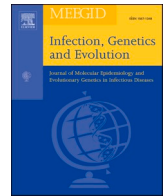**ELSEVIER**

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.

Genetic association of *TMPRSS2* rs2070788 polymorphism with COVID-19 case fatality rate among Indian populations

Rudra Kumar Pandey [*], Anshika Srivastava, Prajjval Pratap Singh, Gyaneshwer Chaubey [*]

Cytogenetics Laboratory, Department of Zoology, Banaras Hindu University, Varanasi 221005, India

A B S T R A C T

SARS-CoV-2, the causative agent for COVID-19, an ongoing pandemic, engages the *ACE2* receptor to enter the host cell through S protein priming by a serine protease, *TMPRSS2*. Variation in the TMPRSS2 gene may account for the disparity in disease susceptibility between populations. Therefore, in the present study, we have used next-generation sequencing (NGS) data of world populations from 393 individuals and analyzed the TMPRSS2 gene using a haplotype-based approach with a major focus on South Asia to study its phylogenetic structure and their haplotype sharing among various populations worldwide. Our analysis of phylogenetic relatedness showed a closer affinity of South Asians with the West Eurasian populations therefore, host disease susceptibility and severity particularly in the context of *TMPRSS2* will be more akin to West Eurasian instead of East Eurasian. This is in contrast to our prior study on the ACE2 gene which shows South Asian haplotypes have a strong affinity towards West Eurasians. Thus *ACE2* and *TMPRSS2* have an antagonistic genetic relatedness among South Asians. Considering the significance of the TMPRSS2 gene in the SARS-CoV-2 pathogenicity, COVID-19 infection and intensity trends could be directly associated with increased expression therefore, we have also tested the SNPs frequencies of this gene among various Indian state populations with respect to the case fatality rate (CFR). Interestingly, we found a significant positive association between the rs2070788 SNP (G Allele) and the CFR among Indian populations. Further our cis eQTL analysis of rs2070788 shows that the GG genotype of the rs2070788 tends to have a significantly higher expression of TMPRSS2 gene in the lung compared to the AG and AA genotypes thus validating the previous observation and therefore it might play a vital part in determining differential disease vulnerability. We trust that this information will be useful in understanding the role of the *TMPRSS2* variant in COVID-19 susceptibility and using it as a biomarker may help to predict populations at risk.

1. Introduction

COVID-19 is an ongoing pandemic that has cost millions of lives worldwide, caused by the SARS-CoV-2 virus of the Beta Family. Along with *ACE2* (Angiotensin-converting enzyme 2) which acts as a receptor, *TMPRSS2* (Transmembrane protease, serine 2), a serine protease, is also involved in virus entry into the host cell through S Protein priming (Hoffmann et al., 2020; Zhou et al., 2020). The impact of the COVID-19 crisis is not uniform across ethnic groups. Patients from different ethnic backgrounds suffer disproportionately (Webb Hooper et al., 2020). Discrepancies in infection, as well as CFR, could be due to multiple reasons e.g., existing comorbidities, differences in quarantine and social distancing policies, access to medical care, reliability & coverage of epidemiological data, and population age structure, which shows that mortality is greater among the elderly and those with comorbidity (Ejaz et al., 2020; Sanyaolu et al., 2020). However, many young and healthy people have also lost their lives due to rapid cytokine storms (Muschitz et al., 2021). It is important to note that these factors do not appear to account for all the disparities noticed among groups, and there are significant knowledge gaps. However, countries with strict standards for the collection and presentation of epidemiological data suggest that human variation in genetic makeup may account for differential susceptibility and severity in disease outcomes among different populations (SeyedAlinaghi et al., 2021). There is evidence that supports the role of ACE2 gene variations in COVID-19 susceptibility among Indian populations (Srivastava et al., 2020a; Srivastava et al., 2020b). However, little is known regarding the genetic structure of the TMPRSS2 gene among South Asian populations, a detailed analysis of the sequence data of the TMPRSS2 gene from world populations may unveil its haplotype sharing, which may help understand the role of *TMPRSS2* in disease

susceptibility globally.

TMPRSS2 has been linked to a variety of physiological and pathological processes. Along with SARS-CoV-2 and Influenza virus, various other human coronaviruses such as HCoV-229E, MERS-CoV, and SARS-CoV, have been also identified to utilize this protein for cell entrance (Shen et al., 2017). Androgenic hormones were shown to upregulate this gene in prostate cancer cells, while androgen-independent prostate cancer tissue was found to downregulate it (Mollica et al., 2020). Northern blots analysis has revealed that in mice *TMPRSS2* is mainly expressed in the kidney and prostate, whereas in humans, *TMPRSS2* is largely expressed in the prostate, salivary gland, stomach, and colon (Vaarala et al., 2001). *TMPRSS2* is also expressed in the epithelia of the respiratory, urogenital and gastrointestinal tracts according to in-situ hybridization investigations performed on mice embryos and adult tissues (Vaarala et al., 2001). Given the relevance of the TMPRSS2 gene in the SARS-CoV-2 infection process, COVID-19 infection and severity pattern may be directly linked to elevated TMPRSS2 gene expression, resulting in varying disease susceptibility outcomes in various communities. However, the role of *TMPRSS2* polymorphism in disease susceptibility in the Indian populations is largely unexplored. Therefore, in the current study, we analyzed the haplotype structure of the TMPRSS2 gene focusing on South Asia and its genetic variants that could be responsible for changes in the gene expression in the lungs tissue and, tested its correlation with epidemiological data available on COVID-19, for any existing association among Indian population.

2. Material and methods

The Haplotype analysis of the TMPRSS2 gene among various world populations was performed using NGS data from Pagani et al. (Pagani et al., 2016). PLINK 1.9 was used to extract complete coordinates of the TMPRSS2 gene from the dataset for different populations (Purcell et al., 2007). Based on Principal component analysis (PCA) results performed by Pagani et al., ethnic outliers samples from Sahul and Africa were excluded, as well relatives up till second-degree, from which a total of 393 samples and 795 SNVs (Single nucleotide variants) were observed and used subsequently for our study (Supplementary Table 1 and 2). The plink file was converted to fasta (ped to IUPAC) by a customized script (Budak, 2020). For phasing, *Fst* calculation, and generation of Network and Arlequin input file, DNAsp (ver 6.0) was used (Rozas et al., 2017). MEGA X was used to construct an *Fst* based Neighbour-joining tree (Kumar et al., 2018). To calculate Nei's genetic and average pairwise distance, Arlequin 3.5 was used and plotted on a graph by R V3.1 (Excoffier and Lischer, 2010; R: The R Project for Statistical Computing [Internet]. [cited, 2021]). Network v5 and network publisher were employed to draw the median-joining network while total and prevalent haplotypes in the TMPRSS2 gene for each population were calculated using XML file generated through Arlequin 3.5 (Excoffier and Lischer, 2010; Bandelt et al., 1999).

For the association study, we searched for the studies on TMPRSS2 gene variants reported in the literature elsewhere in relation to COVID-19 susceptibility (Mollica et al., 2020; Andolfo et al., 2021; Asselta et al., 2020; Bhattacharyya et al., 2020; Darbani, 2020; Hou et al., 2020; Irham et al., 2020; Iyer et al., 2020; Jeon et al., 2020; Kim and Jeong, 2021; Latini et al., 2020; Paniri et al., 2021; Piva et al., 2021; Ragia and Manolopoulos, 2020; Senapati et al., 2020; Sharma et al., n.d.; Singh et al., 2021; Strope et al., 2020; Torre-Fuentes et al., 2021; Vargas-Alarcón et al., 2020; Wang et al., 2020; Wulandari et al., 2021; Saih et al., 2021; Schönfelder et al., 2021; Kehdy et al., 2021) from this a total of 5 SNPs (rs2070788, rs734056, rs12329760, rs2276205, and rs3787950) was observed in our data and studied subsequently in detail. Data from the Estonian Biocentre (Chaubey et al., 2017; Pathak et al., 2018; Re3data, 2014; Tätte et al., 2019), phase 3 of the 1000 Genomes Project (Durbin et al., 2010), and our new genotyped samples from several Indian states were used to calculate the frequency of each of these SNPs among various Indian populations using plink 1.9. We

obtained state-wise frequency of cases and deaths from https://www.mygov.in/corona-data/covid19-statewise-status/. State-wise frequency maps for rs2070788 and COVID-19 CFR among the Indian population were made using https://www.datawrapper.de/. and worldwide spatial distribution of rs2070788 was generated from the *PGG.SNV* toolkit using 1000 genome samples (Zhang et al., 2019). The linear regression calculations for statewise allele frequency Vs the CFR were done using Microsoft excel and results were further validated by SPSS (ver 26) at a 95% confidence interval and 1000 bootstrapping (2,000,000 seeds) for a two-tailed significance. The composite plots for all variables were produced using a customized R script (ver 4) (R: The R Project for Statistical Computing [Internet]. [cited, 2021]). Expression quantitative trait loci (eQTL) were examined utilizing the GTEx database (http://www.gtexportal.org/home/) to determine the relationship between genetic variations and TMPRSS2 gene expression profile (THE GTEX CONSORTIUM et al., 2015). Haploview was used to generate a Linkage disequilibrium (LD) map and to calculate aggregate haplotypes frequency carrying rs2070788 (G allele) in each population (Barrett et al., 2005).

3. Result and discussion

TMPRSS2 is a serine protease enzyme that is encoded in humans by the TMPRSS2 gene located on chromosome 21q22.3. (Paoloni-Giacobino et al., 1997). This protein aids in the entry of various viruses into the host cells, like the influenza virus, and human coronaviruses such as HCoV-229E, MERS-CoV, SARS-CoV, and SARS-CoV-2 by proteolytically cleaving and then activating the viral envelope glycoproteins (Huggins, 2020), and thus can be inhibited by TMPRSS2 inhibitor (Hoffmann et al., 2020). Genetic variation in this gene may account for differential vulnerability for COVID-19 disease among diverse populations. Therefore, in the present study with our major focus on South Asia, we examine the genetic relatedness, using TMPRSS2 gene sequence data among world populations by haplotype-based approach for comparison among the various groups. *F*st based Neighbour-Joining (NJ) tree showed the clustering of South Asians with the West Eurasian populations (Caucasus, West Asia, Europe, and Central Asia) therefore, suggesting the closer affinity of South Asians with West Eurasians populations for the TMPRSS2 gene. (Fig. 1A). Similarly, the Average Pairwise differences analysis showed smaller diversity and genetic distance between populations, among East and West Eurasians, while greater diversity and genetic distance was observed between East and West Eurasian populations. The lowest diversity was found in West Asia & the American populations (Fig. 1B). A median-joining (MJ) network analysis of the TMPRSS2 gene revealed that there are 499 haplotypes throughout this gene among the examined populations, with five prevalent haplotypes (Hap 34, Hap 48, Hap 75, Hap 98, and Hap 260), each having ≥10 individuals. Haplotypes 48 and 75 were found to be more common in Europe, while haplotypes 98 and 260 were frequent in Siberia. Haplotype 34 was predominant in Southeast Asia, followed by Central Asia (Supplementary Table 3A and Supplementary Fig. 1). South Asian populations carry 47 haplotypes, among which 6 are shared (Hap_34, Hap_48, Hap_78, Hap_112, Hap_219, and Hap_260) with other continental populations while the rest are unique to South Asia. Among the shared haplotypes, five are shared majorly with the West Eurasian populations i.e. (Hap_34, Hap_48, Hap_78, Hap_112, Hap_219), whereas only a single haplotype (Hap_260) is majorly shared with the East Eurasian populations. (Fig. 1C and Supplementary Table 3B). Thus haplotype sharing, as well as *Fst* analysis, are consistent with the West Eurasian affiliation of the majority of South Asian *TMPRSS2* haplotypes (Fig. 1C and Fig. 1A). Therefore, the host susceptibility of SARS-CoV-2 for the TMPRSS2 gene among South Asians is most likely expected to be similar to West Eurasians rather than that of East Eurasians. In contrast with this, our previous study on the ACE2 gene has shown the strong genetic affinity of South Asian haplotypes with the East Eurasians (Srivastava et al., 2020a; Srivastava et al., 2020b). Thus, for the South
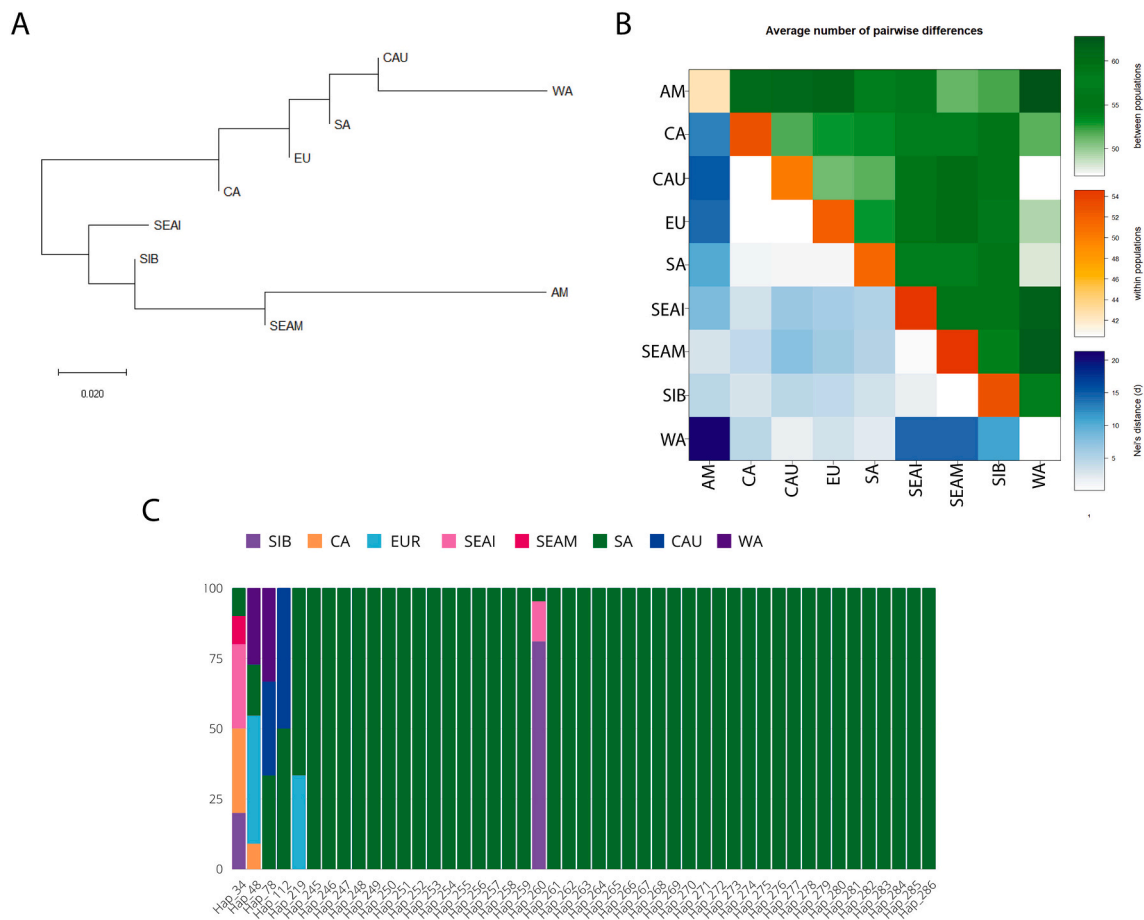
Fig. 1. (A) Neighbour-Joining (NJ) tree based on Fst distance, showing genetic relationship for TMPRSS2 gene among the studied population. (B) Matrix showing average paired variation for TMPRSS2 gene, between the population (green) in the upper triangle, within-population (orange) along diagonal, and Nei's distance between populations are shown (blue) in the lower triangle. The obtained value for different variables is directly proportional to the color gradient. (C) The stacked bar-plot represents 47 haplotypes observed in TMPRSS2 Gene among South Asian populations. Frequency and sharing for each haplotype with South Asia and to other geographic regions are indicated with different colored bars.

Asians, *ACE2* and *TMPRSS2* have an antagonistically genetic relatedness. As a result, it is worth proposing that the South Asian population's susceptibility to SARS-CoV-2 will fall somewhere between West and East Eurasian people, which is most likely the cause of the moderate susceptibility seen in first and second waves.

Recognizing the importance of the TMPRSS2 gene in the SARS-CoV-2 infection cycle, the disparity in COVID-19 infection and severity could be attributed to enhanced TMPRSS2 gene expression thus, variation in this gene may play such part and therefore may result in a difference in disease susceptibility in different communities around the world. The effect of *TMPRSS2* polymorphism in COVID-19 susceptibility in Indian people is largely unknown and this needs to be investigated. Therefore, we calculated state-wise allele frequencies in Indian populations for all the 5 SNPs (rs2070788, rs734056, rs12329760, rs2276205, and rs3787950) observed in our data. The linear regression analysis was carried out to test the association of these SNP's frequency with COVID-19 CFR among various Indian states (Supplementary Table 4 A, B, and 5). The linear regression analysis performed for rs2070788 (G allele) allele frequency showed a significant positive correlation with CFR among Indian populations;i.e.- $p = 0.029$ (at significance threshold of $p < 0.05$). Among the Indian populations, higher CFR was observed where the allele frequency is higher and vice versa (Fig. 2A and B). The goodness of fit (R^2) explained 33.82% of the variation (Figs. 2C) with an effect size (β) of 0.582, suggesting a large effect of this allele in the Indian population. As this is an active pandemic with changing numbers of infected and dead patients, we confirmed our findings at different

timelines (latest up to August 2021). The recent data reciprocate the previous observation with no substantial difference between the outcomes. Thus, supporting the previous observation of strong positive association (Table 1).

rs2070788 was previously reported to have significantly associated with the susceptibility to influenza infection, therefore, have clinical significance in infectious diseases (Cheng et al., 2015). We used the publicly accessible GTEx database to confirm our associations between rs2070788 (G) allele and *TMPRSS2* expression in the lungs. We obtained expression of the TMPRSS2 gene across different genotypes in eQTL from 515 samples, which shows that the rs2070788 GG genotype had significant higher *TMPRSS2* expression (p-value threshold <0.05) in the lung as compared to the AG and AA genotypes, i.e. GG > AG > AA (p-value = 8.9e-9, A allele normalized effect size = −0.115) (Fig. 3). Thus, the G allele may contribute to severe consequences in SARS-CoV-2 infection in populations with high frequency. We found that G allele frequency in India ranges from 20% to 50%, with the mean frequency of 39%, lowest being in Arunachal Pradesh and highest in Bihar which is in accordance as per data observed which clearly shows Arunachal Pradesh is among those states that show lowest CFR while Bihar and other states are among higher CFR (Supplementary Table 4A and B). Thus this may explain the disparity in severity of the COVID-19 pandemic among various Indian states (Fig. 2 B). Being an androgen-sensitive gene *TMPRSS2* is known to mediate sex-related effects and rs2070788 SNP is known to play an important role (Alshahawey et al., 2020). Higher expression of *TMPRSS2* in males might make them more prone to virus
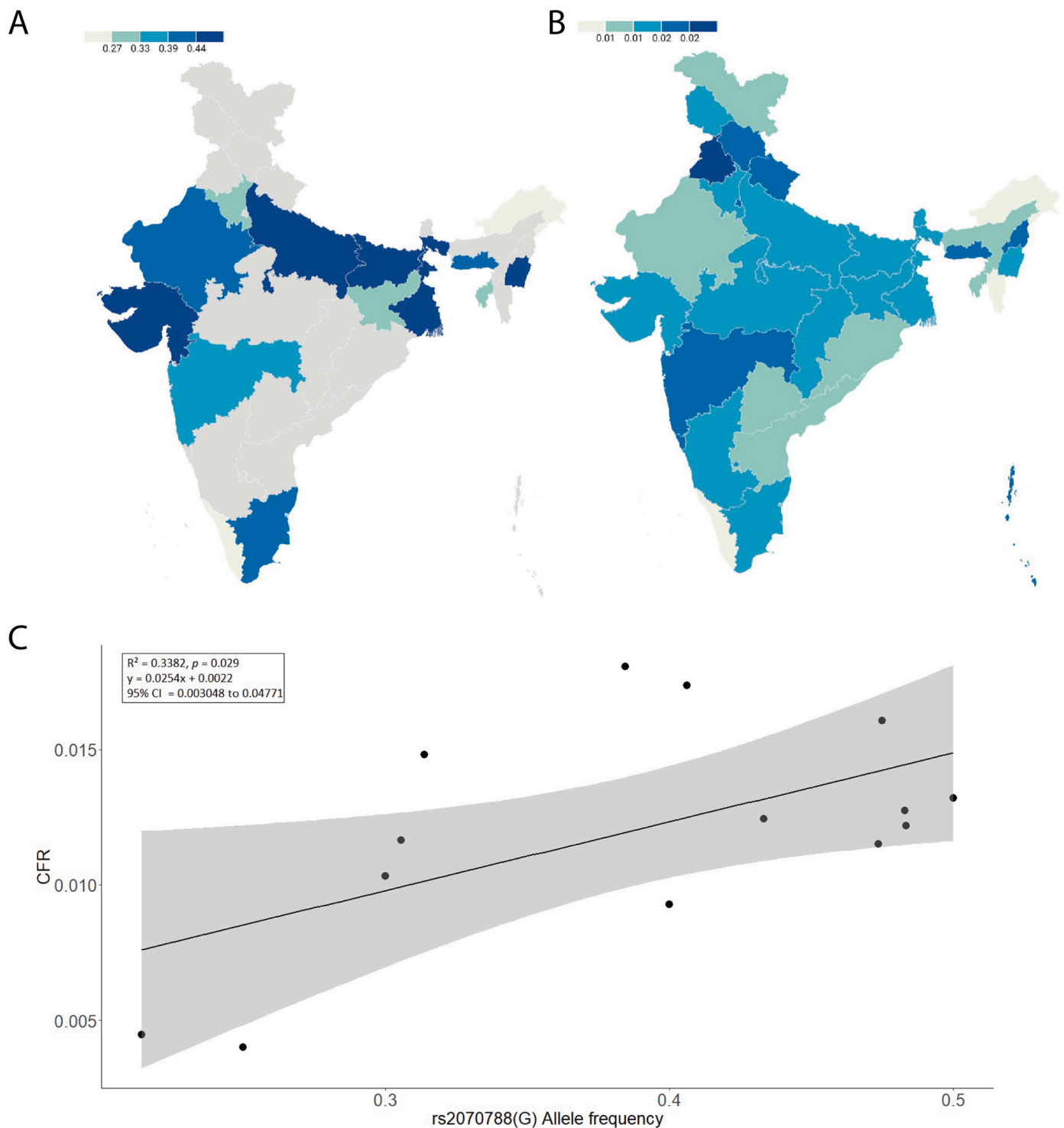
Fig. 2. (A) frequency map (%) showing the spatial distribution of allele rs2070788 among Indian populations. Grey color marks the absence of data. (B) The Map of state-wise frequency (%) of case-fatality rate (CFR) (updated till 30th August 2021). (C) The linear regression graph shows the relationship between rs2070788 (G) allele frequency and COVID-19 CFR of the TMPRSS2 gene.

fusion and could explain high COVID-19 mortality in males (Lamy et al., 2020; Peckham et al., 2020).

For Linkage disequilibrium (LD) analysis, LD plots were made for each population focussing on rs2070788 and nearby SNPs on that haplotype. LD blocks of various sizes were observed among Central Asians, Caucasians, Europeans, South Asians, Siberians, and West Asians. The highest LD level was found in Americans (Supplementary Fig. 2). We also calculated aggregate haplotypes frequency carrying

rs2070788 (G allele), in each population presented in (Supplementary Table 6). Considerable levels of variation in haplotype frequency were observed among the populations. The highest haplotype frequency was observed in America (0.654), while the lowest haplotype frequency was recorded in Southeast Asia Island (0.322), these findings are consistent with epidemiological data available on COVID-19 which clearly shows that the American population has the most number of cases and death while Southeast Asians are much below in the list. We also looked for

Table 1
Results of tests conducted for statistical significance at different timelines of the pandemic in India.

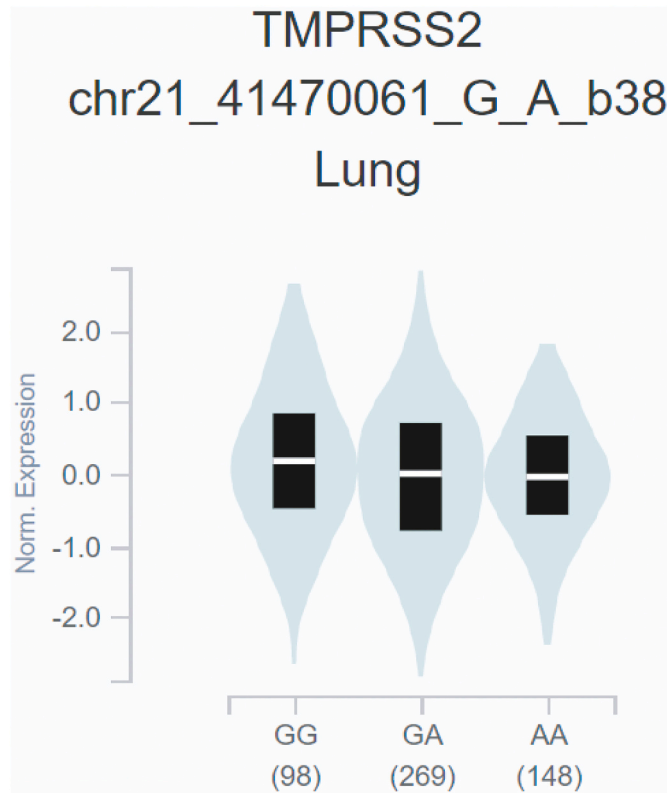| Observation	Linear regression analysis parameters			
rs2070788	R^2	p-value	Effect size (β)	Std. Error (S_e)
June 2021_CFR	0.338	0.029	0.582	0.080467
July 2021_CFR	0.310	0.039	0.557	0.082181
August 2021_CFR	0.289	0.047	0.537	0.083416

Fig. 3. Violin plot of TMPRSS2 expression for *cis* eQTL rs2070788 in Lung tissue. The allelic effect of rs2070788 on normalized *TMPRSS2* gene expression levels is presented by boxplots inside violin plots. G and A alleles indicate the major and minor allele types, respectively with the number of subjects shown under each genotype. The plots indicate the density distribution of the samples in each genotype. The white stripe in the box plot (black) illustrates the median value of the gene expression at each genotype.

worldwide distribution of rs2070788 (G allele) from 1000 genome and gnomAD database (Supplementary Table 7 and Supplementary Fig. 3) and found consistent with the previous observation, rs2070788 (G allele) frequency was highest in Americans, while lowest in African and East Asians populations, this may explain high fatality among Americans populations while African and East Asians being least affected. Low severity among East Asians could be due to adaptation at many genes that engage with coronaviruses, also including the SARS-CoV-2, which began 25,000 years back for coronaviruses, or a related virus outbreak in East Asia at that time (Souilmi et al., 2021).

We caution that this SNP (rs2070788) to be one of the several factors affecting COVID-19 severity; however, the major limitations of the present study is the lack of additional data on covariates such as age, sex, existing comorbidity, virus strain, access to healthcare facilities, and precautionary measures which can alter the CFR.

4. Conclusion

In conclusion for the first time, we have shown a closer affinity of South Asians with the West Eurasian populations for the TMPRSS2 gene. Hence, host disease susceptibility in the context of the *TMPRSS2* gene will be more likely similar to West Eurasian populations. This is in contrast to our prior study on the ACE2 gene, which showed a closer genetic affinity of South Asian haplotypes with East Eurasians. Thus, for South Asians, *ACE2* and *TMPRSS2* have an antagonistic genetic relationship. So, it is worth proposing that the susceptibility of the South Asian population to SARS-CoV-2 will fall somewhere between West and East Eurasian populations, which may be most likely the cause of the moderate susceptibility. We also found a significant genetic association between the rs2070788 (G) allele and COVID-19 CFR among Indian populations. *Cis* eQTL of rs2070788 across different genotypes from the GTEx database further confirms associations between rs2070788(G) allele and *TMPRSS2* expression, which shows G allele is associated with higher expression in the lungs. This information could be used as a genetic biomarker to predict susceptible populations, which may be very useful during the epidemic in policymaking and making better resource allocation.

Data availability statement

All datasets generated for this study are included in the article/Supplementary Material.

Author contributions

GC and RKP conceived and designed this study. RKP, AS, and PPS analyzed the data. RKP, AS, PPS, and GC wrote the manuscript. All authors contributed to the article and approved the submitted version.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2022.105206.

References

Alshahawey, M., Raslan, M., Sabri, N., 2020 Nov. Sex-mediated effects of ACE2 and TMPRSS2 on the incidence and severity of COVID-19; the need for genetic implementation. Curr Res Transl Med. 68 (4), 149–150.

Andolfo, I., Russo, R., Lasorsa, V.A., Cantalupo, S., Rosato, B.E., Bonfiglio, F., et al., 2021 Apr 23. Common variants at 21q22.3 locus influence MX1 and TMPRSS2 gene expression and susceptibility to severe COVID-19. iScience 24 (4), 102322.

Asselta, R., Paraboschi, E.M., Mantovani, A., Duga, S., 2020 Jun 5. ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. Aging. 12 (11), 10087–10098.

Bandelt, H.J., Forster, P., Röhl, A., 1999 Jan 1. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16 (1), 37–48.

Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005 Jan 15. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 21 (2), 263–265.

Bhattacharyya, C., Das, C., Ghosh, A., Singh, A.K., Mukherjee, S., Majumder, P.P., et al., 2020 May. 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 [Internet]. https://doi.org/10.1101/2020.05.04.075911v1 [cited 2021 Aug 28] p. 2020.05.04.075911. Available from:

Budak, G., 2020. ped2fasta - PED to FASTA converter [Internet] [cited 2021 Nov 26]. Available from: https://github.com/gungorbudak/ped2fasta/blob/32c30 f7755be7d07f924aa4d37e082eda0e3fa49/ped2fasta.pl.

Chaubey, G., Ayub, Q., Rai, N., Prakash, S., Mushrif-Tripathy, V., Mezzavilla, M., et al., 2017 Jun 14. "Like sugar in milk": reconstructing the genetic history of the Parsi population. Genome Biol. 18 (1), 110.

Cheng, Z., Zhou, J., To, K.K.-W., Chu, H., Li, C., Wang, D., et al., 2015 Oct 15. Identification of TMPRSS2 as a susceptibility gene for severe 2009 pandemic a (H1N1) influenza and a(H7N9) influenza. J. Infect. Dis. 212 (8), 1214–1221.

Darbani, B., 2020 Jan. The expression and polymorphism of entry machinery for COVID-19 in human: juxtaposing population groups, gender, and different tissues. Int. J. Environ. Res. Public Health 17 (10), 3433.

Durbin, R.M., Altshuler, D., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., et al., 2010 Oct. A map of human genome variation from population-scale sequencing. Nature. 467 (7319), 1061–1073.

Ejaz, H., Alsrhani, A., Zafar, A., Javed, H., Junaid, K., Abdalla, A.E., et al., 2020 Dec 1. COVID-19 and comorbidities: deleterious impact on infected patients. J Infect Public Health. 13 (12), 1833–1839.

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. Mol. Ecol. Resour. 10 (3), 564–567.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al., 2020 Apr 16. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell. 181 (2), 271–280.e8.

Hou, Y., Zhao, J., Martin, W., Kallianpur, A., Chung, M.K., Jehi, L., et al., 2020 Jul 15. New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis. BMC Med. 18 (1), 216.

Huggins, D.J., 2020 Nov. Structural analysis of experimental drugs binding to the SARS-CoV-2 target TMPRSS2. J Mol Graph Model. 1 (100), 107710.

Irham, L.M., Chou, W.-H., Calkins, M.J., Adikusuma, W., Hsieh, S.-L., Chang, W.-C., 2020 Aug 20. Genetic variants that influence SARS-CoV-2 receptor TMPRSS2 expression among population cohorts from multiple continents. Biochem. Biophys. Res. Commun. 529 (2), 263–269.

Iyer, G.R., Samajder, S., Zubeda, S., Dsn, S., Mali, V., Pv, S.K., et al., 2020. Infectivity and progression of COVID-19 based on selected host candidate gene variants. Front. Genet. 11, 861.

Jeon, S., Blazyte, A., Yoon, C., Ryu, H., Jeon, Y., Bhak, Y., et al., 2020. Ethnicity-dependent allele frequencies are correlated with COVID-19 case fatality rate [internet]. Preprints. Oct [cited 2021 Aug 28]. Available from: https://www.authore a.com/users/367817/articles/487091-ethnicity-dependent-allele-frequencies-are-correlated-with-covid-19-case-fatality-rate?commit=92f9ba974af4c5e0ff31 2d7dd9994aa1b1589975.

Kehdy, F.S.G., Pita-Oliveira, M., Scudeler, M.M., Torres-Loureiro, S., Zolini, C., Moreira, R., et al., 2021 Aug. Human-SARS-CoV-2 interactome and human genetic diversity: *TMPRSS2*-rs2070788, associated with severe influenza, and its population genetics caveats in Native Americans. Genet Mol Biol [Internet] 25, 44 [cited 2021 Nov 26]. Available from: http://www.scielo.br/j/gmb/a/LZ5N4QRCYzHRHwh XGFmQCJB/abstract/?lang=en.

Kim, Y.-C., Jeong, B.-H., 2021 Jan. Strong correlation between the case fatality rate of COVID-19 and the rs6598045 single nucleotide polymorphism (SNP) of the interferon-induced transmembrane protein 3 (IFITM3) gene at the population-level. Genes. 12 (1), 42.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018 Jun. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35 (6), 1547–1549.

Lamy, P.-J., Rébillard, X., Vacherot, F., de la Taille, A., 2020 Jun. Androgenic hormones and the excess male mortality observed in COVID-19 patients: new convergent data. World J. Urol. 2, 1–3.

Latini, A., Agolini, E., Novelli, A., Borgiani, P., Giannini, R., Gravina, P., et al., 2020 Sep. COVID-19 and genetic variants of protein involved in the SARS-CoV-2 entry into the host cells. Genes. 11 (9), 1010.

Mollica, V., Rizzo, A., Massari, F., 2020 Sep 1. The pivotal role of TMPRSS2 in coronavirus disease 2019 and prostate cancer. Future Oncol. 16 (27), 2029–2033.

Muschitz, C., Trummert, A., Berent, T., Laimer, N., Knobloch, L., Bodlaj, G., et al., 2021 Apr 27. Attenuation of COVID-19-induced cytokine storm in a young male patient with severe respiratory and neurological symptoms. Wien Klin Wochenschr [Internet]. https://doi.org/10.1007/s00508-021-01867-2 [cited 2021 Aug 28]; Available from.

Pagani, L., Lawson, D.J., Jagoda, E., Mörseburg, A., Eriksson, A., Mitt, M., et al., 2016 Oct. Genomic analyses inform on migration events during the peopling of Eurasia. Nature. 538 (7624), 238–242.

Paniri, A., Hosseini, M.M., Akhavan-Niaki, H., 2021 Jul 3. First comprehensive computational analysis of functional consequences of TMPRSS2 SNPs in susceptibility to SARS-CoV-2 among different populations. J. Biomol. Struct. Dyn. 39 (10), 3576–3593.

Paoloni-Giacobino, A., Chen, H., Peitsch, M.C., Rossier, C., Antonarakis, S.E., 1997 Sep 15. Cloning of the TMPRSS2 gene, which encodes a novel serine protease with

transmembrane, LDLRA, and SRCR domains and maps to 21q22.3. Genomics. 44 (3), 309–320.

Pathak, A.K., Kadian, A., Kushniarevich, A., Montinaro, F., Mondal, M., Ongaro, L., et al., 2018 Dec 6. The genetic ancestry of modern Indus Valley populations from Northwest India. Am. J. Hum. Genet. 103 (6), 918–929.

Peckham, H., de Gruijter, N.M., Raine, C., Radziszewska, A., Ciurtin, C., Wedderburn, L. R., et al., 2020 Dec 9. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. Nat. Commun. 11 (1), 6317.

Piva, F., Sabanovic, B., Cecati, M., Giulietti, M., 2021 Feb 1. Expression and co-expression analyses of TMPRSS2, a key element in COVID-19. Eur. J. Clin. Microbiol. Infect. Dis. 40 (2), 451–455.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., et al., 2007 Sep 1. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81 (3), 559–575.

R: The R Project for Statistical Computing [Internet]. [cited, 2021 Aug 28. Available from: https://www.r-project.org/.

Ragia, G., Manolopoulos, V.G., 2020 Dec 1. Assessing COVID-19 susceptibility through analysis of the genetic and epigenetic diversity of ACE2-mediated SARS-CoV-2 entry. Pharmacogenomics. 21 (18), 1311–1329.

Re3data, 2014. Org. Estonian Biocentre Public Data [cited 2021 Aug 28]; Available from: http://service.re3data.org/repository/r3d100010986.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., et al., 2017 Dec 1. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34 (12), 3299–3302.

Saih, A., Bouqdayr, M., Baba, H., Hamdi, S., Moussamih, S., Bennani, H., et al., 2021 Oct. Computational analysis of missense variants in the human transmembrane protease serine 2 (TMPRSS2) and SARS-CoV-2. Biomed. Res. Int. 19 (2021), e9982729.

Sanyaolu, A., Okorie, C., Marinkovic, A., Patidar, R., Younis, K., Desai, P., et al., 2020 Aug 1. Comorbidity and its impact on patients with COVID-19. SN Compr Clin Med. 2 (8), 1069–1076.

Schönfelder, K., Breuckmann, K., Elsner, C., Dittmer, U., Fistera, D., Herbstreit, F., et al., 2021 Apr. Transmembrane serine protease 2 polymorphisms and susceptibility to severe acute respiratory syndrome coronavirus type 2 infection: a German case-control study. Front. Genet. 21 (12), 667231.

Senapati, S., Kumar, S., Singh, A.K., Banerjee, P., Bhagavatula, S., 2020. Assessment of risk conferred by coding and regulatory variations of TMPRSS2 and CD26 in susceptibility to SARS-CoV-2 infection in human. J. Genet. 99, 53.

SeyedAlinaghi, S., Mehrtak, M., MohsseniPour, M., Mirzapour, P., Barzegary, A., Habibi, P., et al., 2021 May 20. Genetic susceptibility of COVID-19: a systematic review of current evidence. Eur. J. Med. Res. 26 (1), 46.

Sharma S, Singh I, Haider S, Malik MZ, Ponnusamy K, Rai E. ACE2 Homo-dimerization, Human Genomic variants and Interaction of Host Proteins Explain High Population Specific Differences in Outcomes of COVID19 [Internet]. 2020 Apr [cited 2021 Aug 28] p. 2020.04.24.050534. Available from: https://doi.org/10.1101/2020.04.24 .050534v1.

Shen, L.W., Mao, H.J., Wu, Y.L., Tanaka, Y., Zhang, W., 2017 Nov 1. TMPRSS2: a potential target for treatment of influenza virus and coronavirus infections. Biochimie. (142), 1–10.

Singh, H., Choudhari, R., Nema, V., Khan, A.A., 2021 Jan. ACE2 and TMPRSS2 polymorphisms in various diseases with special reference to its impact on COVID-19 disease. Microb. Pathog. 1 (150), 104621.

Souilmi, Y., Lauterbur, M.E., Tobler, R., Huber, C.D., Johar, A.S., Moradi, S.V., et al., 2021 Aug 23. An ancient viral epidemic involving host coronavirus interacting genes more than 20,000 years ago in East Asia. Curr. Biol. 31 (16), 3504–3514.e9.

Srivastava, A., Bandopadhyay, A., Das, D., Pandey, R.K., Singh, V., Khanam, N., et al., 2020a. Genetic association of ACE2 rs2285666 polymorphism with COVID-19 spatial distribution in India. Front. Genet. 11, 1163.

Srivastava, A., Pandey, R.K., Singh, P.P., Kumar, P., Rasalkar, A.A., Tamang, R., et al., 2020 Sep 16. Most frequent south Asian haplotypes of ACE2 share identity by descent with east Eurasian populations. PLoS One 15 (9), e0238255.

Strope, J.D., PharmD, C.H.C., Figg, W.D., 2020 May 21. TMPRSS2: potential biomarker for COVID-19 outcomes. J. Clin. Pharmacol. https://doi.org/10.1002/jcph.1641.

Tätte, K., Pagani, L., Pathak, A.K., Kõks, S., Ho Duy, B., Ho, X.D., et al., 2019 Mar. The genetic legacy of continental scale admixture in Indian Austroasiatic speakers. Sci. Rep. 7 (9), 3818.

THE GTEX CONSORTIUM, Ardlie, K.G., Deluca, D.S., Segrè, A.V., Sullivan, T.J., Young, T.R., et al., 2015 May 8. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 348 (6235), 648–660.

Torre-Fuentes, L., Matías-Guiu, J., Hernández-Lorenzo, L., Montero-Escribano, P., Pytel, V., Porta-Etessam, J., et al., 2021 Feb. ACE2, TMPRSS2, and Furin variants and SARS-CoV-2 infection in Madrid, Spain. J. Med. Virol. 93 (2), 863–869.

Vaarala, M.H., Porvari, K.S., Kellokumpu, S., Kyllönen, A.P., Vihko, P.T., 2001. Expression of transmembrane serine protease TMPRSS2 in mouse and human tissues. J. Pathol. 193 (1), 134–140.

Vargas-Alarcón, G., Posadas-Sánchez, R., Ramírez-Bello, J., 2020 Nov. 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. Life Sci. 1 (260), 118313.

Wang, F., Huang, S., Gao, R., Zhou, Y., Lai, C., Li, Z., et al., 2020 Nov 10. Initial whole-genome sequencing and analysis of the host genetic contribution to COVID-19 severity and susceptibility. Cell Discov. 6 (1), 1–16.

Webb Hooper, M., Nápoles, A.M., Pérez-Stable, E.J., 2020 Jun 23. COVID-19 and racial/ethnic disparities. JAMA. 323 (24), 2466–2467.

Wulandari, L., Hamidah, B., Pakpahan, C., Damayanti, N.S., Kurniati, N.D., Adiatmaja, C. O., et al., 2021. Initial study on TMPRSS2 p.Val160Met genetic variant in COVID-19 patients. *Hum Genomics.* 15, 29.

Zhang, C., Gao, Y., Ning, Z., Lu, Y., Zhang, X., Liu, J., et al., 2019. PGG.SNV: understanding the evolutionary and medical implications of human single nucleotide variations in diverse populations. Genome Biol. 20, 215.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al., 2020 Mar. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 579 (7798), 270–273.