

## Asians and other races express similar levels of and share the same genetic polymorphisms of the SARS-CoV-2 cell-entry receptor

Ying Chen<sup>1,2</sup>, Kejia Shan<sup>1,2,3</sup>, and Wenfeng Qian<sup>1,2,3\*</sup>

<sup>1</sup> State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

<sup>2</sup> Key Laboratory of Genetic Network Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

<sup>3</sup> University of Chinese Academy of Sciences, Beijing 100049, China

\* Correspondence to:

Wenfeng Qian

Institute of Genetics and Developmental Biology

Chinese Academy of Sciences

Beijing 100101, China

Email: [wfqian@genetics.ac.cn](mailto:wfqian@genetics.ac.cn)

## ABSTRACT

The recurrent coronavirus outbreaks in China (SARS-CoV and its relative, SARS-CoV-2) raise the possibility that Asians are more susceptible to coronavirus. Here, we test this possibility with the lung expression of *ACE2*, which encodes the cell-entry receptor of both SARS-CoV and SARS-CoV-2. We show that *ACE2* expression is not affected during tumorigenesis, suggesting that the transcriptome data from the more than 1000 lung cancer samples in The Cancer Genome Atlas (TCGA) can be used to study *ACE2* expression among people without cancer. The expression of *ACE2* increases with age, but is not associated with sex. Asians show a similar *ACE2* expression to other races. Furthermore, the frequencies of *ACE2* alleles in Asians are not significantly deviated from those in other races. These observations indicate that individuals of all races need the same level of personal protection against SARS-CoV-2.

## KEYWORDS

SARS-CoV-2; cell-entry receptor; *ACE2* expression; The Cancer Genome Atlas; susceptibility; demographic factors; race

## INTRODUCTION

The epidemic caused by SARS-CoV-2 has led to significant illnesses and deaths and has been designated a global health emergency by the World Health Organization. It is clear that SARS-CoV-2 is a close relative of SARS-CoV (Xu et al., 2020), which has caused severe acute respiratory syndrome in 2003. However, the demographic factors that predict the susceptibility to these coronaviruses among individuals remain poorly understood. In particular, since both SARS-CoV and SARS-CoV-2 epidemics broke out from China, a possibility has been raised that East Asians are more susceptible to these coronaviruses (Zhao et al., 2020).

SARS-CoV enters cells through ACE2 (Kuba et al., 2005; Li et al., 2003), whose native function is to play a role in the renin-angiotensin system (Donoghue et al., 2000). Susceptibility to SARS-CoV was associated with *ACE2* expression among cells (Hofmann et al., 2004; Jia et al., 2005). Overexpressing *ACE2* in cell lines promoted efficient replication of SARS-CoV while neutralizing ACE2 by antibodies inhibited viral replication in a dose-dependent manner (Li et al., 2003). Furthermore, *ACE2* expression was detected in the alveolar epithelial cells (Hamming et al., 2004), which were the primary target of SARS-CoV in the lung (Kuiken et al., 2003); knocking-out *ACE2* in mice suppressed SARS-CoV infection in the lung (Kuba et al., 2005). ACE2 is also used by SARS-CoV-2 to enter cells. The spike protein of SARS-CoV-2 had a strong binding affinity with ACE2 based on structural modeling (Xu et al., 2020). Furthermore, HeLa cells were infected by SARS-CoV-2 only when *ACE2* was expressed (Zhou et al., 2020). Collectively, *ACE2* expression in lung cells is likely by far one of the most reliable indicators of the susceptibility to SARS-CoV and SARS-CoV-2 among individuals.

The genotype-tissue expression (GTEx) project failed to identify any quantitative trait loci for *ACE2* expression or any Asian-biased gene expressions in the lung (Mele et al., 2015), probably due to the small number of samples from Asians (1.3% of all samples). In contrast, there are thousands of lung samples from individuals of various

ages, sexes, and races in TCGA and other cancer genomics studies. In this study, we investigate the demographic determinants of *ACE2* expression from the cancer transcriptome data. We show that *ACE2* expression is not affected during tumorigenesis and that Asians do not have higher *ACE2* expression than other races. Furthermore, Asians do not have any unique *ACE2* genetic polymorphisms. The results of this study can help develop global protection strategies against SARS-CoV-2.

## RESULTS AND DISCUSSION

TCGA includes lung transcriptomes of two cancer types: lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). We focused on the 594 LUAD samples in this study since LUAD is more likely to stem from alveolar cells (Lin et al., 2012; Xu et al., 2012) and more highly expresses *ACE2* ( $P = 0.001$ , the two-tailed Mann-Whitney  $U$  test). The expression level of *ACE2* was similar between primary tumors and solid normal tissues (**Figure S1A–B**) and was not associated with survival probability or pathologic stage (**Figure S1C–D**), indicating that *ACE2* expression is not affected during tumorigenesis. In fact, the expression level of *ACE2* in primary tumor samples genuinely reflected that in the adjacent normal tissues across 19 cancer types in TCGA (Pearson's correlation coefficient  $R = 0.8$ ,  $P = 3 \times 10^{-5}$ , **Figure S1E** and **Table S1**). Therefore, the transcriptome data from LUAD samples can be used to study *ACE2* expression among cancer-free individuals. However, there were only seven Asian LUAD samples in TCGA; to solve this problem, we expand the sample size of Asians by including 260 Chinese LUAD transcriptomes (Chen et al., 2020a).

The *ACE2* expression was positively correlated with age among middle-aged and older adults in LUAD samples (Pearson's correlation coefficient  $R = 0.13$ ,  $P = 2 \times 10^{-4}$ ,  $N = 835$ , **Figure 1A**), increasing by  $\sim 1.2$  times with every 10-year increase in age. This may partly explain the observation that the elderly are more susceptible to SARS-CoV-2 (Chen et al., 2020b; Huang et al., 2020; Li et al., 2020). Sex was not associated with *ACE2* expression ( $P = 0.24$ , the two-tailed Mann-Whitney  $U$  test,

**Figure 1B**); the observation that more males were infected by SARS-CoV-2 in the epidemic (Chen et al., 2020b; Huang et al., 2020; Li et al., 2020; Yang et al., 2020) could be caused by their more exposure to viruses due to occupation or lifestyle.

The average expression level of *ACE2* among Asians was not significantly different from that among individuals of African or European ancestry (hereinafter referred to as African and European,  $P = 0.42$  and  $0.96$ , respectively, the two-tailed Mann-Whitney  $U$  tests, **Figure 1C**). Such absence of difference held in a linear model that predicted *ACE2* expression from multiple demographic factors including age, sex, and race (**Table S2**); the variance in the  $\log_2$ -transformed *ACE2* expression level among all LUAD samples was 3.53 (i.e., standard deviation = 1.88), and race explained only  $\sim 0.1\%$  of it ( $P = 0.64$ , the analysis of variance). This observation refutes the previous finding using a single Asian sample (Zhao et al., 2020).

There are some caveats in this study, though. First, we have been focusing on *ACE2* expression in the lung, but *ACE2* is expressed in other tissues/organs as well (**Figure S2**), especially in the gastrointestinal tract and kidney. These organs were also reported to be infected by SARS-CoV (Gu et al., 2005), and SARS-CoV-2 was detected from patients' stool (Holshue et al., 2020). With that in mind, we further tested if Asians expressed *ACE2* at a higher level in other cancer types; *ACE2* was similarly expressed between Asians and others in all these cancer types with *ACE2* expression higher than LUAD (**Figure 2**). Nevertheless, we could not rule out the possibility that *ACE2* expresses at a higher level in Asians in some other tissues/organs. Second, we have been focusing on gene expression level, but Asian-specific genetic variation in the coding sequence of *ACE2* may also affect the cell-entry efficiency of viruses. To test this possibility we retrieved all genetic variation data of *ACE2* from the 1000 Genomes Project (The 1000 Genomes Project Consortium, 2015). All the 22 missense or stop-gained variations were present at a low frequency in East Asians ( $< 1\%$ , **Figure 3**). Other variants ( $N = 690$ , **Table S3**) are mostly located in introns, and the frequencies of them in East Asians did not show a significant deviation from those in all populations; among the 37 polymorphisms

with the alternate allele frequency  $> 10\%$  in East Asians, none was two times as frequent as in all populations (**Figure 3**). Yet it is still plausible that genetic variations in other genes may affect the modification, folding, or subcellular localization of ACE2 *in trans*. Third, we have been focusing on the mRNA level; future investigations are required to determine if difference among races exists at the ACE2 protein level. While ACE2 protein levels are largely correlated with mRNA levels among tissues (<https://www.proteinatlas.org/ENSG00000130234-ACE2/tissue>) (Uhlen et al., 2015), whether these ACE2 proteins are located on the cell membrane remains unclear. It would also be of importance to investigate if co-receptors of SARS-Cov or SARS-CoV-2 exist and if their abundance varies among races.

Based on the available data, we conclude that Asians do not express *ACE2* at a higher level and do not bear unique genetic polymorphisms in *ACE2*. The recurrent coronavirus outbreaks in China may be better explained by the high diversity of coronaviruses and their animal hosts, as well as the Chinese food culture (Fan et al., 2019). We, therefore, caution any use of race to predict the susceptibility to SARS-CoV-2 among individuals; individuals of all races require the same level of personal protection against SARS-CoV-2.

## METHODS

All available level-3 RNA-seq data and clinical data of LUAD and LUSC samples in TCGA were retrieved from <https://www.cancer.gov/tcga>. The RNA-seq data were in the unit of FPKM-UQ (fragments per kilobase of transcript per million mapped reads of the upper quartile gene). Transcriptomes of additional Asian samples (Chen et al., 2020a) were retrieved from OncoSG (<https://src.gisapps.org/OncoSG/>) under dataset Lung Adenocarcinoma (GIS, 2019); FPKM-UQ was calculated from the numbers of reads mapped to individual genes. FPKM-UQ values of expressed genes were globally higher in the data from OncoSG (**Figure S3A**), and therefore, median normalization was further performed to ensure that expression levels from two sources are comparable (**Figure S3B**). The annotation of clinical information for each sample and the statistical analyses were performed with in-house *R* scripts. The

Kaplan-Meier curves were generated with UCSC Xena (<http://xena.ucsc.edu>) (Goldman et al., 2019). The gene expression levels of *ACE2* across cancer types were also downloaded from UCSC Xena.

The genetic variation data of *ACE2* were retrieved from the 1000 Genomes Project (The 1000 Genomes Project Consortium, 2015) at the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/>). The allele frequencies among East Asians were calculated from samples from Chinese Dai in Xishuangbanna, Han Chinese in Beijing, Southern Han Chinese, Japanese in Tokyo, and Kinh in Ho Chi Minh City. The annotation of the locations or types of variants was retrieved from the Ensembl Genome Browser (<https://www.ensembl.org/index.html>).

**Code availability.** All codes used to analyze the data and to generate the figures are available at <https://github.com/YingChen10/ACE2-Expression>.

## DECLARATIONS

### Acknowledgments

We thank Drs. Di Liu (Wuhan Institute of Virology, CAS), Weiwei Zhai (Institute of Zoology, CAS), and Yuting Zhao (UT Southwestern Medical Center) for their critical comments on the manuscript. The results shown in this study are in part based upon data generated by the TCGA Research Network (<https://www.cancer.gov/tcga>), the GTEx Project (<https://gtexportal.org/home/>), and the 1000 Genomes Project (<https://www.internationalgenome.org/>).

### Funding

This work was done by the authors during holidays and was not supported by grants.

### Authors' contributions

Y.C., K.S., and W.Q. designed the study, performed data analyses, and wrote the manuscript.

### Competing interests

The authors declare that they have no competing interests.

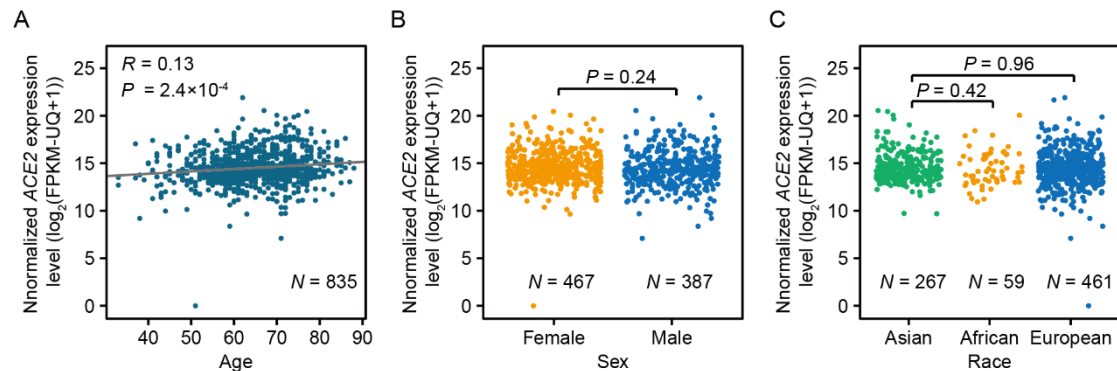
## REFERENCES

- Chen, J., Yang, H., Teo, A.S.M., Amer, L.B., Sherbaf, F.G., Tan, C.Q., Alvarez, J.J.S., Lu, B., Lim, J.Q., Takano, A., *et al.* (2020a). Genomic landscape of lung adenocarcinoma in East Asians. *Nat Genet*.
- Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., *et al.* (2020b). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*.
- Donoghue, M., Hsieh, F., Baronas, E., Godbout, K., Gosselin, M., Stagliano, N., Donovan, M., Woolf, B., Robison, K., Jeyaseelan, R., *et al.* (2000). A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circ Res* 87, E1-9.
- Fan, Y., Zhao, K., Shi, Z.L., and Zhou, P. (2019). Bat Coronaviruses in China. *Viruses* 11.
- Goldman, M., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., *et al.* (2019). The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *bioRxiv*, 326470.
- Gu, J., Gong, E., Zhang, B., Zheng, J., Gao, Z., Zhong, Y., Zou, W., Zhan, J., Wang, S., Xie, Z., *et al.* (2005). Multiple organ infection and the pathogenesis of SARS. *J Exp Med* 202, 415-424.
- Hamming, I., Timens, W., Bulthuis, M.L., Lely, A.T., Navis, G., and van Goor, H. (2004). Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol* 203, 631-637.
- Hofmann, H., Geier, M., Marzi, A., Krumbiegel, M., Peipp, M., Fey, G.H., Gramberg, T., and Pohlmann, 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, 1216-1221.
- Holshue, M.L., DeBolt, C., Lindquist, S., Lofy, K.H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A., *et al.* (2020). First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med*.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., *et al.* (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*.
- Jia, H.P., Look, D.C., Shi, L., Hickey, M., Pewe, L., Netland, J., Farzan, M., Wohlford-Lenane, C., Perlman, S., and McCray, P.B., Jr. (2005). ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. *J Virol* 79, 14614-14621.
- Kuba, K., Imai, Y., Rao, S., Gao, H., Guo, F., Guan, B., Huan, Y., Yang, P., Zhang, Y., Deng, W., *et al.* (2005). A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat Med* 11, 875-879.
- Kuiken, T., Fouchier, R.A., Schutten, M., Rimmelzwaan, G.F., van Amerongen, G., van Riel, D., Laman, J.D., de Jong, T., van Doornum, G., Lim, W., *et al.* (2003). Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet* 362, 263-270.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau,



- E.H.Y., Wong, J.Y., *et al.* (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*.
- Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C., *et al.* (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426, 450-454.
- Lin, C., Song, H., Huang, C., Yao, E., Gacayan, R., Xu, S.M., and Chuang, P.T. (2012). Alveolar type II cells possess the capability of initiating lung tumor development. *PLoS One* 7, e53817.
- Mele, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., *et al.* (2015). Human genomics. The human transcriptome across tissues and individuals. *Science* 348, 660-665.
- The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* 526, 68-74.
- Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., *et al.* (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419.
- Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., Zhong, W., and Hao, P. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci*, 1674-7305.
- Xu, X., Rock, J.R., Lu, Y., Futtner, C., Schwab, B., Guinney, J., Hogan, B.L., and Onaitis, M.W. (2012). Evidence for type II cells as cells of origin of K-Ras-induced distal lung adenocarcinoma. *Proc Natl Acad Sci U S A* 109, 4910-4915.
- Yang, Y., Lu, Q., Liu, M., Wang, Y., Zhang, A., Jalali, N., Dean, N., Longini, I., Halloran, M.E., Xu, B., *et al.* (2020). Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. *medRxiv*, 2020.2002.2010.20021675.
- Zhao, Y., Zhao, Z., Wang, Y., Zhou, Y., Ma, Y., and Zuo, W. (2020). Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCov. *bioRxiv*, 2020.2001.2026.919985.
- Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., *et al.* (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*.

## FIGURES

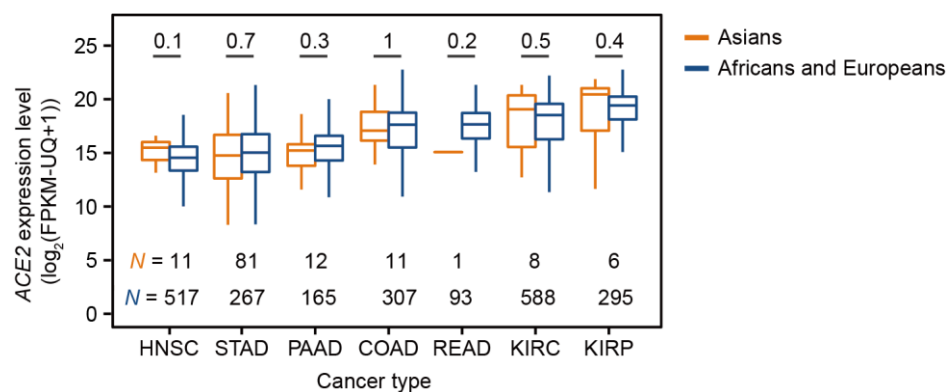


**Figure 1. The expression of *ACE2* is associated with age, but not sex or race, among LUAD samples.**

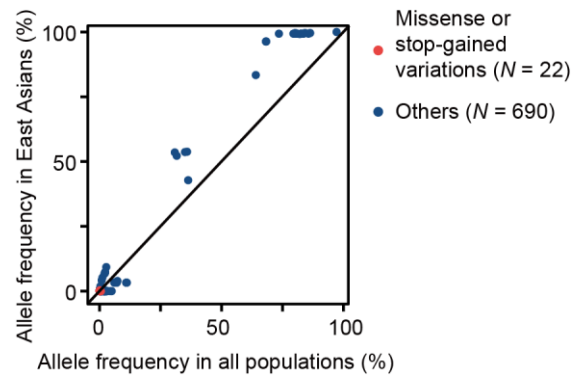
(A) Relationship between age and *ACE2* expression level. Pearson's correlation coefficient ( $R$ ) and the corresponding  $P$ -value are shown.

(B) Comparison of *ACE2* expression between sexes.  $P$ -value is calculated with a two-tailed Mann-Whitney  $U$  test.

(C) Comparison of *ACE2* expression among races.  $P$ -values are calculated with two-tailed Mann-Whitney  $U$  tests. The single sample of American Indian or Alaska Native was not shown.



**Figure 2. Comparison of *ACE2* expression between Asians and other races among cancer types.** Seven cancer types with *ACE2* expression higher than LUAD are shown.  $P$ -values are calculated with the two-tailed Mann-Whitney  $U$  tests. The full names of cancer types can be found in Table S1. Outliers are not shown for the boxplots.



**Figure 3. Allele frequencies of the genetic variants of *ACE2* collected by the 1000 Genomes Project, in East Asians (*y*-axis) or in all populations (*x*-axis).** Each dot represents a polymorphic site, and the frequency of the alternate allele is shown. Only dimorphic genetic variations are shown. Detailed information is provided in Table S3.

## SUPPLEMENTARY INFORMATION

Figure S1. The expression of *ACE2* in primary lung tumors can reflect that in normal lung tissues.

Figure S2. The *ACE2* expression among tissues and between sexes.

Figure S3. Distribution of the expression level among genes within a sample, before (A) and after (B) median normalization.

Table S1. The median expression level of *ACE2* ( $\log_2(\text{FPKM-UQ} + 1)$ ) among cancer types.

Table S2. A linear model that predicts *ACE2* expression among LUAD samples from the race, age, and sex.

Table S3. Variations of *ACE2* on human genome assembly GRCh37.