Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A survey of genetic variants in SARS-CoV-2 interacting domains of ACE2, TMPRSS2 and TLR3/7/8 across populations

In-Hee Leea,1, Ji-Won Leeb,c,1, Sek Won Konga,d,⁎

a Computational Health Informatics Program, Boston Children's Hospital, Boston, MA 02115, USA
b Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA
c Department of Nephrology, Boston Children's Hospital, Boston, MA 02115, USA
d Department of Pharmacology, Graduate School of Dental Medicine, Hokkaido University, Sapporo 060-8549, Japan

ABSTRACT

The COVID-19 pandemic highlighted healthcare disparities in multiple countries. As such morbidity and mortality vary significantly around the globe between populations and ethnic groups. Underlying medical conditions and environmental factors contribute higher incidence in some populations and a genetic predisposition may play a role for severe cases with respiratory failure. Here we investigated whether genetic variation in the key genes for viral entry to host cells—ACE2 and TMPRSS2—and sensing of viral genomic RNAs (i.e., TLR3/7/8) could explain the variation in incidence across diverse ethnic groups. Overall, these genes are under strong selection pressure and have very few nonsynonymous variants in all populations. Genetic determinant for the binding affinity between SARS-CoV-2 and ACE2 does not show significant difference between populations. Non-genetic factors are likely to contribute differential population characteristics affected by COVID-19. Nonetheless, a systematic mutagenesis study on the receptor binding domain of ACE2 is required to understand the difference in host-viral interaction across populations.

1. Introduction

Coronavirus disease 2019 (COVID-19) caused by SARS-CoV-2 is a pandemic as of Mar. 2020. Initial reports from China revealed diverse risk factors, courses, and outcome for a relatively homogenous population (Zhou et al., 2020a). Morbidity and mortality vary between populations (Yancy, 2020). African Americans and Latinos are disproportionately affected by COVID-19 and show significantly higher mortality compared to the other race and ethnic groups in the US (Wadhera et al., 2020) and in the UK (Kirby, 2020). A “healthcare disparity” must be responsible for the high incidence among minorities although socioeconomic factors, underlying medical conditions, and the difference in genetic susceptibility to SARS-CoV-2 infection may contribute (Chen et al., 2020). Of note, a 3p21.31 gene cluster—SLC6A20, LZTFL1, CCR9, FYCO1, CXCR6 and XCR1—is associated with genetic susceptibility for severe COVID-19 cases with respiratory failure (Ellinghaus et al., 2020). To find allelic variation across populations in the genes that are known to be involved in viral entry to the host cells and sensing of viral RNA in host immune cells, we surveyed publicly available databases of genomic variants.

SARS-CoV-2 is an enveloped and positive single-stranded RNA (ssRNA) virus and initiates human cell entry by binding of spike (S) protein present on the viral envelope to angiotensin converting enzyme 2 (ACE2) receptor on the host cells (Zhou et al., 2020b). The SARS-CoV S protein/ACE2 interface has been elucidated at the atomic level, and the ACE2 was found to be a key factor of SARS-CoV transmission (Li et al., 2005b). The binding mode of SARS-CoV-2 receptor binding domain (RBD) to ACE2 is nearly identical to SARS-CoV (Lan et al., 2020). The S protein is cleaved into S1 and S2 by the type 2 transmembrane serine protease (TMPRSS2) and endosomal cysteine proteases cathepsin B and L (CatB/L) (Du et al., 2009). TMPRSS2 is believed to be of utmost importance for SARS-CoV-2 entry into host cells. Recent studies demonstrated that an inhibitor of the protease activity of TMPRSS2—camostat mesylate—attenuated SARS-CoV-2 entry into lung epithelial cells suggesting a promising candidate for potential intervention against COVID-19 (Hoffmann et al., 2020). The C-terminal domain of S1 subunit is responsible for binding of SARS-CoV-2 to ACE2 and the S2 subunit undergoes a conformational change that result in virus-membrane fusion and entry into the target cell (Du et al., 2009). Viral genomic RNA is then released and translated into viral polymerase...
proteins for viral replication. Innate immune response is the first line of host defense mechanism for SARS-CoV-2 infection. Toll-like receptors recognize the viral RNA – double-stranded RNA (dsRNA) by TLR3 and ssRNA by TLR7 and TLR8 — and trigger innate immune responses such as the expression of inflammatory factors for type I interferons and pro-inflammatory cytokines (Iwasaki and Pillai, 2014; Iwasaki and Yang, 2020).

Here we surveyed the genetic variants in functional residues of ACE2, TMPRSS2, CTSL/L (CatB/L), and TLR3/7/8 to investigate the difference in the genetic predisposition to the susceptibility of SARS-CoV-2 infection and the initiation of innate immune response. For ACE2, we investigated genetic variants in the residues on the interface to SARS-CoV-2 RBD from recent structural analyses (Hussain et al., 2020; Lan et al., 2020; Shang et al., 2020; Wrapp et al., 2020; Yan et al., 2020). Given the high sequence similarity between S proteins of SARS-CoV-2 and SARS-CoV, we also investigated the residues shown to inhibit interactions from in vitro mutagenesis analysis (Li et al., 2005b). We checked two residues reported to cause loss of cleavage activity of TMPRSS2 (Afar et al., 2001) and the enzymatically active sites for CatB/L. A total of 16 residues of TLR7 that are necessary for ssRNA-induced activation (Zhang et al., 2016) and the residues affecting re-action to ssRNAs from in vitro mutagenesis studies for TLR3 (Bell et al., 2006; de Boutelleur et al., 2005; Sarkar et al., 2007) and for TLR8 (Tanji et al., 2015) were checked for sequence variation. Additionally, we searched for nonsynonymous variants that would cause loss of gene function (i.e., frameshift, in-frame insertion/deletion, stop-gain, splice-disrupting, start-lost and stop-lost). The list of reported genetic variants in the genes and their allele frequencies (AFs) were compiled from three population-scale genomic variants databases—gnomAD (Karczewski et al., 2020), Korean Reference Genome Database (Jung et al., 2020), and TogoVar (a Japanese genetic variation database available at https://togovar.biosciencedbc.jp/) —and three whole-genome sequencing datasets (i.e., 1000 Genomes Project (Clarke et al., 2017), Genome Tissue Expression (Consortium et al., 2017), and Simons Genome Diversity Project (Mallick et al., 2016)).

ACE2 is highly conserved with few nonsynonymous variants in the interacting domain with the SARS-CoV-2 RBM (Lan et al., 2020). Of 370 coding variants in ACE2, 248 were nonsynonymous variants with the highest AF of 1.6% (rs41303171). Within 33 residues interfacing the SARS-CoV-2 RBM, 19 variants (including 4 synonymous variants) were found with average AF of 0.03% (ranges 0–0.39%) (Table 1). Only one of the 19 variants (rs4646116; K26R) had global AF greater than 0.1% (AF = 0.39%). Rs4646116 (NC_000009.11:g.90343780A>C) is a common allele (global AF of 0.029%, similar to gnomAD). However, it was found with higher AF of 0.23% at NC_000009.11:g.90343780A>C (East Asian). Further studies are required to test whether rs12329760 could exert functional impact on TMPRSS2 activity. Thus, differences in TMPRSS2 activity caused either by variants at critical loci or by loss-of-function variants are unlikely. SARS-CoV-2 uses both TMPRSS2 and the endosomal cysteine proteases cathepsin B and L (CTSL and CTSL) for priming S protein (Hoffmann et al., 2020). UniProt entries for human CTSL and CTSL report 3 active sites. We found 3 variants in the active sites for CTSL (two missense variants and one synonymous variant), and one missense variant for CTSL (Table 1 and Fig. 1B). Although all missense variants on active sites of CTSL/L are predicted deleterious, they were of very low allele frequencies (AF < 0.01%). CTSL has 429 nonsynonymous variants including 51 loss-of-function variants (all with AF < 0.01%). CTSL has 211 nonsynonymous variants including 17 loss-of-function variants. Of note, one of 17 variants in CTSL (rs2378757, NC_000009.11:g.90343780A>C) is a common allele (global AF of 70.32%, population AF ranges from 62.66% to 98.48%). The variant changes stop codon to serine for one CTSL transcript isoform (ENST00000342020.5) but falls in intron for the other transcript isoforms.

Next we checked genetic variants in TLRs that sense viral RNAs and initiate innate immune responses. There were 7 variants—4 synonymous and 3 nonsynonymous—in the 16 residues of ssRNA interacting domain of TLR7 (Table 1 and Fig. 1C). Most variants were of extremely low frequencies (AF < 0.01%) except for one synonymous variant, rs769401373 (D135D), found only in east Asian population (AF = 0.46%). TLR7 harbors 232 nonsynonymous variants including 8 loss-of-function variants. As in TMPRSS2, AFs of loss-of-function variants were very low (AF < 0.01%). The UniProt entries for TLR3 and TLR8 list 10 sites (6 for TLR3 (Bel et al., 2006; de Boutelleur et al., 2005; Sarkar et al., 2007) and 4 for TLR8 (Tanji et al., 2015)) from in vitro mutagenesis study that impact their response to viral infection (sensing of dsRNA or ssRNA, respectively). For these loci, two missense variants on TLR3 and one missense variant with one synonymous variant on TLR8 were found (Table 1 and Fig. 1C). All of these variants in TLRs were very rare (AF < 0.01%) across all populations.

To summarize, the critical loci for host-viral interaction and sensing viral genomic RNA are highly conserved in all populations with few very rare variants. Especially, ACE2 and TLR7 seem to be under strong selection pressure as reflected in their relatively lower number of loss-of-function variants than expected in large variant databases such as gnomAD (Karczewski et al., 2020): three observed variants out of 31 expected ones for ACE2 and two observed variants out of 20.7 expected ones for TLR7. Moreover, nonsynonymous variants in these genes were mostly of very low frequencies which suggests the chance of gene function altered by these variants would be unlikely, compared to the incidence of COVID-19 around the globe. Other factors such as existing medical conditions and environmental risk factors could contribute the regulation of expression of these key genes in susceptible individuals; however, further studies are required to elucidate potential associations.

The majority of infected individuals experience no or mild symptoms of upper respiratory tract infection; however, for some
| Genes | Residues | AA changes from mutagenesis studies | Residue loci (b:37) | Reported variants within the residues | Variant allele frequencies |
|-------|----------|-------------------------------------|---------------------|--------------------------------------|--------------------------|
|       |          |                                     |                     |                                      |                          |
|       |          |                                     |                     | g:15618980A > G                     |                          |
| ACE2  | S19      | X:15618978–15,618,980 NC_000023.10:| rs73635825 Missense |                                       |
|       |          | - g:15618980A > G                   |                     |                                       |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs761614932 Synonymous                |                          |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs4646116 Missense                    | K > R 0.388% 0.095%     |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs1299103394 Missense                | K > E 0.001%            |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs781255386 Missense                 | T > A 0.001%            |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs758278442 Synonymous                | = 0.002%                |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs36865410 Synonymous                 | = 0.063%                |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs1348114695 Missense                | E > K 0.002%            |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs146676783 Missense                 | E > K 0.004% 0.111%    |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs755691167 Missense                 | K > E 0.001%            |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs766996587 Missense                 | M > I 0.002% 0.026%    |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs759134032 Missense                 | P > T 0.001%            |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs143936283 Missense                 | E > G 0.003%            |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs96130700 Missense                  | D > N 0.001%            |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs762890235 Missense                 | P > H 0.004%            |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs1238146879 Missense                | P > A 0.001%            |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs1335386721 Synonymous              | = 0.001%                |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs1316056737 Missense                | D > Y 0.001% 0.015%    |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs101877725 Missense                 | R > S 0.001%            |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs200549906 Synonymous                | = 0.002%                |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs141939304 Missense                 | L > F 0.002%            |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs185622178 Synonymous                | = 0.001%                |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs773554811 Synonymous                | = 0.001%                |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs769401373 Synonymous                | = 0.033%                |
|       |          |                                      |                     |                                       |
|       |          |                                     |                     | rs868177091 Missense                 | R > Q 0.001%            |

(continued on next page)
| Genes | Residues | AA changes from mutagenesis studies | Residue loci (b37) | Reported variants within the residues | Variants | RS ID | Impact | AA Change | Variant allele frequencies |
|-------|----------|-----------------------------------|-------------------|-------------------------------------|----------|-------|--------|-----------|--------------------------|
|       |          |                                   |                   |                                     |          |       |        |           |                          |
| R473  | R473A    |                                   | X:12905044–12,905,046 | NC_000023.10: g.12905045G > A | rs754381606 | Missense | R > K  | 0.001%    |                          |
|       |          |                                   |                   |                                     |          |       |        |           |                          |
| CTSB  | C108     |                                   | 8:11708378–11,708,380 | NC_000008.10: g.11708783G > A  | rs759843078 | Synonymous | =      | 0.002%    |                          |
|       |          |                                   |                   |                                     |          |       |        |           |                          |
| H278  | H278A    |                                   | 8:11703258–11,703,260 | NC_000008.10: g.11703259T > C  | rs1373655221 | Missense | H > R  | 0.0004%   | Only found in Finnish population (0.005%) |
|       |          |                                   |                   |                                     |          |       |        |           |                          |
| CTSB  | C138     |                                   | 9:90343515–90,343,517 | NC_000008.10: g.90343520T > A | rs1225109229 | Missense | H > Y  | 0.0004%   |                          |
|       |          |                                   |                   |                                     |          |       |        |           |                          |
| TLR3  | H539     | H539E                             | 4:187004455–187,004,465 | NC_000041.11: g.187004466A > G | rs776387492 | Missense | H > R  | 0.001%    |                          |
|       |          |                                   |                   |                                     |          |       |        |           |                          |
| TLR7  | Y759     | Y759F                             | 4:187005115–187,005,115 | NC_000041.11: g.187005116T > C | rs768605211 | Missense | Y > H  | 0.001%    | Only found in Finnish population (0.006%) |
|       |          |                                   |                   |                                     |          |       |        |           |                          |
| TLR8  | Y348     | Y348A                             | X:12938201–12,938,203 | NC_000023.10: g.12938203T > C  | rs768675789 | Synonymous | =      | 0.001%    |                          |

Table 1 (continued)
| Genes          | Variant allele frequencies | gnomAD[1] | 1KGP[2] | SGDP[3] | GTEx[4] | KRGDB[5] | TogoVar[6] |
|---------------|---------------------------|---------|-------|--------|--------|--------|----------|
|               |                           | Latino  | European | East Asian | South Asian |       |          |
|               |                           | 0.001%  | 0.458% | 0.005% | 0.005% | 0.013% |          |
| CTSB (CatB)[13] | Only found in Finnish population (0.005%) | 0.001%  |        | 0.003% |        |        |          |
| CTSL (CatL)[14] |                           | 0.003%  |        | 0.005% | 0.003% | 0.007% | 0.029%   |
| TLR3[15]      |                           | 0.005%  |        |        |        |        |          |
| TLR8[16]      | Only found in Finnish population (0.006%) |        |        |        |        | 0.007% |          |

[1] The genome aggregate database (gnomAD), v2.1.1. [https://gnomad.broadinstitute.org](https://gnomad.broadinstitute.org). Allele frequencies for European are from Non-Finnish European population.
[2] 1000 Genomes Project (1KGP), phase 3. [https://www.internationalgenome.org](https://www.internationalgenome.org)
[3] Simons Genome Diversity Project (SGDP). [https://www.simonsfoundation.org/simons-genome-diversity-project/](https://www.simonsfoundation.org/simons-genome-diversity-project/)
[4] Gene-Tissue Expression project (GTEx), v8 whole genomes. [https://gtexportal.org/home/](https://gtexportal.org/home/)
[5] Korean Reference Genome Database (KRGDB). [http://coda.nih.go.kr/coda/KRGDB/index.jsp](http://coda.nih.go.kr/coda/KRGDB/index.jsp)
[6] NBDCs integrated database of Japanese genomic variation (TogoVar). [https://togovar.biosciencedbc.jp](https://togovar.biosciencedbc.jp)
[7] Shang et al., Nature, 2020
[8] Yan et al., Science, 2020
[9] Lan et al., Nature, 2020
[10] Hussain et al., J Med Vir, 2020
[11] Based on mutagenesis studies from Uniprot protein information for Q9BYF1 (ACE2_HUMAN). [https://www.uniprot.org/uniprot/Q9BYF1](https://www.uniprot.org/uniprot/Q9BYF1)
[12] The ligand-binding sites for small ligands and ssRNA from Zhang et al., Immunity, 2016
[13] Based on active sites from Uniprot protein information for P07858 (CATB_HUMAN). [https://www.uniprot.org/uniprot/P07858](https://www.uniprot.org/uniprot/P07858)
[14] Based on active sites from Uniprot protein information for P07711 (CATL1_HUMAN). [https://www.uniprot.org/uniprot/P07711](https://www.uniprot.org/uniprot/P07711)
[15] Based on mutagenesis studies from Uniprot protein information for O15455 (TLR3_HUMAN). [https://www.uniprot.org/uniprot/O15455](https://www.uniprot.org/uniprot/O15455)
[16] Based on mutagenesis studies from Uniprot protein information for Q9NR97 (TLR8_HUMAN). [https://www.uniprot.org/uniprot/Q9NR97](https://www.uniprot.org/uniprot/Q9NR97)
individuals, the consequence of SARS-CoV-2 infection could be fatal. One of the contributing factors may be the viral load due to differential affinity of viral spike proteins to ACE2 and the efficiency of cleavage by TMPRSS2 that are essential for virus to enter and replicate inside of host cells. We did not find genetic variation between populations while there is a significant difference in incidence and mortality between race and ethnic groups in the U.S. Therefore, underlying medical conditions, age, environmental factors (e.g., air pollution, smoking, and humidity), and a healthcare disparity influence morbidity and mortality from COVID-19 considering the allelic spectrum for the key genes associated with viral entry. Nonetheless, genetic susceptibility may play a role for severe cases with respiratory failure (Ellinghaus et al., 2020).

The population-scale genotype databases and datasets used in this study have limitations from relatively small sample size and imbalanced and incomplete representation of various human populations. Thus, there could be unreported variants in ACE2, TMPRSS2, and TLR3/7/8 that may be associated with change of susceptibility to COVID-19. With additional population-scale genomic databases for diverse populations, it will be possible to identify the individuals with rare genetic variants such as rs758278442 in the interacting domain of ACE2 and the genetic predisposition to cytokine storm that causes an acute progress of illness in young people. In parallel, a systematic mutagenesis analysis of the RBM of ACE2 is highly required to understand the difference in host-viral interaction across populations (Lan et al., 2020).

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2020.104507.

Declaration of Competing Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Acknowledgements

SWK was supported in part by the grants from the National Institution of Health (R01MH107205, R24OD024622 and U01TR002623). JWL was supported by Grant-in-Aids for Scientific Research from the Japan Society for the Promotion of Science (19K10044).
