

1 **The digestive system is a potential route of 2019-nCov infection: a bioinformatics**  
2 **analysis based on single-cell transcriptomes**

3 Hao Zhang<sup>1,4,9</sup>†, Zijian Kang<sup>1,9</sup>†, Haiyi Gong<sup>4,9</sup>†, Da Xu<sup>6,9</sup>†, Jing Wang<sup>5</sup>, Zifu Li<sup>5</sup>,  
4 Xingang Cui<sup>6</sup>, Jianru Xiao<sup>4</sup>, Tong Meng<sup>7,8,9</sup>\*, Wang Zhou<sup>2,9</sup>\*, Jianmin Liu<sup>5</sup>\*, Huji  
5 Xu<sup>1,2,3</sup>\*

6 <sup>1</sup> Department of Rheumatology and Immunology, Changzheng Hospital, Second  
7 Military Medical University, 200003 Shanghai, China

8 <sup>2</sup> Peking-Tsinghua Center for Life Sciences, Tsinghua University, Beijing, P.R. China

9 <sup>3</sup> Beijing Tsinghua Changgeng Hospital, School of Clinical Medicine, Tsinghua  
10 University, 100084 Beijing, China

11 <sup>4</sup> Department of Orthopaedic Oncology, Changzheng Hospital, Second Military  
12 Medical University, 200003 Shanghai, China

13 <sup>5</sup> Department of Neurosurgery, Changhai Hospital, Second Military Medical  
14 University, 200003 Shanghai, China

15 <sup>6</sup> Department of Urology, The Third Affiliated Hospital of Second Military Medical  
16 University, 201805 Shanghai, China

17 <sup>7</sup> Division of Spine, Department of Orthopedics, Tongji Hospital affiliated to Tongji  
18 University School of Medicine, 200065 Shanghai, China

19 <sup>8</sup> Tongji University Cancer Center, School of Medicine, Tongji University, 200092  
20 Shanghai, China

21 <sup>9</sup> Qiu-Jiang Bioinformatics Institute, 200003 Shanghai, China

22

23 \*Correspondence to: [huji\\_xu@tsinghua.edu.cn](mailto:huji_xu@tsinghua.edu.cn)

24 [chstroke@163.com](mailto:chstroke@163.com)

25 [brilliant212@163.com](mailto:brilliant212@163.com)

26 [mengtong@medmail.com.cn](mailto:mengtong@medmail.com.cn)

27

28

1 **Abstract**

2 Since December 2019, a newly identified coronavirus (2019 novel coronavirus,  
3 2019-nCov) is causing outbreak of pneumonia in one of largest cities, Wuhan, in  
4 Hubei province of China and has draw significant public health attention. The same as  
5 severe acute respiratory syndrome coronavirus (SARS-CoV), 2019-nCov enters into  
6 host cells via cell receptor angiotensin converting enzyme II (ACE2). In order to  
7 dissect the ACE2-expressing cell composition and proportion and explore a potential  
8 route of the 2019-nCov infection in digestive system infection, 4 datasets with  
9 single-cell transcriptomes of lung, esophagus, gastric, ileum and colon were analyzed.  
10 The data showed that ACE2 was not only highly expressed in the lung AT2 cells,  
11 esophagus upper and stratified epithelial cells but also in absorptive enterocytes from  
12 ileum and colon. These results indicated along with respiratory systems, digestive  
13 system is a potential routes for 2019-nCov infection. In conclusion, this study has  
14 provided the bioinformatics evidence of the potential route for infection of 2019-nCov  
15 in digestive system along with respiratory tract and may have significant impact for  
16 our healthy policy setting regards to prevention of 2019-nCoV infection.

17 **Introduction**

18 At the end of 2019, a rising number of pneumonia patients with unknown pathogen  
19 has been emerging in one of largest cities of China,Wuhan, and quickly spread  
20 throughout whole country[1]. A novel coronavirus was then isolated from the human  
21 airway epithelial cells and was named 2019 novel coronavirus (2019-nCoV)[2]. The  
22 complete genome sequences has reveled that 2019-nCoV sharing 86.9% nucleotide  
23 sequence identity to a severe acute respiratory syndrome (SARS)-like coronavirus  
24 detected in bats (bat-SL-CoVZC45, MG772933.1). This suggested that 2019-nCoV is  
25 the species of SARS related coronaviruses (SARSr-CoV) by pairwise protein  
26 sequence analysis[2, 3].

27 As for the clinical manifestations of 2019-nCoV infection, fever and cough are most  
28 common symptoms at onset[4, 5]. In addition, it frequently induces severe enteric

1 symptoms, such as diarrhea and nausea, which are even graver than those of  
2 SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV)[6, 7].  
3 However, a little was known why and how the 2019-nCov induced enteric symptoms.  
4 In addition, it is unknown yet whether 2019-nCoV can be transmitted through the  
5 digestive tract besides respiratory tract[5].

6 The prerequisite of coronaviruses infection is its entrance into the host cell. During  
7 this process, the spike (S) glycoprotein recognizes host cell receptors and induces the  
8 fusion of viral and cellular membranes[8]. In 2019-nCoV infection, a  
9 metalloproteinase, angiotensin converting enzyme II (ACE2) is proved to be the cell  
10 receptor, the same as SARS-CoV infection[9-11]. 2019-nCoV can enter into  
11 ACE2-expressing cells, but not into cells without ACE2 or cells with other  
12 coronavirus receptors, such as aminopeptidase N and dipeptidyl peptidase[10]. Thus,  
13 ACE2 plays an vital role in the 2019-nCoV infection.

14 In order to explore the infection routes of 2019-nCov and the roles of ACE2 in  
15 digestive system infection, we identified the ACE2-expressing cell composition and  
16 proportion in normal human lung and gastrointestinal system by single-cell  
17 transcriptomes based on the public databases. A striking finding is that ACE2 was not  
18 only expressed in lung AT2 cells, but also found in esophagus upper and stratified  
19 epithelial cells and absorptive enterocytes from ileum and colon. In addition, the  
20 enteric symptoms of 2019-nCov may be associated with the invaded  
21 ACE2-expressing enterocytes. These findings indicate that the digestive systems  
22 along with respiratory tract may be potential routes of 2019-nCov infection may have  
23 significant impact for our healthy policy setting regards to prevention of 2019-nCoV  
24 infection..

## 25 **Materials and Methods**

### 26 **Data Sources**

27 Single-cell expression matrices for the lung, esophagus, stomach, ileum and colon  
28 were obtained from the Gene Expression Omnibus (GEO;

1 <https://www.ncbi.nlm.nih.gov/>)[12], Single Cell Portal  
2 ([https://singlecell.broadinstitute.org/single\\_cell](https://singlecell.broadinstitute.org/single_cell)) and Human Cell Atlas Data Portal.  
3 (<https://data.humancellatlas.org>). Single-cell data for the esophagus and lung were  
4 obtained from the research by E Madisson et al which contained 6 esophageal and 5  
5 Lung tissue samples[13], which contained 6 esophageal and 5 lung tissue samples..  
6 The data of gastric mucosal samples from 3 non-atrophic gastritis and 3 chronic  
7 atrophic gastritis patients were obtained from GSE134520[14]. GSE134809[15] was  
8 comprised of 22 ileal specimens from 11 ileal Crohn’s disease patients and only  
9 non-inflammatory samples were selected for analysis. The research by Christopher S  
10 et al[16] included 12 normal colon samples.

### 11 **Quality Control**

12 Low quality Cells with expressed genes were lower than 200 or larger than 5000 were  
13 removed. We further required the percentage of UMIs mapped to mitochondrial or  
14 ribosomal genes to be lower than 20%.

### 15 **Data Integration, Dimension Reduction and Cell Clustering**

16 Different data processing methods were performed for different single-cell projects  
17 according to the downloaded data.

18 *Esophagus and lung datasets:* Seurat [17] rds data was directly download from  
19 supplementary material in the research by E. Madisson al [13]. Uniform Manifold  
20 Approximation and Projection (UMAP) visualization were performed for gaining  
21 clusters of cells.

22 *Stomach and ileum datasets:* Single cell data expression matrix was processed with  
23 the R package Seurat (version 3.0)[17]. We first utilized “NormalizeData” normalize  
24 and the single-cell gene expression data. UMI counts were normalized by the total  
25 number of UMIs per cell, multiplied 10000 for the normalization and were  
26 transformed to the log-transformed counts. The highly variable Genes (HVGs) were  
27 identified using the function “FindVariableGenes”. We then used  
28 “FindIntegrationAnchors” and “Integratdata” function to merge multiple sample data

1 within each dataset. After removing unwanted sources of variation from a single-cell  
2 dataset such as cell cycle stage, or mitochondrial contamination, we used the  
3 “RunPCA” function to perform the principle component analysis (PCA) on the  
4 single-cell expression matrix with significant HVGs. Then we constructed a  
5 K-nearest-neighbor graph based on the euclidean distance in PCA space using the  
6 “FindNeighbors” function and applied Louvain algorithm to iteratively group cells  
7 together by “FindClusters” function with optimal resolution. UMAP was used for  
8 visualization purposes.

9 *Colon Dataset* Single cell data expression matrix was processed with the R package  
10 LIGER[18] and Seurat[17]. We first normalized the data to account for differences in  
11 sequencing depth and capture efficiency among cells. Then we used “selectGenes”  
12 function to identify variable genes on each dataset separately and took the union of  
13 the result. Next integrative non-negative matrix factorization was performed to  
14 identify shared and distinct metagenes across the datasets and the corresponding  
15 factor loadings for each cell using “optimizeALS” function in LIGER. We selected a  
16 k of 15 and lambda of 5.0 get a plot of expected alignment. We then identified clusters  
17 shared across datasets and aligned quantiles within each cluster and factor using  
18 “quantileAlignSNF” function. Next nonlinear dimensionality reduction was  
19 performed using “RunUMAP” function in Seurat and the results were visualized with  
20 UMAP.

## 21 **Identification of cell types and Gene expression analysis**

22 We annotated cell clusters based on the expression of known cell marker and the  
23 clustering information provided in the articles. Then we used “RunALRA” function in  
24 Seurat to impute dropped out values in scRNA-seq data. Feature plots and violin plots  
25 were generated using Seurat to show imputed gene expression. In order to compare  
26 gene expression in different datasets, we used “Quantile normalization” in R package  
27 preprocessCore (R package version 1.46.0.  
28 <https://github.com/bmbolstad/preprocessCore>) to preprocess data. Then gene

1 expression data were further denoised by adding random generation for the normal  
2 distribution with mean equal to mean and standard deviation equal to sd.

### 3 **Results**

#### 4 **Annotation of cell types**

5 The gastrointestinal system is composed of esophagus, stomach, ileum, colon and  
6 cecum. In this study, 4 datasets with single-cell transcriptomes of esophagus, gastric,  
7 ileum and colon were analyzed, along with lung (Additional file). Based on Cell  
8 Ranger output, the gene expression count matrices were used to present sequential  
9 clustering of cells according to different organs or particular clusters. The cell type  
10 identity in each cluster was annotated by the expression of the known cell type  
11 markers.

12 In the esophagus, 14 cell types were identified through 87,947 cells. Over 90% cells  
13 fall into four major epithelial cell types: upper, stratified, suprabasal, and dividing  
14 cells of the suprabasal layer (Fig. 1A). The additional cells from the basal layer of  
15 epithelia clustered more closely to the gland duct and mucous secreting cells. Lymph  
16 vessel and endothelial cells are associated with vessel tissues. Immune cells in the  
17 esophagus include T cells, B cells, monocytes, macrophages, dendritic cells (DCs),  
18 and mast cells.

19 A total of 29,678 cells and 10 cell types were identified in the stomach after quality  
20 control with a high proportion of gastric epithelial cells, including antral basal gland  
21 mucous cells (GMCs), pit mucous cells (PMCs), chief cells and enteroendocrine cells  
22 (Fig. 1B). The non-epithelial cell lineages were composed of T cells, B cells, myeloid  
23 cells, fibroblasts and endothelial cells.

24 After quality controls, 50,286 cells and 10 cell types were identified in the ileum (Fig.  
25 1C). The detected cell types included epithelia, endothelial, fibroblast and  
26 enteroendocrine cells. The identified immune cell types were myeloid, CD4<sup>+</sup>T,  
27 CD8<sup>+</sup>T and natural killer T (NKT) cells, along with plasma and B cells. Among

1 11,218 epithelial cells, 5 cell types were identified, namely, absorptive enterocytes,  
2 progenitor absorptive, goblet, Paneth and undifferentiated cells (Fig. 1D).  
3 All the 47,442 cells from the colon were annotated after quality controls (Fig. 1E).  
4 Absorptive and secretory clusters were identified in epithelial cells. The absorptive  
5 clusters included further sub-clusters for transit amplifying (TA) cells (TA 1, TA 2),  
6 immature enterocytes, and enterocytes. The secretory clusters included sub-clusters  
7 for progenitor cells (secretory TA, immature goblet) and for mature cells (goblet, and  
8 enteroendocrine). Ganglion cells and cycling TA cells were also identified in the final  
9 UMAP.

#### 10 **Cell type-specific ACE2 expression**

11 With regard to stomach, the expression of ACE2 is relatively low in all the clusters  
12 (Fig. 2B, C). The selected cell type-specific marker genes were used to identify each  
13 cluster in the stomach (Fig. 2C). MUC6 and TIFF1 were highly expressed in all the  
14 clusters. PGA4 was used to identify chief cells, along with CHGB for enteroendocrine  
15 cells, CD34 for endothelial cells, CD79A for B cells, CD8A and PRF1 for T cells,  
16 VCAN and COL1A1 for fibrous blast and CLEC10A for myeloid cells.

17 As for esophagus, ACE2 was highly expressed in upper and stratified epithelial cells  
18 (Fig. 3B, C). The glands also have a low expression of ACE2 (Fig. 3C). The selected  
19 cell type-specific marker genes were used to identify each cluster in the esophagus  
20 (Fig. 3C). ECM1 was highly expressed in upper epithelial cells. KRT4 and 5 were  
21 mainly found in stratified epithelial cells. KI67 was used to identify dividing  
22 epithelial cells, with MUC5B and KRT23 for glands, COL1A1 and DCN for stroma  
23 cells, VWF and PECAM1 for lymph vessel and endothelial cells. TPSB2, FCN1,  
24 CD79A, GNLY, CD27 and CD3E were used for immune cells, such as myeloid, DC,  
25 B, T and mast cells.

26 In the epithelial cells of the ileum, ACE2 was highly expressed in absorptive  
27 enterocytes and less expressed in progenitor absorptive cells, which was similar to  
28 those in the colon (Fig. 4B, C). The selected cell type-specific marker genes were also

1 used to identify the epithelial cells of the ileum (Fig. 4C). SEC2A5 was found mainly  
2 in the absorptive enterocytes and progenitor absorptive cells. CD24 was found in all  
3 epithelial cells except absorptive enterocytes. MIII67 and AMACR were highly  
4 expressed in undifferentiated and Paneth cells, respectively. BCAS1 was used to  
5 identify undifferentiated cells, with AMACR for Paneth.

6 In the colon, ACE2 was mainly found in enterocytes and less expressed in immature  
7 enterocytes (Fig. 5B, C). The selected cell type-specific marker genes were used to  
8 identify each cluster in the colon (Fig. 5C). AQP8 was mainly found in enterocytes  
9 and immature enterocytes. Additionally, ZG16 and ITLN1 was highly expressed in  
10 goblet and immature goblet. The expression of APOE was in TA2 and secretory TA.  
11 CD27 and TPH1 were used to identify enteroendocrine, with SPC25 for cycling TA.

12 After initial quality controls, 57,020 cells and 25 cell types were identified in the lung  
13 (Fig. 6A). The detected cell types included ciliated, alveolar type 1 (AT1) and  
14 alveolar type 2 (AT2) cells, along with fibroblast, muscle, and endothelial cells. The  
15 identified immune cell types were T, B and NK cells, along with macrophages,  
16 monocytes and dendritic cells (DC). ACE2 was mainly expressed in AT2 cells and  
17 could also be found in AT1 and fibroblast cells (Fig. 6B).

18 Among all the ACE2-expressing cells in normal digestive system and lung, the  
19 expression of ACE2 was more in ileum and colon than that in the lung and esophagus  
20 (Fig. 6C).

## 21 **Discussion**

22 The coronavirus is the common infection source of upper respiratory,  
23 gastrointestinal and central nervous system in humans and other mammals[19]. At the  
24 beginning of the twenty-first century, two betacoronaviruses, SARS-CoV and  
25 MERS-CoV, caused persistent public panics and became the most significant public  
26 health events[20]. In December 2019, a novel identified coronavirus (2019-nCov)  
27 induced an ongoing outbreak of pneumonia in Wuhan, Hubei, China with arising  
28 number of infected patients[4]. Till now, its infection routes and digestive system

1 infection are still unclear. In this study, we found the high expressions of ACE2, the  
2 cell entry receptor of 2019-nCov, in the lung AT2 cells, esophagus upper and stratified  
3 epithelial cells and absorptive enterocytes from ileum and colon, indicating that not  
4 only respiratory system but also digestive system are potential routes of infection. In  
5 addition, the enteric symptoms of 2019-nCov may be associated with the invaded  
6 ACE2-expressing enterocytes.

7 Generally, many respiratory pathogens, such as influenza, SARS-CoV and  
8 SARSr-CoV, cause enteric symptoms, so is 2019-nCov[4, 5]. As a classic respiratory  
9 coronavirus, SARS often causes enteric symptoms along with respiratory symptoms.  
10 Moreover, transmission with stool is also a neglected risk for SARS[21]. During the  
11 infection of SARS and highly pathogenic strains of influenza, their enteric symptoms  
12 are associated with the increased permeability to intestinal lipopolysaccharide (LPS)  
13 and bacterial transmigration through gastrointestinal wall[22, 23]. However, the  
14 mechanism of 2019-nCov-induced enteric symptom is still unknown.

15 A recent study revealed that similar to SARS-CoV and MERS-CoV, ACE2 was the  
16 cell entry receptor for 2019-nCov[10]. Previously, ACE2 was isolated from  
17 SARS-CoV-permissive Vero E6 cells[24]. It could interact with a defined  
18 receptor-binding domain (RBD) of CTD1 in SARS-CoV and facilitate efficient  
19 cross-species infection and person-to-person transmission[9, 25]. The “up” and “down”  
20 transition of CTD1 allows ACE2 binding by regulating the relationship among CTD1,  
21 CTD2, S1-ACE2 complex and S2 subunit[26]. With regard to human HeLa cells,  
22 expressing ACE2 from human, civet, and Chinese horseshoe bat can help many kinds  
23 of SARSr-CoV, including 2019-nCov, to enter into the cells, indicating the important  
24 role of ACE2 in cellular entry [10, 27-29]. Therapeutically, anti-ACE2 antibody can  
25 block viral replication on Vero E6 cells[24].

26 By analyzing the expression of ACE2 in normal human gastrointestinal system and  
27 lung, we found high expression of ACE2 in the lung AT2 cells, esophagus upper and  
28 stratified epithelial cells and absorptive enterocytes from ileum and colon. Similar to

1 the previous study, ACE2 was more expressed in AT2 cells and less expressed in AT1  
2 cells in normal lung[30]. In lung alveoli, AT1 epithelial cells are responsible for gas  
3 exchange and AT2 cells are in charge of surfactant biosynthesis and  
4 self-renewing[31]. In SARS-CoV infection, AT2 is the major infected cell types by  
5 viral antigens and secretory vesicles detection. Its expression in AT2 cells is variable  
6 in different donors, which may be associated with different susceptibility and  
7 seriousness[30]. Thus, we suppose that AT2 cells might be the key  
8 2019-nCov-invaded cell in lung and its number might be associated with the severity  
9 of respiratory symptoms, which can explain the existence of asymptomatic  
10 2019-nCov carrier.

11 ACE2 was also highly expressed in the esophagus upper and stratified epithelial cells.  
12 Histologically, both esophagus and respiratory system organs, such as trachea and  
13 lung are originated from the anterior portion of the intermediate foregut[32]. After  
14 being separated from the neighboring respiratory system, the esophagus undergoes  
15 subsequent morphogenesis of a simple columnar-to-stratified squamous epithelium  
16 conversion[33]. The stratified squamous epithelium can be nourished by submucosal  
17 glands and sustain the passing of the abrasive raw food. In Barrett's oesophagus (BE),  
18 acid reflux-induced oesophagitis and the multilayered epithelium (MLE) are  
19 associated with both upper and stratified epithelial cells[34].

20 In the digestive system, besides esophagus upper and stratified epithelial cells, ACE2  
21 was also found in the absorptive enterocytes from ileum and colon, the most  
22 vulnerable intestinal epithelial cells. In microbe infections, the intestinal epithelial  
23 cells function as a barrier and help to coordinate immune responses[35]. The  
24 absorptive enterocytes can be infected by coronavirus, rotavirus and noroviruses,  
25 resulting in diarrhea by destructing absorptive enterocytes, malabsorption, unbalanced  
26 intestinal secretion and activated enteric nervous system[36-38]. Thus, we suppose  
27 that the enteric symptom of diarrhea might be associated with the invaded  
28 ACE2-expressing enterocytes. In addition, due to the high expression of cell receptor

1 ACE2 in esophagus upper and stratified epithelial cells and absorptive enterocytes  
2 from ileum and colon, we suppose that digestive system can be invaded by  
3 2019-nCov and serve as a route of infection.

#### 4 **Conclusion**

5 This study provides the bioinformatics evidence for the potential respiratory and  
6 digestive systems infection of 2019-nCov and assists clinicians in preventing and  
7 treating the 2019-nCoV infection.

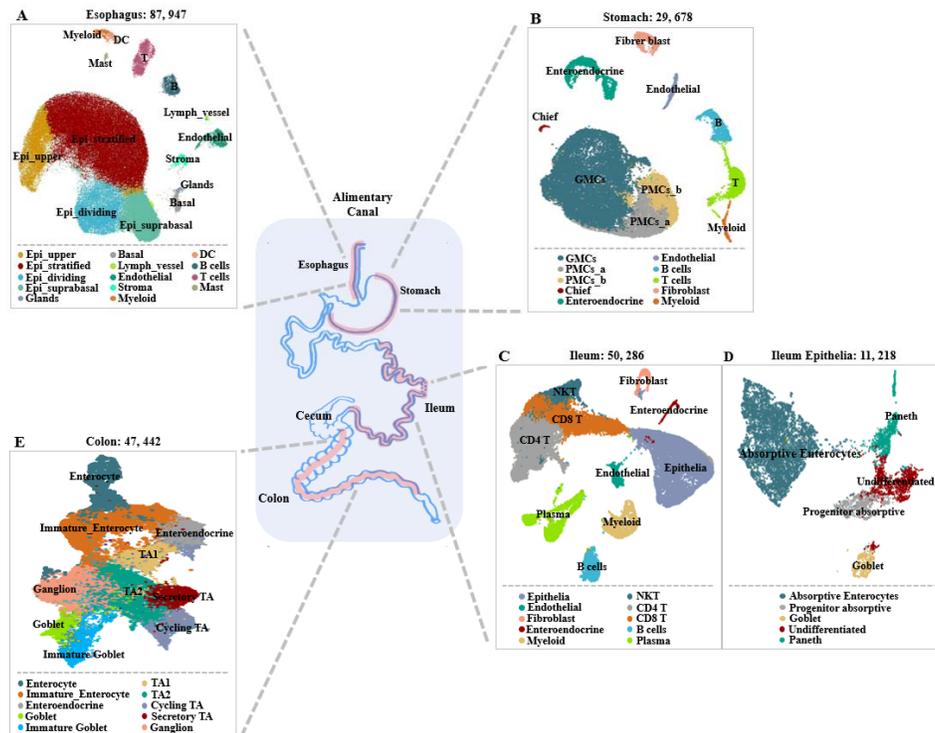
#### 8 **References**

- 9 1. The L. Emerging understandings of 2019-nCoV. *Lancet*. 2020.
- 10 2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R,  
11 Niu P, Zhan F, Ma X, et al. A Novel Coronavirus from Patients with Pneumonia  
12 in China, 2019. *N Engl J Med*. 2020.
- 13 3. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, Yuen KY. Genomic  
14 characterization of the 2019 novel human-pathogenic coronavirus isolated from a  
15 patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect*.  
16 2020; 9: 221-36.
- 17 4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X,  
18 Cheng Z, Yu T, Xia J, et al. Clinical features of patients infected with 2019 novel  
19 coronavirus in Wuhan, China. *Lancet*. 2020.
- 20 5. Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, Xing F, Liu J, Yip CC, Poon  
21 RW, Tsoi HW, Lo SK, Chan KH, et al. A familial cluster of pneumonia associated  
22 with the 2019 novel coronavirus indicating person-to-person transmission: a  
23 study of a family cluster. *Lancet*. 2020.
- 24 6. Zhou J, Li C, Zhao G, Chu H, Wang D, Yan HH, Poon VK, Wen L, Wong BH,  
25 Zhao X, Chiu MC, Yang D, Wang Y, et al. Human intestinal tract serves as an  
26 alternative infection route for Middle East respiratory syndrome coronavirus. *Sci*  
27 *Adv*. 2017; 3: eaao4966.
- 28 7. Openshaw PJ. Crossing barriers: infections of the lung and the gut. *Mucosal*  
29 *Immunol*. 2009; 2: 100-2.
- 30 8. Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu*  
31 *Rev Virol*. 2016; 3: 237-61.
- 32 9. Gui M, Song W, Zhou H, Xu J, Chen S, Xiang Y, Wang X. Cryo-electron  
33 microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite  
34 conformational state for receptor binding. *Cell Res*. 2017; 27: 119-29.
- 35 10. Zhou P YX, Wang XiG , Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B,  
36 Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y,  
37 Shen XR, Wang X, Zheng XS, Zhao Ka, Chen QJ, Deng F, Liu LL, Yan B,  
38 Zhan FX, Wang YY, Xiao GF, Shi ZL. Discovery of a novel coronavirus

- 1 associated with the recent pneumonia outbreak in humans and its potential bat  
2 origin. bioRxiv. 2020.
- 3 11. Xintian Xu PC, Jingfang Wang, Jiannan Feng, Hui Zhou, Xuan Li, Wu Zhong, Pei  
4 Hao. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and  
5 modeling of its spike protein for risk of human transmission Science China. 2020.
- 6 12. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene  
7 expression and hybridization array data repository. Nucleic Acids Res. 2002; 30:  
8 207-10.
- 9 13. Madisson E, Wilbrey-Clark A, Miragaia RJ, Saeb-Parsy K, Mahbubani KT,  
10 Georgakopoulos N, Harding P, Polanski K, Huang N, Nowicki-Osuch K,  
11 Fitzgerald RC, Loudon KW, Ferdinand JR, et al. scRNA-seq assessment of the  
12 human lung, spleen, and esophagus tissue stability after cold preservation.  
13 Genome Biol. 2019; 21: 1.
- 14 14. Zhang P, Yang M, Zhang Y, Xiao S, Lai X, Tan A, Du S, Li S. Dissecting the  
15 Single-Cell Transcriptome Network Underlying Gastric Premalignant Lesions  
16 and Early Gastric Cancer. Cell Rep. 2019; 27: 1934-47.e5.
- 17 15. Martin JC, Chang C, Boschetti G, Ungaro R, Giri M, Grout JA, Gettler K,  
18 Chuang LS, Nayar S, Greenstein AJ, Dubinsky M, Walker L, Leader A, et al.  
19 Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular  
20 Module Associated with Resistance to Anti-TNF Therapy. Cell. 2019; 178:  
21 1493-508.e20.
- 22 16. Smillie CS, Biton M, Ordovas-Montanes J, Sullivan KM, Burgin G, Graham DB,  
23 Herbst RH, Rogel N, Slyper M, Waldman J, Sud M, Andrews E, Velonias G, et al.  
24 Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis.  
25 Cell. 2019; 178: 714-30.e22.
- 26 17. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao  
27 Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell  
28 Data. Cell. 2019; 177: 1888-902.e21.
- 29 18. Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ.  
30 Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain  
31 Cell Identity. Cell. 2019; 177: 1873-87.e17.
- 32 19. Perlman S, Netland J. Coronaviruses post-SARS: update on replication and  
33 pathogenesis. Nat Rev Microbiol. 2009; 7: 439-50.
- 34 20. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent  
35 insights into emerging coronaviruses. Nat Rev Microbiol. 2016; 14: 523-34.
- 36 21. Peiris JS, Chu CM, Cheng VC, Chan KS, Hung IF, Poon LL, Law KI, Tang BS,  
37 Hon TY, Chan CS, Chan KH, Ng JS, Zheng BJ, et al. Clinical progression and  
38 viral load in a community outbreak of coronavirus-associated SARS pneumonia:  
39 a prospective study. Lancet. 2003; 361: 1767-72.
- 40 22. Powers JH, 3rd, Bacci ED, Guerrero ML, Leidy NK, Stringer S, Kim K, Memoli  
41 MJ, Han A, Fairchok MP, Chen WJ, Arnold JC, Danaher PJ, Lalani T, et al.  
42 Reliability, Validity, and Responsiveness of InFLUenza Patient-Reported

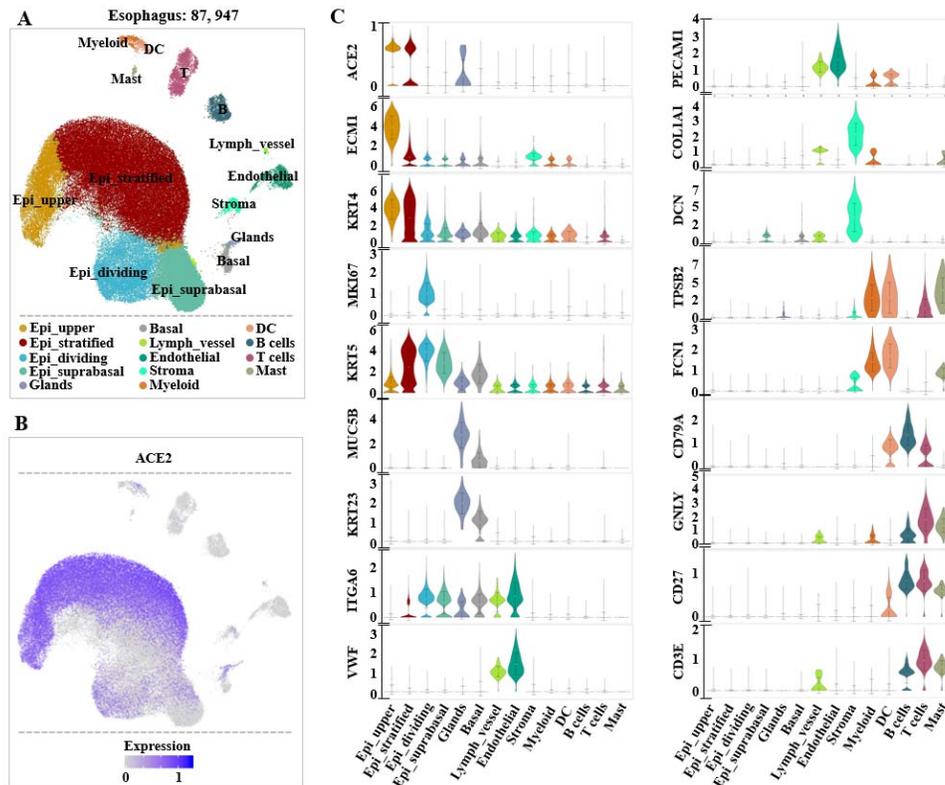
- 1 Outcome (FLU-PRO(c)) Scores in Influenza-Positive Patients. *Value Health*.  
2 2018; 21: 210-8.
- 3 23. To KF, Tong JH, Chan PK, Au FW, Chim SS, Chan KC, Cheung JL, Liu EY, Tse  
4 GM, Lo AW, Lo YM, Ng HK. Tissue and cellular tropism of the coronavirus  
5 associated with severe acute respiratory syndrome: an in-situ hybridization study  
6 of fatal cases. *J Pathol*. 2004; 202: 157-63.
- 7 24. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M,  
8 Sullivan JL, Luzuriaga K, Greenough TC, Choe H, Farzan M.  
9 Angiotensin-converting enzyme 2 is a functional receptor for the SARS  
10 coronavirus. *Nature*. 2003; 426: 450-4.
- 11 25. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike  
12 receptor-binding domain complexed with receptor. *Science*. 2005; 309: 1864-8.
- 13 26. Song W, Gui M, Wang X, Xiang Y. Cryo-EM structure of the SARS coronavirus  
14 spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog*.  
15 2018; 14: e1007236.
- 16 27. Yang XL, Hu B, Wang B, Wang MN, Zhang Q, Zhang W, Wu LJ, Ge XY, Zhang  
17 YZ, Daszak P, Wang LF, Shi ZL. Isolation and Characterization of a Novel Bat  
18 Coronavirus Closely Related to the Direct Progenitor of Severe Acute Respiratory  
19 Syndrome Coronavirus. *J Virol*. 2015; 90: 3253-6.
- 20 28. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B,  
21 Zhang W, Peng C, Zhang YJ, Luo CM, Tan B, et al. Isolation and characterization  
22 of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*. 2013; 503:  
23 535-8.
- 24 29. Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, Li B, Xie JZ, Shen XR, Zhang YZ,  
25 Wang N, Luo DS, Zheng XS, Wang MN, et al. Discovery of a rich gene pool of  
26 bat SARS-related coronaviruses provides new insights into the origin of SARS  
27 coronavirus. *PLoS Pathog*. 2017; 13: e1006698.
- 28 30. Qian Z, Travanty EA, Oko L, Edeen K, Berglund A, Wang J, Ito Y, Holmes KV,  
29 Mason RJ. Innate immune response of human alveolar type II cells infected with  
30 severe acute respiratory syndrome-coronavirus. *Am J Respir Cell Mol Biol*. 2013;  
31 48: 742-8.
- 32 31. Nabhan AN, Brownfield DG, Harbury PB, Krasnow MA, Desai TJ. Single-cell  
33 Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science*. 2018;  
34 359: 1118-23.
- 35 32. Que J, Okubo T, Goldenring JR, Nam KT, Kurotani R, Morrissey EE, Taranova O,  
36 Pevny LH, Hogan BL. Multiple dose-dependent roles for Sox2 in the patterning  
37 and differentiation of anterior foregut endoderm. *Development*. 2007; 134:  
38 2521-31.
- 39 33. Zhang Y, Jiang M, Kim E, Lin S, Liu K, Lan X, Que J. Development and stem  
40 cells of the esophagus. *Semin Cell Dev Biol*. 2017; 66: 25-35.
- 41 34. Jiang M, Li H, Zhang Y, Yang Y, Lu R, Liu K, Lin S, Lan X, Wang H, Wu H, Zhu  
42 J, Zhou Z, Xu J, et al. Transitional basal cells at the squamous-columnar junction

- 1 generate Barrett's oesophagus. *Nature*. 2017; 550: 529-33.
- 2 35. Haber AL, Biton M, Rogel N, Herbst RH, Shekhar K, Smillie C, Burgin G,  
3 Delorey TM, Howitt MR, Katz Y, Tirosh I, Beyaz S, Dionne D, et al. A single-cell  
4 survey of the small intestinal epithelium. *Nature*. 2017; 551: 333-9.
- 5 36. Crawford SE, Ramani S, Tate JE, Parashar UD, Svensson L, Hagbom M, Franco  
6 MA, Greenberg HB, O'Ryan M, Kang G, Desselberger U, Estes MK. Rotavirus  
7 infection. *Nat Rev Dis Primers*. 2017; 3: 17083.
- 8 37. Ettayebi K, Crawford SE, Murakami K, Broughman JR, Karandikar U, Tenge VR,  
9 Neill FH, Blutt SE, Zeng XL, Qu L, Kou B, Opekun AR, Burrin D, et al.  
10 Replication of human noroviruses in stem cell-derived human enteroids. *Science*.  
11 2016; 353: 1387-93.
- 12 38. Desmarests LM, Theuns S, Roukaerts ID, Acar DD, Nauwynck HJ. Role of sialic  
13 acids in feline enteric coronavirus infections. *J Gen Virol*. 2014; 95: 1911-8.
- 14



15  
16 **Figure 1: Single-Cell Atlas of digestive tract samples**  
17 (A). The UMAP plot of 87947 esophageal cells to visualize cell-type clusters  
18 (B). The UMAP plot of 29678 gastric mucosa cells to visualize cell-type clusters.  
19 (C). The UMAP plot of 50286 ileal cell cells to visualize cell-type clusters.  
20 (D). The UMAP plot of 11218 ileal epithelial cells to visualize finer clusters.  
21 Epithelial cells in ileum were further divided into finer cell subsets because of the  
22 heterogeneity within the cell population according to transcription characteristics.  
23 (E). The UMAP plot of 47442 colon cells to visualize cell-type clusters.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14



15  
16  
17  
18  
19  
20  
21  
22

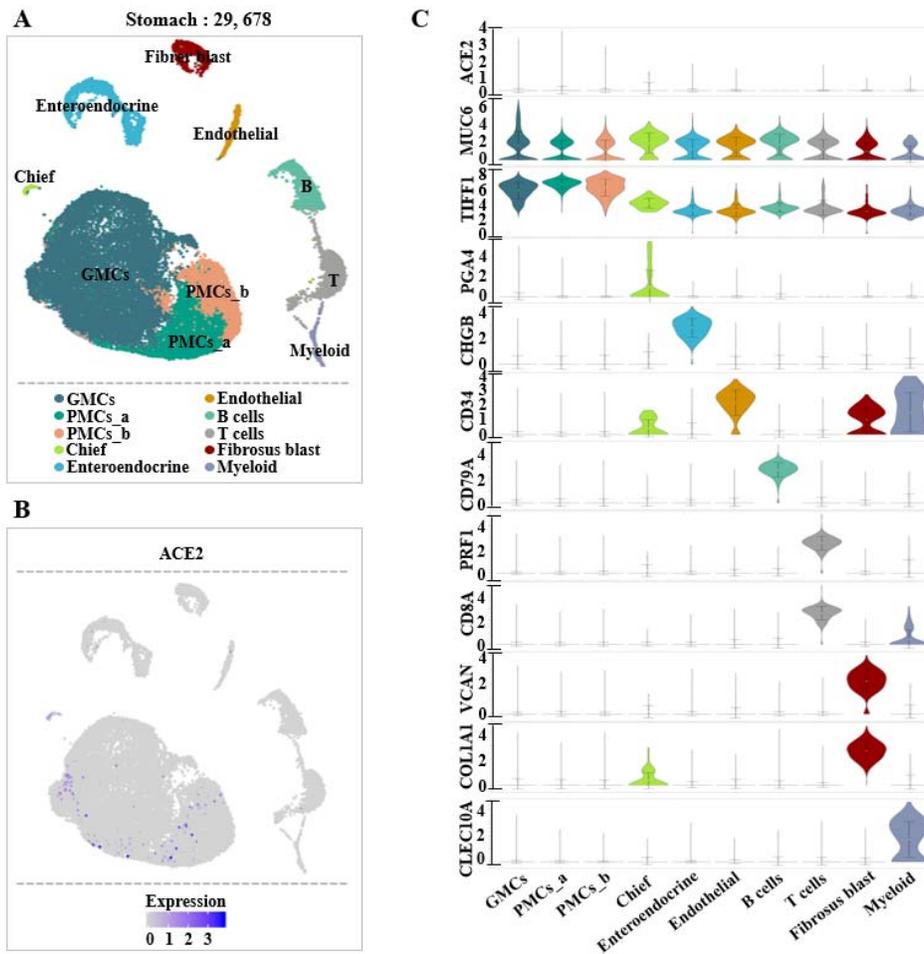
Figure 2. Single-cell analysis of esophageal cells

(A). UMAP plots showing the landscape of esophageal cells. 14 cell clusters were identified across 87947 cells.

(B). UMAP plots showing the expression of *ACE2* across clusters.

(C). Violin plots for esophageal clusters showing marker genes and *ACE2* across clusters. The expression is measured as the  $\log_2(\text{TP10K}+1)$ .

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14



15

16 Figure 3. Single-cell analysis of gastric mucosal cells

17 (A). UMAP plots showing the landscape of gastric mucosal tissue. 10 cell clusters were

1 identified across 29678 cells after quality control, dimensionality reduction and  
2 clustering.

3 (B). UMAP plots showing the expression (grey to blue) of gene *ACE2* across clusters.

4 (C). Violin plots for gastric mucosal clusters marker genes and *ACE2* across clusters.

5 The expression is measured as the  $\log_2$  (TP10K+1).

6

7

8

9

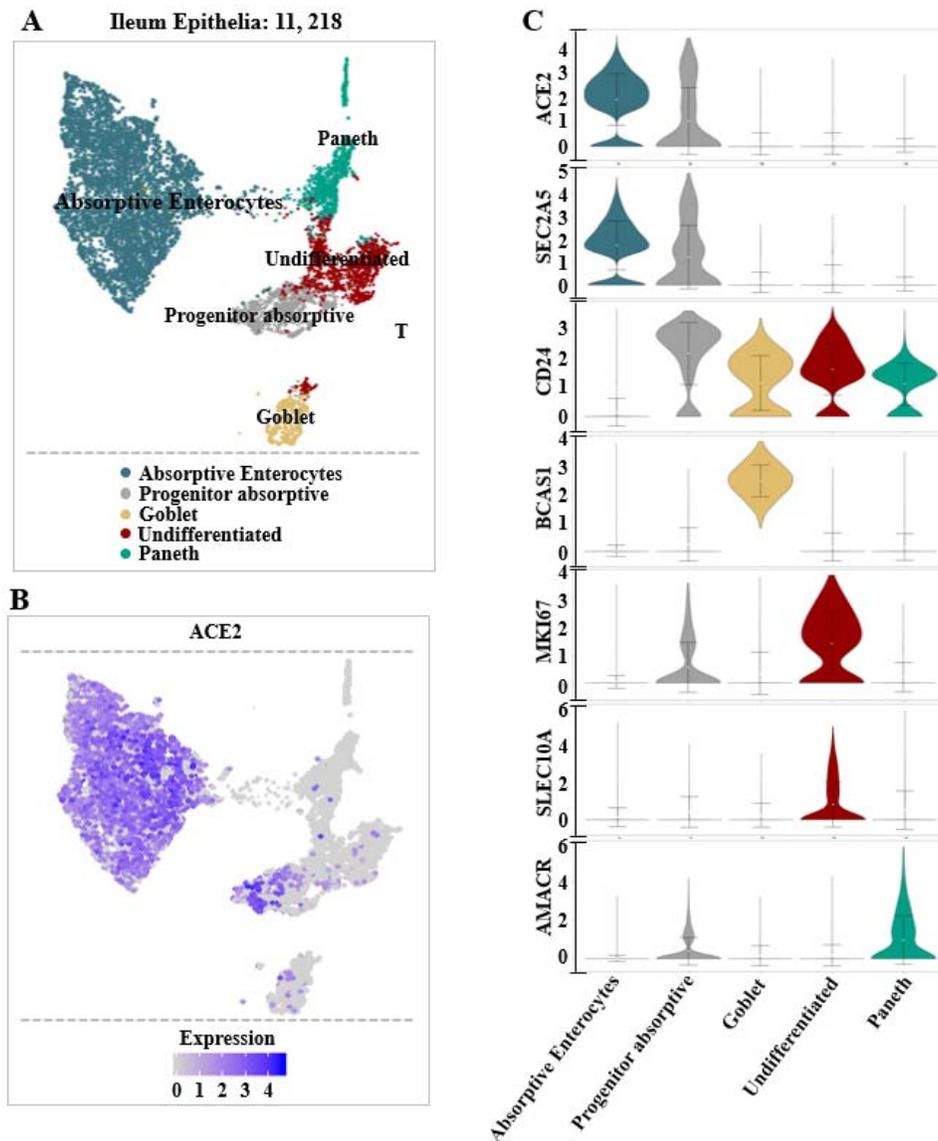
10

11

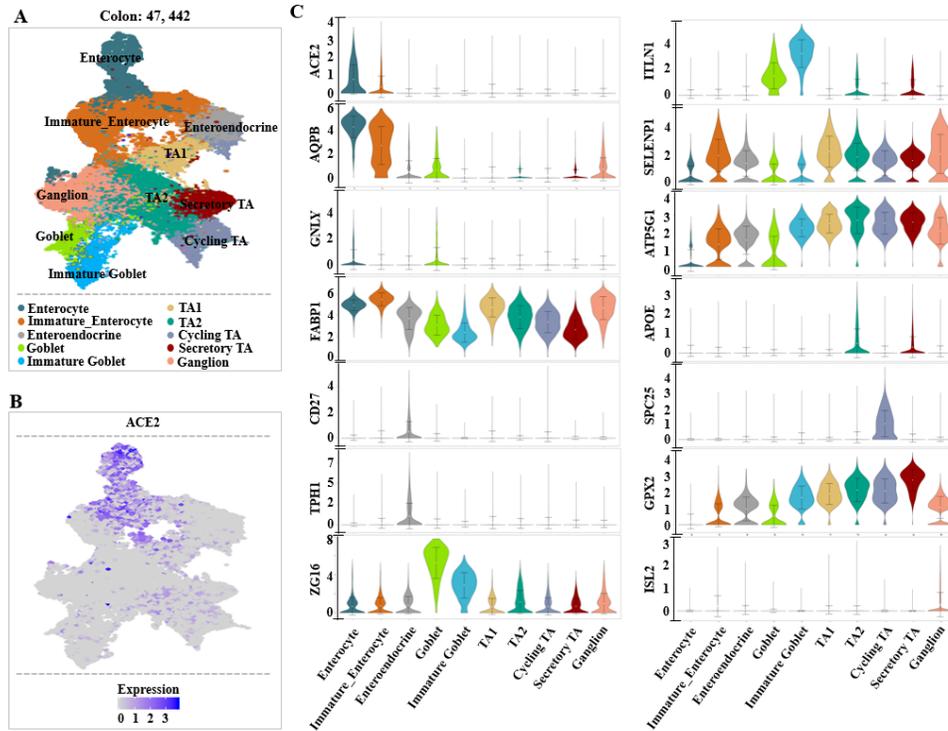
12

13

14

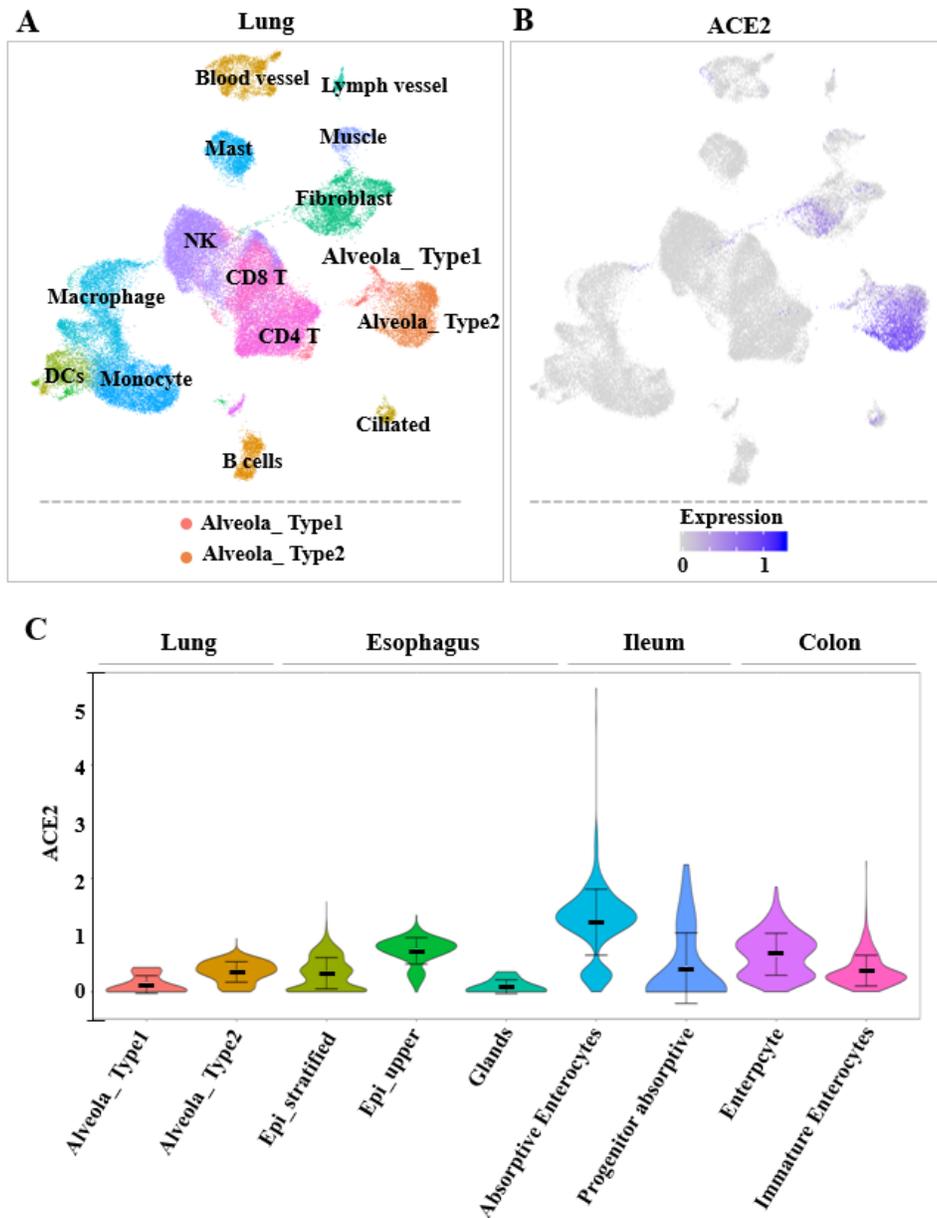


1  
2 Figure 4. Single-cell analysis of ileal epithelial cells  
3 (A). UMAP plots showing the landscape of ileal epithelial cells. 10 cell clusters were  
4 identified across 11218 cells after quality control, dimensionality reduction and  
5 clustering.  
6 (B). UMAP plots showing the expression of *ACE2* across clusters.  
7 (C). Violin plots for ileal epithelial marker genes and *ACE2* across clusters. The  
8 expression is measured as the  $\log_2(TP10K+1)$ .  
9  
10  
11  
12  
13



1  
2 Figure 5. Single-cell analysis of colon cells  
3 (A). UMAP plots showing the landscape of colon cell cells. 10 cell clusters were  
4 identified across 47442 cells after quality control, dimensionality reduction and  
5 clustering.  
6 (B). UMAP plots showing the expression of *ACE2* across clusters.  
7 (C). Violin plots for colon clusters marker genes and *ACE2* across clusters. The  
8 expression is measured as the  $\log_2(TP10K+1)$ .

9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

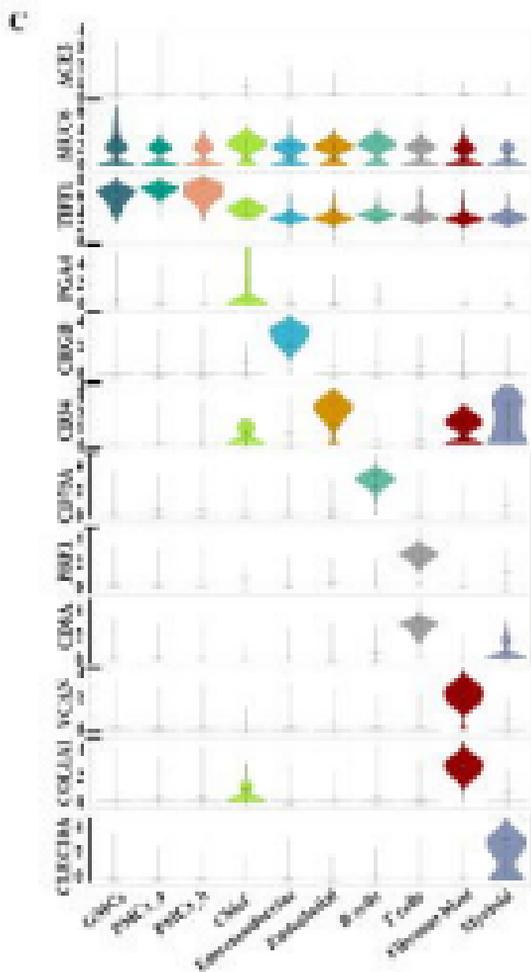
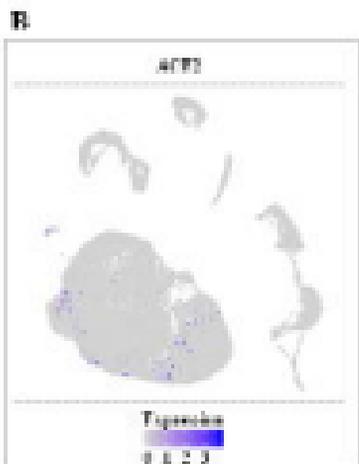


1  
2 Figure 6. Single-cell analysis of lung cells  
3 (A). UMAP plots showing the landscape of lung cells. 16 cell clusters were identified  
4 across 57020 cells.  
5 (B). UMAP plots showing the expression of *ACE2* across lung clusters.  
6 (C). Violin plots for *ACE2* across 2 lung clusters and 7 digestive tract clusters. Gene  
7 expression matrix was normalized and denoised to remove unwanted technical  
8 variability across 4 datasets.

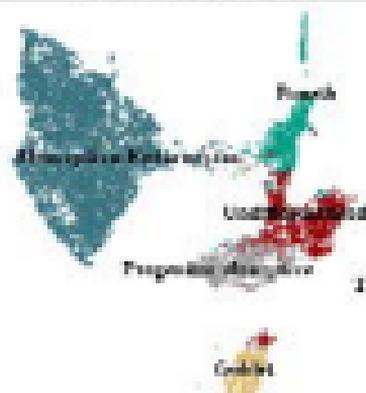
9  
10  
11







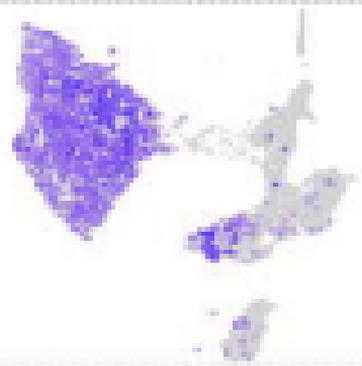
**A** *Human Epithelial 10, 218*



- Alveolar Macrophages
- Progenitor Alveolar
- Collet
- Lymphoid/endothelial
- Paneth

**B**

ACE2



**C**

