

The Major Genetic Risk Factor for Severe COVID-19 Does Not Show Any Association Among Indian Populations

Prajval Singh

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

Anshika Shrivastava

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

Gazi Nurun Nahar Sultana

Centre for Advanced Research in Sciences (CARS), Genetic Engineering and Biotechnology Research Laboratory, University of Dhaka, Dhaka, 1000, Bangladesh

Nargis Khanam

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

Abhishek Pathak

Department of Neurology, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India

Prashanth Suravajhala

Department of Biotechnology and Bioinformatics, Birla Institute of Scientific Research Statue Circle, Jaipur, Rajasthan, India

Royana Singh

Department of Anatomy, Institute of Science, Banaras Hindu University, Varanasi, Uttar Pradesh, India-221005

Pankaj Shrivastava

Department of Home (Police), DNA Fingerprinting Unit, State Forensic Science Laboratory, Government of MP, Sagar, India

Kumarasamy Thangaraj

Centre for DNA Fingerprinting and Diagnostics

Gyaneshwer Chaubey (✉ gyaneshwer.chaubey@bhu.ac.in)

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

George van Driem

Institut für Sprachwissenschaft, Länggassstrasse 49, Universität Bern, 3012 Bern, Switzerland

Research Article

Keywords: variable human susceptibility, realtime infection rate, case fatality ratio, risk for COVID-19

DOI: <https://doi.org/10.21203/rs.3.rs-269492/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

With the growing evidence on the variable human susceptibility against COVID-19, it is evident that some genetic loci modulate the severity of the infection. Recent studies have identified several loci associated with greater severity. More recently, a study has identified a 50kb segment introgressed from Neanderthal adding a risk for COVID-19, and this trait is present among 16% and 50% people of European and South Asian origin respectively. Contrary to this finding, our studies on *ACE2* identified a haplotype present among 20% and 60% of European and South Asian populations respectively, which appears to be responsible for the low case fatality ratio among South Asian populations. This result was also consistent with the realtime infection rate and case fatality ratio among various states of India. We readdressed this issue using both of the contrasting datasets and compared them with the realtime infection rates and case fatality ratio in India. We found out that the polymorphism present in the 50kb introgressed segment (rs10490770) did not show any significant correlation with the realtime infection and case fatality ratio in India.

Introduction

Since the beginning of COVID-19 pandemic, it has been observed that people with a different ethnic background and country or continent of origin have variable degrees of susceptibility. Though there are a few well known factors for higher susceptibility, e.g. age and comorbidity ^{1,2}, the variability of the disease has been also reported among healthy people ³. Recent genome wide association study has identified several loci on chromosome 3 associated with the severe risk factor for COVID-19 among Europeans ^{4,5}. Subsequently, Zeberg & Pääbo ⁶ have identified a risk haplotype of 50kb introgressed from Neanderthals, which they called the 'Neanderthal core haplotype'. This risk haplotype was found to be present with an allele frequency of 30% among South Asians, 8% in Europeans and 4% among African Americans. The peak carrier frequency was estimated among the Bangladeshi population, where 63% carried at least one copy of this haplotype. The study also cited twice the risk of mortality among people of Bangladeshi extraction living in the UK as opposed to the native population of Brittanic pedigree ⁷.

Conversely, three of our studies on *ACE2*, the gateway of SARS-CoV-2, identified a haplotype, shared among South Asians and East Eurasians, likely protecting them from severe risk ⁸⁻¹⁰. Additionally, the spatial distribution of this haplotype showed strong association with the low infection as well as low case fatality rate (CFR) ¹⁰. To resolve this discrepancy between the two sets of findings and the associated claims, we have extracted a SNP (rs10490770) reported to be associated with the high risk for COVID-19 ⁶ from our published and unpublished genome wide datasets (Supplementary Table 1), and looked for existing association with the state-wise COVID-19 data of India.

Materials And Method

The genome-wide genotype data by Illumina tagged rs2285666 and rs10490770 SNPs in their panel. Therefore, we searched the genotype datasets generated by this platform. The frequency data for both of the SNPs from various Indian populations were extracted by using Plink 1.9¹¹, from 1000 genome project data phase 3¹² data published by the Estonian Biocentre¹³⁻¹⁶ and our newly genotyped samples for various Indian states and Bangladesh (Supplementary Table 1). In addition to our previous study¹⁰, more samples were added for the SNP rs2285666. The state-wise COVID-19 infection and CFR datasets were extracted from <https://www.covid19india.org/>. The regression estimations and plots were built by <https://www.graphpad.com/quickcalcs/linear1/> and reverified by the Microsoft Excel regression calculations. We have also used Pearson's correlation coefficient test¹⁷ to evaluate the effect of both of the SNPs. The spatial distribution of both the SNPs were drawn by using web tool available on <https://www.datawrapper.de/>.

Results And Discussion

In contrast to the conclusions drawn by Zeberg & Pääbo⁶, our work on *ACE2* identified a haplotype frequent among South Asians and East Eurasians⁸⁻¹⁰. This haplotype is derived by a polymorphism rs2285666 responsible for elevated expression of *ACE2*. We have found high inverse correlation of this haplotype with the state-wise cases as well as case fatality ratio among Indian populations¹⁰. This correlation was significant at various timelines of the pandemic in India (Table 1). We verified the statistical tests with the updated data up to December 2020 and found these data to be consistent with previous observations (Fig. 1 and Supplementary Fig. 1). Thus, it is likely that the *ACE2* SNP rs2285666 has played a significant role in modulating the susceptibility to the disease among Indian populations.

In our search of the SNPs reported to be associated with high risk by Zeberg & Pääbo⁶, we found rs10490770 from genome-wide datasets^{14,15,18-20}. We applied the same tests done for the *ACE2* SNPs (Fig. 1). The state-wise frequency variation of this SNP did not show any association either with the number of cases or the case fatality ratio (Table 1 and Supplementary Figure 1). We repeated these regression tests for the number of cases as well as the case fatality ratio data, obtained during all the three months. However, none of them showed any association with the rs10490770 ($p > 0.3$) (Table 1). The lack of association is striking and suggests instead a complex susceptibility response among Indian populations.

Zeberg & Pääbo⁶ have used the data of higher susceptibility to the disease among the Bangladeshi population living in UK²¹ to support their findings. The higher mortality rate for Bangladeshi population in the UK needs more detailed investigation on comorbidity, relative age, genetic admixture as well as local environment and socio-economic circumstances in their particular British context. More importantly, a similar trend has also been observed among African Americans, where some of the same qualifications may apply *mutatis mutandis*²²⁻²⁴. Furthermore, it is notable that among the Bangladeshi samples analysed by us, the tribal populations of Bangladesh showed almost three times less frequency of rs10490770 (Supplementary Table 1). Therefore, it is advised to explicitly mention the caste and tribal

populations while making any statement about South Asian populations. Significantly, our data also show that the incidence of the allele rs2285666 has been found to occur in the highest frequency of 100% in Indian populations such as the Nishi and Kokborok (Tripuri), who represent Trans-Himalayan language communities (Supplementary Fig. 1 and Supplementary Table 1). As a linguistic phylum, the Trans-Himalayan language family is widespread in parts of eastern Eurasia and includes languages such as Tibetan, Burmese, Mandarin, Cantonese and Hokkien.

Thus, our extensive analyses on realtime data did not show any association of the SNP rs10490770 with the state-wise infection rates as well as CFRs, suggesting that the risk allele for COVID-19 in Europe does not play a significant role in COVID-19 severity in India.

Declarations

Data Availability Statement

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

Author Contributions

GC, KT, RS and PS conceived and designed this study. PPS, AS, GS, NK, AP, PrS, RS, and PS collected the data for alleles and COVID-19. PPS, AS, PrS, and GC analyzed the data. PPS, PrS, KT, GvD and GC wrote the manuscript from the inputs of other co-authors. All authors contributed to the article and approved the submitted version.

Conflict of Interest

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.

References

1. Alberca, R. W., Oliveira, L. de M., Branco, A. C. C. C., Pereira, N. Z. & Sato, M. N. Obesity as a risk factor for COVID-19: an overview. *Crit. Rev. Food Sci. Nutr.* 1–15 (2020).
2. Fang, L., Karakiulakis, G. & Roth, M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir. Med.* (2020).
3. Godri Pollitt, K. J. *et al.* COVID-19 vulnerability: the potential impact of genetic susceptibility and airborne transmission. *Hum. Genomics* **14**, 1–7 (2020).
4. Ellinghaus, D. *et al.* Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* (2020).

5. COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet.* 1 (2020).
6. Zeberg, H. & Pääbo, S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* 1–3 (2020).
7. Public Health England. COVID-19: review of disparities in risks and outcomes. (2020).
8. Srivastava, A. *et al.* Most frequent South Asian haplotypes of ACE2 share identity by descent with East Eurasian populations. *PLOS ONE***15**, e0238255 (2020).
9. Singh, K. K., Chaubey, G., Chen, J. Y. & Suravajhala, P. Decoding SARS-CoV-2 Hijacking of Host Mitochondria in Pathogenesis of COVID-19. *Am. J. Physiol.-Cell Physiol.* (2020) doi:10.1152/ajpcell.00224.2020.
10. Srivastava, A. *et al.* Genetic Association of ACE2 rs2285666 Polymorphism With COVID-19 Spatial Distribution in India. *Front. Genet.***11**, 1163 (2020).
11. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer data sets. *BMC Biol***4**, (2015).
12. 1000 Genomes Project Consortium *et al.* A map of human genome variation from population-scale sequencing. *Nature***467**, 1061–73 (2010).
13. Estonian Biocentre Public_Data. <https://evolbio.ut.ee/>.
14. Chaubey, G. *et al.* “Like sugar in milk”: reconstructing the genetic history of the Parsi population. *Genome Biol.***18**, 110 (2017).
15. Pathak, A. K. *et al.* The Genetic Ancestry of Modern Indus Valley Populations from Northwest India. *Am. J. Hum. Genet.***103**, 918–929 (2018).
16. Tätte, K. *et al.* The genetic legacy of continental scale admixture in Indian Austroasiatic speakers. *Sci. Rep.***9**, 3818 (2019).
17. Benesty, J., Chen, J., Huang, Y. & Cohen, I. Pearson correlation coefficient. in *Noise reduction in speech processing* 1–4 (Springer, 2009).
18. Chaubey, G. *et al.* Population Genetic Structure in Indian Austroasiatic speakers: The Role of Landscape Barriers and Sex-specific Admixture. *Mol. Biol. Evol.***28**, 1013–24 (2011).
19. Metspalu, M. *et al.* Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am. J. Hum. Genet.***89**, 731–44 (2011).
20. Tamang, R. *et al.* Reconstructing the demographic history of the Himalayan and adjoining populations. *Hum. Genet.***137**, 129–139 (2018).
21. Aldridge, R. W. *et al.* Black, Asian and Minority Ethnic groups in England are at increased risk of death from COVID-19: indirect standardisation of NHS mortality data. *Wellcome Open Res.***5**, 88 (2020).
22. Kullar, R. *et al.* Racial disparity of coronavirus disease 2019 in African American communities. *J. Infect. Dis.***222**, 890–893 (2020).
23. Yancy, C. W. COVID-19 and African Americans. *Jama* (2020).

24. Hooper, M. W., Nápoles, A. M. & Pérez-Stable, E. J. COVID-19 and racial/ethnic disparities. *Jama* (2020).

Tables

Table 1. Estimates of Person correlation coefficient for the rs2285666 and rs10490770 with the realtime COVID-19 cases as well as case fatality ratio among Indian populations. The significant values are shown in bold letters.

| | | rs2285666 (T) | rs10490770 (C) | FreqMay2020 | FreqAug2020 | FreqDec2020 | CFRAug2020 | CFRDec2020 |
|-------------------|---------------------|---------------|----------------|---------------|---------------|---------------|----------------|---------------|
| rs2285666 | Pearson Correlation | 1 | -0.428 | -.588* | -.558* | -.580* | -.649** | -.568* |
| | Sig. (2-tailed) | | 0.098 | 0.017 | 0.025 | 0.018 | 0.007 | 0.022 |
| rs10490770 | Pearson Correlation | -0.428 | 1 | 0.204 | 0.164 | 0.266 | 0.244 | -0.026 |
| | Sig. (2-tailed) | 0.098 | | 0.450 | 0.543 | 0.319 | 0.362 | 0.924 |

*. Correlation is significant at the 0.05 level (2-tailed).

** . Correlation is significant at the 0.01 level (2-tailed).