

Digital Object Identifier:10.11989/JEST.1674-862X.20224017

Article Number: 20224017

# COVID-19 Related Research by Data Mining in Single Cell Transcriptome Profiles

Zi-Wei Wang\* | Chi-Chang Chang\* | Quan Zou

**Abstract**—The novel pneumonia in Wuhan, China has outbroken and drawn public attention all over the world. As a newly emerging area, single cell sequencing also exerts its power in the battle over the epidemic. In this review, the up-to-date knowledge of COVID-19 and its receptor is summarized, followed by a collection of the mining of single cell transcriptome profiling data for the information in aspects of the vulnerable cell types in humans and the potential mechanisms of the disease.

**Index Terms**—Coronavirus Disease 2019 (COVID-19), bioinformatics, data mining, single cell sequencing.

## 1. Introduction

A pneumonia caused by coronavirus attacked Wuhan, China by the end of the year 2019 and has been declared as a Public Health Emergency of International Concern (PHEIC) by the World Health Organization (WHO). The pathogen causing the infection has been identified as a novel kind of coronavirus, later named as COVID-19 (Coronavirus Disease 2019), which could spread through contacting patients or contaminated surface, or respiratory droplets in the air. The symptoms of this infection resembles severe acute respiratory syndrome (SARS) and middle east respiratory syndrome (MERS) in aspects of fever, cough, diarrhea, nausea and even multi organs failure.

In the battle between humans and infectious epidemic, the newly emerged technique of single cell transcriptome profiling also has its place. Single cell ribonucleic acid sequencing (scRNA-seq) has been developed at an exciting speed and applied in an increasing scale, which offers an unprecedentedly powerful tool to clarify more detailed mechanism of biological function as well as disease at cellular resolution. There is no scRNA-seq data from COVID-19 patients up to date, though, as the receptor of COVID-19 has already been identified, the existing single cell transcription profiles in database can serve as samples of susceptible population for data mining to unveil the vulnerability of different parts of the human body, the route of propagation, and the possible mechanism of this newly emerged disease.

---

\*Corresponding author

Manuscript received 2020-02-24; revised 2020-04-02.

The work was supported by the National Key R&D Program of China under Grant No. 2018YFC0910405 and the National Natural Science Foundation of China under Grants No.61922020, No. 61771331, and No. 91935302.

Z.-W. Wang and Q. Zou are with the Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054 (e-mail: zouquan@nclab.net)

C.-C. Chang is with the School of Medical Informatics, Chung Shan Medical University & IT office, Chung Shan Medical University Hospital, Taichung 40201

Publishing editor: Xuan Xie

## 2. Prerequisite of COVID-19 and Its Receptor in Humans

The knowledge about the COVID-19 virus and its route into the human body has been deciphered rapidly, based on which the mining of scRNA-seq data could find its way into the research of COVID-19. As a self-replicable single-strand RNA virus, its RNA has been sequenced right after the outbreak of the epidemic, with the result showing that the COVID-19 RNA shares 76.47% sequence identity with SARS-Cov<sup>[1]</sup>. COVID-19 has been found to invade the human body through its association with angiotensin I converting enzyme 2 (ACE2) by its spike glycoprotein<sup>[2]</sup>, indicating that ACE2 is the putative receptor of COVID-19. ACE2 is a member of the angiotensin-converting enzyme family of dipeptidyl carboxy-dipeptidases, and it plays a role in the regulation of cardiovascular and renal function, as well as fertility<sup>[3]</sup>.

ACE2 is also the receptor of SARS-Cov. According to the structural analysis of COVID-19 spike glycoprotein using cryo-electron microscopy, the dissociation constant  $K_D$  of COVID-19 to its receptor ACE2 is 15 nM<sup>[4]</sup>, indicating an affinity around 10 to 20 folds higher than that of SARS-Cov of which the  $K_D$  is 325.8 nM. Therefore, studies using single cell profiles based on ACE2 might catch some hints or even reliable clues for further medical research like drug or vaccine development. Based on this knowledge, single cell bioinformatics studies in COVID-19 have already generated several meaningful discoveries.

## 3. COVID-19 Research Using Single Cell Transcriptome Profiles

### 3.1. Overview of ACE2 Expression throughout Human Body

Although the previous study has detected the expression of ACE2 in different human tissues by real-time PCR<sup>[5]</sup>, the exact cell types which express ACE2 and thus might be more vulnerable to the infection of COVID-19 could not be revealed by such kind of techniques. To overcome this gap, Zou *et al.*<sup>[6]</sup> have analyzed published scRNA-seq datasets for the expression of ACE2, thus have generated a whole-body map for the infection risk of COVID-19 in different cell types in humans. According to their results, respiratory track, heart, esophagus, ileum, kidney, and bladder are vulnerable to COVID-19 due to relatively high expression of ACE2 in respiratory epithelial cells, myocardial cells, esophagus epithelial cells, ileal epithelial cells, kidney proximal tubule cells and urothelial cells, respectively.

This study covers most of the systems with the major organs in the human body and reveals a rough overview of the vulnerability to the infection of COVID-19, however, this is merely a description of the vulnerability itself, with little mechanism behind it. In the following text, more studies about COVID-19 in different organs and tissues will be reviewed, in which more detailed findings in aspects of the injury that is caused to patients' health, the possible route of propagation and the putative mechanisms are discussed.

### 3.2. Further Exploration of ACE2 Expression in Different Organs or Tissues

#### 1) Lung

Zhao *et al.*<sup>[7]</sup> have analyzed the single cell transcription profiles of lungs from seven normal adult human donors, finding that only a small proportion of type II alveolar (AT2) cells exhibit the expression of ACE2, which is in concordance with the whole-body vulnerability study mentioned above<sup>[6]</sup>. For the further mechanism of the infection in these cells, gene ontology (GO) has been performed, showing that viral process related GO terms, including general positive regulation of viral process as well as viral replication and assembly, are highlighted in these ACE2-positive AT2 cells, which might be assistance for COVID-19 to exploit the functional machinery in these AT2 cells for the viral proliferation, further infection and spread into other parts of the human body. Some

severe infection cases could even induce sepsis and acute respiratory syndrome, causing a fatality rate of 15% approximately.

The fortes of single cell analysis in resolution and specification are prominent in these results. Interestingly, another conclusion in this study seem to be even more attractive to public attention: Males might be more vulnerable than females due to a higher ratio of ACE2-expressing cells. Considering the small number of individuals (2 men, 6 women) used in the study, this may need more evidence to finally draw a solid conclusion.

## 2) Respiratory Tract

The gate of respiratory system as it is, the upper respiratory tract is the first to bear the brunt during the virus making its way into lungs. Wu *et al.*<sup>[8]</sup> have investigated the ACE2 expression of different parts of the respiratory tract including nasal, bronchial, turbinate and lungs. The scRNA-seq data analysis indicates that the nasal epithelial cells, in which the ACE2-expressing cell population is comparable to AT2, could be a putative host of COVID-19, which is consistent with the PCR result in which the concentration of the virus is higher in nasal-swab than that in the throat-swab.

## 3) Digestive Tract

As COVID-19 patients exhibit diarrhea and nausea along with typical symptoms of pneumonia, Zhang *et al.*<sup>[9]</sup> have analyzed the expression of ACE2 in the scRNA-seq profiles of the human digestive system, finding that the expression of ACE2 is at a relatively high level in the absorptive enterocytes of ileum, upper and stratified epithelial cells of the esophagus, as well as the enterocytes in the colon, indicating that the COVID-19 virus could also invade the human body through the digestive tract.

## 4) Testis and Kidney

Since COVID-19 patients exhibit kidney damage after infection, the expression of ACE2 in the kidney has been investigated, showing that renal tubular cells have ACE2 expressed at a relatively high level. The authors of the same study also have examined the expression of ACE2 in testis, finding a high level of expression in Leydig cells and seminiferous duct cells, which indicates potential damage to fertility in patients<sup>[10]</sup>. These results imply that special attention should be paid to the risk of testicular lesions in patients during clinical work, and also some follow-up might be needed.

# 4. Discussion

Among all kinds of studies of COVID-19, scRNA-seq related studies are highly conspicuous due to its detailed resolution to shed light on the possible mechanism of the disease and potential targets for the development of medical treatment. The results of these studies also inform not only medical staff but also the public of the possible route of infection and propagation, the injure that the infection could do to the human body, as well as the potential sequelae of prognosis.

One of the major drawbacks of these studies is that the samples are too small. There is still no scRNA-seq data from COVID-19 patients, let alone collecting and running scRNA-seq with samples from patients at different stages of the infection during this hectic time fighting with the pestilence. However, still, we have numerous publicly available datasets from databases to feed the data analyzing workflows in the studies mentioned above and to draw conclusions that could be more solid. Take the research using the respiratory tract data<sup>[7]</sup> for instance, the sample of eight individuals is far less from drawing a statistically significant conclusion, thus causing over-interpretation of the sample and the analysis results to some extent.

Furthermore, the data analysis should not be restricted to the commonly-used workflow for single cell data analysis, like what was used in the papers mentioned above. At the moment, the mining of public single cell

transcriptome profiling data is obviously informative, however, most of them stop at some preliminary results and the interpretations are restricted to the canonical paradigm which is mainly the identification of vulnerable cell types. Actually, there are plenty of data analyzing tools dedicated to single cell RNA profiling far more than what has been adopted in existing research, therefore, besides cell type identification, functional representation and interpretation are also feasible<sup>[11],[12]</sup>. Further studies that incorporate more functional analysis for a thorough understanding of the mechanism of the disease would be undoubtedly more helpful for humans to win the battle to COVID-19 as well as other rampant infectious epidemics.

## References

- [1] J. F. W. Chan, K. H. Kok, Z. Zhu, H. Chu, K. K. W. To, S.-F. Yuan, *et al*, "Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan," *Emerging Microbes & Infections*, vol. 9, no. 1, pp. 221-236, Jan. 2010.
- [2] X.-T. Xu, P. Chen, J.-F. Wang, J.-N. Feng, H. Zhou, X. Li, *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*, vol. 63, no. 3, pp. 456-460, Jan. 2020.
- [3] M. Donoghue, F. Hsieh, E. Baronas, K. Godbout, M. Gosselin, N. Stagliano, *et al*, "A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9," *Circulation Research*, vol. 87, no. 5, pp. E1-E9, Sept. 2000.
- [4] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, *et al*, "A new coronavirus associated with human respiratory disease in China," *Nature*, vol. 579, no. 7798, pp. 256-269, Feb. 2020.
- [5] I. Hamming, W. Timens, M. L. C. Bulthuis, A. T. Lely, G. J. Navis, and H. van Goor, "Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis," *The Journal of Pathology*, vol. 203, no. 2, pp. 631-637, Jun. 2004.
- [6] X. Zou, K. Chen, J.-W. Zou, P.-Y. Han, J. Hao, and Z.-G. Han, "Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection," *Frontiers of Medicine*, 2020.
- [7] Y. Zhao, Z.-X. Zhao, Y.-J. Wang, Y.-Q. Zhou, Y. Ma, and W. Zuo, "Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCoV," *bioRxiv*, Jan. 2020. DOI: [10.1101/2020.01.26.919985](https://doi.org/10.1101/2020.01.26.919985). (请核对文献类型)
- [8] C. Wu, S.-F. Zheng, Y. Chen, and M. Zheng, "Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCoV, in the nasal tissue," *medRxiv*, Feb. 2020. DOI: [10.1101/2020.02.11.20022228](https://doi.org/10.1101/2020.02.11.20022228). (请核对文献类型)
- [9] H. Zhang, Z.-J. Kang, H.-Y. Gong, D. Xu, J. Wang, Z.-F. Li, *et al*, "The digestive system is a potential route of 2019-nCoV infection: a bioinformatics analysis based on single-cell transcriptomes," *bioRxiv*, Jan. 2020. DOI: [10.1101/2020.01.30.927806](https://doi.org/10.1101/2020.01.30.927806). (请核对文献类型)
- [10] C.-B. Fan, K. Li, Y.-H. Ding, W.-L. Lu, and J.-Q. Wang, "ACE2 expression in kidney and testis may cause kidney and testis damage after 2019-nCoV infection," *medRxiv*, Feb. 2020. DOI: [10.1101/2020.02.12.20022418](https://doi.org/10.1101/2020.02.12.20022418).
- [11] B. Duan, C. Zhou, C.-Y. Zhu, Y.-F. Yu, G.-Y. Li, S.-H. Zhang, *et al*, "Model-based understanding of single-cell CRISPR screening," *Nature Communications*, vol. 10, no. 1, p. 2233, May 2019.
- [12] D. DeTomaso, M. G. Jones, M. Subramaniam, T. Ashuach, C. J. Ye, and N. Yosef, "Functional interpretation of single cell similarity maps," *Nature Communications*, vol. 10, no. 1, p. 4386, Sept. 2019.



**Zi-Wei Wang** was born in Shandong in 1988. She received the B.S. degree in biology from Shandong University (SDU), Jinan in 2011 and the Ph.D. degree in cell biology from Peking University (PKU), Beijing in 2017. Her research interests include molecular biology and bioinformatics, especially those related to single cell RNA sequencing or single cell multi-omics data analysis



**Chi-Chang Chang** received his PhD degree from Yuan Ze University in 2006. He is the Director of Medical Decision Science Laboratory, Chun Shan Medical University. His research interests include multiple primary malignant neoplasms, recurrent gynecologic cancer, shared medical decision making, decision analysis and simulation modeling, performance measurements in health services, health service innovation and management, and ethical aspects in health care



**Quan Zou** received his Ph.D. degree from Harbin Institute of Technology in 2009. He worked with Xiamen University and Tianjin University from 2009 to 2018. Currently, he is a professor with the Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology (UESTC), Chengdu. His research interests include bioinformatics, machine learning and parallel computing. Now he is putting the focus on protein classification, genome assembly, annotation and functional analysis from the next generation sequencing data with parallel computing methods