

HYPOTHESIS

Genetic gateways to COVID-19 infection: Implications for risk, severity, and outcomes

Monojit Debnath¹ | Moinak Banerjee² | Michael Berk^{3,4}

¹Department of Human Genetics, National Institute of Mental Health and Neurosciences, Bangalore, India

²Human Molecular Genetics Laboratory, Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram, India

³IMPACT - the Institute for Mental and Physical Health and Clinical Translation, School of Medicine, Barwon Health, Deakin University, Geelong, VIC, Australia

⁴Florey Institute for Neuroscience and Mental Health, Department of Psychiatry and Orygen, The National Centre of Excellence in Youth Mental Health, The University of Melbourne, Melbourne, VIC, Australia

Correspondence

Monojit Debnath, Department of Human Genetics, National Institute of Mental Health and Neurosciences (NIMHANS), Hosur Road, Bangalore 560029, India.
Email: monozeet@gmail.com

Abstract

The dynamics, such as transmission, spatial epidemiology, and clinical course of Coronavirus Disease-2019 (COVID-19) have emerged as the most intriguing features and remain incompletely understood. The genetic landscape of an individual in particular, and a population in general seems to play a pivotal role in shaping the above COVID-19 dynamics. Considering the implications of host genes in the entry and replication of SARS-CoV-2 and in mounting the host immune response, it appears that multiple genes might be crucially involved in the above processes. Herein, we propose three potentially important genetic gateways to COVID-19 infection; these could explain at least in part the discrepancies of its spread, severity, and mortality. The variations within Angiotensin-converting enzyme 2 (ACE2) gene might constitute the first genetic gateway, influencing the spatial transmission dynamics of COVID-19. The Human Leukocyte Antigen locus, a master regulator of immunity against infection seems to be crucial in influencing susceptibility and severity of COVID-19 and can be the second genetic gateway. The genes regulating Toll-like receptor and complement pathways and subsequently cytokine storm induced exaggerated inflammatory pathways seem to underlie the severity of COVID-19, and such genes might represent the third genetic gateway. Host-pathogen interaction is a complex event and some additional genes might also contribute to the dynamics of COVID-19. Overall, these three genetic gateways proposed here might be the critical host determinants governing the risk, severity, and outcome of COVID-19. Genetic variations within these gateways could be key in influencing geographical discrepancies of COVID-19.

KEYWORDS

ACE, COVID-19, cytokine storm, genetics, HLA, immunity, SARS-CoV-2

Abbreviations: ACE-2, angiotensin-converting enzyme 2; ADE, antibody dependent enhancement; BCG, Bacillus Calmette-Guérin; COVID-19, coronavirus disease 2019; DM, diabetes mellitus; eQTL, quantitative trait loci; GWAS, genome wide association study; HLA, human leukocyte antigen; MASP2, Mannan binding lectin serine peptidase 2; MHC, major histocompatibility complex; OR, olfactory receptor; SARS-CoV-2, severe acute respiratory syndrome coronavirus-2; SNPs, single nucleotide polymorphisms; TLR, Toll-like receptor; WHO, world health organization.

This article was fast-tracked under a recently instituted interim policy in which editors may, at their discretion, accept coronavirus-related manuscripts submitted for the Review, Perspectives, and Hypotheses categories without additional review.

1 | INTRODUCTION

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has emerged as one of the most notable epidemics in the history of mankind. The World Health Organization (WHO) has declared it a pandemic and as of 4th May, 2020, COVID-19 affected 3 442 234 people, and 239 740 people have died due to COVID-19 infection across 215 countries and territories (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>). However, the spread of COVID-19 infection shows extreme geographical variations. The highest number of infections have been reported in Europe (1 544 145), followed by America (1 433 756), the Eastern Mediterranean (211 555), Western Pacific (152 774), South East Asia (68 756), and Africa (30 536). Despite high population densities, countries in South East Asia and Africa to date have witnessed comparatively lesser number of COVID-19 infections. The reasons behind this geographical variation is not known. Besides this, there exists significant variation in the symptom profile and severity of COVID-19 infection. Global estimate suggests that only 3.4% of the infected individuals have succumbed to the COVID-19 infection, but here too we see significant variation across continents.

Several factors and mechanisms are proposed to influence COVID-19 pathogenesis. The most notable risk factor is age, followed by co-morbidities, including diabetes, obesity, cardiovascular and cerebrovascular diseases,¹ while the most widely anticipated pathogenic mechanism is a cytokine storm.² Besides this, genetic variations in immune function-related genes such as human leukocyte antigen (HLA) are emerging as a critical determinant of COVID-19 infection.³ These components and mechanisms may partly explain the progression of COVID-19, however, do not shed insights into geographical variation. More interestingly, a strange relationship of COVID-19 spread is seen with certain latitudinal zones. Although climatic conditions and presence of sunlight in these zones could be a factor, it is not known whether there exists genetic correlates to these environmental variables. Lastly, prior exposure to other pathogens may play a role in resilience, including immunisation with the Bacillus Calmette-Guérin (BCG) vaccine, which is a more common in countries with endemic tuberculosis.^{4,5}

It is a well-known fact that host genetic background plays a pivotal role in determining susceptibility or resilience to viral infections. This role begins with the host recognition of the viral particles, presentation of viral peptides to the host immune system and neutralization of the viral infection. In fact, there are several important genetically mediated gateways through which each of these events are executed. The entry of and a full-blown infection from an agent depends on complex interplay between multiple host components, encoded by genes involved in controlling entry of the virus,

peptide presentation and the immune response against it. While the host immune components primarily reflect the differential response to virus, other factors such as comorbid conditions, or environmental variables like temperature and exposure to sunlight might have important mediating roles.

2 | ROLE OF POPULATION GENOMICS IN COVID-19 PANDEMICS

Population genomics plays an important role in genetic landscaping of populations and providing information about susceptibility and protection against infectious diseases. Besides this, host genes influence the severity of infection, virus replication, and inflammation in outbred populations. Therefore, urgent efforts are needed globally to understand the host genetic background, identify risk, or protective genes and their variants and develop genetic networks that would be useful to provide insights into COVID-19 outcomes. Currently, there are no genetic data to support ethnic/geographical variation of COVID-19 on global basis. An international COVID-19 host genetics initiative (<https://www.covid19hg.org/>) has recently been launched aiming to identify genetic determinants of COVID-19 susceptibility, severity, and outcomes. Preliminary analyses suggest an important role of host genetics, especially genes regulating the immune response in the risk and severity of SARS-CoV-2 infection.⁶ Herein, we aim to provide insights into the host genetic background that might play an eminent role in COVID-19 susceptibility, severity, and outcomes as well as in geographical variation.

2.1 | Role of angiotensin-converting enzyme 2 gene in entry of SARS-CoV-2: The first genetic gateway

It has now become apparent that SARS-CoV-2 uses angiotensin-converting enzyme 2 (ACE2), also a receptor for SARS-CoV for entry into the host cell.^{7,8} The expression of ACE2 influences SARS-CoV infection.⁹ The ACE2 gene is located in X chromosome; potential functional variants of ACE2 gene are shown to alter its transcriptional activity. However, the population distribution pattern and its influence on differential susceptibility to SARS-CoV-2 as well as the genetic basis of its differential expression and functional implications in various populations are inadequately known.

Converging evidence based on comparative genetic analysis of data from different databases suggest low frequency missense variants and a lack of natural resistant mutations in different populations.^{10,11} Interestingly, some single nucleotide polymorphisms (SNPs) located in the coding regions of the ACE2 gene exhibit varying allele frequencies

in different populations.¹⁰ The SNPs like rs758278442 and rs759134032 in the region of protective variants (K31R and Y83H) of ACE2 gene show relatively higher frequency of mutant alleles in Asian populations in comparison to global average, comprising of populations mostly from European and American descent. In contrast to this, the frequency of alternate allele of rs763395248 SNP in T92I risk variant is relatively higher in those of European descent compared to a global average, comprising populations of Asian and African descent (<https://www.ncbi.nlm.nih.gov/snp/>). A structural modeling study predicted that certain ACE2 variants might provide potential resistance to SARS-CoV-2 infection.¹² Additionally, the expression quantitative trait loci (eQTL) analysis of ACE2 variants revealed association of some eQTLs with higher expression of ACE2 in tissues and few eQTLs had higher allele frequencies in East Asian Populations than the European populations.¹⁰ This study further demonstrated that differences in allele frequencies of eQTLs of ACE2 observed in different populations could account for diversity of expression pattern of ACE2 gene in populations.¹⁰ Another study indicated significant allele frequency differences of four missense mutations (K26R, 1468V, N720D, and N638S) of ACE2 gene in various populations. Of these, K26R mutated more frequently among Caucasians while 1468V mutated more frequently in Asians. In addition, the authors reported variable ACE2 gene expression among various populations, however, magnitude of differences seems to be small.¹³ Contrary to this, Chen et al¹⁴ suggested that ACE2 allele frequencies are not significantly different among various populations and people of Asian descent exhibit similar ACE2 expression like other groups.¹⁴ In the absence of functional studies, the observations from in silico studies are largely conflicting. Further analysis of population genetics data related to ACE2 gene variations and their impact on the relative levels of its splice variants from individuals inhabiting COVID-19 hotspots should provide more insights into the role of ACE2 gene in susceptibility or resistance to SARS-CoV-2 infection.

Hypertension and Diabetes mellitus (DM) are the most common comorbid conditions in COVID-19; both these conditions are modulated by ACE2. Loss of ACE2 disrupted balance of the renin-angiotensin system, impaired vascular function, and exacerbated diabetic cardiovascular complications.¹⁵ It seems that increased severity of COVID-19 among individuals suffering from hypertension and DM might be driven at least in part by pathological alterations of the ACE2 pathway. Based on this, it can be assumed that besides its role in susceptibility, ACE2 seems to be crucial in outcomes of COVID-19.

A key role of ACE2 has also been demonstrated in inflammatory processes.¹⁶ Genetic deficiency of ACE2 up-regulates the expression of cytokines and induces vascular inflammation in ApoE knockout mouse model.¹⁷ In a recent

study, ACE2 expression was associated with multiple immune signatures such as markers of T cells, B cells, NK cells, and interferon response across various human tissues.¹⁸ These findings suggest that ACE2 not only acts as a receptor for SARS-CoV-2, but is also involved in mediating the post-infection downstream processes including inflammatory responses.

2.2 | HLA, antigen presentation, and protective immunity in SARS-CoV-2 infection: The second genetic gateway

HLA molecules are important immune regulatory components encoded by Major Histocompatibility Complex (MHC) genes. Classical HLA molecules serve as the leading candidates in conferring susceptibility to infectious diseases. HLA genes exhibit extreme diversity and have several thousand reported polymorphisms. It is now well-documented that genetic difference at HLA genes account for individual variations to the immune response against pathogens.¹⁹ Whether distinct HLA allele distribution in various populations confers protection or vulnerability to SARS-CoV-2 is a topic of debate and discussion.

MHC molecules serve as receptors for viral peptides, as they present the peptides to the virus-specific cytotoxic T lymphocytes. CD8+ T cells recognize the conformational structure of the peptide binding groove of the MHC class I molecules, bound to an antigenic peptide. Variation in the conformational structure of the peptide binding groove and its binding to varied peptides are determined by variations within MHC class I genes. Therefore, HLA genes are crucial in MHC-peptide interactions and in differential susceptibilities to viral infection. This phenomenon of antigen presentation helps to gain insights into the pathogenesis of infectious diseases. In a recent Genome Wide Association study (GWAS), HLA was found to act as susceptibility loci for several common infectious diseases.²⁰ The authors performed fine-mapping analysis of the HLA region and suggested a critical role of amino acid polymorphisms in the antigen-binding clefts in influencing association of HLA genes.

Several studies on HLA association were also carried out on SARS-CoV-1, but the results across studies have been inconsistent and conflicting. Some of the risk HLA alleles reported for SARS-CoV-1 included HLA-B*46:01,²¹ HLA-B*07:03,²² HLA-C*08:01,²³ and HLA-DRB1*1202,²⁴ while several alleles also provided protection, such as HLA-DRB1*03:01,²² HLA-C*15:02, and HLA-DRB1*03:01.²⁵ In addition to this, multiple functional studies identified HLA-A*0201 T cell epitopes from SARS-CoV nuclear capsid and spike proteins.^{26,27} Contrary to this, some studies could not replicate these findings and refuted any association between

HLA and risk of SARS-CoV-1.²⁸ So far, there are no empirical studies on the role of HLA genetic diversity in determining the geographical distribution pattern, risk, severity, and outcomes of SARS-CoV-2 infection. However, there is a growing recognition that HLA, being the critical regulator of immune response to viral infections, may play a crucial role in differential susceptibility to SARS-CoV-2 infection.²⁹

In a study on 28 COVID-19 patients with severe respiratory failure, the expression of HLA-DR was very low and this was accompanied by profound reduction of CD4 lymphocytes, CD19 lymphocytes, and natural killer (NK) cells,³⁰ indicating that HLA might play an important immune-regulatory role in COVID-19. An in silico study on genetic variability across HLA class I genes suggested that HLA-A, -B, and -C genes might affect susceptibility and severity of SARS-CoV-2 infection. The authors performed in silico analysis of binding affinity of MHC class I molecules with viral peptide across 145 HLA-A, -B, and -C genotypes for all the peptides of SARS-CoV-2. This study observed that the HLA-B*46:01 allele could increase susceptibility to COVID-19, as this allele had fewest predicted binding peptides for SARS-CoV-2. HLA-B*15:03 in contrast could provide T cell-based protective immunity as this allele displayed highest capacity to present highly conserved SARS-CoV-2 peptides.³ The authors also stated that at haplotype level, HLA-A*02:02, HLA-B*15:03, and HLA-C*12:03 exhibited highest and HLA-A*25:01, HLA-B*46:01, and HLA-C*01:02 displayed lowest predicted repertoire of epitopes from SARS-CoV-2.

Efforts are also being made to predict the MHC class I epitope landscape using SARS-CoV as well as SARS-CoV-2 viral proteomes. In MHC binding assays, some epitopes of SARS-CoV-2 were associated with five distinct HLA alleles, such as HLA-A*02:01, HLA-B*40:01, HLA-DRA*01:01, HLA-DRB1*07:01, and HLA-DRB1*04:01.³¹ Another study also identified previously validated HLA class I peptides and suggested the presence of additional HLA types for SARS-CoV.³² A bioinformatic prediction and molecular modeling study identified highly immunogenic epitopes of SARS-CoV-2 and their corresponding HLA alleles. The authors predicted HLA-A*02:03 and A*31:01 as effective antigen presenters for SARS-CoV-2, implying that these would provide protection, while, HLA-A*03:02 appeared as a risk allele.³³ Subsequent to this, another in silico analysis suggested higher binding tendency of SARS-CoV peptide to HLA-A*02:01.³⁴ Some immunoinformatic analyses suggest that MHC could play important role in the development of epitope-based peptide vaccine against SARS-CoV-2.³⁵ Based on these understanding, HLA molecules could be considered as second genetic gateway for SARS-CoV-2 infection.

Given the distinct pattern of geographical variation in the incidence of COVID-19, there is a growing recognition that population-specific HLA alleles could act as a key intrinsic determinant of protective immunity against

SARS-CoV-2 and make some individuals and population groups resistant or vulnerable to COVID-19. Some of the HLA alleles that were linked to SARS-CoV-1 and the ones that are predicted to influence SARS-CoV-2 display remarkable diversity in certain populations. One such example is HLA-A2 and this has several molecular subtypes specific to Caucasian, African, Oriental, and Asian populations.³⁶ In a recent study, in North and central Indian populations certain alleles of HLA-A*02 were shown to have high frequencies. This included A*02:01, *02:03, *02:05, *02:06, *02:07, and *02:11, of which A*02:11 exhibited highest occurrence at the repertoire level.³⁷ Notably, A*02:11 allele seems to be common in Indian populations, while it is completely absent in Caucasian and oriental populations. The frequency and diversity of certain alleles of HLA-A*02 were also reported to be high in African populations.³⁸

Nguyen et al³ suggests that the HLA-B*46:01 allele could increase susceptibility to COVID-19.³ It is interesting to note that HLA-B*46:01 originated in people of South East Asian descent, and has high distribution in South East Asia,³⁹ while it is completely absent in Indian and African populations and rarely present in European populations (<http://www.allelefrequencies.net/hla6006a.asp>). A functional study indicated that HLA-B*46:01 has low cell surface expression and low-diversity peptidome, possibly due to its prolonged association with HLA-specific chaperons and intracellular retention.^{40,41} Data generated from epidemiological studies indicate that HLA-B*46:01 carriers are more susceptible to tuberculosis, malaria, HIV, and SARS coronavirus.^{21,42} However, a study from China reported that HLA-B46 provides protection against *Mycobacterium leprae*, causative agent of leprosy.⁴³ Similarly, the protective allele, HLA-B*15:03 is completely absent in East Asian gene pool while it is the most frequent allele in populations of African descent (<http://www.allelefrequencies.net/hla6006a.asp>). The susceptibility allele, HLA-C*12:03 seems to be the most frequent allele in the European descent. This indicates that susceptibility based on HLA loci seems to vary in people from different backgrounds.

Based on their comprehensive in silico analyses, Nguyen et al³ did not find any correlation between the frequency of HLA alleles in the populations and the binding capacity of HLA alleles with SARS-CoV-2.³ It is premature to conclude that HLA diversity may not have any influence on the spread of SARS-CoV-2 across various populations. HLA-A*02 subtypes that are predicted to be protective exhibits higher frequency in Indian and several African populations (<http://www.allelefrequencies.net/hla6006a.asp>). Incidentally, the risk and severity of COVID-19 in Indian and African populations have been comparatively less than European and other world populations, although in the USA, people of African descent have been hardest hit.⁴⁴ Clearly, genetic factors interact with social factors complicating interpretation. Furthermore, HLA frequency

data available in the database represent a small subset of the studied populations, and thus, may not reflect the actual gene pool of the populations. Therefore, in order to discern the influence of HLA genes on the risk, severity, and outcome of SARS-CoV-2, comprehensive HLA genotype data from all major populations of the world need to be generated. An additional mechanism that facilitate entry of the coronavirus and mount downstream deleterious effects could be Antibody Dependent Enhancement (ADE).⁴⁵ ADE influences immune responses and leads to sustained inflammation and a cytokine storm; this mechanism could be a biological explanation for the severity of SARS-CoV-2 in individuals who had prior exposure to other coronaviruses.⁴⁶ However, the precise role of ADE in geographical discrepancy and severity of SARS-CoV-2 remains to be elucidated empirically.

HLA genes may also affect the phenotypes associated with COVID-19 infection. Olfactory dysfunction has been reported to be a clinical presentation of mild to moderate forms of COVID-19.⁴⁷ Subsequent to this, a study from Europe suggest that half of the patients with COVID-19 had olfactory impairment,⁴⁸ while another study from America suggest that almost all the patients exhibited decreased smell function and this has been proposed as a phenotype of COVID-19.⁴⁹ HLA genes are essential in individual olfactory perception.^{50,51} Importantly, olfactory receptor (OR) gene appear to be MHC-linked and polymorphisms within OR gene contribute to extended HLA/OR-haplotypes.^{52,53} The association between HLA and olfactory dysfunction has not yet been tested in COVID-19 but it would be interesting to examine the particular role of HLA in olfactory dysfunction.

2.3 | Genetic underpinning of the cytokine storm: The 3rd genetic gateway

It is becoming increasingly apparent that SARS-CoV-2 infection leads to a cytokine storm and then an exacerbated inflammatory response. In a recent study, COVID-19 patients admitted to ICU were found to have markedly elevated plasma levels of IL2, IL7, IL10, GCSF, IP10, MCP1, MIP1A, and TNF α .⁵⁴ Another study demonstrated that higher levels of IL-6 (≥ 80 pg/mL) were correlated with a 22 times greater risk of respiratory failure in COVID-19 patients.⁵⁵ Given the strong implications of the cytokine storm and accompanying inflammation, use of anti-inflammatory agents is increasingly being envisaged.^{56,57}

Cytokine production depends on a cascading immunological event. T lymphocytes are prolific producers of cytokines. The development, differentiation, and activation of T lymphocytes and production of cytokines are controlled by genetic as well as epigenetic processes. Polymorphisms

within the cytokine genes are associated with the production of serum levels of cytokines. It is now well-established that ethnicity influences the distribution pattern of cytokine gene polymorphisms.⁵⁸⁻⁶⁰ Contextually, cytokine gene polymorphisms confer risk or protection to infectious diseases. Several studies have shown associations between cytokine gene polymorphisms, such as *IFN- γ* +874A allele,⁶¹ *IL12RB1*,⁶² and SARS. Until now, there are no studies on the impact of cytokine gene polymorphisms on the risk of SARS-CoV-2. However, given an important role of genetic background in determining the pattern of inflammatory response, and a critical role of cytokines in influencing the severity of COVID-19, genetic factors regulating inflammation seems to play an overarching role in COVID-19.

Interestingly, our findings as well as data obtained from other world populations suggest that allelic variations in cytokine genes in global populations show a strong latitudinal impact.⁶³ The proximity of Indian, Mexican, and African populations can be attributed to the geographical position of these populations with respect to latitude. Geographical latitude is a strong environmental factor that is influenced by our evolutionary history with respect to environmental selection. Thus, latitude seems to be a proxy variable which is related to a broad range of factors including genetic background (eg, genetic variation between populations at different latitudes reflects a complex mix of ancient and recent genetic admixture), biometeorological variables (eg, temperature, ultraviolet radiation, and rainfall), and socio-economic factors (eg, developed countries tend to be clustered within higher latitude bands). Regarding the role of biometeorological factors, the impact of sunlight in tropical and temperate regions differs significantly and sunlight is required for synthesis of Vitamin D, which in turn plays an important role in maintaining the immune homeostasis. Production of vitamin D3 requires the action of UVB radiation on 7-dehydrocholesterol present in the skin, which mainly occurs in the Northern Hemisphere between April and September, as sunlight is low during winter months. Genetic factors are known to account up to 28% of inter-individual variability in serum 25(OH)D concentrations.⁶⁴ Genetic as well as population differences of vitamin D status have been reported across various populations.⁶⁵ There is a possibility that vitamin D status in the populations across continents might have some influence on geographical variance of COVID-19.

Deficiency in vitamin D can result in increased autoimmunity and also increase the susceptibility to infection. Vitamin D diminish the production of proinflammatory cytokines including TNF- α , IFN- γ , etc. and induce the expression of anti-inflammatory cytokines. Vitamin D status plays a role in susceptibility to COVID-19, with deficiency conferring greater risk and supplements of potential value.⁶⁶ Possibly, vitamin D is critically involved in regulating cytokine storm and the outcome of COVID-19.

Taken together, cytokine genes could be considered as the third genetic gateway of COVID-19. However, cytokine expression depends on various upstream regulators, such as TLR and also potentially interact with other components of innate immunity, such as complement components. The importance of TLR and complement pathways with respect to cytokine storm is discussed in the following sections.

2.4 | TLR-pathway in SARS-CoV-2 infection

TLRs are a family of innate immune sensor proteins and play important role in infection and immunity. TLR signaling plays critical role in the regulation of cytokine expression in the immune system⁶⁷; hence TLR signaling could be crucially involved in cytokine storm in SARS-CoV-2 infection.

Activation of TLR pathway leads to inflammation and/or immune activation. Previous studies suggest that TLRs are required for initiating the innate immune response to SARS-CoV infection. Mice deficient in certain TLRs such as TLR3 as well as TLR3/TLR4 adaptor TRIF are more susceptible to SARS-CoV infection. Increased numbers of inflammatory cell types and alterations in inflammation accompany this infection.⁶⁸ In addition, allelic variation in the Toll like receptor adaptor protein, Ticam2 influences susceptibility to SARS-CoV infection in mice; Ticam2^{-/-} mice had high susceptibility to SARS-CoV infection.⁶⁹ Currently, there are no studies on the role of TLR pathway in SARS-CoV-2 infection. However, previous studies suggest that genetic variation within TLR or components of TLR pathway influenced SARS-CoV infection.

It is a well-documented fact that the expression of TLR molecules are also determined by genetic variation within the TLR genes. It is noteworthy that the TLR genes exhibit a distinct population distribution pattern and are the target of selection pressure. In addition, different TLRs differ in their immunological redundancy, suggesting that they have distinct contribution to host defence.⁷⁰ Given such understanding that TLR genes have a distinct distribution pattern in various ethnic populations and with important role in innate immune signaling, TLR genes could emerge as a crucial determinant of differential susceptibility to SARS-CoV-2 infection and severity. However, this needs to be tested empirically.

2.5 | Complement activation pathway in SARS-CoV-2 infection

The complement system plays an important role in host defence against infection, including viral infections.⁷¹ Complement components interact with TLRs and promote the development of inflammatory immune cells, especially

Th17 cells.⁷² A recent study suggested that the complement system acts as an important host mediator of SARS-CoV infection. SARS-CoV-infected C3^{-/-} mice showed less respiratory dysfunction and reduced levels of cytokines and chemokines in both the sera and lungs.⁷³ Further to this, complement hyper-activation was seen in COVID-19 patients and highly pathogenic coronavirus N protein aggravated MASP-2-mediated complement activation.⁷⁴ This suggests that complement components have important implications in the induction of the cytokine storm and inflammation in SARS-CoV-2 infection. Polymorphisms of complement genes are associated with the risk of various diseases, including infectious diseases. There are no data on the impact of complement gene variation on the risk and severity of SARS-CoV-2 infection. However, along with TLR and cytokines, genes controlling the function of the complement pathway could provide important pointers to the magnitude of inflammatory responses in COVID-19.

All these genetic pathways are likely to interact among each other and the quantum of interaction will determine the risk. This quantum of interaction will depend on the prevalence of risk variants and environmental exposure.

3 | CONCLUSION

There is a growing recognition that genes, especially those regulating the host immune response might confer differential susceptibility and influence the severity and outcomes of SARS-CoV-2 infection. Multiple in silico and molecular prediction studies indicate an important role of various genes coding ACE2, HLA, cytokine, TLR, and complement components in COVID-19. Many of these genes display distinct geographical, population-specific variation and confers susceptibility and/or resistance to various viral diseases. The current pandemic nature of COVID-19 also indicates distinct geographical pattern with respect to incidence and mortality. It is also uncertain what role prior exposure to other coronavirus species might have in influencing geographical discrepancy and severity of SARS-CoV-2, and such exposure might be regionally specific. microRNAs seem to play important role in the pathogenesis of COVID-19. miRNAs also exhibit population differences and might also provide important clues on the susceptibility or protection to COVID-19.

Some preliminary reports suggest that use of certain drugs such as anti-retro viral, anti-malarial and even vaccines, such as BCG are effective against COVID-19 in certain populations and might explain population variance. These observations might have some relevance to the host genetic background and the above genes might be the key determinants of such benefits. Given this understanding, these genes might play significant role in population-specific incidence of COVID-19, however, it is premature to

make such a claim. Further functional validation of in silico and structural prediction-based studies are required to substantiate the notion that genes predispose to vulnerability to COVID-19. Evaluation of host genetic signatures of COVID-19 would be fundamental to diagnosis, phenotypic evaluations, medication, and therapeutic response.

ACKNOWLEDGMENTS

MB is supported by a NHMRC Senior Principal Research Fellowship (1156072). Dr Sangeeta Maity critically read the first draft and provided some opinions, which were helpful in improving the content of this manuscript. The authors are grateful to Dr Maity.

CONFLICT OF INTEREST

The authors have no conflicts to declare.

AUTHOR CONTRIBUTIONS

M. Debnath conceived the idea and wrote the first draft; M. Banerjee and M. Berk reviewed the literature, manuscript draft, provided critical comments, and added further dimension to the first draft.

REFERENCES

- Yang J, Zheng Y, Gou X, et al. Prevalence of comorbidities and its effects in coronavirus disease 2019 patients: a systematic review and meta-analysis. *Int J Infect Dis.* 2020;94:91-95.
- Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet.* 2020;395:1033-1034.
- Nguyen A, David JK, Maden SK, et al. Human leukocyte antigen susceptibility map for SARS-CoV-2. *J Virol.* 2020.
- Floc'h F, Werner GH. Increased resistance to virus infections of mice inoculated with BCG (Bacillus calmette-guerin). *Ann d'immunol.* 1976;127:173-186.
- Gursel M, Gursel I. Is global BCG vaccination-induced trained immunity relevant to the progression of SARS-CoV-2 pandemic? *Allergy.* 2020.
- Tanigawa Y, Rivas M. Initial review and analysis of COVID-19 host genetics and associated phenotypes. Preprints. 2020;2020030356. <https://doi.org/10.20944/preprints202003.0356.v1>
- Li W, Moore MJ, Vasilieva N, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature.* 2003;426(6965):450-454.
- Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell.* 2020;181(2):271-280.e8.
- Hoffmann H, Geier M, Marzi A, et al. Susceptibility to SARS coronavirus S protein-driven infection correlates with expression of angiotensin converting enzyme 2 and infection can be blocked by soluble receptor. *Biochem Biophys Res Comm.* 2004;319:1216-1221.
- Cao Y, Li L, Feng Z, et al. Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. *Cell discovery.* 2020;6:11.
- Stawiski EW, Diwanji D, Suryamohan K, et al. Human ACE2 receptor polymorphisms predict SARS-CoV-2 susceptibility. *BioRxiv.* 2020. <https://doi.org/10.1101/2020.04.07.024752>
- Hussain EW, Diwanji D, Suryamohan K, et al. Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. *J Med Virol.* 2020. doi:10.1002/jmv.25832
- Li Q, Cao Z, Rahman P. Genetic variability of human angiotensin-converting enzyme 2 (hACE2) among various ethnic populations. *BioRxiv.* 2020. <https://doi.org/10.1101/2020.04.14.041434>
- Chen Y, Shan K, Qian W. Asians and other races express similar levels of and share the same genetic polymorphisms of the SARS-CoV-2 cell-entry receptor. Preprints. 2020;2020020258. <https://doi.org/10.20944/preprints202002.0258.v1>
- Patel VB, Bodiga S, Basu R, et al. Loss of angiotensin-converting enzyme-2 exacerbates diabetic cardiovascular complications and leads to systolic and vascular dysfunction: a critical role of the angiotensin II/AT1 receptor axis. *Circ Res.* 2012;110:1322-1335.
- Gaddam RR, Chambers S, Bhatia M. ACE and ACE2 in inflammation: a tale of two enzymes. *Inflamm Allergy Drug Targets.* 2014;13:224-234.
- Thomas MC, Pickering RJ, Tsorotes D, et al. Genetic Ace2 deficiency accentuates vascular inflammation and atherosclerosis in the ApoE knockout mouse. *Circ Res.* 2010;107:888-897.
- Li M-Y, Li L, Zhang Y, Wang X-S. Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. *Infect Dis Poverty.* 2020;9:45.
- Blackwell JM, Jamieson SE, Burgner D. HLA and infectious diseases. *Clin Microbiol Rev.* 2009;22:370-385.
- Tian C, Hromatka BS, Kiefer AK, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat Commun.* 2017;8:599.
- Lin M, Tseng HK, Trejaut JA, et al. Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Med Genet.* 2003;4:9.
- Ng MH, Lau KM, Li L, et al. Association of human-leukocyte-antigen class I (B*0703) and class II (DRB1*0301) genotypes with susceptibility and resistance to the development of severe acute respiratory syndrome. *J Infect Dis.* 2004;190:515-518.
- Chen YM, Liang SY, Shih YP, et al. Epidemiological and genetic correlates of severe acute respiratory syndrome coronavirus infection in the hospital with the highest nosocomial infection rate in Taiwan in 2003. *J Clin Microbiol.* 2006;44:359-365.
- Keicho N, Itoyama S, Kashiwase K, et al. Association of human leukocyte antigen class II alleles with severe acute respiratory syndrome in the Vietnamese population. *Hum Immunol.* 2009;70:527-531.
- Wang SF, Chen KH, Chen M, et al. Human-leukocyte antigen class I Cw 1502 and class II DR 0301 genotypes are associated with resistance to severe acute respiratory syndrome (SARS) infection. *Viral Immunol.* 2011;24:421-426.
- Tsao YP, Lin JY, Jan JT, et al. HLA-A*0201 T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus nucleocapsid and spike proteins. *Biochem Biophys Res Comm.* 2006;344:63-71.
- Wang B, Chen H, Jiang X, et al. Identification of an HLA-A*0201-restricted CD8+ T-cell epitope SSp-1 of SARS-CoV spike protein. *Blood.* 2004;104:200-206.
- Xiong P, Zeng X, Song MS, et al. Lack of association between HLA-A, -B and -DRB1 alleles and the development of SARS:

- a cohort of 95 SARS-recovered individuals in a population of Guangdong, southern China. *Int J Immunogenet.* 2008;35:69-74.
29. Sanchez-Mazas A. HLA studies in the context of coronavirus outbreaks. *Swiss Med Wkly.* 2020;150:w20248.
 30. Giamarellos-Bourboulis EJ, Netea MG, Rovina N, et al. Complex immune dysregulation in COVID-19 patients with severe respiratory failure. *Cell Host Microbe.* 2020.
 31. Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses.* 2020;12(no.3), 254
 32. Campbell KM, Steiner G, Wells DK, Ribas A, Kalbasi A. Prediction of SARS-CoV-2 epitopes across 9360 HLA class I alleles. *BioRxiv.* 2020. <https://doi.org/10.1101/2020.03.30.016931>
 33. Romero-Lopez JP, Carnalla-Cortes M, Pacheco-Olvera DL, et al. Prediction of SARS-CoV spike protein epitopes reveals HLA-associated susceptibility. *Res Square.* 2020. <https://doi.org/10.21203/rs.3.rs-25844/v1>
 34. Hyun-Jung Lee C, Koohy H. In silico identification of vaccine targets for 2019-nCoV. *F1000Research.* 2020;9:145.
 35. Bhattacharya M, Sharma AR, Patra P, et al. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): immunoinformatics approach. *J Med Virol.* 2020.
 36. Middleton D, Williams F, Meenagh A, et al. Analysis of the distribution of HLA-A alleles in populations from five continents. *Hum Immunol.* 2000;61:1048-1052.
 37. Saxena A, Sharma G, Tyagi S, et al. HLA-A*02 repertoires in three defined population groups from North and Central India: Punjabi Khatris, Kashmiri Brahmins and Sahariya Tribe. *Hla.* 2019;93:16-23.
 38. Cao K, Moormann AM, Lyke KE, et al. Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. *Tissue Antigens.* 2004;63:293-325.
 39. González-Galarza F, Takeshita L, Santos E, et al. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res.* 2015;43:D784-D788.
 40. Sibilio L, Martayan A, Setini A, et al. A single bottleneck in HLA-C assembly. *J Biol Chem.* 2008;283:1267-1274.
 41. Hilton HG, McMurtrey CP, Han AS, et al. The intergenic recombinant HLA-B *46:01 has a distinctive peptidome that includes KIR2DL3 ligands. *Cell Rep.* 2017;19:1394-1405.
 42. Huang X, Ling H, Mao W, et al. Association of HLA-A, B, DRB1 alleles and haplotypes with HIV-1 infection in Chongqing, China. *BMC Infect Dis.* 2009;9:201.
 43. Wang LM, Kimura A, Satoh M, Mineshita S. HLA linked with leprosy in southern China: HLA-linked resistance alleles to leprosy. *Int J Leprosy Other Mycobact Dis.* 1999;67:403-408.
 44. Yancy CW. COVID-19 and African Americans. *JAMA.* 323 2020. (no.19), 1891–1892.
 45. Wan Y, Shang J, Sun S, et al. Molecular mechanism for antibody-dependent enhancement of coronavirus entry. *J Virol.* 2020;94(no.5), e02015–e02019.
 46. Tetro JA. Is COVID-19 receiving ADE from other coronaviruses? *Microbes Infect.* 2020;22:72-73.
 47. Lechien JR, Chiesa-Estomba CM, De Siati DR, et al. Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. *Eur Arch Otorhinolaryngol.* 2020.1–11.
 48. Klopfenstein T, Kadiane-Oussou NJ, Toko L, et al. Features of anosmia in COVID-19. *Méd Mal Infect.* 2020.
 49. Moein ST, Hashemian SMR, Mansourafshar B, Khorram-Tousi A, Tabarsi P, Doty RL. Smell dysfunction: a biomarker for COVID-19. *Int Forum Allergy Rhinol.* 2020.
 50. Secundo L, Snitz K, Weissler K, et al. Individual olfactory perception reveals meaningful nonolfactory genetic information. *Proc Natl Acad Sci USA.* 2015;112:8750-8755.
 51. Milinski M, Croy I, Hummel T, Boehm T. Major histocompatibility complex peptide ligands as olfactory cues in human body odour assessment. *Proc Biol Sci.* 2013;280:20122889.
 52. Younger RM, Amadou C, Bethel G, et al. Characterization of clustered MHC-linked olfactory receptor genes in human and mouse. *Genome Res.* 2001;11:519-530.
 53. Ehlers A, Beck S, Forbes SA, et al. MHC-linked olfactory receptor loci exhibit polymorphism and contribute to extended HLA/OR-haplotypes. *Genome Res.* 2000;10:1968-1978.
 54. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020;395:497-506.
 55. Herold T, Jurincovic V, Arnreich C, et al. Level of IL-6 predicts respiratory failure in hospitalized symptomatic COVID-19 patients. *MedRxiv.* 2020. <https://doi.org/10.1101/2020.04.01.20047381>
 56. Zhao M. Cytokine storm and immunomodulatory therapy in COVID-19: Role of chloroquine and anti-IL-6 monoclonal antibodies. *Int J Antimicrob Agents.* 2020;105982.
 57. Zhang W, Zhao Y, Zhang F, et al. The use of anti-inflammatory drugs in the treatment of people with severe coronavirus disease 2019 (COVID-19): the perspectives of clinical immunologists from China. *Clin Immunol.* 2020;214:108393.
 58. Hoffmann SC, Stanley EM, Cox ED, et al. Ethnicity greatly influences cytokine gene polymorphism distribution. *Am J Transplant.* 2002;2:560-567.
 59. Manchanda PK, Bid HK, Mittal RD. Ethnicity greatly influences the interleukin-1 gene cluster(IL-1b promoter, exon-5 and IL-1Ra) polymorphisms: a pilot study of a north Indian population. *Asian Pac J Cancer Prev.* 2005;6:541-546.
 60. Cox ED, Hoffmann SC, DiMercurio BS, et al. Cytokine polymorphic analyses indicate ethnic differences in the allelic distribution of interleukin-2 and interleukin-6. *Transplantation.* 2001;72:720-726.
 61. Chong WP, Ip WK, Tso GH, et al. The interferon gamma gene polymorphism +874 A/T is associated with severe acute respiratory syndrome. *BMC Infect Dis.* 2006;6:82.
 62. Tang F, Liu W, Zhang F, et al. IL-12 RB1 genetic variants contribute to human susceptibility to severe acute respiratory syndrome infection among Chinese. *PLoS One.* 2008;3:e2183.
 63. Srinivas L, Vellichirammal NN, Alex AM, Nair C, Nair IV, Banerjee M. Pro-inflammatory cytokines and their epistatic interactions in genetic susceptibility to schizophrenia. *J Neuroinflammation.* 2016;13:105.
 64. Shea MK, Benjamin EJ, Dupuis J, et al. Genetic and non-genetic correlates of vitamins K and D. *Eur J Clin Nutr.* 2009;63:458-464.
 65. Lips P, Cashman KD, Lamberg-Allardt C, et al. Current vitamin D status in European and Middle East countries and strategies to prevent vitamin D deficiency: a position statement of the European Calcified Tissue Society. *Eur J Endocrinol.* 2019;180:P23-P54.
 66. Grant WB, Lahore H, McDonnell SL, et al. Evidence that vitamin D supplementation could reduce risk of influenza and COVID-19 infections and deaths. *Nutrients.* 2020;12(no.4), 988

67. Ozato K, Tsujimura H, Tamura T. Toll-like receptor signaling and regulation of cytokine gene expression in the immune system. *BioTechniques*. 2002;33:Suppl, 66–68, 70, 72 passim.
68. Totura AL, Whitmore A, Agnihothram S, et al. Toll-like receptor 3 signaling via TRIF contributes to a protective innate immune response to severe acute respiratory syndrome coronavirus infection. *mBio*. 2015;6:e00638-1615.
69. Gralinski LE, Menachery VD, Morgan AP, et al. Allelic variation in the Toll-like receptor adaptor protein Ticam2 contributes to SARS-coronavirus pathogenesis in mice. *G3 (Bethesda)*. 2017;7:1653-1663.
70. Barreiro LB, Ben-Ali M, Quach H, et al. Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genet*. 2009;5:e1000562.
71. Stoermer KA, Morrison TE. Complement and viral pathogenesis. *Virology*. 2011;411:362-373.
72. Fang C, Zhang X, Miwa T, Song WC. Complement promotes the development of inflammatory T-helper 17 cells through synergistic interaction with Toll-like receptor signaling and interleukin-6 production. *Blood*. 2009;114:1005-1015.
73. Gralinski LE, Sheahan TP, Morrison TE, et al. Complement activation contributes to severe acute respiratory syndrome coronavirus pathogenesis. *mBio*. 2018;9(no.5.), e01753–18.
74. Gao T, Hu M, Zhang X, et al. Highly pathogenic coronavirus N protein aggravates lung injury by MASP-2-mediated complement over-activation. *MedRxiv*. 2020. <https://doi.org/10.1101/2020.03.29.20041962>

How to cite this article: Debnath M, Banerjee M, Berk M. Genetic gateways to COVID-19 infection: Implications for risk, severity, and outcomes. *The FASEB Journal*. 2020;00:1–9. <https://doi.org/10.1096/fj.202001115R>