Mapping the single-cell transcriptomic response of murine diabetic kidney disease to therapies

Haojia Wu
Washington University School of Medicine in St. Louis
Romer Gonzalez Villalobos
CVM Janssen Research & Development
Xiang Yao
Tox LJ Janssen Research & Development
Dermot Reilly
CVM Janssen Research & Development
Tao Chen
PSTS Janssen Research & Development

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/oa_4

Part of the Medicine and Health Sciences Commons

Recommended Citation
Wu, Haojia; Gonzalez Villalobos, Romer; Yao, Xiang; Reilly, Dermot; Chen, Tao; Rankin, Matthew; Myshkin, Eugene; Breyer, Matthew D.; and Humphreys, Benjamin D., "Mapping the single-cell transcriptomic response of murine diabetic kidney disease to therapies." Cell Metabolism. 34, 7. 1064 - 1078.e6. (2022). https://digitalcommons.wustl.edu/oa_4/1596

This Open Access Publication is brought to you for free and open access by the Open Access Publications at Digital Commons@Becker. It has been accepted for inclusion in 2020-Current year OA Pubs by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Mapping the single-cell transcriptomic response of murine diabetic kidney disease to therapies

Highlights
- A 1-million-cell atlas shows heterogeneity in kidney cell responses to DKD and treatments
- Mono- and combination therapies target different cells and induce non-overlapping changes
- Combination therapies are more effective in rescuing DKD-associated transcriptional changes
- SGLT2i signature reveals a role for modulation of alternative splicing by this drug class

Authors
Haojia Wu, Romer Gonzalez Villalobos, Xiang Yao, ..., Eugene Myshkin, Matthew D. Breyer, Benjamin D. Humphreys

Correspondence
humphreysbd@wustl.edu

In brief
Wu et al. generate a single-cell atlas of murine diabetic kidney disease and its treatment. They provide a useful resource for understanding the cell-specific kidney transcriptional responses to DKD and to five common treatment regimens.

Wu et al., 2022, Cell Metabolism 34, 1064–1078
July 5, 2022 © 2022 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cmet.2022.05.010
Diabetic kidney disease (DKD) occurs in ~40% of patients with diabetes and causes kidney failure, cardiovascular disease, and premature death. We analyzed the response of a murine DKD model to five treatment regimens using single-cell RNA sequencing (scRNA-seq). Our atlas of ~1 million cells revealed a heterogeneous response of all kidney cell types both to DKD and its treatment. Both monotherapy and combination therapies targeted differing cell types and induced distinct and non-overlapping transcriptional changes. The early effects of sodium-glucose cotransporter-2 inhibitors (SGLT2i) on the S1 segment of the proximal tubule suggest that this drug class induces fasting mimicry and hypoxia responses. Diabetes downregulated the spliceosome regulator serine/arginine-rich splicing factor 7 (Srsf7) in proximal tubule that was specifically rescued by SGLT2i.

In vitro proximal tubule knockdown of Srsf7 induced a pro-inflammatory phenotype, implicating alternative splicing as a driver of DKD and suggesting SGLT2i regulation of proximal tubule alternative splicing as a potential mechanism of action for this drug class.

INTRODUCTION

Diabetes mellitus is estimated to affect over 10% of the US population, and diabetic kidney disease (DKD) is the leading cause of end-stage kidney disease (ESKD) worldwide (Oshima et al., 2021). Despite recent treatment advances, only a limited reduction in rates of ESKD from DKD have been observed (Brenner et al., 2001; Heerspink et al., 2020; Lewis et al., 2001; Perkovic et al., 2019; Wanner et al., 2016). While sodium-glucose cotransporter-2 inhibitors (SGLT2i) have emerged as a promising drug class for the treatment of DKD, their mechanism of action remains unknown (DeFronzo et al., 2021; Packer, 2020; Sen and Heerspink, 2021). Understanding the renoprotective effects of SGLT2i will inform the development of future therapeutics. For these reasons, deciphering the cellular and molecular drivers of DKD, and how best to target the main drivers of poor outcomes therapeutically, is a top priority.

The pathogenesis of DKD is multifactorial, including endothelial dysfunction, glomerular hyperfiltration, hemodynamic effects, and inflammation. Proteinuria often accompanies these changes, leading to tubular damage, interstitial fibrosis, and glomerulosclerosis. Systemic hypertension occurs frequently in diabetics and worsens intraglomerular hypertension and kidney microvascular damage. Bulk transcriptomic studies have been informative in both experimental models and human DKD (Ban- sal et al., 2020; Fan et al., 2019; Komers et al., 2014; Woroniecka et al., 2011). These studies have revealed important insights but are also limited because signals from minority cell types are lost in the integrated dataset. scRNA-seq addresses this limitation, allowing construction of detailed cell atlases depicting cell type and state in health and disease. We recently applied scRNA-seq in early human diabetic nephropathy, revealing a distal nephron potassium secretory signature as well as a pro-angio-genic glomerular profile (Wilson et al., 2019). Given the multitude of cell types affected in DKD, we reasoned that scRNA-seq could provide further insight into the cellular and molecular drivers of DKD as well as cell-specific responses to different drug treatments, including SGLT2 inhibition.

In this study, we utilized the homozygous Lepr knockout (db/db) mouse model with uninephrectomy and renin-induced hypertension mouse model. We compared vehicle to angiotensin-converting enzyme inhibitor (ACEi), Rosiglitzone, the SGLT2i TA-1887 (Nomura et al., 2014), ACEi + Rosiglitzone, and ACEi + SGLT2i at two time points (2 days and 2 weeks). We generated single nucleus RNA-seq (snRNA-seq) datasets from all groups comprising nearly 1 million cells. Analysis of the resulting atlas revealed that the different medications affected strikingly different cell types. Combination therapy had the largest effects, but these effects...
were also largely non-overlapping. By focusing on the earliest transcriptional effects of SGLT2i on the S1 segment of the proximal tubule (PT), we infer that this drug class induces fastig mimicry and hypoxia responses. More specifically, diabetes downregulated the spliceosome regulator serum/arginine-rich splicing factor 7 (Srsf7) in PT, and this was specifically rescued by SGLT2i. In vitro PT knockdown of Srsf7 triggered differential expression of ~2,000 genes and induced a pro-inflammatory phenotype, suggesting a role for PT alternative splicing in driving DKD progression and of SGLT2i in rescuing it. This study highlights the cellular complexity of DKD and the effects of different classes of therapy. The results support the development of combination therapies, since different drug classes target different cell types in kidney.

RESULTS

Cellular heterogeneity in mouse kidneys

We induced diabetic DKD by adeno-associated virus delivery of renin (ReninAAV) to uninuphetomized, female db/db mice (Figure 1A). This model mimics human progressive DKD and has been validated for use in evaluation of several drug treatments (Harlan et al., 2015, 2018a, 2018b). Physiological measurements confirmed that the ReninAAV model developed more severe proteinuria, hypertension, and histopathological changes compared to the conventional db/db model (Figures 1B, 1C, and S1A). Our study design included seven groups: db/m (control), db/db (vehicle), db/db + lisinopril (ACEi), db/db + rosiglitazone (hypoglycemic drug, Rosi), db/db + JNJ-39933673 (SGLT2i) (Nomura et al., 2014), db/db + Lisinopril + Rosiglitazone (ACEi + Rosi), and db/db + Lisinopril + JNJ39933673 (ACEi + SGLT2i). Our histological data suggested that combination treatments achieve better outcomes in rescuing the kidney pathological changes compared to single treatments (Figure S1B). From the physiological data, we confirmed that after 2 weeks of treatment, mice that received ACEi either alone or in combination with either Rosi or SGLT2i and mice that received Rosi alone experienced significant reductions in urinary albumin to creatinine ratio (Figure 1D). Mice that received ACEi alone or in combination all had a significant reduction in blood pressure, whereas mice that received either Rosi or SGLT2i alone did not (Figure 1D). Finally, a significant reduction in serum glucose was achieved only in the Rosi + ACEi and SGLT2i treatment groups (Figure 1D). These results overall are consistent with expected treatment effects for each drug class.

We collected tissues at two early time points (2 days and 2 weeks post treatment) in order to identify the most proximate effects of these therapies and performed snRNA-seq on the kidneys of 70 mice from 14 groups using the 10x Chromium platform. We implemented several quality control steps during data processing, including the use of CellBender (Fleming et al., 2019) to remove ambient RNA contamination, DoubletFinder (McGinnis et al., 2019) to eliminate cell doublets, and Seurat data integration (Stuart et al., 2019) to correct for batch effect (STAR Methods). We obtained a total of 946,660 high-quality single cells that were classified into 18 major kidney cell types (Figure 2). These included the rare cell types such as 2,421 macular densa (MD) cells and 2,783 renin + cells from the juxtaglomerular apparatus (JGA) (Figure 1). Every experimental group, mouse, and time point contributed cells to every cluster, indicating the absence of dissociation bias across groups (Figures 2 and S2A–S2C). Each cell type uniquely expressed cell-defining marker genes (Figure S2D), also reflecting the high quality of our data after quality control. Overall, we detected an average of 1,167 unique genes and 2,105 transcripts per cell (Figures S2E–S2G).

We observed substantial heterogeneity amongst endothelial cells, immune cells, the thick ascending limb (TAL) of the loop of Henle, and fibroblasts (Figures 2 and S2H). In particular, we could identify six endothelial subtypes, including glomerular endothelial cells (GECs) (Ehd3+)(Patrakka et al., 2007)(Figures 2 and S2H), which were resolved by mapping the molecular signatures of the subtypes to those reported in a public endothelial cell single-cell atlas (Kalucka et al., 2020). Unexpectedly, a total of seven distinct subtypes were identified in TAL. These included three separate medullary TAL subclusters, three distinct cortical subclusters, and Nos1+ MD cells, all expressing unique markers (Figures 2 and S2I). Three fibroblast subclusters were identified, including an Acta2+ myofibroblast cluster and two fibroblast subtypes expressing Dkk2 and Col8a1 that have also been found in other organs such as the bladder (Muhl et al., 2020)(Figures 2 and S2J). Immune cells are relatively underrepresented in snRNA-seq datasets but despite this, we were able to identify T cells, macrophages, and dendritic cells (Figures 2 and S2K).

Ki nucle cell responses to DKD

To assess gene expression changes during progression of DKD in our model, we compared snRNA-seq profiles from control mice (db/m) to db/db and db/db + reninAAV at 2 days and 2 weeks. This analysis identified 2,422, 2,683, and 3,143 unique differentially expressed genes respectively (false discovery rate [FDR] < 0.05, log fold change > 0.5). All cell types exhibited gene expression changes, but the number of differentially expressed genes varied dramatically—from over 500 in parietal cells to fewer than 25 in JGA (Figure 3A). The cell types with the highest number of differentially expressed genes during disease progression included principal cells (PCs), TAL, parietal epithelial cells (PECs), PT, and endothelial cells (ECs). The reduced number of differentially expressed genes in JGA, MD, and immune cell clusters was smaller, possibly reflecting reduced statistical power in lower abundance cell types (cell fractions for JGA, MD, and immune cells are 0.29%, 0.26%, and 0.44%, respectively).

More than half of the differentially expressed genes were only detected in a single cell type, highlighting the cell-specific and heterogeneous response to diabetes (Figure 3B). In line with this, gene ontology (GO) terms enriched in each cell type also showed cell-type-specific patterns (Figure S3A). For example, the GO term “glomerular basement membrane development” was only significant in the podocyte, whereas the GO term for vasculature development was only enriched in EC. For cell-specific differentially expressed genes, about one-third were unique to day 2, one-third were unique to day 14, and one-third were
Figure 1. Experimental plan, histology and physiologic readouts

(A) Experimental scheme.
(B) Urinary albumin to creatinine ratio (UACR), systolic blood pressure (SBP), and glucose in control (db/m), db/db, or db/db + Renin-AAV at 2 days and 2 weeks.
(C) Periodic acid–Schiff (PAS) staining at 2 weeks across groups.
(D) Effect of treatment on UACR, SBP, and glucose at baseline and 2 weeks.

*p < 0.05; **p < 0.01 (paired t test, multiple comparisons with Benjamini-Hochberg correction).
common to both time points (Figure S3A). This heterogeneity of responses between the two time points provides the rationale for using treatment control groups at both time points. We identified the top two cell-specific up- and downregulated genes across the time course (Figures 3D and 3E). This analysis identified a number of genes that might serve as disease biomarkers. For example, the receptor tyrosine kinase EphA6 is specifically upregulated in podocytes, and its upregulation was also reported in the podocytes from other types of focal segmental glomerulosclerosis (Bukosza et al., 2020). AVPR2 was specifically downregulated in principle cells, suggesting a defect in water reabsorption in DKD.

Given cell-specific expression patterns for the majority of genes altered in DKD, we next asked whether chronic kidney disease (CKD)-associated common variants identified by genome-wide association studies (GWAS) also associate with specific kidney cell types. We used an approach similar to Li et al. (2020) to relate heritability from GWAS of estimated glomerular filtration rate (eGFR), proteinuria, and blood urea nitrogen (BUN) to the disease program in each cell type using MAGMA (STAR Methods). The analysis revealed that endothelial cells and collecting duct cells (PC and distal convoluted tubule [DCT]) are significantly associated with eGFR (Figure 3E). For CKD variants, PC were significantly associated. It is striking that both PC and EC are also in the top three cell types with the most differentially expressed genes (DEGs) in DKD. PC are not typically thought to drive the DKD phenotype, but we have previously reported a strong potassium secretory gene signature in this cell type in human DKD (Wilson et al., 2019). We compared these previous results to our mouse DKD atlas using the same informatic pipeline. This revealed consistent cell types between mouse and human but relatively few shared cell-specific DEGs (Figures S3C–S3E), likely reflecting the differing severity of disease in these two datasets, although mouse late distal tubule...
Figure 3. DKD disease genes corrected according to treatment and heritability enrichment for kidney disease traits in DKD
(A) Number of DEG induced during DKD according to cell type. Bar colors indicate different comparisons: light blue, db/db versus db/m; medium blue, AAV 2 days versus db/m; dark blue, AAV 2 weeks versus db/m.
(B) Number of genes detected in only one cell type versus detected in multiple cell types.
(C) Dotplot depicting top two cell-specific, upregulated genes during DKD.
(D) Same plot depicting top two downregulated genes during DKD.
(E) Gene hits from GWAS traits related to CKD, UACR, and eGFR were mapped to DKD cell states at 2 weeks. The dashed blue line denotes Bonferroni-adjusted significance (FDR = 0.05).
Figure 4. Differentially expressed genes normalized by treatment according to cell type and cell-cell communication between injured proximal tubule and other cell types.

(A) Percentage of DEG induced by diabetes rescued by treatment according to cell type.

(B) Heatmap of gene expression changes in PEC, TAL, PC, and ICA at week 2 in phosphate-buffered saline (PBS) (left) versus Rosi (right).

(legend continued on next page)
and collecting duct exhibited the highest magnitude of DKD-induced transcriptional changes, similar to our findings in human.

**Cell-specific responses to treatment of DKD**

We next examined the extent to which different therapeutic regimens rescued the gene expression changes induced by DKD in each cell type. Distinct cell types responded differently to each therapeutic regimen, both at day 2 and day 14. Transcriptional changes in some cell types occur early, but not late (for example, podocytes after ACEi + Rosi treatment), whereas other cell types have greater responses later, but not earlier (such as PEC and TAL after ACEi + Rosi treatment [Zenodo data]). We asked what percentage of genes altered by disease (up- or downregulated) were corrected by treatment both at day 2 and day 14. This revealed cell-type- and time-specific effects across therapeutic regimens (Figure 4A). For example, a higher fraction of disease genes were rescued at day 2 in podocytes, DCT, and connecting tubule (CNT) compared to day 14, whereas 14 days of treatment rescued more disease genes in TAL and DTL compared to day 2 (Figure 4A). Intriguingly, Rosi had maximal therapeutic effect on PEC, TAL, PC, and type A intercalated cells (ICAs) at week 2 treatment. However, the actual genes and pathways being targeted by Rosi treatment in PEC, TAL, PC, and ICA were quite distinct, and all of them were ameliorated by treatment (Figures 4B and S4A). In line with a highly heterogeneous response, the actual genes and pathways altered by the different therapies differ substantially with very few genes shared by all therapeutic regimens (Figure S4B). For example, principal cells exhibited many more unique gene changes than conserved changes across treatment regimens (Figures S4C and S4D). These results indicate a distinct and heterogeneous transcriptional response of kidney cell types both to the diabetic milieu and to different treatment regimens.

To study patterns of intercellular communication and changes therein during disease and its treatment, we compared interaction strengths inferred by CellChat (Jin et al., 2021) between all kidney cell types. DKD induced substantial changes in cell-cell communication between all pairs of kidney cell types (Figure S4E). We calculated an incoming and outgoing signaling strength score for each cell type and compared the degree to which those scores changed in DKD compared to control kidney. Cell types for which diabetes induced the largest change in intercellular communication included podocytes, PT, descending thin limbs of Henle’s loop (DTL), and TAL (Figure S4F). We then asked how each treatment regimen normalized the signaling score in podocytes, PT, and TAL. This analysis revealed that ACEi had the least effect in all three cell types, and SGLT2i and Rosi alone had similar effects (Figure S4G). For podocytes and Rosi, ACEi + Rosi did the best job of normalizing the signaling strength score, whereas for PT, SGLT2i alone was best (Figure S4G). This analysis specifically identified osteopontin (Spp1) secreted by injured PT signaling with podocytes and fibroblasts (Figure 4C). Spp1 has already been shown to play an important role in mediating albuminuria through modulation of podocyte signaling in models of DKD (Lorenzen et al., 2008), and serum osteopontin levels strongly predict DKD and all-cause mortality in patients with type 1 diabetes mellitus (Gordin et al., 2014). Osteopontin intercellular signaling was best suppressed by combination therapy, either ACEi + Rosi or ACEi + SGLT2i, and this correlated with reduced expression of the injury marker Havcr1, further emphasizing the beneficial effects of combination therapies (Figure 4D).

**Synergistic effect of combination drug treatments on injured PT**

The integrated analysis of all groups identified a subset of proximal tubular cells that adopted an injured state (we annotated as “inj.PT”), expressing typical injury PT markers such as Havcr1, Vcam1, and C3 (Figures 2 and 5A). Combination drug treatments, but not single drug treatments, significantly attenuated the aberrant up-regulation of the PT injury marker Havcr1 in DKD (Figures 5B and 5C). The number of inj.PT was significantly increased in our DKD models (Figure 5D; p = 0.038 for AAV2d versus db/m; p = 0.001 for AAV2w versus db/m, one-way ANOVA with post-hoc Tukey multiple test). We additionally performed bulk RNA-seq on our mouse samples, deconvoluted the fraction of inj.PT, and confirmed that combination therapy was most effective at reducing the proportion of these cells (Figure 5D). To generalize these results, we asked whether a similar population of injured PT is present in human DKD by deconvoluting cell type compositions from two public bulk RNA-seq datasets using the cell-type-specific molecular signatures identified from our dataset. In late stage DKD, we could detect a significant proportion of inj.PT in human DKD (Figures S5A and S5B).

To further characterize these drug responses, we compared the gene expression changes in the injured PT from the diabetic kidneys between the combo treatment groups (ACEi + Rosi and ACEi + SGLT2i) versus their corresponding single treatment groups (ACEi, Rosi, and SGLT2i). To quantify PT injury in an unbiased way, we applied a similar algorithm developed by Melms et al. (2021) to score each individual PT cell using the top 50 up-regulated genes from the comparison of vehicle treated versus control (db/m). This analysis revealed no drastic change in the injury score after single drug treatments, but substantial decrease in the scores after 2 weeks of combination therapy (Figure 5E). As a complementary analysis, we applied a computational algorithm (scId) (Boufea et al., 2020) to measure the fraction of cells after treatment mapping to the cells from healthy versus diabetic kidneys (STAR Methods). Consistent with the injury score analysis, a greater proportion of injured PT cells mapped to the cells from healthy kidney after either combination therapy (Figure 5F). In addition, we observed that the change of injured PT gene regulatory networks, as well as the dysregulated pathways such as fatty acid oxidation in diabetes, was also...
corrected by combination drug treatments to a greater degree compared to single drug treatments (Figures S5C–S5E). Together, these results emphasize that combination therapies have stronger transcriptional effects than monotherapy.

**Drug-induced glomerular responses**

The size of our atlas allowed a detailed analysis of glomerular cell types and responses in DKD and its treatment. We identified glomerular cell types based on a score developed using marker

---

**Figure 5. Superiority of combination therapies in ameliorating proximal tubule injury responses**

(A) Specific expression of injury genes Havcr1, Vcam1, and C3 in the injured PT subcluster.

(B) Expression of injured PT marker Havcr1 in disease and after week 2 treatments.

(C) Validation of the Havcr1 expression by immunofluorescence staining. Cells were co-stained with PT marker (LTL, white).

(D) Percentage of injured PT across the week two groups by single-cell quantification (upper) and bulk RNA-seq deconvolution. Each dot is an individual mouse.

(E) Density plot to show the injury scores across groups. Cells were scored by the top 50 upregulated genes from injured PT in comparing Renin-AAV versus db/m.

(F) Fraction of cells after treatment mapping to healthy or diseased state as measured by scID.
Figure 6. Glomerular communication networks in diabetes and its treatment

(A) Reclustering of the glomerular cells.

(B) Cell-specific marker genes define glomerular subtypes.

(C) Human glomerulus DKD genes mapped to mouse glomerulus cell types. The left heatmap shows bulk glomerulus expression in healthy versus DKD. The dot plot in the middle assigns those same genes to mouse glomerulus cell types. The dot plot on the right shows that some of these genes are normalized by therapy.

(D) Ligand-receptor scoring for glomerular endothelial cell signaling to podocytes.

(E) Bmp6 and its receptor Bmpr1a expression in glomerulus during diabetes and its treatment.
Figure 7. Effects of SGLT2i on the S1 versus the S2 segment
(A) Reclustering of the PCT into S1 and S2 segments.
(B) Density plot showing that Slc5a2 expression is limited to the S1 segment.
genes identified from a recent glomerulus scRNA-seq atlas (He et al., 2021) (STAR Methods). This allowed identification of podocyte, mesangial cells (MCs), GECs, and PEC (Figures 6A and 6B). For example, Ehd3, a regulator of capillary fenestration was strongly expressed in the GEC cluster (Figure 6B). Itga8 and Gata3, two putative MC genes, are strongly expressed in the MC cluster (Figure 6B). Diabetes induced substantial cell-type-specific transcriptional changes in glomerular cell types at both time points with fewer than 20% of these genes shared across cell types (Figures S6A and S6B). As observed before, drug treatments target non-overlapping genes across glomerular cell types (Figures S6C and S6D).

In recent work, human healthy and diabetic glomeruli underwent bulk transcriptional profiling by microdissection (Levin et al., 2020). We sought to assign the relevant cell type to a manually curated list of genes that were in common in both human and mouse DKD. For example, thymosin j4 (Tmsb4x) (Vasilopoulos et al., 2016) and the GTPase Rasl11a were both exclusively upregulated in podocytes (Figure 6C). Sorbs2, which is involved in the formation of stress fibers (Anekal et al., 2015) and has been associated with DKD by proteomic analysis (Nakatani et al., 2011), was upregulated only in MCs (Figure 6C). We could additionally assess the effect of different treatment regimens on expression of these human DKD-relevant genes. Once again, this revealed a heterogeneous response. For example, the upregulation of Tmsb4x and Rasl11a was attenuated by ACEi + Rosi or ACEi + SGLT2i treatment (Figure 6C), but not by monotherapy. We then used CellChat (Jin et al., 2021) to identify the top changes in intercellular glomerular communication in diabetes and after therapy (STAR Methods). During progression of diabetes, signals sent from GECs to podocytes were upregulated in DKD at both time points, and these were normalized only by combination therapies but not by any monotherapy (Figure S6E). We next asked which ligand-receptor pairs mediate GEC to podocyte signaling. This revealed that endothelial-derived Bmp6 signaling was strongly induced by diabetes, and this signaling was most strongly attenuated by Sglt2i alone or either combination therapy (Figure 6D). Unlike Bmp6, which was specifically upregulated in GECs by diabetes, its receptor Bmpr1a is constitutively expressed in podocytes (Figure 6E).

**Inferring a mechanism for SGLT2i by comparing effects on S1 versus S2 PT**

SGLT2i target the sodium–glucose cotransporter Slc5a2, which is exclusively expressed on the apical membrane of the S1 segment of the PT (Ghezzi et al., 2017). We reasoned that since the S1 and S2 segments are similar transcriptionally and adjacent to each other in the nephron, but since SGLT2i have no direct effect on the S2 segment, effects specific to the S1 segment would represent direct drug effects. In this way, we might dissociate direct drug effects from systemic effects such as glycosuria, hemodynamic changes, and local inflammation. We first confirmed that Slc5a2 expression was specific to the proximal convoluted tubule (containing both S1 and S2 segments) and not expressed in other segments (Figure S7A). We then re-clustered the proximal convoluted tubule into S1 and S2 segments and confirmed that Slc5a2 expression was exclusive to the S1 segment (Figures 7A and 7B). We next asked how each treatment, both alone and in combination, altered gene expression in the S1 versus S2 segments at 2 weeks. This revealed a striking difference in the number of unique genes (e.g., not shared with other therapies) for SGLT2i in the S1 versus 2 segments (67 versus 21, respectively; Figures 7C and 7D). This is consistent with the S1 segment as the direct target for SGLT2i inhibition.

We next performed GO analysis on the full set of DEGs in S1 versus S2. Terms related to starvation and hypoxia responses were specifically rescued by SGLT2i in the S1 segment, but not the S2 segment (Figures S7C and S7D). By contrast, terms related to fatty acid oxidation were rescued by treatment in both segments (Figures S7E and S7F). These results suggest that SGLT2i directly induce both starvation mimicry and hypoxia responses in the S1 segment, which has been one of several previously proposed mechanisms of action for this drug class (Sen and Heerspink, 2021) (Packer, 2020). Terms that were normalized in both segments, however, may reflect secondary systemic effects (such as lower glucose levels, glycosuria, and hemodynamic effects) of SGLT2i inhibition.

To generate hypotheses for how SGLT2i inhibition may induce protective starvation mimicry and hypoxia responses in the S1 segment, we inspected the individual DEGs in the S1 versus S2 dataset (Table S1). We reasoned that changes in gene expression at day 2 that are specific to the S1 segment might reflect the most proximate transcriptional effects of SGLT2i inhibition and could shed light on the molecular mechanism for their protective effects. There were 53 upregulated and 36 downregulated S1-specific genes in SGLT2i versus PBS at day 2. The second most highly upregulated gene is Srsf7, which encodes a member of the serine/arginine (SR) spliceosome protein family that regulates mRNA splicing. Srsf7 has been shown in mice to promote a juvenile transcriptome gene program by regulating the alternative splicing of genes that promote anabolism and inhibit senescence pathways (Kadota et al., 2020). Another SR family gene upregulated in the S1 by SGLT2 inhibition, Srsf5, is involved in the regulation of glucose metabolism and acetyl coenzyme A (acetyl-CoA) production (Chen et al., 2018). These observations suggest the possibility that one mechanism of action of SGLT2i may be to regulate alternative splicing to activate a protective metabolic switch.
We confirmed that only SGLT2i, and not ACEi or Rosi, rescued Srsf7 expression in the S1 segment (Figure 7D). We next sought to develop an in vitro cell culture model to study Srsf7 regulation of PT. We found that human primary renal PT cells (RPTEC) express Srsf7 and that exposure to glycated albumin downregulated Srsf7 mRNA in a dose-dependent fashion (Figure S7G). We then knocked down Srsf7 by small interfering RNA (siRNA) in RPTEC, achieving >80% knockdown of both mRNA and protein (Figures 7E and 7F). Bulk RNA-seq of scramble versus Srsf7 siRNA in RPTEC revealed dramatic changes in gene expression, with almost 2,000 genes changed (Figure 7G). Pathway analysis of these gene changes revealed upregulation of a variety of pro-inflammatory terms after Srsf7 knockdown, suggesting that Srsf7 normally promotes a healthy and anti-inflammatory phenotype in the S1 segment (Figures 7H and S7H). Importantly, a number of genes that were rescued by SGLT2i were also differentially expressed after Srsf7 knockdown, providing a mechanistic link between our model and the in vivo results (Figures 7I and 7J).

DISCUSSION

Success in identifying mechanisms of DKD and how different treatments are protective has been complicated by the large number of tissue compartments affected. While murine models of DKD have been useful, they also have been limited by the variable degrees to which they reflect the full spectrum of DKD phenotypes. scRNA-seq offers a powerful and unbiased way to survey kidney responses during the progression of DKD and how different therapies act to protect against progression. We combined scRNA-seq with a recent murine model of DKD that recapitulates the major clinical and histologic features of DKD, including hyperglycemia, hypertension, proteinuria, progressive loss of kidney function, and glomerulosclerosis (Harlan et al., 2018a). To our knowledge, this is the largest mouse kidney single-cell atlas to date. The data herein provide a useful resource for understanding the cell-specific kidney transcriptomic responses to DKD and to five common treatment regimens. The DEGs identified in this atlas of DKD and its treatment can be used to gain mechanistic insight as well as serving as potential biomarkers for diagnostics.

With the addition of SGLT2i for the treatment of DKD, the past 3 years have witnessed a revolution in therapy for this disease. However, it should not be ignored that while dapagliflozin and canagliflozin reduce ESKD risk by one-third, two-thirds of “at-risk” trial subjects on both ACEi and SGLT2i continue to progress to ESKD (Perkovic et al., 2019). This residual unmet need requires novel therapies. The present studies identify non-overlapping cell types and transcriptional changes induced by each treatment regimen. Furthermore, combination therapy of ACEi + rosiglitazone or ACEi + SGLT2i had the strongest normalizing effect, and this was not simply additive of either monotherapy alone. These non-overlapping transcriptional effects may reflect residual disease mechanisms not targeted by current therapies that might be exploited through future combination therapies to treat patients unresponsive to current standards of care.

There are a variety of ways that our single-cell atlas may be productively mined. As an example, the large size of our study allowed the construction of a detailed intercellular communication network within the glomerulus. This suggested GEC-to-podocyte BMP-6 signaling induced by DKD. A member of the transforming growth factor-β superfamily, BMP-6 exerts tissue and cell-specific effects. In liver, endothelial-derived BMP-6 acts in a paracrine fashion to activate hepatocyte hepcidin transcription and thereby maintain iron homeostasis (Canali et al., 2017). Global deletion of BMP-6 exacerbated renal fibrosis, suggesting that its induction in DKD may play a protective role (Dendooven et al., 2011).

A major challenge for cell-specific DEG analyses such as this one is distinguishing between gene expression that is causing versus a consequence of the DKD phenotype (Porcu et al., 2021). We attempted to address this by comparing the S1 and S2 segment responses to SGLT2i, reasoning that SGLT2i targets S1 and not S2, so early transcriptional changes may reflect the most proximate signaling events induced by this drug class. This analysis suggested that SGLT2i induce starvation mimicry and hypoxia responses (Packer, 2020). A hypothesis generated from this analysis is that SGLT2i accomplish their protective effects in part by regulating alternative splicing in the S1 segment. This is because SGLT2 inhibition specifically upregulated two splicing genes, Srsf5 and Srsf7, that have been previously shown to regulate inflammation and metabolism. Alternative splicing increases transcript diversity by generating multiple variant mRNAs from one gene, and ~95% of multieXonic genes undergo alternative splicing (Pan et al., 2008). Our finding that Srsf7 knockdown promotes a pro-inflammatory PT transcriptional state strongly supports a role for alternative splicing in both DKD pathogenesis and in mediating the protective effect of SGLT2 inhibition. Lending support to this hypothesis, insulin itself regulates alternative splicing in some contexts (Chalfant et al., 1995), and both obesity and the ZSF1 rat model of DKD are associated with alternative splicing (Pihlajamaki et al., 2011; Zhang et al., 2018).

This resource should be of value to the DKD research community. We have generated a data visualization tool that can be accessed online (http://humphreyslab.com/SingleCell/). In addition, all codes and the R object can be downloaded from Github and GEO. In conclusion, our single-cell atlas of DKD and its treatment should allow detailed exploration of the mechanisms underlying disease progression and yield a deeper understanding of how current therapies act to slow progression of DKD.

Limitations of study

The cell-specific transcriptional responses to this model of DKD and its treatment may differ between different strains and ages of mice. Furthermore, our single-cell atlas of DEGs cannot distinguish between gene transcription that is causal versus a consequence of DKD. A recent report suggests that SGLT2 inhibition may attenuate mitochondrial dysfunction in PT (Wu et al., 2021), but we cannot directly assess this because our nuclear preparations lack mitochondrial RNA. Finally, through single nucleus dissociation, contextual information from neighboring cells and tissue microenvironment is lost. This complicated our ability to infer intercellular communication networks. Future spatially resolved transcriptomic studies will be required to complement our single-cell atlas of DKD and its treatment.
STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mouse model of diabetic nephropathy and drug treatments
- METHOD DETAILS
  - Nuclei isolation from mouse kidney
  - Human primary tubular epithelial cell culture and treatment
  - SRSF7 siRNA transfection
- REAL-TIME PCR
  - Western blot analysis
  - Immunofluorescence
  - Single nucleus RNA-seq
  - Ambient RNA and doublets removal, batch effect correction, and cell clustering
  - snRNA-seq differential gene analysis
  - Pathway and gene ontology analysis
  - Bulk RNA-seq and data analysis
  - Cell type deconvolution of bulk RNA-seq data
  - Measurement of the drug effect on injured PT
  - GRN inference and metabolic activity analysis
  - Association of cell type with GWAS traits by integrative analysis
  - Cell–cell interaction network analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cmet.2022.05.010.

ACKNOWLEDGMENTS

This study was supported by a sponsored research agreement from Janssen Pharmaceuticals, by the Washington University Diabetes Research Center NIH DK20579 and by NIH DK103740.

AUTHOR CONTRIBUTIONS


DECLARATION OF INTERESTS


# STAR METHODS

## KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRSF7 Polyclonal Antibody</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# PA5-39482; RRID:AB_2556056</td>
</tr>
<tr>
<td>HAVCR1 Antibody</td>
<td>R &amp; D Systems</td>
<td>Cat# AF1817; RRID: AB_2116446</td>
</tr>
<tr>
<td>Streptavidin, Alexa Fluor 647 conjugate</td>
<td>Invitrogen</td>
<td>Cat# S21374</td>
</tr>
<tr>
<td>SPP1 antibody</td>
<td>Abcam</td>
<td>Cat# ab218237; RRID: AB_2732079</td>
</tr>
<tr>
<td>Tubulin antibody</td>
<td>Proteintech</td>
<td>Cat# 66240; RRID: AB_2881629</td>
</tr>
<tr>
<td>Donkey Anti Rabbit, Alexa Fluor 488 conjugate</td>
<td>Invitrogen</td>
<td>Cat# A21206</td>
</tr>
<tr>
<td>Donkey Anti Goat, Alexa Fluor 568 conjugate</td>
<td>Invitrogen</td>
<td>Cat# A11057</td>
</tr>
<tr>
<td>Goat anti-rabbit HRP</td>
<td>Dako</td>
<td>Cat# P0448</td>
</tr>
<tr>
<td>Goat anti-mouse HRP</td>
<td>Dako</td>
<td>Cat# P0447</td>
</tr>
<tr>
<td><strong>Chemicals, peptides, and recombinant proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1x DPBS</td>
<td>Gibco</td>
<td>Cat# 14190144</td>
</tr>
<tr>
<td>Nuclei EZ Lysis Buffer</td>
<td>Sigma</td>
<td>Cat# NUC101</td>
</tr>
<tr>
<td>Complete ULTRA Tablets, Mini, EDTAfree, EASYpack</td>
<td>Roche</td>
<td>Cat# 05 892 791 001</td>
</tr>
<tr>
<td>RNasin Plus Ribonuclease Inhibitors</td>
<td>Promega</td>
<td>Cat# N2615</td>
</tr>
<tr>
<td>SUPERaseIN RNase Inhibitor</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# AM2696</td>
</tr>
<tr>
<td>RNaseZap</td>
<td>Ambion</td>
<td>Cat# AM9780</td>
</tr>
<tr>
<td>RNase free H2O</td>
<td>Thermo Scientific</td>
<td>Cat# AM9938</td>
</tr>
<tr>
<td>Primary RPTEC</td>
<td>Lonza</td>
<td>Cat# CC-2553</td>
</tr>
<tr>
<td>REGM BulletKit</td>
<td>Lonza</td>
<td>Cat# CC-3190</td>
</tr>
<tr>
<td>Glycated albumin</td>
<td>Sigma</td>
<td>Cat# A8301</td>
</tr>
<tr>
<td>Periodic Acid-Schiff (PAS) Kit</td>
<td>Sigma</td>
<td>Cat# 3958</td>
</tr>
<tr>
<td>siGENOME Human SRSF7</td>
<td>Horizon Discovery</td>
<td>Cat# M-015909-01-0020</td>
</tr>
<tr>
<td>siGENOME Non-Targeting siRNA</td>
<td>Horizon Discovery</td>
<td>Cat# D-001210-05-20</td>
</tr>
<tr>
<td>Lipofectamine™ RNAiMAX</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# 13778075</td>
</tr>
<tr>
<td>Biotinylated Lotus Tetragonolobus Lectin (LTL)</td>
<td>Vector Laboratories</td>
<td>Cat# B-1325</td>
</tr>
<tr>
<td>BCA Protein Assay Kit</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# 23252</td>
</tr>
<tr>
<td><strong>Critical commercial assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromium Single Cell 3’ GEM, Library &amp; Gel Bead Kit v3</td>
<td>10x Genomics</td>
<td>Cat# PN-1000075</td>
</tr>
<tr>
<td>Chromium Single Cell B Chip Kit</td>
<td>10x Genomics</td>
<td>Cat# PN-1000153</td>
</tr>
<tr>
<td><strong>Deposited data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw and analyzed snRNA-seq data</td>
<td>This study</td>
<td>GSE184652</td>
</tr>
<tr>
<td>Raw and analyzed bulk RNA-seq data</td>
<td>This study</td>
<td>GSE199437</td>
</tr>
<tr>
<td>Data S1 – Source Data</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>Zenodo data</td>
<td>This study</td>
<td><a href="https://zenodo.org/record/6344528">https://zenodo.org/record/6344528</a></td>
</tr>
<tr>
<td>Nuclei isolation protocol</td>
<td>Protocols.io</td>
<td><a href="https://dx.doi.org/10.17504/protocols.io.nahdab6">https://dx.doi.org/10.17504/protocols.io.nahdab6</a></td>
</tr>
<tr>
<td>Kidney Interactive Transcriptomics</td>
<td>The Humphreys Lab</td>
<td><a href="http://humphreyslab.com/SingleCell/">http://humphreyslab.com/SingleCell/</a></td>
</tr>
<tr>
<td><strong>Experimental models: Organisms/strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BKS.Cg-Dock7m +/+ - Leprdb/J Leprdb- heterozygous</td>
<td>The Jackson Lab</td>
<td>Lot 000642</td>
</tr>
<tr>
<td>BKS.Cg-Dock7m +/- Leprdb/J Leprdb- homozygous</td>
<td>The Jackson Lab</td>
<td>Lot 000642</td>
</tr>
</tbody>
</table>

(Continued on next page)
RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Benjamin D. Humphreys (humphreysbd@wustl.edu).

Materials availability
This study did not generate new unique reagents.

Data and code availability
- The accession number for the RNA sequencing data reported in this paper is NCBI GEO: GSE184652 for snRNA-seq and GSE199437 for bulk RNA-seq. A searchable database, including gene expression in all kidney cell types, glomerular cell types, and PCT cell types is available at our Kidney Interactive Transcriptomics (K.I.T.) Website: http://humphreyslab.com/SingleCell/.
- DEG lists were deposited in Zenodo (https://zenodo.org/record/6344528).
- Original codes for data analysis were deposited on GitHub at https://github.com/TheHumphreysLab/Mouse_DKD_Rx_Atlas. An R package to visualize the single cell RNA-seq data with complex group design is available on GitHub (https://github.com/TheHumphreysLab/plot1cell).
- All values used to generate the graphs of the paper and original western blot images can be found in the file Data S1 - Source Data. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse model of diabetic nephropathy and drug treatments
The in vivo experiments (conducted at Plato Biopharma, Inc. Fort Collings, CO) were performed in the ReninAAV db/db mouse after uninephrectomy (Unx), a murine model of severe progressive DKD. As in the human condition, mice subject to this model display
severe renal dysfunction characterized by increased urinary albumin excretion, reduced GFR, increased sCr, BUN, and hypertension. Further, the kidneys of these mice display prominent features of DKD including mesangial expansion, glomerular sclerosis, tubular degeneration as well as tubulointerstitial inflammation and fibrosis (Harlan et al., 2015; 2018b).

Mice were maintained and treated in accordance with the guidelines approved by the Association for Assessment and Accreditation of Laboratory Animal Care. Mice were given ad libitum access to Formulab Diet 5008 (LabDiet, Cat. #: 0,001,325) and water, housed under standard conditions, and allowed to acclimatize for at least one week prior to any procedures. To establish the model, female db/db mice (BKS.Cg-Dock7m +/+ Leprdb/J from Jackson, BKS background, Lot 642) were subject to left nephrectomy at 5 weeks of age and, at 12 weeks, infected with a renin expressing adeno-associated virus (ReninAAV, 109 genomic copies) to induce hypertension and accelerate disease progression. For weeks later, a battery of tests (UACR, sCR, BUN, blood glucose, body weight, SBP, and HR) was conducted to randomize the mice into several treatment as follows:

1. Vehicle (drinking water)
2. Lisinopril (An ACE inhibitor, 10 mg/kg/d in the drinking water),
3. Rosiglitazone (A PPARγ activator, 0.007% w/w via admix diet, p.o. ~10 mg/kg/d)
4. JNJ-39933673 (A SGLT2 inhibitor, 0.0075% w/w via admix diet, p.o. 1 ~ mg/kg/d)
5. ACEi plus SGLT2i (dosed as before)
6. ACEi plus rosiglitazone (dosed as before)

In addition to the ReninAAV Unx db/db mice, db/m and Unx db/db mice infected with an empty viral vector (LacZ AAV) were included as healthy and diabetic non-hypertensive controls, respectively.

Terminal samples were collected 48 h and 2 weeks after treatment inception (n = 5/group). For this, mice were placed in metabolic cages to collect urines and later, under general anesthesia with inhalational isoflurane, mice were exsanguinated via cardiac puncture and the kidneys harvested. The right kidneys were cut transversally in roughly three equal size sections. The top section was fixed in 10% neutral buffered formalin and further processed for histopathology analysis. The mid-section was cut in half through the papilla, and the kidneys harvested. The right kidneys were cut transversally in roughly three equal size sections. The top section was fixed in 10% neutral buffered formalin and further processed for histopathology analysis. The mid-section was cut in half through the papilla, one-half was snap-frozen, stored at ~80°C and transferred to Washington University in St. Louis for nuclear transcriptomic analysis. The remaining mid- and bottom sections were stored in RNA later.

The SBP and HR were measured at baseline and after 2 weeks of treatment via tail-cuff plethysmography using a CODA High Throughput Noninvasive Blood Pressure System (Kent Scientific; Cat. #: CODA-HT8). Plasma and urine samples underwent standard clinical chemistry analysis on an Olympus AU400e Clinical Chemistry Analyzer (Beckman Coulter, Inc.) for the determination of Laboratory Animal Care. Mice were given ad libitum access to Formulab Diet 5008 (LabDiet, Cat. #: 0,001,325) and water, housed under standard conditions, and allowed to acclimatize for at least one week prior to any procedures. To establish the model, female db/db mice (BKS.Cg-Dock7m +/+ Leprdb/J from Jackson, BKS background, Lot 642) were subject to left nephrectomy at 5 weeks of age and, at 12 weeks, infected with a renin expressing adeno-associated virus (ReninAAV, 109 genomic copies) to induce hypertension and accelerate disease progression. For weeks later, a battery of tests (UACR, sCR, BUN, blood glucose, body weight, SBP, and HR) was conducted to randomize the mice into several treatment as follows:

1. Vehicle (drinking water)
2. Lisinopril (An ACE inhibitor, 10 mg/kg/d in the drinking water),
3. Rosiglitazone (A PPARγ activator, 0.007% w/w via admix diet, p.o. ~10 mg/kg/d)
4. JNJ-39933673 (A SGLT2 inhibitor, 0.0075% w/w via admix diet, p.o. 1 ~ mg/kg/d)
5. ACEi plus SGLT2i (dosed as before)
6. ACEi plus rosiglitazone (dosed as before)

METHOD DETAILS

Nuclei isolation from mouse kidney

Single nuclei were isolated using a protocol based upon (Habib et al., 2017) with adaptions for adult mouse kidney including: tissue mincing, homogenization strokes, addition of protease inhibitor and RNasin and adjustments to strainer size and sequence. Mouse kidney were cut into <2 mm pieces and homogenized using a Dounce homogenizer (Kimble Chase #885302-0002) in 2mL of ice-cold Nuclei EZ Lysis buffer (Sigma-Aldrich NUC101) supplemented with protease inhibitor (Roche #5892791001) and RNase inhibitor (Promega #N2615, Life Technologies #AM2696). The samples were then incubated on ice for 5 min with an additional 2mL of lysis buffer. The homogenate was filtered through a 40-μm cell strainer (pluriSelect #43-50040-51) and then centrifuged at 500 x g for 5 min at four°C. The pellet was resuspended and washed with 4 mL of the buffer and incubated on ice for 5 min. After another centrifugation, the pellet was resuspended with Nuclei Suspension Buffer (1x PBS, 0.07% BSA, 0.1% RNase inhibitor), filtered through a 5-μm cell strainer (pluriSelect 43-50,005-01) and counted using a disposable hemecytometer (InCYTO, DHC-F01-5). A full step-by-step protocol was submitted to protocols.io (dx.doi.org/10.17504/protocols.io.nahdabad).

Human primary tubular epithelial cell culture and treatment

Primary human renal proximal tubular cells (RPTECs) were purchased from Lonza (CC-2553) and cultured with Renal Epithelium Cell Growth Medium (CC-3190) supplemented with 10 ng/mL EGF, 5% v/v fetal calf serum, provided with the medium kit. Cells were maintained in a humidified 5% CO2 atmosphere at 37°C. All experiments were carried out on passage three of the cells. Cells were treated with glycated albumin or siRNA after a 24-h serum starvation.
SRSF7 siRNA transfection

hRPTECs were grown to 50–60% confluence and transfected with 100 and 200 nmol/L SRSF7 siRNA (siGenome) or negative control siRNA (siGenome) using Lipofectamine RNAiMAX (Life Technologies) following the manufacturer’s protocol. Cells were harvested at day 1 and day 3 after transfection for protein and RNA isolation in order to validate knockdown. Day 3 mRNA samples were submitted for bulk RNA-seq.

REAL-TIME PCR

RNA was extracted using the RNEASY Mini prep kit (Qiagen) following the manufacturer’s instructions. The extracted RNA (600 ng) was reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit (Life Technologies). qPCR (RT-PCR) was performed using iTaq Universal SYBR Green Supermix (Bio-Rad). Expression levels were normalized to GAPDH and data analyzed using the 2^(-ΔΔCt) method. Cycling conditions were 95°C for 3 min then 40 cycles of 95°C for 15 s and 60°C for 1 min, followed by one cycle of 95°C for 10 s. Primers used in this study were: SRSF7 forward: 5'-GAAGAAGCAGGTCACGTCT-3', SRSF7 reverse: 5'-CGACGGGAGTTGGAATACCT-3'; GAPDH forward: 5'-GACAGTCAGCCGCATCTTCT-3', GAPDH reverse: 5'-GCGCCCAATACGACCAAATC-3'.

Western blot analysis

RPTEC cell lysates were lysed with RIP-A lysis buffer containing protease inhibitor cocktails (Roche) on ice. Protein was collected after centrifugation at 12,000g and concentration was determined by the BCA method (Thermo Fisher). Twenty micrograms of total protein was separated by SDS electrophoresis and transferred to an Immobilon PVDF membrane (Millipore). Membrane was blocked with 5% milk in TBST and probed overnight at 4°C with the primary antibody. After washing the membrane with TBST, it was incubated for 1 h at room temperature with HRP-conjugated secondary antibody (Dako). The primary antibodies used in this assay were: anti-SRSF7 from Abcam and anti-Tubulin from Proteintech. The membrane was incubated by secondary antibodies (Dako) and developed using the ECL Detection System (GE Healthcare). Image was captured by ChemiDoc imaging system (Bio-Rad).

Immunofluorescence

Formalin-fixed kidney blocks were microtome sectioned at a thickness of 5 μm and mounted on Superfrost slides (Thermo Fisher Scientific). Sections were deparaffinized by xylene and rehydrated by series of ethanol in decreasing concentration. Immunofluorescent staining was performed as follows: kidney sections were washed with 1× PBS for 5 min and permeabilized with 0.25% Triton X for 15 min. Blocking was done with 5%BSA in PBS for 1 h. Primary antibodies were incubated for overnight at cold room and sections rinsed with 1× PBS for 5 min × 3. Secondary antibodies were incubated for 1 h at room temperature and rinsed with 1× PBS for 5 min × 3. DAPI was used for counterstaining. The following primary antibodies were used: anti-SPP1 from Abcam and anti-HAVCR1 from R&D Systems. Images were obtained by confocal microscopy (Nikon C2+ Eclipse).

Single nucleus RNA-seq

Droplet-based single-cell RNA sequencing was performed using the 10x Genomics Chromium Single Cell Kit v3. Single nucleus suspensions at a concentration of 1000 nuclei/ul were loaded into the Chromium Controller as per manufacturer instructions, which target 10,000 nuclei per sample. 12 PCR cycles were used for cDNA synthesis. Each library was sequenced using an Illumina NovaSeq instrument, eight samples per S4 lane, with one 150 bp read located near the 3’ end of the mRNA. Illumina runs were demultiplexed and the resulting fastq files were processed by the 10x Genomics cellranger pipeline (version 6.0.1). Both exonic and intronic reads were incorporated in the final molecule counts.

Ambient RNA and doublets removal, batch effect correction, and cell clustering

We used CellBender to remove the ambient RNA contamination (Fleming et al., 2019). In brief, the unfiltered count matrices sampleID.h5 from cellranger were input into CellBender. CellBender remove-background was run (on cuda mode) with parameters “expected-cells” and “total-droplets-included” were chosen for each dataset based on the total UMI per cell vs. cell barcode curve in accordance with CellBender documentation. Other parameters were left at their default values. The CellBender corrected count matrices were then input to DoubletFinder (McGinnis et al., 2019) to remove the potential doublets. We used default parameters to run DoubletFinder except for the nExp, which is estimated based on the total cell number of the sample (nNuc) and a linear regression equation derived from the Demuxlet study (Kang et al., 2018), i.e. nExp = nNuc * (0.018 + 8.4 × 10^-6 * nNuc). The ambient RNA and doublet-free count matrices were input into Seurat for clustering analysis. Batch effect was corrected by using the integration pipeline from Seurat. Due to the large size of our dataset, we used a reference-based approach to make the analysis less computationally intensive (https://satijalab.org/seurat/articles/integration_large_datasets.html). One sample from the healthy group (sample A3020) was used as the ‘reference’ with the remainder designated as ‘query’ datasets. After data integration, the combined data proceeded with joint analysis such as principal component analysis (PCA), clustering and 2D projection by UMAP. We annotated the cell clusters based on their marker gene expression. Cell clusters with confused marker gene expression of multiple cell types were manually removed. We also excluded the tAL cell type from downstream analysis because our initial analysis indicated that the number of cells in tAL varies dramatically across the 14 groups ranging from 12 cells in AAV2v to 1,468 cells in db/m, which greatly skewed the differential gene analysis. Alternative ‘reference’ sample (sample E3019) and data integration method (Harmony) were used to validate the clustering results.
snRNA-seq differential gene analysis
Cell type specific markers were identified from the whole dataset after clustering using Seurat function FindAllMarkers, with significant cutoff FDR set to 0.05, and log fold change set to 0.25. Cell class were annotated based on the well-known kidney cell markers. To find cell type specific disease genes for DN, we compared the gene expression between db/db and db/m, AAV 2days and db/m, and AAV 2w and db/m using the Seurat function FindMarkers, with significant cutoff FDR set to 0.05, and log fold change set to 0.5. The DE genes were categorized by cell type where they were detected. To identify the disease genes that were rescued by each drug treatment, we compared the gene expression between treatment versus DN and cross the gene list to the disease gene list. The overlapped DE genes in opposite directions are considered the genes that were rescued by treatment.

Pathway and gene ontology analysis
Pathway analysis of the differentially expressed genes were performed using fgsea R package (Korotkevich et al., 2021). Multiple pathway databases were used for this analysis, including MSigDB, Hallmark, and Reactome. The returning enrichment scores and p values were employed to compare the pathway activity across cell types and drug treatments. For gene ontology analysis, topGO package (Alexa et al., 2006) was used for testing the gene-sets related to biological processes on the DE gene lists generated by various comparisons. GO terms with greater than 500 genes or less than five genes were excluded from the analysis.

Bulk RNA-seq and data analysis
Total RNA integrity was determined using Agilent 4200 Tapestation. Library preparation was performed with 500ng to 1ug of total RNA. rRNA was removed by an RNase-H method using RiboErase kits (Kapa Biosystems). mRNA was then fragmented in reverse transcriptase buffer and heating to 94°C for 8 min mRNA was reverse transcribed to yield cDNA using Super-Script III RT enzyme (Life Technologies, per manufacturer’s instructions) and random hexamers. A second strand reaction was performed to yield ds-cDNA. cDNA was blunt ended, had an A base added to the 3’ ends, and then had Illumina sequencing adapters ligated to the ends. Ligated fragments were then amplified for 15 cycles using primers incorporating unique dual index tags. Fragments were sequenced on an Illumina NovaSeq-6000 using paired end reads extending 150 bases. Basecalls and demultiplexing were performed with Illumina’s bcl2fastq software and a maximum of one mismatch in the indexing read was used in demultiplexing. Fastq files were processed using an in-house pipeline (https://github.com/HaojiaWu/Bulk_RNAseq_analysis_pipeline). RNA-seq reads were then aligned to the mm10 (for mouse kidneys) and GRCh38 (for human RPTEC) with STAR version 2.7.3a. Gene counts were derived from the number of uniquely aligned unambiguous reads by Subread:featureCount version 2.0.3. The count matrices were then input to edgeR for data normalization and differential gene analysis.

Cell type deconvolution of bulk RNA-seq data
We used BisqueRNA (Jew et al., 2020) to deconvolute the mouse DKD dataset (this study) and publicly available bulk RNA-seq data from human DKD (Fan et al., 2019; Levin et al., 2020). For the microdissection dataset, raw count matrix was directly downloaded from the link that the authors provided (Levin et al., 2020). For the early and advanced DKD dataset (Fan et al., 2019), fastq files were download from GEO: GSE142025, and processed using our inhouse scripts to generate the count matrix (https://github.com/HaojiaWu/Bulk_RNAseq_analysis_pipeline). RNA-seq reads were then aligned to the mm10 (for mouse kidneys) and GRCh38 (for human RPTEC) with STAR version 2.7.3a. Gene counts were derived from the number of uniquely aligned unambiguous reads by Subread:featureCount version 2.0.3. The count matrices were then input to edgeR for data normalization and differential gene analysis.

Measurement of the drug effect on injured PT
We used three different approaches to assess the drug effect in ameliorating the injured PT. First, we inspected the expression level of the well-known injured markers (such as Havcr1, Cx, and Cxcl1) before and after drug treatments. Second, we used the top 50 upregulated genes in comparing AAV versus healthy PT to score the cells. PT injury after treatment was measured by the scores assigned to each group. Third, we used a discriminant analysis tool, scID (Boufea et al., 2020), to qualify the percent of cells mapping to disease or healthy PT after treatments. We set a cutoff of logFold-change = 0.6 for a gene to be selected as marker genes to input in to scID.

GRN inference and metabolic activity analysis
We inferred GRNs independently on the injured PT subsets from different groups of the scRNA-seq data using SCENIC (Aibar et al., 2017; Van de Sande et al., 2020). In brief, we log2-transformed the CPM-normalized counts of each subset with a prior of one and used those normalized counts for pySCENIC in a docker container. We create cell GRN networks from injured PT single nuclei data, with the mouse mm10 genome for cis-regulatory analysis and the gene-motif rankings 10 kilobases around the transcription start site (TSS). AUCell values of the regulons from the SCENIC output file were used for differential test and to generate the dot plot. For metabolic analysis, we ran COMPASS (Wagner et al., 2021) on the normalized count data from the injured PT of different groups. Reaction scores for each single cell were processed using the R script provided by the authors. Reactions from fatty acid oxidation were selected and visualized.

Association of cell type with GWAS traits by integrative analysis
MAGMA (de Leeuw et al., 2015) was used to test for a positive association between the binned fractions in each cell type and the gene-level associations. For a given kidney cell type, this tested whether increasing tissue specificity of gene expression was
associated with increasing common-variant genetic findings for eGFR using information from all of the genes. This analysis was composed of the following steps. First, GWAS summary statistics of the four kidney function traits eGFR, UACR, BUN, and serum urate were downloaded from the CKDGen Consortium Data website (https://ckdgen.imbi.uni-freiburg.de). To match the European linkage disequilibrium (LD) score, we used the European ancestry-specific GWAS summary statistics only. Second, the summary stats were preprocessed following the public tutorial (https://github.com/NathanSkene/MAGMA_Celltyping). MAGMA v1.0 was used to calculate gene-level association statistics using a window of 10 kb upstream and 1.5 kb downstream of each gene, accounting for LD, gene length, gene density, sample size, and minor allele count. Significant heritability enrichment for kidney GWAS traits and the kidney cell types was reported based on the p value with adjusted by multiple tests (i.e. FDR <0.05).

Cell–cell interaction network analysis
We used CellChat (Version 1.1.3) to infer the cell–cell interactions across all kidney cell types and among the glomerular cell types using the expression of known ligand-receptor pairs. To identify the changes of the cell–cell communication networks induced by diabetic nephropathy and assess how drug treatments can rescue the perturbations, we followed the tutorial for comparison analysis of multiple datasets in the CellChat github repository (https://htmlpreview.github.io/?https://github.com/sqjin/CellChat/blob/master/tutorial/Comparison_analysis_of_multiple_datasets.html) and computed the major signaling changes between DKD and healthy as well as the treated and untreated groups by joint manifold learning and quantitative contrasts of multiple cell-cell communication networks. For the analysis on the whole kidney dataset, we selected the “Secreted Signaling pathways” and used the precompiled mouse protein–protein-Interactions as a priori network information. For the analysis on the glomerular dataset, we selected all databases curated in CellChat including the “Secreted Signaling”, “Cell-Cell contact” and “ECM-Receptor”. We then compared the outgoing and incoming interaction strength of each pair of cell types to identify significant changes in sending or receiving signals between groups, and used the interaction strength to qualify the drug treatment effect. We then visualized the most significant ligand-receptor pairs that mediate the cell-cell interaction changes across comparisons by using the netVisual_bubble function in CellChat.