



Identification of gene expression location of angiotensin-converting enzyme-2 SNPs as a receptor for SARS-CoV-2 in different populations by using various databases

Dyah Aryani Perwitasari¹, Rita Maliza^{2*}, Bayu Tri Murti³, Haafizah Dania¹, Athika Darumas Putri³

¹Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia 55164, Indonesia

²Study Program of Biology, Faculty of Science and Applied Technology, Universitas Ahmad Dahlan, Yogyakarta 55166, Indonesia

³Semarang College of Pharmaceutical Science (STIFAR), Semarang, Jawa Tengah 50192, Indonesia

*Corresponding author: ritamaliza@bio.uad.ac.id

SUBMITTED 22 January 2021 REVISED 28 April 2021 ACCEPTED 18 May 2021

ABSTRACT The World Health Organization (WHO) has announced that Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) and Coronavirus disease (COVID-19) is considered a worldwide pandemic. Rapidly rising numbers of patients have been reported in almost every country, along with the growing mortality rates. Uncontrolled growth in patient numbers may be due to reasons such as treatment options and vaccine availabilities and unidentified targets of SARS-CoV-2. Previous study has revealed that the molecular target of SARS-CoV-2 is analogous to SARS (2003), i.e. angiotensin-converting enzyme-2 (ACE-2). Therefore, the determination of ACE-2 may enrich existing information and facilitate development of drugs targeted toward SARS-CoV-2. This study aims to screen the expression of ACE-2 genes and their relationship to the types of SNP variants in SARS-CoV-2. We explored a series of observations using powerful databases, e.g. GTEx portal, HaploReg, 1000 Genome and Ensembl, to identify the gene variant of ACE-2. We showed that ACE-2 is highly expressed in the testes and small intestine, while its lowest level is observed in lymphocytes. Subsequently, we observed 17 gene variants containing a missense mutation potentially damaging protein level. Among these genes, single nucleotide polymorphism (SNP) rs370187012 shows the highest damage-level score, while the lowest effect is in SNP rs4646116. The highest frequency of the C allele was observed in European populations (1%). In addition to showing that ACE-2 is expressed in several organs, we concluded that the ACE-2 gene variation can be found in African, American, Southeast and East Asian, and European populations. The polymorphisms of ACE-2 impact on the ACE2 protein structure and the binding capacity of the ACE-2 receptor with the S-Protein of SARS-CoV-2.

KEYWORDS ACE-2; COVID-19; expression genes; SARS-CoV-2; Single Nucleotide Polymorphism

1. Introduction

Continuous increment of patients suffering from SARS-CoV-2 infection across continents, i.e., Asia Pacific, Europe, America, and Africa, has threatened humans globally. Enforced by an uncontrolled number of SARS-CoV-2 patients, WHO as per-11 March 2020 announced that the status of global SARS-CoV-2 had reached an outbreak pandemic. According to an update from Worldometers (Worldometer 2020a) as per-17 January 2021: 94,949,414 confirmed cases; 2,030,914 deaths; and 67,773,034 cured patients COVID-19 were reported among 215 countries, since the first case allegedly reported in Wuhan, Hubei, China in December 2019. Each country has conducted various strategies to uplift the prevention of the disease, mainly by restraining the air flights towards pandemic sides and vice versa.

This coronavirus is firstly known as “2019 novel

coronavirus”. Furthermore, it was officially changed to “COVID-19” and included the type of SARS-CoV-2 virus. The SARS-CoV-2 is regarded as the cause, while COVID-19 is referring to the disease caused by SARS-CoV-2. Providing the disease name is the authority of WHO through standardized International Classification of Disease (ICD). The International Committee on Taxonomy of Viruses (ICTV) of WHO legitimately announced the rise of SARS-CoV-2 on 11 February 2020, considering the virus is genetically analogous to the type of SARS coronavirus spread in 2003.

Given the current surge of cases that occur mainly in the Asian population, scientists put the suspect on them and further observe the factors. One may provide the insight answer to the question is the genetics/genomic science. Genetics science studied the overall information about the genetic who owned a cell or organism, while the

genome is a plural form of genetics. Thus far, scientists are trying their best to cope with the SARS-CoV-2, i.e. identify the origin of the virus through sequencing/mapping of SARS-CoV-2 and compare with some animals considered or suspected as the virus origin as bat found in the traditional market in Wuhan.

Scientists have been identifying various possibilities of genes damaged by the virus. Instead, one was also attempting to identify the receptor of SARS-CoV-2 itself. Chao and co-workers analyzed the comparison of human gene characteristics in several populations and correlated it with their vulnerability against COVID-19 (Cao et al. 2020). They elucidated the angiotensin-converting-2 enzyme (ACE-2) as one of the receptors of SARS-CoV-2. ACE-2 receptors are generally found in the respiratory tract. Some previous studies reported reinforcing the hypothesis that ACE-2 is strongly evident as the receptor or supreme target of SARS-Cov-2 (Zhou et al. 2020; Lu et al. 2020).

Additionally, an in-vitro study also represented the positive correlations between the ACE-2 expression with SARS-CoV in 2003 (Hofmann et al. 2004; Li et al. 2007). The SARS-Cov-2 enables binding ACE-2 with a strong affinity towards S-protein (Li et al. 2005). There is an urgent necessity to screen the expression of ACE-2 genes and their relation to SNP variants types in SARS-CoV-2. Previous studies have not explored the types of variations in the ACE-2 gene, which are susceptible to SARS-CoV-2 and especially to the outbreak pandemic. Regarding the fundamental role of ACE2 in SARS-CoV-2 cell entry, and as a potential therapeutic target for antiviral therapy, in this study, we used several databases to investigate related expression profiles and SNPs of ACE2. Using the various databases, we found that the results were consistent with some previous studies (Hikmet et al. 2020; Paniri et al. 2021) that ACE-2 expressed in several organs, including the respiratory tract. The polymorphisms of ACE-2 impact the ACE2 protein structure, function and this also affected to ACE2 -dependent cell entry of SARS-CoV-2. This information will be beneficial for further research in regards to SARS-CoV-2.

2. Materials and Methods

In order to find the expression of ACE -2 genes, the GTEx portal database (GTExPortal 2020), a site that contains all genes expression, was used to identify the types of SNPs in the ACE -2 gene. The GTEx portal database was retrieved on 20 June 2020. Afterward, to confirm the type of mutated genes that affect the level of protein or so-called the gene coding protein (variation type of genes of missense mutase), we then use Haploreg version 4.1 (HaploReg 2020) and 1000 Genome project (AsiaEnsemble 2020). Upon identifying the types of non-synonym SNPs or missense mutase, we employ the Ensembl database to compare the SNP among the worldwide population. Ensembl encompasses a database that more specifically predicts protein changes, i.e. PolyPhen-2. This database di-

vided the changes of protein into three categories based on score and level of SNP damage in protein: score 0.00 - 0:15 (benign), 0.15 - 0.85 (possibly damaging), and 0.85 - 1.00 (probably damaging), with the range between 0 to 1. This study investigated several affected populations by SARS-Cov-2, including Europe, Southeast and South Asia, America, and Africa.

3. Results and Discussion

3.1. Identification of ACE-2 gene expression over various organs

The well-known database used to contain all gene expressions of human organs is expression quantitative trait loci (eQTLs), which is loaded on the GTEx database (Ardlie et al. 2015). By comparing them with accessible genome databases such as the 1000 Genome database, the outcome allows us to identify the types of genes in a certain population, i.e., Asia, when most SARS-CoV-2 outbreaks occurred. Using a website database (GTExPortal 2020), results showed that ACE-2 genes positively associate with ACE-2 expression in several organs. (Figure 1.). Figure 1 shows the expression of ACE-2 in several organs, which were scoring with log₁₀ (Transcripts Per Million (TPM) +1). ACE-2 was highly expressed in the testis (median TPM in male was 46.53), small intestine or the terminal ileum (median TPM in female 50,06; male 24.45), adipose visceral (median TPM in female 9.409; male 8.543), kidney (median TPM in female 10.70; male 7.676), heart - left ventricle (median TPM in female 8.991; male 6.896), thyroid (median TPM in female 6.172; male 6.414), heart - atrial appendage (median TPM in female 5.706; male 5.422), colon - transverse (median TPM in female 4.679; male 2.775), while the lowest level observed in lymphocytes (median TPM in female 0.02405; male 0.01637). We also identified that ACE-2 was expressed in the human lung (median TPM in females 0.94; male 1.044), which this organ knew as the host for SARS-CoV-2. It is also confirmed that the ACE-2 gene tended to be expressed in the female gender (cervix - ectocervix, cervix - endocervix, uterus, fallopian tube, ovary, and vagina) compared to the male (testis, kidney - medulla and prostate). However, sequencing/ RNA mapping showed that the male Asian population showed a higher ACE-2 gene expression than the female (Zhao et al. 2020). This finding is in accordance with the data reported by Worldometers (Worldometer 2020b), indicates that the death rate of male patients is merely higher (61.8%) than that of female (38.2%).

3.2. Identification of ACE-2 gene variation affecting the protein change

Single nucleotide polymorphisms (SNPs) are able to impact protein function, structure, stability, and abundance (Calcagnile et al. 2020). For identifying the gene variation affecting the protein change, we used Haploreg ver. 4.1 and 1000 genome project databases that contain 2.6

million gene variations of SNPs (Devuyst 2015). We observe 128 SNPs that have a variety of missense mutations. Through further Ensembl database, we identified 17 SNPs that may impact the protein alteration types with the following types: probably damaging, possibly damaging, and hyperplasia (Table 1). We further identified that 17 gene variations of ACE-2 have an alleles variety of <1% over the population of African, American, and Asian. This data indicates that the ACE-2 is rarely observed in those populations. However, for SNP rs4646116, the allele frequency of European showed over 1%. This study also demonstrated that the ACE-2 gene with SNP rs370187012 is the most potent contributing to protein levels alteration through a predicted probably damaging. Ensembl database observation also indicates the rs370187012 allele frequency of <0.1% among the Southeast Asian population in which is one of the most affected regions of SARS-CoV-2 and COVID-19 outbreak (Table 1).

3.3. Discussion

Increase evidence of COVID-19 has been a massive threat around the world. Scientists are still searching for a break-

through to overcome the SARS-CoV-2 attack by identifying the causes and biological receptors targeted by this virus. For instance, as the SARS receptor has been identified (ACE-2) in 2003 (Hofmann et al. 2004), the initial step to encounter the new virus can be probably initiated by using the similar target receptor to correlate with the binding to the new virus (Zhang et al. 2020). In this study, we used a gene expression database to identify the ACE-2 most-expressing organs. Several previous studies have been conducted and mentioned that ACE-2 is utmost-expressed in the respiratory system, thus may become one of the possible pathways to cure COVID-19 (Zhang et al. 2020). However, from our data, we found that the expression of ACE-2 in the respiratory system was not very high (Figure 1). The higher expression of ACE2 may be due to comorbidity, including hypertension, coronary artery disease, chronic obstructive pulmonary disease, diabetes mellitus, chronic kidney disease, obesity and smoking patients (Guan et al. 2020). Furthermore, changes in the expression levels of ACE2 correlate with SNPs, splicing, and transcription processing which could raise the vulnerability of individuals to COVID-19 infection. We further as-

TABLE 1 Distribution of allele frequencies of ACE-2 along with the impact for proteins.

ID SNP	Chr	Gene	Protein changing	Deleterious (Score)	Consequence	Allele		Allele frequency				
						REF Allele	ALT Allele	Africa	America	EAS	SAS	EUR
rs370187012	chrX	ACE2	R710H	0.999	Probably Damaging	C	T	0	0	0.000163	0	0
rs776995986	chrX	ACE2	R708W	0.997	Probably Damaging	G	A	0	0	0	0	0
rs758142853	chrX	ACE2	V184A	0.996	Probably Damaging	T	C	0	0	0	0	0
rs140016715	chrX	ACE2	R768W	0.995	Probably Damaging	G	A	0.000081	0	0.000085	0	0
rs147311723	chrX	ACE2	L731F	0.941	Probably Damaging	G	A	0.014754	0.000496	0	0	0
rs754511501	chrX	ACE2	G173S	0.876	Possibly Damaging	A	G	0	0	0	0	0
rs373153165	chrX	ACE2	D785N	0.838	Possibly Damaging	T	C	0	0	0	0	0
rs191860450	chrX	ACE2	I468V	0.747	Possibly Damaging	G	A	0	0	0	0	0
rs73635825	chrX	ACE2	S19P	0.74	Possibly Damaging	A	G	0.003423	0	0	0	0
rs149039346	chrX	ACE2	S692P	0.723	Possibly Damaging	A	G	0.004946	0.000037	0	0.000053	0
rs748163894	chrX	ACE2	A627V	0.713	Possibly Damaging	T	C	0	0	0	0	0
rs140473595	chrX	ACE2	A501T	0.323	Benign	G	A	0	0	0	0	0
rs183135788	chrX	ACE2	N638S	0.054	Benign	T	C	0	0	0	0	0
rs41303171	chrX	ACE2	N720D	0.02	Benign	T	C	0.003053	0.002695	0	0.00434	0
rs373252182	chrX	ACE2	N149S	0.015	Benign	T	C	0	0	0	0	0
rs200180615	chrX	ACE2	E668K	0.005	Benign	C	T	0	0	0	0	0
rs4646116	chrX	ACE2	K26R	0.001	Benign	T	C	0.000988	0.003286	0.000072	0.001313	0.01

EAS: East Asia; SAS: Southeast Asia; EUR: Europe

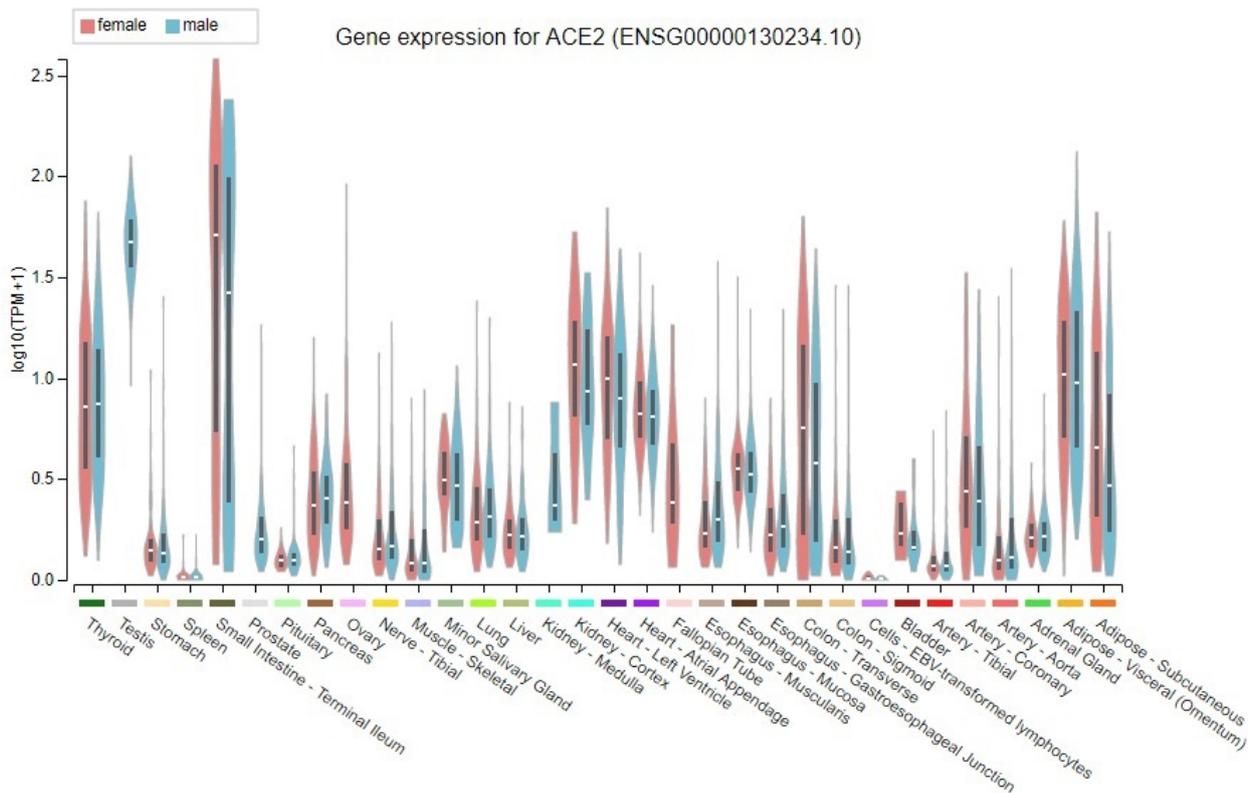


FIGURE 1 Gene expression of ACE-2 in several organs (The GTExPortal 2020)

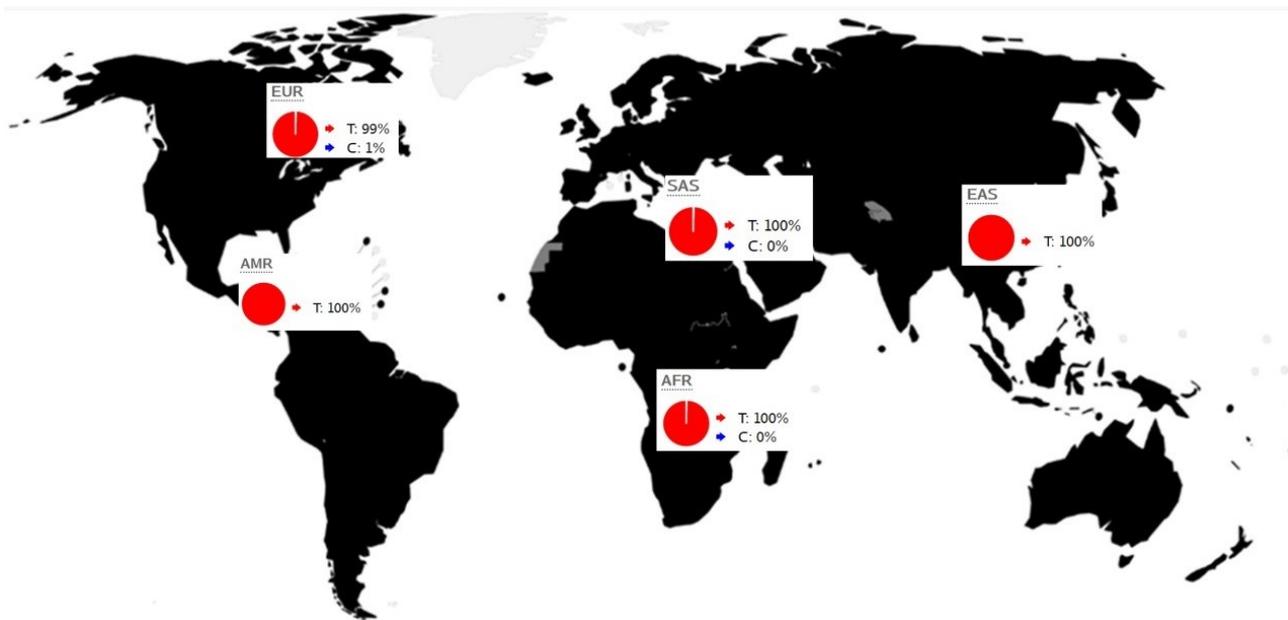


FIGURE 2 Distribution of allele frequencies of ACE-2 type-SNP rs4646116 type in some populations. Abbreviations: AMR; America, AFR; Africa, EUR: Europe, EAS; East Asia, SAS; Southeast Asia.

essed the correlation of population with COVID-19 with the gene variant. A study conducted by Chao and co-workers showed that SNP with rs4646127 located in intron ACE-2 is the highest frequency of an allele in China population 0.997 and Southeast Asia (0.994) (Cao et al. 2020). Compared to Europe and the US data, the SNP of

rs4646127 showed much lower expression of an allele in China, 0.651 and 0.754, respectively. Nevertheless, this SNP type suggests an intone, which is not a coding type of protein that can trigger a change in protein level.

In this study, we employed several gene variety databases, 1000 genome database, and GTEx database that

may enhance the SNP signal present and potentially detected the phenotype changes possibility or, in this case, correlated highly to COVID-19. The presence of SNPs within the coding region of ACE-2 can result in the alteration of amino acid sequence. This change in amino acid sequence at the interaction site can affect the binding capacity of the ACE-2 receptor with the S-Protein of SARS-CoV-2. Furthermore, SNPs within promoter or 3'-UTR can cause downregulation of the ACE-2 gene resulting in lower levels of ACE-2 receptor at the cell surface for interaction with virus particles (Chaudhary 2020). We found 17 types of ACE-2 variation, which are the potential to give impact down to the protein level. This fact may help raise clues to assess the COVID-19 severity in correlation with the ACE-2 gene and its coexistence with diseases, such as pneumonia. The SNP rs370187012 with 0.999 scores showed potential damage for R710H protein type, while SNP rs4646116 showed the lowest score of 0.001 along with benign potency, or rarely inducing any protein mutation. However, further responsible steps are required to ensure the vulnerability of the ACE-2 gene of SARS-CoV2 infected patients toward the COVID-19. Furthermore, a noteworthy fact is underlined herein, that SNP rs4646116 appears among Europe population with more than 1% occurrence probability of the ACE-2; meanwhile, the other populations, viz. The US and Asia show lower probability frequency (less than 1%). The rs4646116 (K26R), rs149039346 (S692P), and rs41303171 (N720D) located in the helix, helix and coil structure, respectively. Phyre2 software indicated that S692P protein changing probably causes disorder with relatively high confidence. Change in ACE2 structure affected COVID-19 cell entry, and therefore its replication in host cells, especially lung (Liu et al. 2020; Sommerstein et al. 2020). The polymorphisms of ACE-2 impact the ACE2 protein structure, affecting the binding capacity of the ACE-2 receptor with the S-Protein of SARS-CoV-2.

Overall, our data is fairly restricted to early identification of gene variation of ACE-2 using the databases study and prioritizing the gene variety of mutated type, missense. Further validation of such experiments associated with SNP types among- and inter-populations of COVID-19 patients are highly required to verify the allele difference among the different populations of COVID-19. Thus, it may provide more data in regards to ACE-2 gene vulnerability for various SNPs of the COVID-19 patients.

4. Conclusions

A vigorous database search has successfully provided an insight into ACE-2 gene variation and evidence of ACE-2 expressed in the testis, small intestine or the terminal ileum, adipose visceral, kidney heart - left ventricle, thyroid, heart - atrial appendage, colon - transverse and the respiratory tract, which has been well-known as the SARS-CoV-2 target. Furthermore, we identified 17 variants of SNPs missense mutation that play a potential role in protein changes, along with SNP rs370187012 and SNP

rs4646116 regarded as the highest and lowest protein alterations, respectively.

Acknowledgments

The authors were thankful to Lalu Muhammad Irham and Imaniar Noor Faridah of Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia, for their valuable technical suggestions and help.

Authors' contributions

Conceptualization: DAP, RM Methodology: RM, ADP Software: RM, BTM Validation: DAP, RM Formal analysis: BTM, ADP Investigation: DAP, RM Resources: DAP, RM Data Curation: RM, ADP Writing - Original Draft: DAP, RM, BTM, ADP, HD Writing - Review & Editing: DAP, RM, BTM, ADP, HD Visualization: DAP, HD Supervision: DAP Project administration: HD Funding acquisition: DAP. All authors read and approved the final version of the manuscript.

Competing interests

The author declare that they have no competing interest.

References

- Ardlie KG, DeLuca DS, Segrè AV, Sullivan TJ, Young TR, Gelfand ET, Trowbridge CA, Maller JB, Tukiainen T, Lek M, et al. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*. 348(6235):648–660. doi:10.1126/science.1262110.
- AsiaEnsemble. 2020. Asia Ensemble. URL <http://asia.ensembl.org/index.html>.
- Calcagnile M, Forgez P, Iannelli A, Bucci C, Alifano M, Alifano P. 2020. ACE2 polymorphisms and individual susceptibility to SARS-CoV-2 infection: insights from an in silico study. *BioRxiv*.
- Cao Y, Li L, Feng Z, Wan S, Huang P, Sun X, Wen F, Huang X, Ning G, Wang W. 2020. Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. *Cell Discovery*. 6(1):1–4. doi:10.1038/s41421-020-0147-1.
- Chaudhary M. 2020. COVID-19 susceptibility: potential of ACE2 polymorphisms. *Egypt J Med Hum Genet*. 21(1). doi:10.1186/s43042-020-00099-9.
- Devuyst O. 2015. The 1000 genomes project: Welcome to a new world. *Peritoneal Dial Int*. 35(7):676–677. doi:10.3747/pdi.2015.00261.
- GTExPortal. 2020. Gene Page ACE2. URL <https://gtexportal.org/home/gene/ACE2>.
- Guan Wj, Liang Wh, Zhao Y, Liang Hr, Chen Zs, Li Ym, Liu Xq, Chen Rc, Tang Cl, Wang T, et al. 2020. Comorbidity and its impact on

- 1590 patients with COVID-19 in China: a nationwide analysis. *Eur Respir J.* 55(5):2000547. doi:10.1183/13993003.00547-2020.Supp1.
- HaploReg. 2020. HaploReg v4.1. URL <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>.
- Hikmet F, Méar L, Edvinsson Å, Micke P, Uhlén M, Lindskog C. 2020. The protein expression profile of ACE2 in human tissues. *Mol Syst Biol.* 16(7). doi:10.15252/msb.20209610.
- Hofmann H, Geier M, Marzi A, Krumbiegel M, Peipp M, Fey GH, Gramberg T, Pöhlmann S. 2004. 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 Commun.* 319(4):1216–1221. doi:10.1016/j.bbrc.2004.05.114.
- Li W, Sui J, Huang IC, Kuhn JH, Radoshitzky SR, Marasco WA, Choe H, Farzan M. 2007. The S proteins of human coronavirus NL63 and severe acute respiratory syndrome coronavirus bind overlapping regions of ACE2. *Virology.* 367(2):367–374. doi:10.1016/j.virol.2007.04.035.
- Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, Wong SK, Huang IC, Xu K, Vasilieva N, Murakami A, He Y, Marasco WA, Guan Y, Choe H, Farzan M. 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* 24(8):1634–1643. doi:10.1038/sj.emboj.7600640.
- Liu M, Wang T, Zhou Y, Zhao Y, Zhang Y, Li J. 2020. Potential role of ACE2 in coronavirus disease 2019 (COVID-19) prevention and management. *J Transl Int Med.* 8(1):9–19. doi:10.2478/2Fjtim-2020-0003.
- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, et al. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 395(10224):565–574. doi:10.1016/S0140-6736(20)30251-8.
- Paniri A, Hosseini MM, Moballegh-Eslam M, Akhavan-Niaki H. 2021. Comprehensive in silico identification of impacts of ACE2 SNPs on COVID-19 susceptibility in different populations. *Gene Rep.* 22. doi:10.1016/j.genrep.2020.100979.
- Sommerstein R, Kochen MM, Messerli FH, Gräni C. 2020. Coronavirus disease 2019 (Covid-19): Do angiotensin-converting enzyme inhibitors/ angiotensin receptor blockers have a biphasic effect? *J Am Heart Assoc.* 9(7). doi:10.1161/JAHA.120.016509.
- Worldometer. 2020a. Coronavirus Age, Sex, Demographics (COVID-19) - Worldometer. URL <https://www.worldometers.info/coronavirus/coronavirus-age-sex-demographics/>.
- Worldometer. 2020b. COVID-19 CORONAVIRUS PAN-DEMIC. URL <https://www.worldometers.info/coronavirus/coronavirus-age-sex-demographics/>.
- Zhang H, Penninger JM, Li Y, Zhong N, Slutsky AS. 2020. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Medicine* 46(4):586–590. doi:10.1007/s00134-020-05985-9.
- Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo W. 2020. Single-Cell RNA Expression Profiling of ACE2, the Receptor of SARS-CoV-2. *Am J Respir Crit Care Med.* 202(5):756–759. doi:10.1164/rccm.202001-0179LE.
- Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, et al. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579(7798):270–273. doi:10.1038/s41586-020-2012-7.