

# Single cell analysis of ACE2 expression reveals the potential targets for 2019-nCoV

Yanyan Zhu<sup>1,a</sup>, Miaomiao Jiang<sup>2,a</sup>, Liang Gao<sup>3,\*</sup>, Xiaoyun Huang<sup>2,\*</sup>

<sup>1</sup> Department of Oncology, Henan Provincial People's Hospital; Zhengzhou University People's Hospital, Zhengzhou, Henan, 450003, China

<sup>2</sup> Zhiyu Center for Systems Biology, Zhiyu Inc., Shenzhen, China

<sup>3</sup> The Brain Cognition and Brain Disease Institute, ShenZhen Institutes of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Avenue, Shenzhen 518055, China

<sup>a</sup>Those authors contribute equally.

\*Corresponding author: liang.gao@siat.ac.cn (Liang Gao) and stemidea@gmail.com (Xiaoyun Huang)

## Abstract

ACE2, the putative receptor for the novel coronavirus (2019-nCoV), played an important role in cell entry of 2019-nCoV. However, it is not yet clear what cell types within the human body express ACE2. Here, a systematic analysis was undertaken using published single cell datasets. In total, our study analyzed 229652 cells, from five different organs, derived from 88 donors. The top ACE2 expressing cells include proximal tubule cells in the kidney and enterocytes in the intestine. Other major ACE2 expressing cells in the kidney include podocytes, intercalated cells and endothelial cells. Our results offer a comprehensive atlas of ACE2 expression at the single cell level and unravel the enormous potential targets of 2019-nCoV infection beyond the lung.

## Introduction

Recently, a novel coronavirus (2019-nCoV) outbreak has claimed around 40261 cases of infection and 909 cases of deaths until the preparation of our manuscript. Spike proteins of the SARS coronaviruses associate with angiotensin-converting enzyme 2 (ACE2) to mediate cell entry[1]. The putative progenitor virus of SARS coronavirus was isolated from bat fecal samples and was reported to harness ACE2 from human, civet and Chinese horseshoe bat for cell entry[2]. Analysis of the sequence of 2019-nCoV receptor-binding domain (RBD) revealed that 2019-nCoV and SARS shared sequence similarity in RBD and ACE2 might function as the receptor for cell entry by 2019-nCoV[3]. Isolation and co-localization analysis also suggested 2019-nCoV used ACE2 for cell entry[4]. However, a systematic analysis of ACE2 expression in different organs has not yet been performed.

In the past decade, we have witnessed the transformation of biomedical sciences by single cell analysis[5, 6]. One notable example is the application of single cell RNA-seq to revisit the century old questions in developmental biology and cancer biology. To understand the cellular composition of the human body at the single cell level, the Human Cell Atlas Project was initiated in Oct. of 2016[7]. Numerous single cell experiments have been performed to profile cells from different organs of the human body, and those massive datasets were made available publicly. Thus, the time is ripe now to ask what cell types express a gene of interest and to what extent this gene is expressed.

Our study aims to undertake a comprehensive analysis of ACE2 expression at the single cell level using publicly available datasets and shed light on the potential mechanisms for the cell entry and pathogenicity of 2019-nCoV. In total, our study analyzed 229652 single cells, from five different organs, derived from 88 donors. Our results offer a comprehensive atlas of ACE2 expression at the single cell level and unravel the enormous potential targets of 2019-nCoV infection beyond the lung.

## Result

First, ACE2 expression was analyzed in the lung, the major organ affected by the 2019-nCoV. 57020 cells from five donors were used for the analysis[8]. Three compartments were recovered in the dataset, including epithelial, immune and stromal compartments

(**Figure 1A**). ACE2 was mainly expressed in the epithelial compartment, with around 1.7% Type II alveolar cells expressing ACE2 (**Figure 1B**). Very few fibroblasts (13/4624) also expressed ACE2 (**Supplementary Figure 1**).

The expression of ACE2 was investigated in the liver using a dataset generated with mCEL-Seq2[9]. In total, eight distinct cell types were recovered, including hepatocyte, cholangiocyte, B cell, NK+T cell, Kupffer cell, monocyte, hepatic stellate cell, liver macrovascular endothelial cell (LVEC) and liver sinusoidal endothelial cell (LSEC). All three distinct compartments were represented: epithelial, immune and stromal compartment (**Figure 2A**). ACE2 was detected in less than 1% of hepatocytes and cholangiocytes (**Figure 2B**).

The intestine was an important organ in the digestive system[10]. One dataset consisting of cells from ileum, colon and rectum was first employed for analysis. In total, 14537 cells were detected, including TA cell, stem cell, progenitor cell, Paneth like cell, goblet cell, enterocyte and Enteroendocrine cell (**Figure 2C**). The cell annotation from the original study was used. ACE2 was expressed in enterocytes, Paneth like cells, Enteroendocrine cells and goblet cells. Using publicly available data at the Human Protein Atlas[11, 12], the expression of ACE2 at the protein level was confirmed (**Supplementary Figure 4**).

To obtain a glimpse of ACE2 expression in the central nervous system, a dataset of single cell transcriptomes derived from human prefrontal cortex was used[13]. In total, 80660 cells from 48 donors with varying degrees of AD pathology were analyzed. The cell types were annotated with classical cell markers (**Figure 3A**). Eight cell types were identified: Excitatory neuron, inhibitory neuron, oligodendrocyte progenitor, oligodendrocyte, microglia, astrocyte, pericyte, endothelial cell. ACE2 was detected in around 0.2% of cells in the prefrontal cortex (**Figure 3B, 3C**).

Next the expression of ACE2 was analyzed in the kidney using a recently generated human kidney cell atlas[14]. In total, 40268 cells from mature kidneys and 27203 cells from fetal kidneys were included in the atlas (**Figure 4**). The cell types were annotated as in the original study. The clustering results and the expression of ACE2 were visualized on a UMAP dimensionality reduced plots (**Figure 4A, 4C**). Surprisingly, ACE2 expression was detected in 8.445% of proximal tubule cells in the mature kidney and 42.1499% of proximal tubule cells in the fetal kidney (**Figure 4B, 4D**). Other cell types with similar expression percentage as compared to Type II alveolar cells included podocyte, intercalated cells, pelvic epithelium, peritubular capillary endothelium 2 and descending vasa recta endothelium (**Supplementary Figure 2**). In the fetal kidney, medial S shaped body contained around 10% of ACE2 expressing cells (**Supplementary Figure 3**). Interestingly, the expression of ACE2 at the protein level was confirmed with publicly available data at the Human Protein Atlas (**Supplementary Figure 5**).

We then asked what distinguished ACE2 positive cells from ACE2 negative cells. The human kidney cell atlas data was employed in the analysis as ACE2 positive cells constituted a significant portion of the total population in both mature and fetal kidneys. Differentially expressed genes were identified. The top genes significantly upregulated in ACE2 positive cells in the mature kidney were enriched for regulated exocytosis, protein processing in endoplasmic reticulum (**Figure 5A**). The protein-protein interaction network

was constructed and MCODE components were identified, revealing the potential biological functions enriched: antigen processing and presentation of endogenous antigen, transport to the Golgi and subsequent modification, response to ER stress and protein folding (**Figure 5B, 5C**). Those results revealed the potential mechanisms for coronavirus to hijack ACE2 positive cells for viral infection and propagation.

## Methods

### Datasets

Lung cell atlas dataset was available through the Human Cell Atlas Data Coordination Platform and NCBI BIOPROJECT (accession code PRJEB31843). Liver cell atlas dataset was downloaded from GEO (GSE124395). Precortex cell atlas was downloaded from Synapse (syn18485175). The intestine cell atlas dataset was downloaded from GEO (GSE125970). The kidney cell atlas dataset was downloaded from the data portal of the Human Cell Atlas project ( www.kidneycellatlas.org.).

### Data pre-processing

Seurat and Scanpy were used to analyze the single cell datasets. In general, a quality control step was undertaken to remove low-quality cells with minimal number of genes detected, maximal number of genes detected, minimal number of cells in which the gene was detected. Second, a normalization step was performed to scale the gene-cell matrix. The top 2000 highly variable genes were identified by the FindVariableGenes function.

### Cell clustering and visualization

Seurat and Scanpy were used to analyze the single cell datasets. In general, a quality control step was undertaken to remove low-quality cells with minimal number of genes detected, maximal number of genes detected, minimal number of cells in which the gene was detected. Second, a normalization step was performed to scale the gene-cell matrix. In most cases, the original cell type annotations were used as previously published.

### Analysis of differentially expressed genes

Differentially expressed genes were identified by wilcoxon test and  $p$ -value was corrected with benjamini-hochberg (BH) method.  $\log_2$ foldchange was calculated with this formula:  $\log_2(e^{(mean1+1e-9)} / e^{(mean2+1e-9)})$ . Top significantly differentially expressed genes were filtered by BH corrected  $p$ -value  $< 0.05$  and  $|\log_2FC| > 0.5$ .

### Gene list analysis

Gene list analysis was performed with metascape[15]. Differentially expressed gene list was input into the online tool and analyzed using default parameters. Statistically enriched terms were identified and filtered. Remaining significant terms were then hierarchically clustered into a tree based on Kappa-statistical similarities among their gene memberships.

### Protein-protein interaction network

All protein-protein interactions (PPI) among the input gene list were extracted from PPI databases and formed a PPI network. MCODE components were identified from the merged network. Each MCODE network is assigned a unique color. GO enrichment analysis was applied to the original PPI network and its MCODE network components to assign biological “meanings”, where top three best  $p$ -value terms were retained. All

input gene lists were also merged into one list and resulted in a PPI network.

## Discussion

To our knowledge, this study is the first comprehensive profiling of ACE2 expression in the human body across different systems. The expression of ACE2 was analyzed in the respiratory system, the digestive system, the urinary system, the nervous system. Our result suggested that various cell types expressed ACE2 at different level, rendering them potential targets of 2019-nCoV infection. Our study further extends previous attempts to analyze ACE2 expression at the single cell level using publicly available datasets[16, 17].

Besides fever and cough, less common symptoms of 2019-nCoV associated pneumonia also included sputum production, headache, haemoptysis and diarrhoea[18]. The first patient diagnosed with 2019-nCoV infection in the United States presented dry cough, nausea and vomiting, loose stool very early in the disease progression, suggesting that the digestive system might be affected directly by viral infection[19]. The finding that ACE2 is highly expressed in several cell types within the kidney suggests the kidney might also be an important target of the 2019-nCoV infection.

Cell-to-cell variability is a common biological phenomenon. On one hand, single cell analysis enabled an unprecedented resolution for us to analyze gene expression and thus revealed information that could be potentially masked by population averaged measurements. On the other hand, viral infection might be influenced by population context, leading to a more complex picture of viral infection in vivo[20].

The infection of 2019-nCoV might be a lot more complex. It is not yet clear whether other receptors exist to enable the cell entry for 2019-nCoV. Furthermore, the expression of ACE2 might not be sufficient enough to drive cell entry of 2019-nCoV. Studies have shown that the cooperation of other proteins on the cell membrane might significantly enhance viral infection.

Patient-to-patient variability was observed in the pneumonia associated with 2019-nCoV. Around 10% of patients presented as critically ill patients, with failures in multiple organs. It is yet to be determined whether organ failures are also associated with viral infections of affected organs.

Our study served as a proof-of-concept that mining the publicly available datasets generated by the Human Cell Atlas is a powerful means to deepen our understanding of the human body at both physiological and pathological circumstances.

## References

- [1] W. Li, M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, H. Choe, and M. Farzan. Angiotensin-converting enzyme 2 is a functional receptor for the sars coronavirus. *Nature*, 426(6965):450–4, 2003.
- [2] X. Y. Ge, J. L. Li, X. L. Yang, A. A. Chmura, G. Zhu, J. H. Epstein, J. K. Mazet, B. Hu, W. Zhang, C. Peng, Y. J. Zhang, C. M. Luo, B. Tan, N. Wang, Y. Zhu, G. Cramer, S. Y. Zhang, L. F. Wang, P. Daszak, and Z. L. Shi. Isolation and characterization of a bat sars-like coronavirus that uses the ace2 receptor. *Nature*, 503(7477):535–8, 2013.
- [3] Y. Wan, J. Shang, R. Graham, R. S. Baric, and F. Li. Receptor recognition by novel coronavirus from wuhan: An analysis based on decade-long structural studies of sars. *J Virol*, 2020.
- [4] Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, Hui-Dong Chen, Jing Chen, Yun Luo, Hua Guo, Ren-Di Jiang, Mei-Qin Liu, Ying Chen, Xu-Rui Shen, Xi Wang, Xiao-Shuang Zheng, Kai Zhao, Quan-Jiao Chen, Fei Deng, Lin-Lin Liu, Bing Yan, Fa-Xian Zhan, Yan-Yi Wang, Geng-Fu Xiao, and Zheng-Li Shi. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 2020.
- [5] A. Haque, J. Engel, S. A. Teichmann, and T. Lonnberg. A practical guide to single-cell rna-sequencing for biomedical research and clinical applications. *Genome Med*, 9(1):75, 2017.
- [6] A. Tanay and A. Regev. Scaling single-cell genomics from phenomenology to mechanism. *Nature*, 541(7637):331–338, 2017.
- [7] A. Regev, S. A. Teichmann, E. S. Lander, I. Amit, C. Benoist, E. Birney, B. Bodenmiller, P. Campbell, P. Carninci, M. Clatworthy, H. Clevers, B. Deplancke, I. Dunham, J. Eberwine, R. Eils, W. Enard, A. Farmer, L. Fugger, B. Gottgens, N. Hacohen, M. Haniffa, M. Hemberg, S. Kim, P. Klenerman, A. Kriegstein, E. Lein, S. Linnarsson, E. Lundberg, J. Lundberg, P. Majumder, J. C. Marioni, M. Merad, M. Mhlanga, M. Nawijn, M. Netea, G. Nolan, D. Pe’er, A. Phillipakis, C. P. Ponting, S. Quake, W. Reik, O. Rozenblatt-Rosen, J. Sanes, R. Satija, T. N. Schumacher, A. Shalek, E. Shapiro, P. Sharma, J. W. Shin, O. Stegle, M. Stratton, M. J. T. Stubbington, F. J. Theis, M. Uhlen, A. van Oudenaarden, A. Wagner, F. Watt, J. Weissman, B. Wold, R. Xavier, N. Yosef, and Participants Human Cell Atlas Meeting. The human cell atlas. *Elife*, 6, 2017.
- [8] E. Madissoon, A. Wilbrey-Clark, R. J. Miragaia, K. Saeb-Parsy, K. T. Mahbubani, N. Georgakopoulos, P. Harding, K. Polanski, N. Huang, K. Nowicki-Osuch, R. C. Fitzgerald, K. W. Loudon, J. R. Ferdinand, M. R. Clatworthy, A. Tsingene, S. van Dongen, M. Dabrowska, M. Patel, M. J. T. Stubbington, S. A. Teichmann, O. Stegle,

and K. B. Meyer. scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biol*, 21(1):1, 2019.

- [9] N. Aizarani, A. Saviano, Sagar, L. Mailly, S. Durand, J. S. Herman, P. Pessaux, T. F. Baumert, and D. Grun. A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature*, 572(7768):199–204, 2019.
- [10] Y. Wang, W. Song, J. Wang, T. Wang, X. Xiong, Z. Qi, W. Fu, X. Yang, and Y. G. Chen. Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. *J Exp Med*, 217(2), 2020.
- [11] M. Uhlen, L. Fagerberg, B. M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. Szigartyo, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P. H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, and F. Ponten. Proteomics. tissue-based map of the human proteome. *Science*, 347(6220):1260419, 2015.
- [12] M. Uhlen, P. Oksvold, L. Fagerberg, E. Lundberg, K. Jonasson, M. Forsberg, M. Zwahlen, C. Kampf, K. Wester, S. Hober, H. Wernerus, L. Bjorling, and F. Ponten. Towards a knowledge-based human protein atlas. *Nat Biotechnol*, 28(12):1248–50, 2010.
- [13] H. Mathys, J. Davila-Velderrain, Z. Peng, F. Gao, S. Mohammadi, J. Z. Young, M. Menon, L. He, F. Abdurrob, X. Jiang, A. J. Martorell, R. M. Ransohoff, B. P. Hafler, D. A. Bennett, M. Kellis, and L. H. Tsai. Single-cell transcriptomic analysis of alzheimer’s disease. *Nature*, 570(7761):332–337, 2019.
- [14] B. J. Stewart, J. R. Ferdinand, M. D. Young, T. J. Mitchell, K. W. Loudon, A. M. Riding, N. Richoz, G. L. Frazer, J. U. L. Staniforth, F. A. Vieira Braga, R. A. Botting, D. M. Popescu, R. Vento-Tormo, E. Stephenson, A. Cagan, S. J. Farndon, K. Polanski, M. Efremova, K. Green, M. Del Castillo Velasco-Herrera, C. Guzzo, G. Collord, L. Mamanova, T. Aho, J. N. Armitage, A. C. P. Riddick, I. Mushtaq, S. Farrell, D. Rampling, J. Nicholson, A. Filby, J. Burge, S. Lisgo, S. Lindsay, M. Bajenoff, A. Y. Warren, G. D. Stewart, N. Sebire, N. Coleman, M. Haniffa, S. A. Teichmann, S. Behjati, and M. R. Clatworthy. Spatiotemporal immune zonation of the human kidney. *Science*, 365(6460):1461–1466, 2019.
- [15] Y. Zhou, B. Zhou, L. Pache, M. Chang, A. H. Khodabakhshi, O. Tanaseichuk, C. Benner, and S. K. Chanda. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*, 10(1):1523, 2019.
- [16] Yu Zhao, Zixian Zhao, Yujia Wang, Yueqing Zhou, Yu Ma, and Wei Zuo. Single-cell rna expression profiling of ace2, the putative receptor of wuhan 2019-ncov. *bioRxiv*, 2020.
- [17] Hao Zhang, Zijian Kang, Haiyi Gong, Da Xu, Jing Wang, Zifu Li, Xingang Cui, Jianru Xiao, Tong Meng, Wang Zhou, Jianmin Liu, and Huji Xu. The digestive

system is a potential route of 2019-ncov infection: a bioinformatics analysis based on single-cell transcriptomes. *bioRxiv*, 2020.

- [18] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, and B. Cao. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *Lancet*, 2020.
- [19] M. L. Holshue, C. DeBolt, S. Lindquist, K. H. Lofy, J. Wiesman, H. Bruce, C. Spitters, K. Ericson, S. Wilkerson, A. Tural, G. Diaz, A. Cohn, L. Fox, A. Patel, S. I. Gerber, L. Kim, S. Tong, X. Lu, S. Lindstrom, M. A. Pallansch, W. C. Weldon, H. M. Biggs, T. M. Uyeki, S. K. Pillai, and V. Case Investigation Team Washington State nCo. First case of 2019 novel coronavirus in the united states. *N Engl J Med*, 2020.
- [20] B. Snijder, R. Sacher, P. Ramo, E. M. Damm, P. Liberali, and L. Pelkmans. Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature*, 461(7263):520–3, 2009.



## Figure legends

### Figure 1

ACE2 expression in the lung at the single cell level. (A) 57020 cells were visualized on UMAP plot. (B) Dotplot visualization of ACE2 expression in the annotated cell types (lung).

### Figure 2

ACE2 expression in the liver and the intestine at the single cell level. (A) 9964 cells were visualized on UMAP plot. Annotated cell types including Cholangiocyte, NK+T, B cell, Monocyte, Kupffer cell, liver sinusoidal endothelial cell (LSEC), liver macrovascular endothelial cell (LVEC), plasma cell, hepatocyte and hepatic stellate cell. (B) Dotplot visualization of ACE2 expression in the annotated cell types (liver). (C) 14537 cells were visualized on UMAP plot. Annotated cell types including Enteroendocrine, Enterocyte, Goblet, Paneth like, Progenitor, Stem cell and Trans-amplifying (TA). (D) Dotplot visualization of ACE2 expression in the annotated cell types (intestine).

### Figure 3

ACE2 expression in the prefrontal cortex at the single cell level. (A) 80660 cells from prefrontal human cortex were visualized on UMAP plot. (B) Feature plot visualization of ACE2 expression in the annotated cell types in the prefrontal human cortex. (C) Violin plot visualization of ACE2 expression in the annotated cell types in the prefrontal human cortex.

### Figure 4

ACE2 expression in the kidney at the single cell level. (A) 40268 cells from mature human kidney were visualized on UMAP plot. (B) Dotplot visualization of ACE2 expression in the annotated cell types in the mature human kidney. (C) 27203 cells from the fetal human kidney were visualized on UMAP plot. (D) Dotplot visualization of ACE2 expression in the annotated cell types in the fetal human kidney.

### Figure 5

Biological processes and network components associated with ACE2 expression. (A) The top 20 clusters enriched in ACE2 positive cells. (B) MCODE components identified in protein-protein interaction network derived from the top genes significantly upregulated in ACE2 positive cells. (C) The top 5 MCODE components identified.



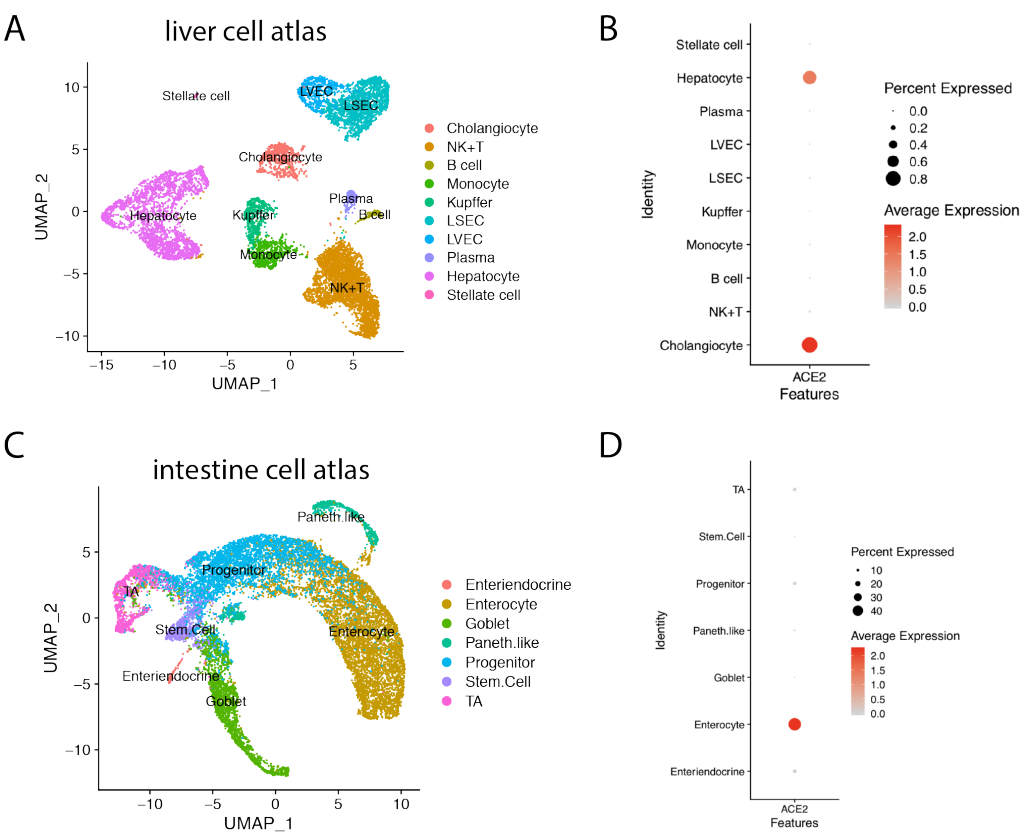


Figure 2: ACE2 expression in the liver and the intestine at the single cell level. (A) 9964 cells were visualized on UMAP plot. Annotated cell types including Cholangiocyte, NK+T, B cell, Monocyte, Kupffer cell, liver sinusoidal endothelial cell (LSEC), liver macrovascular endothelial cell (LVEC), plasma cell, hepatocyte and hepatic stellate cell. (B) Dotplot visualization of ACE2 expression in the annotated cell types (liver). (C) 14537 cells were visualized on UMAP plot. Annotated cell types including Enteriendocrine, Enterocyte, Goblet, Paneth like, Progenitor, Stem cell and Trans-amplifying (TA). (D) Dotplot visualization of ACE2 expression in the annotated cell types (intestine).

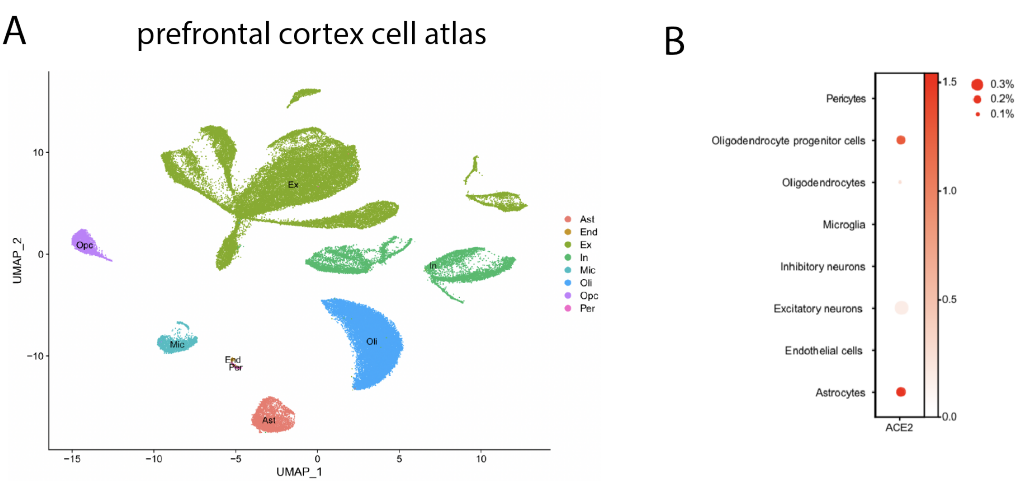


Figure 3: ACE2 expression in the prefrontal cortex at the single cell level. (A) 80660 cells from prefrontal human cortex were visualized on UMAP plot. (B) Feature plot visualization of ACE2 expression in the annotated cell types in the prefrontal human cortex. (C) Violin plot visualization of ACE2 expression in the annotated cell types in the prefrontal human cortex.

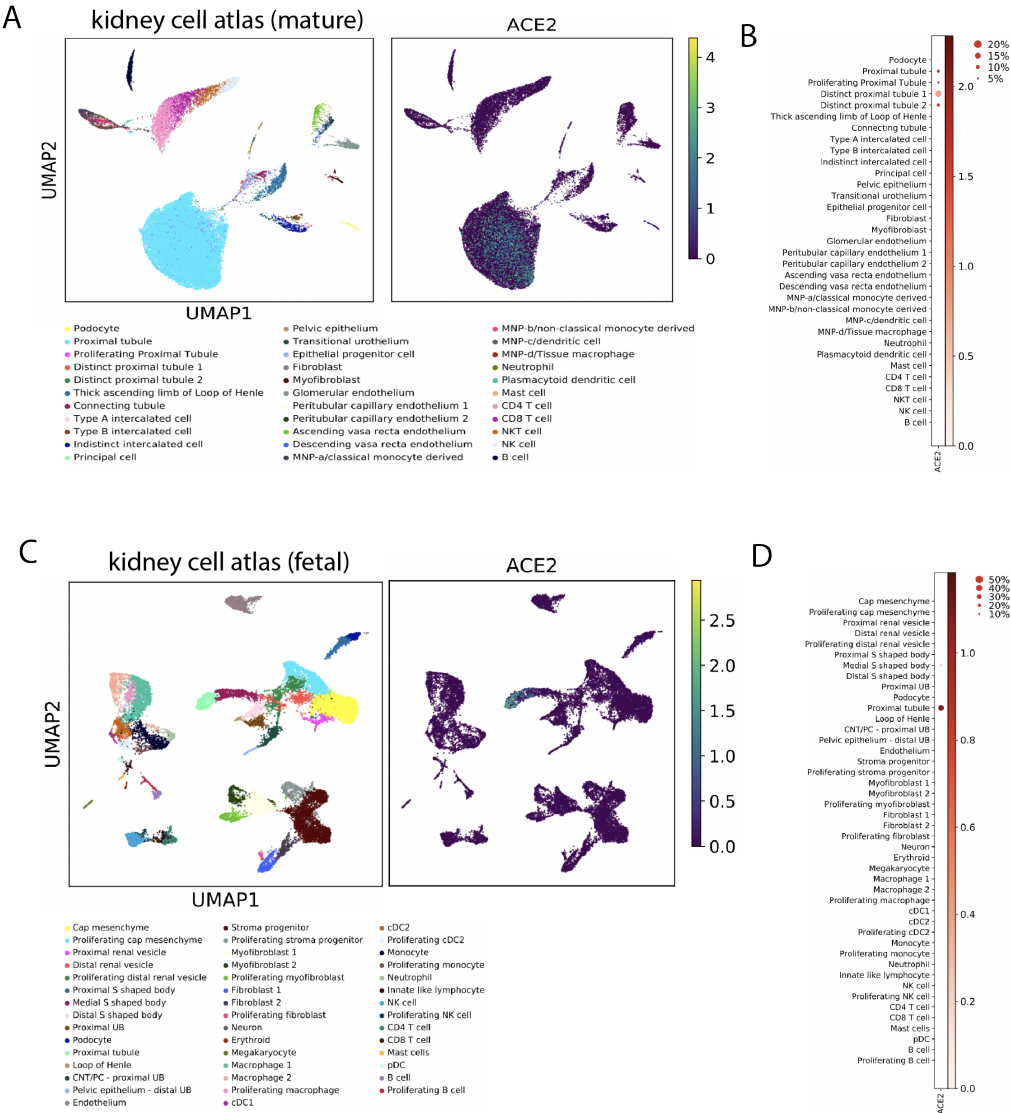


Figure 4: ACE2 expression in the kidney at the single cell level. (A) 40268 cells from mature human kidney were visualized on UMAP plot. (B) Dotplot visualization of ACE2 expression in the annotated cell types in the mature human kidney. (C) 27203 cells from the fetal human kidney were visualized on UMAP plot. (D) Dotplot visualization of ACE2 expression in the annotated cell types in the fetal human kidney.

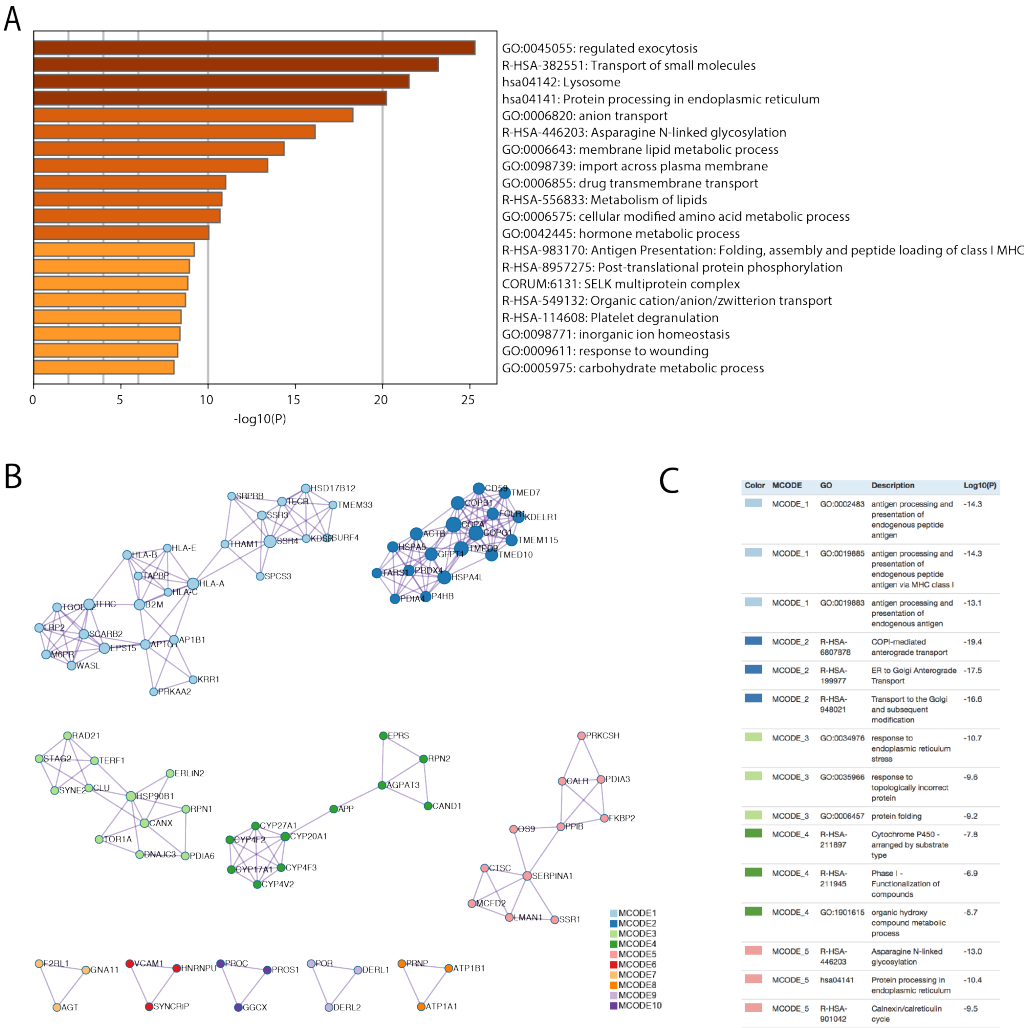


Figure 5: Biological processes and network components associated with ACE2 expression. (A) The top 20 clusters enriched in ACE2 positive cells. (B) MCODE components identified in protein-protein interaction network derived from the top genes significantly upregulated in ACE2 positive cells. (C) The top 5 MCODE components identified.