Supplemental Information

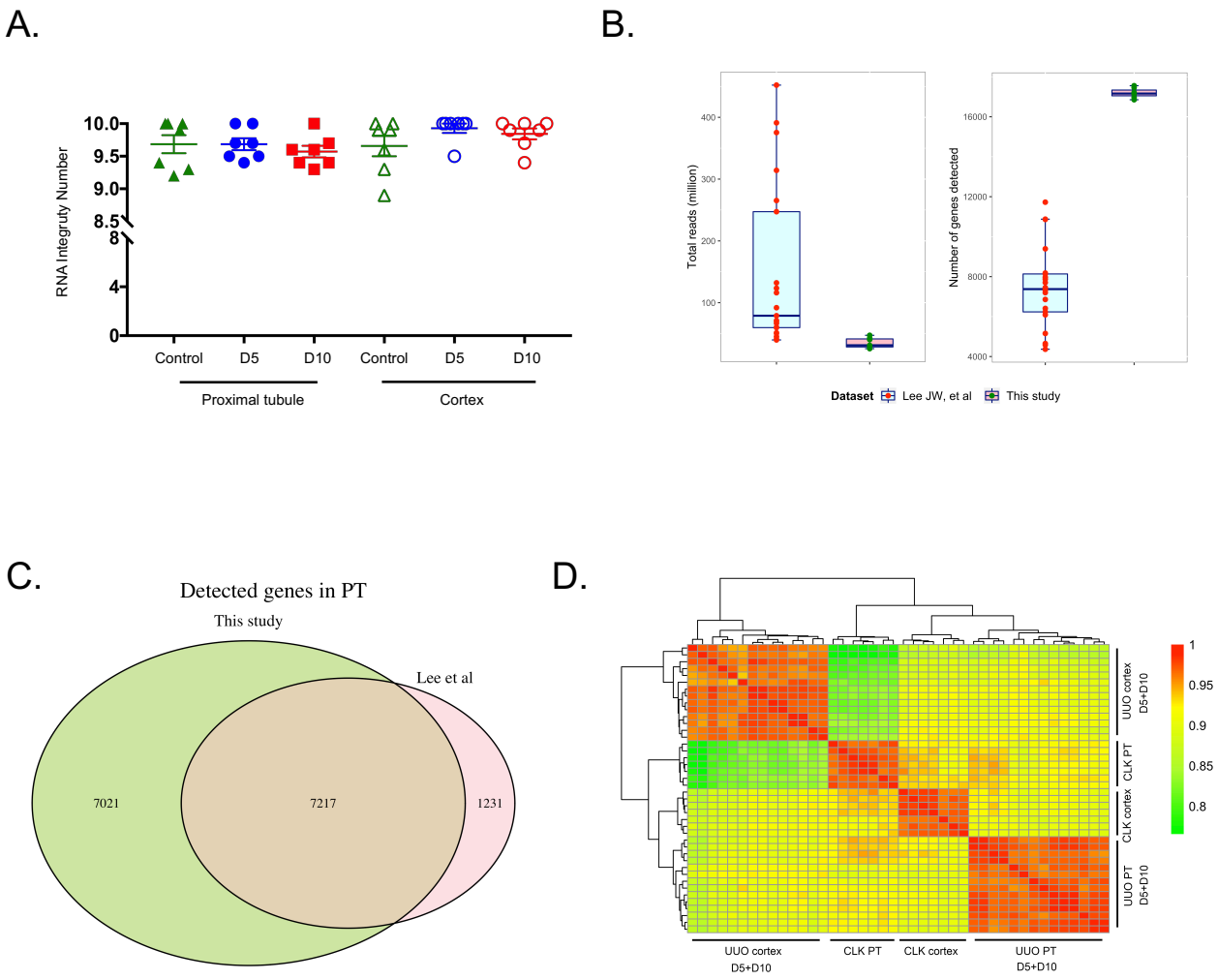
Proximal tubule translational profiling during kidney fibrosis reveals pro-inflammatory and lncRNA expression patterns with sexual dimorphism

Haojia Wu, Chun-Fu Lai, Monica Chang-Panesso and Benjamin D. Humphreys

Supplementary Figures

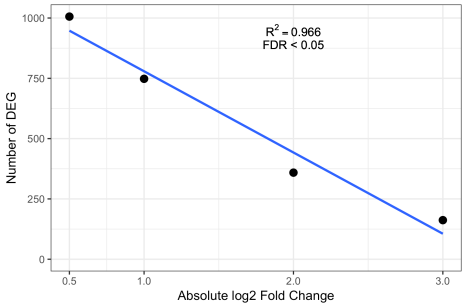
Supplementary Methods

Supplementary References

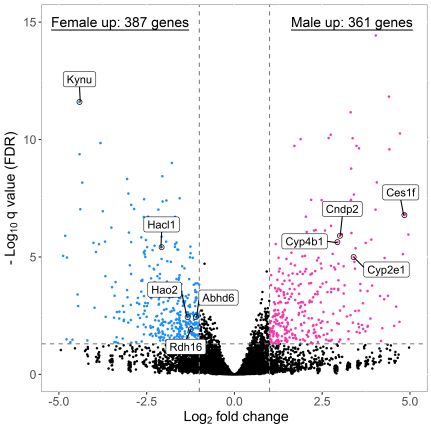


Supplementary Figure S1. TRAP sample quality control.
A. RNA integrity as determined by BioAnalyzer. N=7 for each group. **B.** Read depth and total gene detection in Lee *et al*[1] compared with this study. **C.** Venn diagram revealing the overlapping genes detected in this study and in Lee *et al*.[1] **D.**Heatmap showing the correlation between each two samples. The scale bar represents the Pearson correlation coefficient. PTEC, proximal tubule epithelial cells.

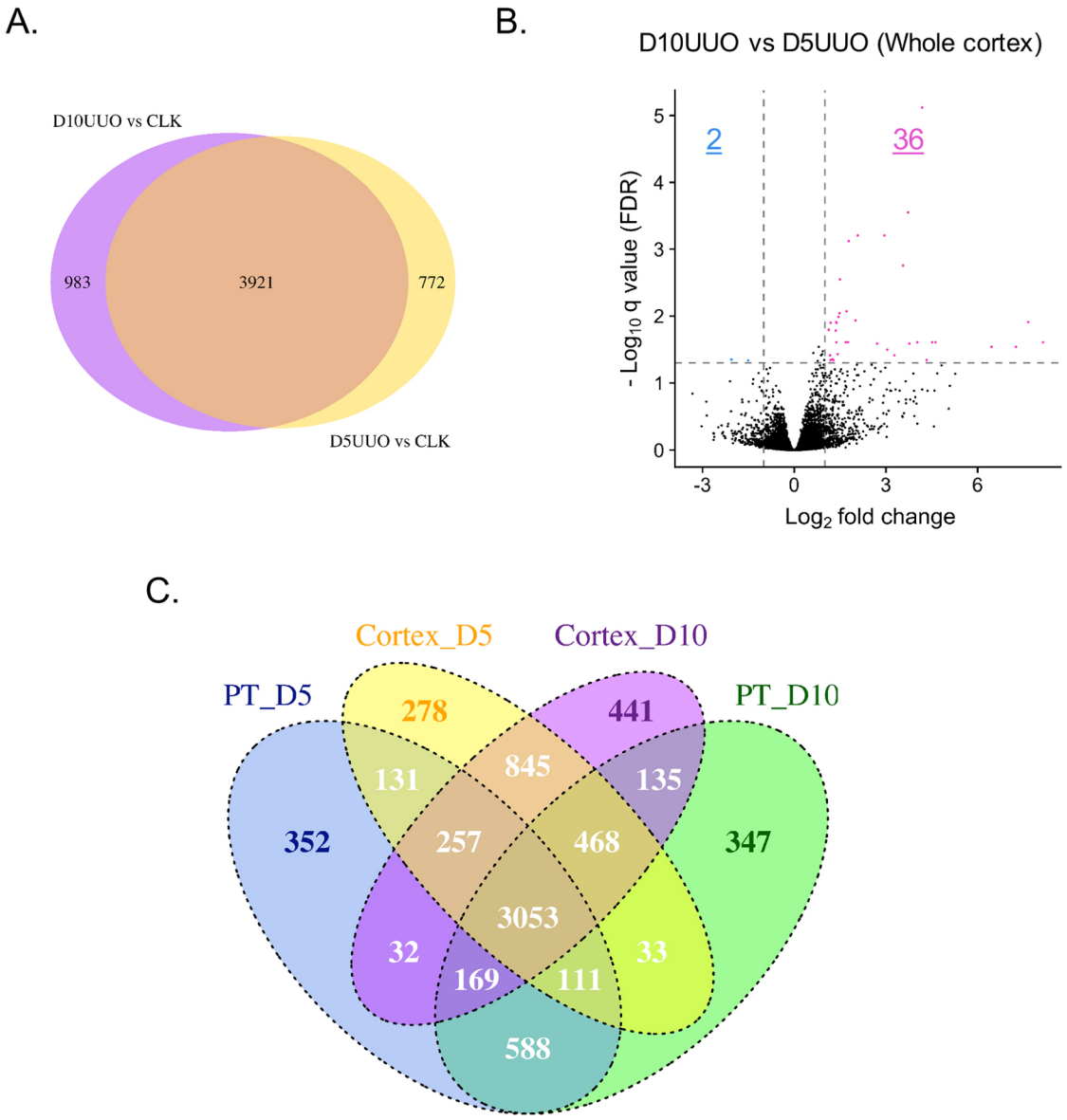
A.



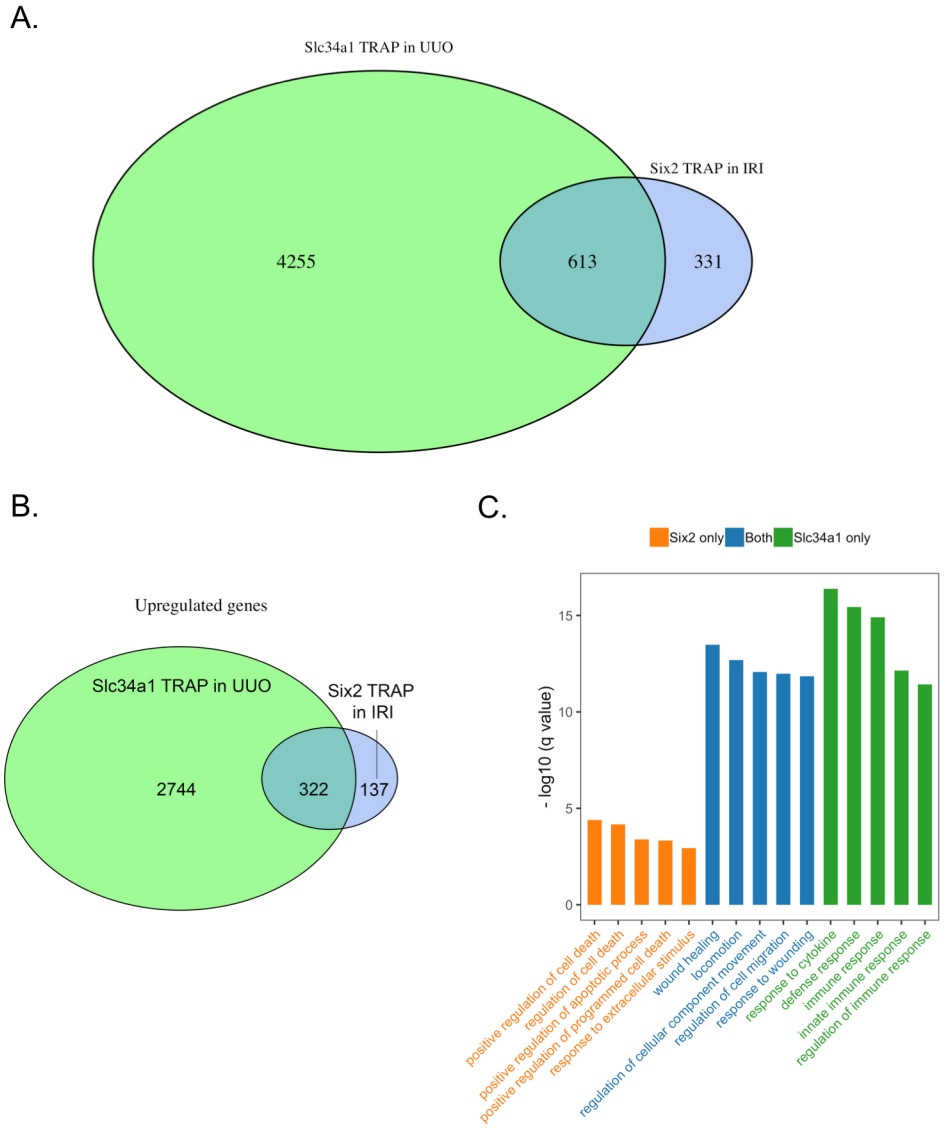
B.



Supplementary Figure S2. Sexually dimorphic genes in proximal tubule.
A. Linear relationship of log fold change cutoffs and the number of detected DEG. **B.** Volcano plot visualizing the sexually dimorphic genes that have been validated by ISH and the published scRNA-seq datasets.



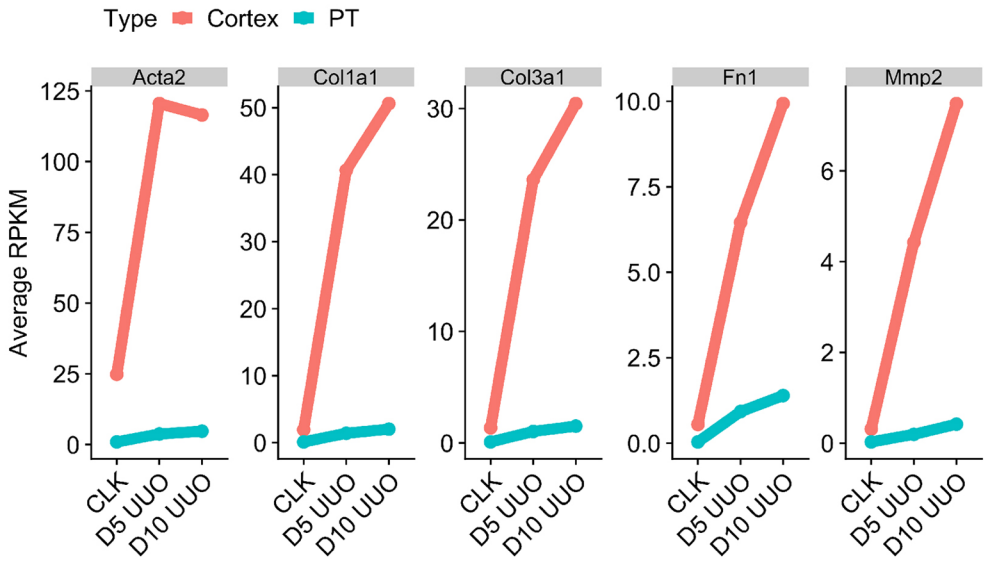
Supplementary Figure S3. Disease signature for kidney fibrosis.
A. Venn diagram showing ~80% of the DE genes identified in fibrotic PT are overlapped at day 5 and day 10 UUO. **B.** Very few differentially expressed genes are identified in the whole cortex from day 10 UUO and day 5 UUO kidneys. **C.** Venn diagram summarizing the overlap of the DE genes across pairwise comparisons.



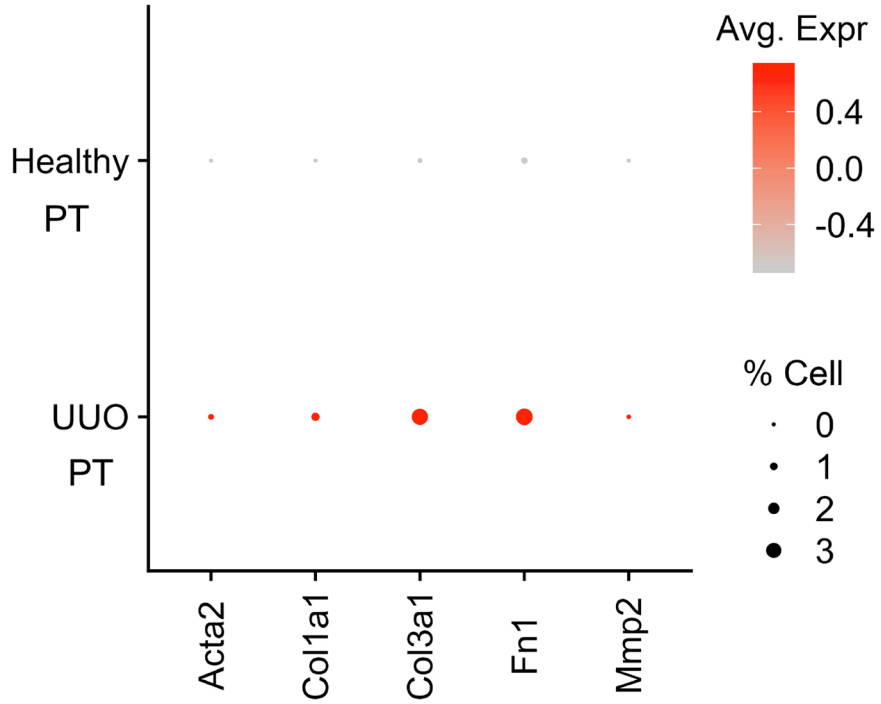
Supplementary Figure S4. Comparison of the DEG identified from this study and the published *Six2*-TRAP dataset

A. Venn diagram showing ~65% of the DEGs identified in this study are overlapped with DEGs reported by Liu *et al*[2] using a *Six2* TRAP line. **B**. Venn diagram showing the upregulated genes unique in each study or shared by both datasets. **C.** GO enrichment analysis on the DEGs unique to each dataset, or shared by both datasets.

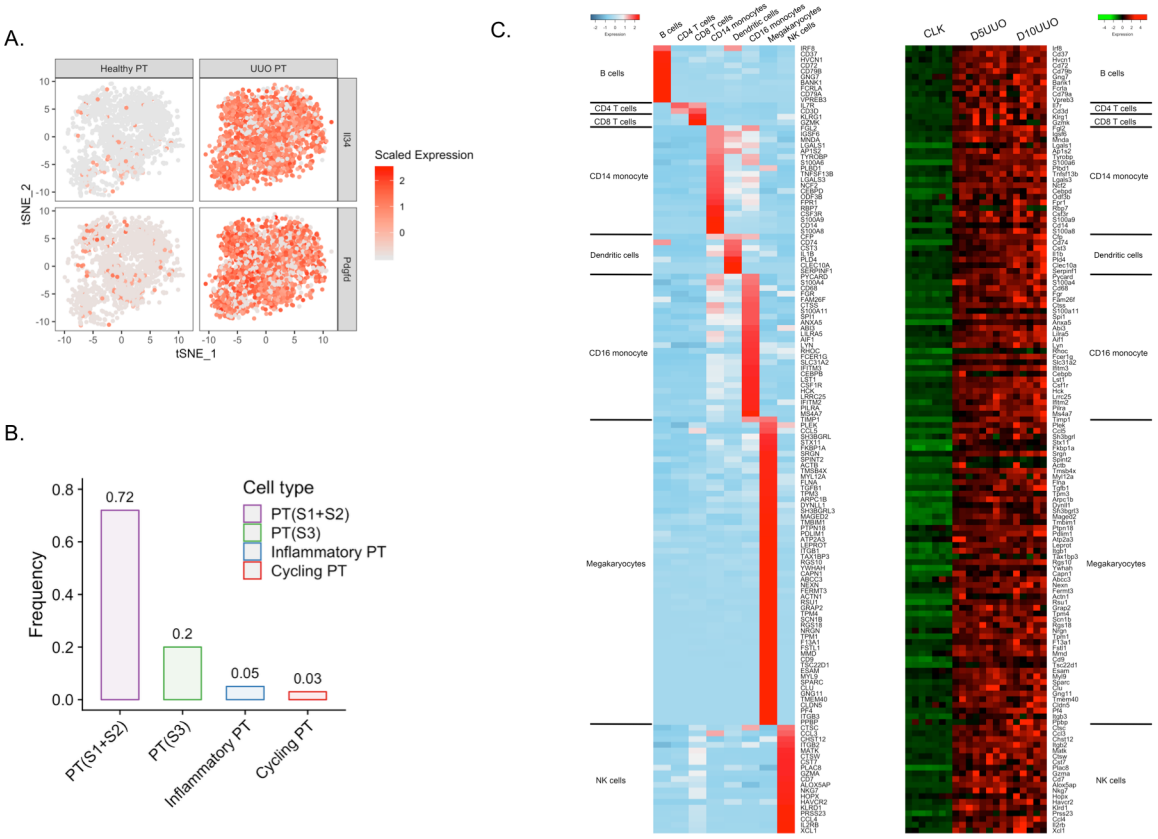
A.



B.

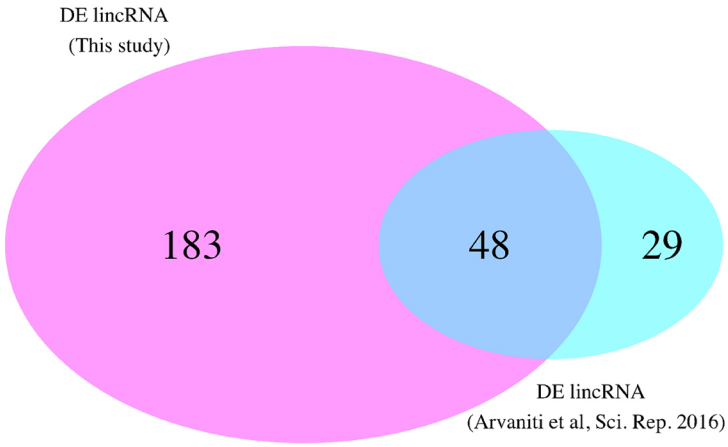


Supplementary Figure S5. EMT marker expression in TRAP-seq and scRNA-seq.
A. Expression of the EMT markers (RPKM value) are much higher in whole cortex than in PT.
B. scRNA-seq[3] revealed that less than 3% PT cells from the day 14 UUO kidney are expressing the EMT markers.

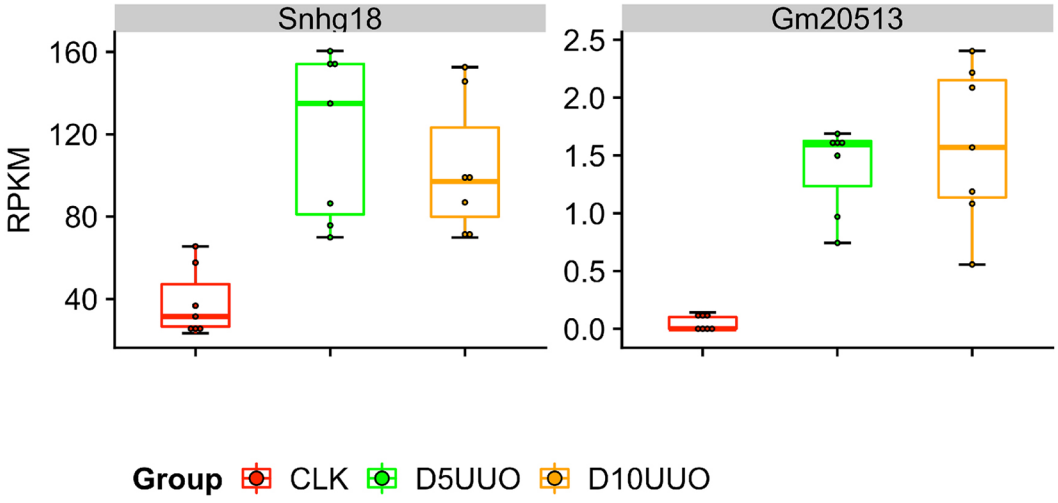


Supplementary Figure S6. Proximal tubule pro-inflammatory gene expression and subtype composition revealed by scRNA-seq.
A. Feature plot confirms the upregulation of Il34 and Pdgfd at day 14 UUO. **B.** Fraction of PT subtypes in the day 14 UUO kidney**. C.** Expression of immune cell markers in total cortex during UUO. Left: heatmap showing the marker expression in a publicly available PBMC dataset (https://satijalab.org/seurat/). Right: heatmap showing the expression of the same set of immune cell markers in our total cortex.

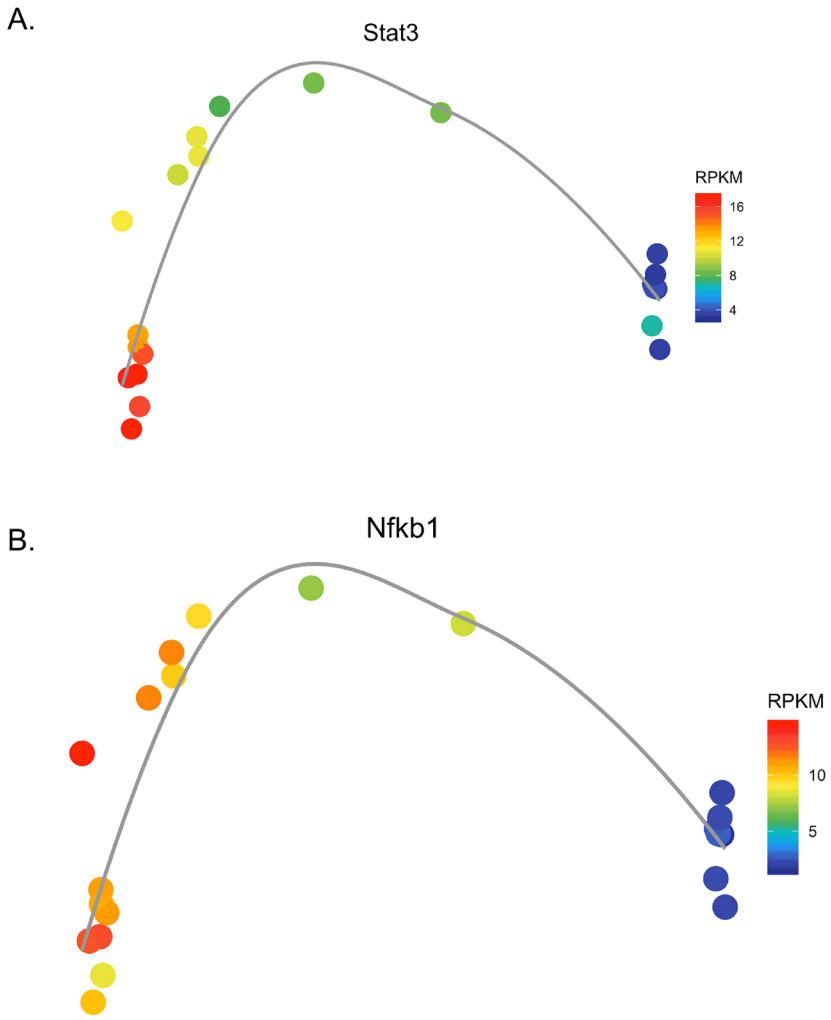
A.



B.



Supplementary Figure S7. lincRNA expression in UUO kidneys.
A. Venn diagram to compare the DE lincRNAs identified from this study (whole cortex) with a published dataset (whole kidney).[4] **B.** Box plot showing the RPKM value for the selected lincRNAs Snhg18 and Gm20513 in proximal tubule.



Supplementary Figure S8. Sample ordering validation by transcription factors known to be upregulated in proximal tubule during kidney fibrosis.
The accuracy of sample ordering can be validated by the expression of the known TFs such as Stat3 **(A)** and Nfkb1 **(B).**

Supplementary Methods

***In Situ* Hybridization**

Primers were as follows: *Cndp2*, SP6: 5'-CATTTAGGTGACACTATAGCTCCCTGCATGGGATCGAAG – 3'; T7: 5' – TAATACGACTCACTATAGGGTTCTCATTCTGCGAGTGGGC – 3'; *Hao2*, SP6; 5' - CATTTAGGTGACACTATAG GGCAGACTTTAAGGCACAAGC – 3'; T7: 5' – TAATACGACTCACTATAGGGCTCTCCAGGGGATCGGAGAT – 3'; *Cstb*, SP6; 5' – CATTTAGGTGACACTATAGGCCAGGTTTTTCTAGGGTCCA – 3'; T7; 5' – TAATACGACTCACTATAGGGCAAAGGAGCCCCGAATCAGA – 3'; *S100a10*, SP6; 5' - CATTTAGGTGACACTATAG GAGTGCTCATGGAACGGGAG – 3'; T7; 5' - TAATACGACTCACTATAGGG TTGAGGGCAATGGGATGCAAA – 3'; *Carhsp1*, SP6; 5' – CATTTAGGTGACACTATAGCTGGAGGAGTAGGACGTGTCG – 3'; T7: 5' – TAATACGACTCACTATAGGGGGTCTCATGCTTGGTCCCTG – 3'; *RhoC*: 5' – CATTTAGGTGACACTATAGATCGAAGTGGATGGCAAGCA – 3'; T7: 5' – TAATACGACTCACTATAGGGCTTGGGGCTGGGAAACTCAT – 3'; *Slc22a8*, SP6; 5' – CATTTAGGTGACACTATAGCTGGCTACAGTTGTCCGTGT – 3'; T7: 5' – TAATACGACTCACTATAGGGGACAGGCATCCCTTCCCAAA – 3'; *Snhg18*, SP6; 5' – CATTTAGGTGACACTATAGTCGCCTGGAAGAGACCTTCT – 3'; T7: 5' – TAATACGACTCACTATAGGGTAAAACCGAAGCAGCACCGA – 3'; *Gm20513*, SP6: 5' – CATTTAGGTGACACTATAGGAGGAGCAACCTTCCCTGGT – 3'; T7: 5' – TAATACGACTCACTATAGGGATGTTCCCCAAATCAACTGCAT – 3'. *In vitro* transcription was then performed to generate digoxigenin (DIG)-labeled sense and antisense riboprobes DIG labeling mixture (cat #11277073910, Roche) by SP6 and T7 RNA polymerases, respectively.

Bioinformatic analysis

1. Differential expression analysis

Genes with fewer than 0.3 RPKM (Reads Per Kilobase of transcript, per Million mapped reads) in one sixth of the samples were excluded from further analysis. The filtered gene-level counts were imported into the R/Bioconductor package edgeR.[5] Raw counts were normalized using the TMM (Trimmed Mean of M-values) technique. P values were reported using a negative binomial test in edgeR, followed by a Benjamini–Hochberg multiple testing correction to derive FDR. Only differentially expressed genes with FDR < 0.05 and absolute log2 fold change >1 were considered for downstream analysis. Performance of the samples was assessed with a Pearson correlation matrix and multi-dimensional scaling (MDS) plots. Unless otherwise noted, we used R packages heatmap.2, VennDiagram and ggplot2 to generate the heatmaps, venn diagrams and all other plots in this study.

2. Gene-set enrichment and pathway analysis

GSEA (http://software.broadinstitute.org/gsea/index.jsp) was used to estimate the enriched gene ontology (GO) terms from the differentially expressed (DE) genes. We used the DE gene list ranked by log2 fold change as a ranked list, and a gene sets database downloaded from Bader Lab (http://download.baderlab.org/EM_Genesets/) that contained all mouse GO terms as gene set file input to GSEA. We then used 1,000 gene label permutations to identify the significantly enriched gene-sets as defined by adjusted p value < 0.05. Pathway analysis results were obtained from the ToppGene suite[6] (https://toppgene.cchmc.org) using the DE genes derived from pairwise comparison as input.

3. Comparative analysis of the differential genes from TRAP-seq, microdissection RNA-seq and TRAP microarray

To compare the differential gene list produced from this study to the genes from the public cell-type specific profiling datasets, we re-analyzed a study reported by Lee *et al*[1] (GSE56743;

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56743) where each tubular segment was profiled using microdissection RNA-seq, and a study published by us using TRAP and microarray to profile the *Six2* cells in the ischemia-reperfusion kidney disease model[2] (GSE52004; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52004). To minimize the bias arising from different analytical pipelines and parameters chosen, we used the same pipeline and cutoffs to generate the differential gene lists. For the microdissection RNA-seq dataset, we removed the lowly expressed genes using the same cutoff reported in this study (i.e. RPKM>0.3 in at least 1/6 samples), followed by the *edgeR* package to produce the differential gene list. For the *Six2*-TRAP microarray data, since *edgeR* was not a suitable tool to analyze microarray data due to the underlying statistical model (negative binomial) but *limma* can analyze both microarray and RNA-seq data (if the normalized counts were transformed by *voom*),[7] we therefore re-analyzed our TRAP data using *limma*. In this analysis, 95% of differential genes generated by *edgeR* and *limma* were overlapped, confirming the similar performance of both pipelines on our TRAP data. We then used the same cutoffs to select the DEGs (log fold change >1 and FDR <0.05) for the comparison.

4. scRNA-seq re-analysis

To validate genes with sexually dimorphic expression, we re-analyzed three independent scRNA-seq datasets generated by three different laboratories.[3, 8, 9] The proximal tubule population was extracted from these datasets and the percentage of cells expressing the sexually dimorphic genes were compared between male and female. In total, we obtained PT single cell transcriptomes from 3 female and 10 male mice (Park et al[8]: 6 male; Wu et al[3]: 2 male + 2 female; and Schaum et al[9]: 2 male + 1 female). Selected genes were reported from the gene list that passed the significance threshold ($p<0.05$, student t-test). To confirm the existence of a proinflammatory PT state, we re-clustered the PT population obtained from a D14UUO snRNA-seq data using the Seurat package. PT subtype identities were annotated based the marker genes identified from each subtype. Pathway analysis was carried out on the proinflammatory PT subtype using the ToppGene suite mentioned above. We then applied the single cell deconvolution algorithm BSeq-sc[10] to estimate the proportion of each PT subtype identified from scRNA-seq in the D5 and D10 UUO PT bulk RNA-seq data. The marker genes for each PT subtype and the RPKM normalized gene expression matrix from TRAP were used as input according to the tutorial from BSeq-sc package (https://shenorrlab.github.io/bseqsc/vignettes/bseq-sc.html). To obtain the marker gene list for each immune cell types, we downloaded the precomputed Seurat R object from PBMC scRNA-seq in Satija's lab (https://satijalab.org/seurat/). We then used *FindAllMarker* function in Seurat to produce the marker gene list.

5. Cytokine-receptor interaction analysis

To study ligand-receptor interactions between PT and other cell types within the UUO kidney, we used a ligand–receptor list comprising 2,557 ligand–receptor pairs curated by the Database of Ligand−Receptor Partners (DLRP), IUPHAR and Human Plasma Membrane Receptome (HPMR).[11, 12] To determine the ligand-receptor pairs to plot on the heatmap, we required that (i) the ligands were significantly upregulated in the PT fraction and receptors were only enriched in the total cortex fraction (FDR<0.05 and |logFC|>1); (ii) Each receptor should have at least one corresponding ligand. To highlight paracrine interactions, we further reduced the gene list using the cytokine-receptor pairs documented in the iTALK package.[13]

6. Long noncoding RNA and G protein-coupled receptor (GPCR) analysis

We extracted lncRNAs from the DE gene list based on the categories in Ensembl NCBIM38 annotations. lncRNA-mRNA interaction was predicted by a publicly available database LncRRIdb (http://rtools.cbrc.jp/LncRRIsearch/). We used network visualization

software Cytoscape (version 3.4.0)[14] to visualize lncRNA-mRNA interactions. To compare the lncRNAs identified from this PT-TRAP line to those reported by a bulk transcriptomic study (14), we set the same cutoffs to select the differentially expressed lncRNAs (|logFC|>1 and FDR <0.05). The number of overlapped and non-overlapped lncRNAs were visualized by venn diagram. To catalog the GPCRs expression in PT, we obtained a mouse GPCR gene list curated by IUPHAR/BPS Guide to PHARMACOLOGY database (http://www.guidetopharmacology.org/GRAC/GPCRListForward?class=A), and crossed it to the whole gene list detected in our TRAP dataset. Differential gene test was performed on the detected GPCRs (UUO vs CLK), and all differentially expressed GPCRs were reported if they passed the statistical threshold (|logFC|>1 and FDR <0.05).

7. Transporters, ion channels and NFκB target gene analysis.

To categorize the genes that were enriched in PT, we obtained a complete mouse list of transporters and ion channels from the PHARMACOLOGY database (http://www.guidetopharmacology.org/targets.jsp). This gene list was crossed to the DEG list produced from the comparison of PT and total cortex. Using a similar approach, we generated the data for the NFκB target genes by crossing the complete NFκB target gene list curated by Thomas Gilmore's lab (http://www.bu.edu/nf-kb/gene-resources/target-genes/) to the DEG list from this study (PT UUO versus PT CLK).

8. Reconstruction of disease progression trajectory with Monocle

We used Monocle2[15] to draw a minimal spanning tree connecting the PT samples from CLK, D5UUO and D10UUO kidneys. As input into Monocle2, we selected the highly variable genes for sample ordering as described in the Monocle2 tutorial (http://cole-trapnell-lab.github.io/monocle-release/). We then reduced the data space to two dimensions using the reduceDimension function with 'DDRTree' method and ordered the samples using the orderCells function in Monocle2. Individual sample were color-coded based on the kidneys where they were collected. To identify the genes whose expression were dynamically changing across the trajectory from healthy to fibrotic kidneys, we performed differentially expression test on the samples across pseudotime using the differentialGeneTest function in Monocle2 with the fullModelFormulaStr parameter set to 'Psudeotime'. The output gene list was used for the downstream TF analysis.

9. Transcription factor and motif enrichment analysis

To identify the TFs, we cross referenced differentially expressed genes obtained from Monocle to the mouse TF list downloaded from AnimalTFDB[16] (http://bioinfo.life.hust.edu.cn/AnimalTFDB/). We selected important TFs that have been reported by the literature as key regulators of proximal tubule injury and visualized them by heatmap or line chart. We then used a mogrify algorithm to score the TFs. The top 25 upregulated and downregulated TFs were selected and visualized by bar chart. To identify the TF binding motif enriched in the 10kbp regions of transcription start site (TSS) for the DE genes, we implemented an R package *RcisTarget*[17] on the differential gene list with the following steps: 1) A precomputed mouse motif database with scores for each gene-motif pair was downloaded from Aerts Lab (https://resources.aertslab.org/cistarget/). In this analysis, we used a motif database named *mm9-tss-centered-10kb-7species.mc9nr.feather* which ranked the motifs enriched in 10kbp upstream and 10kbp downstream of the TSS. 2) A motif annotation database (*motifAnnotations_mgi*) was used to associate the motif to transcription factors. 3) Over-representation of each motif on the gene-set is estimated by computing the Area Under the Curve (AUC) for each pair of motif-geneset using the *calcAUC* function provided by the RcisTarget package. 4) The significantly enriched motifs with normalized enrichment score greater than 3 were reported and annotated to

transcription factors using the *RcisTarget* functions *addMotifAnnotation* and *addSignificantGenes.*

1. Lee, JW, Chou, CL, Knepper, MA: Deep Sequencing in Microdissected Renal Tubules Identifies Nephron Segment-Specific Transcriptomes. *J Am Soc Nephrol,* 26**:** 2669-2677, 2015.
2. Liu, J, Krautzberger, AM, Sui, SH, Hofmann, OM, Chen, Y, Baetscher, M, Grgic, I, Kumar, S, Humphreys, BD, Hide, WA, McMahon, AP: Cell-specific translational profiling in acute kidney injury. *J Clin Invest,* 124**:** 1242-1254, 2014.
3. Wu, H, Kirita, Y, Donnelly, EL, Humphreys, BD: Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *J Am Soc Nephrol,* 30**:** 23-32, 2019.
4. Arvaniti, E, Moulos, P, Vakrakou, A, Chatziantoniou, C, Chadjichristos, C, Kavvadas, P, Charonis, A, Politis, PK: Whole-transcriptome analysis of UUO mouse model of renal fibrosis reveals new molecular players in kidney diseases. *Sci Rep,* 6**:** 26235, 2016.
5. Robinson, MD, McCarthy, DJ, Smyth, GK: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26**:** 139-140, 2010.
6. Chen, J, Bardes, EE, Aronow, BJ, Jegga, AG: ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res,* 37**:** W305-311, 2009.
7. Ritchie, ME, Phipson, B, Wu, D, Hu, Y, Law, CW, Shi, W, Smyth, GK: limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res,* 43**:** e47, 2015.
8. Park, J, Shrestha, R, Qiu, C, Kondo, A, Huang, S, Werth, M, Li, M, Barasch, J, Susztak, K: Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science,* 360**:** 758-763, 2018.
9. Tabula Muris, C, Overall, c, Logistical, c, Organ, c, processing, Library, p, sequencing, Computational data, a, Cell type, a, Writing, g, Supplemental text writing, g, Principal, i: Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature,* 562**:** 367-372, 2018.
10. Baron, M, Veres, A, Wolock, SL, Faust, AL, Gaujoux, R, Vetere, A, Ryu, JH, Wagner, BK, Shen-Orr, SS, Klein, AM, Melton, DA, Yanai, I: A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst,* 3**:** 346-360 e344, 2016.
11. Ramilowski, JA, Goldberg, T, Harshbarger, J, Kloppmann, E, Lizio, M, Satagopam, VP, Itoh, M, Kawaji, H, Carninci, P, Rost, B, Forrest, AR: A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat Commun,* 6**:** 7866, 2015.
12. Wu, H, Malone, AF, Donnelly, EL, Kirita, Y, Uchimura, K, Ramakrishnan, SM, Gaut, JP, Humphreys, BD: Single-Cell Transcriptomics of a Human Kidney Allograft Biopsy Specimen Defines a Diverse Inflammatory Response. *J Am Soc Nephrol,* 29**:** 2069-2080, 2018.
13. Wang, Y, Wang, R, Zhang, S, Song, S, Jiang, C, Han, G, Wang, M, Ajani, J, Futreal, A, Wang, L: iTALK: an R Package to Characterize and Illustrate Intercellular Communication. *bioRxiv***:** 507871, 2019.
14. Shannon, P, Markiel, A, Ozier, O, Baliga, NS, Wang, JT, Ramage, D, Amin, N, Schwikowski, B, Ideker, T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res,* 13**:** 2498-2504, 2003.
15. Qiu, X, Mao, Q, Tang, Y, Wang, L, Chawla, R, Pliner, HA, Trapnell, C: Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods,* 14**:** 979-982, 2017.

16. Hu, H, Miao, YR, Jia, LH, Yu, QY, Zhang, Q, Guo, AY: AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res,* 47**:** D33-D38, 2019.
17. Aibar, S, Gonzalez-Blas, CB, Moerman, T, Huynh-Thu, VA, Imrichova, H, Hulselmans, G, Rambow, F, Marine, JC, Geurts, P, Aerts, J, van den Oord, J, Atak, ZK, Wouters, J, Aerts, S: SCENIC: single-cell regulatory network inference and clustering. *Nat Methods,* 14**:** 1083-1086, 2017.