JOURNAL OF
MEDICAL VIROLOGY WILEY

Functional prediction and frequency of coding variants in human *ACE2* at binding sites with SARS-CoV-2 spike protein on different populations

To the Editor,

Hussain et al[1] identified 17 natural coding variants for human angiotensin-converting enzyme 2 (*ACE2*) that were found at positions described as important for binding of ACE2 with SARS-CoV-2 spike protein. The authors suggested that positive prognosis of COVID-19 may be due to the existence of *ACE2* variants like rs73635825 and rs143936283 in some individuals, and that their findings provide clues to screen frequencies of candidate alleles in different populations to predict the prognosis of COVID-19.[1] We would like to contribute with further data for these 17 *ACE2* variants from other function prediction tools, and to debate regarding their use in population genetic studies.

The combination of functional predictors could yield more reliable findings mainly regarding to missense variants,[2] as previously reported.[3] Therefore, we searched for these *ACE2* variants using other recommended predictors. No data were reported by ClinVar and Clinpred, but rs73635825, rs143936283, rs4646116, and rs146676783 were predicted as tolerated by FATHMM, using the inherited disease algorithm to analyze protein missense variants. Compared to SIFT, rs73635825 was also predicted as tolerated, but rs146676783 was predicted as damaging.[1] Both variants were predicted as probably damaging by PolyPhen-2, but as likely benign by CADD and REVEL.[1]

The identified *ACE2* variants have functional importance in binding viral spike protein to ACE2 receptor, but the reported allele frequencies are <1%.[1] Table 1 shows the minor allele frequency (MAF) for different populations where these 17 *ACE2* variants have already been found. Noteworthy, all of them are considered as rare (MAF <1%) genetics variants.[4,5] Rare variants do not necessarily contribute with a large fraction of the genetic variance underlying complex traits.[6] However, one can expect that much of the variance to be due to rare alleles only for diseases that are caused primarily by strongly deleterious mutations.[6] In this context, we understand their goal for selecting coding variants in the binding region with SARS-CoV-2 spike protein to ACE2 receptor.[1] However, this increases the chance of selecting rare and/or low-frequency (MAF <1%-5%) variants,[4,5] which tend to be population or sample specific.[6,7]

Conversely, population genetic studies usually focus on common variants (MAF ≥5%) that are expected to be found in different populations. Moreover, a previous study that analyzed coding variants for *ACE2* showed no direct genetic evidence supporting the existence of coronavirus S-protein binding-resistant *ACE2* mutants in different populations.[8] However, it is important to highlight that this study did not include the variants identified by Hussain et al,[1] and that further investigation are warranted regarding *ACE2* polymorphisms.[8]

Indeed, the majority of associations with low-frequency and rare variants demonstrate relatively small effects on complex traits and disease.[9] Despite limitations of power and resolution, rare variant association studies are becoming increasingly mature, which have been carried out with multiple alleles of different genes in many different populations.[9] For example, the rare-variant analytical approaches allow the identification of genes containing an excess of rare and presumably deleterious variation among cases ascertained for complex disease traits, relative to controls.[10] There is little evolutionary time for SARS-CoV-2 contact with humans, so not enough time for selective pressure. In addition, not all the variants described were described as deleterious,[1] their effects are unknown and they did not passed through the filters of qualifying a variant used by these approaches.[10]

Therefore, although different approaches are being developed to evaluate the use of rare variants in population studies[10], rare variants may not be suitable markers for population genetic studies and demographic differentiation related to COVID-19. For example, the low MAF for these 17 *ACE2* rare variants (Table 1) may hinder analysis such as linkage disequilibrium maps used to assess the nonrandom association of alleles at different loci, which could be used to examine the correlation between *ACE2* variants as a factor resistance or susceptibility against COVID-19 in different populations. Moreover, the absence of reported pathogenicity for these *ACE2* variants according to several predictors suggest they may not provide protection in a level that explains the differences in infection rates and mortality among the affected countries, as described by World Health Organization (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports).

Finally, the differences in incidence and clinical manifestations found among countries, sex and age groups may be due to interactions in genetic pathways in addition to social/economics factors. Nevertheless, COVID-19 is a recent and major health problem with serious economic and social consequences, which requires efforts in the various fields of science, including further population genetic

TABLE 1 dbSNP and GnomAD data for the 17 variants for human *ACE2* located in binding sites with SARS-CoV-2 spike protein

dbSNP ID build 153	Amino acid change	Chromosomal location	Reference/ Alternative	Allele frequency in dbSNP	Population in GnomAD
rs755691167	K68E	chrX:15613110	T/C	C = 0.000011 (ExAC) C = 0.000011 (GnomAD_exome)	South Asian
rs961360700	D355N	chrX:15599350	C/T	T = 0.000012 (GnomAD_exome)	European (non-Finnish)
rs778500138	E35D	chrX:15618929	T/A	A = 0.0003 (TWINSUK)	NA
rs762890235	P389H	chrX:15596342	G/T	T = 0.000024 (TOPMED) T = 0.000034 (ExAC) T = 0.000038 (GnomAD)	Latino European (non-Finnish)
rs143936283	E329G	chrX:15599427	T/C	C = 0.000023 (ExAC) C = 0.000028 (GnomAD_exome) C = 0.00004 (TOPMED) C = 0.000091 (GnomAD) C = 0.000189 (GoESP)	European (non-Finnish) Other
rs4646116	K26R	chrX:15618957	T/C	C = 0.002119 (1000Genomes) C = 0.00315 (GnomAD) C = 0.003677 (ExAC) C = 0.003971 (GnomAD_exome) C = 0.004579 (TOPMED) C = 0.005112 (GoESP) C = 0.006203 (TWINSUK) C = 0.009346 (ALSPAC)	Ashkenazi Jewish European (Finnish) European (non-Finnish) Latino South Asian African East Asian Other
rs1316056737	D427Y	chrX:15596229	C/A	A = 0.000011 (GnomAD_exome) A = 0.000024 (TOPMED)	African
rs759134032	P84T	chrX:15613062	G/T	T = 0.000005 (GnomAD_exome) T = 0.000011 (ExAC)	South Asian
rs766996587	M82I	chrX:15613066	C/A, T	T = 0.000011 (ExAC) T = 0.000011 (GnomAD_exome) T = 0.000048 (TOPMED) T = 0.000136 (GnomAD)	African
rs1396769231	M383T	chrX:15596360	A/G	G = 0.000008 (TOPMED)	NA
rs73635825	S19P	chrX:15618979	A/G	G = 0.000795	African Other
rs1299103394	K26E	chrX:15618958	T/C	G = 0.000252 (GnomAD_exome) G = 0.000345 (ExAC) G = 0.000605 (TOPMED) G = 0.000795 (1000Genomes) G = 0.000824 (GnomAD) G = 0.00142 (GoESP)	European (non-Finnish)
rs781255386	T27A	chrX:15618955	T/C	C = 0.000011 (ExAC) C = 0.000011 (GnomAD_exome)	Latino
rs146676783	E37K	chrX:15618925	C/T	T = 0.000023 (ExAC) T = 0.000033 (GnomAD_exome) T = 0.00004 (TOPMED) T = 0.000091 (GnomAD) T = 0.000284 (GoESP)	European (Finnish) African
rs1238146879	P426A	chrX:15596232	G/C	C = 0.000005 (GnomAD_exome)	European (non-Finnish)
rs1016777825	R559S	chrX:15589906	C/G	G = 0.000006 (GnomAD_exome)	Latino
rs1348114695	E35K	chrX:15618932	C/T	T = 0.000016 (GnomAD_exome)	East Asian European (non-Finnish)

Abbreviation: GnomAD, The Genome Aggregation Database.

studies on the relationship of SARS-COV-2 and *ACE2* to better assess the eventual susceptibility to SARS-COV-2 infection.

CONFLICT OF INTERESTS
The authors declare that there is no conflict of interests.

AUTHOR CONTRIBUTIONS
JOC, IMCAC, SMBS, and MRL made substantial contributions to the analysis or interpretation of data; JOC, SMBS, and MRL drafted the manuscript. All the authors read and approved the final version of manuscript.

Juliana de O. Cruz MS[1]
Izabela M. C. A. Conceição BS[2]
Sandra Mara B. Sousa PhD[3]
Marcelo R. Luizon PhD[1,2] (iD)

[1]*Genetics Graduate Program, Institute of Biological Sciences, Federal University of Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil*
[2]*Department of Genetics, Ecology and Evolution, Institute of Biological Sciences, Federal University of Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil*
[3]*Department of Natural Sciences, State University of Southwest Bahia (UESB), Vitória da Conquista, Bahia, Brazil*

Correspondence
Marcelo Rizzatti Luizon, PhD, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Av. Pres. Antônio Carlos, 6627-Pampulha, Belo Horizonte, Minas Gerais 31270-901, Brazil.
Email: luizonmr@ufmg.br

ORCID
Marcelo R. Luizon (iD) http://orcid.org/0000-0002-8331-3525

REFERENCES
1. Hussain M, Jabeen N, Raza F, et al. Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein [published online ahead of print April 6, 2020]. *J Med Virol*. 2020:0-1. https://pubmed.ncbi.nlm.nih.gov/32249956/, https://onlinelibrary.wiley.com/doi/10.1002/jmv.25832
2. Pejaver V, Mooney SD, Radivojac P. Missense variant pathogenicity predictors generalize well across a range of function-specific prediction challenges. *Hum Mutat*. 2017;38:1092-1108.
3. Linhares ND, Conceicao I, Sandrim VC, Luizon MR. Candidate genes identified by whole-exome sequencing in preeclampsia families: insights into functional annotation and in-silico prediction of deleterious variants. *J Hypertens*. 2020;38(2):372-374.
4. Consortium TUK. The UK10K project identifies rare variants in health and disease. *Nature*. 2015;526:82-90.
5. Gazal S, Loh P-R, Finucane HK, et al. Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nature Genet*. 2018;50:1600-1607.
6. Simons YB, Turchin MC, Pritchard JK, Sella G. The deleterious mutation load is insensitive to recent population history. *Nature Genet*. 2014;46:220-224.
7. Consortium TGP. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061-1073.
8. 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 Discov*. 2020;6:11.
9. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. *Genome Biol*. 2017;18(1):77.
10. Povysil G, Petrovski S, Hostyk J, Aggarwal V, Allen AS, Goldstein DB. Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nat Rev Genet*. 2019;20(12):747-759.