Supplementary experimental and analytical procedures

Data and Patient Sources. We got fifteen adult human(Supplementary Table 4) single-cell expression matrix of normal kidney including five kidney tissue samples of Occidental and ten kidney samples of Asian donors. The five expression matrix of Occidental were obtained from the research published by Matthew D. Young et al[1]. The three of the ten expression matrix of Asian race were obtained from the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/) uploaded by Liao jinling et al[2] with the GEO ID of GSE131685. The seven remaining expression matrix of Asian race are derived from the scRNA-seq data of seven normal kidney tissues from the Gongli Hospital of Second Military Medical University (Shanghai, China).

Ethical approval. We received approval from the Institution Review Board (IRB) of the Gongli Hospital of Second Military Medical University, and signed informed consent was obtained from seven patients.

Preparation of normal kidney tissue suspension. Seven fresh normal human kidney tissues were collected from the Gongli Hospital of Second Military Medical University. Three out of the seven kidney samples were obtained from the patient undergone partial nephrectomy and the other four kidney samples were obtained from patient undergone radical nephrectomy.

The kidney samples removed from the patient in operating room were immediately put into the 4℃ physiological saline and sent to the laboratory in 30 minutes. Samples were cut into tiny sections less than 1mm^3 in volume on ice. After washed twice with cold phosphate buffer saline (PBS), the samples were moved into 15mL centrifuge tube containing 10mL digestive medium with 0.75 mg/ml collagenase I (Sigma), 2mg/ml collagenase Ⅳ (Sigma), 0.0025 mg/ml DNase Ⅳ (Sigma), and 0.2mg/ml hyaluronidase I-S (Sigma) in 0.25% Trypsin (ThermoFisher Scientific). The mixed digestive medium were

incubated in 37℃ thermostatic shaker in 100 rounds per minute until the mediums became sticky. Then the medium were filtered by 40-μm nylon meshes (Corning) to remove cell debris and clamps then centrifuged at 300g for ten minutes in 4℃. The cell precipitation were resuspended with 1X red blood lysis buffer and incubated on ice for 5 minutes. Cell suspension were centrifuged again under same condition and the cell precipitation were resuspended in 50μL PBS. Cell concentration were determined in trypan blue(0.4%, Gibco) with hematocytometer. If the cell viability were above 70%, cells were performed under the 10X Genomics protocol.

ScRNA-seq and quality control. Cell suspension were sequenced on Illumia Hiseq X and labeled in the human genome(build Grch38). The cells were treated with the single-cell 3' Reagents Kit v2 and the target cell capture number was 10000. The protocol was strictly performed under the guideline of 10x Genomics official website (https://www.10xgenomics.com/). Preliminary sequencing data were transferred to FASTQ files with CellRanger(version 3.0.2) and were imported to R(version 3.6.1) and integrated with Seurat R package(version 2.3.4)[3]. To obtain high quality sequencing data ,we performed the preliminary quality control. Low quality cells with fewer than 200 or greater than 5000 genes and cells with more than 20% mitochondrial and ribosomal genes were eliminated.

Dimension Reduction and Classification of principal cell types. After quality control, the highly variable genes(HVGs) were screened under the help of function "FindVariableFeatures". Different datasets from different source were merged into one dataset with the function of "FindIntegrationAnchors" and "IntegrateData". Principal component analysis(PCA) then was conducted to reduce dimension and the Louvain graph-clustering method[4] was used to distinguish the cell clusters and the Uniform manifold approximation and projection (UMAP) was used for visualization. We annotated all cell clusters

based on the known marker gene in published articles. Then, we used the "RunALRA" function in Seurat to impute missing values in the scRNA-seq data. All the data of lung and digestive duct was provided by the author [5] and pathways related to endocytosis or exocytosis were obtained from Harmonizome dataset [6].

Geneset variation analysis(GSVA). Pathway analysis were conducted by GSVA package(version 1.30.0) with default setting to estimate the pathway activity of each cells. The annotation genesets were loaded from the molecular signature database(MSigDB, version 7.0).

The analysis of differentiation of Renal epithelial cells by monocle3 and Velocity. The future state and pseudotime trajectories were ploted by Package Monocle3, and monocle3 cluster the cells into subclusters after dimension reduction. The expression level of genes were calculated, scaled by log transition. The package velocyto.R [7] was used to annotate the spliced and unspliced RNA to dertermine the direction of differentiation of PT cells and the velocity field was plotted on the UMAP plot.

Statistical analysis. Violinplot, boxplot and dotplot were used to display the expression of corresponding genes. All single-cell sequencing data statistical analysis was performed in R (version 3.6.1) and Statistical significance was accepted for $P < 0.05$. In scRNA-seq data, we use Limma package to fit a generalized linear model and use Bayes moderated F-statistic to determine statistical significance. The expression value of RNA was downloaded from Genotype-Tissue Expression (GTEx) [8] and be plotted by graphpad prism(version 8.2.1).

[1] Young MD, Mitchell TJ, Vieira Braga FA, Tran MGB, Stewart BJ, Ferdinand JR, et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. Science. 2018;361:594-9.

[2] Liao J, Yu Z, Chen Y, Bao M, Zou C, Zhang H, et al. Single-cell RNA sequencing of human kidney. Sci Data. 2020;7:4.

[3] Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019;177:1888-902 e21.

[4] Blondel VD, Guillaume JL, Hendrickx JM, de Kerchove C, Lambiotte R. Local leaders in random networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2008;77:036114.

[5] Meng T, Cao H, Zhang H, Kang Z, Xu D, Gong H, et al. The insert sequence in SARS-CoV-2 enhances spike protein cleavage by TMPRSS. bioRxiv. 2020:doi: https://doi.org/10.1101/2020.02.08.926006.

[6] Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Ma'Ayan A. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database the Journal of Biological Databases & Curation. 2016;2016:baw100.

[7] Gioele LM, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018.

[8] Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648-60.