Genetic Association of ACE2 rs2285666 Polymorphism With COVID-19 Spatial Distribution in India

Anshika Srivastava1, Auddittiya Bandopadhyay1, Debashurti Das4, Rudra Kumar Pandey1, Vanya Singh1, Nargis Khanam1, Nikhil Srivastava1, Prajval Pratap Singh1, Pavan Kumar Dubey2, Abhishek Pathak3, Pranav Gupta4, Niraj Rai5, Gazi Nurun Nahar Sultana6 and Gyaneshwer Chaubey1*

1 Cytogenetics Laboratory, Department of Zoology, Banaras Hindu University, Varanasi, India, 2 Prosthodontics Unit, Faculty of Dental Sciences, Institute of Medical Sciences, Varanasi, India, 3 Department of Neurology, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India, 4 Scangene Labs Private Limited, New Delhi, India, 5 Birbal Sahni Institute of Paleosciences, Lucknow, India, 6 Centre for Advanced Research in Sciences (CARS), Genetic Engineering and Biotechnology Research Laboratory, University of Dhaka, Dhaka, Bangladesh

OPEN ACCESS

INTRODUCTION

The progressive and rapid spread of the novel coronavirus SARS-CoV-2, has created a worldwide wave of crisis by profoundly affecting the human health, and global economic stability. Recent researches have shown that the ACE2 (encoding Angiotensin-Converting Enzyme 2) is the main host cell receptor of spike glycoprotein responsible for the infection (Hoffmann et al., 2020; Lu et al., 2020; Zhou et al., 2020). It plays a crucial role for the entry of the virus into the cell to cause the final infection (Lu et al., 2020). ACE2 is a type 1 transmembrane metalloproteinase with homology to ACE, an enzyme long-known to be a key player in the Renin-Angiotensin system (RAS), and a target for the treatment of hypertension (Li et al., 2003; Shi et al., 2014). The secreted protein catalyzes the cleavage of the C-terminal dipeptide of Angiotensin I to produce Angiotensin 1–9 and Angiotensin II. The ACE2 is mainly expressed in vascular endothelial cells, the renal tubular epithelium, and in Leydig cells of the testes (Riordan, 2003; Kuba et al., 2010). PCR based analyses have revealed that...
ACE2 is also expressed in the lung, kidney, gastrointestinal tract, and blood vessels, tissues (Harmer et al., 2002; Ksiazek et al., 2003; Jiang et al., 2014). More recent reports have suggested ACE2 expression in the mucosa of the oral cavity may grant easy access to the virus for a new susceptible host (Xu et al., 2020). This explains the high incidence of pneumonia and bronchitis in those with a severe SARS-CoV-2 infection (Zhou et al., 2020). The ACE2 regulates blood volume, systemic vascular resistance, and thus cardiovascular homeostasis. ACE2 has previously been found to be associated with hypertension, stroke, dyslipidemia, cardiovascular diseases, and kidney diseases (Wang et al., 2014; Pan et al., 2018; Wu et al., 2018; Zhang et al., 2018). People on ACE inhibitors and ARBs (angiotensin II type I receptor blockers) produce more numbers of receptors, raising the question of increased susceptibility to the infection, as the SARS-CoV-2 attaches itself to the ACE receptors on heart and lung tissues (Mehra et al., 2020). The expression of ACE2 is also substantially increased in patients with type 1 or type 2 diabetes, who are treated with ACE inhibitors and ARBs (Fang et al., 2020).

Contemporary studies on the large number of samples have analyzed the genomic variations present among populations worldwide (Cao et al., 2020; Stawiski et al., 2020). It was unanimously shown that a polymorphism rs2285666 present in ACE2, varied significantly among European and Asians (Asselta et al., 2020; Cao et al., 2020). The expression experiments have suggested that the alternate allele (TT-plus strand or AA-minus strand) of rs2285666 elevated the expression of this gene upto 50%, hence may play a role in SARS-CoV-2 susceptibility (Wu et al., 2017; Asselta et al., 2020). Therefore, we studied the haplotype structure as well as association of this allele for COVID-19 susceptibility in India. In particular, we have analyzed the haplotypes downstream to rs2285666 among worldwide populations as well as compared the frequencies of this allele and number of cases and case-fatality rate in India for any existing association.

**MATERIALS AND METHODS**

The worldwide data for haplotype analysis was extracted from recent complete genome analysis (Pagani et al., 2016). NETWORK 5 (Bandelt et al., 1999) was used to construct the median joining network of haplotypes derived from rs2285666 polymorphism. All the SNV’s observed downstream to rs2285666 have been listed in Supplementary Table 1. Illumina HumanHap 730K genotype chip has listed rs2285666 SNV in their panel.
therefore we looked the genotype data generated on this platform. The frequency data of SNV rs2285666, of various populations were extracted by using Plink 1.9 (Chang et al., 2015), from 1,000 genome project data phase 3 (1000 Genomes Project Consortium et al., 2010), data published by Estonian Biocentre (Chaubey et al., 2017; Pathak et al., 2018; Tätte et al., 2019; Estonian Biocentre Public _Data, 2020) and our newly genotyped samples for various Indian states and the Bangladesh (Supplementary Table 2). The frequency maps were generated by https://www.datawrapper.de/. The regression plots were constructed by https://www.graphpad.com/quickcalcs/linear1/ and verified by the Microsoft excel regression calculations. We have also used Pearson’s correlation coefficient test (Benesty et al., 2009) to validate our results. SPSS (ver 25) was used to estimate the Person correlation for two tailed significance test at 95% CI and 1,000 bootstrapping (2,000,000 seeds). The joint plots for all parameters were obtained from customized script of program R (ver 4) (R Core Team, 2012).

RESULTS AND DISCUSSION

Studies have shown that angiotensin-converting enzyme 2 (ACE2) acts as an entry receptor for coronavirus (Li et al., 2003). The interaction between SARS-S and ACE2 has already been expounded at molecular level in detail suggesting ACE2 as key determinant of SARS-CoV transmission (Li F. et al., 2005; Li W. et al., 2005). It has been revealed that human recombinant soluble ACE2 (hrsACE2) inhibits growth of SARS-CoV-2 and interrupts early stages of infections (Monteil et al., 2020). The variable susceptibility to the SARS-CoV-2 infection may be associated with the certain genomic variants within ACE2, that modulate its function or expression.

Among all the common exonic variants, some very recent studies done on ACE2 variants reported population-based frequency differences for a single nucleotide variant (SNV) rs2285666 (also called G8790A) (Asselta et al., 2020; Cao et al., 2020; Strafella et al., 2020). This variant of ACE2 was significantly different for Europeans (0.2), than the East Asians (0.55) (Asselta et al., 2020; Cao et al., 2020). Our analysis on Indian population has revealed mean frequency ~0.6 of this allele (Srivastava et al., 2020). Moreover, our haplotype analysis for this gene revealed excessive sharing of the frequent South Asian haplotypes with East Eurasian populations, rather than West Eurasian populations (Srivastava et al., 2020). Similar to the East Asians, we have also noted that the frequency of this allele is significantly higher (two tailed \( p < 0.0001 \)) among Indian populations in comparison with either of European, American, or African.

Since rs2285666 has already been proven to be a potential risk factor for hypertension, type 2 diabetes, and coronary artery disease (Chaoxin et al., 2013; Asselta et al., 2020), therefore, may possibly be a predisposing factor associated with the SARS-CoV-2 infection.
with the comorbidities observed in COVID-19 patients. Variant rs2285666 is located at the beginning of the intron 3, theoretically affecting gene expression with alternative splicing mechanisms (Li, 2012; Yang et al., 2015). A study also reported for the association of three rs2285666 genotypes with ACE2 protein level measured in serum by ELISA, with the A/A genotype having an expression level almost 50% higher than the G/G genotype (Li, 2012). More recently, it is shown that the substitution of G with A is predicted to increase the strength of the splice site of about 9.2%, resulting higher expression of ACE2 protein (Asselta et al., 2020). It has been also noted that patients characterized by higher ACE1 activity (a protein similar to ACE2) in conjunction with reduced ACE2 activity (i.e., CC/GG females or hemizygous C/G-males for rs2285666) account for increased susceptibility to hypertension, mainly in association with classical cardiovascular risk factors such as old age, dyslipidemia, and diabetes (Pinheiro et al., 2019; Ghafouri-Fard et al., 2020). Thus, it is clear that decreased ACE2 level contributes to severe consequences of SARS-CoV-2 infection (Samavati and Uhal, 2020; Verdecchia et al., 2020).

There has not been any study so far on this SNV among Indian Populations. Therefore, we first looked the haplotype sharing, derived after variant rs2285666 among worldwide populations (Figure 1). Most of the haplotypes downstream to variant rs2285666 were belonging to South Asian, Central Asian, and East Eurasian populations. The starlike structure of rs2285666 derived haplotype indicate a case of positive selection among Asian populations, which needs further exploration (Figure 1 and Supplementary Table 1). Subsequently, our spatial analysis showed that in India, frequency of alternate allele of this SNV (rs2285666) varied between 33% and 100% (Figure 2A and Supplementary Table 2). The frequency gradient (lower to higher) is observed from Northwestern and Western region to Northeastern part of the subcontinent. In order to understand the correlation of allelic frequency with respect to the frequency of cases among Indian populations (Figure 2B), we performed linear regression and Pearson’s correlation coefficient analyses for variant rs2285666 and frequency of cases as well as case-fatality rate (CFR) (Figures 2C,D and Table 1). The regression analysis showed a significant correlation between allele frequency and number of cases ($p < 0.05$) (Figure 2C and Table 1). More number of cases are observed where frequency of this allele is lower and vice versa. The goodness of fit ($R^2$) explained 34.6% of the variation. This suggests that the effect size of this allele for Indian populations is large. Since this is an ongoing pandemic and the number of infected people changes with time, we tested this result by adding the latest number of cases (August 2020) as well as the CFR (Figure 2 and Table 1). The latest data is also consistent with the older observation. We didn’t find any significant difference between both of the results. Moreover, the CFR data showed stronger association with the allele frequency of rs2285666 (Table 1). Further, in order to confirm our findings, we have also performed the Pearson’s correlation coefficient test. The genetic variation (frequency of rs2285666) and number of cases are negatively correlated with $r = -0.571$, $p = 0.05$ (August 2020), as well as frequency of rs2285666 and CFR are also negatively correlated with $r = -0.699$, $p = 0.005$ (Figures 2C,D and Table 1). The Pearson correlation analyses supported previous observations by showing a strong negative correlation of rs2285666 (allele T/A) with the number of cases and case-fatality ratio.

Thus, for the first time, we showed a strong correlation of alternate allele (allele T on plus strand or allele A on minus strand) of variant rs2285666 with the lower infection rate as well as lower CFR among Indian populations. Although whole genome sequencing of a considerably large sample of cases and control individuals in India need to be performed to secure a robust genetic information on susceptibility for the disease, we here establish a possibility of the SNV (rs2285666) being associated with a protective role against COVID-19.

We caution that this is just one of the factors affecting the transmission, however there are several other elements (e.g., variation of rs2285666 among diverse ethnic groups of a state, sex of a person, comorbidity, virus strain, temperature, humidity, population density, social distancing, lockdown, etc.), which can perturb the infection rate and CFR substantially. If more of genetic factors or polymorphisms are recognized that may have played a significant impact on the variability of SARS-CoV-2 course, it would be worthwhile to design a cheap and accurate DNA based test for coronavirus susceptibility.

**DATA AVAILABILITY STATEMENT**

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

**AUTHOR CONTRIBUTIONS**

GC concived and designed this study. AS, AB, DD, RP, VS, NK, NS, PS, PD, AP, PG, NR, and GS collected the data for allele and COVID-19. AS, AB, PS, PD, AP, and GC analyzed the data. AS, AB, DD, PS, and GC wrote the manuscript from the inputs of other co-authors. All authors contributed to the article and approved the submitted version.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: [https://www.frontiersin.org/articles/10.3389/fgene.2020.564741/full#supplementary-material](https://www.frontiersin.org/articles/10.3389/fgene.2020.564741/full#supplementary-material)

**Supplementary Table 1** | The details of SNV’s downstream to the rs2285666 polymorphism. This data is obtained from the Network analysis.

**Supplementary Table 2** | The summary of data used in present analyses.
REFERENCES

1000 Genomes Project Consortium, Durbin, R. M., Abecasis, G. R., Altshuler, D. L., Auton, A., Brooks, L. D., et al. (2010). A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073. doi: 10.1038/nature09534

Asselot, R., Paraboschi, E. M., Mantovani, A., and Duga, S. (2020). ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. medRxiv [Preprint]. doi: 10.1101/2020.02.25.20022518

Bandelt, H. J., Forster, P., and Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16:2060–2069. doi: 10.1093/molbev/16.10.2060

Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). “Pearson correlation coefficient,” in Noise Reduction in Speech Processing, Springer Topics in Signal Processing, vol. 2, Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-00269-0_5

Cao, Y., Li, L., Feng, Z., Wan, S., Huang, P., Sun, X., et al. (2020). Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. Cell Discov. 6, 1–4. doi: 10.1038/s41421-020-0147-14

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer data sets. BMC Bioinformatics 16:2. doi: 10.1186/s12859-014-0047-8

Chaoxin, J., Daili, S., Yanxin, H., Ruwei, G., Chenlong, W., and Yaboin, T. (2013). “Like sugar in milk”: reconstructing the genetic history of the Parsi population. Genome Biol. 18, 110. doi: 10.1186/s13059-017-1244-1249

Estonian Biocentre Public _Data (2020). Available online at: https://evolbio.ut.ee/

Fang, L., Karakiulakis, G., and Roth, M. (2020). Are patients with hypertension and essential hypertension in the Chinese population: a meta-analysis involving 5260 subjects. Front. Physiol. 3:364. doi: 10.3389/fphys.2012.00364

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, M., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181, 271–280. doi: 10.1016/j.cell.2020.02.052

Jiang, F., Yang, J., Zhang, Y., Dong, M., Wang, S., Zhang, Q., et al. (2014). Association between ACE 2 gene polymorphisms with susceptibility to essential hypertension and dyslipidemia in Xinjiang, China. Lipids Health Dis. 17:241.

Pathak, A. K., Kadian, A., Kushneriarevi, A., Montinaro, F., Mondal, M., Onorgo, L., et al. (2018). The genetic ancestry of modern indus valley populations from Northwest India. Am. J. Hum. Genet. 103, 918–929. doi: 10.1016/j.ajhg.2018.10.022

Pinheiro, D. S., Santos, R. S., Jardim, P. C. V., Silva, E. G., Reis, A. A., Pedrino, G. R., et al. (2019). The combination of ACE I/D and ACE G8790A polymorphisms reveals susceptibility to hypertension: a genetic association study in Brazilian patients. PLoS One 14:e0221248. doi: 10.1371/journal.pone.0221248

R Core Team (2012). R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.

Riordan, J. F. (2003). Angiotensin-I-converting enzyme and its relatives. Genome Biol. 4:225.

Samavit, L., and Uhal, B. D. (2020). ACE2. much more than just a receptor for SARS-CoV-2. Front. Cell. Infect. Microbiol. 10:317. doi: 10.3389/fcimb.2020.00317

Shi, C.-S., Qi, H.-Y., Boularan, C., Huang, N.-N., Abu-Asab, M., Shellerame, J. H., et al. (2014). SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome. J. Immunol. 193, 3080–3089. doi: 10.4049/jimmunol.1303196

Srivastava, A., Pandey, R. K., Singh, P. P., Kumar, P., Rasalakra, A. A., Tamang, R., et al. (2020). Most frequent South Asian haplotypes of ACE2 share identity by descent with East Eurasian populations. PLoS One 15:e0238255. doi: 10.1371/journal.pone.0238255

Tätte, K., Pagani, L., Pathak, A. K., Kõks, S., Duy, B. H., Ho, X. D., et al. (2019). The genetic legacy of continental scale admixture in Indian austroasiatic speakers. Mol. Biol. Evol. 36, 1–4. doi: 10.1093/molbev/msz115

Wang, M., Zhang, W., Zhou, Y., and Zhou, X. (2014). Association between ACE gene polymorphisms and the stroke recurrence in Chinese population. J. Stroke Cerebrovasc. Dis. 23, 2770–2780. doi: 10.1016/j.jstrokecerebrovasdis.2018.06.001

Wei, Y., Li, J., Wang, C., Zhang, L., and Qiao, H. (2017). The ACE 2 G8790A polymorphism: involvement in Type 2 diabetes mellitus combined with cerebral stroke. J. Clin. Lab. Anal. 31:e22033. doi: 10.1002/jcla.22033

Frontiers in Genetics | www.frontiersin.org 5 September 2020 | Volume 11 | Article 564741
Xu, H., Zhong, L., Deng, J., Peng, J., Dan, H., Zeng, X., et al. (2020). High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int. J. Oral Sci.* 12, 1–5.

Yang, M., Zhao, J., Xing, L., and Shi, L. (2015). The association between angiotensin-converting enzyme 2 polymorphisms and essential hypertension risk: a meta-analysis involving 14,122 patients. *J. Renin. Angiotensin Aldosterone Syst.* 16, 1240–1244.

Zhang, Q., Cong, M., Wang, N., Li, X., Zhang, H., Zhang, K., et al. (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. doi: 10.1097/MD.0000000000012917

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2017

**Conflict of Interest:** PG was employed by the company Scangene Labs Private Limited.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Copyright © 2020 Srivastava, Bandopadhyay, Das, Pandey, Singh, Khanam, Srivastava, Singh, Dubey, Pathak, Gupta, Rai, Sultana and Chaubey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.**