudong et al., 2006). Thus, people with existing ACE2 deficiency may be affected more severely by the loss of ACE2 following SARS-CoV-2 infection (Dijkstra et al., 2012; Verdecchia et al., 2020a). The opposite may be true as well.

In the Indian population, a strong correlation exists between the variant allele at rs2285666 locus and lower COVID-19 infection rate as well as lower case fatality rates (CFR) (Srivastava et al., 2020). The variant allele at rs2285666 is present with >0.5 frequencies in the EAS populations and three South Asian populations (BEB, STU, and ITU) (Table 2). The homozygous variant genotype at rs2285666 increases expression of ACE2 by 50% compared to the homozygous wild-type genotype (Asselta et al., 2020; Srivastava et al., 2020). Using the Human Splicing Finder (Desmet et al., 2009), Asselta et al. predicted the presence of variant allele at this splice region variant to increase the splice site strength by about 9.2% (Asselta et al., 2020). But in other studies, no significant splicing alteration by the variant allele at rs2285666 was found using the same tool (Novelli et al., 2020; Srafella et al., 2020). This calls for further analysis to have a clear understanding of the potential role of rs2285666 variant allele in the enhancement of ACE2 protein production.

In Han Chinese females, the heterozygous genotype at rs2285666 locus serves a protective role against essential hypertension (Zhang et al., 2018). Increased circulating Ang-(1–7) level was observed in females with the variant allele at the rs2106809 locus in Chinese Han population (Liu et al., 2016). The haplotype T_C_A_G_T_G_T_C_A_G_C_G_C with the variant alleles at rs4646142, rs2285666 and rs2106809 is highly prevalent in the EUR (0.429) and the SAS (0.429), but much lower in the EUR (0.165) populations.

rs4646188 variant allele is associated with hypertension, left ventricular hypertrophy and dyslipidemia (Luo et al., 2019; Pan et al., 2018; Patel et al., 2012). This variant allele is absent in the EAS populations (Table 2 and Fig. 1), which might be another candidate variant associated with protective role against severity of COVID-19.

Variant alleles at rs4240157, rs1978124, rs2048683, rs4646188, rs4646156 and rs879922 are associated with type 2 diabetes mellitus (Liu et al., 2018). Variant alleles at rs4830542, rs2106809, rs4240157, rs4646188, and rs879922 loci are associated with essential hypertension as well as ras activation (Luo et al., 2019; Pan et al., 2018). rs4646156, rs4830542 and rs4240157 variants are associated with hypertension-related atrial fibrillation, and left atrial remodeling (Luo et al., 2019; Zhang et al., 2018). rs4646156 is a dyslipidemia associated chromatin interactive region related regulatory variant (Table 4). rs2074192 variant allele increases susceptibility to left ventricular hypertrophy (Pan et al., 2019). The variant allele at rs233575 is associated with higher left ventricular mass index (Liu et al., 2018) and progression to type 2 diabetes mellitus (Liu et al., 2018). A strong LD block is seen in the EUR super-population among rs4830542, rs4240157, rs2074192, rs233575, rs1514280, rs879922 and rs4646156 loci (Fig. 1), all of which have modifier effect (Supplementary Table 2). A similar but weaker LD block is observed in the AMR populations. The VAFs at this LD block is >0.2 in two AMR (PUR and CLM) populations and the EUR populations (Table 2) and the haplotype with variant alleles at all these positions, except rs2074192, (C_C_G_A_C_A_C_G_C_A_G_T_G_A_G_T) has >0.01 frequencies in the EUR (0.277) and the AMR (0.179) populations. This might indicate that the people in these regions are at higher risk of hypertension and other phenotypic effects associated with the variant alleles at rs4830542, rs4240157, rs2074192, rs233575, rs1514280, rs879922, and rs4646156 loci.

Admixed Latin American populations- CLM (Colombian), MXL (Mexican), PEL (Peruvians), and PUR (Puerto Ricans)- have distinct patterns of continental genetic admixture (Norris et al., 2018). The Puerto Ricans and the Colombians inherited more genetic content from the European ancestry than the Peruvians and the Mexicans (Homburger et al., 2015; Norris et al., 2018; Ruiz-Linares et al., 2014; Salzano and Sans, 2014). It is not unlikely to expect similar phenotypic effects in the European, Puerto Ricans and Colombians by the variants of ACE2 and PIR. In a previous study on drug response-related genetic variants, CLM and PUR formed a closer cluster with the Europeans in a population dendrogram (Ahsan et al., 2020). Fig. 3 shows a similar clustering of the populations based on the VAFs of the disease-associated and modifier variants of ACE2 and PIR.

The T_C_G_C haplotype consisting of the variant alleles at rs464156, rs879922, rs4240157, and rs233575 may be associated with higher values for left ventricular mass index and septal wall thickness with a higher odds ratio for left ventricular hypertrophy in men (Lieb et al., 2006). Interestingly, the haplotypes #8 and #9 in Table 3 contain this specific combination of the variant alleles. #9 is found only in the European super-population (0.020) and #8 is most frequent in the European (0.277) followed by the admixed Americans (0.179).

Among the disease associated variants of ACE2, rs1514280 appears as an outlier at the first look based on its association with conditions like negative affectivity, temporal pain and psychological symptoms (Table 1). Long-lasting headache in COVID-19 patients is a frequent symptom, which is often severe and resistant to analgesics, in conjunction with anosmia (loss of the sense of smell) and/or aguesia (loss of taste functions of the tongue) (Rocha-Filho and Magalhães, 2020; Uygun et al., 2020). The haplotype with 3 regulatory variants at rs1514280, rs4240157 and rs4646156 loci (A_C_A_C_G) has a high frequency in the EUR (0.3146) and very low frequency in the EAS (0.0026) populations (Table 5). rs1514280 and rs4240157 loci are associated with eQTL formation, circRNA dysregulation is observed in multiple viral infections (Nahand et al., 2020). rs4646156 is an eQTL locus and this locus can modulate the expression of PIR gene in multiple tissues including esophagus, thyroid, adipose and artery (GTEx Consortium, 2013).

Strong LD and high VAFs at rs4830542, rs4240157, rs2074192, rs233575 and rs879922 loci in the EUR and two AMR (PUR and CLM) populations (Table 2) may be associated with more critical illness and higher death rates in the COVID-19 patients in these regions. On the other hand, very low frequencies of these alleles in the EAS populations might play protective role against progression to severe state by the comorbidities in SARS-CoV-2 infection. Although the VAFs at rs4830542, rs2074192, rs4240157 and rs879922 are >0.3 in the AFR populations, no significant LD block is present among these variants in AFR populations. Another variant with high frequencies (>0.35) in the EUR populations as well as PUR and CLM is rs1978124. People with rs1978124 variant have a high risk of developing type 2 diabetes (Liu et al., 2018). This variant allele is also associated with left ventricular remodeling in women (Patel et al., 2012).

COVID-19 is often associated with cytokine release syndrome (CRS) and its severity is associated with an elevated level of inflammatory cytokines and chemokines such as interleukin (IL)-2, IL-6, IL-1β, IL-7, IL-10, tumor necrosis factor (TNF), granulocyte colony-stimulating factor (G-CSF), monocyte chemoattractant protein-1 (MCP-1), and macrophage inflammatory protein 1 alpha (MIP1α) upon SARS-CoV-2 infection (Hojyo et al., 2020; Liu et al., 2020; Tay et al., 2020). The cytokine storm induces ARDS following infection with SARS-CoV-2 (Hirano and Murakami, 2020). IL-6 level elevates with progressive deterioration of illness and is higher in non-survivors compared to the survivors (Chen et al., 2020a; Zhou et al., 2020).

As a consequence of ACE2 internalization following SARS-CoV-2 infection, homeostatic responses try to restore ACE2 level by up-regulating ACE2 expression along with the other genes (including PIR) residing in a co-regulated cluster (Shovlin and Vizcaychipi, 2020). Double elite associations exist between two ACE2 enhancers (GH0X1015596 and GH0X1015579) and the promoter of PIR, which encodes the Pirin protein (Shovlin and Vizcaychipi, 2020). Pirin substantially facilitates binding of NF-κB p65 to DNA (Liu et al., 2013). Increased activation of NF-κB p65 pathway is associated with multiple diseases involving chronic inflammation (Girdharan and Srinivasan, 2018). Hyperactivation of NF-κB pathways leads to cytokine storm.
syndrome and increases COVID-19 severity (Harirhan et al., 2020). Furthermore, Pirin protein can bind to NF-κB, and modulate its DNA binding properties (Liu et al., 2013). Additionally, Pirin is overexpressed in response to oxidative stress (Brzoska et al., 2011). Oxidative stress can play a role in COVID-19 pathogenesis, perpetuate the cytokine storm cycle, blood clotting mechanism, and exacerbate hypoxia (Cecchini and Cecchini, 2020). In addition, cigarette smoke up-regulates PIR gene expression and the overexpressed Pirin protein may direct NF-

Cecchini, R., Cecchini, A., 2020. SARS-CoV-2 infection pathogenesis is related to oxidative stress as a response to aggression. Med. Hypotheses 143, 110102.

Chen, G., Wu, D., Guo, W., Cao, Y., Huang, D., Wang, H., Wang, W., Zhang, X., Chen, H., Yu, H., Zhang, X., Zhang, M., Wu, S., Song, J., Chen, T., Han, M., Li, S., Luo, X., Zhao, J., Ning, Q., 2020. Clinical and immunologic features of severe and moderate coronavirus disease 2019. J. Clin. Invest. 130, 2620–2629.

Chen, N., Zhou, M., Dong, X., Qiu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J.A., Yu, T., Zhang, X., Zhang, L., 2020b. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet (London, England) 395, 507–513.

Cooper, D.N., 2010. Functional intronic polymorphisms: buried treasure awaiting discovery within our genes. Hum Genomics 4, 284–288.

Craighead, M.A., Sarao, R., Oditi, G.V., Yagil, C., Krzakradzki, I., Scanga, S.E., Oliveira-dos-Santos, A.J., da Costa, J., Zhang, L., Pei, Y., Scholey, J., Ferrari, C.M., Manoukian, A.S., Chappell, M.C., Backs, P.H., Yagil, Y., Penninger, J.M., 2002. Angiotensin-converting enzyme 2 is an essential regulator of heart function. Nature 417, 822–828.

Das, S., K. R., A., Birangal, S.R., Nikam, A.N., Pandey, A., Mastalik, S., Joseph, A., 2020. Role of comorbidities like diabetes on severe acute respiratory syndrome coronavirus-2: a review. Life Sci. 225, 39–79.

Debo, J.L., Bangalore, S., Benetton, A., Davis, A.M., Ichosun, E.D., Muntner, P., Rossing, P., Zoungas, S., Bakris, G., 2017. Diabetes and hypertension: a position statement by the American Diabetes Association. Diabetes Care 40, 1273.

Desmet, F.O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M., Beroud, C., 2009. Human splicing finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 37, e67.

Devaux, C.A., Rolain, J.M., Raoult, D., 2020. ACE2 receptor polymorphism. Life Sci. 225, 39–44.

Dvir, O., Rabinovici, D., Glickman, Y., Shani, G., Yerushalmi, R., Weiss, N., Gabizon, A., 2010. Functional intronic polymorphisms: buried treasure awaiting discovery within our genes. Hum Genomics 4, 284–288.

EJ. Gen. Virol. 93, 1924–1929.

Eskandari, H., Nabi-Afjadi, M., Fakkar-Afjadi, J., Farahmandian, N., Mireaesamali, S.M., Bahrei, E., 2020. A comprehensive review of COVID-19 characteristics. Biological Procedures Online 22, 19.

Fan, Z., Wu, G., Yue, M., Ye, J., Chen, Y., Xu, B., Shu, Z., Zhu, J., Lu, N., Tan, X., 2019. Hypertension and hypertensive left ventricular hypertrophy are associated with ACE2 genetic polymorphism. Life Sci. 225, 39–79.

Feng, Q., Lu, Y., Liu, Z., 2020a. The interplay of renin-angiotensin system and toll-like receptor 4 in the inflammation of diabetic nephropathy. J Immunol Res 2020, 6193407.

Feng, Z., Li, S., Wang, D., Cai, J., Gao, C.L., Liu, Y., 2020b. Replication-dependent downregulation of cellular angiotensin-converting enzyme 2 protein expression by human coronavirus NL63. J. Gen. Virol. 93, 1924–1929.

Fikotko, M., Kaplon-Cieślak, A., Hoh, M., Dobrev, D., Linz, D., 2020. COVID-19 associated arial inflammation: incidence, putative mechanisms and potential clinical implications. JUC Heart Vasc. 30, 100631.

Gelbmann, B.D., Heguy, A., O’Connor, T.P., Zabner, J., Crystal, R.G., 2007. Upregulation of PIR gene expression by chronic cigarette smoke is associated with bronchial epithelial cell apoptosis. Respir. Res. 8, 10.

Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kong, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R., 2010. A global reference for human genetic variation. Nature 506, 68–74.

Ghafouri-Fard, S., Noroozi, R., Omrani, M.D., Branicki, W., Pontiea, E., Sayad, A., Pyrc, K., Labaj, P.P., Vaifaei, R., Taberi, M., Sanak, M., 2020. Angiotensin converting enzyme 2: a review on expression and its association with human disorders with special focus on SARS-CoV-2 infection. Vasc. Pharmacol. 130, 106680.
variability among populations highlights a possible link with COVID-19-related neurological complications. Genes (Basel) 11, 741.

Szabo, L., Salaman, J., 2016. Detecting circular RNA: bioinformatic and experimental challenges. Nat. Rev. Genet. 17, 679–692.

Tan, W.L.W., Lim, B.T.S., Anene-Nzelu, C.G.O., Ackers-Johnson, M., Dashi, A., See, K., Tiang, Z., Lee, D.P., Chua, W.H., Loo, T.D.A., Li, P.Y.Q., Richards, A.M., Foo, R.S.Y., 2017. A landscape of circular RNA expression in the human heart. Cardiovasc. Res. 113, 298–309.

Tay, M.Z., Poh, C.M., Rénia, L., MacAry, P.A., Ng, L.F.P., 2020. The trinity of COVID-19: immunity, inflammation and intervention. Nat. Rev. Immunol. 20, 363–374.

Thomas Merlin, C., Pickering Raelene, J., Tsorotes, D., Koitka, A., Sheehy, K., Bernardi, S., Toffoli, B., Nguyen-Huu, T.-P., Head Geoffrey, A., Fu, Y., Chin-Dusting, J., Cooper Mark, E., Tikellis, C., 2010. Genetic Ace2 deficiency accentuates vascular inflammation and atherosclerosis in the ApoE knockout mouse. Circ. Res. 107, 888–897.

Uygun, Ö., Ertas, M., Ekizoglu, E., Bolay, H., Özge, A., Kocasoy Orhan, E., Çağatay, A.A., Baykan, B., 2020. Headache characteristics in COVID-19 pandemic—a survey study. The Journal of Headache and Pain 21, 121.

Verdecchia, P., Cavallini, C., Spanevello, A., Angeli, F., 2020a. The pivotal link between ACE2 deficiency and SARS-CoV-2 infection. European Journal of Internal Medicine 76, 14–20.

Verdecchia, P., Cavallini, C., Spanevello, A., Angeli, F., 2020b. The pivotal link between ACE2 deficiency and SARS-CoV-2 infection. Eur J Intern Med. 76, 14–20.

Wang, K., Long, B., Liu, F., Wang, J.-X., Liu, C.-Y., Zhao, B., Zhou, L.-Y., Sun, T., Wang, M., Yu, T., Gong, Y., Liu, J., Dong, Y.-H., Li, N., Li, P.-F., 2016. A circular RNA protects the heart from pathological hypertrophy and heart failure by targeting miR-223. Eur. Heart J. 37, 2602–2611.

World Health Organization, 2020. WHO Coronavirus Disease (COVID-19) Dashboard, Wu, Z., McGoogan, J.M., 2020. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 323, 1239–1242.

Xia, J., Psychogiou, N., Young, N., Wishart, D.S., 2009. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res. 37, W652–W660.

Xudong, X., Junzhu, C., Xingsiang, W., Furong, Z., Yanrong, L. 2006. Age- and gender-related difference of ACE2 expression in rat lung. Life Sci. 78, 2166–2171.

Yamamoto, N., Bauer, G., 2020. Apparent difference in fatalities between Central Europe and East Asia due to SARS-COV-2 and COVID-19: four hypotheses for possible explanation. Med. Hypotheses 144, 110160.

Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Arzman, I.M., Azov, A.G., Bennett, R., Bhai, J., Billis, K., Boddi, S., Marugán, J. C., Cummins, C., Davidson, C., Dodici, K., Fatima, R., Gall, A., Giron, C.G., Gil, L., Grego, T., Haggerty, L., Hsxen, E., Hourlier, T., Iwanga, O.G., Janacek, S.H., Juetermann, T., Koy, M., Lavidas, I., Le, T., Lemos, D., Martinez, J.G., Maurel, T., McDowall, M., McMahon, A., Mohanan, S., Moore, B., Nuhn, M., Oheb, D.N., Parker, A., Parton, A., Patricio, M., Sakthivel, M.P., Abdul Salam, A.I., Schmitt, B.M., Schillenburg, H., Sheppard, D., Sycheva, M., Szabo, M., Taylor, K., Thomann, A., Threadgold, G., Vullo, A., Walts, B., Winterbottom, A., Zadina, A., Chakiachvili, M., Flint, B., Frankish, A., Hunter, E., Isles, G., Konstandina, M., Langridge, N., Loveland, J.E., Martin, F.J., Morales, J., Mudge, J.M., Mustafa, M., Perry, E., Ruffler, M., Trevanian, S.J., Cunningham, E., Howe, K.L., Zerbino, D.R., Flicek, P., 2020. Ensembl 2020. Nucleic Acids Res. 48, D682–D688.

Zhang, Q., Cong, M., Wang, N., Li, X., Zhang, H., Zhang, K., Jin, M., Wu, N., Qiu, C., Li, J., 2018. Association of angiotensin-converting enzyme 2 gene polymorphism and enzymatic activity with essential hypertension in different gender: a case-control study. Medicine 97, e12917.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., Cao, B., 2020. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 395, 1054–1062.