Supplemental Material for

Urinary single-cell sequencing captures intrarenal injury and repair processes in human acute kidney injury

Jan Klocke[1,2], Seung Joon Kim[3], Christopher M. Skopnik[1,2], Christian Hinze[1,4,5], Anastasiya Boltengagen[3], Diana Metzke[1,2], Emil Grothgar[1,2], Luka Prskalo[1,2], Leonie Wagner[1,2], Paul Freund[1,2], Nina Görlich[1,2], Frédéric Muench[1], Kai M. Schmidt-Ott[1,4,5], Mir-Farzin Mashreghi[6], Christine Kocks[3], Kai-Uwe Eckardt[1]*, Nikolaus Rajewsky[3]*, Philipp Enghard[1,2]*

[1] Department of Nephrology and Medical Intensive Care, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany.

[2] Deutsches Rheumaforschungszentrum, an Institute of the Leibniz Foundation, Berlin, Germany.

[3] Systems Biology of Gene-Regulatory Elements, Berlin Institute for Medical Systems Biology (BIMSB), Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

[4] Molecular and Translational Kidney Research, Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

[5] Department of Nephrology and Hypertension, Hannover Medical School, Hannover, Germany

[6] Therapeutic Gene Regulation, Deutsches Rheuma-Forschungszentrum, an Institute of the Leibniz Foundation, Berlin, Germany

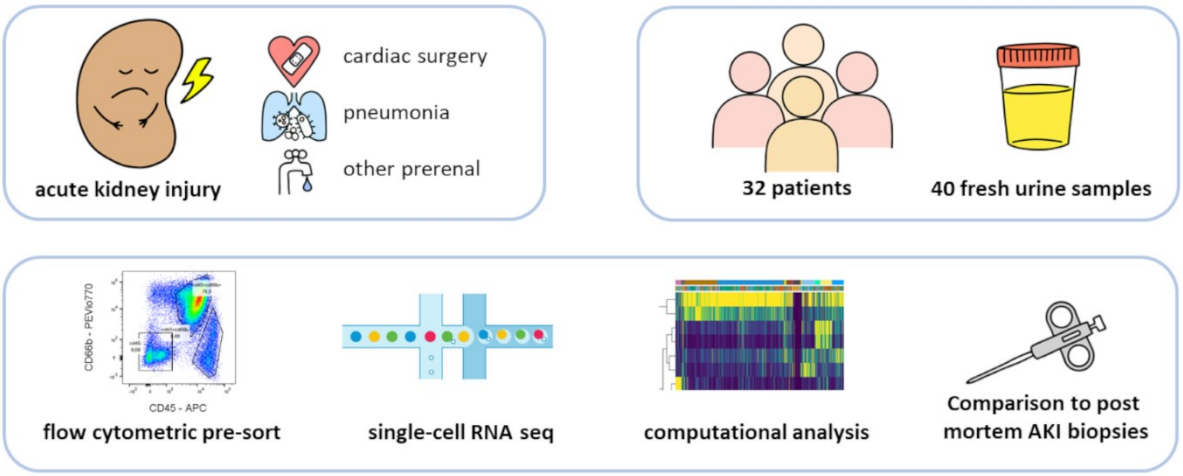
* These authors jointly supervised the work and contributed equally

Corresponding author

Jan Klocke, Department of Nephrology and Medical Intensive Care, Charité-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany, 0049/30450614624, jan.klocke@charite.de

Supplemental Figures

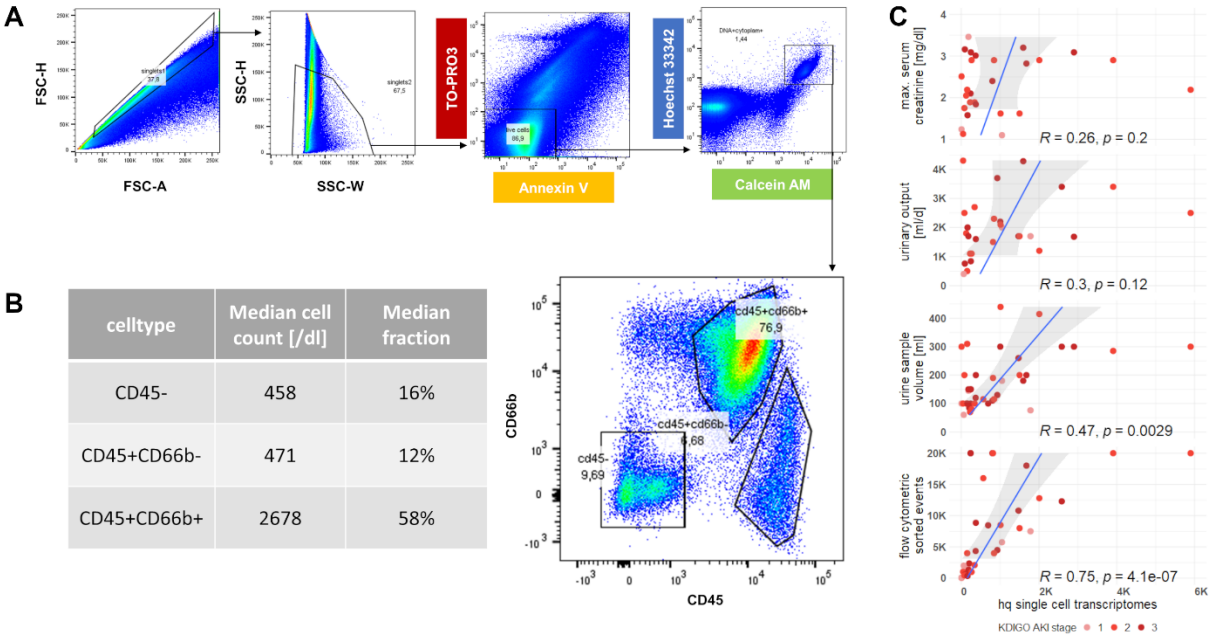
A



B

AKI etiology	cardiac surgery	pneumonia	prerenal	total
patients (samples)	7 (13)	15 (17)	10 (10)	32 (40)
mean age (range)	69 (51-83)	60 (33-78)	68 (24-84)	64 (24-84)
sex (f/m)	2/5	2/13	3/7	7/25
KDIGO AKI stage (1/2/3)	0/4/3	2/5/8	3/2/5	5/11/16
Total post-QC sc transcriptomes	20249	15582	6777	42608

Figure S1. Study design and patient characteristics.



B

celltype	Median cell count [/dl]	Median fraction
CD45-	458	16%
CD45+CD66b-	471	12%
CD45+CD66b+	2678	58%

Figure S2. Flow cytometry can provide viable pre-sorted urine cell fractions
A. *Gating strategy for sorting single viable cells from urine. After singlet gating in forward and sideward-scatter, dead (TO-PRO3+) and apoptotic (Annexin V) cells are excluded, before viable (Calcein AM+) events with DNA (Hoechst 33342+) are singled out. For single-cell suspension, CD45-CD66b- and CD45+CD66b- cells were sorted.* ***B.*** *Table of urinary cell type fractions. Granulocytes (CD45+CD66b+) provide the largest fraction of viable urine cells.* ***C.*** *Correlation of captured high-quality single cell transcriptomes with patient metrics. The absolute number of captured cells is dependent on urine sample size.*

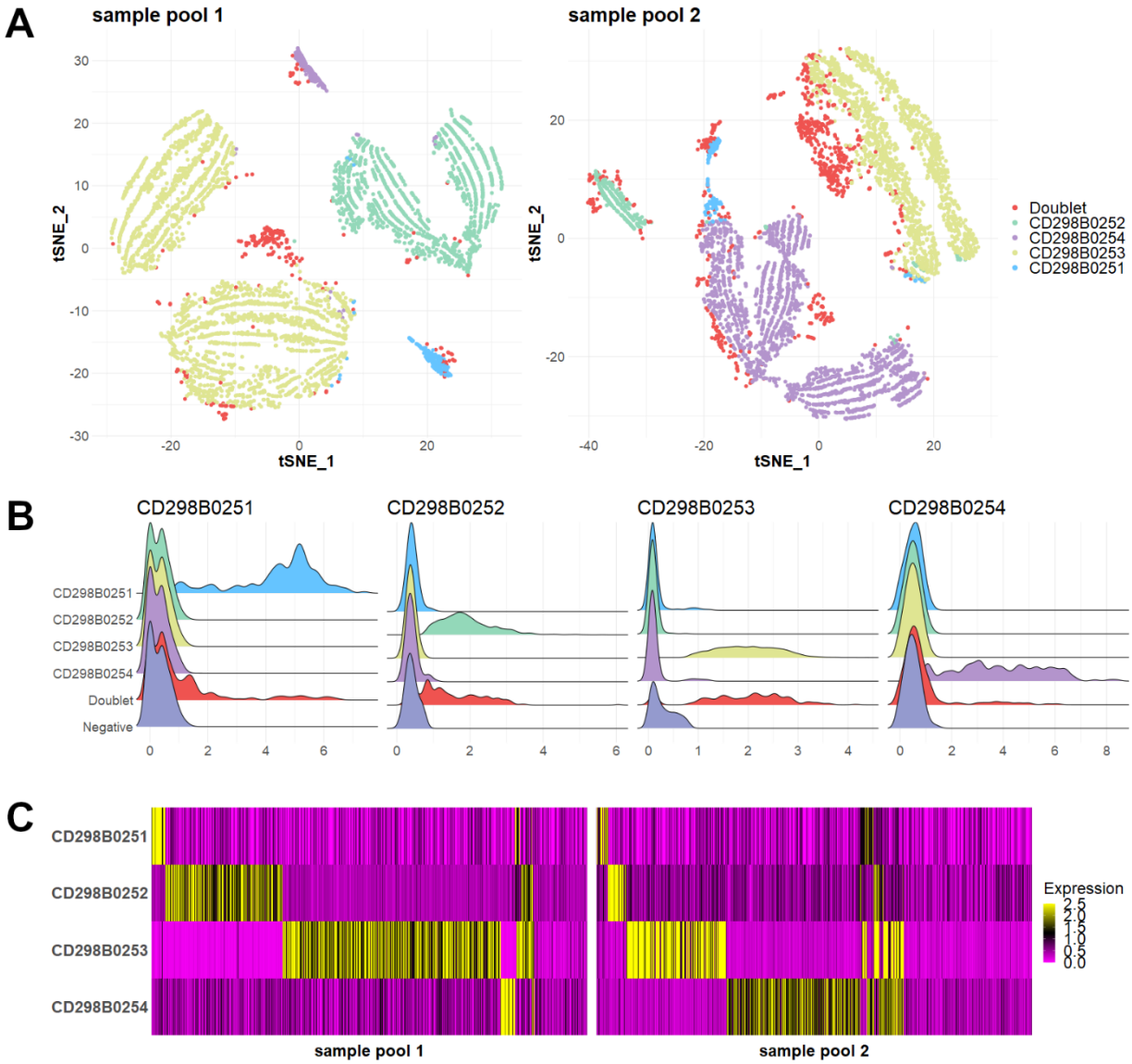


Figure S3. Urine single cell libraries can be demultiplexed after hashing and pooling of samples. ***A.*** *t-Distributed Stochastic Neighbor Embedding (tSNE) dimensional reduction of pooled patient samples that were labeled by barcode cell surface antibodies (CD298B0251-4).* ***B.*** *Ridgeplots of barcode antibody detection per sample. Inter-sample doublets (red) can be easily identified.* ***C.*** *Heatmap of barcode antibody detection per cell. Doublets were excluded, cells without barcode detection were further analyzed as pool samples.*

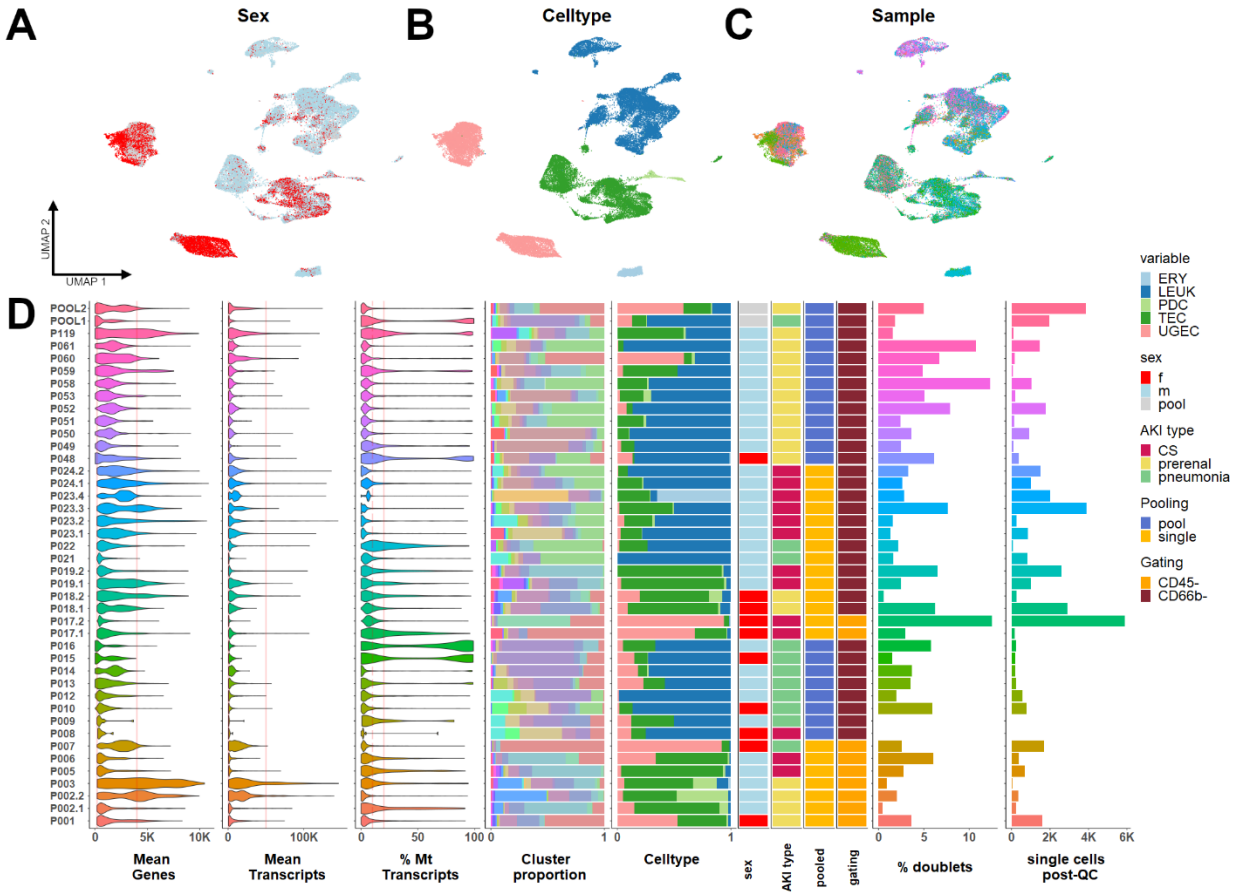


Figure S4. Amount of captured urinary cells is diverse.
Uniform manifold approximation and projection (UMAP) of 42608 *scRNA-seq urine cells from 32 individuals with AKI**. A-C.** *Distribution of cells in UMAP by sex (A), celltype (B) and etiology of AKI (C). Females (red) excrete more UGEC via urine (also Fig. S5).* ***D.*** *Sample metrics by sample. Patients with multiple samples are numbered (1.1, 1.2 …). Amount of captured urinary cells is diverse (panel on the right), but major cell types (kidney cells (PDC – podocytes, TEC – tubular epithelial cells), urogenital epithelial cells (UGEC) and leukocytes (LEUK)) are regularly featured. (Color code – see Legend on the right)*

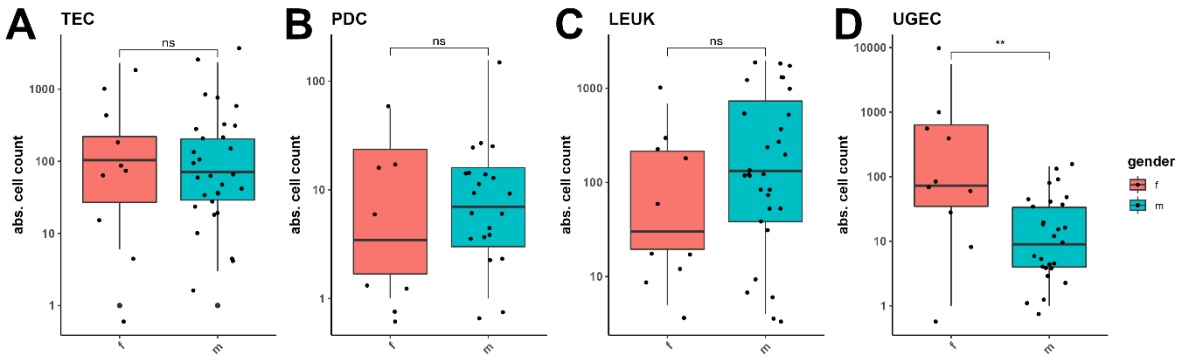


Figure S5. Urogenital cell abundance in urine is higher in females
*Boxplots of absolute counts and relative proportions of urinary cells per sample in female (red) and male (green) patients. PDC – podocytes, TEC – tubular epithelial cells, UGEC – urogenital epithelial cells, LEUK – leukocytes. Boxplot: line = median, hinges = first -third quartile, whiskers = 1.5 interquartile range, large dots = outliers, *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001, unpaired two-samples Wilcoxon test. ns, not significant.*

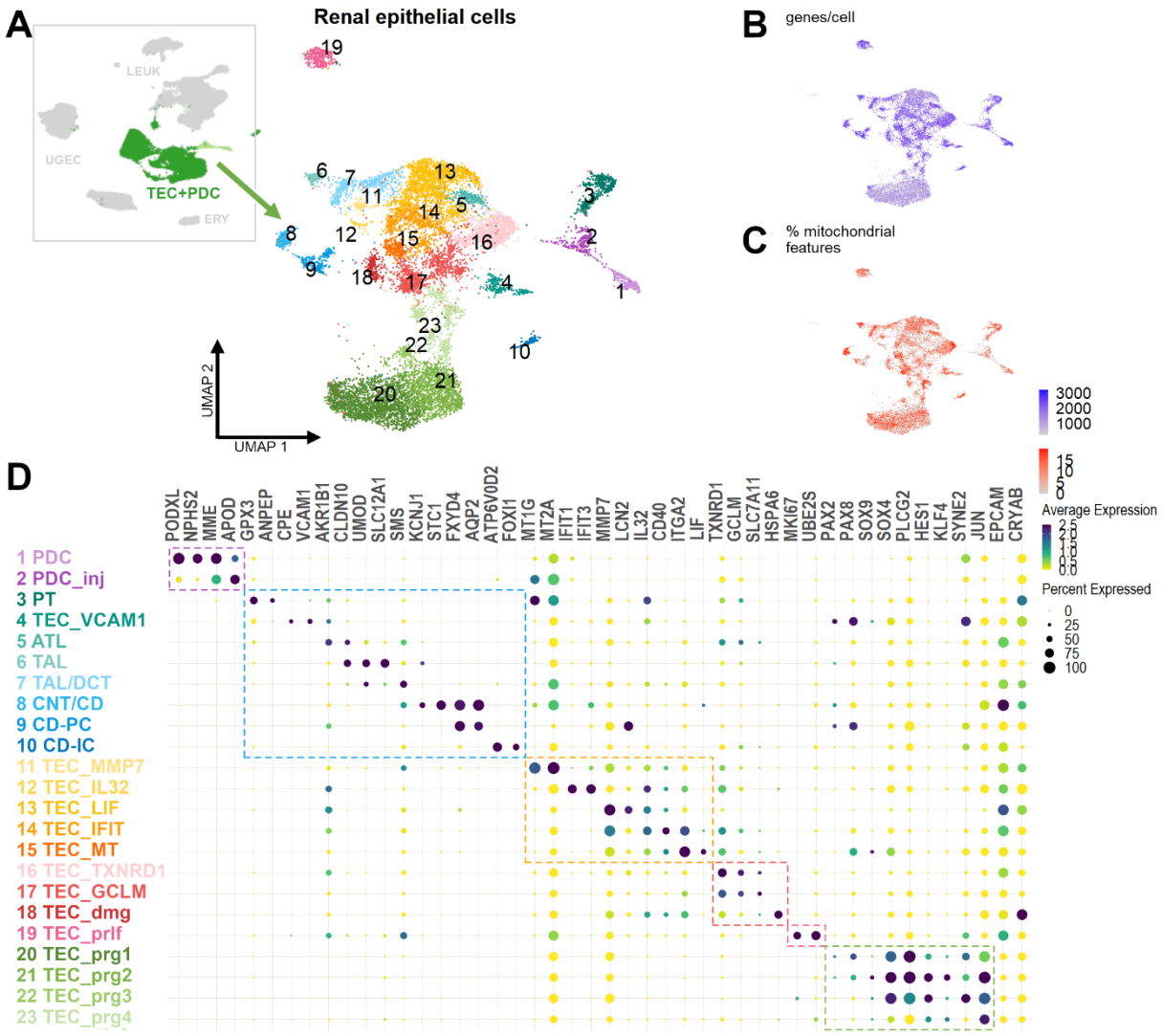


Figure S6. Diverse tubular cell reactions to AKI

A. *UMAP of 12853 urinary renal parenchymal scRNAseq transcriptomes in 23 distinct clusters.* ***B+C.*** *Quality metrics genes/cell (B) and percentage of mitochondrial genes per cell (C) plotted in UMAP.* ***D.*** *Dotplot of marker gene expression for each cell type. Next to podocytes (PDC, cluster 1-2) and segment specific tubular epithelial cells (TEC, clusters 3-10), several injury related subtypes of TEC (clusters 11-23) can be distinguished. PT – proximal tubule, ATL – ascending thin limb, TAL – thick ascending limb, DT – distal tubule, CNT – connecting tubule, CD-PC – collecting duct principal cells, CD-IC – collecting duct intercalated cells, inj – injured, dmg – damaged, prlf – proliferating, prg – progenitor-like.*

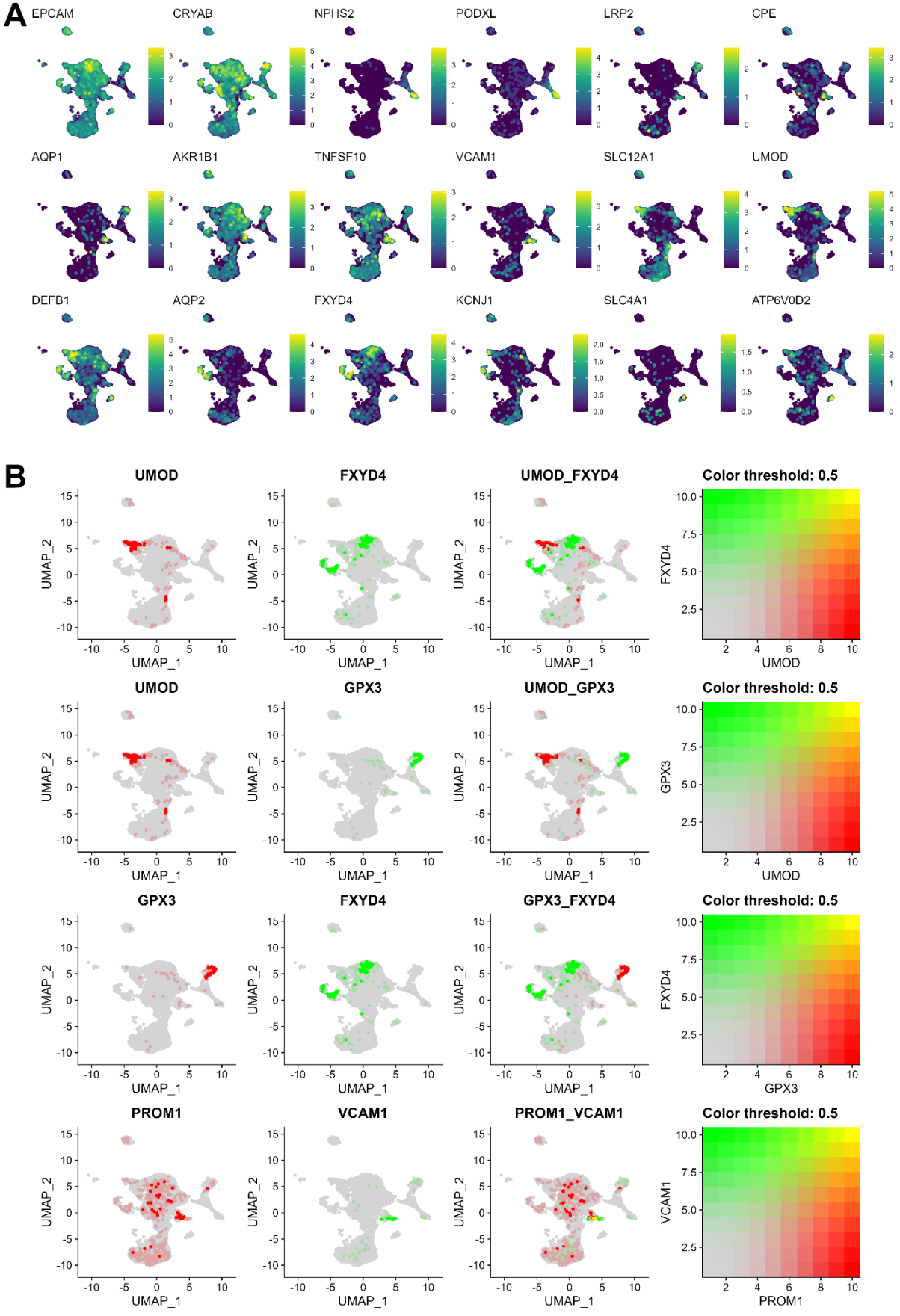


Figure S7. Injured cell states are of mixed tubular identity

A. *Tubular cell marker expression in urinary renal cells plotted in UMAP. Injured clusters (top to bottom middle) dimly express several segment specific marker genes.* ***B.*** *Co-Expression of select markers*

shows no overlap between segment specific markers. One cluster has few cells with co-expression of PROM1 and VCAM1 (lower right panel).

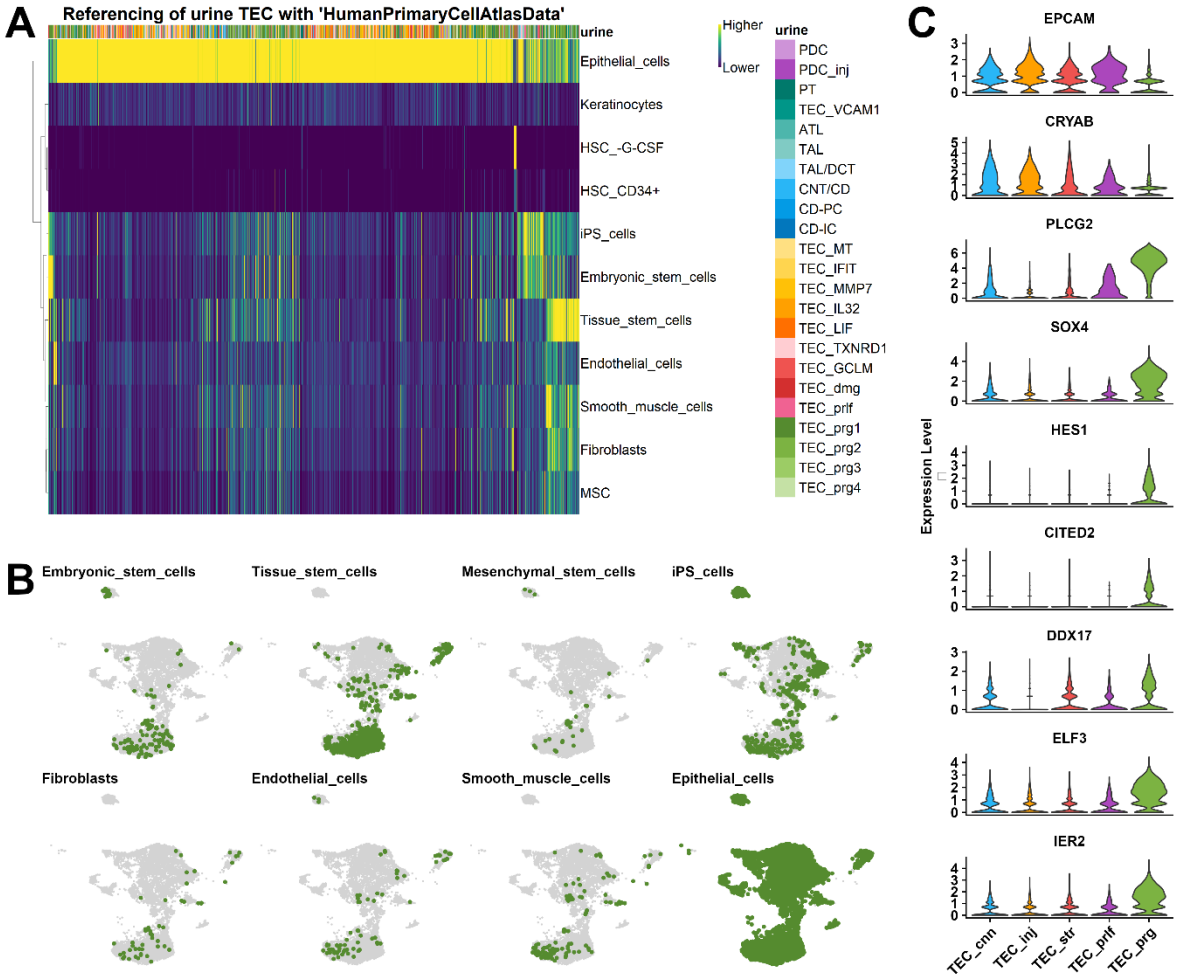


Figure S8. Urinary tubular epithelial cells partially resemble stem cells
A. *Heatmap of automatic annotation scores (SingleR) of urinary tubular epithelial cells (TEC) with a human primary cell atlas* [S1]*. Most cells are annotated as epithelial, while some are annotated as stem- or mesenchymal cells, pointing towards dedifferentiation.* ***B.*** *UMAP Distribution of automatic cell labels of A in urine TEC. Most stem labels occur in progenitor-like (TEC_prg) subsets (bottom middle clusters).* ***C.*** *Violin plot of marker gene expression across TEC subgroups. Tubular markers (EPCAM, CRYAB) are downregulated in TEC_prg, while stemness markers (PLCG2, HES1) are abundantly expressed.*

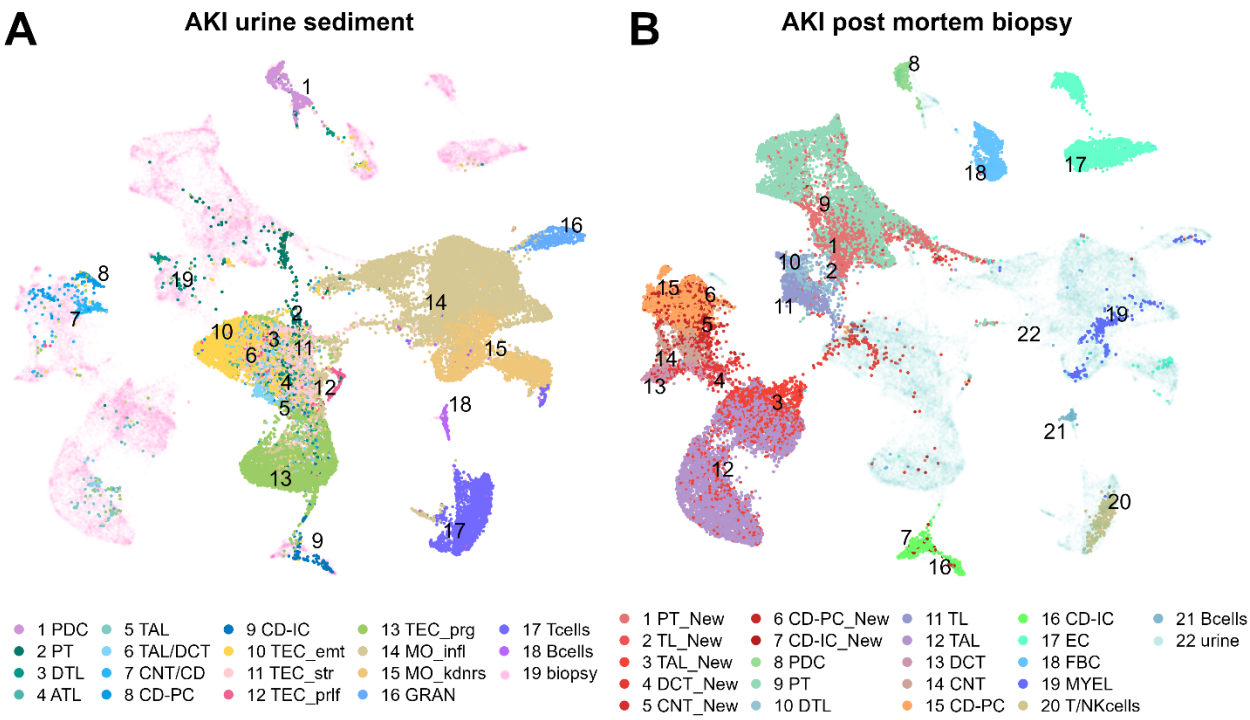


Figure S9. Urinary cells from AKI patients mirror AKI post-mortem biopsies.
A+B. *Integrated AKI urinary scRNAseq (**A**) and AKI post-mortem biopsy single nuclei RNAseq*[S2] *(**B**) datasets. Note how kidney tissue injury related cell types ("_New", red shadings in B) partially overlap with injury related urine TEC (transparent turquois clusters in middle) and how few segment specific urine TEC (green and blue shadings in A) cluster with respective tissue cell types.*

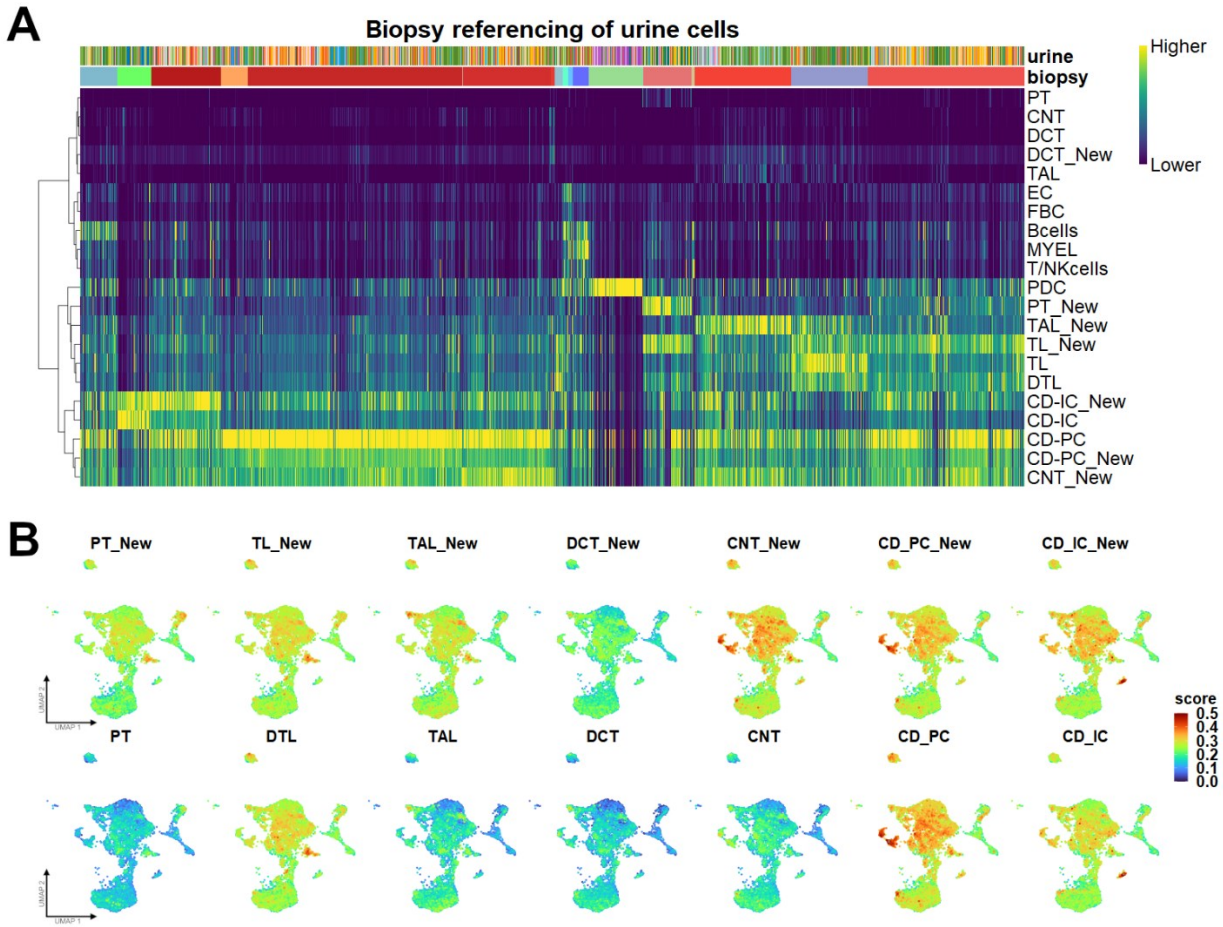


Figure S10. Urinary tubular cells resemble injured and distal tubules

A. *Heatmap of automatic annotation scores (SingleR) of urinary tubular epithelial cells (TEC) with human post-mortal biopsy AKI tissue* [S2]*. Most cells are most similar to injury reactive cell states from AKI kidney tissue ("_New") and distal tubular segments like collecting duct.* ***B.*** *UMAP Distribution automatic annotation scores of A in urine TEC. PDC – podocytes, PCT – proximal convoluted tubule, DTL – descending thin limb, ATL – ascending thin limb, TAL – thick ascending limb, DCT – distal convoluted tubule, CNT – connecting tubule, CD-PC – collecting duct principal cells, CD-IC – collecting duct intercalated cells.*

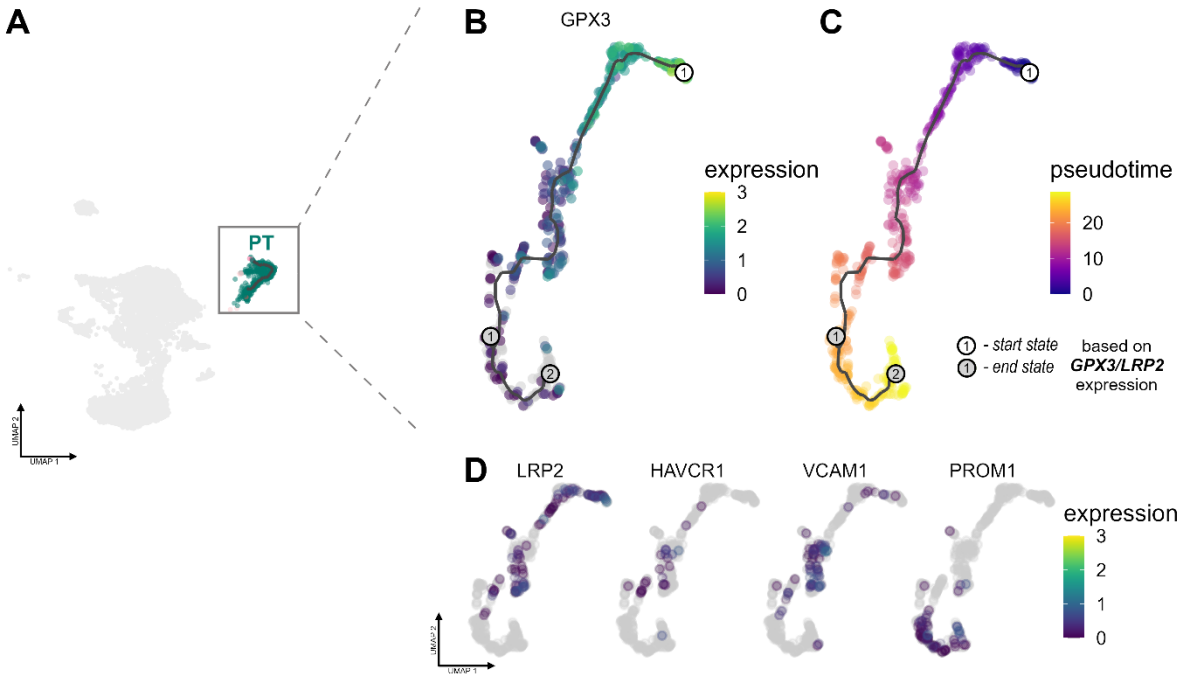


Figure S11. Urinary proximal tubular epithelial cell dedifferentiation trajectory
A. *UMAP overview on all urinary TEC and subsetted proximal tubule (PT).* ***B.*** *Pseudotime trajectory analysis of urinary PT cells after AKI. A region with high PT marker expression (LRP2/GPX3) and low injury marker expression (HAVCR1) was set as a starting state. Trajectory graph shows two endpoints in potentially adaptive cell states.* ***C.*** *Expression of relevant marker genes across the trajectory (LRP2 = megalin, PT marker; HAVCR1 = hepatitis A virus cellular receptor 1, injury marker; VCAM1 = Vascular cell adhesion protein 1, marker for "failed-repair" state in PT; PROM1 = prominin1 (CD133), stem cell marker).*

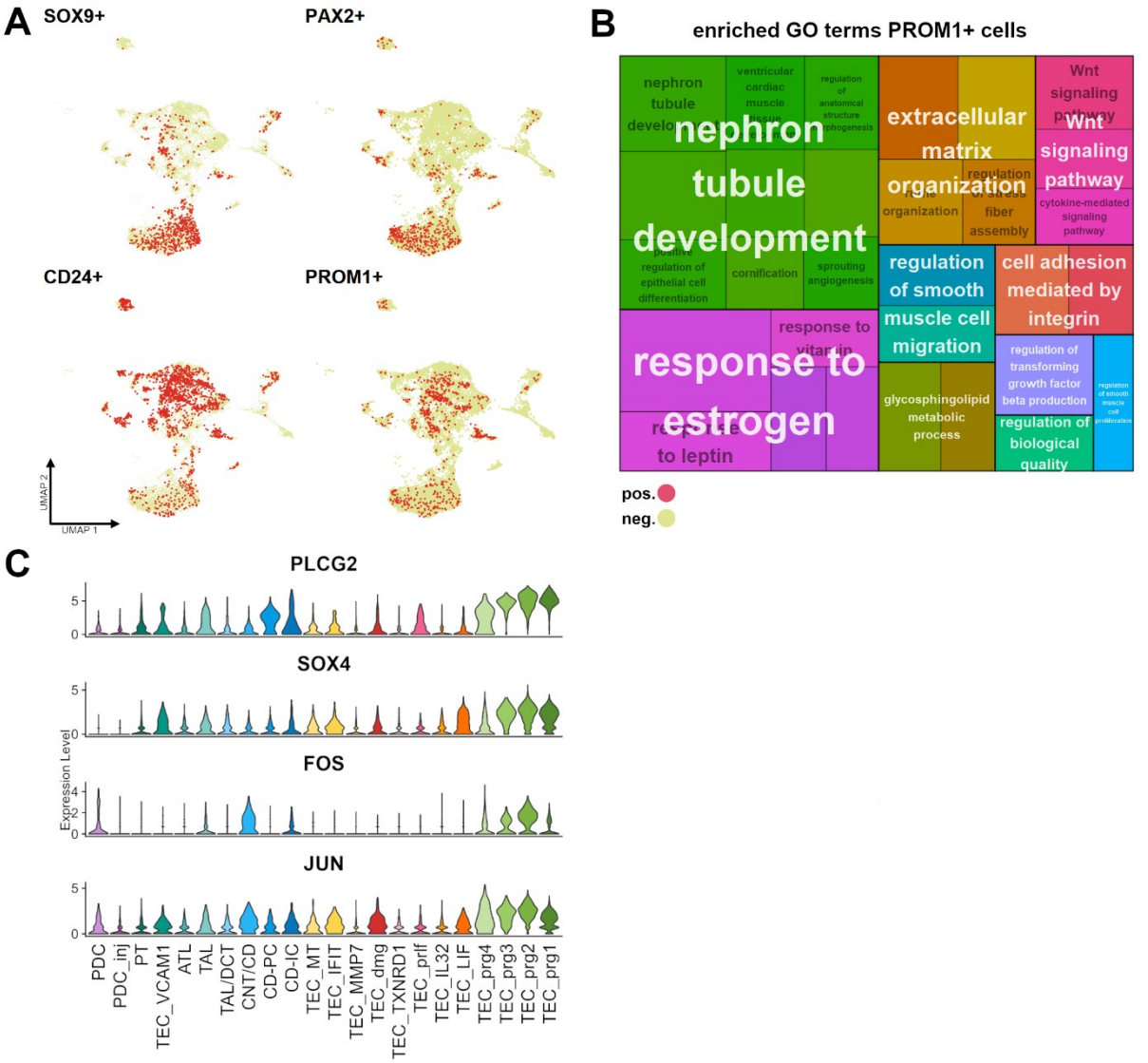


Figure S12. Urinary tubular epithelial cells show regenerative phenotypes.
A*. Distribution of stem cell marker positive cells across UMAP representation.* ***B****. Treemap plot of enriched gene ontology (GO) terms in PROM1+ cells. Each rectangle is a single GO term (black text), sized based on -log10(adj. pvalue). The terms are joined into GO clusters by similarity, with the largest rectangle of the cluster providing the group name (white text). visualized with different colors.* ***C.*** *Violin plot of cell differentiation marker expression across renal tubular cell (TEC) clusters. proximal tubule (PT), descending thin limb (DTL), ascending thin limb (ATL), thick ascending limb (TAL), distal convoluted tubule (DCT), connecting tubule (CNT), collecting duct principal cells (CD-PC), collecting duct intercalated cells (CD-IC), as well as podocytes (PDC), prg – progenitor-like, prlf – proliferating, dmg – damage.*

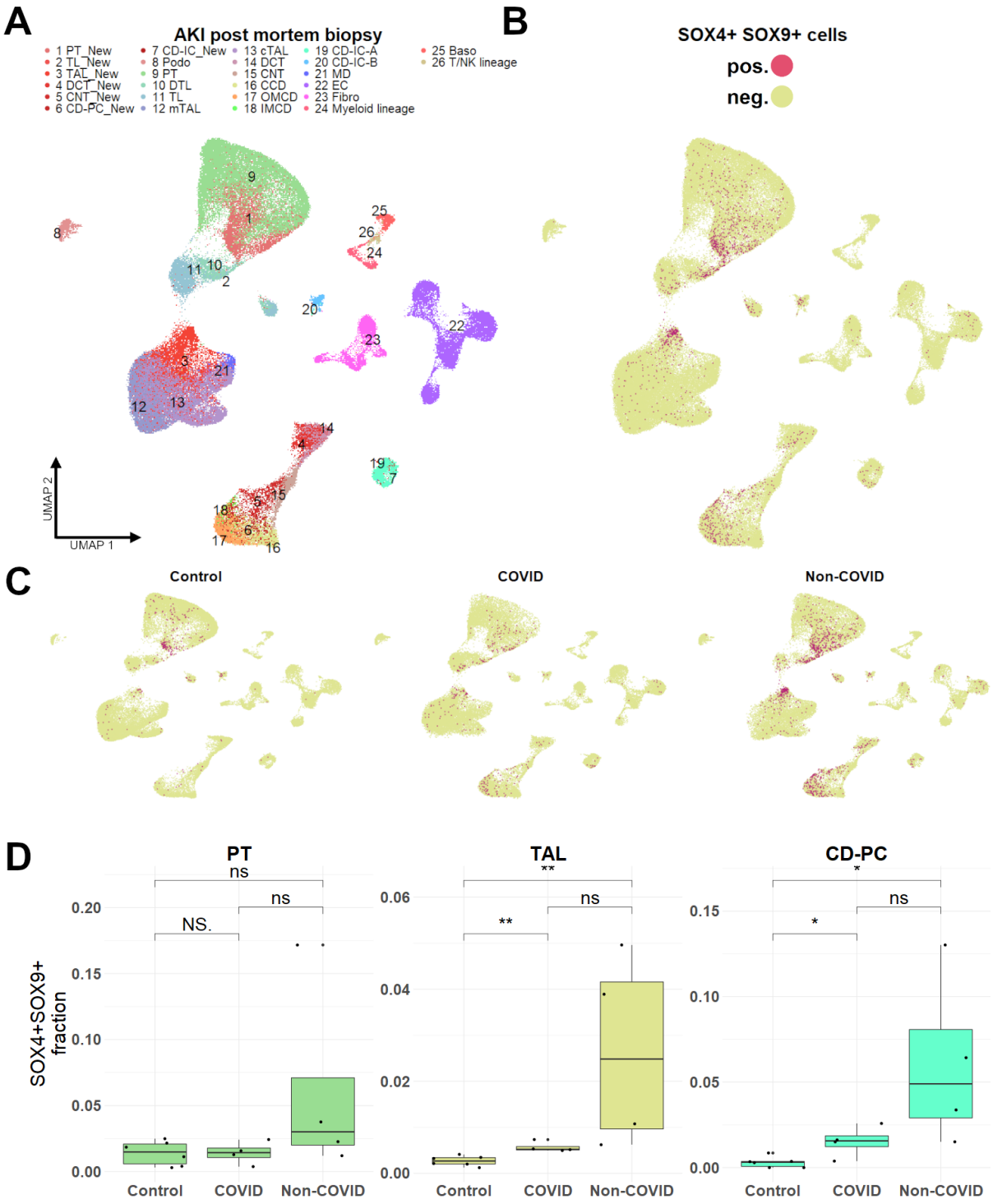


Figure S13. SOX4+SOX9+ progenitor-like cells occur in kidney AKI tissue.
A. *UMAP dimensional reduction of AKI post-mortem biopsy snRNAseq dataset. "_New" clusters in red shadings indicate injury reactive tubular cell states occurring in AKI. PDC – podocytes, PCT – proximal convoluted tubule, DTL – descending thin limb, ATL – ascending thin limb, TAL – thick ascending limb, DCT – distal convoluted tubule, CNT – connecting tubule, CD-PC – collecting duct principal cells, CD-IC – collecting duct intercalated cells, MYEL – myeloid cells, EC – endothelial cells, FBR – fibrocytes, MO – monocytes/macrophages, GRAN – granulocytes, emt – epithelial-mesenchymal transition, str – stressed, prlf – proliferating, prg – progenitor-like* ***B+C.*** *Localization of SOX4+SOX9+ cells in dimensional reduction of all samples (**B**) and by disease group (**C**).* ***D.*** *Fraction of SOX4+SOX9+ TEC in proximal tubule (PT), thick ascending limb (TAL) and collecting duct principal cells (CD-PC) divided by disease group. In both COVID-AKI and non-COVID-AKI the relative SOX4+SOX9+ cell amount increases compared to healthy tissue. Boxplot: line = median, hinges = first -third quartile, whiskers = 1.5 interquartile range, large dots = outliers, *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001, unpaired two-samples Wilcoxon test. ns, not significant.*

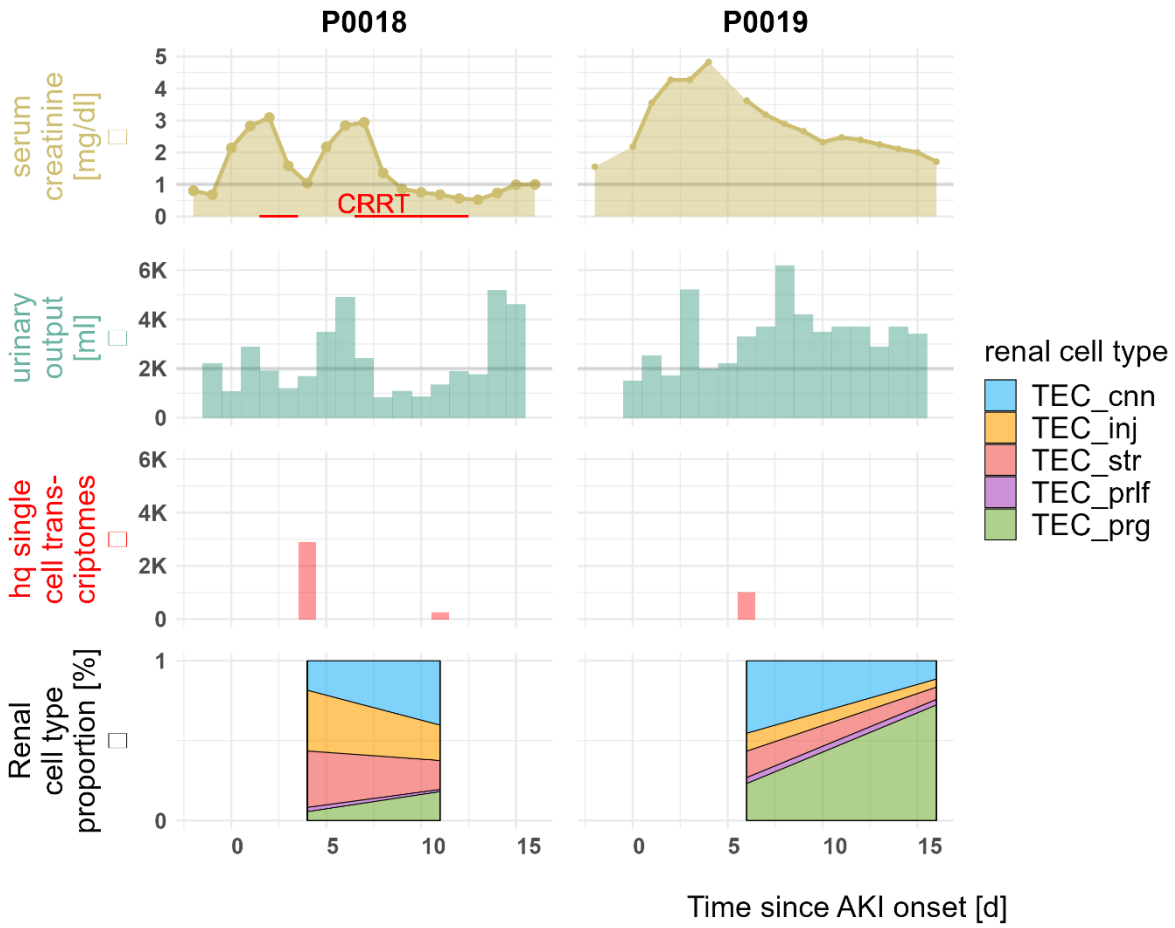


Figure S14. Urinary AKI tubular epithelial cell abundance and cell type proportions at different timepoints.
Individual AKI course of two additional repeatedly sampled patients with serum creatinine, urinary output, single cell transcriptome yield and relative amount of urinary renal parenchymal cell subsets (compare figure 6).

Supplemental Methods

Patients

Between 2019 and 2021 we collected 40 urine samples of 32 patients with AKI as defined by Kidney Disease: Improving Global Outcomes (KDIGO) criteria (Fig. S1, Table S1). Patients were sampled at a variable timepoint within the first 21 days after AKI onset. Five patients were sampled at two separate timepoints and a single patient on four occasions.

Seven patients underwent cardiac surgery within max. 48 h prior to AKI onset, 15 patients were admitted to intensive care units (ICU) because of pneumonia (all fulfilling sepsis criteria) and developed AKI during the first five days of their ICU stay; the majority of these (14/15) suffered from COVID-19. An additional 10 patients had other, mostly prerenal causes of AKI, including gastrointestinal bleeding (1), diarrhea (2), exsiccosis (4) or decompensated heart failure (3). Children and patients with kidney transplants, active oncological disease, urinary tract infections or postrenal causes of AKI were excluded from the study. Additionally, all patients were screened via urinary dipstick prior to sample preparation, excluding samples with > 70 leukocytes/µl (one of three "+" in urinary dipstick) and patients with less than 50 ml/ 4 h urinary output.

Sample preparation

Samples (median 150 ml, range 60-440 ml) were collected either as first morning void urine or via urinary catheter (using the pooled urine output of 4 hours) and stored on ice and transported to the lab. After exclusion of leukocyturia via urinary dipstick, samples were centrifuged at 600g / 4 °C, and resuspended with Annexin binding buffer[S3] / 1 % bovine serum albumin (ABB/BSA). Incubation with Actinomycin D for 30 min on ice was applied to prevent subsequent alteration of transcriptomes[4] followed by blocking Fc receptors using human FcR blocking reagent (Miltenyi, Bergisch Gladbach, Germany). Cells were then incubated with fluorescent dyes and fluorochrome-conjugated monoclonal antibodies for fluorescence-activated cell sorting (FACS). The following antibodies and dyes were used for cell labeling: Calcein AM (BD Biosciences, San Jose, CA, USA, 564061), Hoechst 33342 (BD Biosciences, San Jose, CA,

USA, 561908), AnnexinV-PE (Biolegend, San Diego, CA, USA, 640908), CD45-APC/Vio770 and CD66b-PE/Vio770 (both Miltenyi, 130-110-635 and 130-119-768). Cells were washed, filtered through a 70 µm cell strainer and labeled with To-Pro3 iodide (T3605, Life technologies, Eugene, Oregon) for dead cell discrimination immediately prior to sorting.

Samples were sorted using a Sony MA900 (Sony, Tokyo, Japan) with a 100µm nozzle into phosphate buffered saline (PBS) / 1 % BSA gating To-Pro3$^-$AnnexinV$^-$ Hoechst33342$^+$CalceinAM$^+$ singlets for obtaining single viable urinary cells (Fig. S2). Urinary granulocytes, which make up a large fraction of the urinary cellular signal (median 58% in FACS gating, Fig. S2) but for which a renal tissue origin is not verifiable, and which may interfere with downstream analysis, were excluded. In the initial cohort of patients (n=7) only CD45$^-$ cells were obtained, excluding most leukocytes, to maximize the amount of analyzed renal parenchymal cells; for the subsequent patients (n=25) CD45$^-$CD66b$^-$ and CD45$^+$CD66b$^-$ were sorted, including lymphocytes and monocytes/macrophages in the measurements (Fig. S2). To adjust for low total cell counts in urine (median 3455 viable cells /100ml), samples P008-016 and P048-119 (Fig. S4D) were labelled with TotalSeq B Hashtag antibodies 1-4 (Biolegend, clone LNH-94) during the staining procedure described above and pooled in groups of four.

Single cell sequencing

Relative single cell counts in sorted suspensions were counted using a MACSQuant Analyzer (Miltenyi). Cells were centrifuged, resuspended, and re-counted on a hemocytometer. Cell suspensions were subjected to single-cell sequencing following the 10x Genomics protocol for Chromium Next GEM Single Cell 3' v3.1 chemistry targeting between 1000 - 10000 cells depending on sample cell count. In brief, cells were partitioned into a droplet with a barcoded gel bead using the 10x Chromium controller (10x Genomics, Pleasanton, California, USA) and lysed. RNA was reverse transcribed into complementary DNA (cDNA) within each droplet. cDNA was amplified and fragmented, followed by the addition of Illumina adapters using Single

Cell 3′ Library & Gel Bead Kit v3.1 (10x Genomics). Libraries were sequenced on Illumina HiSeq 4000 (Illumina, San Diego, California, USA) sequencers.

Data processing of scRNAseq libraries

Demultiplexing, barcode processing, read alignment and gene expression quantification was carried out using Cell Ranger software (v3.1.0, 10x Genomics). First, Cell Ranger mkfastq demultiplexed the sequencing by sample index. The quality of the data was checked using FastQC (v0.11.5) and all samples showed high quality RNAseq data with good median per-base quality across most of the read length. Cell Ranger count used STAR software to align sequenced reads to the reference genome (GRCh38.p12). Samples from COVID-19 patients included SARS-CoV1/2 genomic information to reference genome.

Sample Pool Demultiplexing

Demultiplexing of unfiltered hashed cell sample pools was performed in Seurat v4 [S5] (Fig. S3). Hashtag oligo (HTO) libraries were added to the Seurat object as an independent assay and normalized ("NormalizeData(SO, assay = "HTO", normalization.method = "CLR")"). Demultiplexing was performed with "HTODemux(SO, assay = "HTO")" with "positive.quantile =" between 0.90-0.99. Cross sample doublets were excluded. HTO negative cells were further analyzed as "POOL" samples (Fig. S3, 4).

scRNAseq data analysis

Seurat v4 was used for downstream analysis. ScDblFinder [S6] was used for removing multiplets. We merged all AKI sample datasets. Initial quality control was done by removing cells with less than 200 or more than 4,000 genes detected as well as cells with > 20% mitochondrial (mt) reads. For renal cells, a cut-off of 20% mt reads was applied to adjust for higher mitochondrial abundance in the kidney [S7], for leukocytes, a mt cut-off of 10% was applied. SCTransform [S8] was used for normalization, scaling, and determining variable features. After principal component analysis ("RunPCA" function, dims 1:45), integration and batch correction were

done by Harmony [S9] ("RunHarmony" function). Dimensional reduction by UMAP [S10] (dims=1:45) and clustering was performed using reduction="harmony". Cluster resolution was determined manually. "FindAllMarkers()" was applied with default settings to determine differentially expressed genes (DEG) in all clusters. Cluster annotation and rearrangement was done based on review of specific lineage markers (Fig. 1).

For further detailed analysis of kidney parenchymal cells in urine, cells of corresponding clusters were subsetted and re-analyzed using the same workflow as describe above. Cluster resolution and annotation was again performed manually based on DEG and known lineage markers. Clustering revealed 25 clusters, two of which were comprised of few remaining leukocytes and urogenital cells, which were excluded from further analysis (Fig. 2).

Reference datasets

For referencing of urinary cells, multiple datasets were used: AKI post-mortem biopsy data was retrieved as an annotated Seurat object from Hinze et al.; data were processed and analyzed as reported in the original publication [S2]. Healthy tumor adjacent kidney tissue data was downloaded from the Kidney Precision Medicine Project (KPMP) [S11] database (participants 17-1606, 18-139, 18-162, 18-342, 18142-5). scRNAseq data of 3 human bladder samples [S12] (GSE129845) was downloaded from the National Center for the Biotechnology Information Database (NCBI) / Gene expression omnibus (GEO) repository. Public urine scRNAseq data including 16 samples from 5 patients with diabetic nephropathy [S13] (GSE157640), 23 samples from 12 patients with focal segmental glomerulosclerosis (FSGS) [S14] (GSE176465) and 9 samples from 9 patients with AKI / COVID infection [S15] (GSE180595) were also downloaded from the NCBI/GEO repository. Data integration/analysis/dimensional reduction and clustering of these datasets was done exactly as with in urine AKI samples.

Automatic cell annotation with SingleR

SingleR [S16] was used for automatic unbiased cell type annotation in the urine TEC dataset using several references:

Human primary cell atlas bulk data [S1] was used to annotate cells in the kidney parenchymal cell Seurat object. Log10-normalized expression data and original annotations from AKI post mortem biopsy data was used to build a reference for all urine cells. Briefly, SingleR computes Spearman's rank correlation coefficient for the gene expression of each cell with set reference groups. Wilcoxon ranked sum test is used to identify the top markers used for each comparison. A score is defined for each cell based on the reference correlation in comparison to the other cells. Scores are determined across all reference groups. Cells are labeled according to their highest score.

Single-cell reference atlas mapping

For referencing of urinary cells with tissue samples, we used AKI post-mortem biopsy data and annotation from Hinze et al. Using symphony [S17] and the buildReference(…, K = 100, topn = 2000, d = 30, …) function, a reference map was constructed with default settings. Log10-normalized expression matrix was extracted from the integrated Seurat object containing all urine samples and mapped onto the reference using mapQuery(). For visualization, original annotations from Seurat objects were used.

Annotation by nephron segment location and disease state

For Fig. 3D, urine cells were annotated based on their reference atlas mapping (Fig. 3B) with symphony (knnPredict(query_obj = query, ref_obj = reference, train_labels = reference$meta_data$celltype, k = 5). Reference tissue cells were annotated as in the original publication [S2]. For the location plot, clusters were grouped based on nephron location (e.g. TL, TAL = loop of Henle). For the state plot, cells were grouped into healthy (PT, TL, TAL …) and injury (PT_New, TL_New, TAL_New, …) states based on the tissue reference annotation.

Gene expression correlation

For gene expression correlation (Fig. 3D, Fig. S21B, normalized mean expressions for the top 500 variable genes each per urine and reference datasets were calculated per cluster. Correlations were calculatd using rcorr() and visualized using corrplot.

Cross-species approach

Mouse ischemia reperfusion data was downloaded from GEO (GSE139107) [S18], meta data of PT subclustering from the original publication was kindly provided by the Humphreys lab. Healthy S1, S2 and S3 as well as injured S1/2 and S3 subclusters were joined to "healthy and "injured" for downstream analyses. Multinomial classification was performed using the glmnet package version 2.0-16. Training were randomly selected cells (2/3 of cells) from injured mouse PT subclusterings (Fig. S6). Genes used in the training were highly variable features of the mouse AKI data. Human orthologous genes were generated using Biomart [S19]. Test data were the remaining 1/3 of mouse PT injured subclusters. Glmnet produces different models for different values of lambda which determines how hard overcomplexity of the respective model gets punished. Each so-generated model was tested on the test data and the model with the highest accuracy on the test data was determined. The so selected model was then applied to our human urine TEC clusters.

Pseudotime trajectory analyses

To infer potential dynamics of urinary TEC after injury, we performed trajectory analyses. We extracted all TEC clusters from our integrated Seurat object of all kidney parenchymal cells. We used SeuratWrappers function as.cell_data_set() to transfer and further analyze the clusters with Monocle 3 [S20–S22]. Using default parameters, trajectory graphs in the dataset were calculated. For pseudotime calculation, starting states were determined as clusters with high expression of marker genes for differentiated TEC (*UMOD/SLC12A1* for TAL, *LRP2/GPX3* for PT) and a lack of injury signatures (*LCN2, HAVCR1*).

Ligand receptor interactions

We used CellPhoneDB (2.1.1) [S23] to assess cellular crosstalk between different cell types. As input data, we calculated normalized data with scale factor 10,000. To identify putative cell-cell interactions via multi-subunit ligand-receptor complex pairs, label permutation was performed. Finally, we conducted statistical analyses by randomly permuting the cluster labels of each cell 10,000 times. Tests were conducted intraindividually.

Gene set and pathway analyses

GO term enrichment analysis was performed with topGO [S24] using all DEG with an average log-fold change of > 0.5 in cells with SCT expression of >= 1 for *PAX2* or *PROM1* ("runTest(GOdata, statistic = "fisher", algorithm = "elim")"). GO set visualization was done with rrvgo [S25]. Rrvgo groups the provided GO terms based on their similarity. GO group representatives were selected by rrvgo based on adjusted p value.

Pathway activity across TEC subsets was calculated with PROGENy [S26]. TEC clusters were subsetted and progeny activity scores were calculated using progeny(object, scale=FALSE, organism="Human", top=500, perm=1) and scaled using ScaleData(). Displayed are average scores per cluster.

Supplemental Results

The renal origin of urine cells cannot be conclusively determined and comparisons of disease with (often cell-free) healthy control samples is difficult, resulting in controversial debates on the validity of urine-derived data. It is therefore all the more important to compare our own data with meaningful references. In this supplement, we include alignments of our own urine data with several public datasets:

- Human urine sample pool from 12 healthy donors [S13] (GSE157640)
- Human urine samples from five patients with diabetic nephropathy [S13] (GSE157640)
- Human urine samples from 12 patients with focal segmental glomerulosclerosis (FSGS) [S14] (GSE176465)
- Human urine samples from four patients with AKI and five patients with AKI + COVID-19 infection [S15] (GSE180595)
- Healthy tumor adjacent, human kidney tissue data of five participants from the Kidney Precision Medicine Project (KPMP) [S11] database (participants 17-1606, 18-139, 18-162, 18-342, 18142-5)
- Healthy human bladder samples of three participants [S12] (GSE129845)

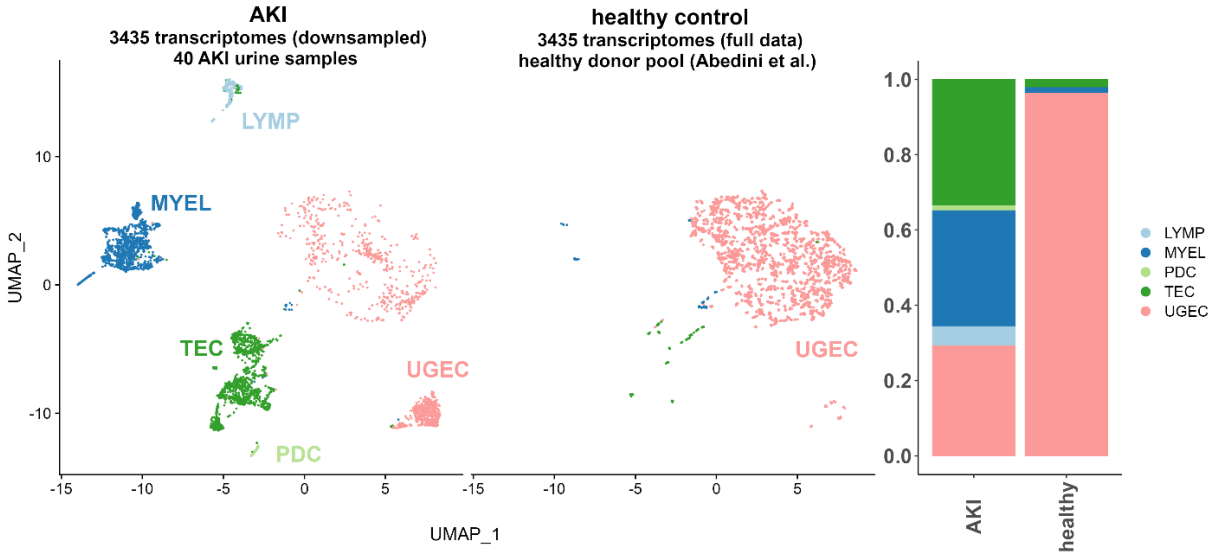


Figure S15. Healthy urine sediment is almost devoid of kidney and immune cells.
Uniform manifold approximation and projection (UMAP) of AKI urine sediment (downsampled from 42608 *scRNA-seq urine cells from 32 individuals) and a public dataset of healthy controls* [S13]

Healthy urine sediment is almost devoid of kidney and immune cells

To investigate whether the urine cell signature and composition found in our data is driven by disease, we compared our data to a public dataset pooled from multiple healthy donors (GSE157640) [S13], including 3435 urinary transcriptomes. This healthy control was comprised almost exclusively of urogenital transcriptomes (>95%), while AKI sediments included equal amounts of renal parenchymal cells (TEC, PDC), leukocytes and UGEC (Fig. S15). As UGEC were not a focus of our study, we didn't investigate transcriptional changes between healthy and AKI UGEC.

Tubular epithelial cell abundance is dependent on AKI etiology

After seeing clearly how kidney injury induces the occurrence of kidney and immune cells in urine samples, we asked whether a differing AKI etiology influences the cellular urine composition in AKI.

Therefore, we examined the data of different AKI entities separately and compared their cellular composition (Fig. S16): Individuals with cardiac-surgery (CS) or pneumonia / COVID-19 associated AKI showed a similar pattern of urinary composition with many renal cells (33 and 36% of all single-cell transcriptomes) and differed only in a more pronounced (B) lymphocyte signal in the pneumonia patient pool (Cluster 17 in Fig. S16B+D). The prerenal AKI pool contributed mostly to myeloid (clusters 10-15) and urogenital (clusters 19, 20) cells and had much less kidney parenchymal cells (7%, clusters 1-9, Fig. S16D), consistent with less severe tubular damage.

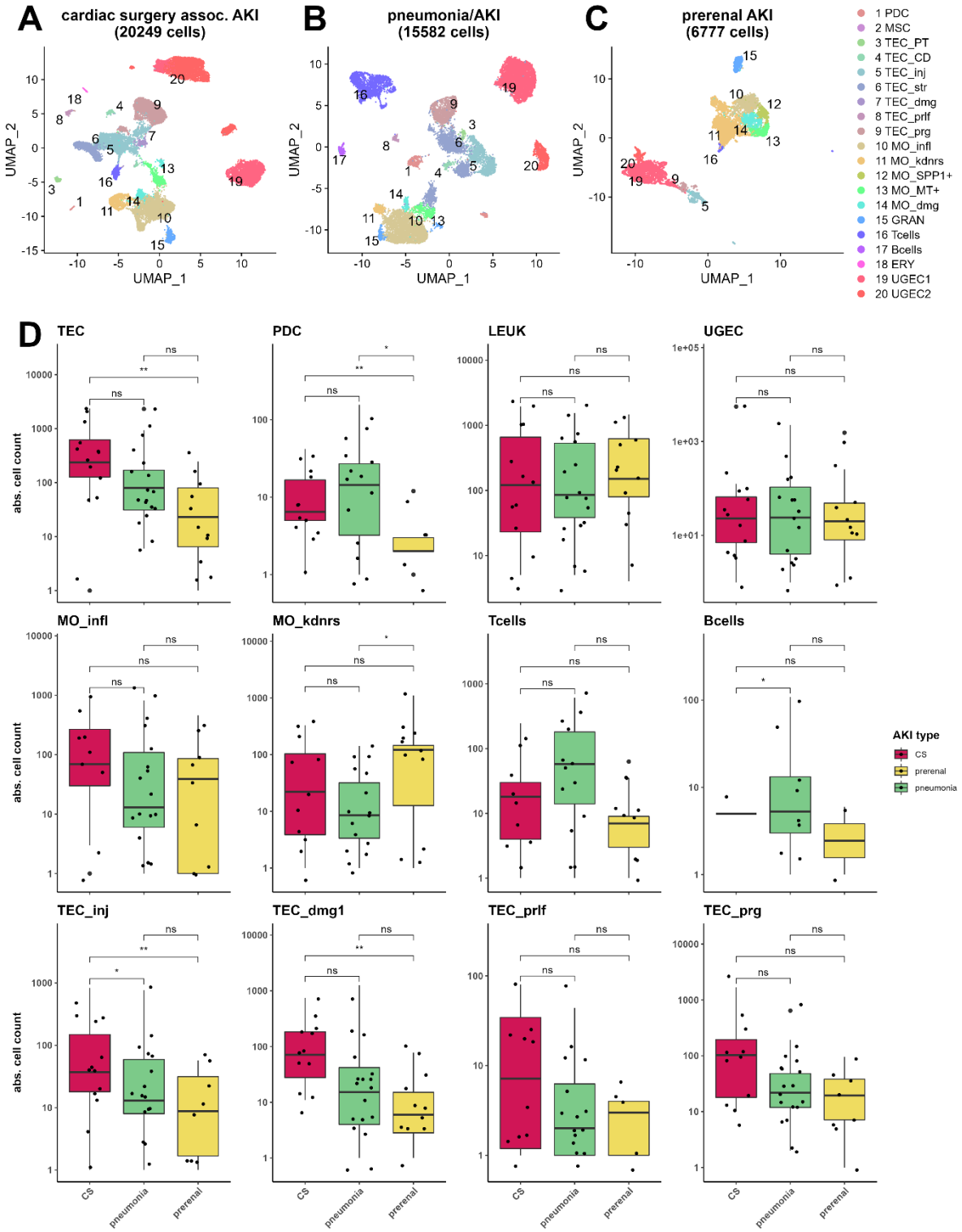


Figure S16. Tubular epithelial cell abundance is dependent on AKI etiology

A-C. *UMAP of urinary cell signatures by AKI etiology* ***D****. Boxplots of absolute counts of urinary cell subsets per sample in different AKI etiologies (cardiac surgery (CS) n=7, – red, pneumonia n=15 – green, prerenal n=10 – yellow). PDC – podocytes, TEC – tubular epithelial cells, UGEC - urogenital epithelial cells, LEUK – leukocytes, PT – proximal tubule, CD – collecting duct, MO –* monocytes/macrophages, *GRAN – granulocytes, inj – injured, dmg – damaged, prlf – proliferating, prg – progenitor-like, infl – inflammatory, kdnrs – kidney resident. Boxplot: line = median, hinges = first -third quartile, whiskers = 1.5 interquartile range, large dots = outliers, *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001, unpaired two-samples Wilcoxon test. ns, not significant.*

SARS-Cov-2 infection alters the urinary transcriptomic signature

Our cohort also included SARS-Cov-2 positive patients. We compared COVID (n=14) and non-COVID AKI (n=18) urine samples and found an altered gene signature, in accordance to recently published data discussing SARS-Cov-2 as an active driver of kidney inflammation and fibrosis [S27]: While urinary COVID-19 AKI patients showed no transcriptional activity for the SARS-Cov-2 genome, previously reported COVID-19 associated genes [S27,S28] (*PLCG2, AFDN*), inflammation (*IL1B*) and damage markers (*HSPA1B*) were differentially expressed in urine of these patients (Fig. S17). This comparison must be interpreted carefully, as non-COVID AKI patients predominantly had cardiac-surgery associated or prerenal causes for their AKI and thus are no ideal control cohort for the septic pneumonia COVID-AKI patients.

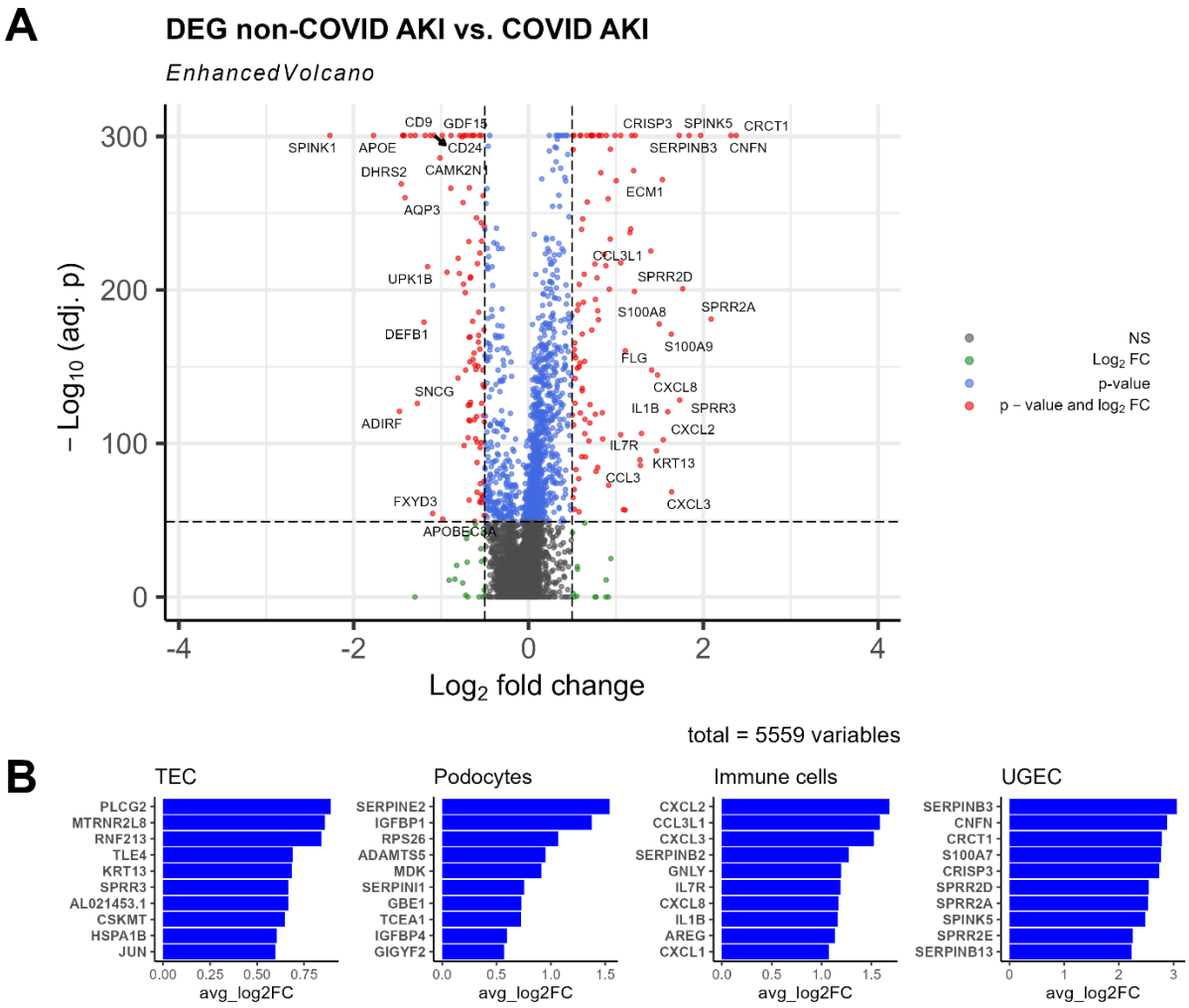


Figure S17. Urine cells show COVID-19 associated transcriptional signatures.
A *Volcano plot indicating significance (y) and differential expression of genes (x) between all urinary cells from COVID-19 AKI patients (n=14) (right) and other AKI patients (n=18) (left).* ***B*** *Bar plots with the top ten most differentially expressed genes per urine cell subgroup. All adj. p values < 10e-5. UGEC – urogenital epithelial cells.*

Disease type influences urine cellular abundance and composition

One of the main opportunities of urinary scRNAseq could be its potential to non-invasively diagnose or stratify diseases. Although data is still scarce, we had the opportunity to compare our AKI data to three different cohorts to test whether any findings are specifically tied to the acute injury: We compared our data with recently published urinary scRNAseq data of chronic glomerular diseases DN [S13] and FSGS [S14] and another cohort of 9 AKI patients both with and without COVID-19 infection [S15] (Fig. S18A-F). Overall, broad cell types like leukocytes, urogenital and tubular epithelial cells were present in all diseases. DN and FSGS urine samples had less TEC and myeloid cells than the more severe forms of AKI studied in out cohort (Fig. S18G), potentially reflecting less acute damage. UGEC were much more prominent in DN and the public AKI samples (Fig. S18G), but this is likely a technical artifact, since our own samples and part of the FSGS samples were pre-sorted, while DN and the second AKI cohort samples were not.

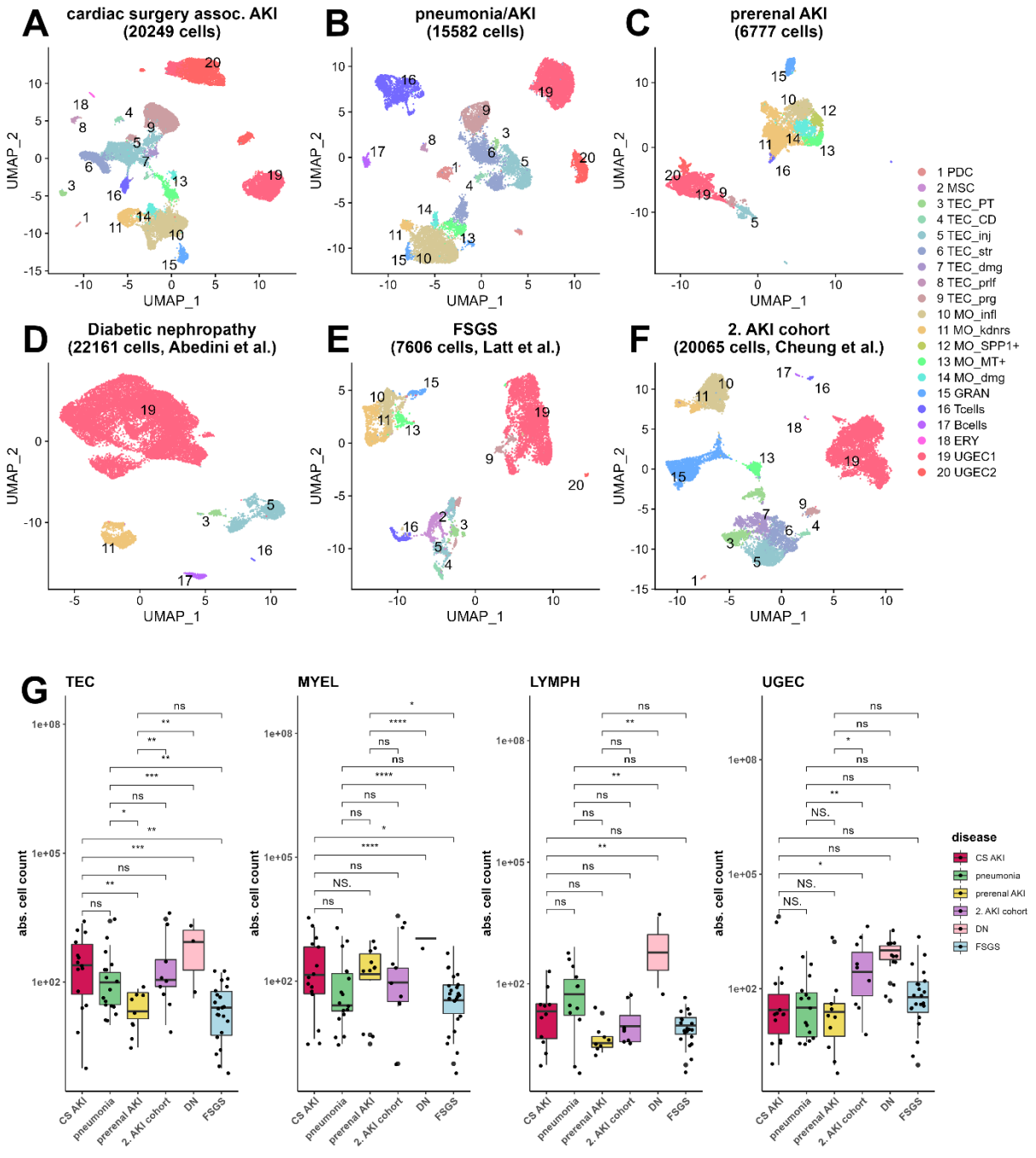


Figure S18. Disease type influences urine cellular abundance and composition

A-C. *UMAP of urinary cell signatures by AKI etiology compared to public urine datasets of diabetic nephropathy*[S13] *(DN,* ***D****),focal segmental glomerulosclerosis*[S14] *(FSGS,* ***E****) and COVID-19 assoc. AKI*[S15] *(AKI/COVID,* ***F****).* ***G.*** *Boxplots of absolute counts of urinary cell subsets per sample in different AKI etiologies (cardiac surgery (CS) n=7, – red, pneumonia n=15 – green, prerenal n=10 – yellow) and public datasets (2. AKI cohort – purple, diabetic nephropathy*[S13] *(DN) – pink, FSGS*[S14] *– light blue). PDC – podocytes, TEC – tubular epithelial cells, LYMP – lymphoid cells, MYEL – myeloid cells, PT – proximal tubule, CD – collecting duct, MO –* monocytes/macrophages, *GRAN – granulocytes, UGEC – urogenital epithelial cells, inj – injured, dmg – damaged, prlf – proliferating, prg – progenitor-like, infl – inflammatory, kdnrs – kidney resident. Boxplot: line = median, hinges = first -third quartile, whiskers = 1.5 interquartile range, large dots = outliers, *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001, unpaired two-samples Wilcoxon test. ns, not significant.*

Adaptive progenitor-like cell abundance is increased in AKI

Looking more specifically at TEC, we subsetted and integrated all kidney parenchymal cells from each disease cohort, finding 12 unique clusters (Fig. S19), which were annotated analogous to the TEC annotation found in Fig. 2 and Fig. S6. Podocytes (PDC) were more frequent in severe AKI forms (in cardiac surgery and pneumonia patients) than in DN and FSGS samples (Fig. S20). Most TEC subsets showed no different abundance between diseases, except for the presumed adaptive cell state TEC_prg, which was significantly increased in all AKI etiologies when compared with FSGS and DN (Fig. S20). The public AKI dataset had similar amounts of TEC_prg per sample, indicating a potential AKI specificity and making an artificial origin of this subset less likely. In conclusion, many cellular features of the urine sediment are similarly excreted in all examined diseases. This analysis however was only concerned with overall cell phenotype and quantity. Going forward, an in-depth analysis of transcriptional signatures will be needed to evaluate disease specific changes/features in more detail.

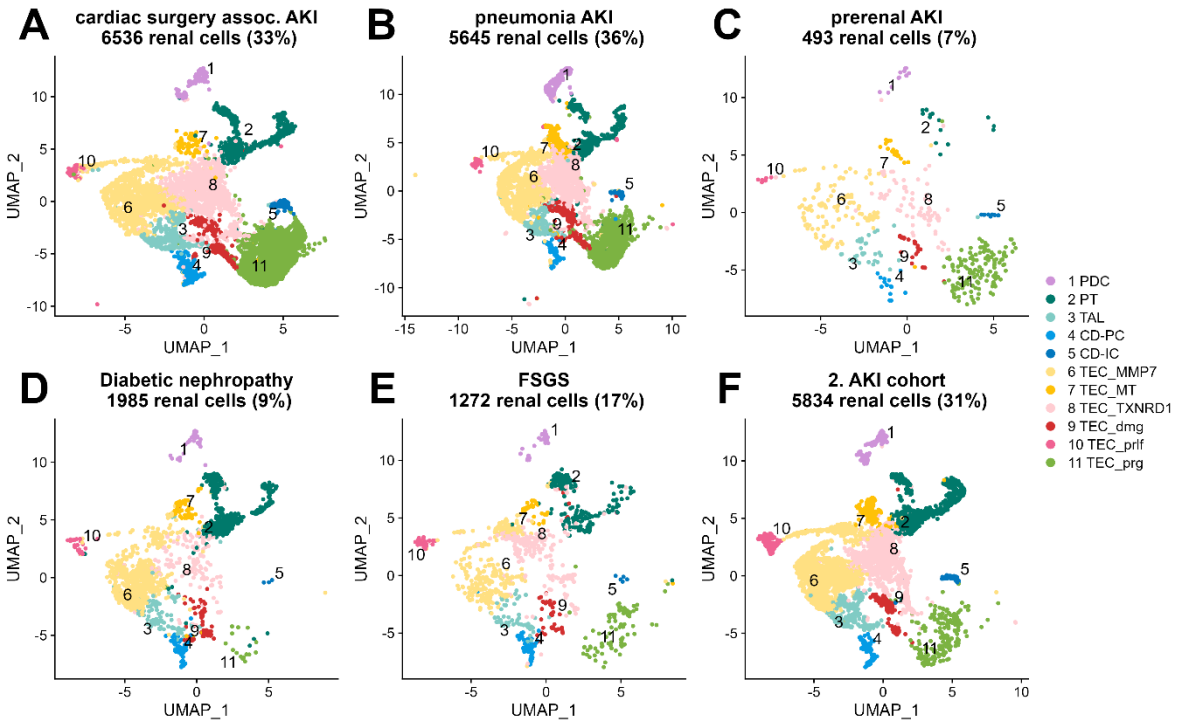


Figure S19. Adaptive progenitor-like cell abundance is increased in AKI

A-C. *UMAP of integrated urinary renal parenchymal cells by AKI etiology compared to public urine datasets of diabetic nephropathy*[S13] *(DN,* ***D****),focal segmental glomerulosclerosis*[S14] *(FSGS,* ***E****) and a second AKI cohort*[S15]*(****F****). PDC – podocytes, PT – proximal tubule, TAL – thick ascending limb, CD –*

collecting duct, PC – principal cells, IC – intercalated cells, dmg – damaged, prlf – proliferating, prg – progenitor-like.

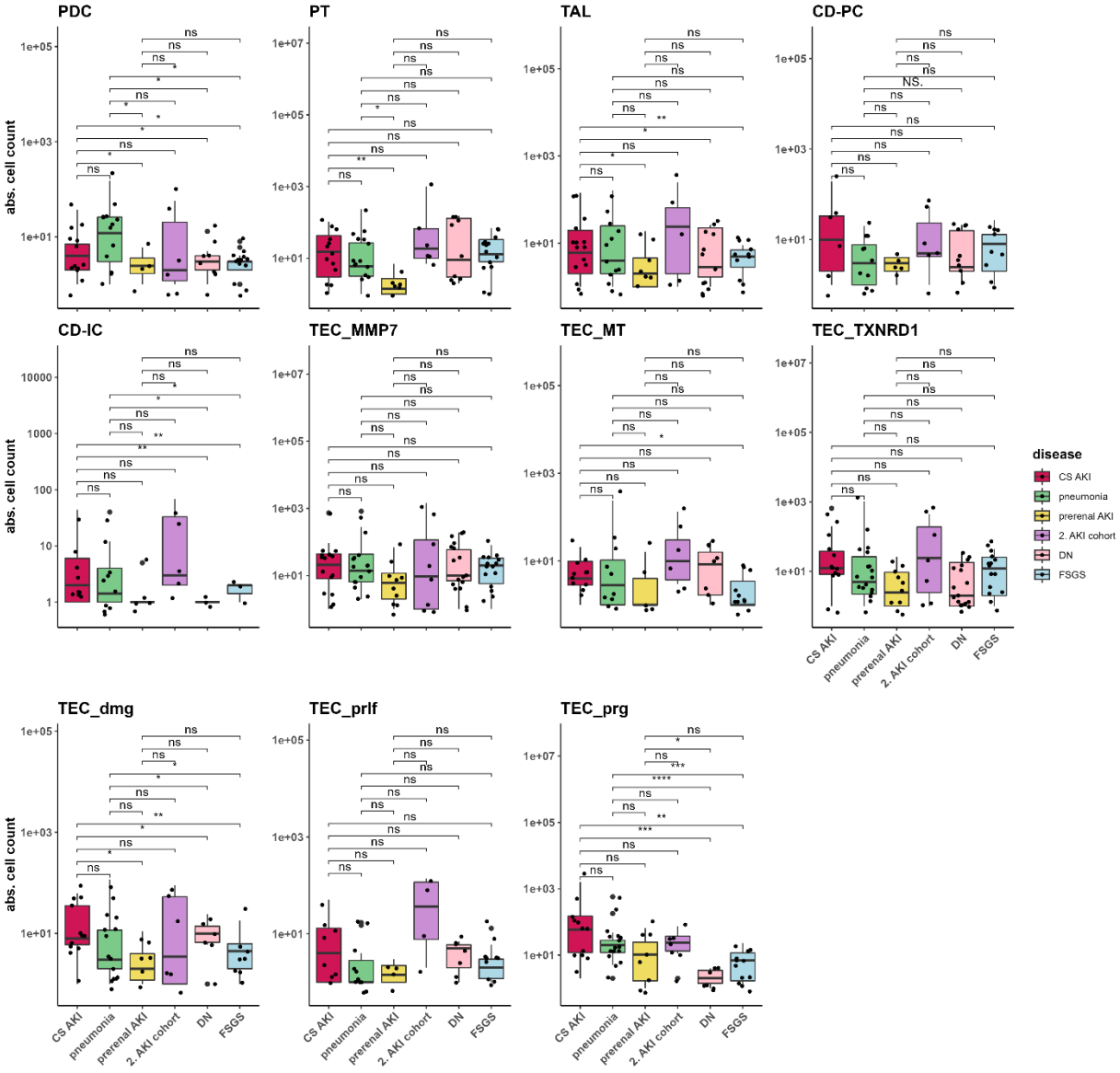


Figure S20. Adaptive progenitor-like cell abundance is increased in AKI
*Boxplots of absolute counts of urinary cell subsets per sample in different AKI etiologies (cardiac surgery (CS) n=7, – red, pneumonia n=15 – green, prerenal n=10 – yellow) and public datasets (second AKI cohort[S15] – purple, diabetic nephropathy[S13] (DN) – pink, FSGS[S14] – light blue). TEC – tubular epithelial cells, PDC – podocytes, PT – proximal tubule, TAL – thick ascending limb, CD – collecting duct, PC – principal cells, IC – intercalated cells, dmg – damaged, prlf – proliferating, prg – progenitor-like. Boxplot: line = median, hinges = first -third quartile, whiskers = 1.5 interquartile range, large dots = outliers, *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001, unpaired two-samples Wilcoxon test. ns, not significant.*

Urine epithelial cells are partly kidney derived

Working with urine derived cells always brings forth two major skeptic questions: “Are those cells really derived from the kidney and not from the urogenital tract?” and “To what extent does prolonged contact to urine alter the phenotype of these cells?”. We addressed these questions by comparison to two further references: In an integrated dataset with tumor adjacent healthy human kidney tissue (KPMP [S11]) and healthy bladder tissue [S12], urinary TEC clustered with kidney derived TEC, while urinary UGEC clustered with bladder cells (Fig. S21), providing further evidence for the renal origin of urinary TEC. We then compared TEC from all tissues and urinary datasets at hand for the occurrence of gene signatures associated with salt-, fluid shear-, osmotic- or pH-stress. Urine samples uniformly had a higher mean expression of the salt-stress signature genes, but all other stress related gene sets were not enriched in the urinary datasets, suggesting that the impact of exposure to urine was confined (Fig. S22). This analysis is limited, as it includes data derived from different techniques (single-cell (urine samples, healthy tissue) and single-nuclei RNA sequencing (AKI tissue) and thus may partly represent technical batch effects.

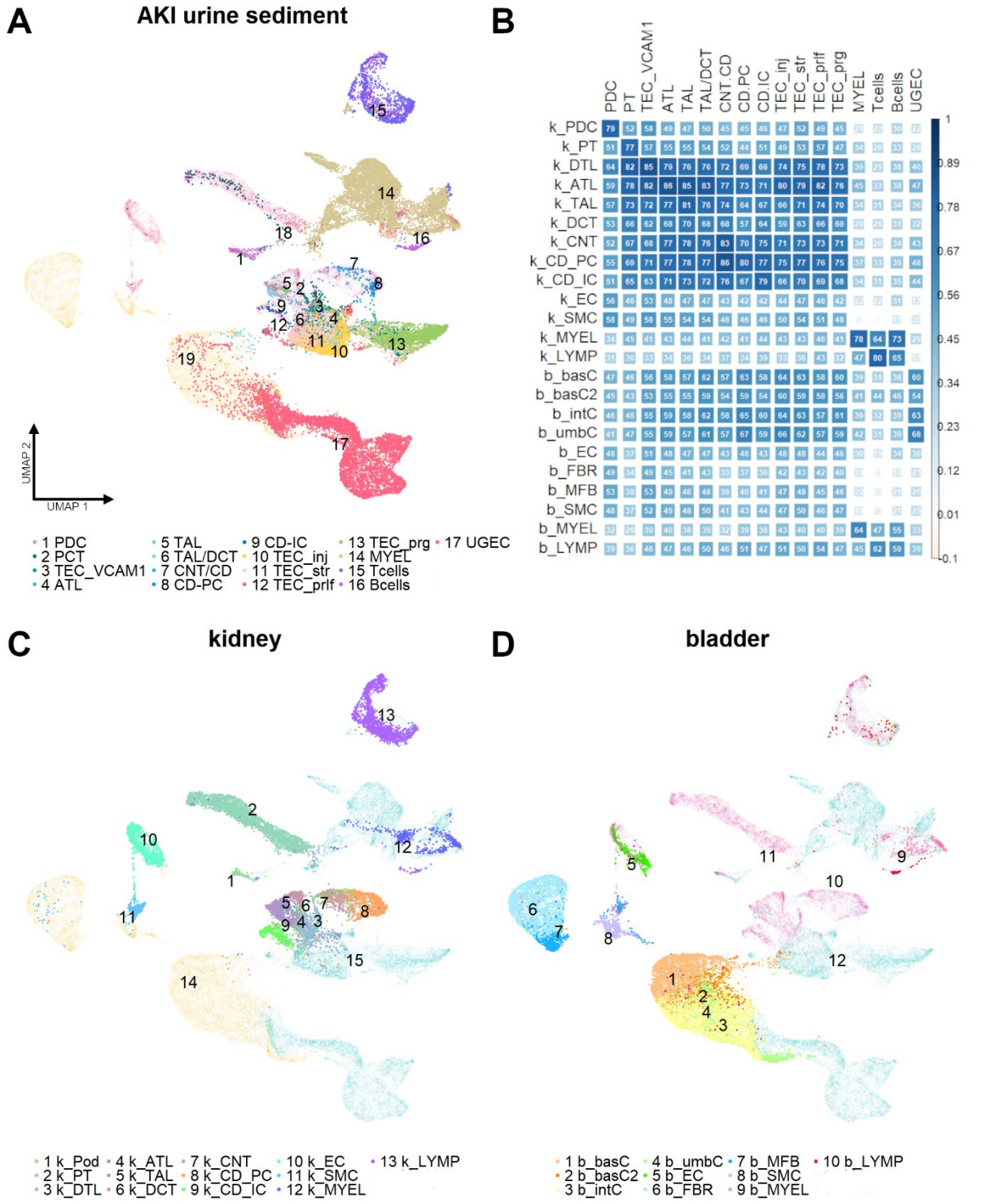


Figure S21. Urine tubular cells cluster with kidney, not bladder tissue

UMAP dimensional reduction of tumor-adjacent healthy kidney tissue scRNAseq[S11] *and healthy bladder tissue*[12] *dataset.* ***A+C-D.*** *Integrated healthy kidney (**C**), bladder (**D**) and AKI urinary scRNAseq (**A**) datasets. Urine tubular epithelial cells (TEC) cluster with kidney epithelial cells while urine urogenital cells cluster with bladder cells, confirming a kidney origin of urinary TEC.* ***B****. Correlation plot for gene expression urine clusters (columns) vs. kidney (k clusters) and bladder (b clusters) tissue clusters (rows). Size and color represent Spearman R (displayed as percentage), all p <0.001. PDC – podocytes, PT – proximal tubule, DTL – descending thin limb, ATL – ascending thin limb, TAL – thick ascending limb, DCT – distal convoluted tubule, CNT – connecting tubule, CD-PC – collecting duct principal cells, CD-IC – collecting duct intercalated cells, MYEL – myeloid cells, LYMP – lymphocytes, EC – endothelial cells, FBR – fibrocytes, MFB – myofibroblasts, SMC – smooth muscle cells, UGEC – urogenital epithelial cells, including basC - basal cells, intC intermediate cells, umbC – umbrella cells, emt – epithelial-*

mesenchymal transition, str – stressed, prlf – proliferating, prg – progenitor-like, infl – inflammatory, kdnrs – kidney resident.

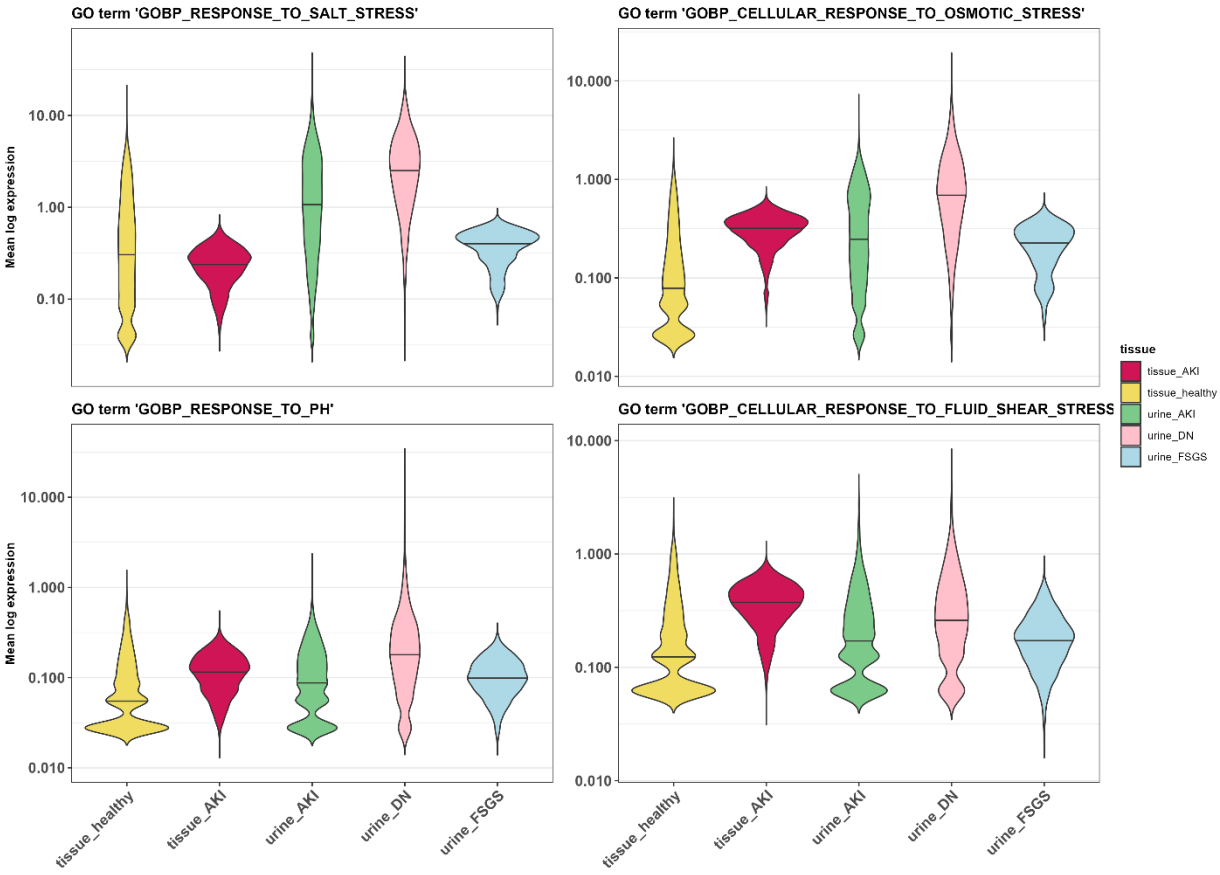


Figure S22. Stress genes are partially upregulated in urine cells
Violin plot of mean log normalized expression per cell of genes contained in named Gene Ontology (GO) term. Violin plot: line = median, widths = frequency, tips = lowest and highest values.

Urinary tubular subsets show distinct interactions with epithelial and immune cells

Emphasizing the difference between TEC and presumed progenitors we used the CellPhone database to search for distinct cellular crosstalk of these cells with other epithelia and leukocytes (Fig. S23): FGF/FGF receptor (FGFR) crosstalk in epithelial cells was pronounced in TEC_prg [S29]. Adrenoreceptor Beta 2 (ADRB2) signaling with IL1B and VEGF which is not usually expressed in epithelia but in mesenchymal stem cells [S30] and renal clear-cell carcinoma [S31], was also expressed exclusively on TEC_prg. In contrast, other TEC contributed to wound healing via the CD44-osteopontin (SPP1) axis [S32(p44),S33(p44),S34(p44)] and EGFR-signaling via various ligands including EGFR-TGFß1 transactivation, known for inducing EMT phenotypes and renal fibrosis [S35–S37]. The epithelial-leukocyte interaction via plexin B (PLXNB) and semaphorin 4D (SEMA4D), which has implications in developmental ureteric branching[S38] are redundantly featured in both TEC and progenitors.

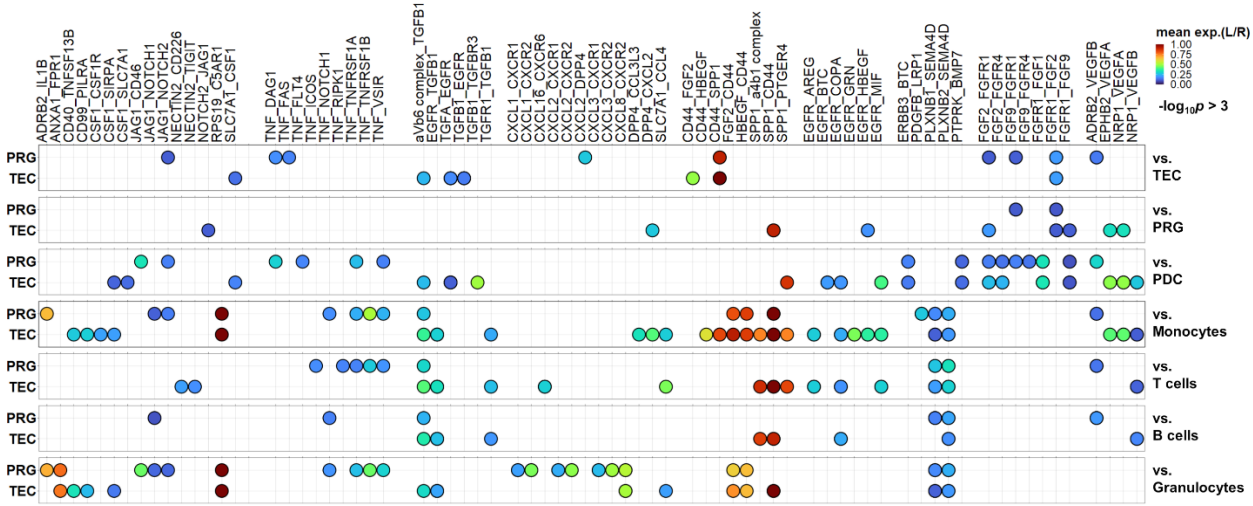


Figure S23. Subset specific cellular crosstalk of urinary tubular cells.
Cellphone DB cell crosstalk of progenitor-like cells (PRG) and tubular epithelial cells (TEC) (right, first interactor) with selected cell types (left, second interactor). Color indicates sum of receptor and ligand expression in respective cells, all interactions -log10(adj. p value) >3.

References

S1. Mabbott NA, Baillie JK, Brown H, Freeman TC, Hume DA. An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics*. 2013;14(1):632. doi:10.1186/1471-2164-14-632

S2. Hinze C, Kocks C, Leiz J, et al. *Transcriptomic Responses of the Human Kidney to Acute Injury at Single Cell Resolution*.; 2021:2021.12.15.472619. doi:10.1101/2021.12.15.472619

S3. Annexin V Binding Buffer. *Cold Spring Harb Protoc*. 2016;2016(11):pdb.rec088443. doi:10.1101/pdb.rec088443

S4. Westendorf K, Okhrimenko A, Grün JR, et al. Unbiased transcriptomes of resting human CD4+CD45RO+ T lymphocytes. *European Journal of Immunology*. 2014;44(6):1866-1869. doi:10.1002/eji.201344323

S5. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184(13):3573-3587.e29. doi:10.1016/j.cell.2021.04.048

S6. Germain PL, Lun A, Macnair W, Robinson MD. Doublet identification in single-cell sequencing data using scDblFinder. Published online September 28, 2021. doi:10.12688/f1000research.73600.1

S7. Schulz S, Lichtmannegger J, Schmitt S, et al. A Protocol for the Parallel Isolation of Intact Mitochondria from Rat Liver, Kidney, Heart, and Brain. In: Posch A, ed. *Proteomic Profiling: Methods and Protocols*. Part of the book series: Methods in Molecular Biology. Springer; 2015:75-86. doi:10.1007/978-1-4939-2550-6_7

S8. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology*. 2019;20(1):296. doi:10.1186/s13059-019-1874-1

S9. Korsunsky I, Fan J, Slowikowski K, et al. *Fast, Sensitive, and Accurate Integration of Single Cell Data with Harmony*.; 2018:461954. doi:10.1101/461954

S10. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:180203426 [cs, stat]*. Published online September 17, 2020. Accessed November 30, 2021. https://doi.org/10.48550/arXiv.1802.03426

S11. de Boer IH, Alpers CE, Azeloglu EU, et al. Rationale and design of the Kidney Precision Medicine Project. *Kidney International*. 2021;99(3):498-510. doi:10.1016/j.kint.2020.08.039

S12. Yu Z, Liao J, Chen Y, et al. Single-Cell Transcriptomic Map of the Human and Mouse Bladders. *J Am Soc Nephrol*. 2019;30(11):2159-2176. doi:10.1681/ASN.2019040335

S13. Abedini A, Zhu YO, Chatterjee S, et al. Urinary Single-Cell Profiling Captures the Cellular Diversity of the Kidney. *JASN*. 2021;32(3):614-627. doi:10.1681/ASN.2020050757

S14. Latt KZ, Heymann J, Jessee JH, et al. Urine single cell RNA-sequencing in focal segmental glomerulosclerosis reveals inflammatory signatures. *Kidney International Reports*. Published online November 25, 2021. doi:10.1016/j.ekir.2021.11.005

S15. Cheung MD, Erman EN, Liu S, et al. Single-Cell RNA Sequencing of Urinary Cells Reveals Distinct Cellular Diversity in COVID-19-Associated AKI. *Kidney360*. 2022;3(1):28-36. doi:10.34067/KID.0005522021

S16. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*. 2019;20(2):163-172. doi:10.1038/s41590-018-0276-y

S17. Kang JB, Nathan A, Weinand K, et al. Efficient and precise single-cell reference atlas mapping with Symphony. *Nat Commun*. 2021;12(1):5890. doi:10.1038/s41467-021-25957-x

S18. Kirita Y, Wu H, Uchimura K, Wilson PC, Humphreys BD. Cell profiling of mouse acute kidney injury reveals conserved cellular responses to injury. *Proc Natl Acad Sci USA*. 2020;117(27):15874-15883. doi:10.1073/pnas.2005477117

S19. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*. 2009;4(8):1184-1191. doi:10.1038/nprot.2009.97

S20. Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*. 2019;566(7745):496-502. doi:10.1038/s41586-019-0969-x

S21. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. 2014;32(4):381-386. doi:10.1038/nbt.2859

S22. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*. 2017;14(10):979-982. doi:10.1038/nmeth.4402

S23. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc*. 2020;15(4):1484-1506. doi:10.1038/s41596-020-0292-x

S24. Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. Accessed 13 September 2022. Published online 2021. doi:10.18129/B9.bioc.topGO

S25. Using the rrvgo package. Accessed February 12, 2022. https://ssayols.github.io/rrvgo/articles/rrvgo.html

S26. Schubert M, Klinger B, Klünemann M, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun*. 2018;9(1):20. doi:10.1038/s41467-017-02391-6

S27. Jansen J, Reimer KC, Nagai JS, et al. SARS-CoV-2 infects the human kidney and drives fibrosis in kidney organoids. *Cell Stem Cell*. Published online December 25, 2021. doi:10.1016/j.stem.2021.12.010

S28. Delorey TM, Ziegler CGK, Heimberg G, et al. COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature*. 2021;595(7865):107-113. doi:10.1038/s41586-021-03570-8

S29. Xie Y, Su N, Yang J, et al. FGF/FGFR signaling in health and disease. *Sig Transduct Target Ther*. 2020;5(1):1-38. doi:10.1038/s41392-020-00222-7

S30. Tyurin-Kuzmin PA, Fadeeva JI, Kanareikina MA, et al. Activation of β-adrenergic receptors is required for elevated α1A-adrenoreceptors expression and signaling in mesenchymal stromal cells. *Sci Rep*. 2016;6(1):32835. doi:10.1038/srep32835

S31. Ha M, Kim DW, Kim J, et al. Prognostic role of the beta-2 adrenergic receptor in clear cell renal cell carcinoma. *Anim Cells Syst (Seoul)*. 2019;23(5):365-369. doi:10.1080/19768354.2019.1658638

S32. Lewington AJP, Padanilam BJ, Martin DR, Hammerman MR. Expression of CD44 in kidney after acute ischemic injury in rats. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*. 2000;278(1):R247-R254. doi:10.1152/ajpregu.2000.278.1.R247

S33. Sibalic V, Fan X, Loffing J, Wüthrich RP. Upregulated renal tubular CD44, hyaluronan, and osteopontin in kdkd mice with interstitial nephritis. *Nephrol Dial Transplant*. 1997;12(7):1344-1353. doi:10.1093/ndt/12.7.1344

S34. Weber GF, Ashkar S, Glimcher MJ, Cantor H. Receptor-ligand interaction between CD44 and osteopontin (Eta-1). *Science*. 1996;271(5248):509-512. doi:10.1126/science.271.5248.509

S35. Gu YY, Liu XS, Huang XR, Yu XQ, Lan HY. Diverse Role of TGF-β in Kidney Disease. *Frontiers in Cell and Developmental Biology*. 2020;8:123. doi:10.3389/fcell.2020.00123

S36. Samarakoon R, Dobberfuhl AD, Cooley C, et al. Induction of renal fibrotic genes by TGF-β1 requires EGFR activation, p53 and reactive oxygen species. *Cellular Signalling*. 2013;25(11):2198-2209. doi:10.1016/j.cellsig.2013.07.007

S37. Harskamp LR, Gansevoort RT, van Goor H, Meijer E. The epidermal growth factor receptor pathway in chronic kidney diseases. *Nat Rev Nephrol*. 2016;12(8):496-506. doi:10.1038/nrneph.2016.91

S38. Perälä N, Jakobson M, Ola R, et al. Sema4C-Plexin B2 signalling modulates ureteric branching in developing kidney. *Differentiation*. 2011;81(2):81-91. doi:10.1016/j.diff.2010.10.001