

Single-cell RNA expression profiling shows that ACE2, the putative receptor of COVID-2019, has significant expression in nasal and mouth tissue, and is co-expressed with TMPRSS2 and not co-expressed with SLC6A19 in the tissues

Chao Wu* and Min Zheng*

State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University. Hangzhou, China.

*Correspondence to: wuchao1984@zju.edu.cn (Chao Wu) or minzheng@zju.edu.cn (Min Zheng)

Abstract

A novel coronavirus (COVID-2019) was first identified in Wuhan, Hubei Province, and then spreads to the other Provinces of China. COVID-2019 was reported to share the same receptor, Angiotensin-converting enzyme 2 (ACE2), with SARS-CoV. But the infection rate of COVID-2019 is much higher than SARS-CoV. The biophysical and structural evidence showed that the COVID-2019 binds ACE2 with 10~20 times affinity than SARS-CoV. TMPRSS2 cleaves ACE2 and facilitates the entry of the virus into host cells. The presence of SLC6A19 may block the access of TMPRSS2 to the cutting site on ACE2 and weaken the entry of COVID-2019 into host cells. Here based on the public single-cell RNA-Seq datasets, we analyzed the ACE2 expression in the nasal, mouth, lung, and colon tissues. We find that the number of ACE2-expressing cells in the nasal and mouth tissues is comparable to the number of ACE2-expressing cells in the lung and colon tissues. We also find that ACE2 tends to be co-expressed with TMPRSS2 and not co-expressed with SLC6A19 in the nasal and mouth tissues. With the results, we infer that nasal and mouth tissues may be the first host cells of COVID-2019 infection. In our previous report in medRxiv and a recurrent report in New England Journal of Medicine, the COVID-2019 load tends to be higher in the nasal-swabs than in throat-swabs. We believe the roles of nasal and mouth tissues in COVID-2019 infection should be investigated, and we need to pay more attention to protect nose and mouth from COVID-2019 infection.

Introduction

Severe infection by COVID-2019 could result in acute respiratory distress syndrome (ARDS) and sepsis, causing death in approximately 2% of infected individuals¹. Once contacted with the human airway, the spike proteins of COVID-2019 can bind the surface receptors of sensitive cells, and mediate the virus to enter into target cells. Recently, Xu et al. modeled the spike protein and indicated that Angiotensin-converting enzyme 2 (ACE2) could be the receptor for the COVID-2019². Zhou et al. showed that ACE2 is essential for COVID-2019 to enter HeLa cells³. Daniel Wrapp et al. analyzed the structure of the COVID-2019 Spike protein and found that the COVID-2019 binds ACE2 with 10~20 times affinity than SARS-CoV⁴. These results indicated that ACE2 is likely to be the main receptor for COVID-2019 infection. Molecular evidence showed that TMPRSS2 cleaves ACE2 and facilitates the entry of SARS-CoV into host cells⁵, while the presence of SLC6A19 may block the access of TMPRSS2 to the cutting site on ACE2 and weaken the entry of COVID-2019 into host cells⁶.

Epidemiological evidence showed that the infection rate of COVID-2019 is much higher than SARS-CoV. Yang Yang et al. used R_0 , defined as the average number of secondary infections a patient can generate in a fully susceptible population, to estimate the basic reproductive number⁷. They found the R_0 of COVID-2019 is 3.77, while the R_0 of SARS-CoV is between 0.85 and 3. With the fact that a large number of COVID-2019 infected people have no or few clinical symptoms, we need to understand the route of COVID-2019 infection, which is a work of great urgency.

Previous studies have investigated the RNA expression of ACE2 in 72 human tissues⁸. They found ACE2 is abundantly present in humans in the epithelia of the lung and small intestines. They also found ACE2 expression in the basal layer of the non-keratinizing squamous epithelium in nasal and oral mucosa and the nasopharynx. However, the experiment is conducted at the bulk level. The number of cells expressing ACE2 in each tissue could not be estimated. In previous work, we found that a significant number of epithelial cells from nasal tissue have ACE2 expression, and nasal-swabs tend to have a higher COVID-2019 load than the throat-swabs⁹. Lirong Zou et al. also found that the COVID-2019 load tends to be higher in nasal-swabs than in throat swabs in the asymptomatic or minimally symptomatic patients¹⁰. Here, we analyzed ACE2 single-cell expression profiles in the non-immune cells of the nasal, mouth, lung, and colon tissues. We find that 2.5% non-immune nasal tissue cells, 2% non-immune mouth tissue cells, 5.6% non-immune lung tissue cells, and 2.8% epithelial cells of the colon have ACE2 expression. For the non-immune tissue cells, the percent of ACE2-expressing cells in the nasal and mouth tissues is comparable to the percent of ACE2-expressing cells in the lung and colon tissues. We also find the non-immune cells expressing ACE2 tend to express TMPRSS2 but not to express SLC6A19 in the nasal and mouth tissues, which suggests they are sensitive to the infection of COVID-2019. With the results, we infer that the nasal and mouth tissues may be the first host cells of COVID-2019 infection. We need to pay more attention to protect the nose and mouth from COVID-2019 infection.

Result

2.5% non-immune nasal tissue cells, 2% non-immune mouth tissue cells, 5.6% non-

immune lung tissue cells, and 2.8% non-immune colon tissue cells express ACE2

Sandra Ruiz García et al. profiled single-cell gene expression of airway epithelium from nasal brushing, bronchial biopsy, and turbinate¹¹. Reyfman, P. et al. sequenced the single-cell gene expression of the lung (bronchioli terminales)¹². The gene expression of the bronchial epithelium was profiled by Duclos GE et al. at single-cell resolution¹³. Puram SV et al. sequenced squamous cell carcinoma tissue of mouth at single-cell resolution¹⁴. Smillie CS et al. profiled single-cell gene expression of colon epithelial cells¹⁵.

Using Seurat, we performed unsupervised graph-based clustering on these single-cell RNA-Seq datasets. Then, We ran the Uniform Manifold Approximation and Projection (UMAP) dimensional reduction technique to visualize the data. Next, we used the violin plot to find the clusters with significant PTPRC (CD45) expression in the datasets and filtered out the clusters. Thus, we got non-immune cells in the datasets.

We calculated the percent of ACE2-expressing cells in each dataset (Table 1). We find that, in nasal tissue, 2.5% non-immune cells from nasal brushing and 1.7% non-immune cells from turbinate have ACE2 expression. 2% non-immune malignant and normal cells from mouth show ACE2 transcription. In lung tissue, 0.2% non-immune cells from the bronchial biopsy, 5.6% non-immune cells from bronchial brushings, and 1.1% non-immune cells from bronchioli terminales have ACE2 transcription. 2.8% non-immune cells from colon epithelial cells express ACE2.

ACE2 tends to be co-expressed with TMPRSS2 in the nasal, mouth, lung, and colon tissues and co-expressed with SLC6A19 in the colon tissue at single-cell resolution

We employed the hypergeometric test to test the co-appearance of ACE2 expression, TMPRSS2 expression, and SLC6A19 expression in the non-immune cells. We find that ACE2-expressing cells of all the single-cell RNA-Seq datasets tend to express TMPRSS2 at single-cell resolution (Table 1; P-value <0.05).

We detected no or few SLC6A19-expressing cells in the single-cell RNA-Seq datasets from nasal, mouth, and lung tissues. We find that ACE2-expressing cells from colon tissue tend to express SLC6A19 at single-cell resolution (Table 1; P-value <0.05). We plotted the distribution of the detected gene number per cell in each single-cell RNA-Seq dataset. The median of the number of detected genes in the colon tissue dataset is around 1000, which is lower than the median of the number of detected genes in most other datasets from nasal, mouth, and lung tissues (Figure 1). We believe that the single-cell RNA-Seq datasets have detected enough genes per cell. The fact that no or few SLC6A19-expressing cells were detected in the nasal, mouth, and lung tissue should not be attributed to the data quality.

Discussion

With single-cell RNA-Seq technology, we precisely calculated the number of ACE2-expressing cells in nasal, mouth, lung, and colon tissues and find the number of ACE2-expressing cells in nasal and mouth tissues is comparable to the number of ACE2-expressing cells in the lung and intestine tissues. Also, we found that ACE2 tends to be co-expressed with TMPRSS2 in the nasal and mouth tissues. Since no or few SLC6A19-expressing cells were detected in the single-cell RNA-Seq datasets from nasal and mouth tissues. We infer that ACE2 tends to be not co-expressed with SLC6A19 in the nasal and mouth tissues. It has been

reported that TMPRSS2 will facilitate the virus into host cells, while SLC6A19 will prevent the virus into host cells. We believe the nasal and mouth tissue could be the first host cells of COVID-2019 infection.

Here, we emphasize the importance of wearing a mask to protect people from COVID-2019 infection. We further emphasize the necessity of detecting COVID-2019 in both the nasal tissue and throat tissue samples, with the fact that a large number of infected people have no or few clinical symptoms. We also think it is not a good idea to let close contacts isolate themselves at homes. The families will become the victim if they don't always wear masks.

We admit the single-cell profiling of mouth tissue with squamous cell carcinoma is not an ideal model to study ACE2 expression in mouth tissue. However, it is the only single-cell RNA-Seq data that we can find having profiled mouth tissue at single-cell resolution. The malignant cells are derived from normal epithelial cells of the mouth; we believe most of the malignant cells should still maintain their tissue specificity.

Method

The single-cell RNA-Seq datasets of mouth, bronchial brushing, lung (bronchioli terminales), and colon epithelial tissues were downloaded from GSE103322, GSE131391, GSE122960, and SCP259 (Single Cell Portal). The single-cell RNA-Seq datasets of turbinate, nasal brushing, and bronchial biopsy were downloaded from GSE121600.

Single-cell RNA-Seq dataset pre-processing

We employed Seurat (3.1.4) to process the single-cell RNA-Seq datasets. At first, we filtered out the cells 1-expressing less than 200 genes, or 2-highly expressing mitochondrial genes, in which mitochondrial genes' reads account for more than 25% of the total reads. We filtered out the genes expressing in less than three samples. Then, we got the processed single-cell RNA-Seq datasets.

Single-cell RNA-Seq dataset clustering and visualization

We employed Seurat in default mode to cluster and visualize cell-clusters (See supplemental file S1.txtc for the R code).

Identification of the non-immune cells

We used the violin plot to check the PTPRC expression in each cluster in each single-cell RNA-Seq dataset. The cluster whose PTPRC expression having a spindle body in the violin plot was filtered. Thus, we got the non-immune cells of each single-cell RNA-Seq dataset.

Test the significance of enrichment of ACE2-expressing cells in TMPRSS2-expressing cells and SLC6A19-expressing cells.

We employed the hypergeometric test to test the significance of enrichment of ACE2-expressing cells in TMPRSS2-expressing cells. Supposed N is the number of total sequenced cells, M is the number of TMPRSS2-expressing cells, K is the number of ACE2-expressing cells, we calculated the possibility (p) of finding x or more than x cells of ACE2 expression and TMPRSS2 expression when we randomly picked K cells from total sequenced cells (N). We used R function `phyper` to calculate p as follow,

$$p = 1 - \text{phyper}((x - 1), M, (N - M), K);$$

where $K > 3$, we do not calculate p for the data whose $K \leq 3$.

We employed the hypergeometric test to test the significance of enrichment of ACE2-expressing cells in SLC6A19-expressing cells in the same way.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of supporting data

All data presented in this manuscript are freely available from GEO database and Single Cell Portal.

Competing interests

Not applicable.

Funding

Not applicable.

Authors' contributions

CW and MZ conceived and designed the study. CW collected and analyzed the data. CW and MZ revised and finalized the manuscript. All authors read and approved the manuscript.

Acknowledgements

This article is dedicated to the people who are fighting with 2019nCoV

Authors' information

CW is a research assistant professor and MZ is a professor. They both work at State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University. Hangzhou, China.

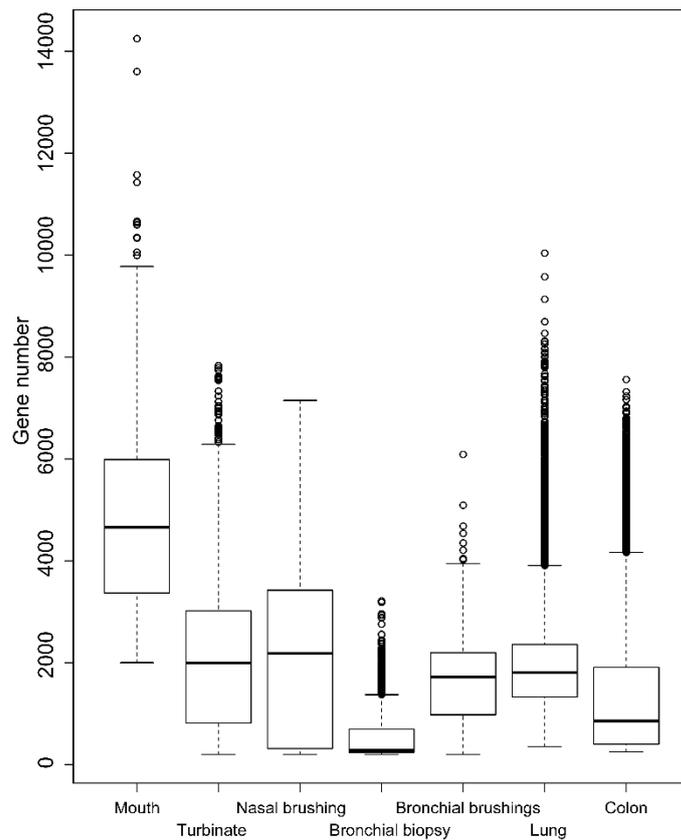


Figure 1 The distribution of the number of detected genes per cell in the seven single-cell RNA-Seq datasets

Table 1: Number of total sequenced cells, ACE2-expressing cells, TMPRSS2-expressing cells and SLC6A19-expressing cells in the nasal, mouth, lung and colon tissues

Samples	Total Cells	ACE2	ACE2/Total cells	TMPRSS2	TMPRSS2 & ACE2	P-value	SLC6A19	SLC6A19 & ACE2	P-value
Mouth	4239	85	0.02	79	5	2.03E-02	4	0	1.00E+00
Turbinate	6066	106	0.017	632	29	6.54E-07	0	0	-
Nasal brushing	6424	163	0.025	1134	61	1.01E-09	0	0	-
Bronchial biopsy	10795	24	0.002	199	3	9.39E-03	0	0	-
Bronchial brushings	921	52	0.056	52	8	6.53E-03	0	0	-
Lung	25291	280	0.011	8576	144	1.02E-09	0	0	-
Colon	30038	828	0.028	6414	503	0.00E+00	328	39	2.04E-14

We calculated the P-value of the hypergeometric test to measure the significance of enrichment of ACE2-expressing cells in TMPRSS2-expressing cells and SLC6A19-expression cells. However, we omitted the data whose ACE2-expressing-cells/TMPRSS2-expressing-cells/SLC6A19-expression-cells are less than 4.

Reference

- 1 Huang, C.-I. et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* (2020).
- 2 Xu, X.-t. 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* 63 (2020).

3 Zhou, P. et al., Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *bioRxiv* (2020).

4 Daniel Wrapp et al., Cryo-EM Structure of the COVID-2019 Spike in the Prefusion Conformation. *bioRxiv* (2020).

5 Heurich A et al., TMPRSS2 and ADAM17 Cleave ACE2 Differentially and Only Proteolysis by TMPRSS2 Augments Entry Driven by the Severe Acute Respiratory Syndrome Coronavirus Spike Protein. *J Virol.* 2014 Jan;88(2):1293-307.

6 Qiang Zhou et al., Structure of dimeric full-length human ACE2 in complex with B0AT1. *bioRxiv* (2020).

7 Yang Yang et al., Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. *bioRxiv* (2020).

8 Hamming, I. et al., Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *Journal of Pathology* 203, 631-637 (2004).

9 Chao Wu et al., Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCoV, in the nasal tissue. *bioRxiv* (2020).

10 Lirong Zou et al., SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *New England Journal of Medicine*, February 19, 2020

11 Ruiz Garcia S et al., Novel dynamics of human mucociliary differentiation revealed by single-cell RNA sequencing of nasal epithelial cultures. *Development.* 2019 Oct 23;146(20).

12 Reyfman, P. et al., Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *American Journal of Respiratory and Critical Care Medicine* (2018).

13 Duclos GE et al., Characterizing smoking-induced transcriptional heterogeneity in the human bronchial epithelium at single-cell resolution. *Sci Adv* 2019 Dec;5(12).

14 Puram SV et al., Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell.* 2017 Dec 14;171(7):1611-1624.e24.

15 Smillie CS et al., Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell.* 2019 Jul 25;178(3):714-730.e22.