Computers in Biology and Medicine 140 (2022) 105092

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Single-cell RNA analysis reveals the potential risk of organ-specific cell types vulnerable to SARS-CoV-2 infections

Zilong Zhang [a,b], Feifei Cui [a,b], Chen Cao [b], Qingsuo Wang [c,*], Quan Zou [a,b,**]

[a] Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, 610054, China
[b] Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou, 324000, China
[c] Beidahuang Industry Group General Hospital, Harbin, 150001, China

ARTICLE INFO

Keywords:
COVID-19
Single-cell RNA sequencing
Bioinformatics
Macrophage
Data mining

ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a global pandemic of coronavirus disease 2019 (COVID-19) since December 2019 that has led to more than 160 million confirmed cases, including 3.3 million deaths. To understand the mechanism by which SARS-CoV-2 invades human cells and reveal organ-specific susceptible cell types for COVID-19, we conducted comprehensive bioinformatic analysis using public single-cell RNA sequencing datasets. Utilizing the expression information of six confirmed COVID-19 receptors (*ACE2, TMPRSS2, NRP1, AXL, FURIN* and *CTSL),* we demonstrated that macrophages are the most likely cells that may be associated with SARS-CoV-2 pathogenesis in lung. Besides the widely reported 'chemokine storm', we identified ribosome related pathways that may also be potential therapeutic target for COVID-19 lung infection patients. Moreover, cell-cell communication analysis and trajectory analysis revealed that M1-like macrophages showed the highest relation to severe COVID-19 patients. And we also demonstrated that up-regulation of chemokine pathways generally lead to severe symptoms, while down-regulation of ribosome and RNA activity related pathways are more likely to be mild. Other organ-specific susceptible cell type analyses could also provide potential targets for COVID-19 therapy. This work can provide clues for understanding the pathogenesis of COVID-19 and contribute to understanding the mechanism by which SARS-CoV-2 invades human cells.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a global pandemic of coronavirus disease 2019 (COVID-19) since December 2019 [1–3]. Similar to the other two members of the β coronavirus genus, i.e., severe acute respiratory syndrome coronavirus (SARS-CoV [4]) and Middle East respiratory syndrome coronavirus (MERS-CoV [5]), SARS-CoV-2 can lead to not only serious respiratory tract diseases but also damage to many other human organs [6–8]. As of May 2021, no effective antiviral for COVID-19 is available, and COVID-19 has caused more than 3.3 million global deaths according to the World Health Organization (WHO). Hence, there is an urgent need to understand the physiological and pathological mechanisms by which SARS-CoV-2 infects humans. Various studies have demonstrated that a glycosylated spike (S) protein of coronavirus plays a key role in the process of cell entry [9,10]. As a cell entry receptor of SARS-CoV [11], angiotensin-converting enzyme 2 (*ACE2*) has also been proven to be an important entry receptor for SARS-CoV-2 [1]. Previous studies have elucidated that the spike protein can bind with the *ACE2* receptor in the receptor binding domain (RBD), and the spike protein is primed by the serine protease *TMPRSS2* [12]. Utilizing the expression value of *ACE2* in single cell data, Dey et al. identified several cell types susceptible to the virus and PPAR signaling pathway as key regulator during infection [13].

However, single-cell RNA sequencing data have revealed that the expression of *ACE2* is very low in most human tissues [14]. This indicates that other receptors may exist for target cell entry due to the rapid and multiorgan infection of COVID-19. Recently, many studies have reported several candidate receptors that may facilitate SARS-CoV-2 entry. Cantuti-Castelvetri et al. reported that neuropilin-1 (*NRP1*) could facilitate SARS-CoV-2 cell entry and may serve as a potential target for drug development [15]. Wu et al. demonstrated that a

* Corresponding author.
** Corresponding author. Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, 610054, China.
E-mail addresses: 7200067@uestc.edu.cn (Q. Wang), zouquan@nclab.net (Q. Zou).

https://doi.org/10.1016/j.compbiomed.2021.105092
Received 19 October 2021; Received in revised form 22 November 2021; Accepted 26 November 2021
Available online 29 November 2021

redundant furin cut site is responsible for the stronger infection of SARS-CoV-2 [16]. Zhao et al. reported that cathepsin L (*CTCL*) plays a crucial role during the process of SARS-CoV-2 infection [17]. Wei et al. found that scavenger receptor B type 1 (*SR-B1*) can act as a host factor and promote SARS-CoV-2 entry [18]. Wang et al. revealed that tyrosine-protein kinase receptor UFO (*AXL*) interacts with SARS-CoV-2 [19].

Although various studies have elucidated the cell entry mechanism of ACE2 and TMPRSS2, little is known about the other newly discovered receptors. Since single-cell RNA sequencing technology has shown great advantages in biological studies [20–31], in this study, we utilized five single-cell RNA sequencing public datasets, each for a tissue vulnerable to COVID-19 infection (specifically, lung, heart, kidney, liver and bladder) [32,33]. Organ-specific susceptible cell types to SARS-CoV-2 were identified, and in lung tissue, we demonstrated that up-regulation of chemokine pathways generally lead to severe symptoms, while down-regulation of ribosome and RNA activity related pathways is more likely to be mild [34]. This work can provide clues for understanding the pathogenesis of COVID-19 and contribute to understanding the mechanism by which SARS-CoV-2 invades human cells [35,36].

2. Material and methods

2.1. Data resources

Public single-cell RNA sequencing datasets were downloaded from the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/). The lung dataset used in this study was obtained as NO. GSE145926. This dataset contained the bronchoalveolar lavage fluid (BALF) of nine COVID-19 patients and three healthy controls. To be consistent with the authors [37], we also obtained other healthy lung scRNA-seq data from NO. GSM3660650 [38]. To explore the expression levels of these candidate receptors, we then utilized scRNA-seq datasets of four other healthy tissues to discover the highly expressed cell types. Specifically, healthy heart samples were acquired from NO. GSE109816 [39]. Five heathy liver samples of scRNA-seq data were acquired from NO. GSE115469 [40]. Three scRNA-seq samples of healthy human kidney were obtained from NO. GSE131685 [41]. Finally, we utilized three healthy bladder scRNA-seq datasets from NO. GSE129845 [42]. The details of all single-cell datasets used in this study was shown in Table 1.

2.2. Dataset processing and analysis

Seurat 3.0 was used to process and analyse the lung dataset [43]. Following the authors' description, we removed low-quality cells by selecting feature numbers between 200 and 6000. Moreover, cells with a mitochondrial percentage >10% and library size <1000 were also identified. The data were normalized by the "NormalizeData" function with the "LogNormalize" method. A total of 2000 highly variable genes were obtained for each sample using the "vst" selection method in the "FindVariableFeatures" function. To correct the batch effect, we utilized the "FindIntegrationAnchors" function and integrated the COVID-19 patients and healthy human samples together for further analysis. Principal component analysis (PCA) was utilized to find the k-nearest neighbours. The "FindClusters" function, which employs the "Louvain" algorithm, showed 31 clusters with a resolution parameter setting of 1.2 [44]. The dimension reduction method UMAP (*Uniform Manifold Approximation and Projection*) was used for data visualization [45]. Based on these clusters, cell type annotation was carried out with canonical marker genes.

Downstream analysis started by finding the top 100 differentially expressed genes (DEGs) of the COVID-19 group and healthy group with the Seurat function "FindMarkers". Differentially expressed genes (DEGs) were filtered by (p value <= 0.05 and q value <= 0.2) for enrichment analysis. The DEGs were further utilized for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment to identify biological pathways. The R package clusterProfiler (version 3.16.1) was used for enrichment analysis [46]. CellChat (version 1.1.0) was utilized for cell-cell communication analysis [47]. Gene set variation analysis (GSVA) analysis was conducted using R package GSVA (version 1.42.0), 50 hallmark gene sets in the Molecular Signatures Database (MSigDB) were utilized for enrichment [48]. We utilized Monocle 2 (version 2.18.0) for trajectory inference and GeneSwitches was used for identifying functional changes pathways [49,50].

The other four scRNA-seq datasets were also analysed similar to the process for the lung dataset. We have uploaded all of the code used in this study at GitHub (https://github.com/ZilongZhang44/COVID-19).

Table 1
Summary of the datasets used in this study.

Dataset GEO number	Organ	DOI	URL
GSE145926 GSM3660650	Lung	10.1038/s41591-020-0901-9	https://www.nature.com/articles/s41591-020-0901-9
GSE109816	Heart	10.1038/s41556-019-0446-7	https://www.nature.com/articles/s41556-019-0446-7
GSE115469	Liver	10.1038/s41467-018-06318-7	https://www.nature.com/articles/s41467-018-06318-7
GSE131685	Kidney	10.1038/s41597-019-0351-8	https://www.nature.com/articles/s41597-019-0351-8
GSE129845	Bladder	10.1681/ASN.2019040335	https://pubmed.ncbi.nlm.nih.gov/31462402/

3. Results

3.1. Lung dataset

Unsupervised clustering of the single-cell RNA sequencing lung dataset showed 31 clusters (Fig. 1A). Canonical marker genes provided by the authors were utilized for cell type annotation; subsequently, twelve cell types were identified, as shown in Fig. 1B. Fig. 1C shows the gene expression of the most important canonical marker genes, specifically *TPPP3, KRT18, CD68, FCGR3B, CD1C, CLEC9A, LILRA4, TPSB2, CD3D, KLRD1, MS4A1* and *IGHG4*. Clusters 0, 1, 2, 3, 4, 5, 7, 8, 10, 11, 12, and 16 were annotated as macrophages with the canonical marker gene *CD68*. With *CD3D*, T cells were identified corresponding to clusters 6, 9, 14 and 29. With *TPP3* and *KRT18*, ciliated cells and secretory cells were annotated corresponding to cluster 13 and cluster 15, respectively. Cluster 18 was identified as NK cells with *KLRD1*, while cluster 19 was identified as neutrophil cells with *FCGR3B*. Cluster 21 was annotated to myeloid dendritic cells with marker genes *CD1C* and *CLEC9A*. Subsequently, B cells and mast cells were annotated with *MS4A1* and *TPSB2*, which corresponded to cluster 25 and cluster 30, respectively. With *IFHG4,* clusters 23 and 26 were identified as plasma cells. Cluster 24 was annotated as epithelial cells with marker genes *TPPP3* and *KRT18*. Finally, plasmacytoid dendritic cells were identified as corresponding to cluster 28 with the marker gene *LILRA4*.

Moreover, we divided the lung dataset into a COVID-19 group and a healthy group and compared the expression differences of all potential receptors mentioned above: *ACE2, TMPRSS2, NRP1, AXL, FURIN* and *CTSL* (*SR-B1* gene was not detected). Fig. 1D illustrates significant expression differences in all 6 reported receptors, which is consistent with the original paper. As *ACE2* and *TMPRSS2* have already been investigated in many studies, we mainly focused on the expression of the most recently reported genes *NRP1, AXL, FURIN* and *CTSL*. As shown in Fig. 1D, in the COVID-19 group, *NRP1* and *AXL* were downregulated, while *FURIN* and *CTSL* were upregulated. To explore specific cell types

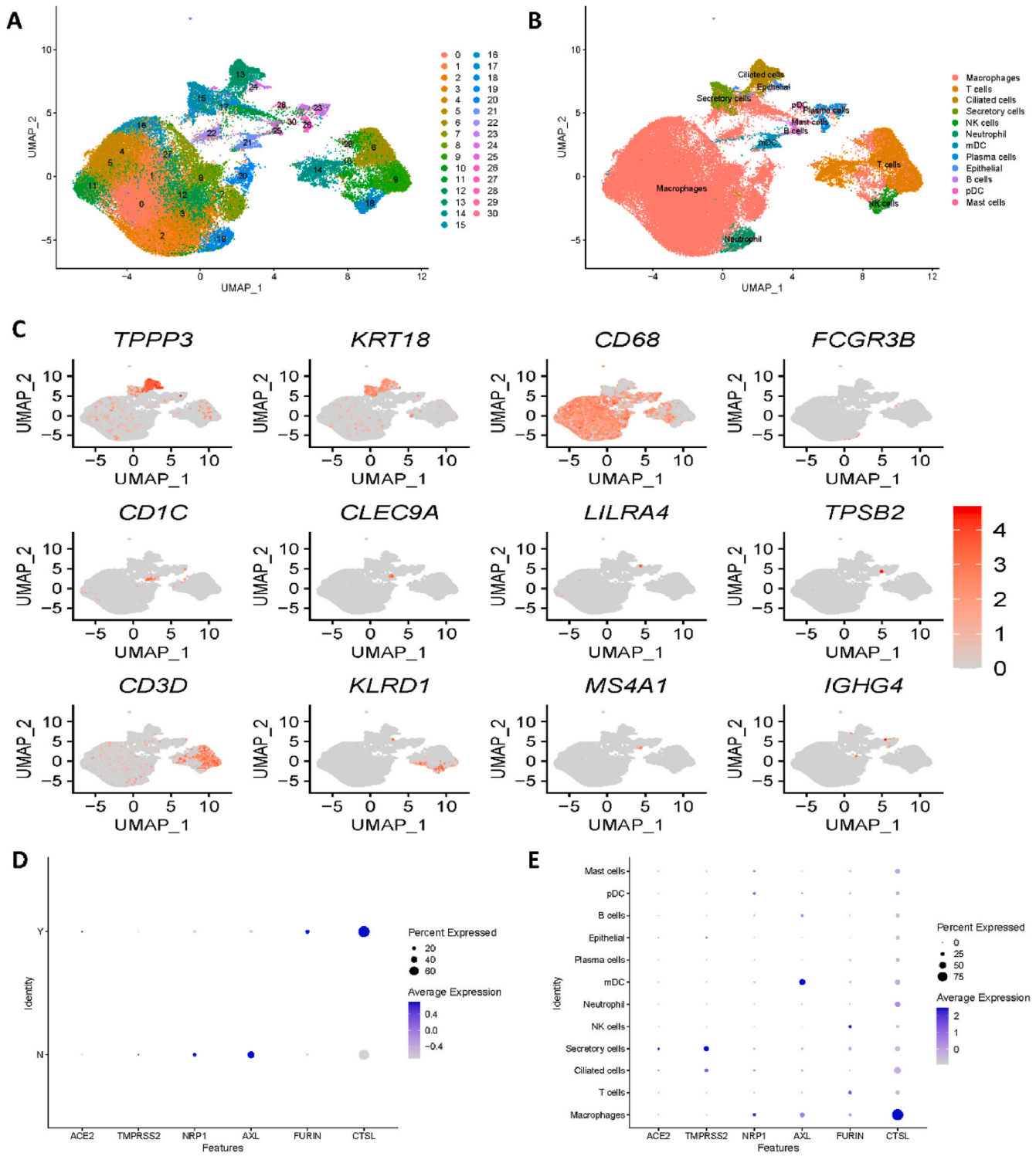

Fig. 1. Single-cell RNA sequencing analysis of the lung dataset. (A) Unsupervised clustering revealed 31 clusters in integrated lung dataset. **(B)** Cell type annotation was conducted and 12 cell types were identified. **(C)** Expression of canonical marker genes used for cell type annotation. **(D)** Expression differences in all 6 reported receptors between COVID-19 patients and healthy controls. Y stands for COVID-19 patients, N stands for healthy controls. **(E)** Dot plot shows the expression of six receptors in all identified cell types.

vulnerable to SARS-CoV-2 infections in the lung, we then revealed the expression of receptors by cell type. Fig. 1E illustrates that all four of the latest reported receptors showed high expression in macrophages. Hence, we concentrated on macrophage cell types for further analysis.

To reveal the heterogeneity among macrophages, the macrophages cells were further clustered into M1-like macrophages and M2-like macrophages based on the canonical marker genes *FCN1, CCL2, CCL3* and *TREM2* (Fig. 2 A, C). Interestingly, we noticed that percentage of M1-like macrophages was highly correlated with COVID-19 progression. Specifically, M1-like macrophages only accounted for 0.2% of macrophages in healthy control group, while increasing to 19.3% and 97.3% in medium COVID-19 cases and severe COVID-19 cases, respectively

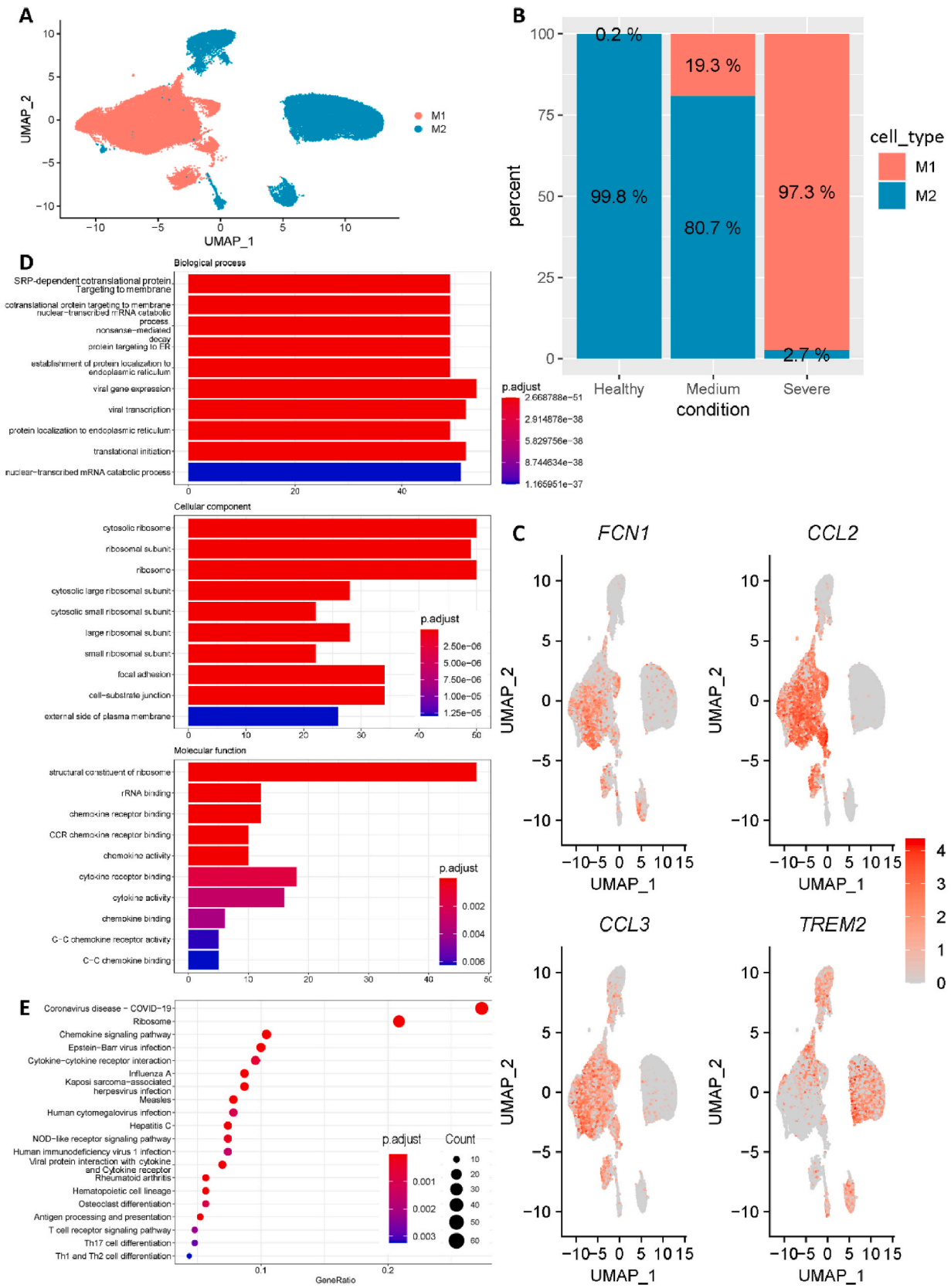

Fig. 2. Macrophages subgroup classification. **(A)** All macrophages were further clustered into M1-like macrophages and M2-like macrophages. **(B)** The fraction of cells originated from different groups (Healthy controls, Medium patients and Severe patients). **(C)** Expression of canonical marker genes used for cell subtype annotation. **(D)** Enriched terms for each of three ontologies revealed by Gene Ontology (GO) enrichment results of M1-like macrophages. **(E)** Most enriched KEGG pathways revealed by differentially expressed genes of M1-like macrophages.

(Fig. 2 B). In addition, all M1-like macrophages were divided into a COVID-19 group and a healthy group, and differential gene expression analysis was applied between these two groups. Differentially expressed genes (DEGs) were filtered by (p value <= 0.05 and q value <= 0.2) for enrichment analysis. Fig. 2D showed the top 10 enriched Gene Ontology (GO) terms for each of three ontologies. Interestingly, besides the widely reported 'chemokine storm', ribosome related terms were strongly enriched. Strikingly, as shown in Fig. 2E, ribosome pathway was the greatest enrichment besides COVID-19 pathway. Collectively, ribosome pathway may be possible therapeutic target for COVID-19 lung infection patients.

We then utilized CellChat to investigate cell-cell communication in COVID-19 patients [47]. Notably, M1-like macrophages showed significantly interactions with T cells, Neutrophil and NK cells (Fig. 3A), and Fig. 3B revealed the dominant role of M1-like macrophages in COVID-19 patients. These findings have also been verified by Lv et al. that M1-like macrophages could facilitate SARS-CoV-2 infection of the lungs by allowing the entry of SARS-CoV-2 RNA from the endosomes into the cytoplasm [51]. In addition, as shown in Fig. 3C, CellChat revealed three communication outgoing patterns. M1-like macrophages were grouped into pattern 1 and *SPP1* (secreted phosphoprotein 1), *NPR2* (Natriuretic Peptide Receptor 2) and *CCL* (Chemokine ligands) may be possible significant pathways. Moreover, GSVA software was utilized to identify differences in hallmark gene sets of Molecular Signatures Database (MSigDB) across different cell types [48,52–55]. As shown in Fig. 3D, complement pathway, KRAS_singaling pathway and Inflammatory response pathway are significantly enriched in M1-like macrophages, which may be considered as possible therapeutic targets for further study.

Finally, we used Monocle 2 to investigate cell trajectory of macrophages in COVID-19 patients [49]. Fig. 4A revealed two major cell trajectories (i.e., path1: from state 4 to state 3, and path2: from state 4 through state 2 to state 1). Notably, we noticed that cell fates belong to path1 (states 4, 3) had a much higher severe rate than path2 (states 4, 2, 1). Subsequently, we utilized GeneSwitches to discover functional events during cell state transitions [50]. Moreover, as shown in Fig. 4B, the top ten significantly changed pathways for path1 are strongly related to previous studies about COVID-19 severe patients, which included interferon alpha, cytokine and immune system. Strikingly, Fig. 4C revealed that the highest functional change pathways in path2 (states 4, 2, 1) are related to ribonucleoprotein and RNA activity, which is consistent with enrichment analysis. Collectively, we demonstrated that up-regulation of Chemokine pathways generally lead to severe symptoms, while down-regulation of ribosome and RNA activity related pathways is more likely to be mild.

3.2. Heart dataset

Cells in the heart dataset were grouped into 15 clusters, as shown in Fig. 5A. We then chose canonical marker genes (*TECRL, VWF, LUM, LAPTM5* and *RGS5*) for biological annotation. The annotation processes and results are shown in Fig. 5B–C. Clusters 0, 5, 6 and 14 were identified as endothelial cells by the marker gene *VWF*. With *TECRL*, clusters 1, 2, 3, and 9 were identified as cardiomyocytes. Cluster 7 was annotated as fibroblasts by the marker gene *LUM*. With *LAPTM5,* clusters 10, 11 and 13 were annotated as macrophages. Finally, clusters 4 and 8 were identified as smooth muscle cells using *RGS5*. Fig. 5D depicts possible heart-specific cell types vulnerable to COVID-19 by showing the gene expression conditions of potential receptors. We found that fibroblasts and smooth muscle cells had higher expression of almost all receptors, which indicates a high risk for COVID-19 infection.

3.3. Kidney dataset

We then reanalysed the expression profiles of COVID-19 receptors in a kidney dataset. Unsupervised clustering by Seurat illustrated 11 clusters, as shown in Fig. 6A. The annotation step for the kidney dataset is shown in Fig. 6B–C. With the canonical marker genes *LYZ* and *CD14*, cluster 6 was annotated as monocytes. NK cells and T cells were identified with *GNLY*, *NKG7* and *CD3D*, which corresponded to cluster 5. Cluster 10 was identified as B cells by the *CD79A* and *CD79B genes,* while cluster 8 was annotated as distal tubule cells by the *UMOD* and *DEFB1* marker genes. With *AQP2* and *CLDN8*, cluster 9 was identified as collecting duct cells. Cluster 11 was identified as collecting duct intercalated cells by the *ATP6V1G3* and *ATP6V0D2* genes. The expression profiles of *KRT8* and *KRT18* indicated that cluster 7 corresponded to glomerular parietal epithelial cells. *SLC22A8* and *SLC22A7* were utilized to annotate cluster 4 and cluster 1 as proximal convoluted tubule cells and proximal straight tubule cells, respectively. Finally, clusters 2 and 3 were identified as proximal tubule cells using marker genes *DCXR* and *GPX3*. Fig. 6D demonstrates that collecting duct cells and proximal tubule cells may be most vulnerable to COVID-19 infection.

3.4. Liver dataset

Similarly, we reanalysed liver scRNA-seq data. As shown in Fig. 7A, all cells were grouped into 22 clusters. With the canonical marker genes *BCHE* and *G6PC*, clusters 0, 1, 2, 17, 18 and 20 were identified as hepatocytes. Clusters 11, 12, and 13 were annotated as limbal epithelial stem cells (LESCs) by *TM4SF1* and *CCL14*. With the marker genes *KRT7* and *KRT19,* cluster 15 was identified as cholangiocytes. Cluster 21 was identified as stellate cells by *ACTA2* and *RBP1*, while cluster 7 was identified as inflammatory macs based on the expression profile of *LYZ* and *HLA-DPB1*. With *CD5L, MARCO* and *VSIG4*, noninflammatory macs were annotated corresponding to cluster 9. Clusters 6, 8 and 19 were all annotated as T cells; specifically, clusters 6 and 19 were identified as γδ T-cells by the marker genes *GNLY* and *STMN1*, while cluster 8 was identified as $CD3^{+}$ αβ T-cells by *GZMK*. B cells were identified corresponding to cluster 14 by *MS4A1* and *CD37*. Finally, plasma cells and erythroid cells were annotated corresponding to clusters 4 and 16 by *IGHG1* and *CA1,* respectively (Fig. 7B–C). Fig. 7D demonstrates that overall, the most specific cell type that is vulnerable to COVID-19 is cholangiocyte, while stellate cells and noninflammatory macs cells may also be potential virus targets based on these four newly reported receptors.

3.5. Bladder dataset

For the bladder dataset, all cells were grouped into 19 clusters (Fig. 8A). The annotation processes are shown in Fig. 8B and C. Clusters 1, 2, 3, and 5 were identified as base cells with genes *KRT5* and *KRT17*. Intermediate cells were annotated with *KRT13* corresponding to cluster 4. With the canonical marker genes UPK1A and *UPK1B*, cluster 15 was annotated as umbrella cells. Interstitial cells were identified corresponding to cluster 17 by *VIM*. With *ACTA2*, cluster 15 was annotated as myofibroblasts. Cluster 8 was annotated as smooth muscle cells with the DES gene. With *SELE* and *PECAM1*, cluster 12 was identified as endothelial cells. Subsequently, T cells and B cells were identified corresponding to cluster 14 and cluster 18, respectively, by marker genes *CD3D* and *MZB1*. With the marker genes *LYZ* and *MS4A7*, cluster 13 was annotated as monocytes. Last, cluster 11 was annotated as fibroblasts with *COL1A2*. However, in Fig. 8D, it was difficult to determine the most vulnerable cell types to COVID-19, and further investigation is needed to obtain a convincing conclusion.

4. Discussion

As of May 2021, the COVID-19 global pandemic has caused more than 160 million confirmed cases, including 3.3 million deaths. Hence, there is an urgent need to understand the physiological and pathological mechanisms by which SARS-CoV-2 infects humans. In this study, we focused on the expression profiles of several reported receptors of SARS-

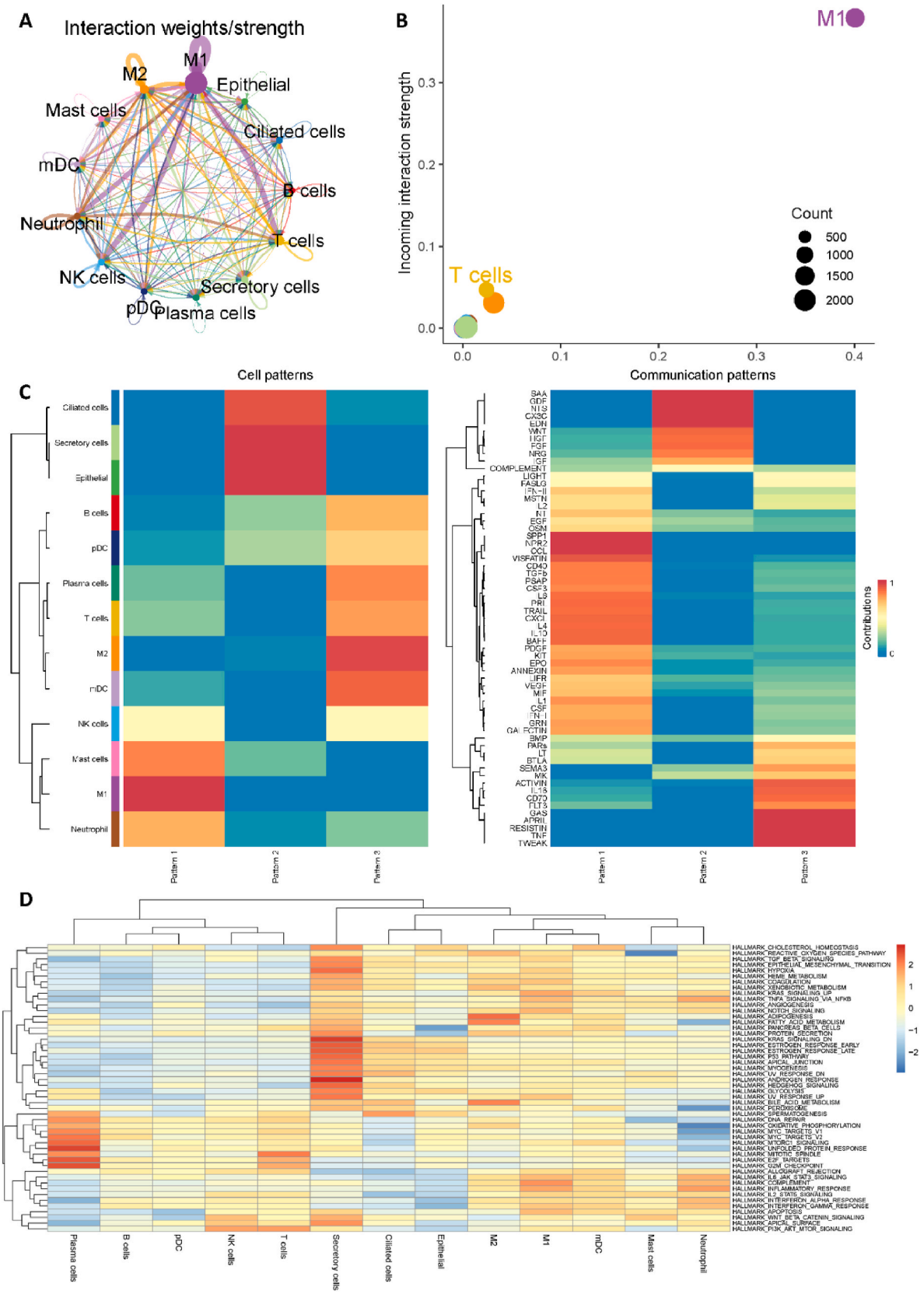

Fig. 3. Cell-cell communication analysis of macrophages. (A) Aggregated cell-cell communication network of total interaction strength (weights) between any two cell groups. **(B)** Signaling role analysis on the aggregated cell-cell communication network from all signaling pathways. **(C)** Global communication pattern of multiple cell types. **(D)** Gene set variation analysis (GSVA) for 50 hallmark gene sets in the Molecular Signatures Database (MSigDB).

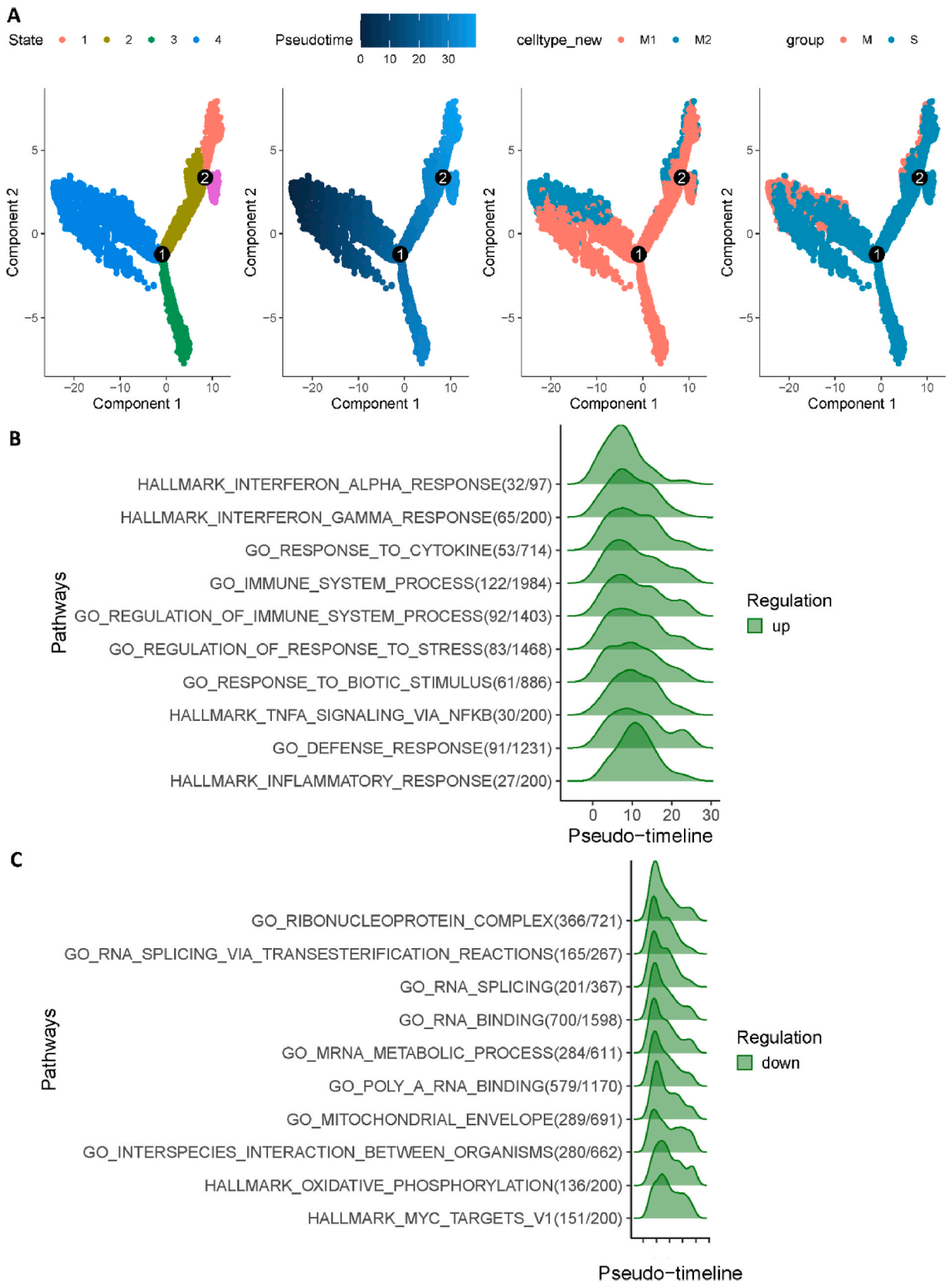

Fig. 4. Trajectory inference of macrophages in COVID-19 patients. (A) Trajectory analysis results of macrophages in COVID-19 patients. (From left to right, cells were coloured by states, cell type, pseudotime and patient group, respectively.) **(B)** Top 10 significantly changed pathways ordered by the switching time of path1 (states 4, 3). **(C)** Top 10 significantly changed pathways ordered by the switching time of path2 (states 4, 2, 1).

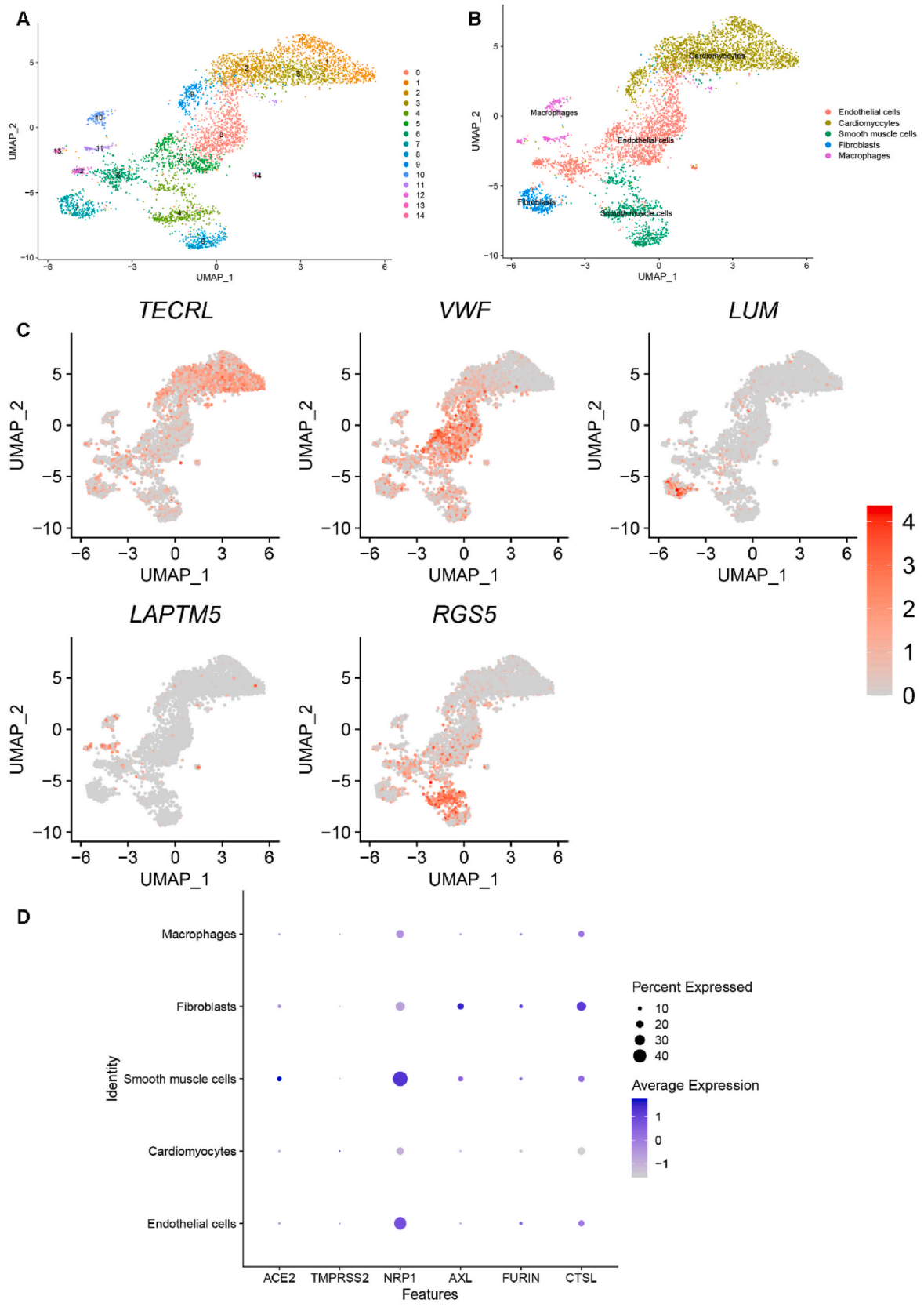

Fig. 5. Single-cell RNA sequencing analysis of the heart dataset. (A) Unsupervised clustering revealed 15 clusters in heart dataset. **(B)** Cell type annotation was conducted and 5 cell types were identified. **(C)** Expression of canonical marker genes used for cell type annotation. **(D)** Dot plot shows the expression of six receptors in all identified cell types.

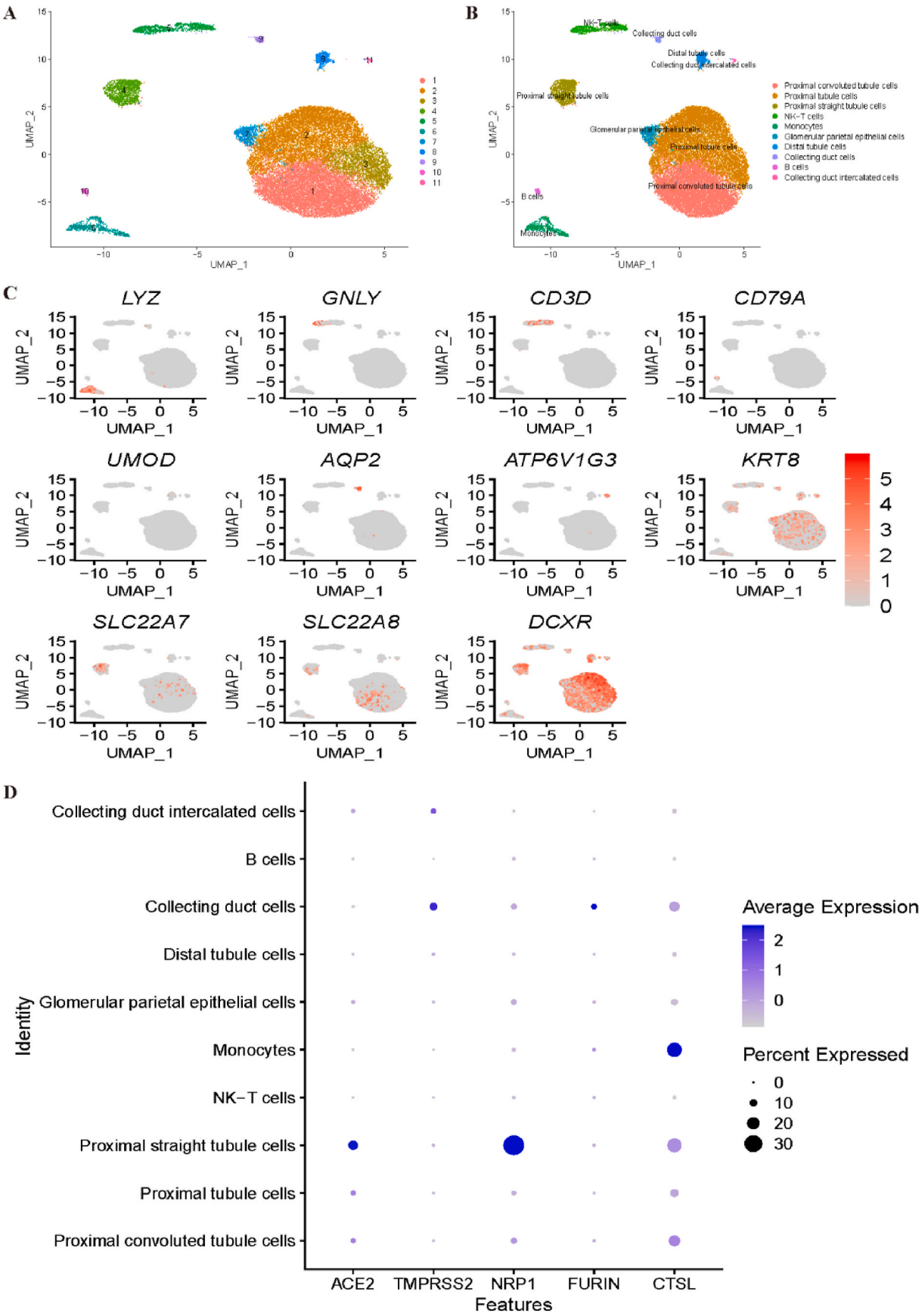

Fig. 6. Single-cell RNA sequencing analysis of the kidney dataset. (A) Unsupervised clustering revealed 11 clusters in kidney dataset. **(B)** Cell type annotation was conducted and 10 cell types were identified. **(C)** Expression of canonical marker genes used for cell type annotation. **(D)** Dot plot shows the expression of six receptors in all identified cell types.

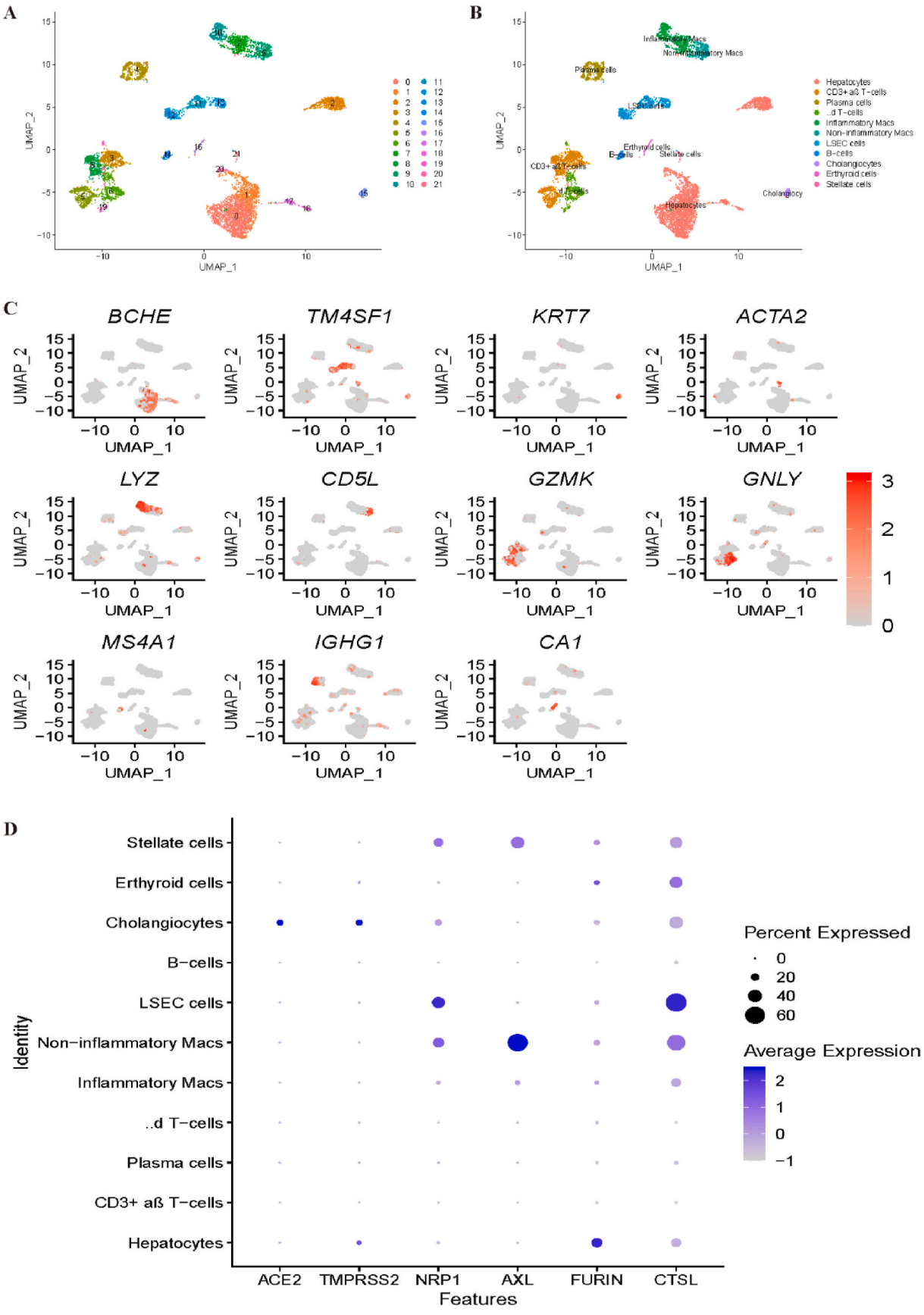

Fig. 7. Single-cell RNA sequencing analysis of the liver dataset. (A) Unsupervised clustering revealed 22 clusters in liver dataset. **(B)** Cell type annotation was conducted and 11 cell types were identified. **(C)** Expression of canonical marker genes used for cell type annotation. **(D)** Dot plot shows the expression of six receptors in all identified cell types.

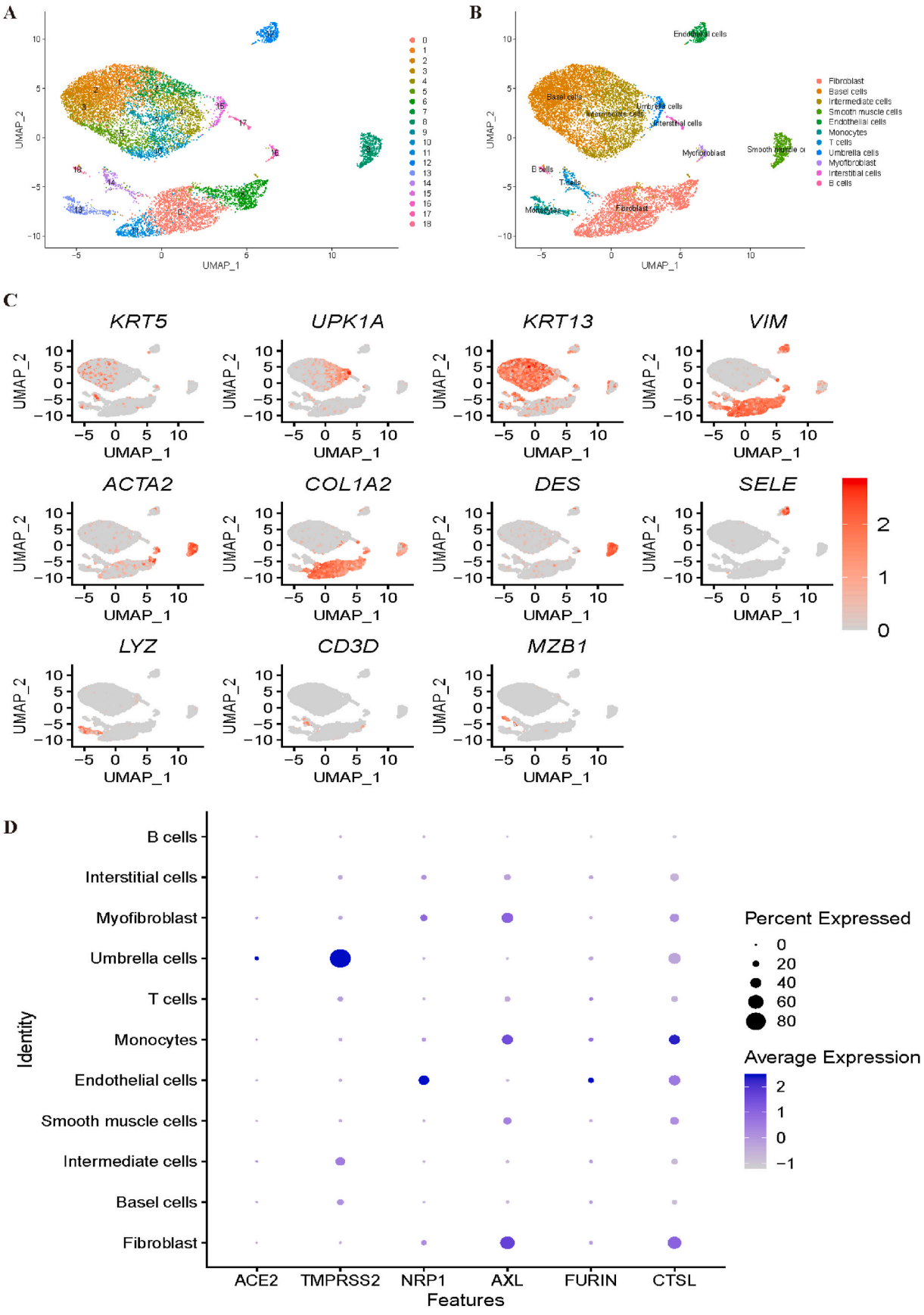

Fig. 8. Single-cell RNA sequencing analysis of the bladder dataset. (A) Unsupervised clustering revealed 19 clusters in bladder dataset. **(B)** Cell type annotation was conducted and 11 cell types were identified. **(C)** Expression of canonical marker genes used for cell type annotation. **(D)** Dot plot shows the expression of six receptors in all identified cell types.

CoV-2 (specifically, *ACE2, TMPRSS2, NRP1, AXL, FURIN* and *CTSL*) in different organs.

Since lung abnormalities are one of the most common symptoms for COVID-19 patients, we first explored the different expression profiles of potential receptors between COVID-19 patients and healthy donors [56]. As expected, all six receptors showed significant transcription differences upon infection with SARS-CoV-2. The mRNA expression levels of ACE2 and TMPRSS2 are highest in secretory cells and ciliated cells, which is consistent with other previous studies [14,57,58]. Interestingly, after focusing on observing the expression profiles of the other four newly confirmed receptors, we found that macrophages showed high expression levels of all of these receptors. Cell-cell communication analysis and trajectory inference were conducted to COVID-19 patients scRNA-seq data, which showed that M1-like macrophages were significantly associated with SARS-CoV-2 pathogenesis in lung. These results were consistent with a recent study which revealed M1 alveolar macrophages facilitate SARS-CoV-2 infection of the lungs [51]. GO and KEGG enrichment were conducted to M1-like macrophages, as expected, the GO enrichment results showed many virus-related pathways. Moreover, various GO terms related to cytokines, such as "cytokine activity" and "chemokine activity", were also significantly enriched, which indicates the important role of the presence of so-called 'cytokine storm' [28,34,59]. Recent study by Jaggi et al. has also proven that M1 macrophages played an essential role in blocking cytokine storm [60]. As we know, the "classically" activated M1-like macrophages are generally pro-inflammatory. Large proportion of M1-like macrophages in the lung could stimulate the SARS-CoV-2 associated cytokine storm, which may be the reason for the serious condition of severity patients. Excitingly, we also noticed that ribosome related pathways were also strongly enriched in both GO and KEGG enrichment, which may be potential therapeutic target for COVID-19 lung infection patients.

Notably, we noticed that up-regulation of chemokine pathways generally lead to severe symptoms, while down-regulation of ribosome and RNA activity related pathways is more likely to be mild. Various studies have already revealed that due to the up-regulation of chemokine, which caused over-production of soluble markers of inflammation [28]. However, down-regulated ribosome pathways have been rarely discussed. As we know, ribosomes are complex molecular inside the living cells for producing proteins. Down-regulated of ribosomes may hinder the SARS-CoV-2 replication and inhibit the progression of the disease. Hence, ribosomes activity may be used as an indicator for COVID-19 disease assessment.

As COVID-19 has been proven to cause many symptoms in tissues other than the lung, we then collected scRNA-seq data from the heart, kidney, liver and bladder in healthy donors [61]. We mainly focused on revealing the organ-specific susceptible cell types for SARS-CoV-2. In the heart dataset, fibroblasts and smooth muscle cells had high expression of all six confirmed receptors, which indicates that they may be potentially associated with SARS-CoV-2 pathogenesis. Similarly, collecting duct cells and proximal tubule cells may be associated with SARS-CoV-2 infection in the kidney. For the liver, stellate cells and noninflammatory macs cells show the highest risk for SARS-CoV-2 infection. However, for bladder single-cell data, it was difficult to identify the possible cell types, which need further investigation.

5. Conclusion

In summary, we reanalysed the public single-cell RNA sequencing dataset to understand the possible tropism of SARS-CoV-2. By comprehensive bioinformatics analysis of a lung single-cell RNA sequencing dataset of COVID-19 patients and healthy donors, macrophages were identified as the most likely cells that may be associated with SARS-CoV-2 pathogenesis. Further analysis of macrophages in lung demonstrates that up-regulation of chemokine pathways generally lead to severe symptoms, while down-regulation of ribosome and RNA activity related pathways is more likely to be mild. Other organ-specific susceptible cell type analyses could also provide a potential target for COVID-19 therapy [62,63]. Of course, we admit that there are certain limitations, such as no experimental validation and limited receptor genes [64,65]. Nonetheless, our studies may provide potential clues for further investigation of the pathogenesis of COVID-19.

Availability

Code used in this study is available from https://github.com/ZilongZhang44/COVID-19.

Funding

The work was supported by the National Natural Science Foundation of China (No. 61922020, No. 61771331) and the Special Science Foundation of Quzhou (2021D004).

Declaration of competing interest

None declared.

References

[1] P. Zhou, X.L. Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B. Li, C. L. Huang, et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature 579 (7798) (2020) 270–273.
[2] L. Cheng, X. Han, Z. Zhu, C. Qi, P. Wang, X. Zhang, Functional alterations caused by mutations reflect evolutionary trends of SARS-CoV-2, Briefings Bioinf. (2021).
[3] Y. Wang, F. Li, Y. Zhang, Y. Zhou, Y. Tan, Y. Chen, F. Zhu, Databases for the targeted COVID-19 therapeutics, Br. J. Pharmacol. 177 (21) (2020) 4999–5001.
[4] T.G. Ksiazek, D. Erdman, C.S. Goldsmith, S.R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J.A. Comer, W. Lim, et al., A novel coronavirus associated with severe acute respiratory syndrome, N. Engl. J. Med. 348 (20) (2003) 1953–1966.
[5] A.M. Zaki, S. van Boheemen, T.M. Bestebroer, A.D. Osterhaus, R.A. Fouchier, Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia, N. Engl. J. Med. 367 (19) (2012) 1814–1820.
[6] L. Zou, F. Ruan, M. Huang, L. Liang, H. Huang, Z. Hong, J. Yu, M. Kang, Y. Song, J. Xia, et al., SARS-CoV-2 viral load in upper respiratory specimens of infected patients, N. Engl. J. Med. 382 (12) (2020) 1177–1179.
[7] M.M. Lamers, J. Beumer, J. van der Vaart, K. Knoops, J. Puschhof, T.I. Breugem, R. B.G. Ravelli, J. Paul van Schayck, A.Z. Mykytyn, H.Q. Duimel, et al., SARS-CoV-2 productively infects human gut enterocytes, Science 369 (6499) (2020) 50–54.
[8] Y. Wang, S. Zhang, F. Li, Y. Zhou, Y. Zhang, Z. Wang, R. Zhang, J. Zhu, Y. Ren, Y. Tan, et al., Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics, Nucleic Acids Res. 48 (D1) (2020) D1031–d1041.
[9] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, et al., Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor, Nature 581 (7807) (2020) 215–220.
[10] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, J.S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, Science 367 (6483) (2020) 1260.
[11] W. Li, M.J. Moore, N. Vasilieva, J. Sui, S.K. Wong, M.A. Berne, M. Somasundaran, J.L. Sullivan, K. Luzuriaga, T.C. Greenough, et al., Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus, Nature 426 (6965) (2003) 450–454.
[12] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.H. Wu, A. Nitsche, et al., SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor, Cell 181 (2) (2020) 271–280, e278.
[13] A. Dey, S. Sen, U. Maulik, Unveiling COVID-19-associated organ-specific cell types and cell-specific pathway cascade, Briefings Bioinf. 22 (2) (2021) 914–923.
[14] W. Sungnak, N. Huang, C. Bécavin, M. Berg, R. Queen, M. Litvinukova, C. Talavera-López, H. Maatz, D. Reichart, F. Sampaziotis, et al., SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes, Nat. Med. 26 (5) (2020) 681–687.
[15] L. Cantuti-Castelvetri, R. Ojha, L.D. Pedro, M. Djannatian, J. Franz, S. Kuivanen, F. van der Meer, K. Kallio, T. Kaya, M. Anastasina, et al., Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity, Science 370 (6518) (2020) 856–860.
[16] C. Wu, M. Zheng, Y. Yang, X. Gu, K. Yang, M. Li, Y. Liu, Q. Zhang, P. Zhang, Y. Wang, et al., Furin: a potential therapeutic target for COVID-19, iScience 23 (10) (2020) 101642.
[17] M.-M. Zhao, W.-L. Yang, F.-Y. Yang, L. Zhang, W. Huang, W. Hou, C. Fan, R. Jin, Y. Feng, Y. Wang, et al., Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development, medRxiv (2020), 2020.2010.2025.20218990.
[18] C. Wei, L. Wan, Q. Yan, X. Wang, J. Zhang, X. Yang, Y. Zhang, C. Fan, D. Li, Y. Deng, et al., HDL-scavenger receptor B type 1 facilitates SARS-CoV-2 entry, Nat. metab. 2 (12) (2020) 1391–1400.

[19] S. Wang, Z. Qiu, Y. Hou, X. Deng, W. Xu, T. Zheng, P. Wu, S. Xie, W. Bian, C. Zhang, et al., AXL is a candidate receptor for SARS-CoV-2 that promotes infection of pulmonary and bronchial epithelial cells, Cell Res. (2021) 1–15.
[20] X. Zhu, H.-D. Li, L. Guo, F.-X. Wu, J. Wang, Analysis of single-cell RNA-seq data by clustering approaches, Curr. Bioinf. 14 (4) (2019) 314–322.
[21] R. Qi, A. Ma, Q. Ma, Q. Zou, Clustering and classification methods for single-cell RNA-sequencing data, Briefings Bioinf. 21 (4) (2020) 1196–1208.
[22] Z. Wang, H. Ding, Q. Zou, Identifying cell types to interpret scRNA-seq data: how, why and more possibilities, Brief. Funct. Geno. 19 (4) (2020) 286–291.
[23] S. Jin, X. Zeng, F. Xia, W. Huang, XJBiB. Liu, Application of deep learning methods in biological networks, Briefings Bioinf. (2020), https://doi.org/10.1093/bib/bbaa043.
[24] X. Liu, Z. Hong, J. Liu, Y. Lin, A. Rodríguez-Patón, Q. Zou, XJBib Zeng, Computational methods for identifying the critical nodes in biological networks, Briefings Bioinf. 21 (2) (2020) 486–497.
[25] IRIS3: Integrated Cell-type-specific Regulon Inference Server from Single-Cell RNA-Seq.
[26] J. Xie, A. Ma, Y. Zhang, B. Liu, S. Cao, C. Wang, J. Xu, C. Zhang, Q. Ma, QUBIC2: a novel and robust biclustering algorithm for analyses and interpretation of large-scale RNA-Seq data, Bioinformatics 36 (4) (2020) 1143–1149.
[27] Y. Li, A. Ma, E.A. Mathé, L. Li, B. Liu, Q. Ma, Elucidation of biological networks across complex diseases using single-cell omics, Trends Genet. 36 (12) (2020) 951–966.
[28] D. Ragab, H. Salah Eldin, M. Taeimah, R. Khattab, R. Salem, The COVID-19 cytokine storm; what we know so far, Front. Immunol. 11 (2020) 1446.
[29] J. Wang, A. Ma, Y. Chang, J. Gong, Y. Jiang, H. Fu, C. Wang, R. Qi, Q. Ma, D. Xu, scGNN: a novel graph neural network framework for single-cell RNA-Seq analyses, bioRxiv (2020), 2020.2008.2002.233569.
[30] Y. Chang, C. Allen, C. Wan, D. Chung, C. Zhang, Z. Li, Q. Ma, IRIS-FGM: an integrative single-cell RNA-Seq interpretation system for functional gene module analysis, bioRxiv (2020), 2020.2011.2004.369108.
[31] Z. Zhang, F. Cui, C. Lin, L. Zhao, C. Wang, Q. Zou, Critical downstream analysis steps for single-cell RNA sequencing data, Briefings Bioinf. (2021).
[32] Z. Zhang, F. Cui, C. Wang, L. Zhao, Q. Zou, Goals and approaches for each processing step for single-cell RNA sequencing data, Briefings Bioinf. (2020).
[33] T. Zhu, J. Guan, H. Liu, S. Zhou, RMDB: an integrated database of single-cytosine-resolution DNA methylation in oryza sativa, Curr. Bioinf. 14 (6) (2019) 524–531.
[34] D.C. Fajgenbaum, C.H. June, Cytokine Storm, N. Engl. J. Med. 383 (23) (2020) 2255–2273.
[35] P. Wang, X. Jin, W. Zhou, M. Luo, Z. Xu, C. Xu, Y. Li, K. Ma, H. Cao, Y. Huang, et al., Comprehensive analysis of TCR repertoire in COVID-19 using single cell sequencing, Genomics 113 (2) (2020) 456–462.
[36] F. Li, M. Luo, W. Zhou, J. Li, X. Jin, Z. Xu, L. Juan, Z. Zhang, Y. Li, R. Liu, et al., Single cell RNA and immune repertoire profiling of COVID-19 patients reveal novel neutralizing antibody, Protein Cell (2020).
[37] M. Liao, Y. Liu, J. Yuan, Y. Wen, G. Xu, J. Zhao, L. Cheng, J. Li, X. Wang, F. Wang, et al., Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19, Nat. Med. 26 (6) (2020) 842–844.
[38] C. Morse, T. Tabib, J. Sembrat, K.L. Buschur, H.T. Bittar, E. Valenzi, Y. Jiang, D. J. Kass, K. Gibson, W. Chen, et al., Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis, Eur. Respir. J. 54 (2) (2019).
[39] L. Wang, P. Yu, B. Zhou, J. Song, Z. Li, M. Zhang, G. Guo, Y. Wang, X. Chen, L. Han, et al., Single-cell reconstruction of the adult human heart during heart failure and recovery reveals the cellular landscape underlying cardiac function, Nat. Cell Biol. 22 (1) (2020) 108–119.
[40] S.A. MacParland, J.C. Liu, X.Z. Ma, B.T. Innes, A.M. Bartczak, B.K. Gage, J. Manuel, N. Khuu, J. Echeverri, I. Linares, et al., Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations, Nat. Commun. 9 (1) (2018) 4383.
[41] J. Liao, Z. Yu, Y. Chen, M. Bao, C. Zou, H. Zhang, D. Liu, T. Li, Q. Zhang, J. Li, et al., Single-cell RNA sequencing of human kidney, Sci Data 7 (1) (2020) 4.
[42] Z. Yu, J. Liao, Y. Chen, C. Zou, H. Zhang, J. Cheng, D. Liu, T. Li, Q. Zhang, J. Li, et al., Single-cell transcriptomic map of the human and mouse bladders, J. Am. Soc. Nephrol. 30 (11) (2019) 2159–2176.
[43] A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species, Nat. Biotechnol. 36 (5) (2018) 411–420.
[44] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theor. Exp. 2008 (10) (2008) P10008.
[45] J.H. Leland McInnes, Melville James, UMAP: uniform Manifold approximation and projection for dimension reduction, 2018 *arXiv:180203426*.
[46] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, Omics 16 (5) (2012) 284–287.
[47] S. Jin, C.F. Guerrero-Juarez, L. Zhang, I. Chang, R. Ramos, C.H. Kuan, P. Myung, M. V. Plikus, Q. Nie, Inference and analysis of cell-cell communication using CellChat, Nat. Commun. 12 (1) (2021) 1088.
[48] S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7.
[49] X. Qiu, A. Hill, J. Packer, D. Lin, Y.A. Ma, C. Trapnell, Single-cell mRNA quantification and differential analysis with Census, Nat. Methods 14 (3) (2017) 309–315.
[50] E.Y. Cao, J.F. Ouyang, O.J.L. Rackham, GeneSwitches: ordering gene expression and functional events in single-cell experiments, Bioinformatics 36 (10) (2020) 3273–3275.
[51] J. Lv, Z. Wang, Y. Qu, H. Zhu, Q. Zhu, W. Tong, L. Bao, Q. Lv, J. Cong, D. Li, et al., Distinct uptake, amplification, and release of SARS-CoV-2 by M1 and M2 alveolar macrophages, Cell discovery 7 (1) (2021) 24.
[52] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U.S.A. 102 (43) (2005) 15545–15550.
[53] Y. Hu, J.Y. Sun, Y. Zhang, H. Zhang, S. Gao, T. Wang, Z. Han, L. Wang, B.L. Sun, G. Liu, rs1990622 variant associates with Alzheimer's disease and regulates TMEM106B expression in human brain tissues, BMC Med. 19 (1) (2021) 11.
[54] Y. Hu, S. Qiu, L. Cheng, Integration of multiple-omics data to analyze the population-specific differences for coronary artery disease, Comput. Math. Meth. Med. 2021 (2021) 7036592.
[55] Y. Hu, H. Zhang, B. Liu, S. Gao, T. Wang, Z. Han, P. International Genomics of Alzheimer's, X. Ji, G. Liu, rs34331204 regulates TSPAN13 expression and contributes to Alzheimer's disease with sex differences, Brain 143 (11) (2020) e95.
[56] W.J. Wiersinga, A. Rhodes, A.C. Cheng, S.J. Peacock, H.C. Prescott, Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review, Jama 324 (8) (2020) 782–793.
[57] Q. Yang, B. Li, J. Tang, X. Cui, Y. Wang, X. Li, J. Hu, Y. Chen, W. Xue, Y. Lou, et al., Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data, Briefings Bioinf. 21 (3) (2020) 1058–1068.
[58] H.P. Jia, D.C. Look, L. Shi, M. Hickey, L. Pewe, J. Netland, M. Farzan, C. Wohlford-Lenane, S. Perlman, P.B. McCray Jr., ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia, J. Virol. 79 (23) (2005) 14614–14621.
[59] Y. Tang, J. Liu, D. Zhang, Z. Xu, J. Ji, C. Wen, Cytokine storm in COVID-19: the current evidence and treatment strategies, Front. Immunol. 11 (2020) 1708.
[60] U. Jaggi, H.H. Matundan, J. Yu, S. Hirose, M. Mueller, F.L. Wormley Jr., H. Ghiasi, Essential role of M1 macrophages in blocking cytokine storm and pathology associated with murine HSV-1 infection, PLoS Pathog. 17 (10) (2021), e1009999.
[61] D.A. Berlin, R.M. Gulick, F.J. Martinez, Severe covid-19, N. Engl. J. Med. 383 (25) (2020) 2451–2460.
[62] X. Zeng, X. Song, T. Ma, X. Pan, Y. Zhou, Y. Hou, Z. Zhang, K. Li, G. Karypis, FJJopr Cheng, Repurpose open data to discover therapeutics for COVID-19 using deep learning, J. Proteome Res. 19 (11) (2020) 4624–4636.
[63] X. Zeng, S. Zhu, W. Lu, Z. Liu, J. Huang, Y. Zhou, J. Fang, Y. Huang, H. Guo, L.J.C. S. Li, Target identification among known drugs by deep learning from heterogeneous networks, Chem. Sci. 11 (7) (2020) 1775–1797.
[64] X. Zeng, Y. Lin, Y. He, L. Lv, X. Min, Deep collaborative filtering for prediction of disease genes, IEEE ACM Trans. Comput. Biol. Bioinf 17 (5) (2020) 1639–1647.
[65] X. Zhang, Q. Zou, A. Rodriguez-Paton, X. Zeng, Bioinformatics, Meta-path methods for prioritizing candidate disease miRNAs, IEEE ACM Trans. Comput. Biol. Bioinf 16 (1) (2017) 283–291.

Zilong Zhang is currently working as a postdoctoral researcher in the University of Electronic Science and Technology of China. He received his Ph.D. degree from the University of Tokyo, Japan in 2020. His research interests include single-cell sequencing data analysis, bioinformatics and machine learning.

Feifei Cui received her Ph.D. degree from the University of Tokyo, Japan. She is currently a postdoctoral researcher at the University of Electronic Science and Technology of China. Her research interests include bioinformatics, deep learning and biological data mining.

Chen Cao is currently working as an associate professor in the Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China. He received his Ph.D. degree from the Jilin University in 2016. His research interests include bioinformatics, machine learning.

Qingsuo Wang is currently working in the Beidahuang Industry Group General Hospital, Harbin 150001, China. His research interests include data mining and bioinformatics.

Quan Zou is a professor at the Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China. He received his Ph.D. from Harbin Institute of Technology, P.R.China in 2009. He is a senior member of IEEE and ACM. His research is in the areas of bioinformatics, machine learning and parallel computing.