Supplemental Material

Title: Urinary single cell profiling captures the cellular diversity of the kidney

Running title: Urinary single cell profiling

Name of Authors:

Amin Abedini[1,2,3]*, Yuan O. Zhu[4]*, Shatakshee Chatterjee[1,2,3]*, Gabor Halasz[4], Kishor Devalaraja-Narashimha[4], Rojesh Shrestha[1,2,3], Michael Balzer[1,2,3], Jihwan Park[1,2,3], Tong Zhou[1,2,3], Ziyuan Ma[1,2,3], Katie Marie Sullivan[1,2,3], Hailong Hu[1,2,3], Xin Sheng[1,2,3], Hongbo Liu[1,2,3], Yi Wei[4], Carine M Boustany-Kari[5], Uptal Patel[6], Salem Almaani[7], Matthew Palmer[8], Raymond Townsend[1], Shira Blady[1], Jonathan Hogan[1], and The TRIDENT Study Investigators, Lori Morton[4] and Katalin Susztak[1,2,3]

Affiliations:

[1]Renal, Electrolyte, and Hypertension Division, Department of Medicine, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA

[2]Institute for Diabetes, Obesity, and Metabolism, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA

[3]Department of Genetics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA

[4]Cardiovascular, Renal and Fibrosis Research, Regeneron Pharmaceuticals Inc., Tarrytown, NY 10591, USA

[5]Cardiometabolic Disease Research department, Boehringer Ingelheim, Ridgefield, CT 06877, USA

[6]Inflammation & Respiratory Therapeutics, Gilead Sciences Inc., Foster City, CA 94404, USA

[7] Division of Nephrology, The Ohio State University Wexner Medical Center, Columbus, OH, 43210, USA

[8]Department of Pathology, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA

*Contributed equally

Correspondence:

Katalin Susztak, MD, PhD

Professor of Medicine

University of Pennsylvania, Perelman School of Medicine

3400 Civic Center Blvd,

Smilow Translational building 12-123,

Philadelphia, PA 19104

Phone: (215) 898-2009

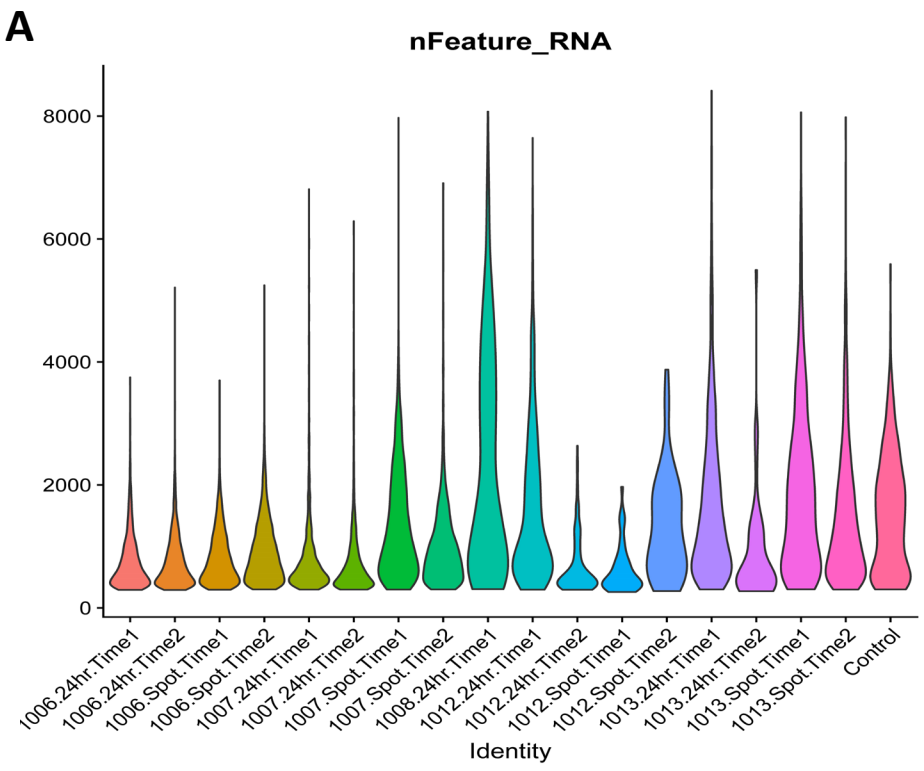
ksusztak@pennmedicine.upenn.edu

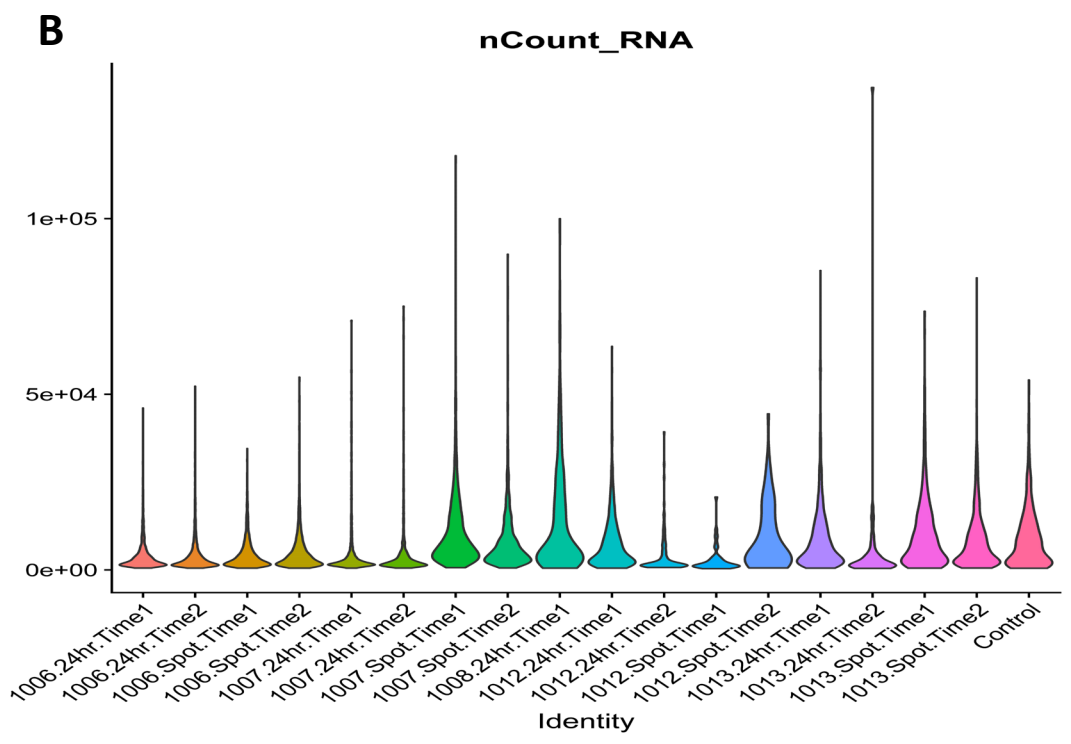
Table of Contents

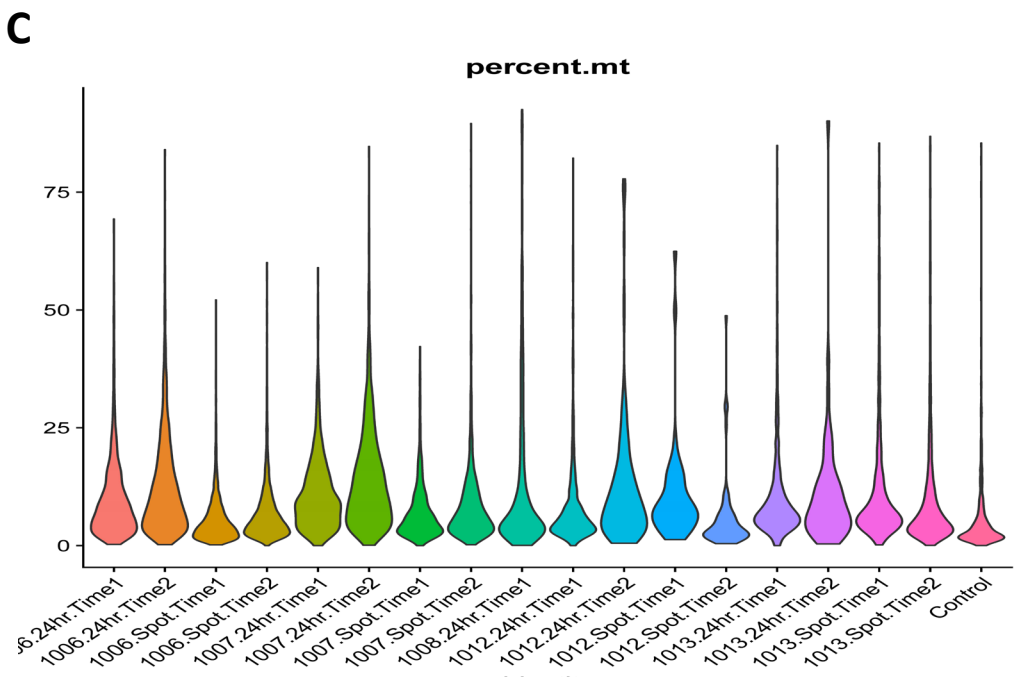
Supplemental Table 1. Quality control metrics of the urine single cell RNA sequencing data.

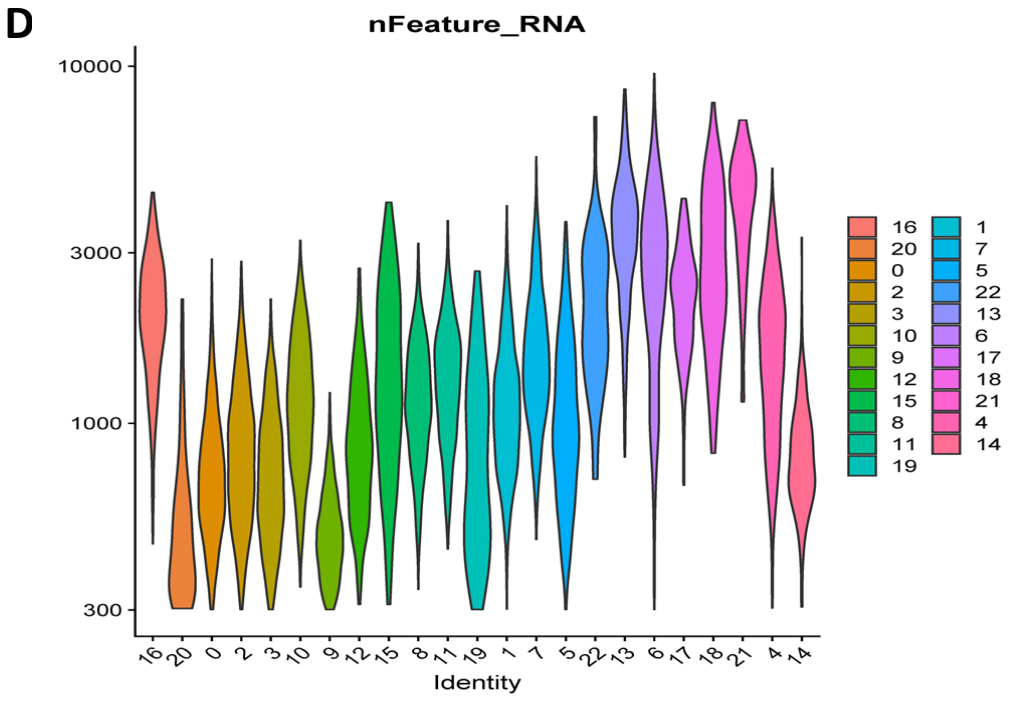
Samples	nFeature_RNA	nCount_RNA
1006-SV1	814.9046 ± 441.4593	4287.279 ± 4081.432
1006-SV2	910.7656 ± 499.4818	4855.382 ± 4881.045
1006-24V1	755.6622 ± 456.17	3816.164 ± 4058.087
1006-24V2	753.5157 ± 428.4311	3766.538 ± 4123.672
1007-SV1	1319.847 ± 851.5852	10130.54 ± 9390.385
1007-SV2	992.6016 ± 636.7436	7956.667 ± 8301.125
1007-24V1	801.1541 ± 674.0588	3434.917 ± 4852.797
1007-24V2	757.9126 ± 644.82	3517.402 ± 5005.482
1008-24V1	2250.244 ± 1737.415	12916.72 ± 14649.32
1012-SV1	626.3878 ± 392.3683	2812.143 ± 3646.245
1012-SV2	1327.446 ± 834.493	8993.386 ± 9102.408
1012-24V1	1583.826 ± 1194.67	7577.219 ± 7727.65
1012-24V2	685.0924 ± 472.3429	3598.13 ± 5075.144
1013-SV1	1888.4 ± 1330.256	9922.629 ± 9866.355
1013-SV2	1466.785 ± 1088.715	8295.849 ± 8234.268
1013-24V1	1651.951 ± 1247.172	8436.437 ± 8708.403
1013-24V2	898.0687 ± 882.9755	5649.186 ± 17857.66
Control	1520.351 ± 925.2031	8742.781 ± 8438.414

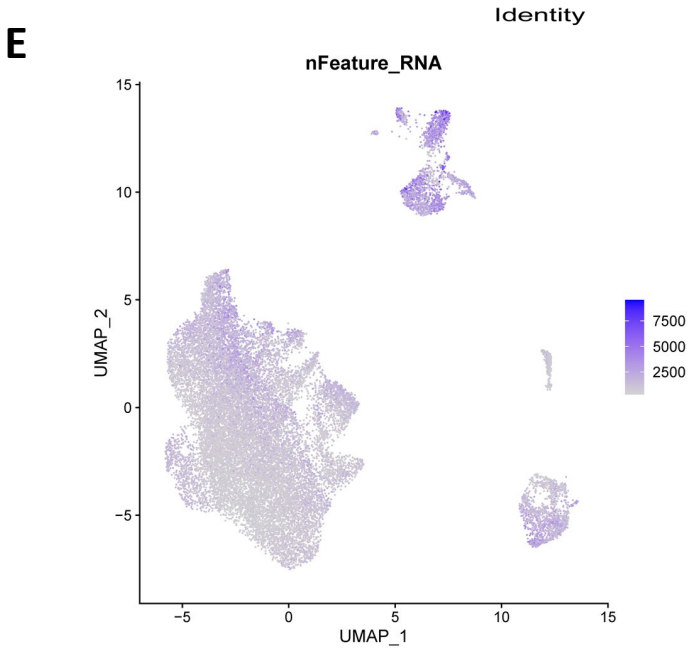
nFeature_RNA shows the number of genes detected in each cell before QC filtering. nCount_RNA indicates the number of genes detected in all cells before QC filtering. All data is shown as mean ± SD. SV1; spot urine sample in visit 1, SV2; spot urine sample In visit 2, 24V1; 24h urine sample in visit 1, 24V2; 24h urine sample in visit 2.

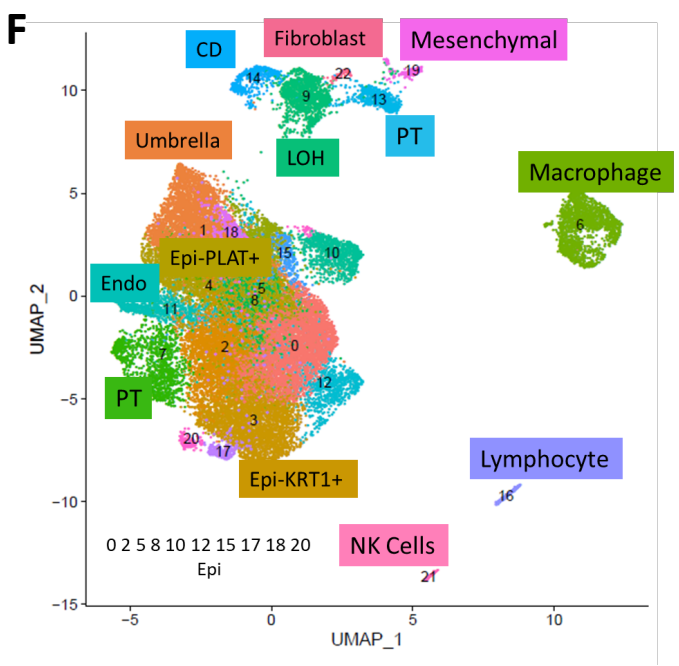


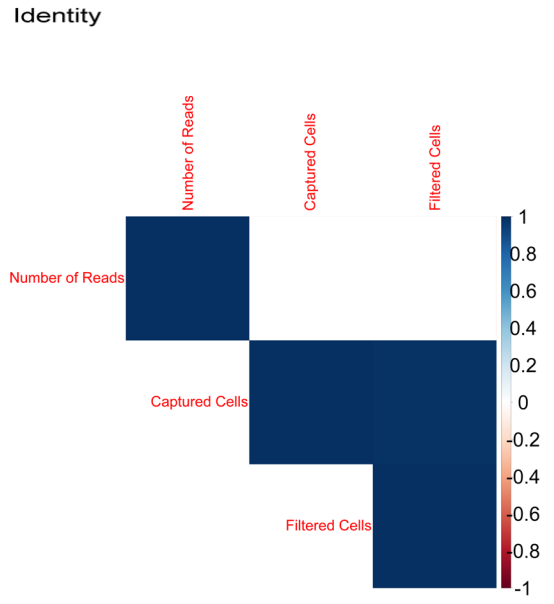


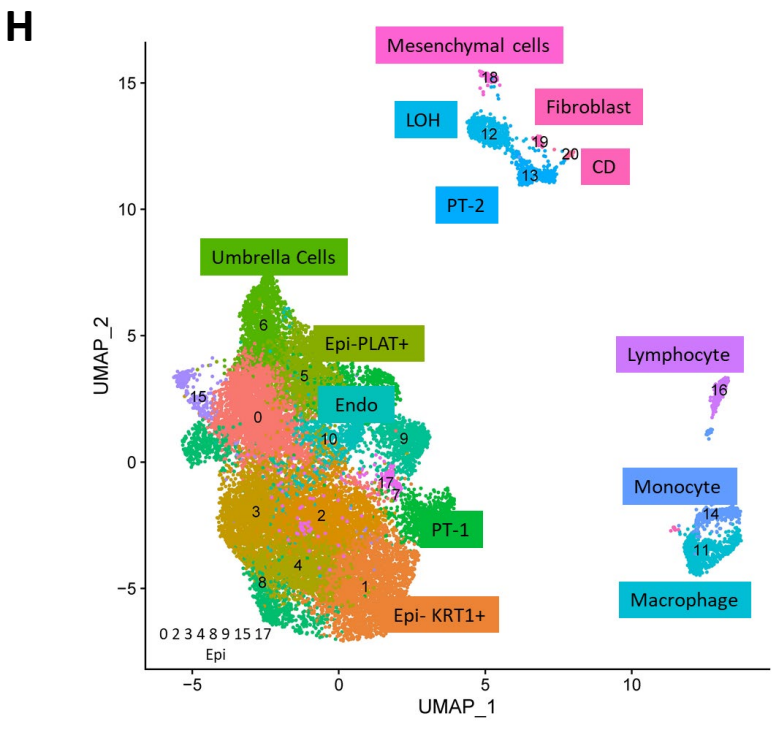


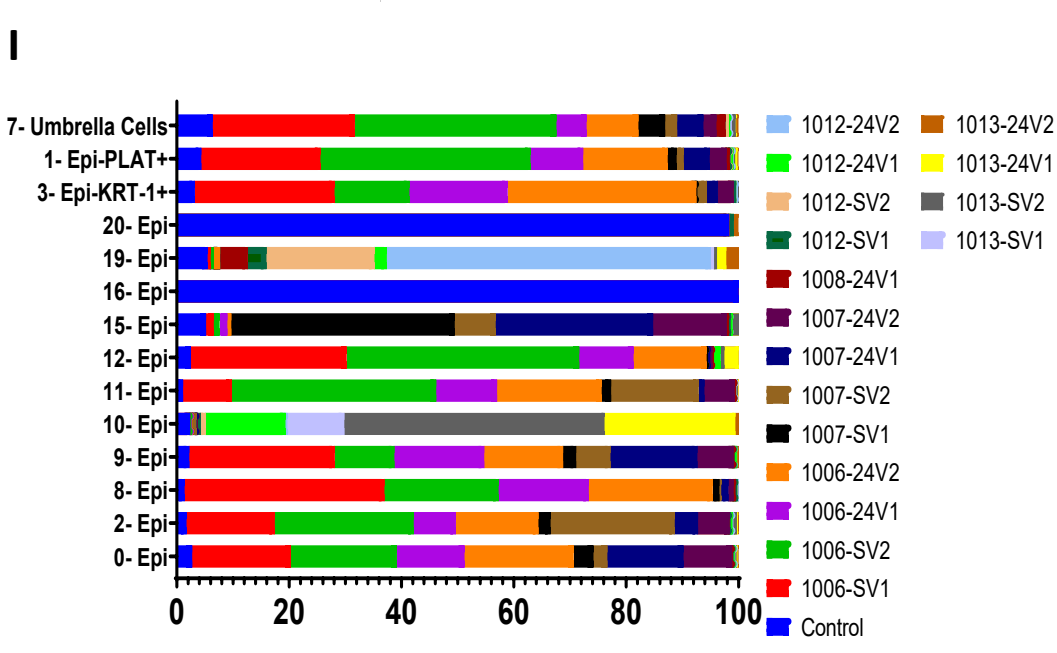












Supplemental figure 1. Quality control metrics of the urine single cell RNA sequencing data.

(**A**)Violin plots of the number of genes detected in each cell (nFeature_RNA) before QC filtering.

(**B**) Violin plots of the number of genes detected in all cells (ncount_RNA) before QC filtering.

(**C**) Mitochondrial percentages in different urine samples before QC filtering.

(**D**) Violin plots of the number of genes detected in each cell clusters (nFeature_RNA) after QC filtering

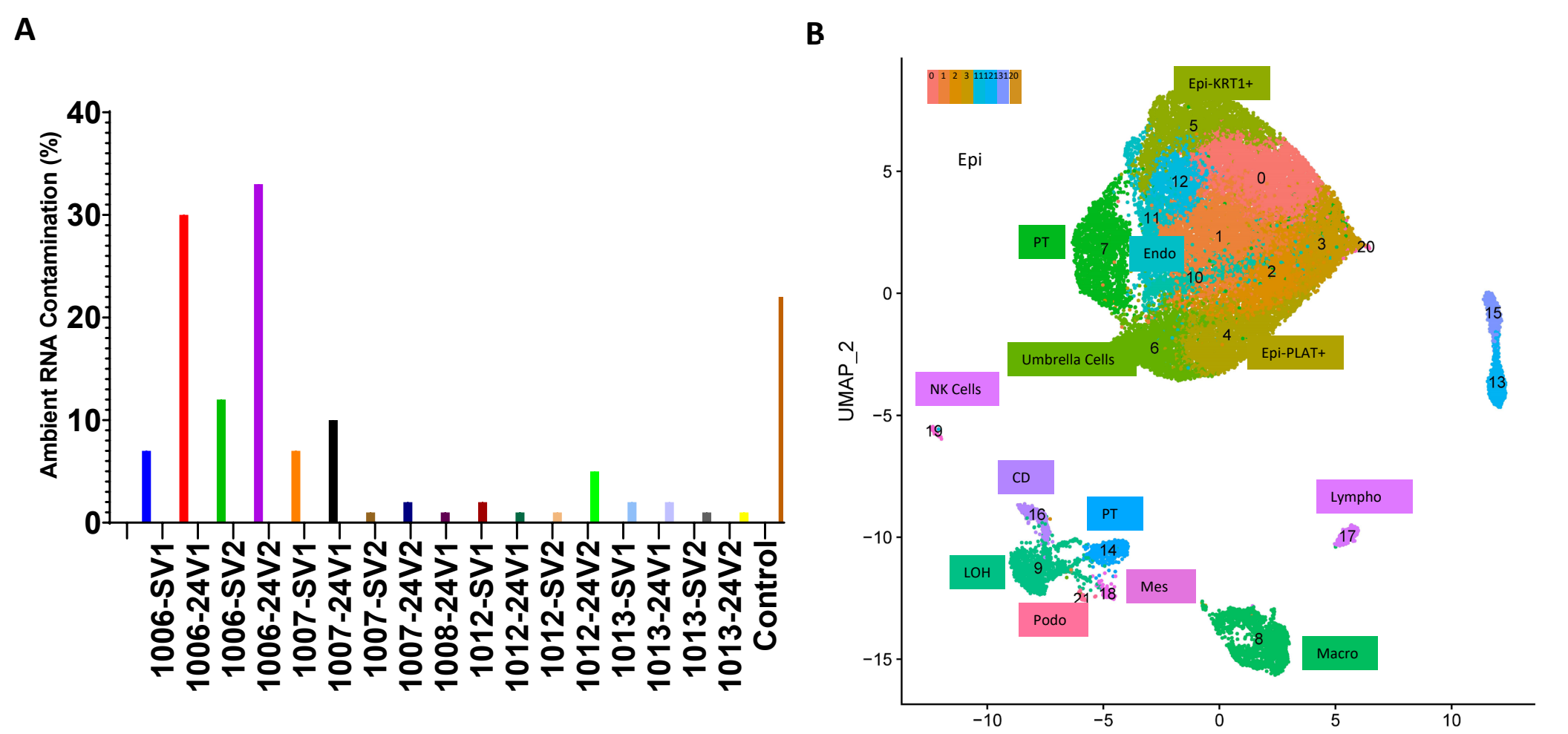
(**E**) Feature plot of nFeature_RNA in cell clusters after QC filtering.

(**F**) UMAP dimension reduction of 31,229 urinary cells with filtering out the cells with greater than 50% mitochondrial genes. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Umbrella; umbrella cells, PT; proximal tubule, LOH; loop of Henle, CD; collecting duct, Endo; endothelial cells, Mesenchymal; mesenchymal cells, NK cells; natural killer cells.

(**G**) Pearson correlation between total number of reads in each urine sample with captured cells and filtered cells. White means no correlation and blue indicates positive correlation.

(**H**) UMAP dimension reduction of 21,351 urinary cells after doublet removal. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, PT; proximal tubule, LOH; loop of Henle, CD; collecting duct, Endo; endothelial cells.

(**I**) The contribution of each sample (as a fraction) to the large epithelial cell group. SV1; spot urine sample in visit 1, SV2; spot urine sample In visit 2, 24V1; 24h urine sample in visit 1, 24V2; 24h urine sample in visit 2, Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells.



Supplemental figure 2. Ambient RNA contamination of urine single cell samples.

(**A**) The fraction of ambient RNA in each urine sample obtained by SoupX package.

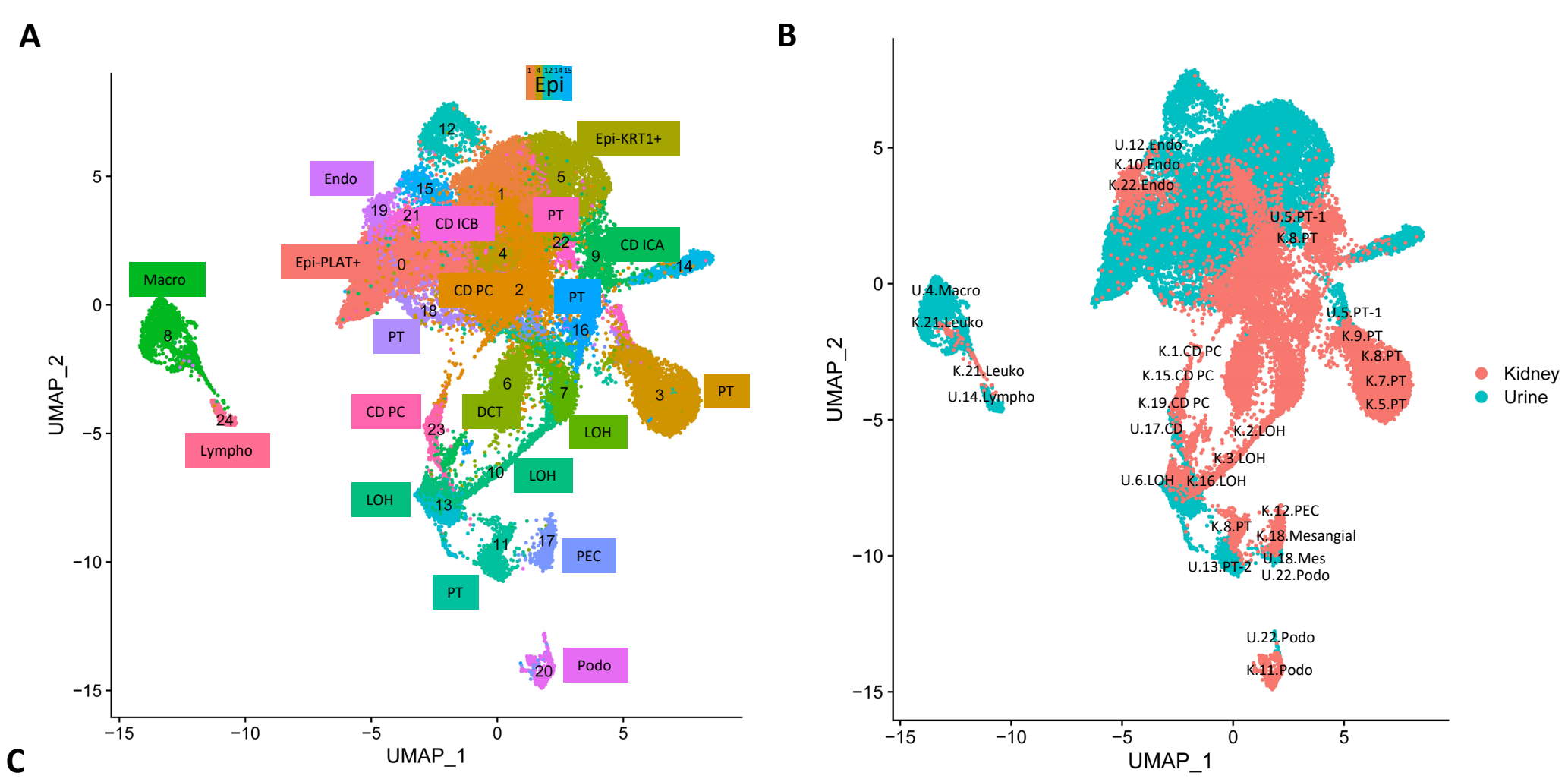
(**B**) UMAP dimension reduction of 23,089 urinary cells after ambient RNA correction. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Endo; endothelial cells, Podo; podocytes, PT; proximal tubule, LOH; loop of Henle, CD; collecting duct principal cell, Mes; mesenchymal cells. Macro; macrophages, Lympho; lymphocytes, NK cells; natural killer cells.

(**C**) UMAP dimension reduction of 23,089 urinary cells after ambient RNA correction using Harmony-based integration. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Endo; endothelial cells, Podo; podocytes, PT; proximal tubule, LOH; loop of Henle, CD; collecting duct principal cell, Macro; macrophages, Lympho; lymphocytes, Mes; mesenchymal cells, NK cells; natural killer cells.

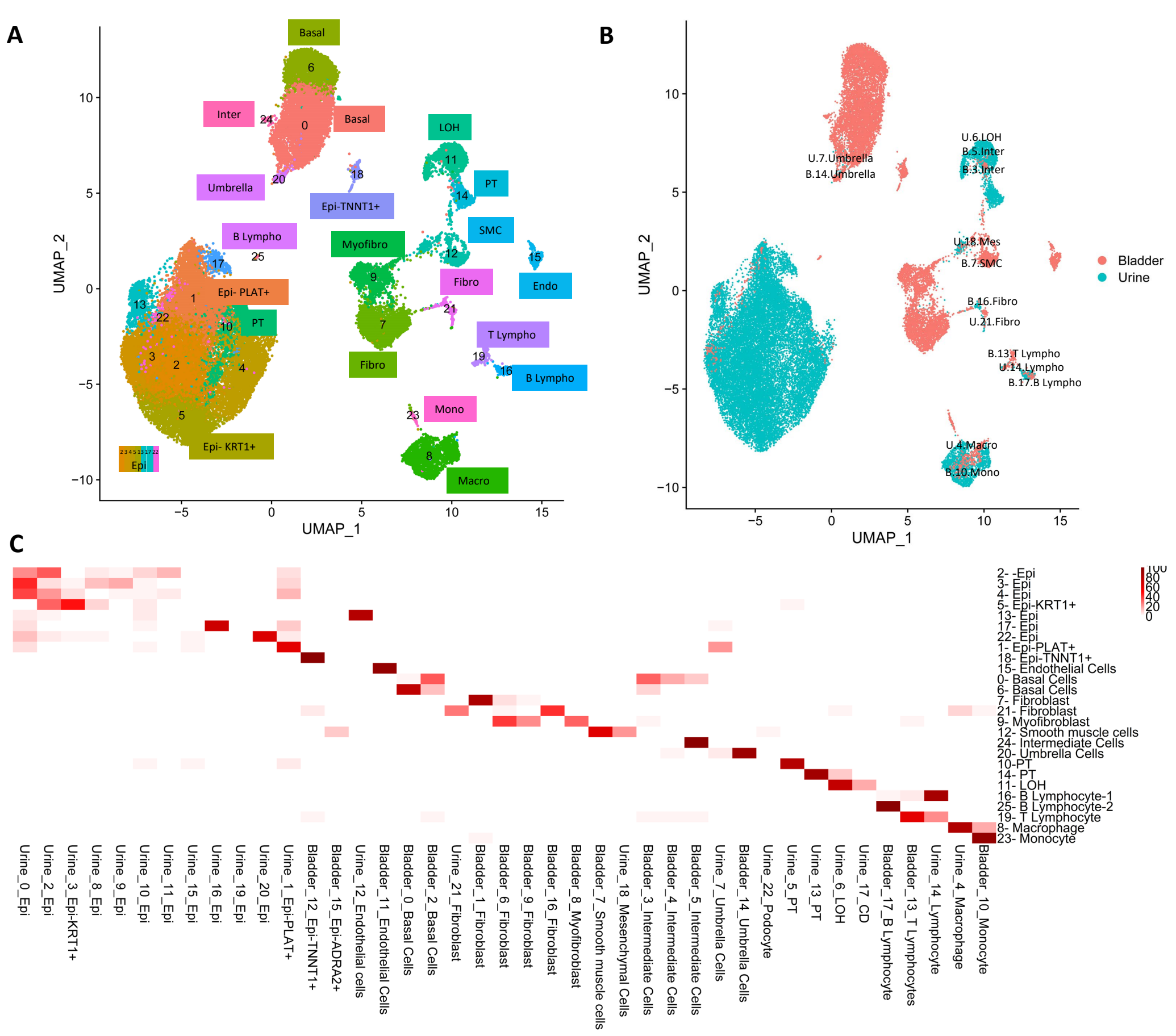
(**D**) Feature plots of *KRT13* (Keratin 13), *S100A8* (S100 calcium binding protein A8, calprotectin*) S100A9* (S100 calcium binding protein A9, calprotectin) without ambient RNA correction.

(**E**) Feature plots of *KRT13* (Keratin 13), *S100A8* (S100 calcium binding protein A8, calprotectin) *S100A9* (S100 calcium binding protein A9, calprotectin) after ambient RNA correction.

(**F**) Feature plots of *KRT13* (Keratin 13), *S100A8* (S100 calcium binding protein A8, calprotectin) *S100A9* (S100 calcium binding protein A9, calprotectin) after ambient RNA correction after harmony-based integration.



Supplemental figure 3. Integration of urine single cell with human DKD and control single nucleus kidney datasets using anchor method.
(**A**) UMAP of anchor-based integration of cells from urine, healthy and DKD kidney samples. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Umbrella; umbrella cells, Endo; endothelial cells, PT; proximal tubule, LOH; loop of Henle, CD PC; collecting duct principal cells, CD ICA; collecting duct intercalated cells A, CD ICB; collecting duct intercalated cells B, DCT; distal convoluted tubule, Podo; podocyte, PEC; parietal epithelial cells, Macro; macrophages, Lympho; lymphocytes.
(**B**) UMAP of anchor-based integration of healthy and DKD kidney urine colored by the sample of origin. Blue indicates urine cells pink shows kidney cells. The origins of the cells are shown on each plot. U; Urine, K; Kidney, Endo; endothelial cells, Podo; podocyte, PT; proximal tubule, LOH; loop of Henle, PEC; parietal epithelial cells, CD PC; collecting duct principal cell, CD; collecting duct, Mesangial; mesangial cells, Mes; mesenchymal cells, Macro; macrophage, Lympho; lymphocyte, Leuko; leukocyte.
(**C**) The percent of cells in each integrated cluster (Y-axis) that came from each original cluster (X-axis). Epi; Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, PT; proximal tubule, LOH; loop of Henle, DCT; distal convoluted tubule, CD PC; collecting duct principal cells, CD ICA; collecting duct intercalated cells A, CD ICB; collecting duct intercalated cells B.
(**D**) Bubble dot plots of the top cell-type specific genes in the integrated clustering of the urine and kidney samples. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Umbrella; umbrella cells, Endo; endothelial cells, PT; proximal tubule, LOH; loop of Henle, CD PC; collecting duct principal cells, CD ICA; collecting duct intercalated cells A, CD ICB; collecting duct intercalated cells B, DCT; distal convoluted tubule, Podo; podocyte, PEC; parietal epithelial cells, Macro; macrophages, Lympho; lymphocytes. The size of the dot indicates the percent positive cells and the darkness of the color indicates average expression.
(**E**) The fraction of cells in each integrated cluster and their sample of origin. Epi; Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, PT; proximal tubule, LOH; loop of Henle, DCT; distal convoluted tubule, CD PC; collecting duct principal cells, CD ICA; collecting duct intercalated cells A, CD ICB; collecting duct intercalated cells B.



Supplemental figure 4. Integration of urine and bladder single cell datasets by Harmony method.

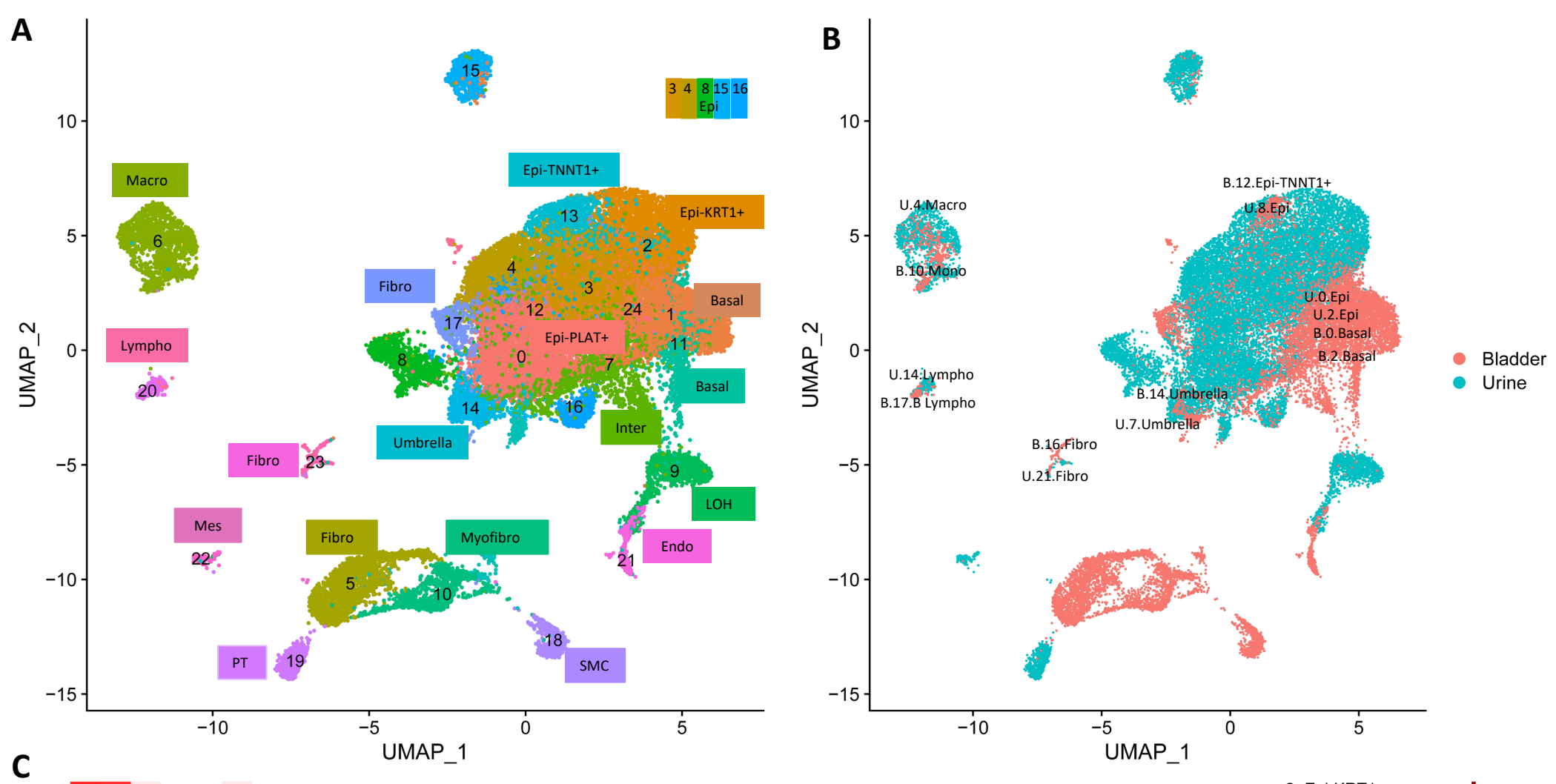
(**A**) UMAP of Harmony-based integration of urinary and bladder cells. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Basal; basal cells, Inter cells; intermediate cells, Epi-TNNT1+; TNNT1 positive cells, Endo; endothelial cells, PT; proximal tubule, LOH; loop of Henle, Fibro; fibroblast, Myofibro; myofibroblast, SMC; smooth muscle cell, Macro; macrophages, Mono; monocyte, Lympho; lymphocytes.

(**B**) UMAP of Harmony-based integration of urinary and bladder cells colored by the sample of origin. Blue indicates urine cells and pink shows bladder cells. The plot is labelled by cell of origin. U; Urine, B; bladder, Umbrella; umbrella cells, LOH; loop of Henle, Inter; intermediate cells, SMC; smooth muscle cell, Mes; mesenchymal cells, Macro; macrophages, Lympho; lymphocytes, Fibro; fibroblast.

(**C**) The percent of cells in each integrated cluster (Y-axis) that came from each original cluster (X-axis). Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Epi-TNNT1+; TNNT1 positive cells, Epi-ADRA2+; ADRA2 positive cells, PT; proximal tubule, LOH; loop of Henle.

(**D**) Bubble dot plots of the top cell-type specific genes in the integrated clusters of urine and bladder samples. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Basal; basal cells, Inter cells; intermediate cells, Epi-TNNT1+; TNNT1 positive cells, Endo; endothelial cells, PT; proximal tubule, LOH; loop of Henle, SMC; smooth muscle cell, Macro; macrophages, Mono; monocyte, Lympho; lymphocytes. The size of the dot indicates expression percentage and the darkness of the color indicates average expression.

(**E**) The fraction of cells in each integrated cluster and their sample of origin. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Epi-TNNT1+; TNNT1 positive cells, PT; proximal tubule, LOH; loop of Henle.



Supplemental figure 5. Integration of urine and bladder single cell datasets by anchor method.

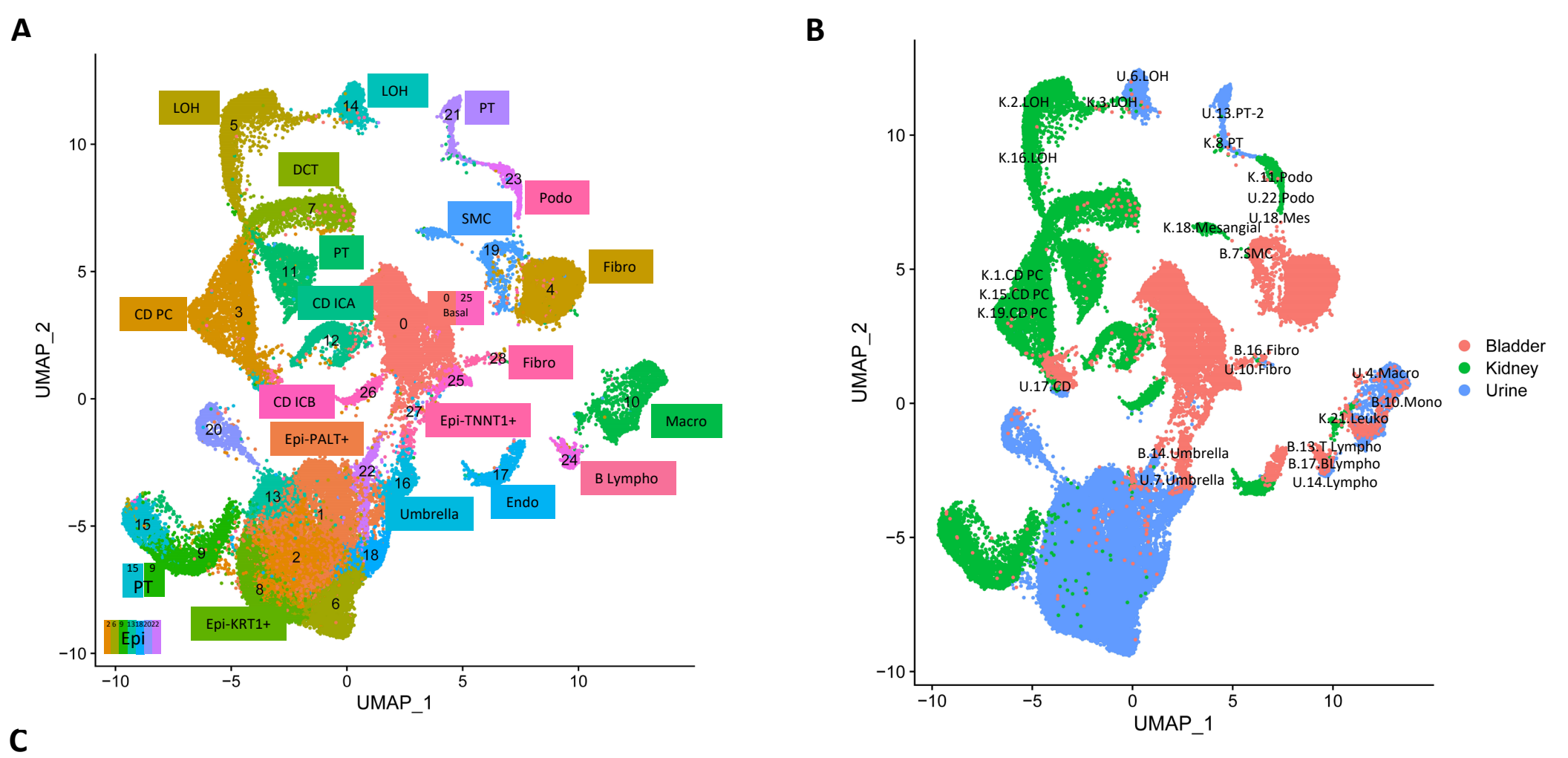
(**A**)UMAP of anchor-based integration urinary and bladder cells. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Basal; basal cells, Inter cells; intermediate cells, Umbrella; umbrella cells, Epi-TNNT1+; TNNT1 positive cells, Endo; endothelial cells, PT; proximal tubule, LOH; loop of Henle, Fibro; fibroblast, Myofibro; myofibroblast, SMC; smooth muscle cell, Macro; macrophages, Lympho; lymphocytes.

(**B**) UMAP of anchor-based integration of urinary and bladder cells, colored by the sample of origin. Blue indicates the urine origin and pink shows the cells originated from bladder. The origins of the cells are written in each plot. U; Urine, B; bladder, Epi; variety epithelial cells, Epi-TNNT1+; TNNT1 positive cells, Umbrella; umbrella cells, Fibro; fibroblast, Lympho; lymphocyte, Mono; monocyte, Macro; macrophage.

(**C**) The percent of cells in each integrated cluster (Y-axis) that came from each original cluster (X-axis). Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Epi-TNNT1+; TNNT1 positive cells, Epi-ADRA2+; ADRA2 positive cells, PT; proximal tubule, LOH; loop of Henle.

(**D**) Bubble dot plots of the top cell-type specific differentially expressed genes in integrated clusters of urine and bladder samples. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Basal; basal cells, Umbrella; umbrella cells, Inter cells; intermediate cells, Epi-TNNT1+; TNNT1 positive cells, Endo; endothelial cells, PT; proximal tubule, LOH; loop of Henle, SMC; smooth muscle cell, Macro; macrophages, Lympho; lymphocytes. The size of the dot indicates expression percentage and the darkness of the color indicates average expression.

(**E**) Heatmap of the fraction of cells in each integrated cluster. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-TNNT1+; TNNT1 positive cells, PT; proximal tubule, LOH; loop of Henle.



Supplemental figure 6. Integration of urine single cell, kidney single nucleus and bladder single cell datasets using the anchor method.

(**A**) UMAP of anchor-based integration of urinary, kidney and bladder cells. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Epi-TNNT1+; TNNT1 positive cells, Basal; basal cells, Inter; intermediate cells, Umbrella; umbrella cells, Endo; endothelial cells, PT; proximal tubule, LOH; loop of Henle, CD PC; collecting duct principal cells, CD ICA; collecting duct intercalated cells A, CD ICB; collecting duct intercalated cells B, DCT; distal convoluted tubule, Podo; podocyte, PEC; parietal epithelial cells, Fibro; fibroblast, Macro; macrophages, Lympho; lymphocytes.

(**B**) UMAP of anchor-based integration of urinary, kidney and bladder cells colored by the sample of origin. Blue, green, and pink indicates the urine, kidney, and bladder origins, respectively. The origins of the cells are written in each plot. U; Urine, K; Kidney, B; bladder, Epi-PLAT+; PLAT positive cells, Umbrella; umbrella cells, Podo; podocyte, PT; proximal tubule, CD PC; collecting duct principal cells, LOH; loop of Henle, Fibro; fibroblast, Mes; mesenchymal cells, Macro; macrophage, Mono; monocyte, Lympho; lymphocyte,

(**C**) The percent of cells in each integrated cluster (Y-axis) that came from each original cluster (X-axis). Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Epi-ADRA2+; ADRA2 positive cells, Epi-TNNT1+; TNNT1 positive cells, PT; proximal tubule, LOH; loop of Henle, CD PC; collecting duct principal cells, CD ICA; collecting duct intercalated cells A, CD ICB; collecting duct intercalated cells B, DCT; distal convoluted tubule.

(**D**) Bubble dot plots of the top cell-type specific differentially expressed genes in integrated clusters of urine, kidney, and bladder samples. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Epi-TNNT1+; TNNT1 positive cells, Basal; basal cells, Inter; intermediate cells, Umbrella; umbrella cells, Endo; endothelial cells, PT; proximal tubule, LOH; loop of Henle, CD PC; collecting duct principal cells, DCT; distal convoluted tubule, Podo; podocyte, PEC; parietal epithelial cells, Macro; macrophages, Lympho; lymphocytes. The size of the dot indicates expression percentage and the darkness of the color indicates average expression.

(**E**) Heatmap of the fraction of cells in each integrated cluster. Epi; variety epithelial cells, Epi-PLAT+; PLAT positive cells, Epi-KRT1+; KRT1 positive cells, Epi-TNNT1+; TNNT1 positive cells, PT; proximal tubule, LOH; loop of Henle, CD PC; collecting duct principal cells, CD ICA; collecting duct intercalated cells A, CD ICB; collecting duct intercalated cells B, DCT; distal convoluted tubule, SMC; smooth muscle cell.