Genetic Variability of Human Angiotensin-Converting Enzyme 2 (hACE2) Among Various Ethnic Populations.

Quan Li[1,2], Zanxia Cao[3], Proton Rahman[1]*

[1]Department of Medicine, Faculty of Medicine, Memorial University, St. John's, Newfoundland, Newfoundland and Labrador, Canada A1B 3X9
[2] Princess Margaret Cancer Centre, University Health Network, University of Toronto, Toronto, Ontario, Canada M5S 1A8
[3] Shandong Provincial Key Laboratory of Biophysics, Institute of Biophysics, Dezhou University, Dezhou 253023, China

*Correspondence should be addressed to
Proton Rahman MD
University Research Professor
Faculty of Medicine, Memorial University of Newfoundland
St. John's Newfoundland, Canada
A1C 5B8
Phone: 709 777-5736 Email: prahman@mun.ca

Key words:
Covid-19; SARS-CoV-2; Angiotensin-converting enzyme 2 (ACE2); gene expression, ethnic variation

Abstract

There appears to be large regional variations for susceptibility, severity and mortality for Covid-19 infections. We set out to examine genetic differences in the human angiotensin-converting enzyme 2 (hACE2) gene, as its receptor serves as a cellular entry for SARS- CoV-2. By comparing 56,885 Non-Finnish European and 9,197 East Asians (including 1,909 Koreans) four missense mutations were noted in the hACE2 gene. Molecular dynamic demonstrated that two of these variants (K26R and I468V) may affect binding characteristics between S protein of the virus and hACE2 receptor. We also examined hACE2 gene expression in eight global populations from the HapMap3 and noted marginal differences in expression for some populations as compared to the Chinese population. However, for both of our studies, the magnitude of the difference was small and the significance is not clear in the absence of further in vitro and functional studies.

Introduction:

As of March 24, 2020, the worldwide tracking for Covid-19 infections reported 1,603,433 cases including 95,716 deaths[1]. Among the 18 countries that have reported more than 20,000 Covid-19 cases, the total number of cases per population ranged from 3,277 per million people in Spain to 57 per million cases in China and the total death rate among these countries ranged from 330 per million in Spain to 2 per million in China[1]. There are numerous potential factors to explain the wide variability in the number of infections and death among the countries including stringency of social isolation and contact tracing, the importation of the SARS-Cov-2, intensity of testing for the virus, as well as the preparedness and capacity of health care system to cope with the pandemic.

Differences among the host, particularly the sex of the patient and presence of selected co-morbid diseases and immunosuppressive state have been identified as risk factors from severe Covid-19[2,3]. Pre-existing medical conditions have been associated with increased prevalence of death including: cardiovascular disease (13.2%); diabetes (9.2%); chronic respiratory disease (8.0%); hypertension (8.4%) and recent diagnosis of cancer (7.6%) [4]. Specifically, hypertensive patients had a hazard ratio of 1.70 for death [5] and 3.05 for in-hospital mortality [4]. An increased association with the use of angiotensin-converting enzyme inhibitors (ACEIs) and angiotensin II receptor blockers (ARBs) has been reported for severe Covid-19 cases[6]. A biologically plausible link between SARS-CoV-2 infections and ACE inhibition has been proposed [7].

Angiotensin-converting enzyme 2 (ACE2) cleaves peptides within the renin-angiotensin system [8]. ACE2 is mostly expressed in kidneys and the GI tract, and smaller amounts are found in type 2 pneumocytes in the lung and peripheral blood. The ACE2

receptor serves as the host cell entry for SARS-CoV and more recently it has been shown to be receptor for cellular entry for SARS- CoV-2 [9,10]. As recently reported by *Hoffman et al [11],* the cellular entry of SARS- CoV-2 can be blocked by an inhibitor of the cellular serine protease TMPRSS2, which is employed by SARS-CoV-2 for S protein priming. Further attention has been drawn to the ACE2 receptor as antimalarials, which can interfere with ACE2 expression, have resulted in increased clearing of the SARS-CoV-2 clearance among Covid-19 patients [12].

To explore the variability in genetic polymorphisms and expression in human ACE2 (hACE2), we set out to determine if there were any differences between the Asian and Caucasian populations for ACE2 polymorphisms and compare the variability of hACE2 expression in peripheral blood among eight different populations.

Comparison of genetic polymorphisms of hACE2 and its possible functional alteration between Asians and Caucasians:

In order to investigate whether differences in genetic variations exist between Caucasians and Asians and if these variants can influence the efficiency of cell entry of SARS-CoV-2, we retrieved the variants in the hACE2 from gnomAD v2.1 exomes[13]. We used 56,885 Non-Finnish European (NFE) and 9,197 East Asians (EAS,including 1,909 Koreans) for our analysis. There were 5,693 mutations of which 177 were missense mutations. In comparing the allele frequency of Caucasian and Asian cohorts, Fisher's exact t tests were performed and adjusted for multiple testing. Four genetic variants reached statistical significance (adjusted p-value less than 0.05), although their effect size was modest (Details in Table1). One variant was located at the a1 helix of ACE2 gene

(K26R), two in the middle of SARS-CoV-2 (N638S and I468V) and one in the tail (N720D).

The K26R variant was of immediate interest, since the spike glycoprotein (S protein) of virus contains the receptor binding domain (RBD), which directly binds to the peptidase domain (PD) of hACE2. The PD of hACE2 mainly engages the a1 helix in the recognition of the RBD[14].

Table.1 Four missense mutations in hACE2 show significant allele frequency difference between EAS and NFE

Chromosome	Position (hg38)	Reference Allele	Alternative Allele	MAF* (EAS)	MAF (NFE)	P-value	Exon	cDNA change	Protein Change
chrX	15564175	T	C	0	0.0258	1.05E-23	18	A2158G	N720D
chrX	15600835	T	C	7.23E-05	0.0059	8.68E-07	2	A77G	K26R
chrX	15567810	T	C	0.0036	0	2.01E-14	16	A1913G	N638S
chrX	15575706	T	C	0.0112	2.50E-05	1.42E-33	11	A1402G	I468V

*MAF stands for minor allele frequency; EAS stands for East Asians; NFE stands for Non-Finnish European

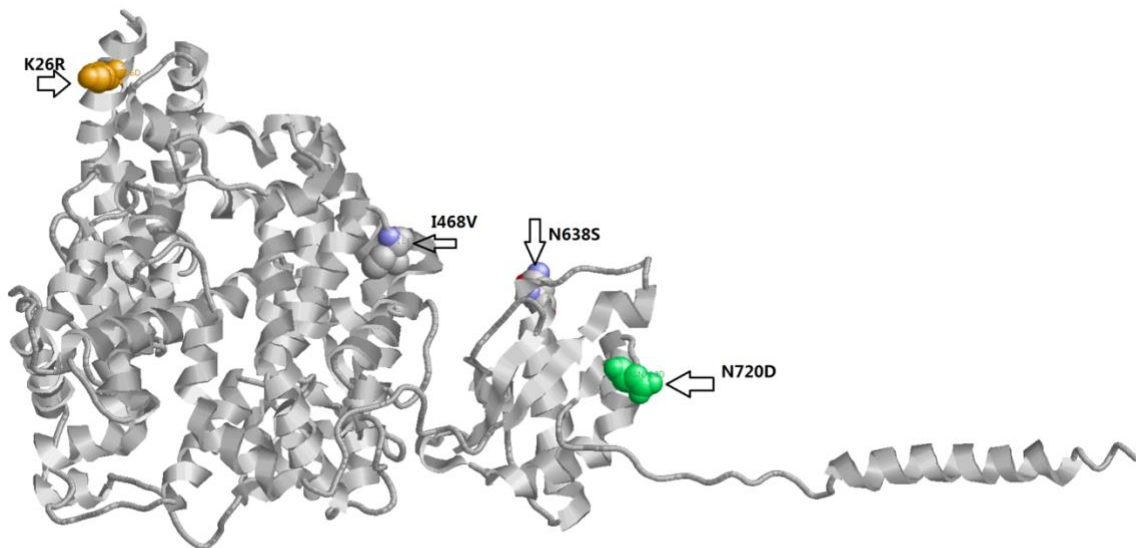We also showed these four mutations in the hACE2 X-ray structure as Figure1.

Figure-1. Human ACE2 structure and four missense mutations that differ between Caucasians and Asians.

Molecular dynamics (MD) simulation which is a canonical method for investigating the influences of mutant on protein structure, was then applied. As mutations N638S and N720 are far away from the RBD, as a result we mainly focused on the K26R as MD simulations demonstrated that this variant mutated more frequently in Non-Finnish European while I468V mutated frequently more in East Asians. The initial wild type structure of the human ACE2 (hACE2) was selected from the X-ray crystal complex structure between hACE2 with SARS-CoV-2 spike (S) protein from Protein Data Bank[15] (PDB ID: 6LZG). The initial structures of mutants were constructed using FoldX (http://foldxsuite.crg.eu/command/PositionScan). The structural minimizations and molecular dynamics simulation was performed using the GROMACS 2018.4[16-18] software package. OPLS-AA force field was adopted for the protein complex and the SPCE water model for solvate the system. The periodic boundary conditions (PBC) in a cubic box was applied, and the distance to the edge of the box was set as 1.0 nm. After the protein complex system was solvated, the appropriate numbers of sodium and chloride ions were added to neutralize the system. Equilibration simulations were done under the NVT ensemble for 2 ns and then the NPT ensemble for 1 ns after an extensive energy minimization by using the steepest descent algorithm. The simulation time step was 2 fs. The temperature was maintained at 300K using the V-rescale algorithm with a tau-t of 0.1 ps. The pressure was maintained at 1 atm by the semi-isotropic Parrinello−Rahman method with a tau-p of 2 ps and a compressibility of 4.5×10^{-5} bar-1. The Particle Mesh Ewald (PME) method was used for calculation the long-range electrostatic interactions.

5

The LINCS algorithm was used for constraining all covalent bonds. Finally, a 100ns long MD simulation was carried out for each system. The stability of the simulation was checked by computing root mean square deviations (RMSDs). RMSDs of Cα were calculated for the mutant K26R/I468V and compared to the wild-type shown in Figure 2. RMSDs for the complexes with K26R changed little during the 100ns MD simulations. Mutant I468V showed RMSD increase during time 20-30 ns, but no obvious RMSD change comparing with wild type after 40ns. The RMSD values for the K26R mutant are slightly higher and structure change compared with wild type.
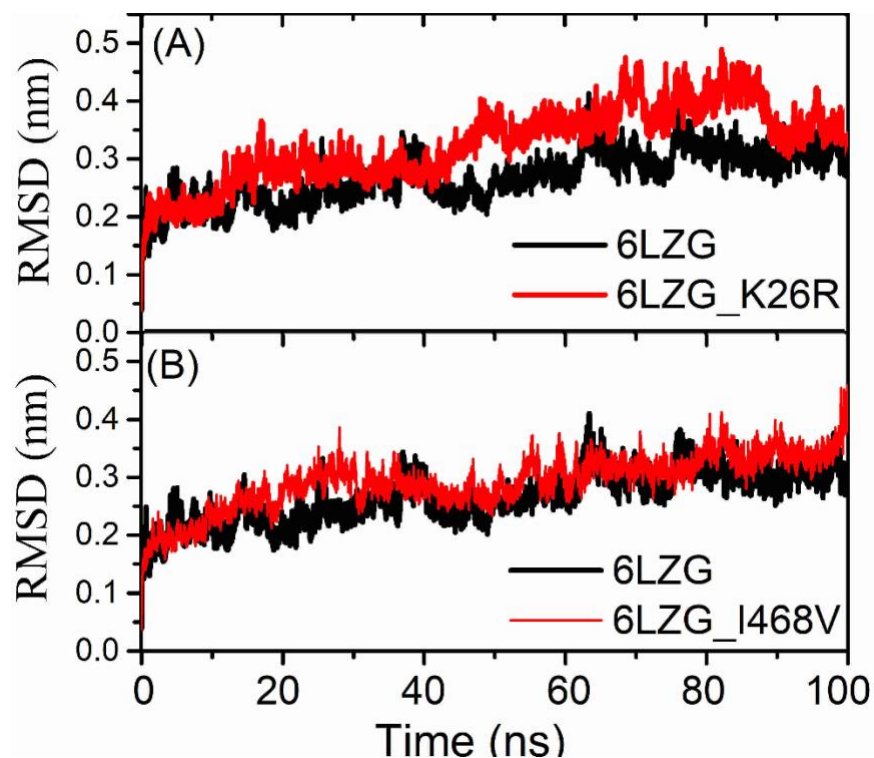

Figure-2 RMSDs of hACE2 wild type (black line) and mutated hACE2 protein(red line) during Molecular dynamics simulation.

The effect of missense on protein-ligand complex can be assessed by various experimental technologies[19], but are very time-consuming. The *in silico* binding free energies calculation can be performed to determine the effects of mutations. Here, the binding free energies between S protein and hACE2 (wild type or mutant) were calculated using the g_mmpbsa (Table 2). These two mutations K26R/I468V were both predicted to slightly increase the binding free energy and may slightly decrease the binding affinity. While, K26R mutated more frequently in Caucasian, and I468V mutated more frequently among Asians.

Table 2. Binding free energy (in units of kJ/mol) for S protein and hACE2 (wild type and mutated) complex

	VDW energy	Electrostattic energy	Polar solvation energy	SASA energy	Binding energy
6LZG_wild type	-343.6±27.4	-1515.1±66.5	716.7±48.4	-45.5±3.1	-1187.5±69.9
6LZG_K26R	-348.2±31.1	-1405.2±63.1	701.4±53.6	-47.5±4.5	-1099.4±54.1
6LZG_I468V	-353.3±23.7	-1498.2±48.9	804.5±71.1	-48.8±4.1	-1095.9±83.1

Variability in gene expression of hACE2 gene among various ethnic populations:
The second goal was to assess the expression variability of hACE2 in eight global populations from the HapMap3 project[20,21]. The following eight populations were compared: Utah residents with Northern and Western European ancestry from the CEPH collection (CEU); Han Chinese in Beijing China (CHB); Gujarati Indians in Houston ,Texas(GIH), Japanese in Tokyo, Japan (JPT); Luhya in Webuye, Kenya (LWK); Mexican ancestry in Los Angeles (MEX), California, Maasai in Kinyawa, Kenya (MKK); and Yoruba in Ibadan, Nigeria (YRI). Raw expression data from Sentrix Human-6 Expression BeadChip version 2 were extracted, background-corrected, log2 scaled and quantile normalized. Expression data from the populations with admixture (GIH, LWK, MEX, and MKK) were also normalized for population admixture using EIGENSTRAT[22].

The normalized expression data were used for the assessing the expression level for different sets of analysis. The hACE2 expression levels of these 8 populations were showed in Figure-3.
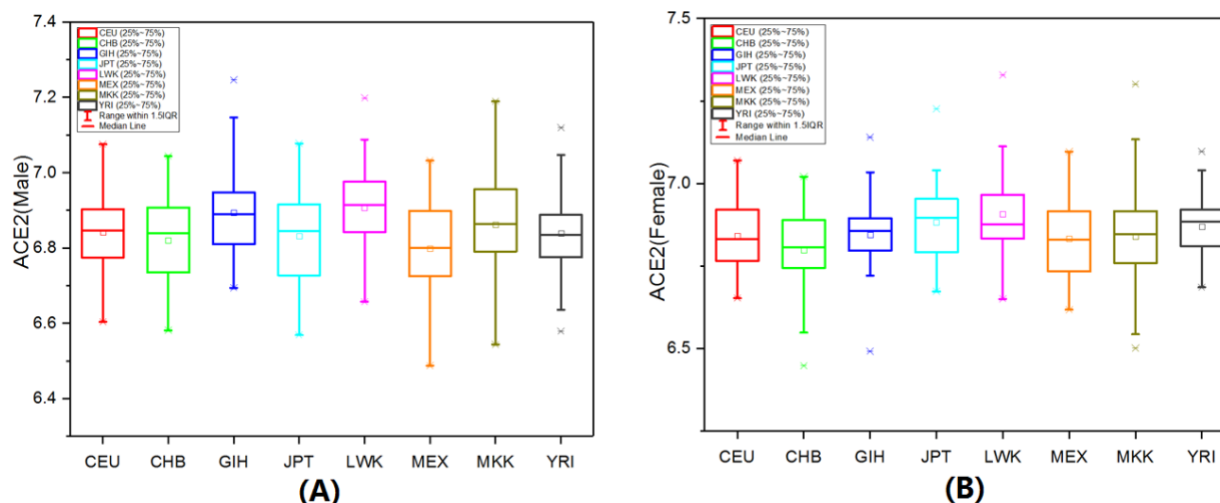

Figure-3. hACE2 expression among 8 Hapmap3 populations. A) for Males and B) for Females.

The hACE2 expression level of CHB population (42 females and 38 males) was then compared to other cohorts using ANOVA analysis, after stratification for sex(Table-3). Marginal differences were noted among the populations. There was a statistically higher hACE2 expression in men for LWK and GIH and JPT populations, LWK and YRI populations for females. However the effect sizes were small and its clinical relevance is not clear.

Table-3. Comparison of ACE2 expression between Asians (CHB) and other 7 populations in Hapmap3

Populations	ACE2(Male)			ACE2(Female)		
	Samples	log2FC	P-value	Samples	log2FC	P-value
CEU	56	0.02	0.43	56	0.042	0.08
GIH	41	0.073	0.007	41	0.046	0.08
JPT	42	0.011	0.69	40	0.083	0.002

LWK	43	0.085	0.0017	40	0.11	6.12E-05
MEX	22	-0.022	0.50	23	0.034	0.27
MKK	70	0.042	0.09	68	0.040	0.09
YRI	55	0.018	0.47	53	0.070	0.005

Discussion:

This study that examined the genetic polymorphisms in the hACE2 gene between Caucasians and Asians. We noted that there were four missense mutations, of which K26R and I468V were of most interest based on their location. The propensity for these variants to mutate were different among these two polymorphisms. Among Caucasians, K26R mutated more, while among Asians I468V mutated more frequently. In silico studies have noted that these two mutations may affecting the binding characteristics between S protein and hACE2, however in the absence of further functional studies, the significance of these alternations is not clear.

We also found that there was some variability of genetic expression of hACE2 among various populations, but the magnitude of the differences was small and so it is unclear if this has any impact based on this subtle differences in susceptibility or severity of Covid-19 among various ethnicities. We also did not notice a differential expression of hACE2 between the sexes, even though it is widely reported that males have a worse prognosis than females for complications of Covid-19 [4]. Our findings are generally consistent with a recently published study by *Chen et al* [23], that reported the ACE2 expression in Asians was similar to that of other races. Our expression study was conducted in peripheral blood and the results for site specific expression (alveolar, GI or renal tissue) may yield different results.

In conclusion, our results do reveal some differences in genetic polymorphisms between Asians and Caucasians which may potentially alter the binding of the virus to the hACE2 receptor and some subtle variation in genetic expression of hACE2 among different populations. However our findings need to replicated and further in vitro studies should performed, before the significance of these findings can be determined. At present the reasons behind the variability in Covid-19 presentation among the various countries needs to be further explored, as we were unable to provide convincing evidence to suggest there are differences in allele frequency or expression of hACE2 among Caucasians and Asians.

Compliance with Ethical Standards:
Conflict of Interest: None of the authors have no conflicts of interest to disclose regarding this manuscript.

Human and Animal Rights and Informed Consent:
Data used for this study was obtained from public databases.

References

1. Worldometers. Covid-19 Coronavirus Pandemic 2020; https://www.worldometers.info/coronavirus/. Accessed 23 March 2020, 2020.
2. China-CCDC;. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19). 2020; http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51. Accessed 23 March 2020, 2020.
3. WorldHealthOrganization;. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). 2020; https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf Accessed 23 March 2020, 2020.
4. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet.* Mar 28 2020;395(10229):1054-1062.
5. Wu C, Chen X, Cai Y, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA internal medicine.* Mar 13 2020.
6. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* Mar 2020;579(7798):270-273.
7. Patel AB, Verma A. COVID-19 and Angiotensin-Converting Enzyme Inhibitors and Angiotensin Receptor Blockers: What Is the Evidence? *Jama.* Mar 24 2020.
8. Batlle D, Wysocki J, Satchell K. Soluble angiotensin-converting enzyme 2: a potential approach for coronavirus infection therapy? *Clinical science.* Mar 13 2020;134(5):543-545.
9. Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S. The spike protein of SARS-CoV--a target for vaccine and therapeutic development. *Nature reviews. Microbiology.* Mar 2009;7(3):226-236.
10. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell.* Mar 6 2020.
11. 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.* Mar 4 2020.
12. Gautret P, Lagier J, Parola P, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *International Journal of Antimicrobial Agents.* 2020;doi.org/10.1016/j.ijantimicag.2020.105949.
13. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* Aug 18 2016;536(7616):285-291.
14. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science.* Mar 27 2020;367(6485):1444-1448.
15. Berman H, Henrick K, Nakamura H, Markley JL. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* Jan 2007;35(Database issue):D301-303.

16.   Kutzner C, Pall S, Fechner M, Esztermann A, de Groot BL, Grubmuller H. More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *Journal of computational chemistry.* Oct 15 2019;40(27):2418-2431.

17.   Kutzner C, Pall S, Fechner M, Esztermann A, de Groot BL, Grubmuller H. Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *Journal of computational chemistry.* Oct 5 2015;36(26):1990-2008.

18.   Pronk S, Pall S, Schulz R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics.* Apr 1 2013;29(7):845-854.

19.   Kastritis PL, Bonvin AM. On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *Journal of the Royal Society, Interface.* Feb 2013;10(79):20120835.

20.   Montgomery SB, Sammeth M, Gutierrez-Arcelus M, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature.* Apr 1 2010;464(7289):773-777.

21.   Stranger BE, Montgomery SB, Dimas AS, et al. Patterns of cis regulatory variation in diverse human populations. *PLoS Genet.* 2012;8(4):e1002639.

22.   Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics.* Aug 2006;38(8):904-909.

23.   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.