Single-cell atlas of a non-human primate reveals new pathogenic mechanisms of COVID-19

Lei Han,[1,2,15] Xiaoyu Wei,[1,2,3,15] Chuanyu Liu,[1,2,14,15] Giacomo Volpe,[4,15] Zhifeng Wang,[1,2,15] Taotao Pan,[1,2,3] Yue Yuan,[1,2,3] Ying Lei,[1,2,14] Yiwei Lai,[4] Carl Ward,[4] Yeya Yu,[1,2,5] Mingyue Wang,[1,2,3] Quan Shi,[1,2,6] Tao Wu,[1,2] Liang Wu,[1,2,3] Ya Liu,[1,2] Chunqing Wang,[1,2,3] Yuanhang Zhang,[1,2,3] Haixi Sun,[1,2] Hao Yu,[1,2] Zhenkun Zhuang,[1,2,7] Tingting Tang,[1,2] Yunting Huang,[2] Haorong Lu,[2] Liqin Xu,[1,2,6] Jiangshan Xu,[1,2,3] Mengnan Cheng,[1,2,3] Yang Liu,[1,2,3] Chi Wai Wong,[8] Tao Tan,[9] Weizhi Ji,[9] Patrick H. Maxwell[12], Huanming Yang,[1,10] Jian Wang,[1,10] Shida Zhu,[1,2] Shiping Liu,[1,2,13] Xun Xu,[1,2,11,*] Yong Hou,[1,2,13,*] Miguel A. Esteban,[4,16,*] and Longqi Liu[1,2,14,*]

[1]BGI-Shenzhen, Shenzhen 518083, China

[2]China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China

[3]BGI Education Center, University of Chinese Academy of Sciences (CAS), Shenzhen 518083, China

[4]Laboratory of Integrative biology, Guangzhou Institutes of Biomedicine and Health, CAS, Guangzhou 510530, China

[5]BGI College, Zhengzhou University, Zhengzhou 450000, China

[6]Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, Copenhagen 2200, Denmark

[7]School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China

[8]Huazhen Biosciences, Guangzhou 510900, China

[9]Yunnan Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming 650500, China

[10]James D. Watson Institute of Genome Science, Hangzhou, 310008, China

[11]Institute for Stem Cells and Regeneration, CAS, Beijing 100101, China

[12]Cambridge Institute for Medical Research, Department of Medicine, University of Cambridge, Cambridge CB2 0XY, United Kingdom

[13]Shenzhen Key Laboratory of Single-Cell Omics, BGI-Shenzhen, Shenzhen 518120, China

[14]Shenzhen Bay Laboratory, Shenzhen 51800, China

[15]These authors contributed equally

[16]Lead Contact

*Correspondence: xuxun@genomics.cn (X.X.), houyong@genomics.cn (Y.H.), miguel@gibh.ac.cn (M.A.E.), liulongqi@genomics.cn (L.L.)

Key words:

Single-cell transcriptome, non-human primate, COVID-19, SARS-CoV-2, ACE2, TMPRSS2, Interleukin 6, STAT transcription factors, immune cell exhaustion

Bullet points:

1-We used a single-cell transcriptome atlas of 9 monkey tissues to study COVID-19.

2-*ACE2*$^{+}$*TMPRSS2*$^{+}$ epithelial cells of lung, kidney and liver are targets for SARS-CoV-2.

3-*ACE2* correlation analysis shows *IDO2* and *ANPEP* as potential therapeutic opportunities.

4-We unveil a link between IL6, STAT transcription factors and boosted SARS-CoV-2 entry.

ABSTRACT

Stopping COVID-19 is a priority worldwide. Understanding which cell types are targeted by SARS-CoV-2 virus, whether interspecies differences exist, and how variations in cell state influence viral entry is fundamental for accelerating therapeutic and preventative approaches. In this endeavor, we profiled the transcriptome at single-cell resolution of nine tissues from a Macaca fascicularis monkey. The distribution of SARS-CoV-2 facilitators, ACE2 and TMRPSS2, in different cell subtypes showed substantial heterogeneity across lung, kidney, thyroid and liver. Co-expression analysis identified immunomodulatory proteins such as IDO2 and ANPEP as potential SARS-CoV-2 targets responsible for immune cell exhaustion. Furthermore, single-cell chromatin accessibility analysis of the kidney unveiled a plausible link between IL6-mediated innate immune responses aiming to protect tissue and enhanced ACE2 expression that could promote viral entry. Our work constitutes a unique resource for understanding SARS-CoV-2 pathophysiology in two phylogenetically close species, which might guide in the development of effective treatments in humans.

INTRODUCTION

As the distance between humans and wild animal habitats diminishes due to uncontrolled human expansion, a series of zoonotic diseases with high mortality rates have emerged. For instance, the recent outbreak of Ebola in Africa, which killed over 5,000 people, was most likely spread from bats and primates to humans (Kock et al., 2019). The current outbreak of coronavirus disease 2019 (COVID-19) caused by the coronavirus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Zhu et al., 2020) is not the only example of coronaviruses that have recently passed from animals to humans. Coronaviruses are a family of RNA viruses that typically cause respiratory tract infections in humans, yet they are frequently in the reservoir of wild animals with no disease (Cui et al., 2019). For example, the common cold is often (10-15%) caused by a coronavirus (e.g. HCoV-229E and HCoV-OC43) (Su et al., 2016). However, coronaviruses can also lead to severe and life-threatening diseases. In the early 2000s a coronavirus called SARS-CoV, believed to be passed from bats to humans in South East Asia, caused more than 700 deaths from around 8,000 confirmed cases worldwide (Lau et al., 2005). Since 2012, another zoonotic coronavirus which is believed to have passed from camels to humans in the Middle East was designated as Middle East Respiratory Syndrome (MERS) (Reusken et al., 2013). To date, there have been over 2,500 confirmed cases of MERS with over 800 deaths. While SARS appears to have been eradicated, MERS cases are sporadic and human to human spread is limited (Su et al., 2016).

As of 2nd April 2020, COVID-19 (Zhu et al., 2020) has become a global pandemic with more than 930,000 confirmed cases and over 47,000 deaths. Due to its high infectivity rate and the high level of intensive care that many patients need, COVID-19 has overwhelmed national health services and destabilized the world. One important reason is that many people who are positive for the virus show mild symptoms (Chan et al., 2020; Rothe et al., 2020), leading to unnoticed spread of the virus. The current worldwide emergency, possibility of expansion to less developed countries, risk of virus mutations and the perpetuation beyond this season has made it imperative to stop the current trajectory of virus spreading. Developing drugs and preventative vaccines are ongoing but to warrant success it is necessary to have more knowledge about the disease mechanisms. So far, little is known except for the viral binding via angiotensin converting enzyme 2 (ACE2) and subsequent priming by type 2 transmembrane serine protease 2 (TMPRSS2) protease, which are shared mechanisms with

SARS and MERS (Hoffmann et al., 2020; Walls et al., 2020). To test experimental treatments, animal models close to humans are necessary due to sequence variation of ACE2 and changes in the proportions of cell subtypes in organs between species. For these reasons, it is essential to have a species close to human to study COVID-19. In this regard, monkey experiments have already shown that reinfection of SARS-CoV-2 is preventable by acquired immunity, thus, largely debunking one of the initial myths about the disease (Bao et al., 2020).

Questions about the proportions of cell types within organs between species and their crosstalk can be addressed effectively through single-cell profiling technologies, in particular single-cell RNA sequencing (RNA-seq) and single-cell assay for transposase accessible chromatin sequencing (ATAC-seq). Yet, although human data are accumulating (Han et al., 2020), monkey data are still scarce. The comparison between human and monkey data will be crucial for advancing our knowledge of COVID-19.

Here, we provide a high-resolution single-cell atlas of nine organs/tissues (lung, kidney, pancreas, brain, parotid, liver, thyroid, aorta artery, and blood) in monkey, encompassing 215,334 cells. By comparing the expression of SARS related targets in monkey and human, we have identified cell-to-cell similarities as expected. Crucially, we also discovered stark differences in *ACE2* expression between these two species, for example in the ciliated vs alveolar type 2 cells of the lung and the hepatocytes in liver. We also observed that *ACE2* is heterogeneous among different epithelial cell subtypes across these organs/tissues, suggesting that variations in cell state could influence viral entry. Supporting this, single-cell ATAC-seq of monkey kidney identified regulatory elements driven by STAT transcription factors and interferon regulatory factor (IRF) in the *ACE2* locus. This suggests that cytokines, particularly IL6, aiming to induce a tissue protective response can exacerbate the disease by aiding viral entry into target cells. Additionally, through correlation analysis with *ACE2* expression, we have identified several potential candidates involved in COVID-19 pathophysiology, such as Transmembrane protein 27 (TMEM27), Indoleamine 2,3-dioxygenase 2 (IDO2), DnaJ heat shock protein family (Hsp40) member C12 (DNAJC12) and Alanyl aminopeptidase N (ANPEP). These targets may offer therapeutic opportunities.

Taken together, our data constitute a unique resource which will aid the scientific community in the fight against SARS-CoV-2. From a wider perspective, this will also be useful for comparative studies aimed at understanding physiological differences between monkey and other species, in particular, human.

RESULTS

Cellular heterogeneity of nine non-human primate tissues assessed by single-cell RNA-seq

We profiled, at the single-cell level, the transcriptome of the model organism cynomolgus monkey *(Macaca fascicularis)*, as it is phylogenetically close to human and this could help advance our knowledge of human physiology and disease. As proof of principle, we decided to use our data to understand how SARS-CoV-2 spreads and triggers the clinical features that have been lethal in a number of patients. For this study, we used a six-year-old female monkey in which we profiled nine different organs (indicated in **Fig. 1A**). These included lung, liver and kidney as the known affected organs by the closely related SARS-CoV infection (Hamming et al., 2004), and have been reported to have high *ACE2* expression in human (Naicker et al., 2020). Peripheral blood mononuclear cells (PBMC) were added because altered immune responses are thought to be detrimental in the disease (Conti et al., 2020). Neocortex was chosen because of the clinical symptoms which involve loss of smell and taste suggesting that the central nervous system may be targeted (Bagheri et al., 2020). The parotid gland was chosen on the basis that saliva is one of the main means of infection spread. Additionally, we selected aorta, thyroid and pancreas.

We employed a high-throughput platform recently developed in-house, DNBelab C4, which is a scalable and cost-effective approach for microfluidic droplet-based approach (Liu et al., 2019). Except for PBMC sequencing, which was performed using cells in suspension, the sequencing for all the other organs was done using single-nucleus library preparations. Following euthanasia, the selected organs were extracted, single-nucleus/cell suspensions were obtained and used for library preparation. A total of 40,226 liver, 45,286 kidney, 36,421 pancreas, 44,355 parotid gland, 12,822 lung, 7,877 thyroid, 6,361 neocortex, 2,260 aorta and 19,726 nuclei/PBMC passed quality control and were used for downstream analysis (**Supplemental table 1, Fig. S1A, B**).

In a global view of our single-cell dataset, each organ clustered separately, with the exception of a few cell types such as macrophages, adipocytes and endothelial cells, which were shared between different organs (**Fig. 1B**). We performed Uniform Manifold Approximation and Projection (UMAP) on the 215,334 cells and identified 44 major clusters by performing unbiased graph-based Louvain clustering (**Supplemental table 1**). Some clusters were largely composed of cells belonging to a specific tissue, such as hepatocytes in cluster 14, pancreas

acinar cells in cluster 27 and parotid acinar cells in cluster 31 (**Fig. 1C, S1C**). We next performed clustering and differential gene expression analysis to dissect the cellular composition of each individual organ. This analysis confirmed the typical patterns of cell heterogeneity for all the organs/tissues. When examining the lung tissue, we defined 10 major clusters with specific molecular markers, including ciliated cells, macrophages, fibroblasts, pericytes, alveolar (pneumocytes) type 1 and type 2, endothelial and club cells (**Fig. S2A**). The kidney consisted of 11 clusters, those being podocytes, thick ascending limb cell, proximal tubule cells, intercalated cells 1 and 2, connecting tubule cells, distal convoluted tubule cells, stomal cells, thin limb cells, principal cells and endothelial cells (**Fig. S2B**). Analysis of liver tissue revealed hepatocytes to be the larger population of cells, while other clusters consisted of cholangiocytes, macrophages (Kupffer cells), natural killer-T (NK-T) cells, endothelial cells and hepatic stellate cells (**Fig. S2C**). In agreement with previous data, inspection of PBMC clustering revealed large populations of B cells, $CD4^+$, $CD8^+$ naïve and $CD8^+$ memory T cells, together with smaller populations of natural killer (NK) cells, dendritic cells, $CD16^+$ and $CD14^+$ monocytes (**Fig. S2D**). We clustered the neocortex and found excitatory neurons, astrocytes, microglia, parvalbumin (PVALB), somatostatin-expressing neurons (SST), synaptic vesicle glycoprotein-expressing cells (SV2C), vasoactive intestinal polypeptide-expressing neurons (VIP), oligodendrocytes and oligodendrocyte precursor cells (**Fig S2E**). Parotid gland instead was composed of a large cluster of serous acinar cells together with small clusters of macrophages, myoepithelial cells, striated duct cells, mucous acinar cells and intercalated duct cells (**Fig. S2F**). Aorta cells could be further divided into adipocytes, endothelial cells, myofibroblasts and a large proportion of smooth muscle cells (**Fig. S2G**). Our clustering also demonstrated that most of the thyroid gland is composed of follicular cells, with smaller populations of adipocytes, endothelial cells, stromal and smooth muscle cells (**Fig. S2H**). Finally, our data showed the largest population of the pancreas to be acinar cells, while smaller clusters were comprised of macrophages, ductal, alpha and beta cells (**Fig. S2I**). In conclusion, we have successfully profiled the transcriptome of nine organs at a single-cell resolution in monkey, which could assist in the study of COVID-19.

***ACE2* and *TMPRSS2* single-cell expression landscape in a non-human primate**

Recent studies have reported that, similarly to SARS-CoV, the capacity of SARS-CoV-2 virus to infect host cells relies on viral spike (S) protein binding to ACE2 entry receptor (Hoffmann et

al., 2020; Walls et al., 2020), which is involved in the control of blood pressure through the renin-angiotensin system (Turner et al., 2004). This phenomenon is primed by the multifunction serine protease TMPRSS2 (Kim et al., 2006). Accordingly, double positive ($ACE2^+$/$TMPRSS2^+$) cells have a higher risk of infection by SARS-CoV-2. Although immunohistological studies have demonstrated localization of these two proteins in the respiratory tract (Jia et al., 2005), it is unclear how many cell subtypes express these genes and how homogenous the expression among a specific cell subtype is. Also, comprehensive information about other cell types and organs that express these two proteins and could be targeted by the virus in human or monkey is lacking.

We inspected our data to see how widespread and homogenous *ACE2* expression was in the monkey tissues. As expected, *ACE2* was detected in several lung clusters, such as ciliated cells, club cells and pulmonary alveolar type 2 cells (**Fig. 2A, 3A upper panel**). In the kidney, *ACE2* was primarily present in proximal tubule cells, though interestingly, the expression was heterogenous (**Fig. 2A, 3B upper panel**). This is consistent with reports that a significant number of COVID-19 patients display altered kidney function (Li et al., 2020; Naicker et al., 2020). In the liver, *ACE2* was mostly expressed in cholangiocytes, with a smaller degree of expression also found in hepatocytes (**Fig. 2A, 3C upper panel**). Notably, the closely related SARS-CoV caused liver injury due to hepatitis in some patients (Peiris et al., 2003), suggesting that the liver may also be a direct target for SARS-CoV-2. Likewise, ACE2 was detected in follicular cells within the thyroid tissue (**Fig. S3**). In contrast, little or no expression was observed in neocortex, pancreas, parotid and PBMC (**Fig. 2A, S3A**).

On the other hand, *TMPRSS2* displayed widespread expression across multiple tissues, although it was highest in kidney cells. However, in contrast to *ACE2*, its expression was highest in the distal convoluted tubule, thin limb, intercalated and principal cell 1 and 2 kidney clusters (**Fig. 2B, 3B lower panel, S3B**). Additionally, significant *TMPRSS2* was observed in both parotid and pancreatic acinar cells, follicular cells and in several lung clusters (**Fig. 2B**). We then determined which cells co-expressed both genes ($ACE2^+$/$TMPRSS2^+$). The largest overlap was observed in the ciliated and club cell clusters of the lung and the proximal tubule cells of the kidney. A smaller overlap was observed in hepatocytes and pancreatic cells (**Fig. 2C, 2E, 3A lower panel, 3C lower panel**).

Therefore, our data show that *ACE2* and *TMPRSS2* are expressed in a variety of cell types, mainly epithelial cells, within the nine monkey organs/tissues (**Supplemental table 2A**). They

also suggest that variations in cell state (e.g. differentiation state, stimulation state or topographical distribution) cause heterogenous expression across an individual tissue. These observations may provide important clues about COVID-19 pathogenesis and symptomatology.

Comparative analysis of *ACE2* and *TMPRSS2* expression in human and non-human primate

Given the heterogeneous nature of *ACE2* and *TMPRSS2* expression in monkey tissues, we decided to investigate similarities and differences between human and monkey. For this purpose, we retrieved publicly available data from single-cell studies in human (see methods). *TMPRSS2* distribution was similar in cell subtypes of lung, kidney and liver between human and monkey (**Fig. 3D-3F**). However, strikingly, *ACE2* showed distinct patterns among cell subtypes in all three organs between the two species (**Fig. 3D-3F**). The biggest differences were observed in ciliated cells of the lung, which had the highest expression of *ACE2* in monkey, and pulmonary alveolar type 2 cells, which had the highest expression of *ACE2* in human. The function of ciliated cells is to move substances (e.g. cell debris and toxic material) across the surface of the respiratory tract and are commonly targeted by respiratory viruses, whereas pulmonary alveolar type 2 cells have regenerative properties, are crucial for alveolar homeostasis and produce surfactant (Hamm et al., 1992; Mason and Williams, 1977). In the kidney of both monkey and human, the highest *ACE2* expression was in proximal tubule cells (**Fig. 3E**), which are responsible for electrolyte and nutrient reabsorption. However, renal endothelial cells had higher expression in monkey compared to human. In liver, cholangiocytes had similarly high *ACE2* expression in monkey and human, but hepatocytes showed higher expression and more positive cells in the human (**Fig. 3F**). After discovering heterogenous expression of *ACE2* within the proximal tubule cells, we revisited the previously analyzed data and were able to sub cluster this population of cells into two (S1 and S3) based on the expression of *SLC5A2* and *SLC7A13* (Lee et al., 2015) (**Supplemental table 2B, Fig. S4**). These two genes are sodium and glucose cotransporters involved in glucose reabsorption in the kidney (Santer and Calado, 2010; Yu et al., 2011). We did not include thyroid, pancreas or aorta in these analyses because of lack of high-quality available human single-cell datasets. As for the neocortex and PBMC, they have little to no expression of *ACE2* in human (data not shown).

These differences in *ACE2* expression across cell subtypes in the lung, kidney and liver in monkey and human raise the possibility that infection with SARS-CoV-2 in the two species will have different effects.

ACE2 correlation analysis across cell types reveals potentially therapeutic targets

To shed light on potential mechanisms that could facilitate *ACE2*-mediated SARS-CoV-2 infection, we performed an analysis of the Pearson's correlation coefficient, based on gene expression in the 44 cell subtypes, to determine what genes are co-regulated with *ACE2* in monkey tissues. Correlated genes were considered as those displaying a coefficient higher than 0.6 with an adjusted p-value < 0.001. Using these criteria, we observed several genes with marked correlation, including genes that belong to metabolic and developmental pathways and genes involved in the cellular response to xenobiotic stimuli (**Fig. 4A, B**). The highest correlation was observed for transmembrane protein 27 (*TMEM27*, cor = 0.84), a protein involved in trafficking amino acid transporters to the apical brush border of kidney proximal tubules (Chu et al., 2017). This is unsurprising considering that *TMEM27* is an important paralog of *ACE2*, and high expression was restricted to kidney cells. DnaJ heat shock protein family (Hsp40) member C12 (*DNAJC12*, cor = 0.78), a gene with a role in immune response processes (Sigdel et al., 2013), had a distribution like *TMEM27*. Importantly, we also observed high correlation with Indoleamine 2,3-dioxygenase 2 (*IDO2*, cor = 0.77), a gene with abundant expression in kidney and liver cells that was also expressed in the lung and other organs. *IDO2* has been shown to function during the early phases of immune responses and to mediate inflammatory autoimmunity (Ball et al., 2009; Lepiller et al., 2015). *ANPEP*, which encodes for alanyl aminopeptidase N, was also co-expressed with *ACE2* in kidney, liver and to a lesser extent in lung too (cor = 0.64), like *IDO2* (**Fig. 4C, D**). Interestingly, *ANPEP* has also been shown to be participate in immune responses, virus receptor activity and in mediating virus entry into host cells (Delmas et al., 1992; Wentworth and Holmes, 2001).

These data highlight potential therapeutic targets to help in the fight against SARS-CoV-2. Due to their potential co-regulation with ACE2, DNAJC12 and ANPEP it is possible that they modulate, or are directly involved in viral entry. Alternatively, depletion of cells expressing *IDO2* and *ANPEP,* through a cytopathic effect of the virus, could trigger an uncontrolled immune response and contribute to the immune cell exhaustion observed in COVID-19 (Guo et al., 2020).

Epigenetic regulation of *ACE2* in each cell subtype of the monkey kidney

To understand whether epigenetic mechanisms underlie the heterogeneity of cellular composition and cell state variation in the kidney, we employed DNBelab C4 technology to perform high-throughput single-cell ATAC-seq (**Fig. 5A**). After filtering, 6,353 nuclei were used for downstream analysis (**Supplemental table 4, Fig. S5A, B**). We integrated these data with the kidney transcriptomic data described in **Fig. 1** and proceeded to perform Louvain clustering to map all the different cell types within the dataset (**Fig. 5B**). Consistent with the transcriptomic data, our epigenomic mapping identified thick ascending limb cells and proximal tubule cells as the largest kidney clusters (**Fig. S2B**). Similarly, smaller clusters of podocytes, principal, intercalated, connected tubule, distal convoluted tubule, thin limb, endothelial and stromal cells were detected (**Fig. 5C, S5C**). Analysis of open chromatin regions revealed discrete peaks in the *ACE2* locus, with the highest signal detected in proximal tubule cells S1 and S3, which are also the highest *ACE2*-expressing cells (**Fig. 5D**). Our approach failed to detect significant signal enrichment in the *ACE2* locus in endothelial cells, possibly related to the low level of expression (**Fig. 5D**). Within the cells of the kidney we observed the highest percentage of $ACE2^{+}$ cells in the proximal tubule S3, with a lower percentage in the proximal tubule S1 and endothelial cells (**Fig. 5E**). Motif analysis within the open chromatin regions in $ACE2^{+}$ cells demonstrated that these regions were preferentially enriched in signal transducer and activator of transcription 1 and 3 (STAT1 and 3) and interferon regulatory factor 1 (IRF1) binding sites (**Fig. 5F**). These findings suggested that tissue protective cytokines including IL5, IL6, EGF and interferons are acting on these proximal tubule cells S3 to induce *ACE2*. We focused on IL6 because a recent clinical trial has been started that uses anti-IL6 receptor (IL6R) antibodies in the treatment of COVID-19 (http://www.chictr.org.cn/showprojen.aspx?proj=49409). IL6 is a potent regulator of immune responses and can be produced by a variety of interstitial cells including fibroblasts, endothelial cells and more importantly tissue macrophages (Heinrich et al., 1990). Interestingly, we also noticed that distribution of *IL6R* correlated well with *ACE2* in proximal tubule cells (**Fig. 5G, Supplemental Fig. S5D**). In human kidney a similar co-expression pattern was detected (**Fig. 5H**).

Our observations suggest a potential positive feedback loop between IL6 and ACE2 expression that can exacerbate COVID-19 disease progression due to increased viral entry and dissemination.

DISCUSSION

Mammalian tissues and organs are composed of many different cell types that can vary in abundance and cell state. Tissue heterogeneity is only beginning to be unraveled thanks to the advent of single-cell profiling technologies that allow us to precisely map transcriptomic and epigenomic programs. These technologies are revolutionizing our view of human physiology and disease. Great efforts are being made to generate the first version of both human and murine atlases (Han et al., 2018; Han et al., 2020); the mouse is among the most commonly used model organisms in biomedical research, but many developmental or pathological aspects are not paralleled in human. Understanding tissue and organ complexity in species that are phylogenetically close to humans is an unmet requirement.

In this study, we have generated a single-cell transcriptomic atlas of nine organs (liver, kidney, lung, pancreas, neocortex, aorta, parotic gland, thyroid and peripheral blood) from cynomolgus monkey. We used this dataset not only to provide fundamental information about the cellular composition of the different tissues tested but also as a platform to dissect the overall expression distribution of the SARS-CoV-2 entry receptor, ACE2, and its serine protease coactivator TMPRSS2 (Hoffmann et al., 2020; Walls et al., 2020). Interestingly, ACE2 was expressed in multiple epithelial tissues besides the lung, specifically the kidney, liver and thyroid. Other organs of epithelial origin such as the gut have also been implicated in the pathogenesis of the disease (Ong et al., 2020). A consequence of this is that the SARS2-CoV-2 virus could infect these organs too, which would explain some of the reported clinical manifestations of COVID-19 (Zhu et al., 2020). By comparing our dataset with publicly available human single-cell RNA-seq data, we have also uncovered significant differences between human and monkey. We showed different expression patterns for *ACE2* in the lung, where the highest levels were detected in ciliated cells in monkey and pulmonary alveolar type 2 cells in human. Similarly, we observed marked differences in liver, in which monkey hepatocytes displayed higher *ACE2* and a larger number of positive cells compared to the human. We do not know whether these differences will affect the pathogenesis of COVID-19 between these two species. Nevertheless, this is a relevant finding considering that monkeys are a preferred model for studying the effectiveness of drug treatments and of vaccines against the impending COVID-19 pandemic.

Through correlation analysis, we identified new potential mechanisms that could facilitate ACE2-mediated viral infection, including genes previously unreported in the context of SARS-CoV-2 that are involved in stimulating different types of immune responses. We observed high expression of *IDO2* and *ANPEP* in kidney, liver and lung. Expression of these can be further induced by viral infection and have been reported to be immune modulators and/or mediate viral entry (Ball et al., 2009; Delmas et al., 1992). These observations are relevant as they highlight new potential therapeutic vulnerabilities in the current emergency. In this respect, a number of inhibitors of ANPEP are currently being developed and could serve to prevent the immune cell exhaustion often observed in many severe COVID-19 cases (Zheng et al., 2020). Similarly, mesenchymal stem/stroma cells (MSC) have immunomodulatory functions that are partly related to IDO2 production (Ball et al., 2009). It is tempting to speculate that cell therapies based on MSC could ameliorate COVID-19 by normalizing immune function and preventing cytokine storms (Wilson et al., 2015).

Intriguingly, in our data, we see heterogenous expression of *ACE2* within the individual cell subtypes in six out of the nine monkey organs that we analyzed, which is also observed in the three human organs analyzed. Interestingly, we noticed two different cell populations in the kidney proximal tubule, one with higher *ACE2* expression than the other. We performed single-cell ATAC-seq of this organ to understand whether this phenomenon has an epigenetic basis. Analysis of open chromatin regions within the ACE2 locus revealed the enrichment of STAT1, STAT3 and IRF1 binding sites. These transcription factors have important immune functions and are direct targets of tissue protective and innate immune responses such as interleukin-6 signaling pathway and interferons. Analysis of IL6R distribution showed broad expression within different the *ACE2*$^{+}$ organs in monkey and human. This suggests a link between paracrine IL6 (e.g. secreted by stromal cells including tissue resident macrophages) and enhanced *ACE2* expression across different organs. Higher and more widespread ACE2 expression could promote increased viral entry. This observation could be of the utmost importance given recent reports describing clinical trials with Tocilizumab, a monoclonal antibody used for IL6R blockade in patients with rheumatoid arthritis (Villiger et al., 2016), for the treatment of COVID-19 (http://www.chictr.org.cn/showprojen.aspx?proj=49409). IL6 has been related to aging and tissue damage (Mosteiro et al., 2018), and this may explain why elderly individuals and those with underlying inflammatory conditions have more severe reactions to SARS-CoV-2 infection (**Fig. 6**). Importantly, high IL6 levels have been detected in

383 plasma from COVID-19 patients (Wan et al., 2020). The proposed enhanced affinity of SARS-
384 CoV-2 for ACE2 compared to SARS-CoV may underlie the differences in the clinical course
385 between the two diseases (Shang et al., 2020).
386 All these observations revealed new potential mechanisms for COVID-19, opening new
387 therapeutic avenues for disease management. However, caution should be exercised when
388 making decisions before additional experimental validation becomes available.
389

METHODS

Ethics statement

This study was approved by the Institutional Review Board on Ethics Committee of BGI (permit no. BGI-IRB19125).

Collection of monkey tissues

A 6-year old female cynomolgus monkey was purchased from Huazhen Laboratory Animal Breeding Centre (Guangzhou, China). The monkey was anesthetized with ketamine hydrochloride (10 mg/kg) and barbiturate (40 mg/kg) administration before being euthanized by exsanguination. Tissues were isolated and placed on the ice-cold board for dissection. Whole organs including lung, kidney, pancreas, liver, brain, thyroid, parotid gland, and aorta were cut into 5-10 pieces, respectively (50-200 mg/piece). Samples were then quickly frozen in liquid nitrogen and stored until nuclear extraction was performed. PBMC were isolated from heparinized venous blood using a Lymphoprep™ medium (STEMCELL Technologies, #07851) according to standard density gradient centrifugation methods. PBMC were resuspended in 90% FBS, 10% DMSO (Sigma Aldrich, #D2650) freezing media and frozen using a Nalgene® Mr. Frosty® Cryo 1°C Freezing Container (Thermo Fisher Scientific, #5100-0001) in a -80°C freezer for 24 hours before being transferred to liquid nitrogen for long-term storage.

Single-nucleus/cell suspension preparation

We isolated nuclei as previously described (Bakken et al., 2018). Briefly, tissues were thawed, minced and added to lysis buffer. Lysates were filtered and resuspended in cell resuspension buffer. Frozen PBMC vials were rapidly thawed in a 37°C water bath for ~2 minutes, then quenched with 10 ml 37°C pre-warmed 1X phosphate-buffered saline (PBS, Thermo Fisher Scientific, #10010031) supplemented with 10% FBS. PBMCs were centrifuged at 500 *R.C.F.* for 5 minutes at room temperature. The supernatant was removed, and the cell pellet was resuspended in 3 ml 37°C pre-warmed 1X PBS containing 0.04% bovine serum albumin (BSA, Sangon Biotech, A600903), passed through a 40 µm cell strainer (Falcon, #352340) and then centrifuged at 500 *R.C.F.* for 5 minutes at room temperature. Nuclei or

421 cells were resuspended with cell resuspension buffer at a concentration of 1,000 cells/μl for
422 single-cell library preparation.
423
424 **Single-nucleus/cell sequencing**
425 The DNBelab C Series Single-Cell Library Prep Set (MGI, #1000021082) was utilized as
426 previously described (Liu et al. 2019). In brief, single-nucleus/cell suspensions were used for
427 droplet generation, emulsion breakage, beads collection, reverse transcription, and cDNA
428 amplification to generate barcoded libraries. Indexed single-cell RNA-seq libraries were
429 constructed according to the manufacturer's protocol. The sequencing libraries were
430 quantified by Qubit™ ssDNA Assay Kit (Thermo Fisher Scientific, #Q10212). Single-cell ATAC-
431 seq libraries were prepared using DNBelab C4 Single-Cell ATAC-seq Library Prep Set (MGI).
432 DNA nanoballs (DNBs) were loaded into the patterned Nano arrays and sequenced on the
433 ultra-high-throughput DIPSEQ T1 sequencer using the following read length: 30 bp for read 1,
434 inclusive of 10 bp cell barcode 1, 10 bp cell barcode 2 and 10 bp unique molecular identifier
435 (UMI), 100 bp of transcript sequence for read 2, and 10 bp for sample index.
436
437 **Single-cell RNA-seq data processing**
438 Raw sequencing reads from DIPSEQ-T1 were filtered and demultiplexed using PISA (version
439 0.2) (https://github.com/shiquan/PISA). Reads were aligned to Macaca_fascicularis_5.0
440 genome using STAR (version 2.7.4a) (Dobin et al., 2013) and sorted by sambamba (version
441 0.7.0) (Tarasov et al., 2015). Cell versus gene UMI count matrix was generated with PISA.
442
443 **Cell clustering and identification of cell types**
444 Clustering analysis of the complete cynomolgus monkey tissue dataset was performed using
445 Scanpy (version 1.4) (Wolf et al., 2018) in a Python environment. Parameters used in each
446 function were manually curated to portray the optimal clustering of cells. In preprocessing,
447 cells or nuclei were filtered based on the criteria of expressing a minimum of 200 genes and
448 a gene which is expressed by a minimum of 3 cells or nuclei. Filtered data were ln (counts per
449 million (CPM)/100 + 1) transformed. 3000 highly variable genes were selected according to
450 their average expression and dispersion. The number of UMI and the percentage of
451 mitochondrial gene content were regressed out and each gene was scaled by default options.
452 Dimension reduction starts with principal component analysis (PCA), and the number of

principal components used for UMAP depends on the importance of embeddings. The Louvain method is then used to detect subgroups of cells. Distinguishing differential genes among clusters were ranked (Benjamini-Hochberg, Wilcoxon rank-sum test). Cell types were manually and iteratively assigned based on overlap of literature, curated and statistically ranked genes.

Each tissue dataset was portrayed using the Seurat package (version 3.1.1) (Stuart et al., 2019) in R environment by default parameters for filtering, data normalization, dimensionality reduction, clustering, and gene differential expression analysis. Finally, we annotated each cell type by extensive literature reading and searching for the specific gene expression pattern.

Gene correlation and Gene Ontology (GO) term enrichment analysis

The correlation between *ACE2* and other genes was drawn using Pearson correlation coefficient (PCC) with gene expression value merged from cells of the same cell types with the R package psych (version 1.9.12.31). To infer the biological function of highly correlated genes (cor > 0.6 and adjusted P value < 0.001), we performed gene set enrichment analysis using Metascape (https://metascape.org/gp/index.html).

Differential gene expression analysis

Differential expression analysis between proximal tubule S1 and proximal tubule S3 was performed using the FindMarkers function of the Seurat package (version 3.1.1).

Single-cell ATAC-seq data processing

Raw sequencing reads from DIPSEQ-T1 were filtered and demultiplexed using PISA (version 0.2) (https://github.com/shiquan/PISA). Peak calling was performed using MACS2 (version 2.1.2) (Feng et al., 2012) with options -f BAM -B -q 0.01 –nomodel. The cell versus peak reads count matrix was generated by custom script. To create a gene activity matrix, we extracted gene coordinates for cynomolgus monkey from NCBI, and extended them to ±2 kb region around TSS. The gene activity score matrix was calculated by custom script.

Single-cell ATAC-seq cell clustering and cell type identification

Cells with low fragments (<1000) and TSS proportion (<0.1) were removed. Then, filtered data were imported into R and the dimensionality was reduced by latent semantic indexing. Anchors between single-cell ATAC-seq and single-cell RNA-seq datasets were identified and used to transfer cell type labels identified from the single-cell RNA-seq data. We embedded the single-cell ATAC-seq and single-cell RNA-seq cells by the TransferData function of Seurat (version 3.1.1).

Transcription factor motif enrichment analysis

To predict the motif footprint in peaks within the *ACE2* promoter, we extracted genome sequences in the peak region with Seqkit (version 0.7.0) (Shen et al., 2016). The sequences were imported into R and were matched with all *Homo sapiens* motifs form JASPAR2018 using matchMotifs function in motifmatchr packages version 1.8.0 with default parameter.

Human single-cell RNA-seq datasets

All human single-cell RNA-seq data matrix were obtained from publicly available dataset as described: (1) Kidney data from Stewart et al. was download from https://www.kidneycellatlas.org/ (Stewart et al., 2019); (2) Lung data from Madissoon et al. was download from https://www.tissuestabilitycellatlas.org/ (Madissoon et al., 2019); (3) Liver data from Aizarani et al. was download from GEO at accession GSE124395 (Aizarani et al., 2019).

Code availability

Computer code used for processing the single-cell RNA-seq and single-cell ATAC-seq will be available upon request.

Data availability

All raw data have been deposited to CNGB Nucleotide Sequence Archive and the accession number will be available upon request.

Acknowledgements

We thank all members of the single-cell omics lab (BGI) and Esteban lab (GIBH) for their support. This work was supported by National Natural Science Foundation of China

(31900466, 31900582), Research and Development Program of China (2018YFA0106903), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16030502) and CAS-JSPS Joint Research Projects (GJHZ2093), Natural Science Foundation of Guangdong Province, China (2018A030313379), the Shenzhen National Key Laboratory of Single-Cell Omics (ZDSYS20190902093613831) and Shenzhen Bay Laboratory (SZBL2019062801012). G.V. is supported by Chinese Academy of Sciences President's International Fellowship for Foreign Experts (2020FSB0002). C.W. is supported by Chinese Academy of Sciences President's International Fellowship Initiative for Postdoctoral Researchers (2019PB0177) and Research Fund for International Young Scientists grant (31950410553).

Author contributions

L.L., M.A.E., Y.H. and X.X. conceived the idea. L.L., L.H., Y.L., S.L., X.W. and Y.Yuan. designed the experiment. L.H., Y.L., S.L., X.W., Y.Y., M.C. and C.W.W. collected the tissue samples. C.L., Z.W., Y.Yuan, Y. Yu, M.W., T.W., Y.L., C.W., Y.Z., T.T., Y.H., H.L., L.X., J.X. and M.C. performed the experiments. X.W., T.P., Q.S., L.W., Z.Z., Y.L., S.Z. and S.L. performed the data analysis. L.L., L.H., X.W., C.L., G.V., T.P., C.W. and Y.L. prepared the figures. P.H.M provided critical review of the manuscript. M.A.E., G.V., C.W., Y.L. and L.L. wrote the manuscript with input from all authors. X.X., Y.H., L.L. and M.A.E supervised the entire study. All other authors contributed to the work. All authors read and approved the manuscript for submission.

Competing interests

Employees of BGI have stock holdings in BGI.

FIGURE LEGENDS

Figure 1. Construction of single-cell atlas across nine tissues of a *Macaca fascicularis* monkey

(**A**) Schematic representation of selected monkey tissues used in this study and description of experimental pipeline for the single-cell sequencing.

(**B**) UMAP visualization of all single cells from the dataset colored by tissue/organ (left) and number of cells from each tissue passing quality control (right).

(**C**) UMAP visualization of each cell type colored according to 44 clusters in the first round of clustering. Cell type annotation is provided in the figure and is associated with a number indicative of every cluster. n = 215,334 individual nuclei/cells.

Figure 2. *ACE2* and *TMPRSS2* expression across 44 cell clusters in monkey

(**A-B**) UMAP projection of *ACE2* (**A**) and *TMPRSS2* (**B**) expression in all single cells within our dataset.

(**C**) UMAP projection of $ACE2^{+}$/$TMPRSS2^{+}$ cells.

(**D**) Barplot indicating the percentage of *ACE2* and *TMPRSS2* expressing cells within each cell cluster.

(**E**) Bubble plots showing the level of expression of *TMPRSS2* and *ACE2* genes and the ratio of expressing cells in the indicated cell types. The color of each bubble represents the level of expression and the size indicates the proportion of expressing cells.

Figure 3. Comparative analysis of *ACE2* and *TMPRSS2* expression between monkey and human

(**A-C**) UMAP projection of *ACE2* (top) and *TMPRSS2* (bottom) expression in single cells of monkey lung (**A**), kidney (**B**) and liver tissues (**C**). The red arrow in this panel indicates cholangiocytes. The color of the cells reflects the expression level as indicated in the scale bar. (**D-F**) Bubble plots showing the ratio and expression of *ACE2 and TMPRSS2* in the indicated cell types of lung (**D**), kidney (**E**) and liver (**F**) in monkey and human. The color of each bubble represents the level of expression and the size indicates the proportion of expressing cells.

Figure 4. Co-expression analysis of *ACE2*

(**A**) Volcano plot of correlation coefficients (Pearson r^2) from association tests between *ACE2* and other individual genes. The correlation coefficient for every gene (x-axis) versus the adjusted *P*-value (using Benjamini-Hochberg correction; y-axis). The genes indicated in the plot are those with a correlation score > 0.6 and an adjusted *P*-value < 0.001.

(**B**) Gene ontology analysis of genes that show high expression correlation with *ACE2*.

(**C**) Scatter plots showing the association between *ACE2* and the indicated genes. The correlation coefficients (Pearson r^2) and adjusted *P*-values are shown in the plots.

(**D**) UMAP projection of expression of the indicated genes in all single cells.

Figure 5. Chromatin accessibility analysis reveals epigenetic regulation of *ACE2* in kidney

(**A**) Schematic of experimental design for single-cell ATAC-seq of monkey.

(**B**) Joint UMAP visualization of kidney single-cell ATAC-seq data with single-cell RNA-seq data.

(**C**) UMAP visualization of single-cell ATAC-seq data.

(**D**) IGV visualization of aggregate single-cell ATAC-seq signal in each cell type.

(**E**) Ratio of ACE^+ cells in each cell type of kidney.

(**F**) The transcription factor motifs predicted based on DNA sequence within those regions of the *ACE2* locus.

(**G-H**) UMAP projection of IL6R expression and cells with $IL6R^+/ACE2^+$ cells in all kidney single cell in monkey (**G**) and human (**H**).

Figure 6. Proposed molecular mechanism for SARS-Cov-2 pathogenesis through reinforced IL6-mediated immune response in monkey and humans.

Schematic representation of potential mechanism of SARS-CoV-2 spreading through lung, kidney and liver. Kidney proximal tubule cells within the nephron have the highest expression of ACE2 receptor which facilitates virus entry. After virus contact, IL6R stimulates an immune response that, through the activation of STATs factors, potentiates the paracrine positive feedback loop that facilitates virus spreading. IL6 expression, which is higher in elderly patients and those with inflammatory conditions, is effectively targeted by anti-IL6R monoclonal antibodies leading to a more favourable disease course.

603 **SUPPLEMENTAL FIGURE LEGENDS**

604

605 **Supplemental Figure 1. Quality control of the single-cell RNA-seq libraries**

606 (**A**) Violin plot showing the number of unique molecular identifiers (UMIs) identified in each
607 tissue.

608 (**B**) Violin plot showing the number of genes identified in each organ. (**C**) Heatmap showing
609 the expression of marker genes of the indicated cell type

610

611 **Supplemental Figure 2. Various cell types identified in each tissue**

612 (**A-I**) UMAP visualization of cell clusters in lung (**A**), kidney (**B**), liver (**C**), PBMC (**D**), neocortex
613 (**E**), parotid (**F**), aorta (**G**), thyroid (**H**) and pancreas (**I**). The name of the population
614 corresponding to each cell cluster is indicated in every plot.

615

616 **Supplemental Figure 3. *ACE2* and *TMPRSS2* expression in each tissue**

617 (**A-B**) UMAP Projection of *ACE2* (**A**) and *TMPRSS2* (**B**) expression in each tissue.

618

619 **Supplemental Figure 4. Specific accessible chromatin in each cluster of kidney**

620 (**A**) UMAP visualization of single cells from the kidney tissue, colored by cell types.

621 **(B)** Volcano plot showing the differentially expressed genes between proximal tubule S1 and
622 proximal tubule S3 cells. Examples of highly variable genes are indicated.

623 **(C)** UMAP projection of expression for the indicated genes in all single cells.

624 **(D)** The structure and specific gene expression in kidney tubules. The specific genes and *ACE2*
625 expression level for proximal tubule S1 and proximal tubule S3 cells are indicated.

626

627 **Supplemental Figure 5. Accessible chromatin analysis in each cluster of kidney**

628 (**A**) Number of fragments captured in all cells of the two single-cell ATAC-seq libraries.

629 (**B**) Proportion of TSS fragments in all cells of the two single-cell ATAC-seq libraries.

630 (**C**) IGV visualization of specific accessible chromatin in each cell type.

631 (**D**) UMAP projection of *ACE2* expression in human kidney.

632

633

634 **REFERENCES**

635

636 Aizarani, N., Saviano, A., Sagar, Mailly, L., Durand, S., Herman, J.S., Pessaux, P., Baumert, T.F.,
637 and Grun, D. (2019). A human liver cell atlas reveals heterogeneity and epithelial progenitors.
638 Nature *572*, 199-204.
639 Bakken, T.E., Hodge, R.D., Miller, J.A., Yao, Z., Nguyen, T.N., Aevermann, B., Barkan, E.,
640 Bertagnolli, D., Casper, T., Dee, N.*, et al.* (2018). Single-nucleus and single-cell transcriptomes
641 compared in matched cortical cell types. PLoS One *13*, e0209648.
642 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M.,
643 and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.
644 Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using
645 MACS. Nat Protoc *7*, 1728-1740.
646 Liu C, Wu T, Fan F, et al. (2019). A portable and cost-effective microfluidic system for massively
647 parallel single-cell transcriptome profiling. BioRxiv. 2019:818450.
648 Madissoon, E., Wilbrey-Clark, A., Miragaia, R.J., Saeb-Parsy, K., Mahbubani, K.T.,
649 Georgakopoulos, N., Harding, P., Polanski, K., Huang, N., Nowicki-Osuch, K.*, et al.* (2019).
650 scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold
651 preservation. Genome Biol *21*, 1.
652 Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for
653 FASTA/Q File Manipulation. PLoS One *11*, e0163962.
654 Stewart, B.J., Ferdinand, J.R., Young, M.D., Mitchell, T.J., Loudon, K.W., Riding, A.M., Richoz,
655 N., Frazer, G.L., Staniforth, J.U.L., Vieira Braga, F.A.*, et al.* (2019). Spatiotemporal immune
656 zonation of the human kidney. Science *365*, 1461-1466.
657 Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y.,
658 Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell
659 Data. Cell *177*, 1888-1902 e1821.
660 Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast
661 processing of NGS alignment formats. Bioinformatics *31*, 2032-2034.
662 Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression
663 data analysis. Genome Biol *19*, 15.
664 Aizarani, N., Saviano, A., Sagar, Mailly, L., Durand, S., Herman, J.S., Pessaux, P., Baumert, T.F.,
665 and Grun, D. (2019). A human liver cell atlas reveals heterogeneity and epithelial progenitors.
666 Nature *572*, 199-204.
667 Bagheri, S.H.R., Asghari, A.M., Farhadi, M., Shamshiri, A.R., Kabir, A., Kamrava, S.K., Jalessi,
668 M., Mohebbi, A., Alizadeh, R., Honarmand, A.A.*, et al.* (2020). Coincidence of COVID-19
669 epidemic and olfactory dysfunction outbreak. medRxiv, 2020.2003.2023.20041889.
670 Bakken, T.E., Hodge, R.D., Miller, J.A., Yao, Z., Nguyen, T.N., Aevermann, B., Barkan, E.,
671 Bertagnolli, D., Casper, T., Dee, N.*, et al.* (2018). Single-nucleus and single-cell transcriptomes
672 compared in matched cortical cell types. PLoS One *13*, e0209648.
673 Ball, H.J., Yuasa, H.J., Austin, C.J., Weiser, S., and Hunt, N.H. (2009). Indoleamine 2,3-
674 dioxygenase-2; a new enzyme in the kynurenine pathway. The international journal of
675 biochemistry & cell biology *41*, 467-471.
676 Bao, L., Deng, W., Gao, H., Xiao, C., Liu, J., Xue, J., Lv, Q., Liu, J., Yu, P., Xu, Y.*, et al.* (2020).
677 Reinfection could not occur in SARS-CoV-2 infected rhesus macaques. bioRxiv,
678 2020.2003.2013.990226.

679 Chan, J.F., Yuan, S., Kok, K.H., To, K.K., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C.C., Poon, R.W., *et al.* (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus
680 indicating person-to-person transmission: a study of a family cluster. Lancet *395*, 514-523.
681 Chu, P.L., Gigliotti, J.C., Cechova, S., Bodonyi-Kovacs, G., Chan, F., Ralph, D.L., Howell, N.,
682 Kalantari, K., Klibanov, A.L., Carey, R.M., *et al.* (2017). Renal Collectrin Protects against Salt-
683 Sensitive Hypertension and Is Downregulated by Angiotensin II. J Am Soc Nephrol *28*, 1826-
684 1837.
685 Conti, P., Ronconi, G., Caraffa, A., Gallenga, C.E., Ross, R., Frydas, I., and Kritas, S.K. (2020).
686 Induction of pro-inflammatory cytokines (IL-1 and IL-6) and lung inflammation by
687 Coronavirus-19 (COVI-19 or SARS-CoV-2): anti-inflammatory strategies. J Biol Regul Homeost
688 Agents *34*.
689 Cui, J., Li, F., and Shi, Z.L. (2019). Origin and evolution of pathogenic coronaviruses. Nat Rev
690 Microbiol *17*, 181-192.
691 Delmas, B., Gelfi, J., L'Haridon, R., Vogel, L.K., Sjostrom, H., Noren, O., and Laude, H. (1992).
692 Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. Nature
693 *357*, 417-420.
694 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M.,
695 and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.
696 Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using
697 MACS. Nat Protoc *7*, 1728-1740.
698 Guo, Y.R., Cao, Q.D., Hong, Z.S., Tan, Y.Y., Chen, S.D., Jin, H.J., Tan, K.S., Wang, D.Y., and Yan,
699 Y. (2020). The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-
700 19) outbreak - an update on the status. Mil Med Res *7*, 11.
701 Hamm, H., Fabel, H., and Bartsch, W. (1992). The surfactant system of the adult lung:
702 physiology and clinical perspectives. Clin Investig *70*, 637-657.
703 Hamming, I., Timens, W., Bulthuis, M.L., Lely, A.T., Navis, G., and van Goor, H. (2004). Tissue
704 distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in
705 understanding SARS pathogenesis. J Pathol *203*, 631-637.
706 Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., *et*
707 *al.* (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. Cell *172*, 1091-1107 e1017.
708 Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., Chen, H., Wang, J., Tang, H., Ge, W., *et al.*
709 (2020). Construction of a human cell landscape at single-cell level. Nature.
710 Heinrich, P.C., Castell, J.V., and Andus, T. (1990). Interleukin-6 and the acute phase response.
711 The Biochemical journal *265*, 621-636.
712 Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens,
713 T.S., Herrler, G., Wu, N.H., Nitsche, A., *et al.* (2020). SARS-CoV-2 Cell Entry Depends on ACE2
714 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell.
715 Jia, H.P., Look, D.C., Shi, L., Hickey, M., Pewe, L., Netland, J., Farzan, M., Wohlford-Lenane, C.,
716 Perlman, S., and McCray, P.B., Jr. (2005). ACE2 receptor expression and severe acute
717 respiratory syndrome coronavirus infection depend on differentiation of human airway
718 epithelia. J Virol *79*, 14614-14621.
719 Kim, T.S., Heinlein, C., Hackman, R.C., and Nelson, P.S. (2006). Phenotypic analysis of mice
720 lacking the Tmprss2-encoded protease. Mol Cell Biol *26*, 965-975.
721 Kock, R.A., Begovoeva, M., Ansumana, R., and Suluku, R. (2019). Searching for the source of
722 Ebola: the elusive factors driving its spillover into humans during the West African outbreak
723 of 2013-2016. Rev Sci Tech *38*, 113-122.
724

Lau, S.K., Woo, P.C., Li, K.S., Huang, Y., Tsoi, H.W., Wong, B.H., Wong, S.S., Leung, S.Y., Chan, K.H., and Yuen, K.Y. (2005). Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A *102*, 14040-14045.

Lee, J.W., Chou, C.L., and Knepper, M.A. (2015). Deep Sequencing in Microdissected Renal Tubules Identifies Nephron Segment-Specific Transcriptomes. J Am Soc Nephrol *26*, 2669-2677.

Lepiller, Q., Soulier, E., Li, Q., Lambotin, M., Barths, J., Fuchs, D., Stoll-Keller, F., Liang, T.J., and Barth, H. (2015). Antiviral and Immunoregulatory Effects of Indoleamine-2,3-Dioxygenase in Hepatitis C Virus Infection. J Innate Immun *7*, 530-544.

Li, Z., Wu, M., Yao, J., Guo, J., Liao, X., Song, S., Li, J., Duan, G., Zhou, Y., Wu, X.*, et al.* (2020). Caution on Kidney Dysfunctions of COVID-19 Patients. medRxiv, 2020.2002.2008.20021212.

Liu, C., Wu, T., Fan, F., Liu, Y., Wu, L., Junkin, M., Wang, Z., Yu, Y., Wang, W., Wei, W.*, et al.* (2019). A portable and cost-effective microfluidic system for massively parallel single-cell transcriptome profiling. bioRxiv, 818450.

Madissoon, E., Wilbrey-Clark, A., Miragaia, R.J., Saeb-Parsy, K., Mahbubani, K.T., Georgakopoulos, N., Harding, P., Polanski, K., Huang, N., Nowicki-Osuch, K.*, et al.* (2019). scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. Genome Biol *21*, 1.

Mason, R.J., and Williams, M.C. (1977). Type II alveolar cell. Defender of the alveolus. Am Rev Respir Dis *115*, 81-91.

Mosteiro, L., Pantoja, C., de Martino, A., and Serrano, M. (2018). Senescence promotes in vivo reprogramming through p16(INK)(4a) and IL-6. Aging Cell *17*.

Naicker, S., Yang, C.W., Hwang, S.J., Liu, B.C., Chen, J.H., and Jha, V. (2020). The Novel Coronavirus 2019 epidemic and kidneys. Kidney Int.

Ong, J., Young, B.E., and Ong, S. (2020). COVID-19 in gastroenterology: a clinical perspective. Gut, Epub ahead of print.

Peiris, J.S., Lai, S.T., Poon, L.L., Guan, Y., Yam, L.Y., Lim, W., Nicholls, J., Yee, W.K., Yan, W.W., Cheung, M.T.*, et al.* (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet *361*, 1319-1325.

Reusken, C.B., Haagmans, B.L., Muller, M.A., Gutierrez, C., Godeke, G.J., Meyer, B., Muth, D., Raj, V.S., Smits-De Vries, L., Corman, V.M.*, et al.* (2013). Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. Lancet Infect Dis *13*, 859-866.

Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W.*, et al.* (2020). Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. The New England journal of medicine *382*, 970-971.

Santer, R., and Calado, J. (2010). Familial renal glucosuria and SGLT2: from a mendelian trait to a therapeutic target. Clin J Am Soc Nephrol *5*, 133-141.

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. Nature.

Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One *11*, e0163962.

Sigdel, T.K., Shoemaker, L.D., Chen, R., Li, L., Butte, A.J., Sarwal, M.M., and Steinberg, G.K. (2013). Immune response profiling identifies autoantibodies specific to Moyamoya patients. Orphanet journal of rare diseases *8*, 45.

770 Stewart, B.J., Ferdinand, J.R., Young, M.D., Mitchell, T.J., Loudon, K.W., Riding, A.M., Richoz,
771 N., Frazer, G.L., Staniforth, J.U.L., Vieira Braga, F.A*., et al.* (2019). Spatiotemporal immune
772 zonation of the human kidney. Science *365*, 1461-1466.
773 Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y.,
774 Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell
775 Data. Cell *177*, 1888-1902 e1821.
776 Su, S., Wong, G., Shi, W., Liu, J., Lai, A.C.K., Zhou, J., Liu, W., Bi, Y., and Gao, G.F. (2016).
777 Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. Trends Microbiol
778 *24*, 490-502.
779 Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast
780 processing of NGS alignment formats. Bioinformatics *31*, 2032-2034.
781 Turner, A.J., Hiscox, J.A., and Hooper, N.M. (2004). ACE2: from vasopeptidase to SARS virus
782 receptor. Trends Pharmacol Sci *25*, 291-294.
783 Villiger, P.M., Adler, S., Kuchen, S., Wermelinger, F., Dan, D., Fiege, V., Butikofer, L., Seitz, M.,
784 and Reichenbach, S. (2016). Tocilizumab for induction and maintenance of remission in giant
785 cell arteritis: a phase 2, randomised, double-blind, placebo-controlled trial. Lancet *387*, 1921-
786 1927.
787 Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D. (2020). Structure,
788 Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell, Epub ahead of print.
789 Wan, S., Yi, Q., Fan, S., Lv, J., Zhang, X., Guo, L., Lang, C., Xiao, Q., Xiao, K., Yi, Z*., et al.* (2020).
790 Characteristics of lymphocyte subsets and cytokines in peripheral blood of 123 hospitalized
791 patients with 2019 novel coronavirus pneumonia (NCP). medRxiv, 2020.2002.2010.20021832.
792 Wentworth, D.E., and Holmes, K.V. (2001). Molecular determinants of species specificity in
793 the coronavirus receptor aminopeptidase N (CD13): influence of N-linked glycosylation. J Virol
794 *75*, 9741-9752.
795 Wilson, J.G., Liu, K.D., Zhuo, H., Caballero, L., McMillan, M., Fang, X., Cosgrove, K., Vojnik, R.,
796 Calfee, C.S., Lee, J.W*., et al.* (2015). Mesenchymal stem (stromal) cells for treatment of ARDS:
797 a phase 1 clinical trial. Lancet Respir Med *3*, 24-32.
798 Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression
799 data analysis. Genome Biol *19*, 15.
800 Yu, L., Lv, J.C., Zhou, X.J., Zhu, L., Hou, P., and Zhang, H. (2011). Abnormal expression and
801 dysfunction of novel SGLT2 mutations identified in familial renal glucosuria patients. Hum
802 Genet *129*, 335-344.
803 Zheng, M., Gao, Y., Wang, G., Song, G., Liu, S., Sun, D., Xu, Y., and Tian, Z. (2020). Functional
804 exhaustion of antiviral lymphocytes in COVID-19 patients. Cell Mol Immunol, Epub ahead of
805 print.
806 Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R*., et al.*
807 (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. The New England
808 journal of medicine *382*, 727-733.
809

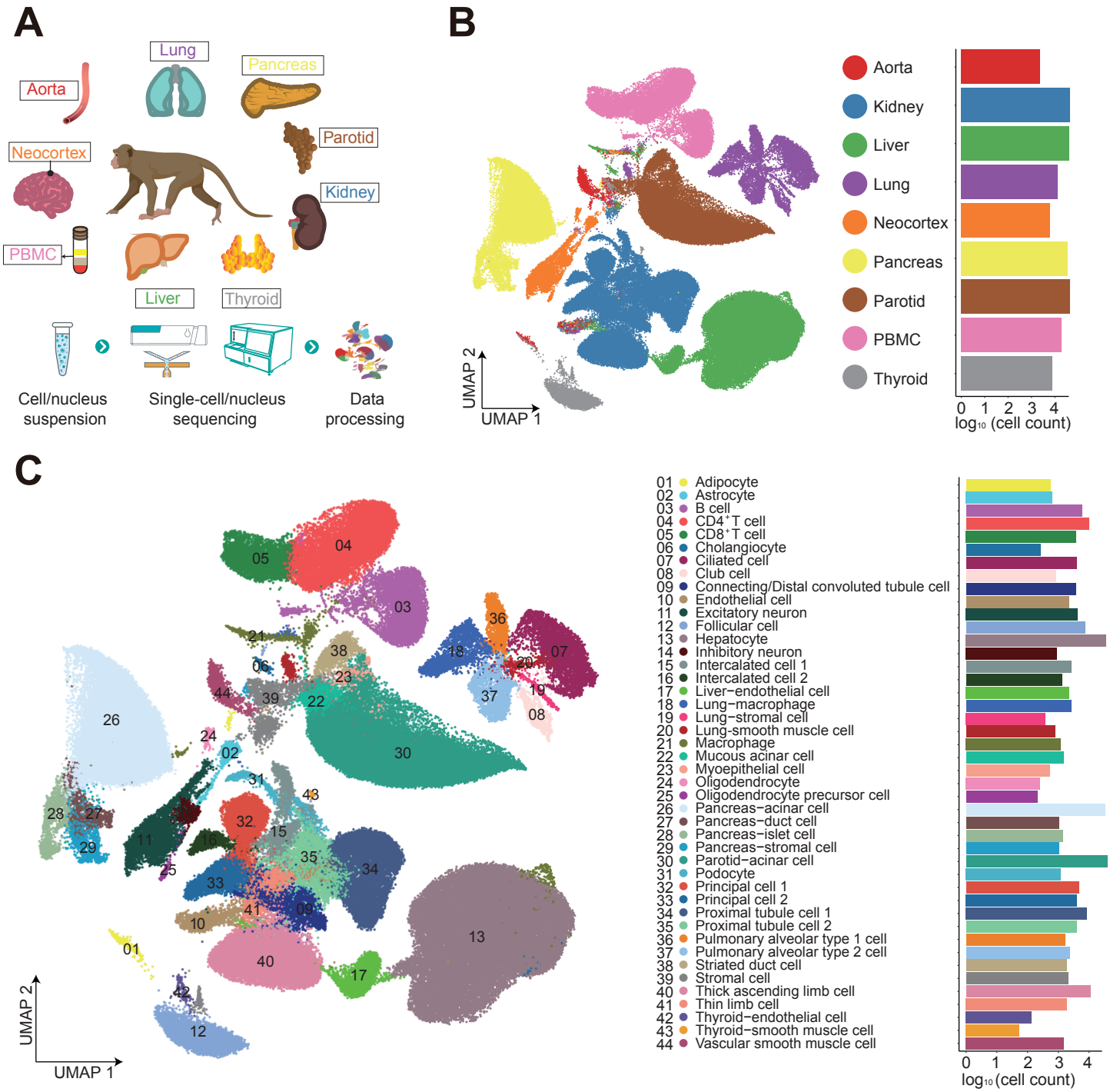


Figure 1

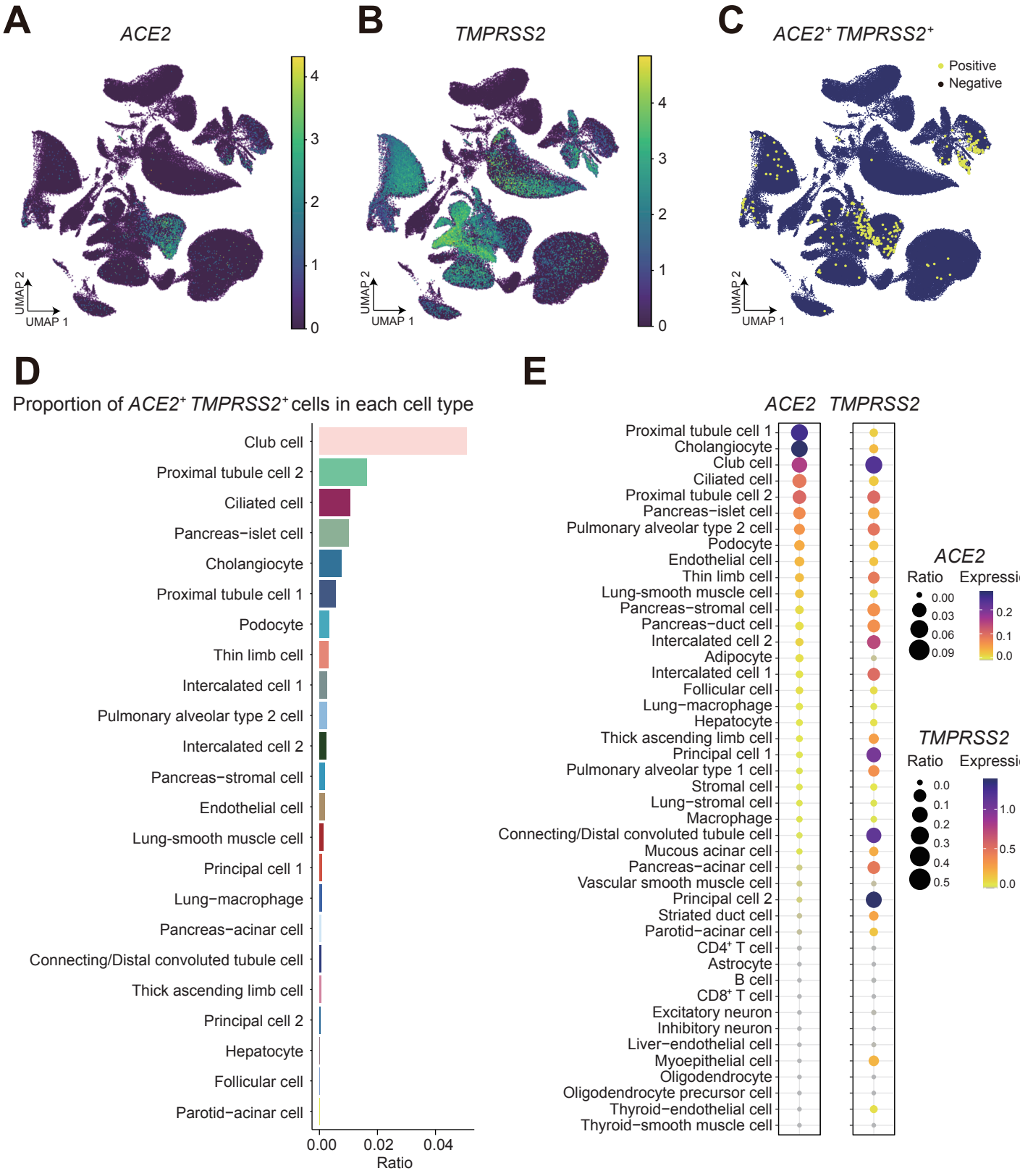


Figure 2

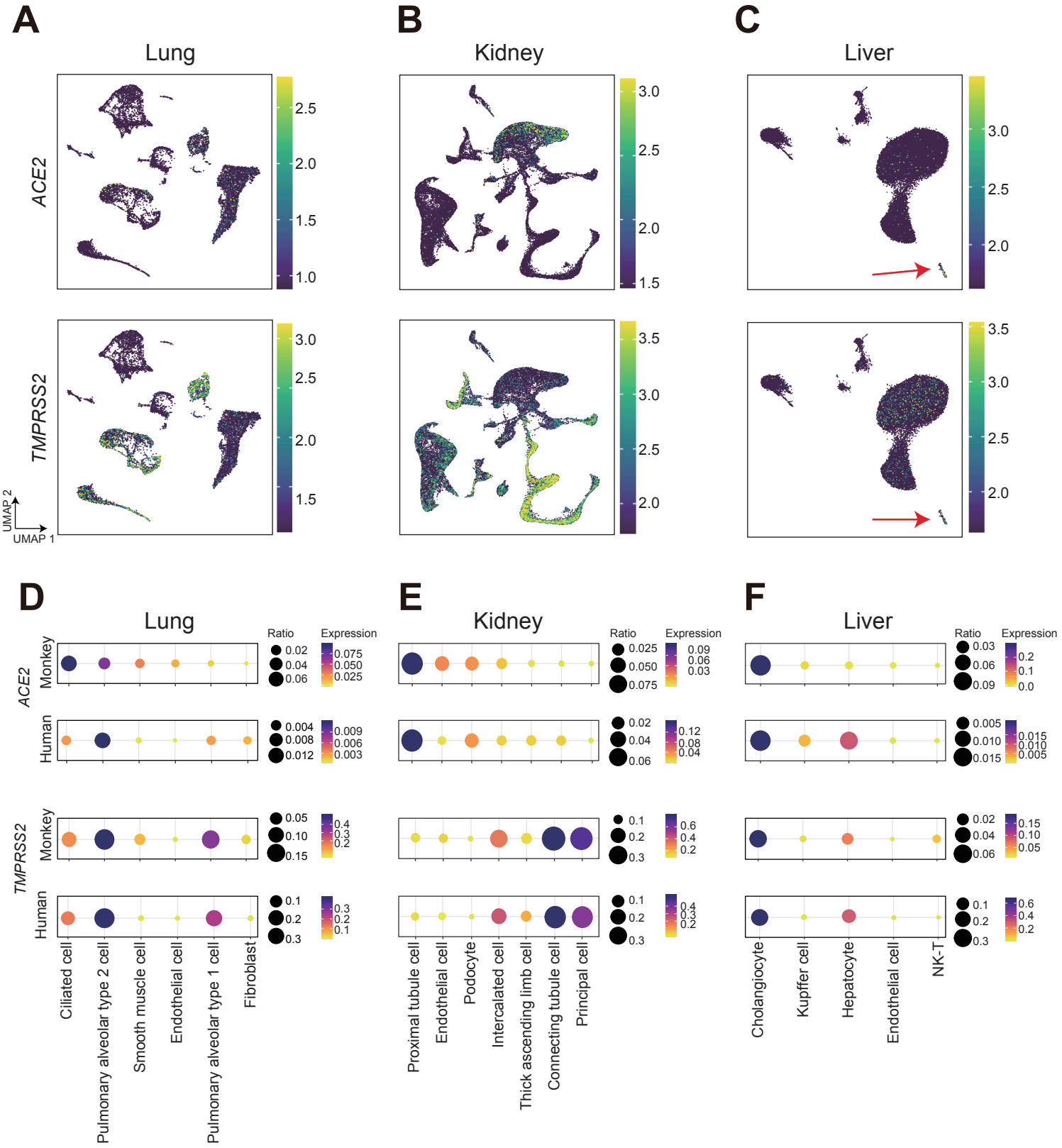


Figure 3

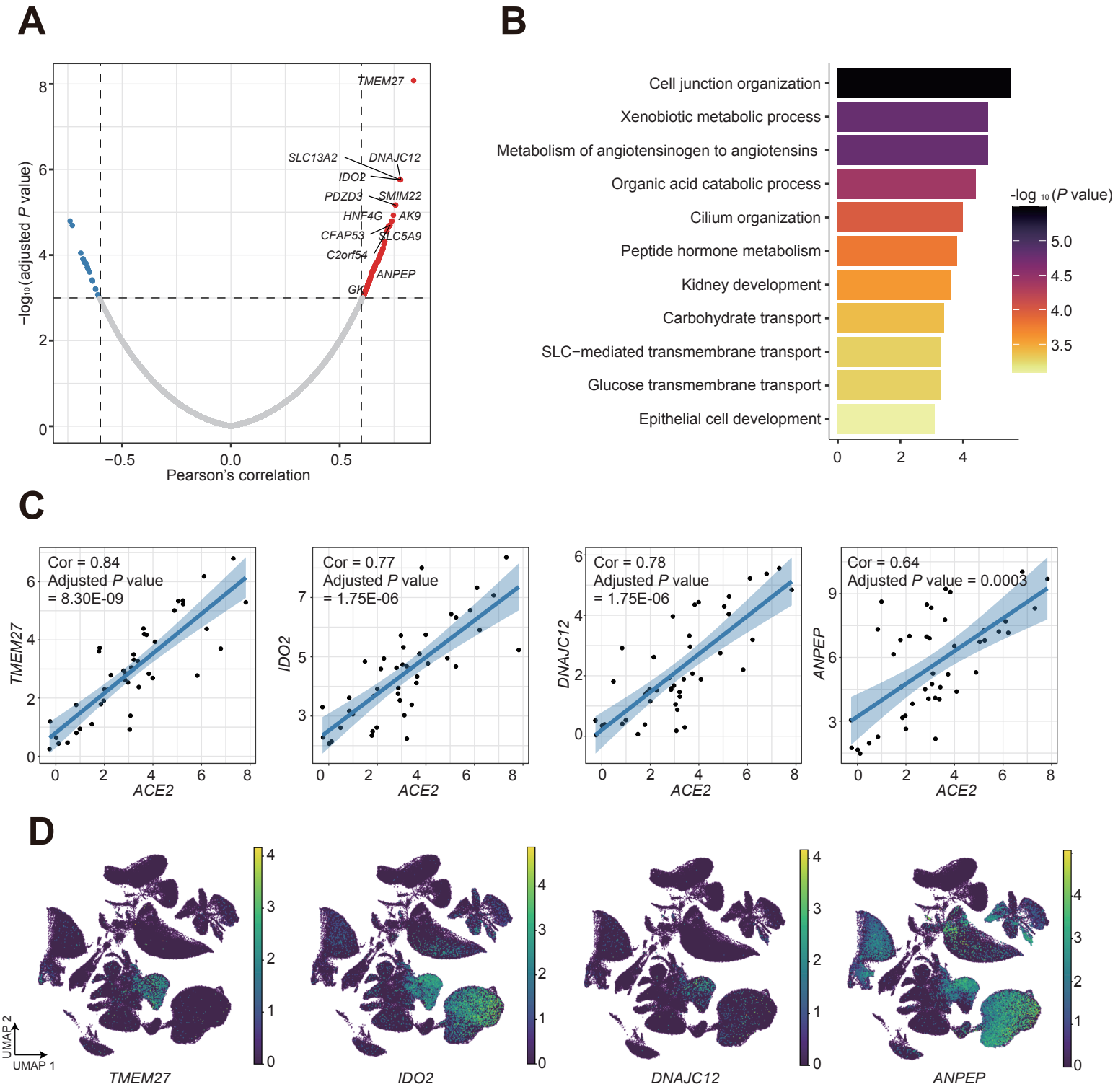
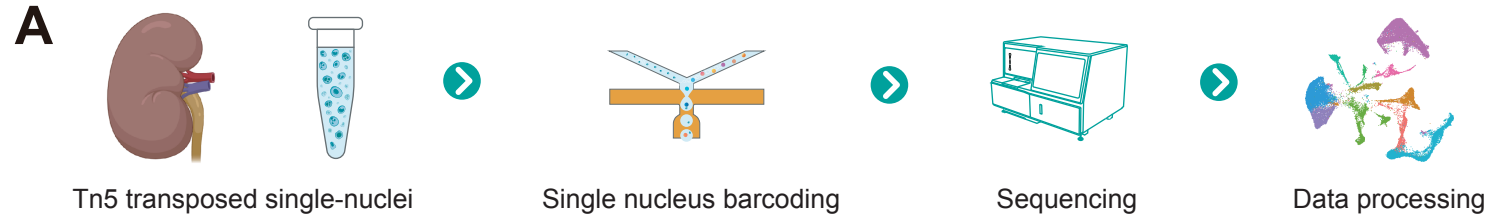
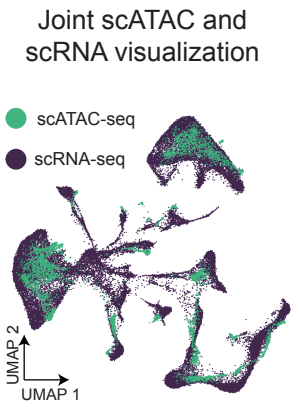
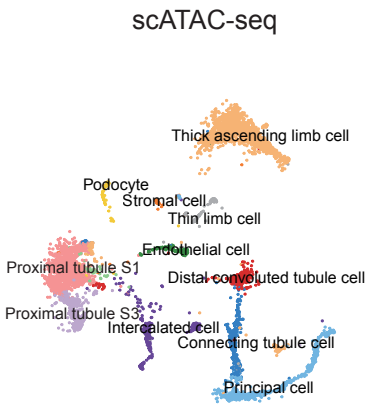
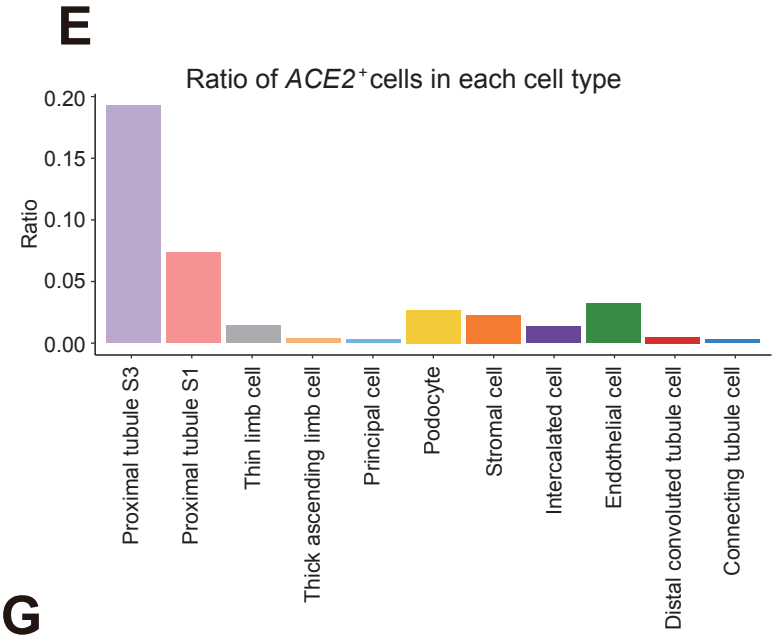


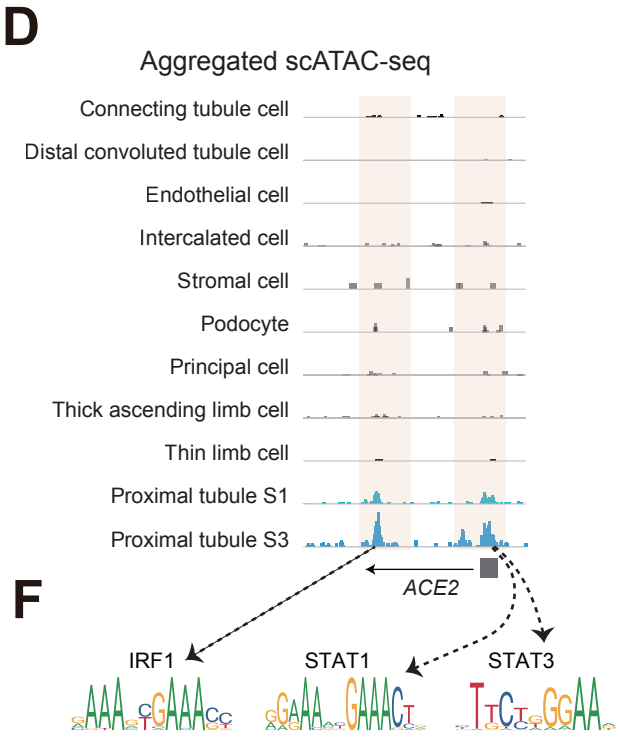
Figure 4











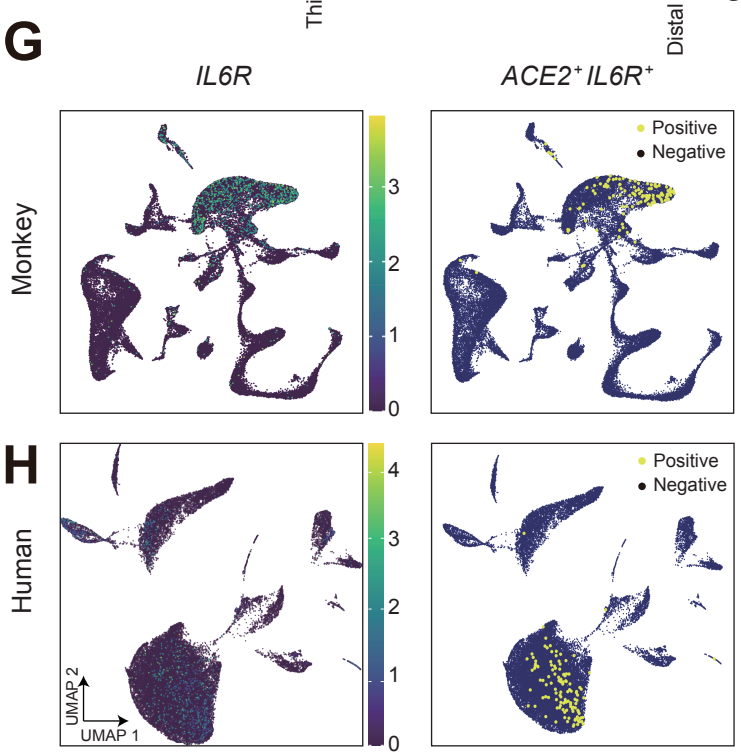


Figure 5

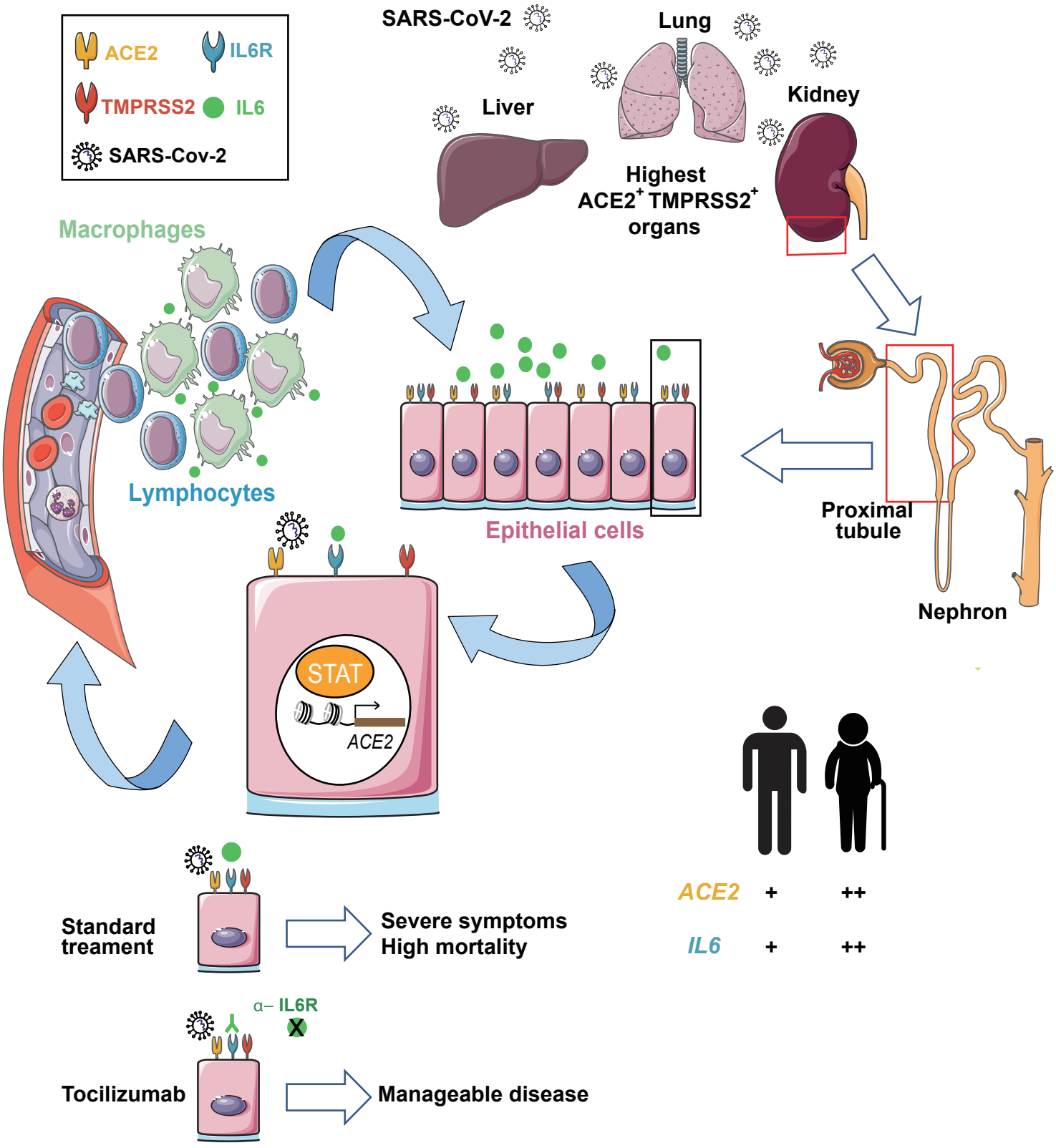
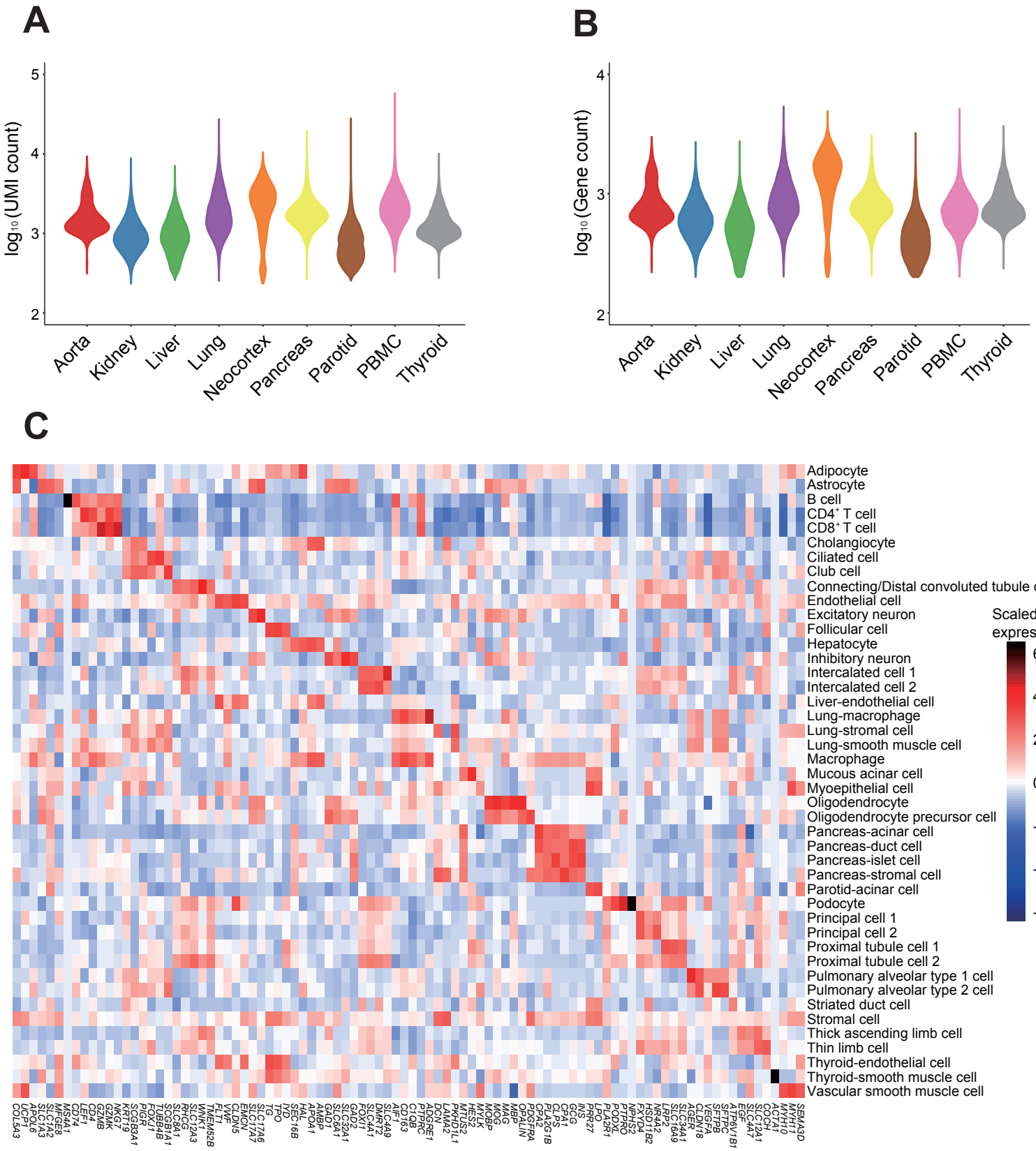
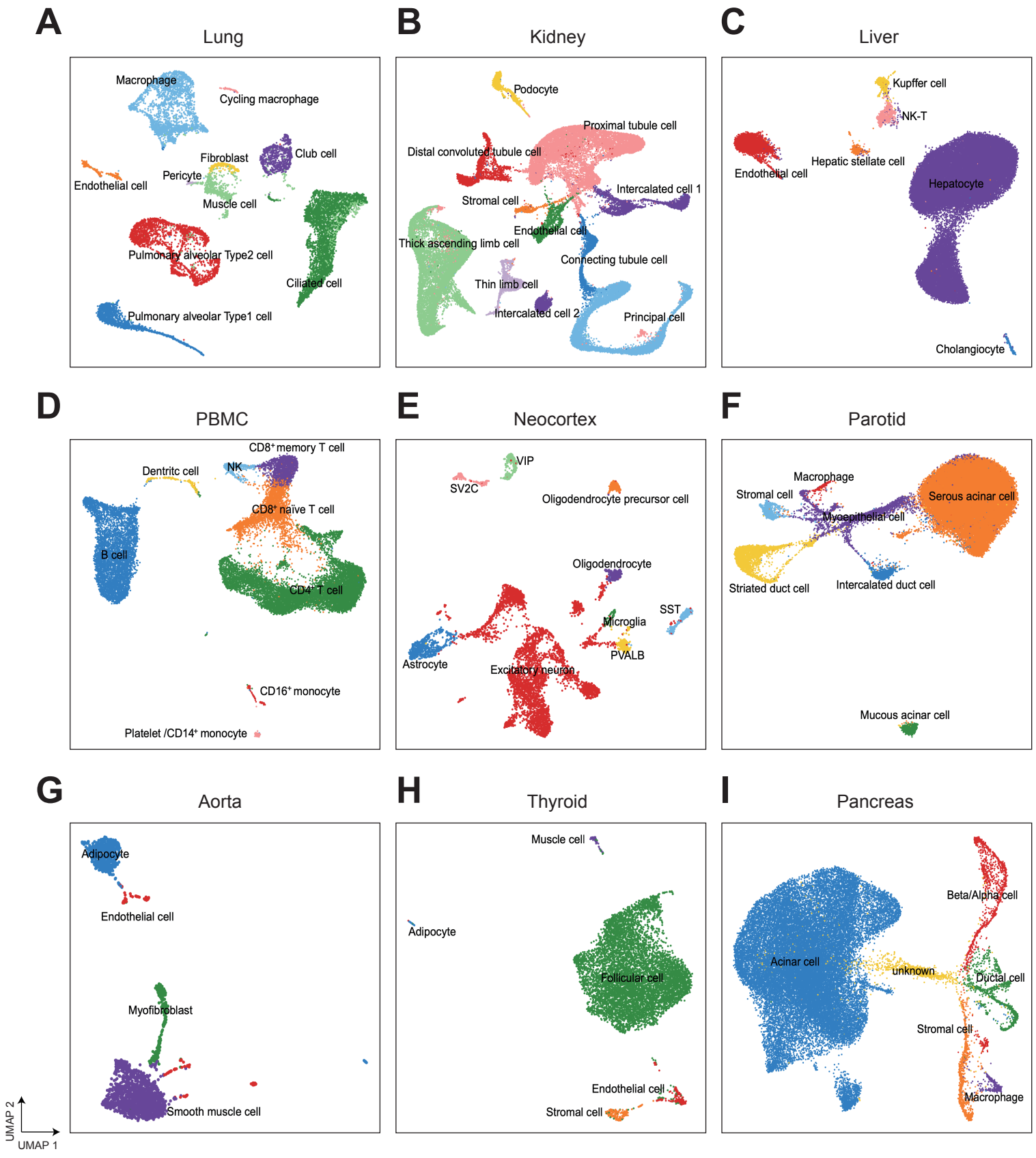


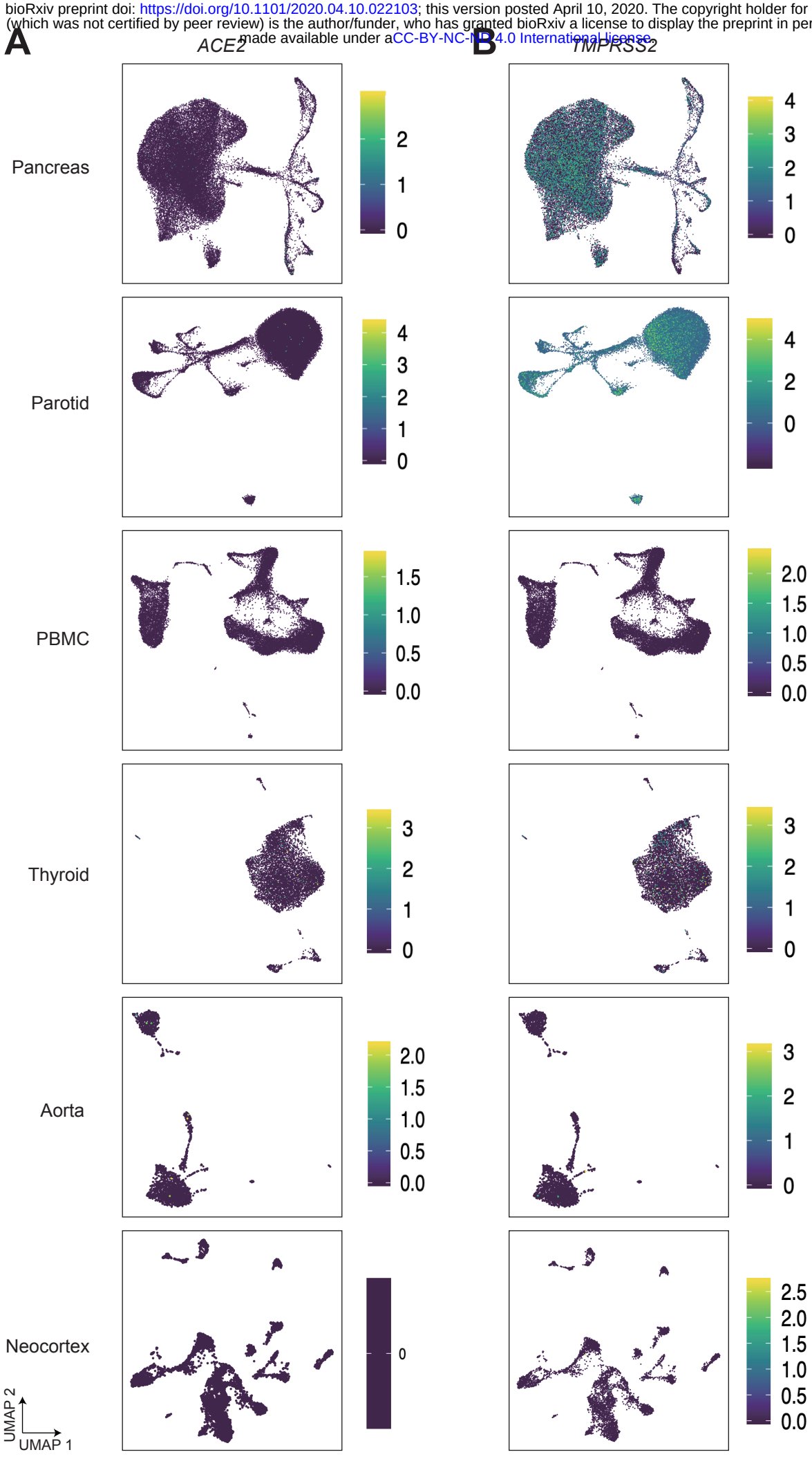
Figure 6



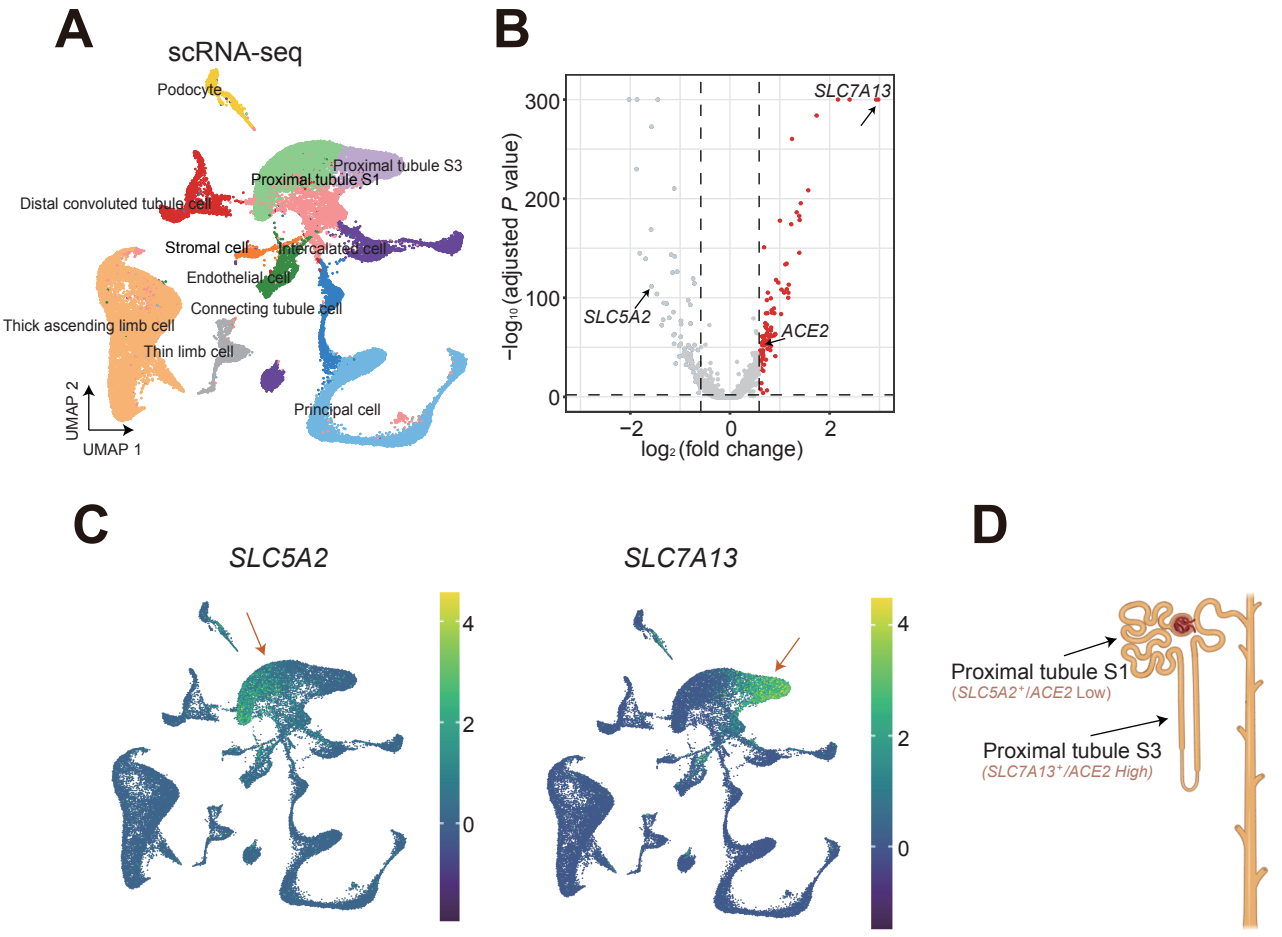
Supplemental Figure 1



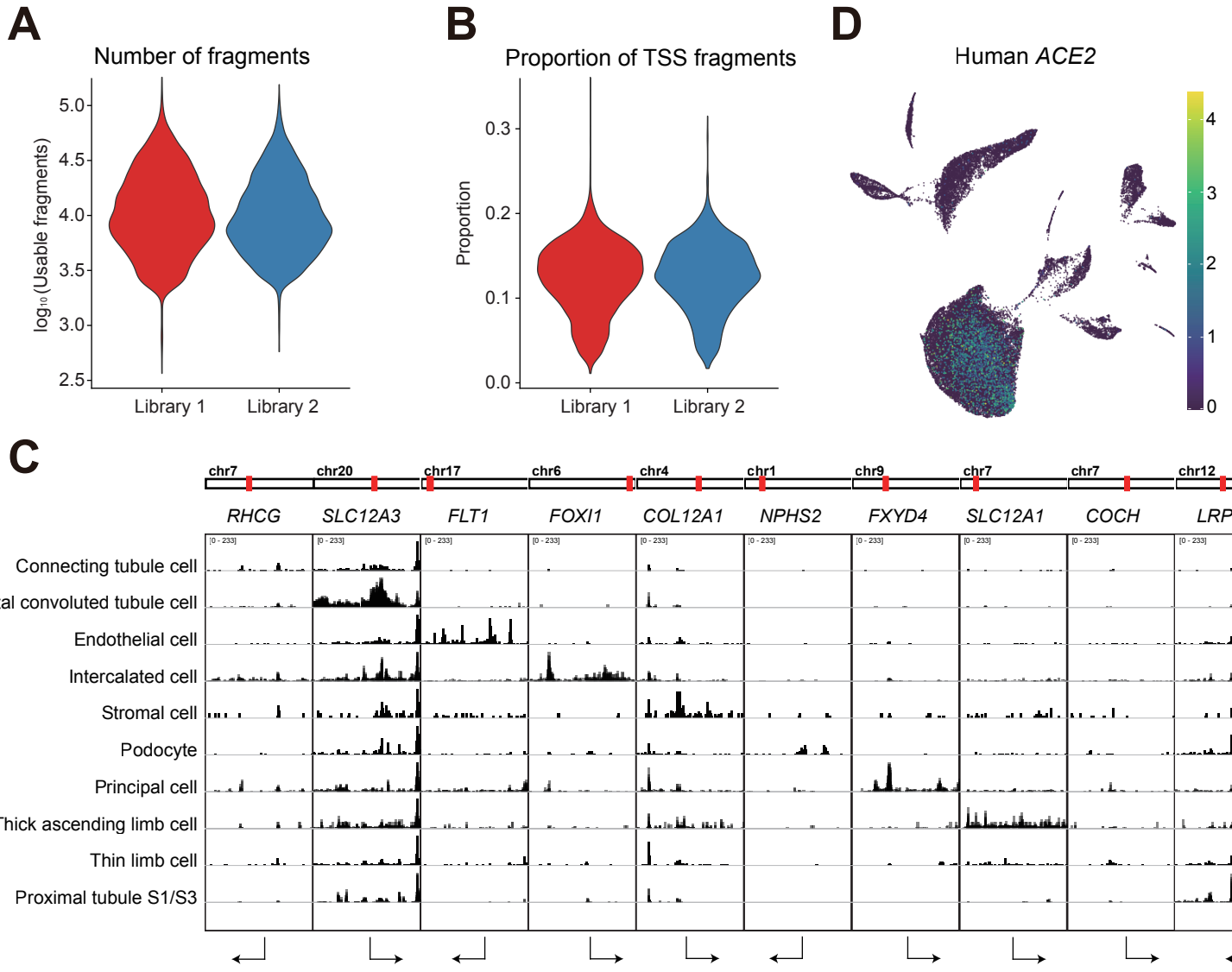
Supplemental Figure 2



Supplemental Figure 3



Supplemental Figure 4



Supplemental Figure 5