Tobacco-use disparity in gene expression of ACE2, the receptor of 2019-nCov

Guoshuai Cai

Department of Environmental Health Sciences, Arnold School of Public Health, University of South Carolina, Columbia, SC 29208

Address for Correspondence:

Guoshuai Cai
Department of Environmental Health Sciences
Arnold School of Public Health
University of South Carolina
915 Greene Street
Discovery 517
Columbia, SC 29204
GCAI@mailbox.sc.edu
Phone: 803-777-4120

Abstract

In current severe global emergency situation of 2019-nCov outbreak, it is imperative to identify vulnerable and susceptible groups for effective protection and care. Recently, studies found that 2019-nCov and SARS-nCov share the same receptor, ACE2. In this study, we analyzed four large-scale datasets of normal lung tissue to investigate the disparities related to race, age, gender and smoking status in ACE2 gene expression. No significant disparities in ACE2 gene expression were found between racial groups (Asian vs Caucasian), age groups (>60 vs <60) or gender groups (male vs female). However, we observed significantly higher ACE2 gene expression in smoker samples compared to non-smoker samples. This indicates the smokers may be more susceptible to 2019-nCov and thus smoking history should be considered in identifying susceptible population and standardizing treatment regimen.

Key words

Wuhan 2019-nCov, ACE2, expression, susceptibility, race, age, gender, smoking

Introduction

In the past two decades, pathogenic coronaviruses (CoVs) have cause epidemic infections, including the server acute respiratory syndrome (SARS)-CoV outbreak in 2003, the Middle East respiratory syndrome (MERS) outbreak in 2012 and the current novel Wuhan 2019-nCov outbreak. We have learned from SARS-Cov and MERS-Cov that human populations showed disparities in susceptibility to these viruses. For example, epidemiology studies found that males had higher incidence and mortality rates than females.[1,2] We believe that the susceptibility to the novel 2019-nCov is also different among population groups. In current severe global emergency situation of 2019-nCov outbreak, it is imperative to identify vulnerable and susceptible groups for effective protection and care.

Recently, Xu et.al. computationally modelled protein interactions and identified a putative cell entry receptor of 2019-nCov, angiotensin-converting enzyme 2 (ACE2), which is also a receptor for SARS-nCov.[3] Zhou et.al. further confirmed this virus receptor in the HELA cell line.[4] Interestingly, Zhao et al. found ACE2 is specifically expressed in a subset of type II alveolar cells (AT2), in which genes regulating viral reproduction and transmission are highly expressed.[5] They also found that an Asian male has much higher ACE2-expressing cell ratio than other seven white and African American donors, which may indicate the higher susceptibility of Asian. However, the sample size was too small to draw conclusion on this racial disparity. Here, we analyzed four large-scale datasets of normal lung tissue to investigate the disparities related to race, age, gender and smoking status in ACE2 gene expression.

Methods

Two RNA-seq datasets and two DNA microarray datasets from lung cancer patients were analyzed in this study, including a Caucasian RNA-seq dataset from TCGA (https://www.cancer.gov/tcga), an Asian RNA-seq dataset from Gene Expression Omnibus (GEO) with the accession number GSE40419[6], an Asian microarray dataset from GEO with the accession number GSE19804[7] and a Caucasian microarray dataset from GEO with the accession number GSE10072[8]. Both RNA-seq datasets were generated with the Illumina

HiSeq platform and both microarray datasets were generated with the Affymetrix GeneChip Human Genome U133 Array. The details and processing of data were described in our previous study[9]. All datasets contain samples from tumor and normal pairs and we only use the normal samples in this study. In total, 54 samples in the TCGA dataset, 77 samples in the GSE40419 dataset, 60 samples in the GSE19804 dataset and 33 samples in the GSE10072 dataset were analyzed. We studied the Reads per kilobase per million mapped reads (RPKM) values for RNA-seq data and Robust Multi-Array Average (RMA)[10] values for microarray data. All data were log2 transformed to improve normality. The data means across samples in each dataset from the same platform were highly correlated (Pearson correlation coefficient r=0.9 for microarray datasets and r=0.97 for RNA-seq datasets, Fig. S1), indicating no significant system variation in datasets from the same platform.

Simple linear regressions were used to test the association of ACE2 gene expression with each single variable of age, gender, race and smoking status. Also, multiple linear regression was used to test the association of ACE2 expression with multiple factors (age, gender, race, smoking status and data platform). All data management, statistical analyses and visualizations were accomplished using R 3.6.1.

Results

Racial Disparity

Inconsistent with the study of Zhao et al.[5], we observed no significant difference in ACE2 expression in Caucasian lung tissue samples compared to Asian lung tissue samples in the RNAseq datasets (*p*-value=0.45, Fig 1A). In the microarray datasets, a higher ACE2 expression was observed in Caucasian samples compared to Asian samples (*p*-value=0.03, Fig 1A). Given that the GSE19804 RNA-seq study focused on female non-smokers while the TCGA dataset includes samples from both males and females and both smokers and non-smokers, we believe that the observed disparity may be due to other factors other than race, such as smoking, gender and unknown factors. Therefore, we performed multiple linear

regression on multiple independent variables (age, gender, race, smoking status and platform) and found no significant difference between racial groups (*p*-value=0.36, Fig. 1E).

Tobacco-related disparity

We found a significant higher ACE2 gene expression in smoker samples compared to non-smoker samples in the TCGA (*p*-value=0.05) and GSE40419 datasets (*p*-value=0.008, Fig. 1B). Smokers in GSE10072 showed a higher mean of ACE2 gene expression than non-smokers. The difference is not significant (*p*-value=0.18), which may be due to its small sample size (n=33) with insufficient power to detect the difference. The GSE19804 data with only smoker samples available was not included into the analysis. Adjusted by other factors (age, gender, race and platforms) in multivariate analysis, smoking still shows a significant disparity in ACE2 gene expression (*p*-value=0.008, Fig. 1E).

Age and Gender

We didn't observe a disparity between age groups (>60 vs <60) or gender groups (male vs female) in ACE2 gene expression in each available study (Fig. 1C, D). Consistently, multivariate analysis didn't reject the null hypothesis that there is no difference between groups of age or gender after other variables (age/gender, race, smoking status and platforms) were adjusted (*p*-value=0.90 for age, *p*-value=0.35 for gender, Fig. 1E). We also consistently found no difference between male and female healthy lung tissue samples from GTEx[12] (Fig. S2).

Discussion

In this study, we investigated the disparities related to race, age, gender and smoking status in ACE2 gene expression and found significantly higher ACE2 gene expression in lung tissue of smokers compared to that of non-smokers. This may explain the reason why more males (56% of 425 cases) were found in a recent epidemiology report of 2019-nCov early transmission by China CDC[11]. We didn't observe significant disparities in ACE2 gene expression between racial groups (Asian vs Caucasian), age groups (>60 vs <60) or gender groups (male vs female).

This study has several limitations. First, the data analysed in this study were from the normal lung tissue of patients with lung adenocarcinoma, which may be different with the lung tissue of healthy people. Although we observed no difference between male and female healthy samples from GTEx, further validation studies are required for other factors. Second, our analysis was based on the average expression from bulk tissue. This may lead to a power loss in detecting the expression from particular cell types such as the AT2 cells in which ACE2 are specifically highly expressed.

Whether ACE2 is the only or major receptor of 2019-nCov is unknown. The reason(s) for the tobacco-related disparity in ACE2 expression is unknown. Despites current limited knowledge, this study indicates the smokers may be more susceptible to 2019-nCov and thus smoking history should be considered in identifying susceptible population and standardizing treatment regimen. Wuhan, stay strong.

Ethical oversight

There is no direct involvement of human subjects in this study. All the data use existing de-identified biological samples and data from prior studies. Therefore, ethical oversight and patient consent were not handled in this project.

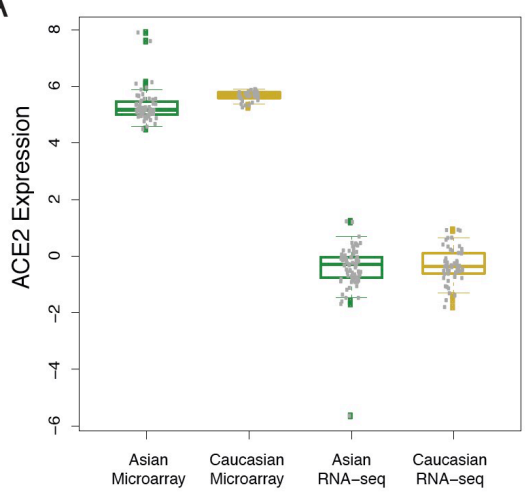
References

1. Karlberg J, Chong DS, Lai WY. Do men have a higher case fatality rate of severe acute respiratory syndrome than women do? Am J Epidemiol 2004;159:229-31.
2. Alghamdi IG, Hussain, II, Almalki SS, Alghamdi MS, Alghamdi MM, El-Sheemy MA. The pattern of Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive epidemiological analysis of data from the Saudi Ministry of Health. Int J Gen Med 2014;7:417-23.
3. Xu X, Chen P, Wang J, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. SCIENCE CHINA Life Sciences 2020.
4. Zhou P, Yang X-L, Wang X-G, et al. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. bioRxiv 2020:2020.01.22.914952.
5. Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo W. Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCov. bioRxiv 2020:2020.01.26.919985.

6. Seo JS, Ju YS, Lee WC, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. Genome research 2012;22:2109-19.
7. Lu TP, Tsai MH, Lee JM, et al. Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2010;19:2590-7.
8. Landi MT, Dracheva T, Rotunno M, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. PloS one 2008;3:e1651.
9. Cai G, Xiao F, Cheng C, Li Y, Amos CI, Whitfield ML. Population effect model identifies gene expression predictors of survival outcomes in lung adenocarcinoma for both Caucasian and Asian patients. PLoS One 2017;12:e0175850.
10. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4:249-64.
11. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020.
12. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580-5.

Figure

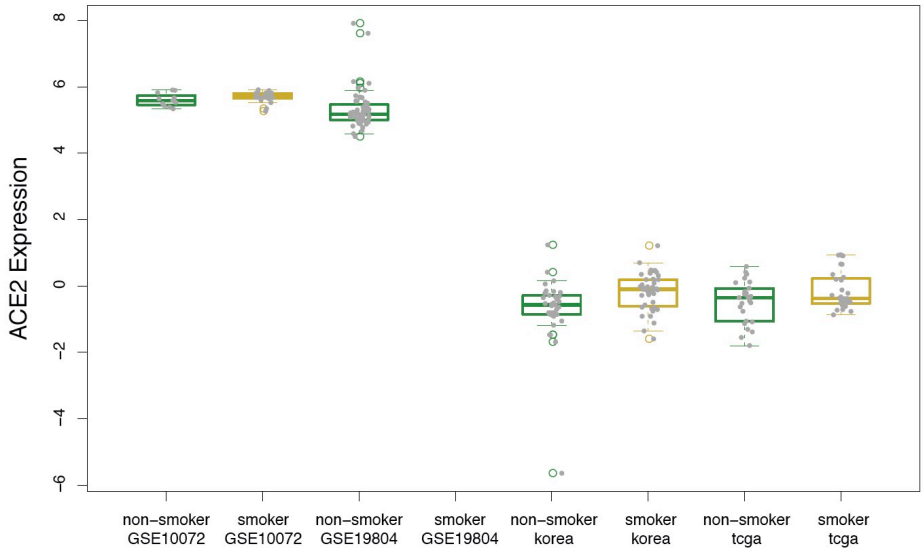
A



E

Variable	Beta	P-value
Platform: RNA-seq	-5.940	<2e-16
Race: Caucasian	0.088	0.359
Age: >60	-0.012	0.900
Gender: Male	0.105	0.352
Smoking: Smoker	0.313	**0.008**

B



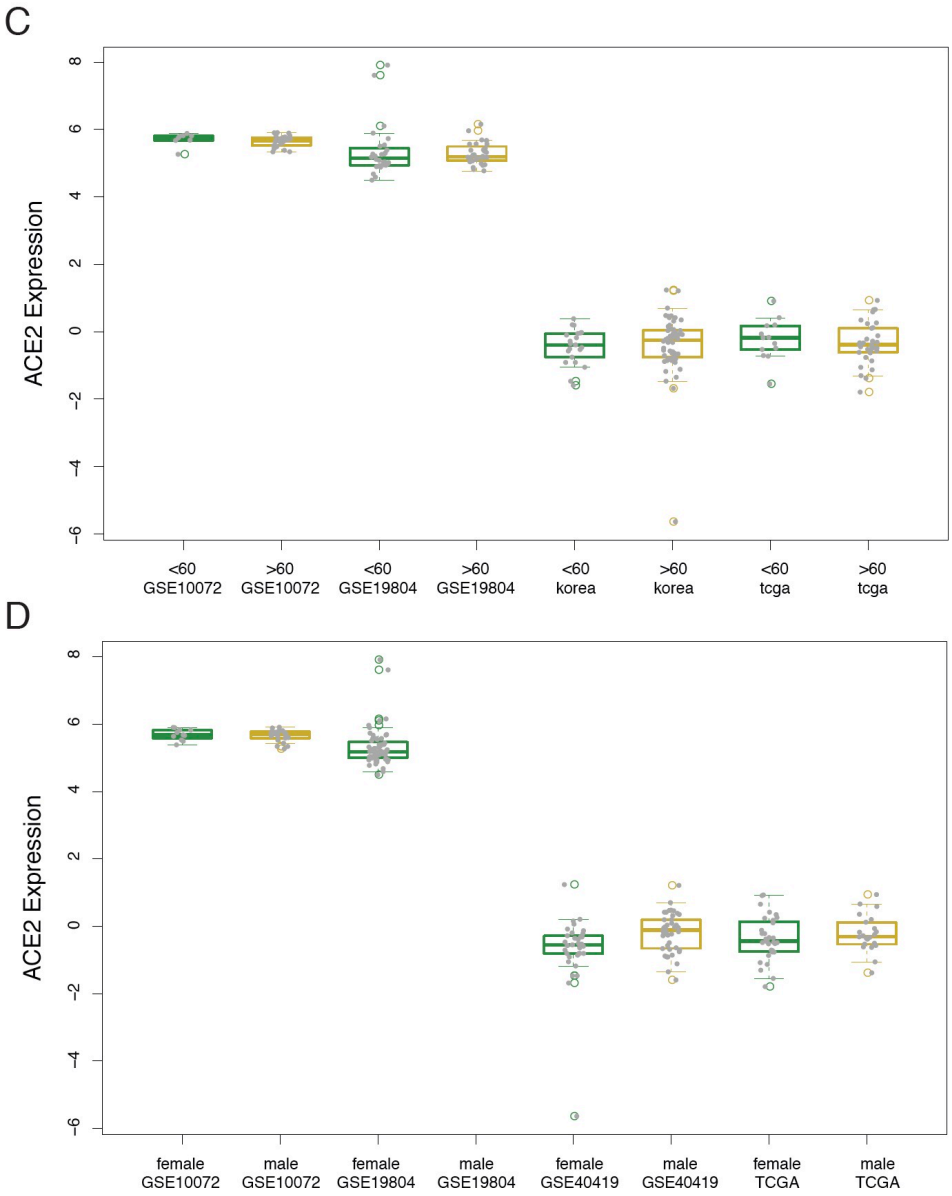


Figure 1. ACE2 gene expression profiling in groups.
A-D shows groups in race (Caucasian vs Asian), smoking (smoker vs non-smoker), age (>60 vs <60) and gender (male vs female). E shows the result from multivariate analysis with all factors including age, gender, race, smoking and platforms.

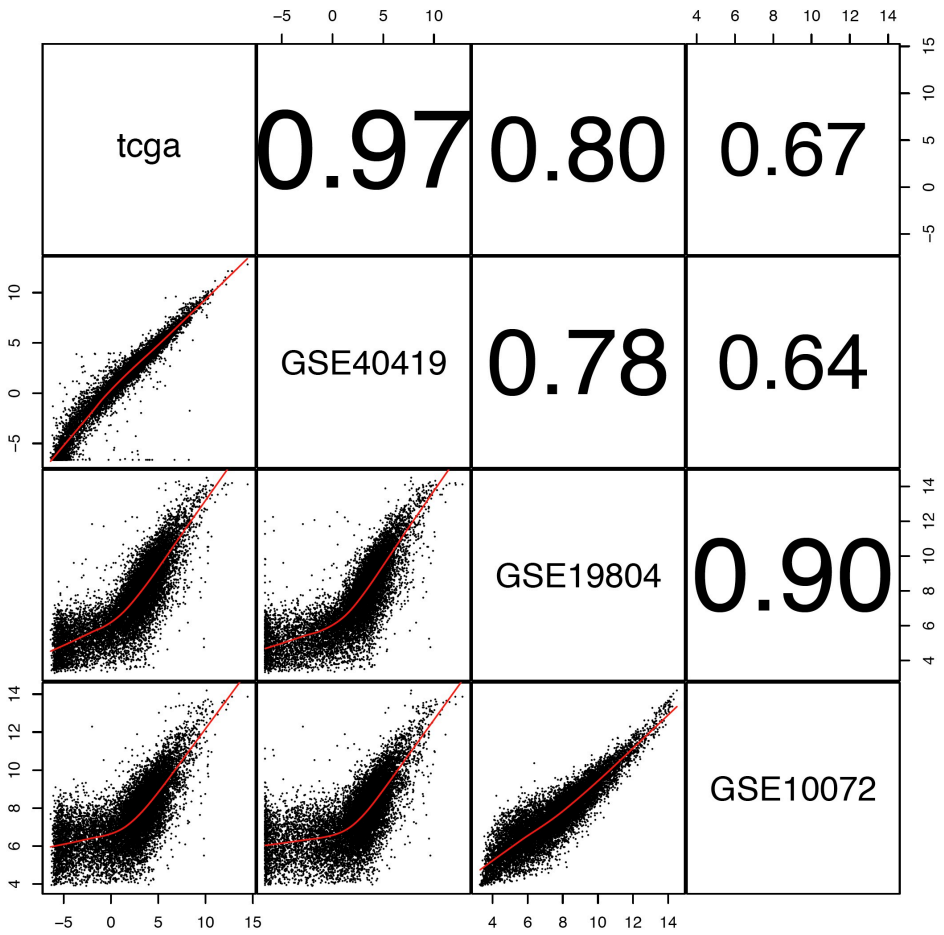


Figure S1. Correlation of four datasets.
Lower panel shows pairwise scatter plots of data mean across samples in each dataset.

Upper panel shows their corresponding Pearson correlation coefficients.

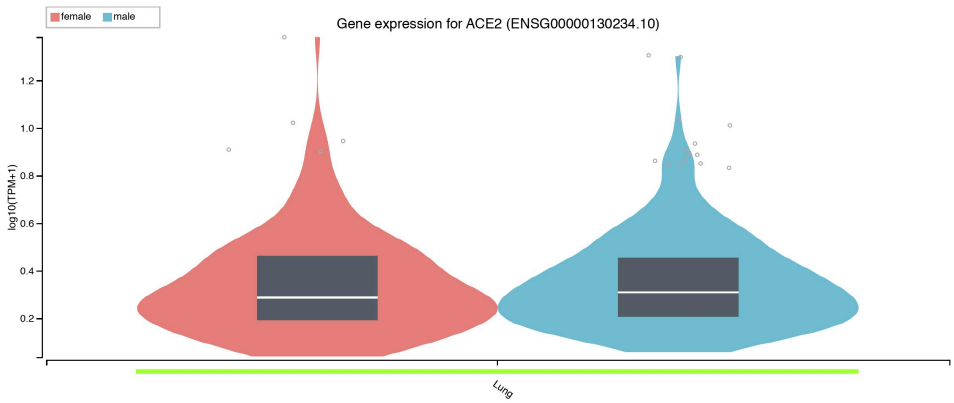


Figure S2. ACE2 gene expression in GTEx female and male lung tissues.
y-axix shows the log10 scaled RNA-seq Transcript Per Million (TPM) values.