Published in final edited form as:

Genes Immun. 2020 August ; 21(4): 269–272. doi:10.1038/s41435-020-0107-7.

Genetic variability in the expression of the SARS-CoV-2 host cell entry factors across populations

Lourdes Ortiz-Fernández, Ph.D.¹, Amr H Sawalha, M.D.^{1,2}

¹Department of Pediatrics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

²Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Abstract

The entry of SARS-CoV-2 into host cells is dependent upon angiotensin-converting enzyme 2 (ACE2), which serves as a functional attachment receptor for the viral spike glycoprotein, and the serine protease TMPRSS2 which allows fusion of the viral and host cell membranes. We devised a quantitative measure to estimate genetic determinants of *ACE2* and *TMPRSS2* expression and applied this measure to >2,500 individuals. Our data show significant variability in genetic determinants of *ACE2* and *TMPRSS2* expression among individuals and between populations, and indicate a genetic predisposition for lower expression levels of both key viral entry genes in African populations. These data suggest that host genetics related to viral entry mechanisms might influence inter-individual variability in disease susceptibility and severity of COVID-19.

Keywords

SARS-CoV-2; COVID-19; Genetic; ACE2; TMPRSS2; eQTL

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel single-stranded RNA virus of the *Coronaviridae* family. This recently emerged virus is the cause of a pandemic infection that can result in severe life-threatening disease (Coronavirus Disease-2019; COVID-19)^{1, 2}. Similar to SARS-CoV (which caused SARS), SARS-CoV-2 entry into target host cells is mediated through binding of the viral spike glycoprotein to angiotensin-converting enzyme 2 (ACE2) on the cell surface^{3–6}. The interaction between coronaviral spike proteins and their attachment receptor ACE2 is believed to be key for viral transmissibility and dissemination of infection to organs and tissues⁷. In addition, the host cell serine protease TMPRSS2 cleaves the spike protein of SARS-CoV-2 to allow fusion of

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Please address correspondence to Amr H. Sawalha, MD. Address: 7123 Rangos Research Center, 4401 Penn Avenue, Pittsburgh, PA 15224, USA. Phone: (412) 692-8140. Fax: (412) 412-692-5054. asawalha@pitt.edu.

Author contribution

LO-F Performed the analysis, interpreted that data, and edited the manuscript.

AHS Conceived and designed the study, interpreted the data, and wrote the manuscript.

Conflict of interest: The authors have declared that no conflict of interest exists

the viral and host cell membranes, which is an essential step in viral entry^{3, 8, 9}. Other molecules have been proposed to play alternative roles in the SARS-CoV-2 entry mechanisms. It has been suggested that cathepsin B and L may cleave the spike protein in the absence of TMPRSS2⁹. The results of other studies proposed that SARS-CoV-2 entry is primarily mediated through endocytosis and that PIKfyve, TPC2, and cathepsin L but not cathepsin B are essential for viral entry^{5, 8}. In addition, it has been described that furin may have an effect along with TMPRSS2 and cathepsins on activating viral entry⁸. However, a previous study revealed inconclusive results since furin pre-activation enhanced or reduced SARS-CoV-2 pseudovirus entry in different cell types^{6, 8}. These data suggest that SARS-CoV-2 may use alternative mechanisms to enter the cells but the role of these mechanisms and molecules involved is less clear and needs further investigation.

To gain insight into genetic determinants of SARS-CoV-2 transmissibility and potential viremia and disseminated infection, we devised a quantitative measure to assess the cumulative effect of genetic variants upon the expression of the two key molecules involved in SARS-CoV-2 viral entry, *ACE2* and *TMPRSS2*, and then applied this measure to 2,504 individuals from 5 different populations around the world.

Results and Discussion

We calculated a cumulative genetic expression score (GES) for *ACE2* and *TMPRSS2*, as a measure to estimate host genetic determinants of viral entry of SARS-CoV-2. We evaluated this measure in 2,504 individuals across the 5 major populations included in the 1000 Genomes project (African, Admixed American, European, East Asian, and South Asian)¹⁰. Because *ACE2* is located on the X chromosome, we analyzed female and male individuals separately. There was a significant difference in the cumulative genetic expression score of *ACE2* between populations (ANOVA $P < 0.0001$ for both male and female groups). Genetic determinants of highest expression of *ACE2* were observed in South Asian and East Asian populations, while African populations were genetically associated with the lowest *ACE2* expression levels (Figure 1).

Similarly, significant differences for *TMPRSS2* were observed in both female and male groups across populations (ANOVA $P < 0.0001$). East Asian populations had the highest values for genetic determinants of *TMPRSS2* expression, and Africans showed genetic predisposition for the lowest *TMPRSS2* expression levels across populations (Figure 1).

As mentioned earlier, *ACE2* is located on the X chromosome (and not on the pseudoautosomal region). Therefore, female individuals will have two copies of the gene while males will only have one. Normally, X-chromosome genes are subject to random X-chromosome inactivation, silencing one gene copy in females to keep gene expression balance between females and males. However, a number of X-chromosome genes, including *ACE2*, are known to escape X-chromosome inactivation¹¹. We did not observe differences in genetically determined *ACE2* expression between male and female individuals. Previous reports showed higher expression of *ACE2* in male compared to female tissues, and that this difference in expression was predominantly attributed to non-genetic factors, consistent with

our findings^{11, 12}. No difference between male and female individuals for *TMPRSS2* were observed in our study.

These data suggest that genetic determinants of *ACE2* and *TMPRSS2* expression might play a role in the variability of transmission and severity of SARS-CoV-2 between populations. African populations showed a genetic predisposition for lower expression levels of both *ACE2* and *TMPRSS2*, which are vital for SARS-CoV-2 entry into host cells. These data suggest that a genetic component might contribute to lower numbers of reported COVID-19 in Africa. However, it remains likely that non-genetic factors such as age and associated comorbidities might play a more important role than host genetic elements, especially in determining disease severity and outcome in infected individuals. In addition, genome-wide association studies will be needed to characterize genetic susceptibility to a more severe disease course in patients infected with COVID-19. Additional studies to replicate and extend our findings and examine expression levels of *ACE2* and *TMPRSS2* in different cell types across populations and in patients infected with COVID-19 are warranted.

Methods

We devised a cumulative genetic expression score to estimate genetically determined expression of *ACE2* and *TMPRSS2*. We used expression quantitative trait loci (eQTL) data for *ACE2* and *TMPRSS2* in tissues included in the Genotype-Tissue Expression project (GTEx, release V8)¹³. All eQTL variants affecting *ACE2* and *TMPRSS2* expression across all cell types and tissues were identified, and then pruned to remove variants in linkage disequilibrium (LD). LD pruning was performed using PLINK v.1.9 and the combined haplotypes of the 1000 Genomes Project populations¹⁴. For variants that demonstrate eQTL effects in multiple tissues, the most significant eQTL normalized effect size value was used. The genetic variants used to calculate GRES values for *ACE2* and *TMPRSS2* are shown in Supplementary Dataset 1. A total of 21 genetic polymorphisms that affect *ACE2* expression, and 14 that affect *TMPRSS2* expression were identified and used in subsequent analysis.

The cumulative genetic expression score was derived using the formula:

$GES = \sum_{i=1}^x (n_i \times NES_i)$, where n is the number of alternative alleles (0, 1 or 2), x is the number of evaluated variants in *ACE2* and *TMPRSS2*, and NES is the normalized effect size which reflects the expression in the alternative allele relative to the reference allele for each variant. We calculated the genetic expression scores for *ACE2* and *TMPRSS2* in a total of 2,504 individuals from the five major populations included in the 1000 Genomes Project phase 3 release: African, $n=661$; Admixed American, $n=347$; East Asian, $n=504$; European, $n=503$; and South Asian, $n=489$ ¹⁰. To determine if the cumulative genetic expression score was different across populations, one-way ANOVA following by Tukey's multiple comparison test was performed using GraphPad Prism version 8.1.1 (GraphPad Software, La Jolla California USA). ANOVA p values < 0.05 and Tukey's adjusted p values < 0.05 were considered significant.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

Dr. Sawalha is funded by the Lupus Research Alliance and the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health grant number R01AR070148.

References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; 395(10223): 497–506. [PubMed: 31986264]
2. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020.
3. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 2020.
4. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; 579(7798): 270–273. [PubMed: 32015507]
5. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun* 2020; 11(1): 1620. [PubMed: 32221306]
6. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 2020; 181(2): 281–292 e6. [PubMed: 32155444]
7. Gu J, Korteweg C. Pathology and pathogenesis of severe acute respiratory syndrome. *Am J Pathol* 2007; 170(4): 1136–47. [PubMed: 17392154]
8. Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A et al. Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A* 2020; 117(21): 11727–11734. [PubMed: 32376634]
9. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen 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. [PubMed: 32142651]
10. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM et al. A global reference for human genetic variation. *Nature* 2015; 526(7571): 68–74. [PubMed: 26432245]
11. Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satija R et al. Landscape of X chromosome inactivation across human tissues. *Nature* 2017; 550(7675): 244–248. [PubMed: 29022598]
12. Liu J, Ji H, Zheng W, Wu X, Zhu JJ, Arnold AP et al. Sex differences in renal angiotensin converting enzyme 2 (ACE2) activity are 17beta-oestradiol-dependent and sex chromosome-independent. *Biol Sex Differ* 2010; 1(1): 6. [PubMed: 21208466]
13. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC et al. Genetic effects on gene expression across human tissues. *Nature* 2017; 550(7675): 204–213. [PubMed: 29022597]
14. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; 4: 7. [PubMed: 25722852]

Figure 1: Differences in the cumulative effect of genetic polymorphisms on the expression of SARS-CoV-2 entry mediators *ACE2* and *TMPRSS2* between populations.

upper left Genetic expression score of *ACE2* in female individuals. ANOVA $P < 0.0001$.

Adjusted P values using Tukey's multiple comparisons test: African vs. European, $P = 0.47$; African vs. Admixed American, $P = 0.043$; African vs. East Asian, $P < 0.0001$; African vs. South Asian, $P < 0.0001$; European vs. Admixed American, $P = 0.71$; European vs. East Asian, $P = 0.0175$; European vs. South Asian, $P < 0.0001$; Admixed American vs. East Asian, $P = 0.56$; Admixed American vs. South Asian, $P < 0.0001$; East Asian vs. South Asian, $P = 0.0017$. upper right Genetic expression score of *ACE2* in male individuals. ANOVA $P < 0.0001$. Adjusted P values using Tukey's multiple comparisons test: African vs. European, $P = 0.14$; African vs. Admixed American, $P = 0.17$; African vs. East Asian, $P < 0.0001$; African vs. South Asian, $P < 0.0001$; European vs. Admixed American, $P = 1$; European vs. East Asian, $P < 0.0001$; European vs. South Asian, $P < 0.0001$; Admixed American vs. East Asian, $P = 0.0004$; Admixed American vs. South Asian, $P = 0.0009$; East Asian vs. South Asian, $P = 1$. lower left Genetic expression score of *TMPRSS2* in female individuals. ANOVA $P < 0.0001$. Adjusted P values using Tukey's multiple comparisons test: African vs. European, $P < 0.0001$; African vs. Admixed American, $P < 0.0001$; African vs. East Asian, $P < 0.0001$; African vs. South Asian, $P < 0.0001$; European vs. Admixed American, $P = 0.97$;

European vs. East Asian, $P=0.95$; European vs. South Asian, $P=0.93$; Admixed American vs. East Asian, $P=1$; Admixed American vs. South Asian, $P=0.66$; East Asian vs. South Asian, $P=0.55$. lower right Genetic expression score of *TMPRSS2* in male individuals. ANOVA $P<0.0001$. Adjusted P values using Tukey's multiple comparisons test: African vs. European, $P<0.0001$; African vs. Admixed American, $P<0.0001$; African vs. East Asian, $P<0.0001$; African vs. South Asian, $P<0.0001$; European vs. Admixed American, $P=0.084$; European vs. East Asian, $P=0.03$; European vs. South Asian, $P=1$; Admixed American vs. East Asian, $P=1$; Admixed American vs. South Asian, $P=0.053$; East Asian vs. South Asian, $P=0.015$.