 doi:10.20944/preprints202002.0408.v1

A Hint on the COVID-19 Risk: Population Disparities in Gene Expression of Three Receptors of SARS-CoV

Guoshuai Cai†, Xiang Cui*, Xia Zhu*, Jun Zhou*

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

†Address correspondence to:

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

*authors contributed equally to this manuscript

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2020 doi:10.20944/preprints202002.0408.v1

Abstract

The current spreading novel coronavirus SARS-CoV-2 is highly infectious and pathogenic and has attracted global attention. Recent studies have found that SARS-CoV-2 and SARS-CoV share around 80% of homology and use the same cell entry receptor, ACE2. These inspired us to study other receptors of SARS-CoV, which may be used for SARS-CoV-2 binding as well. In this study, we screened the gene expression of three receptors (ACE2, DC-SIGN and L-SIGN) in four datasets of normal lung tissue from lung adenocarcinoma patients and two single-cell RNA sequencing datasets from normal lung and bronchial epithelial cells separately. No significant difference in gene expression of these three receptors were found between gender groups (male vs female). We found higher gene expression of DC-SIGN in elder with age>60 and higher gene expression of L-SIGN in Caucasian than Asian. Similar to ACE2, we observed significantly higher DC-SIGN gene expression in the lungs of smokers, especially former smokers. However, smokers upregulate ACE2 and DC-SIGN gene expression in different cell types. In the whole lung, ACE2 is actively expressed in remodeled Alveolar Type II cells of former smokers, while DC-SIGN is largely expressed in monocytes of former smokers and dendritic cells of current smokers. In bronchial epithelium, no obvious gene expression of DC-SIGN and L-SIGN was observed while ACE2 was found to be actively expressed in goblet cells of current smokers and club cells of non-smokers. In conclusion, our findings may indicate that smokers, especially former smokers, and people over 60 have higher risk and are more susceptible to SARS-CoV-2 infection. Also, this study provides hints on possible SARS-CoV-2 pathogenicity mechanisms in lung infection.

Key words

Wuhan SARS-CoV-2, ACE2, DC-SIGN, L-SIGN, expression, susceptibility, race, age, gender, smoking, single cell

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2020 doi:10.20944/preprints202002.0408.v1

Introduction

A novel coronavirus SARS-CoV-2 from Wuhan has been spreading as a pandemic infection. SARS-CoV-2 is highly infectious and pathogenic through animal-to-human and human-to-human transmission and causes severe Coronavirus disease 2019 (COVID-19)[1]. As of February 19, the number of identified cases has hit 75204 globally (74280 and 924 were reported in China and other 25 countries, respectively)[2]. It is urgent to identify the susceptible population of SARS-CoV-2 for effective prevention and treatment. Previously we found upregulated ACE2 in smokers' lungs which may indicate an enhanced susceptibility of smokers to SARS-CoV-2 infection[3].

Coronavirus is a single-stranded RNA virus that can be divided into four main genres including the alpha, beta, gamma and delta coronaviruses[4]. Recent studies showed that SARS-CoV-2 is closely related to SARS-CoV, with around 80% identity of genome[5], belonging to the same β-genres[6] and causing similar symptoms such as fever, malaise, dry cough and acute respiratory response[7]. Moreover, Lu et al. found that SARS-CoV-2 and SARS-CoV have similar receptor-binding domain (RBD) structures[4] and further studies confirmed that they use the same cell entry receptor, ACE2[8]. These findings indicate that SARS-CoV-2 and SARS-CoV may largely share pathogenicity mechanisms.

Many viruses use multiple alternative receptors to enter host cells. Three receptors—ACE2, DC-SIGN and L-SIGN (gene symbol as *ACE2*, *CD209* and *CLEC4M* respectively) have been found to be involved in the pathogenicity of SARS-CoV[9-12]. DC-SIGN and L-SIGN are homologous C-type lectins receptors, which can identify carbohydrate structures of viral glycoproteins and play crucial roles in anchoring enveloped viruses [13]. In vivo, L-SIGN is largely expressed on endothelial cells in liver sinusoids and lymph nodes, whereas DC-SIGN is expressed on dendritic cells (DCs)[14]. They both capture virus and play important roles in virus transmission within the host[15,16]. Studies showed DC-SIGN is an independent receptor and synergistically works with ACE2 on the SARS-CoV viral entry which is mediated by the pH-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2020 doi:10.20944/preprints202002.0408.v1

dependent endocytosis[10,17,18] . Also, Jeffers et.al found L-SIGN is a potential portal of entry for SARS-CoV, similar to Ebola and Sindbis.[12] Paradoxically, L-SIGN can also internalize the virus and promote virus degradation in a proteasome-dependent manner[19]. Indeed, Chan et.al found homozygote L-SIGN in 69-nucleotide tandem repeats in exon 4 may play a protective role during SARS infection[11].

Based on above knowledge of SARS-CoV, we believe that ACE2 is not the only receptor of SARS-CoV-2 as well. Here, we studied the differences related to race, age, gender, and smoking status in the gene expression of three putative SARS-CoV-2 receptors (ACE2, DC-SIGN and CD209L) by analyzing four large-scale datasets of normal lung tissues. Also, we investigated the distribution of their gene expression among cell types by analyzing two lung tissue single-cell transcriptomic datasets. This study helps understand the pathogenicity and susceptibility of SARS-CoV-2 infection.

Methods

Bulk transcriptomics

We used two RNA-seq datasets and two DNA microarray datasets of normal lung tissues from lung adenocarcinoma patients, including a Caucasian RNA-seq dataset from TCGA (54 samples), an Asian RNA-seq dataset GSE40419 (77 samples), a Caucasian microarray dataset GSE10072 (33 samples) and an Asian microarray dataset GSE19804 (60 samples). The details and processing of data were described in our previous study[3].

Simple linear regressions were used to test the association of *ACE2*, *CD209* and *CLCE4M* gene expression with each single variable of age, gender, race and smoking status. Also, multiple linear regression was used to test the association of their gene expression with multiple factors (age, gender, race, smoking status and data platform). P-values and fold changes from group comparisons were visualized in dot plots. Also, ordinal regression was performed to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2020 doi:10.20944/preprints202002.0408.v1

investigate the trend between *ACE2, CD209* and *CLEC4M* expression and ordinal categorical smoking history (current smokers, former smokers and non-smokers). All data management, statistical analyses and visualizations were accomplished using R 3.6.3.

Single-cell transcriptomics

We analyzed two single-cell RNA sequencing datasets GSE122960 and GSE131391. The GSE122960 dataset was from lung tissue of 8 lung transplant donors, including 5 African American non-smokers, 1 Asian former smoker and 2 Caucasian current smokers. The GSE131391 dataset focused on bronchial epithelial cells from 6 never and 6 current smokers. The data normalization, high variable feature selection, data scaling, data visualization and cell type identification and other analyses were performed in the same way as our previous study[3].

Results

***ACE2* and *CD209* are overexpressed in lungs of smokers, especially former smokers**

Same as our previous study, we found smokers (including current smokers and former smokers) had a significantly higher gene expression of *ACE2* than non-smokers in GSE40419 (p-value<0.01, Fig. 1C) and TCGA (p-value=0.05, Fig. 1C) datasets. The GSE19804 microarray study, which focused on female non-smokers, was not included into the analysis. Further, we performed a multivariate analysis of smoking status to adjust the effects from other factors (platform, age, race and gender) and found smoking still shows a significant disparity in gene expression of *ACE2* (p-value<0.01, Table A). We also found that *CD209* was significantly upregulated in smokers after multi-factor adjustment (p-value<0.01, Table A) although the difference in each dataset didn't reach the statistical threshold of significance. No significant difference in *CLEC4M* gene expression between smokers and non-smokers was found either in simple or multiple regression analysis. Further, we studied the expression profiles of non-smoker, former-smoker and current smoker and found no significant trend of gene expression of *CD209* in these three separate datasets. The small sample size in each group might not have

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2020 doi:10.20944/preprints202002.0408.v1

enough power to detect the trend. However, we found higher mean expression of gene *CD209* in former smokers compared with non-smokers and current smokers in both GSE10072 and GSE40419 datasets, which is similar to our previous observation on *ACE2* gene expression (Fig. 2). Consistently, TCGA dataset showed higher average expression of *CD209* in recent quitters (<=15 years) compared to non-smokers, current smokers and smokers who have quit for longer durations (>15 years). Although these differences are not statistically significant which might due to the small sample size, these findings provide scientific hypotheses for further investigation.

Caucasian have higher *CLEC4M* lung gene expression than Asian

We observed higher gene expression of *CLEC4M* in Caucasian lung tissue samples compared with Asian lung tissue samples in both Microarray datasets (*p*-value<0.01) and RNA-seq datasets (*p*-value<0.01). Also, multivariate analysis showed a significant disparity in *CLEC4M* gene expression (*p*-value=3.75E-12, Table A). Differently, we didn't observe that in *CD209* and *ACE2* expression in multivariate analysis with adjustment of platform, age, gender and smoking-status.

***CD209* is upregulated in elder, no gender disparity was observed**

We found higher expression of *CD209* in the age>60 group than the age<60 group by single variable analysis on the GSE40419 dataset (*p*-value=0.05, Fig. 1A) and multivariate analysis on all four datasets (*p*-value=0.03, Table 1). No significant differences between age groups were found in *ACE2* and *CLEC4M* gene expression. And, we didn't find any significantly difference in the expression of any of three interested genes between gender groups.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2020 doi:10.20944/preprints202002.0408.v1

Table 1. Multivariate analysis of three gene expression

	ACE2		*CD209*		*CLEC4M*	
Variable	**Beta**	**P-value**	**Beta**	**P-value**	**Beta**	**P-value**
Race: Caucasian	0.09	0.36	-0.07	0.52	1.89	**3.75E-12**
Age: >60	-0.01	0.90	0.25	**0.03**	-0.28	0.27
Gender: male	0.11	0.35	0.09	0.47	0.05	0.87
Smoking: smoker	0.31	**0.008**	0.41	**0.003**	-0.31	0.31

Smokers have upregulated *ACE2* and *CD209* gene expression in different cell types in lung

In our previous study, we analyzed two single cell datasets and found that *ACE2* was expressed in goblet cells of smokers and club cells of non-smokers, respectively, and upregulated in remodeled Alveolar Type II (AT2) cells of former smokers. Applying the same method, we identified 13 different cell populations from single-cell RNA sequencing of whole-lung tissue and found that *CD209* was highly expressed in monocytes and dendritic cells (DCs), which is consistent with previous reports [15]. We found *CD209* was actively expressed in monocytes in the former smoker but not distinctly expressed in current smokers and non-smokers. And, we observed active *CD209* expression in DCs from current smokers but not from a former smoker and non-smokers (Fig.3). Given the close lineage relationship between monocytes and DCs, this may indicate smokers has upregulated *CD209* in the monocyte-DC lineage. In summary, similarly to *ACE2*, *CD209* may be associated with smoking history in different cell types as well.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2020 doi:10.20944/preprints202002.0408.v1

We didn't observe the expression of *CLEC4M* in smokers, quitters or non-smokers in current analyzed dataset. And, *CD209* and *CLEC4M* were not obviously expressed in human bronchial epithelial cells (Fig.S1).

Discussion

In this study, we investigated disparities of four factors including gender, age, race and tobacco-use in lung gene expression of three SARS-CoV receptors, which are putative receptors of SARS-CoV-2. These receptors might be independent viral entry portal based on our knowledge and we didn't observe significant correlations between genes in current analyzed four datasets (Fig.S2). We found lung tissue samples from smokers especially former smokers had significantly higher ACE2 and DC-SIGN gene expression than those from non-smokers. This is consistent with the recent clinical characteristics of 1,099 cases that the quitters had a higher risk of severe disease[20]. Indeed, cigarette smoking has been confirmed to increase the susceptibility to various respiratory diseases such as pneumococcal disease and pneumonia[21,22]. Severance et.al also identified smoking status as well as race, socioeconomic status as risk factors for exposure of four non-SARS coronavirus strains among US population[23]. Moreover, experimental studies have showed cigarette smoke exposure could upregulate pulmonary ACE2 activities in WT mice and abnormal expression of ACE2-related pathway may induce lung injury in a rat model as well[24,25], which support our findings. Further, our single cell RNA sequencing analysis found that smokers have upregulated ACE2 and DC-SIGN gene expression in AT2 and monocytes/DCs respectively. Based on the knowledge that SARS-CoV infects lung cells expressing ACE2 and could be enhanced by monocytes and DCs transfer through DC-SIGN[26,27], we inferred that smoking may remodel lung cells to be more susceptible to SARS-CoV-2 infection by upregulating ACE2 expression in AT2 cells and further enhance the virus entry by upregulating DC-SIGN expression in DCs. Meanwhile, smoking enhances the inflammatory behavior of DCs and promotes the body's inflammatory response[28]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2020 doi:10.20944/preprints202002.0408.v1

and maybe partially contribute to the cytokine storm in severe patients. Also possibly, with upregulated DC-SIGN, smokers' DCs are more active in presenting viral antigen and invoking the adaptive immunity to target and remove infected cells and pathogens, which may explain that current smokers show lower risk of developing severe disease than former smokers[20].

Also, we found higher gene expression of L-SIGN in Caucasian samples than Asian samples. Studies showed that SARS-CoV can use L-SIGN to mediate virus entry[12]. Controversially, Chan et.al found homozygote L-SIGN (*CLEC4M*) plays a protective role during SARS-CoV infection by binding virus for efficient internalization and degradation[11]. Therefore, L-SIGN might be a double-blade sword during viral infection. This finding may provide another hint on reasons for that coronavirus (including SARS-CoV and SARS-CoV-2) frequently outbreak in China. However, we don't know whether L-SIGN is an important receptor for SARS-CoV-2 and whether L-SIGN is protective to its infection. Also, this difference may be due to the inter-dataset variation although no significant systematic variation was observed[3] and similar results were observed after quantile normalization on samples (data not shown). Further studies are required for conclusions.

Besides, we also observed significant higher gene expression of DC-SIGN in population ages 60 and above. Together with the aging immune system and organ health, this may lead to the higher COVID-19 severity in elderly patients[20,29]. No significant disparities in the gene expression of these three receptors were found between gender groups (male vs female).

Similar with our previous study on ACE2[3], this study has several limitations. First, the data analyzed were from normal lung normal lung tissue of patients with lung adenocarcinoma, which may be different with the lung tissue of healthy people. Second, the small sample size of current single-cell transcriptome datasets may fail to capture inter-subject variation. Third, other factors which are unavailable in current datasets may have confounding effects. Besides, whether DC-SIGN and L-SIGN are the true receptors of 2019-nCoV has not been confirmed.

 doi:10.20944/preprints202002.0408.v1

In this study, we observed disparities of age, race and tobacco use in gene expression of three receptors of SARS-CoV. This provides hints on possible SARS-CoV-2 infection pathogenicity and risk factors. Developing conspiracy theories based on unilaterally interpretation of this study is wrong and unwise.

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.

Reference

1. Chan JF-W, Yuan S, Kok K-H, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020:S0140-6736(20)30154-9.

2. Coronavirus disease 2019 (COVID-19) Situation Report – 30. 2019.

3. Cai G. Bulk and single-cell transcriptomics identify tobacco-use disparity in lung gene expression of ACE2, the receptor of 2019-nCov. medRxiv 2020:2020.02.05.20020107.

4. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. Methods Mol Biol 2015;1282:1-23.

5. Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020:10.1038/s41586-020-2012-7.

6. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. J Virol 2020:JVI.00127-20.

7. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020:S0140-6736(20)30183-5.

8. 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.

9. Li W, Moore MJ, Vasilieva N, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 2003;426:450-4.

10. Yang Z-Y, Huang Y, Ganesh L, et al. pH-dependent entry of severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through DC-SIGN. J Virol 2004;78:5642-50.

11. Chan VSF, Chan KYK, Chen Y, et al. Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. Nat Genet 2006;38:38-46.

12. Jeffers SA, Tusell SM, Gillim-Ross L, et al. CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. Proc Natl Acad Sci U S A 2004;101:15748-53.

13. Lozach PY, Burleigh L, Staropoli I, Amara A. The C type lectins DC-SIGN and L-SIGN: receptors for viral glycoproteins. Methods Mol Biol 2007;379:51-68.

14. Gardner JP, Durso RJ, Arrigale RR, et al. L-SIGN (CD 209L) is a liver-specific capture receptor for hepatitis C virus. Proc Natl Acad Sci U S A 2003;100:4498-503.

15. Khoo U-S, Chan KYK, Chan VSF, Lin CLS. DC-SIGN and L-SIGN: the SIGNs for infection. J Mol Med (Berl) 2008;86:861-74.

16. da Silva RC, Segat L, Crovella S. Role of DC-SIGN and L-SIGN receptors in HIV-1 vertical transmission. Hum Immunol 2011;72:305-11.

17. 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. Sci China Life Sci 2020:10.1007/s11427-020-1637-5.

18. Han DP, Lohani M, Cho MW. Specific asparagine-linked glycosylation sites are critical for DC-SIGN- and L-SIGN-mediated severe acute respiratory syndrome coronavirus entry. J Virol 2007;81:12029-39.

19. Boily-Larouche G, Iscache A-L, Zijenah LS, et al. Functional genetic variants in DC-SIGNR are associated with mother-to-child transmission of HIV-1. PLoS One 2009;4:e7211-e.

20. Guan W-j, Ni Z-y, Hu Y, et al. Clinical characteristics of 2019 novel coronavirus infection in China. medRxiv 2020:2020.02.06.20020974.

21. Huttunen R, Heikkinen T, Syrjänen J. Smoking and the outcome of infection. J Intern Med 2011;269:258-69.

 doi:10.20944/preprints202002.0408.v1

22. Nuorti JP, Butler JC, Farley MM, et al. Cigarette smoking and invasive pneumococcal disease. Active Bacterial Core Surveillance Team. N Engl J Med 2000;342:681-9.

23. Severance EG, Bossis I, Dickerson FB, et al. Development of a nucleocapsid-based human coronavirus immunoassay and estimates of individuals exposed to coronavirus in a U.S. metropolitan population. Clin Vaccine Immunol 2008;15:1805-10.

24. Hung Y-H, Hsieh W-Y, Hsieh J-S, et al. Alternative Roles of STAT3 and MAPK Signaling Pathways in the MMPs Activation and Progression of Lung Injury Induced by Cigarette Smoke Exposure in ACE2 Knockout Mice. Int J Biol Sci 2016;12:454-65.

25. Yilin Z, Yandong N, Faguang J. Role of angiotensin-converting enzyme (ACE) and ACE2 in a rat model of smoke inhalation induced acute respiratory distress syndrome. Burns 2015;41:1468-77.

26. Chen J, Subbarao K. The Immunobiology of SARS*. Annu Rev Immunol 2007;25:443-72.

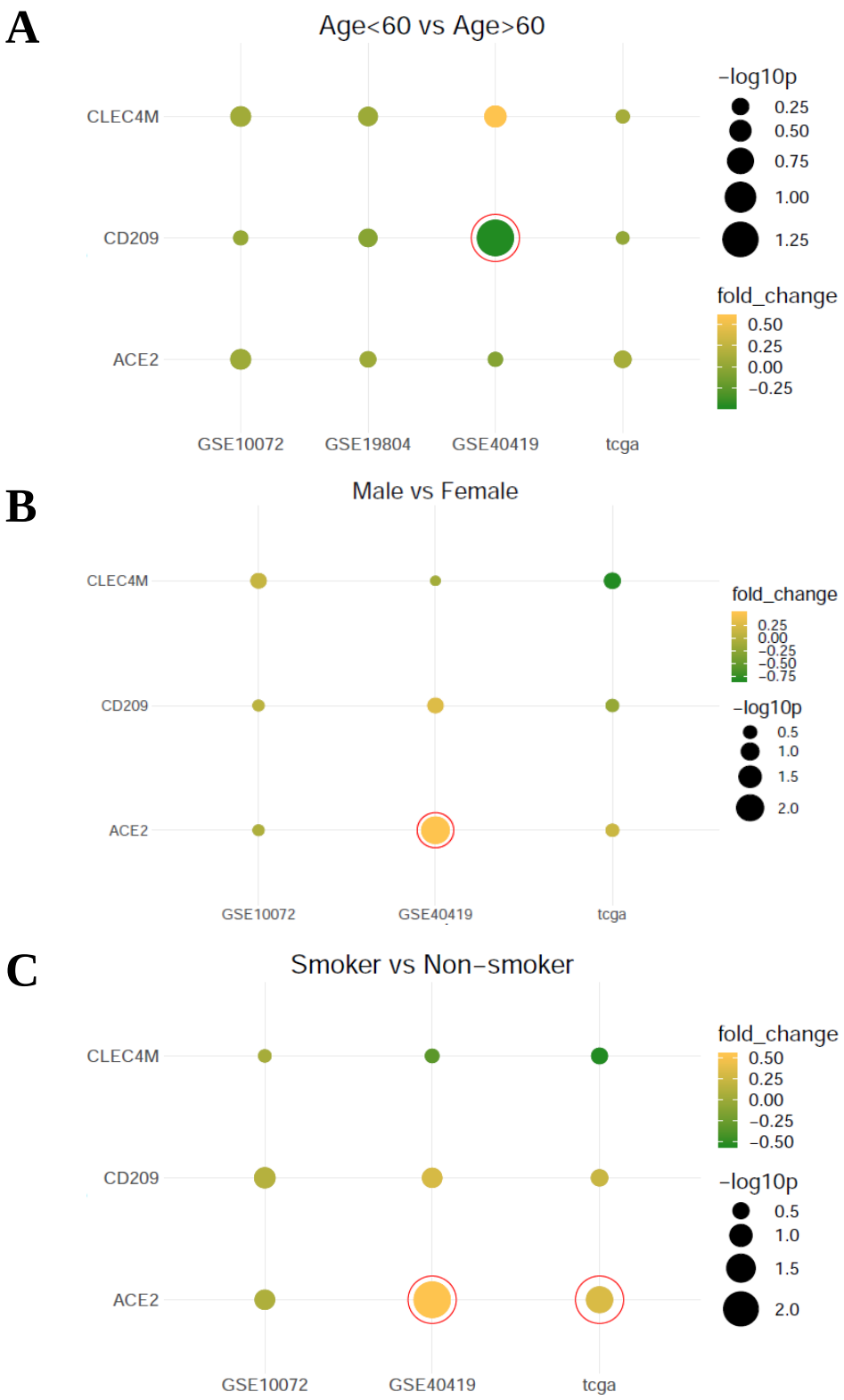
27. Yen Y-T, Liao F, Hsiao C-H, Kao C-L, Chen Y-C, Wu-Hsieh BA. Modeling the early events of severe acute respiratory syndrome coronavirus infection in vitro. J Virol 2006;80:2684-93.

28. Prado C, Iruretagoyena M, Burgos PI, Pacheco R. Smoking promotes exacerbated inflammatory features in dendritic cells of Chilean rheumatoid arthritis patients. Rev Med Chil 2018;146:150-9.

29. Gralinski LE, Menachery VD. Return of the Coronavirus: 2019-nCoV. Viruses 2020;12:E135.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2020 doi:10.20944/preprints202002.0408.v1

Figure



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2020 doi:10.20944/preprints202002.0408.v1

Figure 1. Dot plots of ACE2, CD209 and CLEC4M gene expression in groups.

A, **B** and **C** shows groups in age (>60 vs <60), gender (male vs female) and smoking (non-smoking vs smoking). The color shows fold change comparing groups, while the size indicates the -log10 (*p*-value). The significant difference was indicated by red circles surrounding dots.

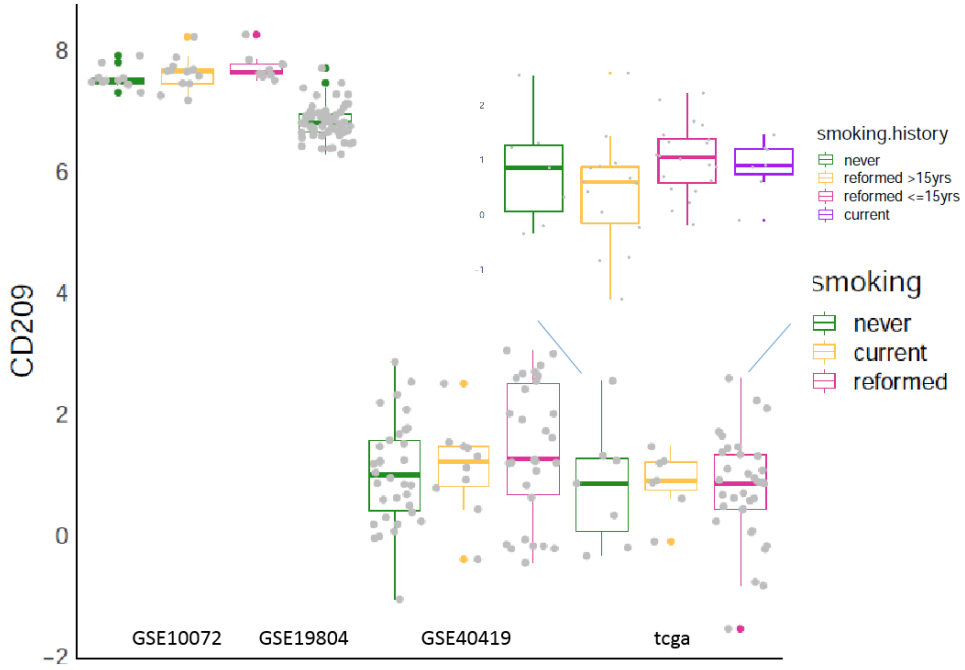


Figure 2. CD209 gene expression profiles in smoking groups.

The figure shows groups of never-smoker/non-smoker, reformed/former smoker and current smoker. TCGA dataset has more categories of smoking history, including never-smoker, smoker reformed more than 15 years, smoker reformed less than 15 years and current smoker.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2020 doi:10.20944/preprints202002.0408.v1

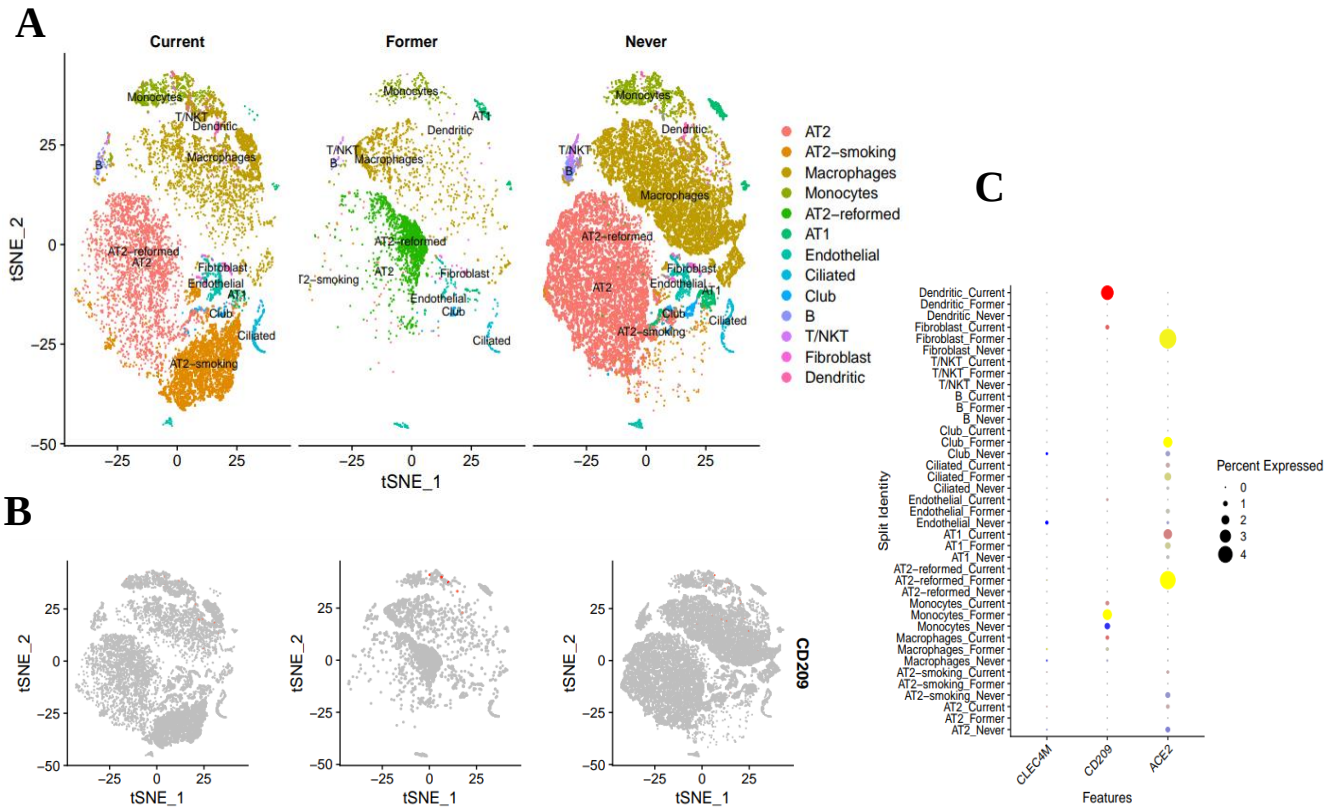


Figure 3. Single cell transcriptomics of lung cells

A shows tSNE plots of single-cell transcriptome profiles and identified cell types from never smokers, former smokers and current smokers. **B** shows *CD209* gene expression in cells from never smokers, former smokers and current smokers separately. **C** shows detection rates of *CD209* in each type of cells from never smokers, former smokers and current smokers

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2020 doi:10.20944/preprints202002.0408.v1

Supplementary File

Figure S1. Single-cell transcriptomics of bronchial epithelium cells.

A shows tSNE plots of single-cell transcriptome profiles from never smokers and current smokers. **B** shows identified cell types. **C** shows detection rates of markers in each cell cluster. **D** and **E** show *ACE2* expression in cells of current smokers and never smokers, separately.

Figure S2. Correlation of genes *ACE2*, *CLEC4M* and *CD209* in four datasets.

Lower panel shows pairwise scatter plots of data mean of each gene expression across samples. Upper panel shows their corresponding Pearson correlation coefficients.