



## Supplementary Information for

### **The Single Cell Transcriptomic Landscape of Early Human Diabetic Nephropathy**

Parker C. Wilson, Haojia Wu, Yuhei Kiritani, Kohei Uchimura, Nicolas Ledru, Helmut G. Rennke, Paul A. Welling, Sushrut S. Waikar, Benjamin D. Humphreys

Benjamin D. Humphreys, MD, PhD  
Division of Nephrology  
Washington University School of Medicine  
660 S. Euclid Ave., CB 8129  
St Louis, MO 63110  
humphreysbd@wustl.edu

#### **This PDF file includes:**

Supplementary text  
Figs. S1 to S7  
Tables S1 to S2

#### **Other supplementary materials for this manuscript include the following:**

Datasets S1 to S10

## **Supplementary Information Text**

### **Methods**

#### **Single-Nuclei Isolation and Library Preparation**

Nuclei were isolated with Nuclei EZ Lysis buffer (NUC-101; Sigma-Aldrich) supplemented with protease inhibitor (5892791001; Roche) and RNase inhibitor (N2615; Promega and AM2696; Life Technologies). Samples were cut into <2-mm pieces and homogenized using a Dounce homogenizer (885302-0002; Kimble Chase) in 2 ml of ice-cold Nuclei EZ Lysis buffer, and they were incubated on ice for 5 minutes with an additional 2 ml of lysis buffer. The homogenate was filtered through a 40- $\mu$ m cell strainer (43-50040-51; pluriSelect) and then centrifuged at 500 $\times$ g for 5 minutes at 4°C. The pellet was resuspended and washed with 4 ml of the buffer, and then, it was incubated on ice for 5 minutes. After another centrifugation, the pellet was resuspended in Nuclei Suspension Buffer (1 $\times$  PBS, 1% BSA, and 0.1% RNase inhibitor), filtered through a 5- $\mu$ m cell strainer (43-50020-50; pluriSelect), and counted. The 10 $\times$  Chromium single-cell protocol provided by the manufacturer (10 $\times$  Genomics) was used to generate high-quality cDNA libraries.

#### **Immunofluorescence Studies**

Formalin-fixed paraffin embedded tissue sections were deparaffinized and underwent antigen retrieval. Sections were washed with PBS, blocked with 10% normal goat serum (Vector Labs), permeabilized with 0.2% Triton-X100 in PBS and incubated overnight with primary antibodies for AQP3 (LSBio, LS-B8185), ATP1A1 (abcam, EP1845Y), AQP2 (Santa Cruz, sc-9882), or V-ATPase (Santa Cruz, sc-55544) followed by staining with secondary antibodies (FITC-, Cy3, or Cy5-conjugated, Jackson ImmunoResearch). Sections were stained with DAPI (4',6'-diamidino-2-phenylindole) and mounted in Prolong Gold (Life Technologies). Images were obtained by confocal microscopy (Nikon C2+ Eclipse; Nikon, Melville, NY). AQP3 and ATP1A1 were quantified in ImageJ by using the color threshold and measure functions to calculate integrated fluorescence density. Total corrected target immunofluorescence was calculated by subtracting the total principal cell area multiplied by the background fluorescence from the integrated fluorescence density. Results were compared between control and diabetic samples using a two-tailed students t-test.

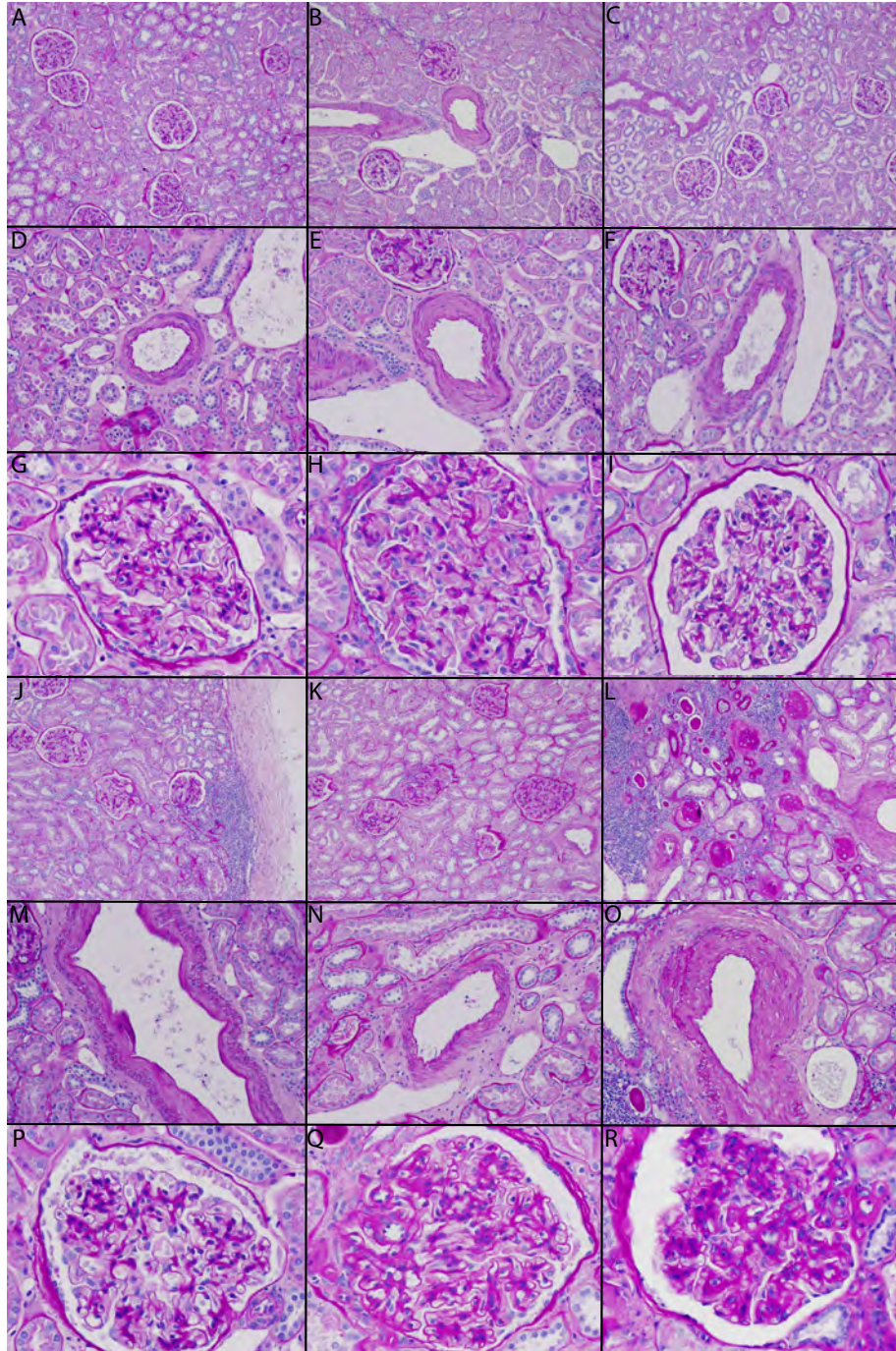
**Generation of count matrices.** Raw paired-end fastq files were aligned and counted using zUMIs v2.0 with default parameters for 10X genomics 5' chemistry. The reference index was generated by STAR v2.6.0 and GRCh38 release 93 downloaded from ensembl.org.

**Creation of Seurat objects.** Seurat v2.3 objects were initialized using intronic and exonic counts generated from zUMIs and CreateSeuratObject() with min.cells=3. Seurat objects were filtered using FilterCells() with low.thresholds=500 and high.thresholds=Inf. They were subsequently preprocessed using NormalizeData() with default parameters, FindVariableGenes() with mean.function=ExpMean, dispersion.function=LogVMR, x.low.cutoff=0.0125, x.high.cutoff=8, y.cutoff=1, and num.bin=30. ScaleData() was run with default parameters. RunPCA() was run with pcs.compute=40. FindClusters() was run with dims.use=1:20 and resolution=0.6.

**Integration of Seurat objects.** The three control and three diabetic Seurat objects were integrated using RunMultiCCA() with num.cc=12. The genes.use list was obtained by identifying the top 1000 variable genes from each object with FindVariableGenes() and merging the list. The integrated object was processed with AlignSubspace() with dims.align=12. TSNE was generated by RunTSNE() dims.use=1:12, and FindClusters() with resolution=0.6 and dims.use=1:12. Cluster reassignment was performed based on manual review of lineage-specific marker expression.

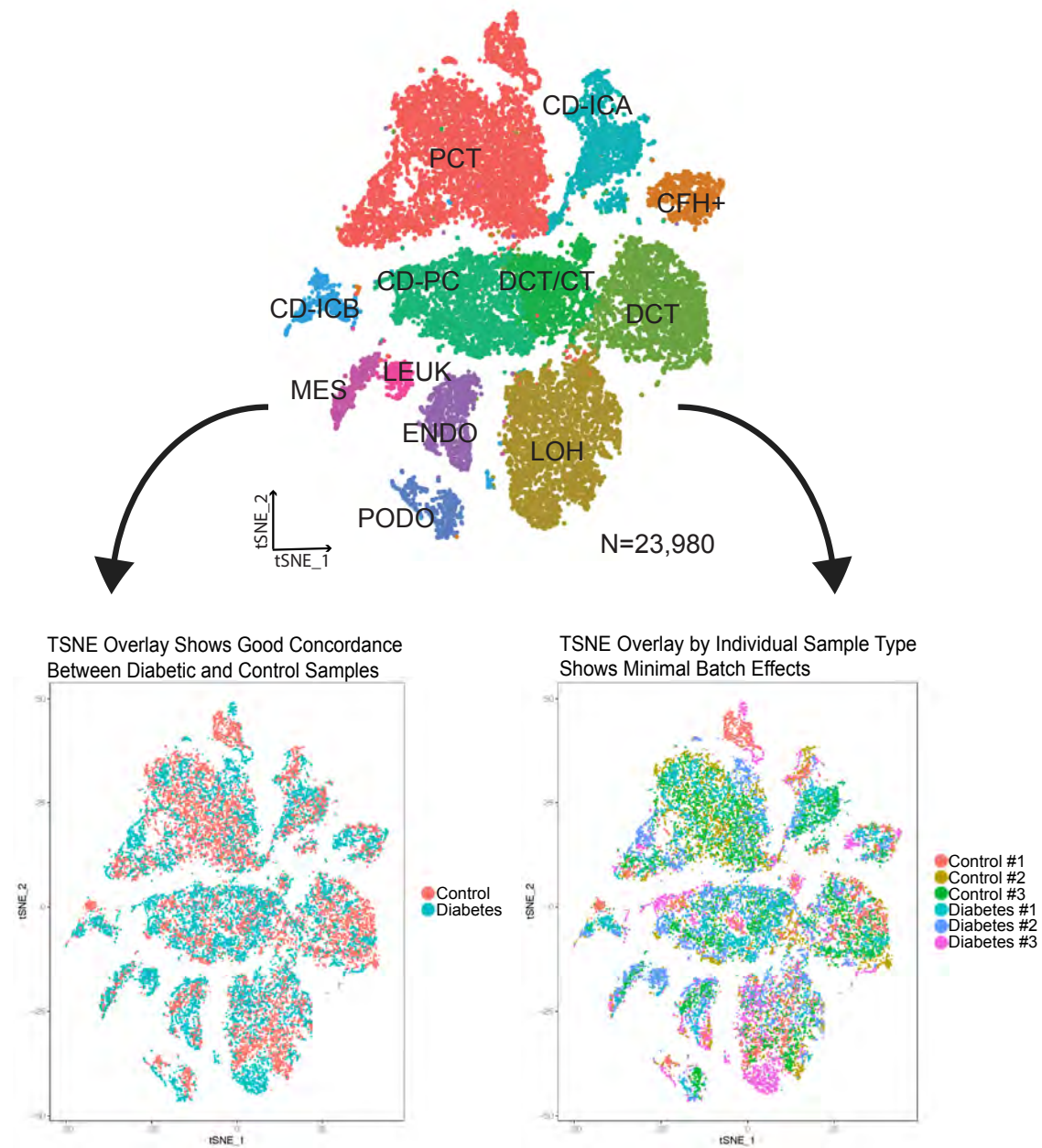
**Identification of differentially expressed genes.** Differentially expressed genes were identified by comparing the transcriptional profile of diabetic and control patients within individual cell clusters in the integrated dataset using FindMarkers(). Markers displayed in figures 2A-C were generated with logfc.threshold=0.25 and min.pct=0.25.

**Identification of overlap between aldosterone-sensitive genes curated from the literature and celltype-specific differentially expressed genes.** We used the R package GeneOverlap to compare a curated list of 908 aldosterone and salt-sensitive genes from the literature to our list of differentially expressed genes in the thick ascending limb, distal convoluted tubule, and collecting duct. We converted gene IDs to hgnc symbols using ensembl and performed the overlap analysis using a hypergeometric distribution with a false discovery threshold of 0.05 and the default background.

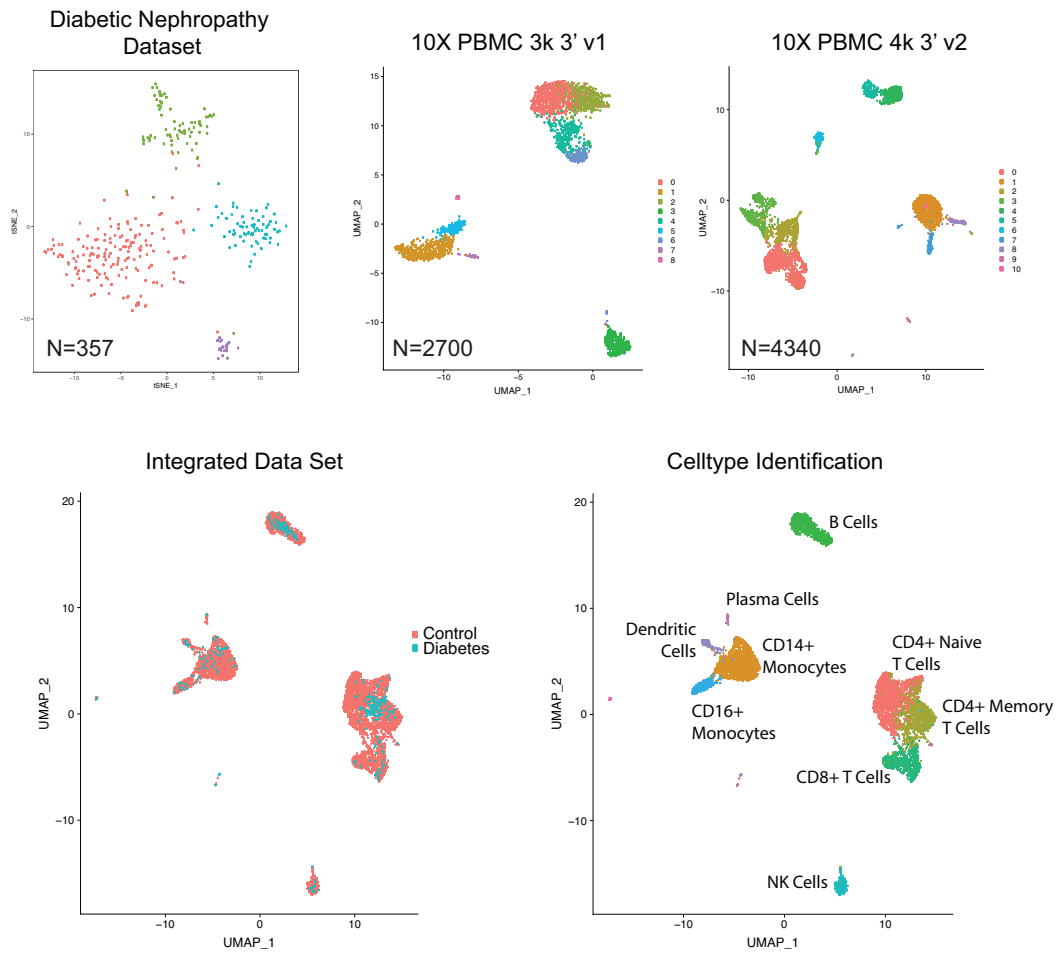


**Fig. S1.** Representative H&E images of control (A-I) and diabetic (J-R) samples. Low-power images (100x) of control samples (A-C) show no evidence of glomerulosclerosis, interstitial fibrosis, or immune cell infiltrate. Low-power images of diabetic samples (J-L) show patchy glomerulosclerosis, interstitial fibrosis, and immune cell infiltrate. Medium-power images (200x) of control (D-F) and diabetic (M-O) vessels show mild to moderate intimal sclerosis. High-power images (400x) of control glomeruli (G-I) appear normal, whereas diabetic glomeruli (P-R) show mesangial expansion and glomerular basement membrane thickening.





**Fig. S2.** TSNE plots of the integrated dataset separated by sample type (Control vs. Diabetes) and sample of origin (Control #1-3 and Diabetes #1-3).



**Fig. S3. Comparison of leukocyte subset obtained from patients with diabetic nephropathy to 2 publicly-available datasets downloaded from 10x genomics.** Leukocytes from the diabetic nephropathy dataset were extracted into a separate Seurat object using the SubsetData function and integrated with pbmc datasets (3k PBMCs from a Health Donor Cell Ranger 1.1.0 and 4k PBMCs from a Health Donor, Cell Ranger 2.1.0) using Seurat 3.0. Differential gene expression was performed within leukocyte subsets.

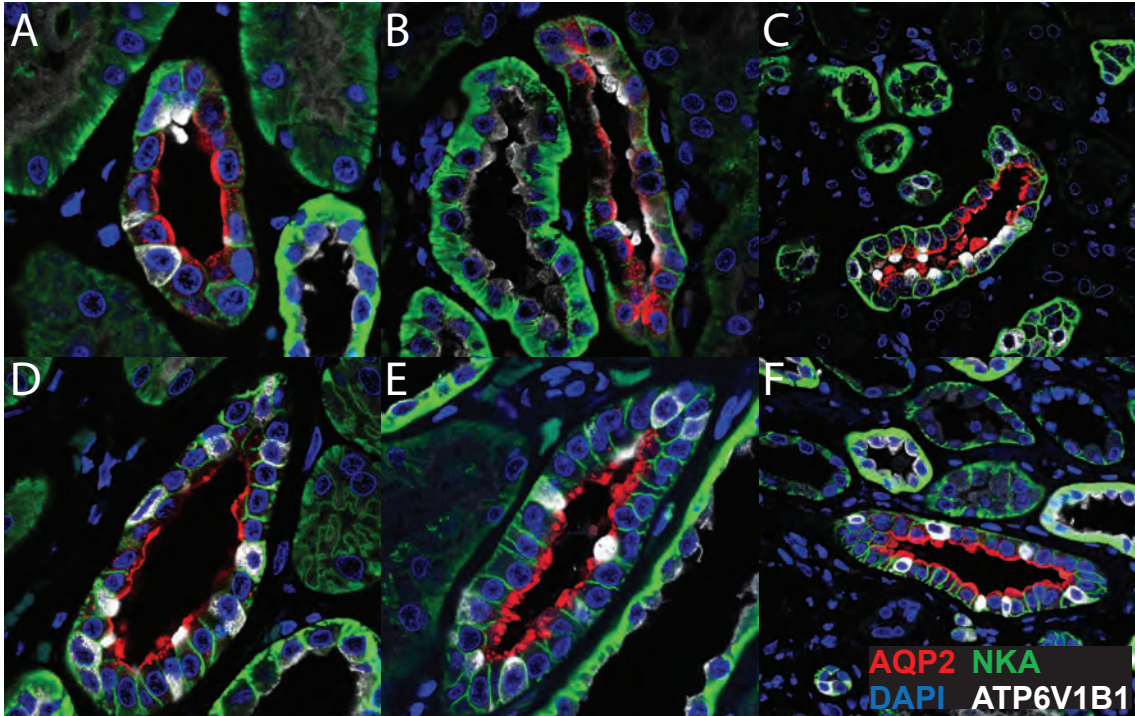


Fig. S4. **Representative images of expression of Na<sup>+</sup>/K<sup>+</sup> ATPase (NKA) in control (A-C) and diabetic (D-F) principal cells.** Formalin-fixed paraffin embedded sections were deparaffinized and stained for AQP2 (red), Na<sup>+</sup>/K<sup>+</sup> ATPase subunit ATP1A1 (green) and ATP6V1B1 (white) following antigen retrieval and imaged on a confocal microscope. Na<sup>+</sup>/K<sup>+</sup> ATPase expression was quantified in AQP2+ principal cells using ImageJ. Each image represents an individual patient.

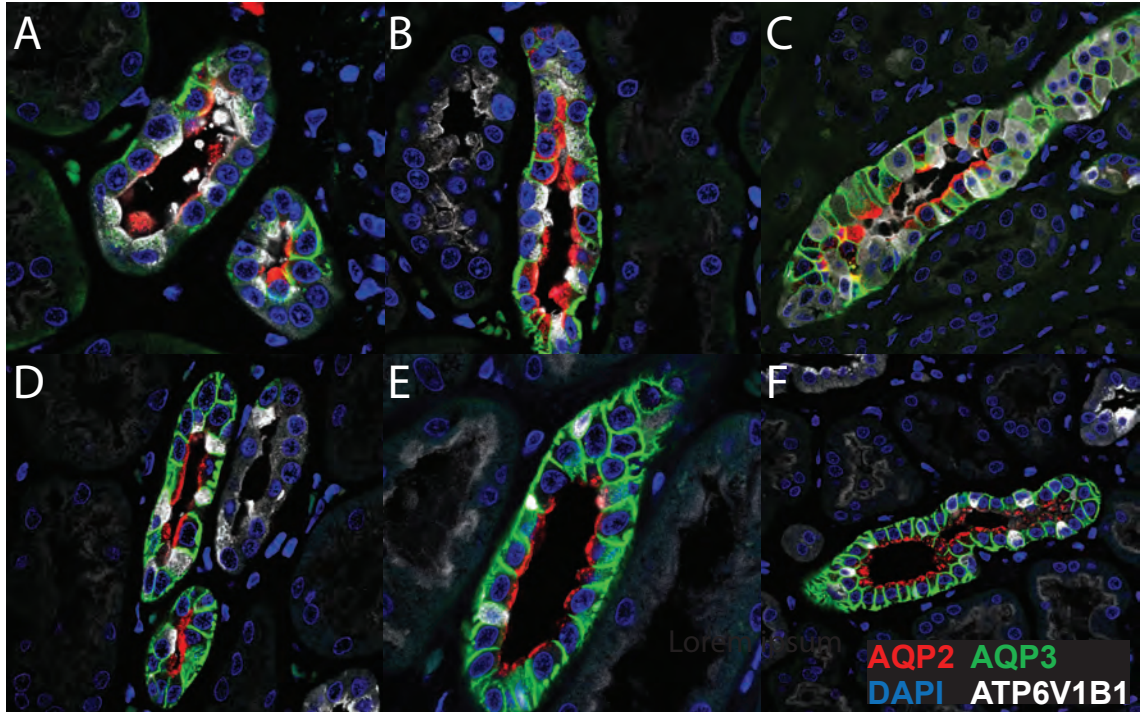


Fig. S5. **Representative images of expression of AQP3 in control (A-C) and diabetic (D-F) principal cells.** Formalin-fixed paraffin embedded sections were deparaffinized and stained for AQP2 (red), AQP3 (green) and ATP6V1B1 (white) following antigen retrieval and imaged on a confocal microscope. AQP3 expression was quantified in AQP2+ principal cells using ImageJ. Each image represents an individual patient.



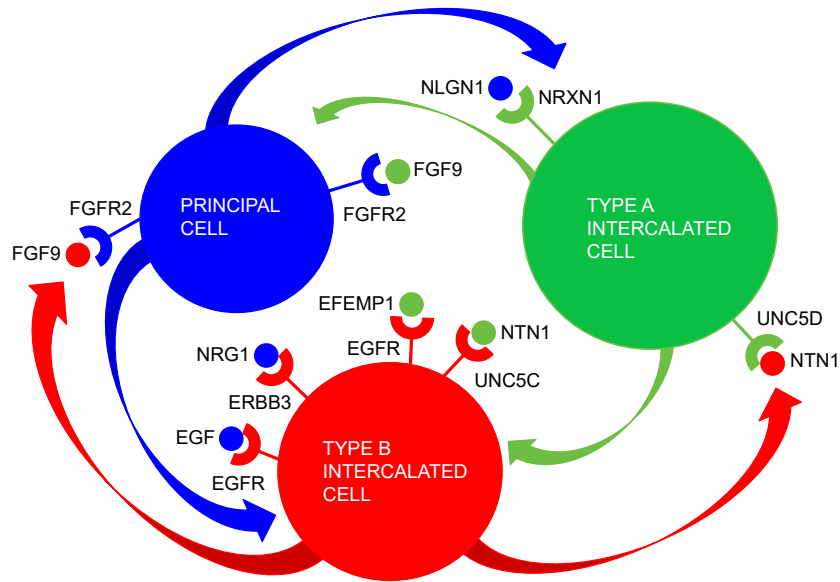


Fig. S6. **Differentially expressed ligand-receptor interactions in the collecting duct.** To study ligand-receptor interactions, we used a draft network published by Ramilowski *et al* (1). We examined collecting duct cell types and required that i) the ligand, receptor, or both were differentially expressed and ii) its cognate pair was expressed in the partner cell type. Differential expression was assessed using the FindMarkers function in Seurat with default settings.

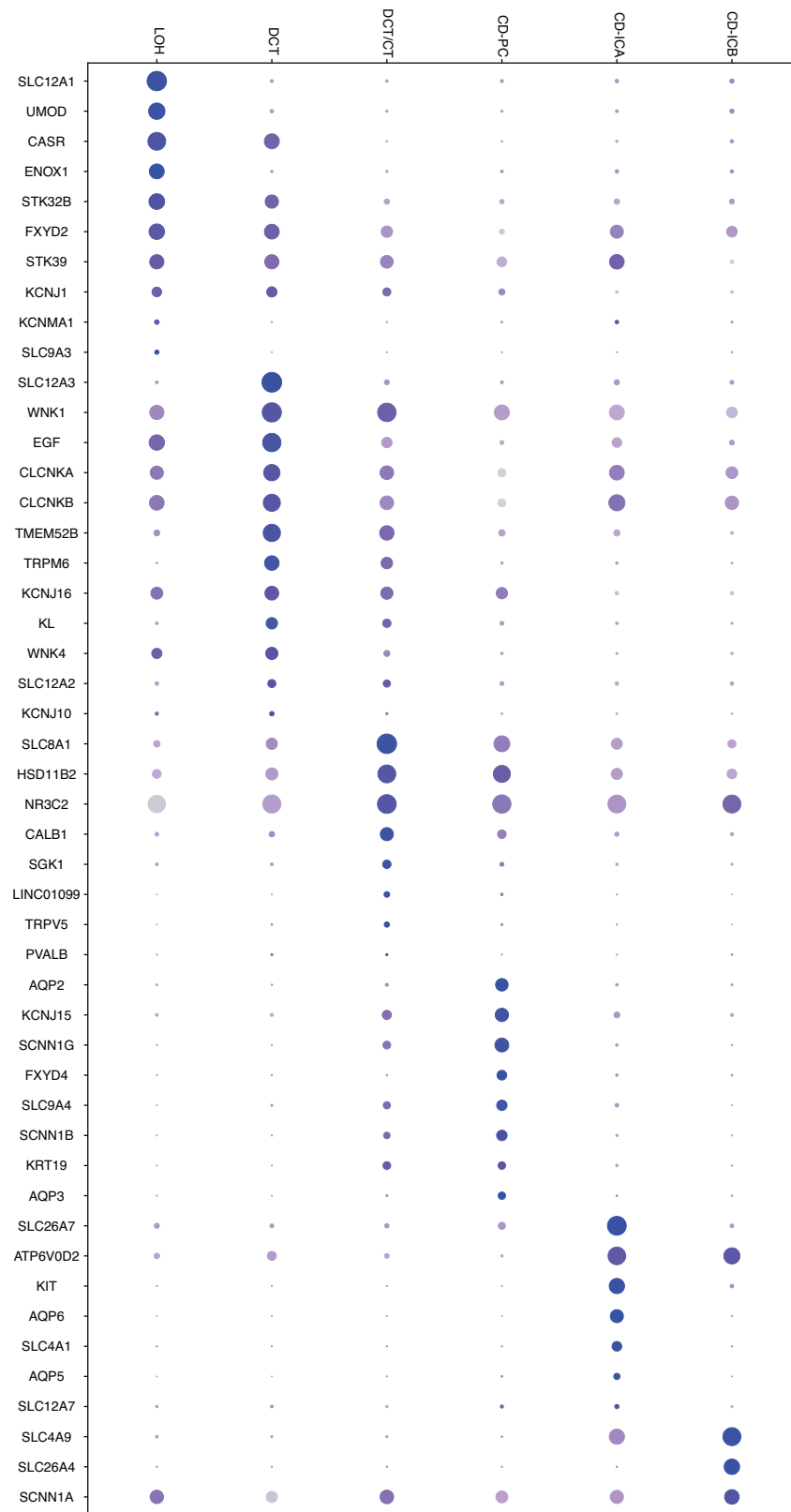


Figure S7. Extended panel of markers in the thick ascending limb, distal tubule, and collecting duct

**Table S1. Demographic, laboratory and pathology data for patient samples (see DatatableS1.xlsx for additional clinical information)**

	Control Patients (n=3)			Diabetic Patients (n=3)		
Age, Gender	54M	62M	61F	74M	52M	57F
A1c	NA	NA	NA	6.7	7.3	9.7
sCr	1.28	1.21	0.89	1.26	1.03	0.7
Proteinuria	NA	NA	NA	Yes	2+	No
Glomerulosclerosis	None (<10%)	None (<10%)	None (<10%)	Moderate (26-50%%)	Mild (11-25%)	None (<10%)
IFTA	1-10%	1-10%	1-10%	11-25%	11-25%	1-10%
Arteriosclerosis	Moderate	Moderate	Mild	Mild	Moderate	Mild

**Table S2. Number of nuclei and number of genes and UMI per nucleus for each sample**

	Control 1	Control 2	Control 3	Diabetes 1	Diabetes 2	Diabetes 3	Total
Nuclei Accepted	3740	3917	6203	3771	3806	2543	23,980
Mean No. Genes per Nucleus	4046	1354	2114	3142	1902	2689	2541
Mean No. UMI per Nucleus	12736	2734	4717	9806	4813	6561	6894



**Additional data table S1 (DatatableS1.xlsx)**

Patient demographics with detailed clinical information and laboratory data.

**Additional data table S2 (DatatableS2.xlsx)**

Differentially expressed genes by celltype comparing the diabetic patient with proteinuria (#3) to the diabetic patients without proteinuria (#1 and #2).

**Additional data table S3 (DatatableS3.xlsx)**

Gene set enrichment analysis for individual cell types performed with the R package fgsea comparing the diabetic patient with proteinuria (#3) to the diabetic patients without proteinuria (#1 and #2) using GO Biologic Processes from msigdb

**Additional data table S4 (DatatableS4.xlsx)**

Gene set enrichment analysis for individual cell types performed with the R package fgsea comparing the diabetic patient with proteinuria (#3) to the diabetic patients without proteinuria (#1 and #2) using GO Molecular Function from msigdb

**Additional data table S5 (DatatableS5.xlsx)**

Differentially expressed genes by celltype comparing the diabetic patients to the control patients.

**Additional data table S6 (DatatableS6.xlsx)**

Gene set enrichment analysis for individual cell types performed with the R package fgsea comparing the diabetic patients with control patients using GO Biologic Processes from msigdb

**Additional data table S7 (DatatableS7.xlsx)**

Gene set enrichment analysis for individual cell types performed with the R package fgsea comparing the diabetic patients with control patients using GO Molecular Function from msigdb

**Additional data table S8 (DatatableS8.xlsx)**

Gene ontology enrichment analysis for individual cell types performed with the R package topGO using differentially expressed genes (DatatableS5.xlsx) and GO Biologic Processes.

**Additional data table S9 (DatatableS9.xlsx)**

Number of cells per sample for each cell type with corresponding p-values testing for a difference in the mean between diabetic and control groups.

**Additional data table S10 (DatatableS10.xlsx)**

Overlap of a curated list of 908 aldosterone-sensitive genes with the differentially expressed genes in individual celltypes using the R package GeneOverlap and the hypergeometric distribution.

**References**

1. Ramilowski JA, *et al.* (2015) A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat Commun* 6:7866.

