Unified Mouse and Human Kidney Single-Cell Expression Atlas Reveal Commonalities and Differences in Disease States

Jianfu Zhou,[1,2,3,4] Amin Abedini [1,2,3,4] Michael S. Balzer [1,2,3,4] Rojesh Shrestha [1,2,3,4] Poonam Dhillon,[1,2,3,4] Hongbo Liu [1,2,3,4] Hailong Hu [1,2,3,4] and Katalin Susztak [1,2,3,4]

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

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

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

[4] Kidney Innovation Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

ABSTRACT

Background Mouse models have been widely used to understand kidney disease pathomechanisms and play an important role in drug discovery. However, these models have not been systematically analyzed and compared.

Methods We analyzed single-cell RNA sequencing data (36 samples) and bulk gene expression data (42 samples) from 18 commonly used mouse kidney disease models. We compared single-nucleus RNA sequencing data from a mouse diabetic kidney disease model with data from patients with diabetic kidney disease and healthy controls.

Results We generated a uniformly processed mouse single-cell atlas containing information for nearly 300,000 cells, identifying all major kidney cell types and states. Our analysis revealed that changes in fractions of cell types are major drivers of differences in bulk gene expression. Although gene expression changes at the single-cell level were mostly model-specific, different disease models showed similar changes when compared at a pathway level. Tensor decomposition analysis highlighted the important changes in proximal tubule cells in disease states. Specifically, we identified important alterations in expression of metabolic and inflammation-associated pathways. The mouse diabetic kidney disease model and patients with diabetic kidney disease shared only a small number of conserved cell type–specific differentially expressed genes, but we observed pathway-level activation patterns conserved between mouse and human diabetic kidney disease samples.

Conclusions This study provides a comprehensive mouse kidney single-cell atlas and defines gene expression commonalities and differences in disease states in mice. The results highlight the key role of cell heterogeneity in driving changes in bulk gene expression and the limited overlap of single-cell gene expression changes between animal models and patients, but they also reveal consistent pathway-level changes.

INTRODUCTION

Chronic kidney disease (CKD) has become a serious global public health concern, affecting about 800 million people worldwide[1] and one in ten people in the United States.[2] The health care cost associated with CKD has also been rapidly increasing worldwide, which poses a

substantial financial burden on both individuals and the society, particularly in low-income and middle-income countries.[1,2] If left untreated, CKD can progress to end stage kidney failure, which is fatal without dialysis or transplantation.[1] Regrettably, the underlying mechanisms of kidney disease are still poorly understood.

Mouse kidney disease models are crucial in improving our understanding of human CKD development. Several mouse CKD models have been developed, including tubule crystal precipitation model by folic acid (folic acid nephropathy, FAN),[3] outflow obstruction by unilateral ureteral obstruction (UUO),[4] and ischemia reperfusion injury (IRI).[5] In addition, genetic manipulation, such as transgenic overexpression of risk variant apolipoprotein L1 (APOL1)[6] or the metabolic regulator peroxisome proliferator-activated receptor gamma coactivator 1-α (PGC1a)[7] in podocyte has been used to induce progressive kidney disease in mice. Although the initial injury is very different in these mouse models, they are frequently used interchangeably as models of human kidney fibrosis to test the function of various genes and inhibitors. On histological examination, the models show similarities including epithelial dedifferentiation, matrix deposition capillary loss, and immune infiltrate, commonly referred to as tubulointerstitial fibrosis. Interstitial fibrosis is considered a common driver pathway for kidney disease progression, but it is unclear whether lesions with similar histological characteristics are caused by the same or similar cell-type gene expression changes.

Single-cell gene expression analysis provides an unbiased approach for characterizing gene expression changes in millions of cells.[8] Previous studies have indicated that fibrosis is associated with a marked increase in cell diversity in the kidney.[3,4] While proximal tubule (PT) cells are lost in CKD, the number and the types of immune cell increase.[3] Recent research has identified proximal tubule plasticity in kidney disease models with potential new cell types referred to as injured or profibrotic PT cells.[3,5,9–12] Changes in PT metabolism seem to be an important driver of PT dedifferentiation,[3] and necrotic cell death pathways have been proposed to play an important role in maladaptive repair after acute kidney injury (AKI).[5] Despite the tremendous success of using single-cell tools to study kidney disease development in mice, a systematic comparison of cell type–specific changes in genes and pathways between various mouse disease models has not been performed. In addition, there is a significant knowledge gap in this field regarding the lack of systematic comparison of mouse models and patient samples to understand the usefulness of animal models in recapitulating the human condition. These questions are particularly important because the single-cell clustering tools use relative spatial distance analysis,[13,14] and clustering and cell-type observations in one study cannot be directly translated to another.

In this study, we combined single-cell RNA sequencing data from 18 different mouse kidney disease models using a

Significance Statement

Mouse models have been widely used to understand kidney disease pathomechanisms and play an important role in drug discovery. However, these models have not been systematically analyzed and compared. The authors characterized 18 different mouse kidney disease models at both bulk and single-cell gene expression levels and compared single-cell gene expression data from diabetic kidney disease (DKD) mice and from patients with DKD. Although single cell–level gene expression changes were mostly model-specific, different disease models showed similar changes when compared at a pathway level. The authors also found that changes in fractions of cell types are major drivers of bulk gene expression differences. Although the authors found only a small overlap of single cell-level gene expression changes between the mouse DKD model and patients, they observed consistent pathway-level changes.

total of 36 mice. To ensure the identified similarities between mouse models were conserved and robust, we intentionally selected mice with different backgrounds or ages for different models. We found that cell fraction changes accounted for most of the bulk gene expression differences. While cell type–specific gene expression changes revealed significant differences between models, we identified conserved pathway activity patterns among various mouse models. In addition, by comparing human and mouse diabetic kidney single-nucleus RNA sequencing data, we demonstrated important pathway conservation between patient samples and mouse models.

METHODS

Mouse Models

A full list of mouse models studied in this article was provided in Supplemental Dataset 1. Specifically, the FAN (GEO accession number *GSE156686*),[3] UUO (GEO accession numbers *GSE182256* and *GSE210716*),[4] IRI (GEO accession number *GSE180420*),[5] LPS (GEO accession number *GSE151658*),[15] APOL1 (GEO accession numbers *GSE181671* and *GSE81492*),[6,16] Cisplatin (GEO accession number *GSE207587*),[17] and DKD (GEO accession number *GSE184652*)[12] models and their respective datasets were published in previous studies. In addition to these published datasets, for this study, we produced the Esrra, Notch1, and PGC1a mouse models and their respective datasets, which we deposited in GEO with accession number *GSE220493*. Animal studies were approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Pennsylvania.

Esrra Knockout Mouse Model

The male C57BL/6 Esrra KO mice were kindly provided by Dr. Liming Pei (University of Pennsylvania).[3] The mice were aged 8 weeks. Mice were housed in the Institute pathogen-free animal house (12 hour dark/light cycle) in a temperature-controlled and humidity-controlled

environment (23±1°C) and fed with standard mouse diet and water ad libitum.

Notch1 Transgenic Mouse Model

We intercrossed Pax8rtTA mice and tetO-ICN1 mice to generate the double-transgenic Pax8rtTA/ICNotch1 mice.[18] All mice were on the FVB background. The resultant mice were fed doxycycline-containing food (Bio Serv S3888) beginning at age 4 weeks and were sacrificed at age 6 weeks. Only male mice were used in this study.

PGC1a Transgenic Mouse Model

We crossed male tetO-Ppargc1a transgenic mice with female Nefta mice carrying the nephrin-rtTA transgene to generate the double-transgenic mice.[7] All mice were on the FVB background. The resultant double-transgenic mice were placed on doxycycline-containing chow (Bio Serv S3888) starting at age 4 weeks and were sacrificed at age 6 weeks. Only male mice were used in this study.

Preparation of Mouse Single-Cell Suspension

Euthanized mice were perfused with chilled $1\times$ PBS through the left ventricle. Kidneys were harvested, minced into approximately 1 mm^3 cubes, and digested using Multi Tissue dissociation kit (Miltenyi, 130-110-201). The tissue was homogenized using 21G and 26.5G syringes. Up to 0.25 g of the tissue was digested with 50 μl of Enzyme D, 25 μl of Enzyme R, and 6.75 μl of Enzyme A in an ml of RPMI and incubated 30 minutes at 37°C. Reaction was deactivated by 10% FBS. The solution was then passed through a 40 μm cell strainer. After centrifugation at 400 g for 5 minutes, cell pellet was incubated with 1 ml of RBC lysis buffer on ice for 3 minutes. Cell number and viability were analyzed using Countess AutoCounter (Invitrogen, C10227). This method generated single-cell suspension with >80% viability.

Single-Cell RNA Sequencing

Ten thousand cells were loaded into the Chromium Controller (10× Genomics, PN-120223) on a Chromium Single Cell B Chip (10× Genomics, PN-120262) and processed to generate single-cell gel beads in the emulsion (GEM) according to the manufacturer's protocol (10× Genomics, CG000183). Libraries were generated using Chromium Single Cell 3′ Reagent Kits v3 (10× Genomics, PN-1000092) and Chromium i7 Multiplex Kit (10× Genomics, PN-120262) according to the manufacturer's manual. Quality control for constructed library was performed by using the Agilent Bioanalyzer High Sensitivity DNA kit (Agilent Technologies, 5067-4626) for qualitative analysis. Quantification analysis was performed by using the Illumina Library Quantification Kit (KAPA Biosystems, KK4824). The library was sequenced on the Illumina HiSeq 4000 system with 2×150 paired-end kits using the following read length: 28 bp Read1 for cell barcode and UMI, 8 bp I7 index for sample index, and 91 bp Read2 for transcript.

Mouse scRNA-seq Data Analysis

Raw Data Collecting and Alignment

We downloaded the raw fastq files from Gene Expression Omnibus (GEO)[19,20] using the Sequence Read Archive (SRA) Toolkit 2.10.8 (https://github.com/ncbi/sra-tools). For each sample in the mouse scRNA-seq atlas, we generated a gene-by-cell count matrix by aligning its fastq files to the mm10 reference dataset (version 3.0.0) using 10× Genomics Cell Ranger 3.1.0.[21]

Quality Control

The quality control (QC) included (1) removing ambient mRNA contamination, (2) eliminating low-quality cells and genes, and (3) excluding doublet-like cells.

We used SoupX 1.4.8 to remove ambient mRNA.[22] Function *autoEstCont* estimated the level of background contamination for each cell with parameter *forceAccept=TRUE*. Function *adjustCounts* removed the contamination by correcting the count matrix with parameter *roundToInt=TRUE*.

We defined low-quality cells as those with (1) the number of unique molecular identifiers (UMIs) ≤500, (2) the number of detected genes ≤250 or ≥2500, (3) the percentage of mitochondrially encoded gene reads (*i.e.*, mt-*) ≥50%, or (4) the ratio of detected genes to UMIs ≤0.25. We excluded these cells using Seurat 3.2.2.[23] We further removed genes that were expressed in <10 cells.

To exclude doublet-like cells, we used DoubletFinder 2.0.3.[24] Function *paramSweep_v3* explored the combinations of parameters *pN* and *pK* with arguments *PCs=1:10* and *sct=FALSE*. Assuming no ground truth, functions *summarizeSweep* and *find.pK* selected *pK* as the value with the greatest mean-variance normalized bimodality coefficient. Function *modelHomotypic* predicted the proportion of homotypic doublets. We estimated an expected number for doublets using a multiplet rate suggested by 10× Genomics[25] and further adjusted this number with the proportion of homotypic doublets. Using the select *pK* and the expected number of doublets, function *doubletFinder_v3* identified doublet-like cells with parameters *PCs=1:10, pN=0.25,* and *sct=FALSE*. After eliminating doublet-like cells, we had 438,686 cells left in total for downstream analysis.

Preprocessing, Integration, and Clustering

To process the data, we used Seurat 4.0.4.[26] (1) For each cell, we normalized and natural log–transformed its expression of each gene. (2) We identified top 2000 highly variable genes (HVGs) within each sample and merged them, to avoid selecting sample-specific HVGs. (3) We scaled and centered the HVGs. (4) We reduced the dimensionality of the data by performing principal component analysis (PCA). (5) To correct the sample-induced batch effect, we integrated the data using harmonypy 0.0.5.[27] (6) We did another dimensional reduction on the data by running UMAP and constructed a shared nearest neighbor (SNN) graph of cells. (7) We directly clustered the SNN graph of cells using the Leiden algorithm

with a range of resolutions.[13,14] (*8*) We selected the clusters produced by resolution 0.8 for downstream analysis.

Cell-Type Correlation between Our and Published Mouse Kidney Data

To compare cell-type annotations between our scRNA-seq and published mouse kidney single-cell/single-nucleus RNA-seq data, we performed the following steps. First, we calculated the averaged expression value of each gene for each cell type using function *AverageExpression* of Seurat 4.0.4.[26] Then, for each gene, we scaled its expression in each cell type by computing the z-scores. Finally, based on the genes shared by our and the published data, we computed the Pearson correlation coefficient for each pair of cell types between our and the published data using function *cor* of R package stats 4.1.2 with parameters *use=“complete.obs”* and *method=“pearson”*.

Tensor Decomposition

To perform tensor decomposition analysis, we used scITD 1.0.2.[28] We used the count matrix of our mouse scRNA-seq atlas along with the condition, model, strain, and age information as the input and formed the pseudobulk tensor using function *form_tensor* with parameters *donor_min_cells=0*, *vargenes_method=“norm_var_pvals”*, *vargenes_thresh=0.1*, *batch_var=“sample”*, and *var_scale_power=2*. In particular, *batch_var=“sample”* acted as a correction to the sample-induced batch effect. We determined the number of extracted factors using function *determine_ranks_tucker* with parameters *num_iter=10* and *var_scale_power=2*. We further validated the factor number by assessing the stability of the factors using function *run_stability_analysis* with parameters *sub_prop=0.95* and *n_iterations=50*, which showed that all the resultant factors had mean donor score correlation close to 1. Then, we ran the Tucker tensor decomposition using function *run_tucker_ica*. Finally, we identified genes that were significantly associated with each factor using function *get_lm_pvals*.

Subclustering of PT Cells

We isolated PT cells from our mouse scRNA-seq atlas and retrieved the corresponding quality-controlled count matrix. To subcluster PT cells, we used Seurat 4.0.4 to run the same steps as those for the entire mouse scRNA-seq atlas. We selected the clusters identified using resolution 0.3 for downstream analysis. Clusters representing contamination by mixed identities and high mitochondrial ratio were removed from further analyses.

Trajectory Analysis

Monocle 2. To build single-cell trajectories for the PT cells, we first did a random sampling of the PT cells. For each of the identified S1, S2, and S2/S3 cells, we randomly selected the same number of cells from each studied mouse model, which picked out roughly 3200 cells from each cell type. Because the total number of the identified S3 cells was

approximately 3200 and the cell counts of the identified injured PT cells were much lower, we included all S3, Injured1, and Injured2 cells. This resulted in a total of 13,252 PT cells for trajectory analysis. Then, we used Monocle 2 (version 2.22.0) to run the following steps.[29] (*1*) We selected genes for ordering cells if they were expressed in ≥10 cells, their mean expression value was ≥0.01, and their empirical dispersion value was ≥2. (*2*) We reduced the dimensionality of the cells using function *reduceDimension* with parameter *reduction_method=“DDRTree”*. (*3*) We calculated the trajectories for the cells using function *orderCells*.

RNA Velocity. We generated a BAM file by aligning the associated fastq files to the mm10 reference dataset (version 3.0.0) using 10× Genomics Cell Ranger 3.1.0[21] for each mouse scRNA-seq sample. We further generated a loom file by analyzing the BAM file using the Python-implemented velocyto command line tool (version 0.17.17).[30] The loom file contains two gene-by-cell count matrices, one for spliced RNA reads and the other for unspliced RNA reads.

To calculate RNA velocity for the PT cells, we isolated the exact same 13,252 cells that were previously selected for Monocle analysis from the loom files. To process the data, we used scVelo 0.2.4 to run the following steps.[31] (*1*) Function *scvelo.pp.filter_and_normalize* did the filtering, normalization, and log transforming of the data. (*2*) Function *scvelo.pp.moments* calculated the first-order and second-order moments for velocity estimation. (*3*) Function *scvelo.tl.recover_dynamics* learned the full transcriptional dynamics of splicing kinetics. (*4*) Function *scvelo.tl.velocity* estimated velocities with parameter *mode=“dynamical”*. (*5*) Function *scvelo.tl.velocity_graph* computed a velocity graph with *mode_neighbors=“connectivities”*. (*6*) Function *scvelo.tl.latent_time* predicted the pseudotime of individual cells.

Weighted Gene Coexpression Network Analysis

We performed the weighted gene coexpression network analysis (WGCNA) of the exact same 13,252 PT cells that were previously selected for Monocle analysis using R packages hdWGCNA 0.1.1.9011[32,33] and WGCNA 1.71.[34] The original WGCNA was designed for analyzing bulk gene expression data rather than the sparse scRNA-seq data. To make WGCNA compatible with the latter, we aggregated transcriptionally neighboring single cells from the same mouse model and cell type into pseudobulk metacells (Supplemental Figure 11A) using function *construct_metacells* of hdWGCNA. Then, we selected an appropriate soft-thresholding power for constructing the coexpression network using function *pickSoftThreshold* of WGCNA with parameters *blockSize=20000*, *corFnc=“bicor”*, and *networkType=“signed”*. With the soft-thresholding power, we built the coexpression network and identified consensus modules using function *blockwiseConsensusModules* of WGCNA with parameters *maxBlockSize=20000*, *corType=“pearson”*, *networkType=“signed”*,

deepSplit=4, *minModuleSize=50* (meaning that a valid module contained ≥50 genes), and *mergeCutHeight=0.2* (indicating that modules with a correlation of >0.8 were merged).

Identification of Marker Genes and Differentially Expressed Genes
We manually collected a list of marker genes from previous publications[3,35–40] for manual cell annotation. To identify DEGs in each cell type (against all the other cell types), we used function *FindMarkers* of Seurat 3.2.2[23] with parameters *test.use*="MAST", *min.pct=0.1*, and *logfc.threshold=0.25*. We further filtered the resultant DEGs using *p_val_adj<0.05*. Within each cell type, we used function *FindMarkers* with the same parameters to calculate DEGs in each disease model against its corresponding control, which were filtered using *p_val_adj<0.05* as well. To identify genes regulated along one cell trajectory of a mouse model, we used function *FindMarkers* with the same parameters to compute DEGs in the end state of the trajectory against the root state, which were further filtered using *p_val_adj<0.05*.

Gene Set Enrichment Analysis
For gene sets of interest, we performed the enrichment analysis using the Over-Representation Analysis (ORA) provided by WebGestalt (version 2019).[41] We filtered the enriched KEGG pathways using FDR <0.05.

Mouse Bulk RNA-seq Data Analysis
Quality Control, Alignment, and Deconvolution
We first pruned low-quality bases and adapter sequences from fastq files using Trim Galore 0.6.6 (https://github.com/FelixKrueger/TrimGalore). Then, we aligned reads to the Gencode mouse genome (mm10) using STAR 2.6.1e.[42] Finally, we quantified gene-level and transcript-level expressions using RSEM 1.3.3.[43] We performed bulk RNA-seq deconvolution and estimated cell fractions using CIBERSORTx[44] with DEGs of each cell type in a published mouse kidney scRNA-seq dataset[3] as the reference.

Identification of Differentially Expressed Genes
For each disease model, we calculated DEGs by comparing disease samples with their corresponding controls using DESeq2.[45] We only used genes that had expression ≥1 transcript per million (TPM) in ≥2 samples. We further filtered the resultant DEGs using *|log2FoldChange|≥1* and *padj<0.05*. To adjust DEGs using the estimated cell fractions, we added the latter when setting up *DESeqDataSet* and removed them when performing differential expression analysis.

Mouse snRNA-seq Data Analysis
Raw Data Collecting and Alignment
We downloaded the raw fastq files from GEO[19,20] using the SRA Toolkit 2.10.8 (https://github.com/ncbi/sra-tools). For each sample, we produced a gene-by-cell count matrix through aligning its fastq files to the mm10 reference dataset (version 2020-A) using 10× Genomics Cell Ranger 7.1.0[21] with parameter –include-introns set to true.

Quality Control
We cleaned each count matrix using the same QC steps as those for the mouse scRNA-seq atlas except that we defined low-quality nuclei as those that met any of the following criteria: (1) the number of UMIs was ≤500; (2) the number of detected genes was ≤250 or ≥2500; (3) the percentage of mitochondrially encoded gene reads (*i.e.*, mt-*) was ≥1%; (4) the ratio of detected genes to UMIs was ≤0.25.

Preprocessing, Integration, and Annotation
To process the data, we used Seurat 4.0.4 to run the following steps. (1) For each cell, we normalized and natural log–transformed its expression of each gene. (2) We identified top 2000 HVGs within each sample. (3) Among these HVGs, we selected the top 2000 that were repeatedly variable across samples for integration. (4) Using the select HVGs, we identified anchors using function *FindIntegrationAnchors* with parameters *reference=NULL*, *reduction*="cca", and *dims=1:30*. (5) Using the identified anchors, we integrated the samples together using function *IntegrateData* with parameter *dims=1:30*. (6) We scaled and centered the select HVGs on the integrated data. (7) We reduced the dimensionality of the integrated data by performing PCA. (8) We did another dimensional reduction on the integrated data by running UMAP and constructed a SNN graph of cells. (9) We directly used the cell annotations downloaded from GEO accession number *GSE184652*.[12]

Trajectory Analysis
To build single-cell trajectories for the PT and injured PT cells, we randomly selected 500 PT and 500 injured PT cells, which resulted in a total of 1000 cells for trajectory analysis. Then, we used Monocle 2 (version 2.22.0)[29] to run the same steps as those for the PT cells in the mouse scRNA-seq atlas.

Identification of Differentially Expressed Genes
To calculate various types of DEGs, we used the exact same ways (*i.e.*, function, parameters, and thresholds) as those for the mouse scRNA-seq atlas.

Weighted Gene Coexpression Network Analysis
To perform the WGCNA of the exact same 1000 PT and injured PT cells that were previously selected for Monocle analysis, we used R packages hdWGCNA 0.1.1.9011[32,33] and WGCNA 1.71[34] to run the same steps as those for the PT cells in the mouse scRNA-seq atlas. Furthermore, to measure the preservation of mouse WGCNA modules in the human genome, we first converted the human gene symbols into the mouse ones using function *convert_human_to_mouse_symbols* of R package nichenetr 1.1.0[46] with parameter *version=1* and then calculated the preservation using function *modulePreservation* of the WGCNA package with parameters *referenceNetworks=1*, *nPermutations=200*, and *quickCor=0*.

Gene Set Enrichment Analysis
For gene sets of interest, we performed the enrichment analysis as we did for the PT cells in the mouse scRNA-seq atlas.

Human snRNA-seq Data Analysis

Raw Data Alignment
For each human sample, we produced a gene-by-cell count matrix by aligning its fastq files to the hg19 reference dataset using 10× Genomics Cell Ranger 6.0.1[21] with the –*include-introns* option.

Quality Control
We cleaned each count matrix using the same QC steps as those for the mouse scRNA-seq atlas except that we defined low-quality nuclei as those that met any of the following criteria: (*1*) the number of detected genes was ≤250 or ≥2500, (*2*) the percentage of mitochondrially encoded gene reads (*i.e.*, MT-*) was ≥5%, or (*3*) the ratio of detected genes to UMIs was ≤0.25.

Preprocessing, Integration, and Clustering
To process the data, we used Seurat 4.0.4[26] to run the same steps as those for the mouse scRNA-seq atlas. We selected the clusters found using resolution 1 for downstream analysis. Clusters representing contamination by mixed identities and high mitochondrial ratio were removed from further analyses.

Trajectory Analysis
To build single-cell trajectories for the PT and injured PT cells, we randomly selected 500 PT and 500 injured PT cells, which resulted in a total of 1000 cells for trajectory analysis. Then, we used Monocle 2 (version 2.22.0)[29] to run the same steps as those for the PT cells in the mouse scRNA-seq atlas.

Identification of Marker Genes and Differentially Expressed Genes
We manually collected a list of marker genes from previous publications[11,47–49] for manual cell annotation. To calculate various types of DEGs, we used the exact same ways (*i.e.*, function, parameters, and thresholds) as those for the mouse scRNA-seq atlas.

Weighted Gene Coexpression Network Analysis
To perform the WGCNA analysis of the exact same 1000 PT and injured PT cells that were previously selected for Monocle analysis, we used R packages hdWGCNA 0.1.1.9011[32,33] and WGCNA 1.71[34] to run the same steps as those for the PT cells in the mouse scRNA-seq atlas. Furthermore, to measure the preservation of human WGCNA modules in the mouse genome, we first converted the human gene symbols into the mouse ones using function *convert_human_to_mouse_symbols* of R package nichenetr 1.1.0[46] with parameter *version=1* and then calculated the preservation using function *modulePreservation* of the WGCNA package with parameters *referenceNetworks=1*, *nPermutations=200*, and *quickCor=0*.

Gene Set Enrichment Analysis
For gene sets of interest, we performed the enrichment analysis as we did for the PT cells in the mouse scRNA-seq atlas.

Integration of Human and Mouse snRNA-seq Data
From the annotated human and mouse snRNA-seq data, we retrieved their respective count matrices. To make the human data compatible with the mouse data, we converted the human gene symbols into the mouse ones using function *convert_human_to_mouse_symbols* of R package nichenetr 1.1.0[46] with parameter *version=1*. We then filtered the human and mouse count matrices by their shared genes, which were further combined to form a single gene-by-cell count matrix. To process this combined count matrix, we used Seurat 4.0.4[26] to run the following steps. (*1*) For each cell, we normalized and natural log–transformed its expression of each gene. (*2*) We identified top 2000 HVGs within each sample. (*3*) Among these HVGs, we selected the top 2000 that were repeatedly variable across samples for integration. (*4*) We scaled and centered the select HVGs on each sample. (*5*) We ran PCA on each sample using the select HVGs. (*6*) Using the select HVGs, we identified anchors using function *FindIntegration Anchors* with parameters *reference=NULL*, *reduction="rpca"*, *dims=1:50*, and *k.anchor=15*. (*7*) Using the identified anchors, we integrated the samples together using function *IntegrateData* with parameter *dims=1:50*. (*8*) We scaled and centered the select HVGs on the integrated data. (*9*) We reduced the dimensionality of the integrated data by performing PCA. (*10*) We did another dimensional reduction on the integrated data by running UMAP and constructed a SNN graph of cells. (*11*) We directly used and combined the cell annotations from the annotated human and mouse snRNA-seq data. Specifically, Endo included GEC and Endo from the human data and Endo from the mouse data; PT included PT from the human data and PCT and PST from the mouse data; DLOH was DTL from the mouse data; ALOH included ALOH from the human data and tAL and TAL from the mouse data; Immune included Macro, B lymph, T lymph, and Immune from the human data and Immune from the mouse data.

Cell-Type Correlation between Human and Mouse snRNA-seq Data
To compare cell-type annotations between human and mouse snRNA-seq data, we performed the following steps. First, we calculated the averaged expression value of each gene for each cell type using function *AverageExpression* of Seurat 4.0.4.[26] Then, for each gene, we scaled its expression in each cell type by computing the z-scores. To make the human data compatible with the mouse data, we further converted the human gene symbols into the mouse ones using function *convert_human_to_mouse_symbols* of R package nichenetr 1.1.0[46] with parameter *version=1*. Finally, based on the genes shared by the human and mouse data, we computed the Pearson correlation coefficient for each pair of cell types between the human and mouse data using function *cor* of

R package stats 4.1.2 with parameters *use="complete.obs"* and *method="pearson"*.

Integration of Human and Mouse PT Cells

We isolated the PT and injured PT cells from the combined human and mouse count matrix. To integrate the PT and injured PT cells, we used Seurat 4.0.4 to run the same steps as those for the mouse snRNA-seq data except that we identified anchors using function *FindIntegrationAnchors* with parameters *reference=c("HK2989","GSE184652-dbmPBSrep2")*, *reduction="cca"*, and *dims=1:30*, indicating that we used two control samples (HK2989 from human and GSE184652-dbmPBSrep2 from mouse) as the reference.

Data Availability

The mouse single-cell RNA sequencing data used in this article are available at GEO accession numbers *GSE107585*, *GSE151658*, *GSE156686*, *GSE180420*, *GSE181671*, *GSE182256*, and *GSE220493*. The processed mouse scRNA-seq data can be viewed using an interactive website at https://susztaklab.com/Mouse_scRNA_Atlas/index.php. The mouse DKD single-nucleus RNA sequencing data used in this article are available at GEO accession number *GSE184652*. The mouse bulk RNA sequencing data used in this article are available at GEO accession numbers *GSE156686*, *GSE207587*, *GSE210716*, and *GSE81492*. The human DKD single-nucleus RNA sequencing data are available at GEO accession number *GSE211785*.

Code Availability

The codes used to perform all the analyses in this study are available at GitHub (https://github.com/jzhou88/mouse_kidney_single_cell).

RESULTS

Single-Cell Atlas of Mouse Kidney Disease Models

To investigate disease specific and shared gene expression changes in various mouse kidney disease models, we created a unified single-cell RNA sequencing (scRNA-seq) atlas for a collection of mouse kidney models. We included 36 (27 disease and nine control) mouse kidney samples (Supplemental Dataset 1), of which 29 were generated in our laboratory[3–5,7,16,18,38] (Methods) and seven were downloaded from publicly available databases.[15] We intentionally included mice with different backgrounds and ages to ensure that the identified commonalities between mouse models were conserved and robust. The 27 disease samples comprised various disease models (Figure 1A, Supplemental Dataset 1), such as FAN,[3] UUO,[4] long and short IRI,[5] and mice with tubule-specific transgenic expression with Notch1 intracellular domain (Notch1)[18] and podocyte-specific expression of PGC1a[7] and APOL1.[6,16] Each sequencing data underwent rigorous quality control (Supplemental Figure 1,

Methods), before we integrated them and performed batch effect correction (Supplemental Figure 2, Methods). Using unsupervised clustering, we identified all previously recognized major cell types within the mouse kidney[3,5,38] (Figure 1, B and D, and Supplemental Figure 3), based on the expression of canonical marker genes[3,5,38] (Figure 1C, Supplemental Dataset 2).
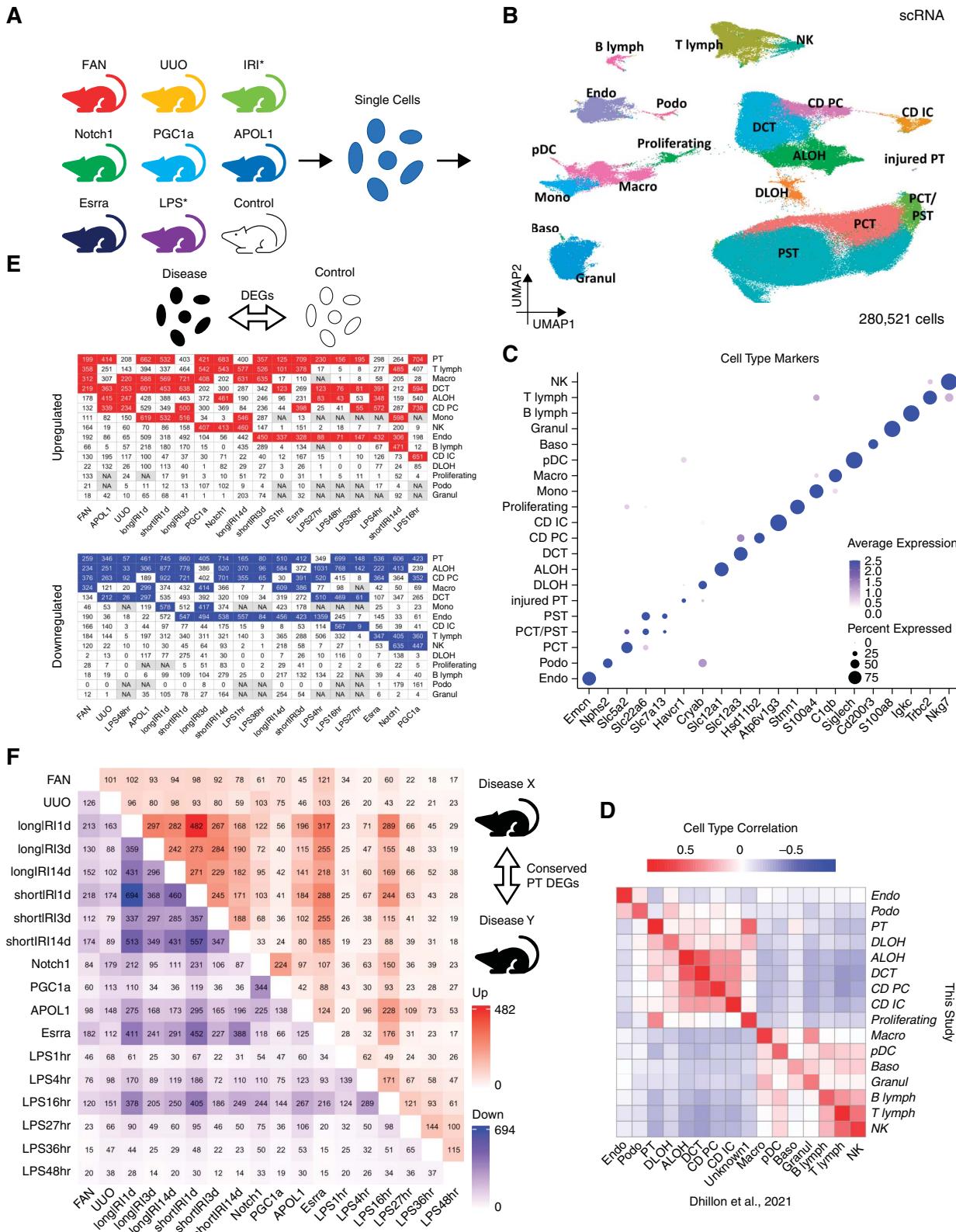
Next, we examined cell type–specific gene expression changes in different kidney disease models. We generated a list of differentially expressed genes (DEGs) for all identified cell types in each studied disease model (Figure 1E, Supplemental Dataset 3, Methods). We observed important gene expression differences in each cell type across different models. Most of the models had at least seven cell types with more than 100 upregulated genes and at least six cell types with more than 100 downregulated genes, when compared with controls. Interestingly, there was no obvious cell type consistently showing the greatest number of DEGs in all models. In general, immune cells showed many genes with higher expression levels in disease while tubule cells had the most genes with lower expression levels in disease. However, different disease models exhibited different degrees of cell type–specific changes in gene expression. Notably, PT cells showed a large number of DEGs in almost all models, which could also partially be explained by their abundance.

To further understand consistent pattern of gene expression changes at the single-cell level, we overlapped the identified DEGs in each cell type between different models. Despite the large number of DEGs, we identified a relatively small number of genes showing consistent changes in all disease models (Figure 1F and Supplemental Figure 4). The IRI models sampled at different time points showed shared gene expression.[5]

In summary, here we generated a large comprehensive mouse kidney single cell atlas by analyzing most commonly used mouse models. We show cell type–specific DEGs in each model, however, we identify a relatively small number of genes that show conserved differential expression in all models.

Cell Fraction Changes Account for Shared Bulk Gene Expression Changes in Mouse Disease Models

We analyzed bulk RNA-seq data from whole-kidney samples of the following disease models previously published by our laboratory (Supplemental Dataset 1, Methods): FAN,[3] UUO,[4] long and short IRI (longIRI1d, longIRI3d, longIRI14d, short-IRI1d, shortIRI3d, and shortIRI14d),[5] Cisplatin,[17] and the APOL1 transgenic (APOL1) model.[6] We performed differential gene expression analysis to understand changes in disease states. Differential expression testing of the bulk RNA-seq data indicated dramatic gene expression differences between disease models and controls (Figure 2A, Supplemental Dataset 4, Methods). All models (except shortIRI3d and shortIRI14d) had more than 1000 genes with higher expression and more than 1000 genes with lower expression when compared with controls. We even found important consistency between the

Figure 1. Single-cell RNA-seq atlas of mouse kidney disease models. (A) Mouse kidney models used to generate the scRNA-seq atlas: APOL1, *Apol1* transgenic; Control, wild-type control; Esrra, *Esrra* knockout; FAN, folic acid nephropathy; IRI, ischemia reperfusion injury; LPS, endotoxin (LPS) injection; Notch1, *Notch1* transgenic; PGC1a, *Pgc1a* transgenic; UUO, unilateral ureteral obstruction. Note, IRI contains long and short IRI samples collected 1, 3, and 14 days after the ischemia; LPS contains samples obtained

Figure 1. *(Continued)* at 1, 4, 16, 27, 36, and 48 hours after the LPS injection. (B) UMAP of 280,521 mouse kidney single cells. Nineteen cell types were identified: ALOH, ascending loop of Henle; B lymph, B lymphocyte; Baso, basophile; CD IC, collecting duct intercalated cell; CD PC, collecting duct principal cell; DCT, distal convoluted tubule; DLOH, descending loop of Henle; Endo, endothelial cell; Granul, granulocyte; injured PT, injured proximal tubule; Macro, macrophage; Mono, monocyte; NK, natural killer cell; PCT, proximal convoluted tubule; pDC, plasmacytoid dendritic cell; Podo, podocyte; Proliferating, proliferating cell; PST, proximal straight tubule; T lymph, T lymphocyte. (C) Dot plot of cell type–specific marker genes (dot size denotes percentage of cells expressing the marker, and color scale represents average gene expression values). (D) Heatmap showing Pearson correlation coefficients of averaged cell type gene expression between mouse kidney scRNA-seq atlas generated in this study and a published mouse scRNA-seq dataset with control and FAN kidney samples.[3] (E) Numbers of upregulated (top) and downregulated (bottom) cell type–specific DEGs (disease versus control). NA (*i.e.*, not applicable) means not enough cells in either disease or control groups for DEG identification. Within each mouse model, the four cell types having the most upregulated and downregulated DEGs are highlighted in red and blue, respectively. (F) Heatmap showing the numbers of upregulated (upper triangle) and downregulated (lower triangle) cell type–specific DEGs (against the control) conserved between any two studied mouse kidney disease models in PT cells. Figure 1 can be viewed in color online at www.jasn.org.

models: hundreds of genes were commonly differentially expressed in several models.

However, we recognized that bulk gene expression changes could reflect changes either in cell fractions or within specific cell types. Therefore, we estimated cell fractions in the bulk RNA-seq data by performing *in silico* deconvolution with a published mouse kidney scRNA-seq data[3] as the reference (Methods). This analysis indicated broad differences in cell fractions within each disease model compared with controls, with lower epithelial cell fractions and higher immune cell fractions in disease models (Figure 2B). To understand the contribution of cell fraction changes to the overall bulk gene expression differences, we corrected the bulk gene expression changes for the estimated cell fraction differences (Methods). This analysis showed that the number of identified DEGs was dramatically reduced in all the models (Figure 2C, Supplemental Dataset 5), indicating that the bulk DEGs mostly reflected cell fraction changes.

In summary, our results indicate marked gene expression changes in disease models when analyzed at whole-kidney level. Bulk gene expression changes are shared between different disease models. Cell fraction changes in disease state drive most observed bulk gene expression changes in mouse kidney disease samples.
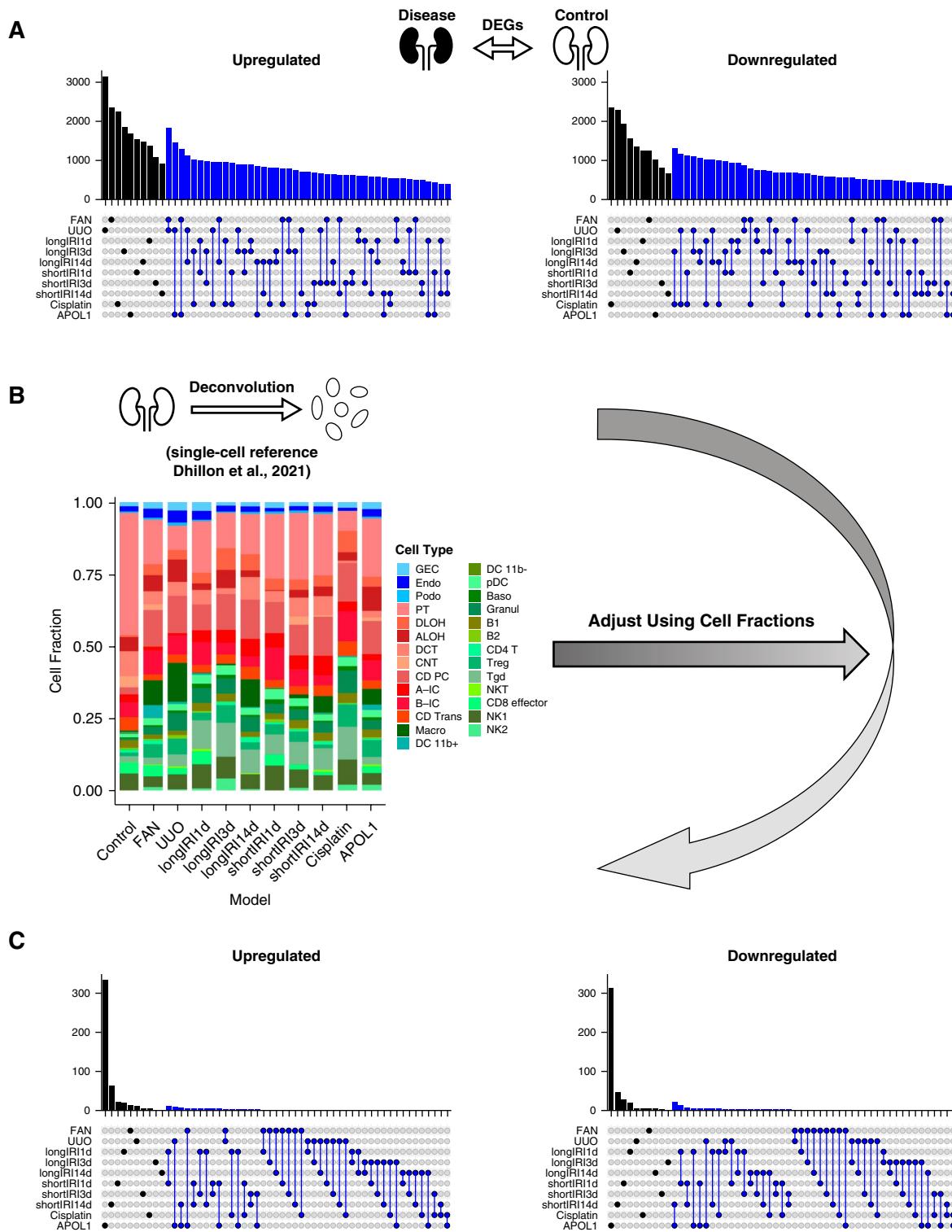
Tensor Decomposition Recognizes PT as Central Cell Type in Mouse Kidney Disease Models

To identify cell types and pathways that play important roles in kidney disease development, we applied Single-Cell Interpretable Tensor Decomposition (scITD)[28] (Figure 3A, Methods). scITD is a computational method capable of extracting multicellular gene expression programs that vary across samples. The approach is premised on the idea that higher-level biological processes often involve the coordinated actions and interactions of multiple cell types. Given single-cell expression data from multiple heterogenous samples, scITD aims to detect these joint patterns of dysregulation affecting multiple cell types.[28] The analysis highlighted five factors (Figure 3, B–D and Supplemental Figure 5). Most variation was explained by factors 1 and 2, but factor 1 did not
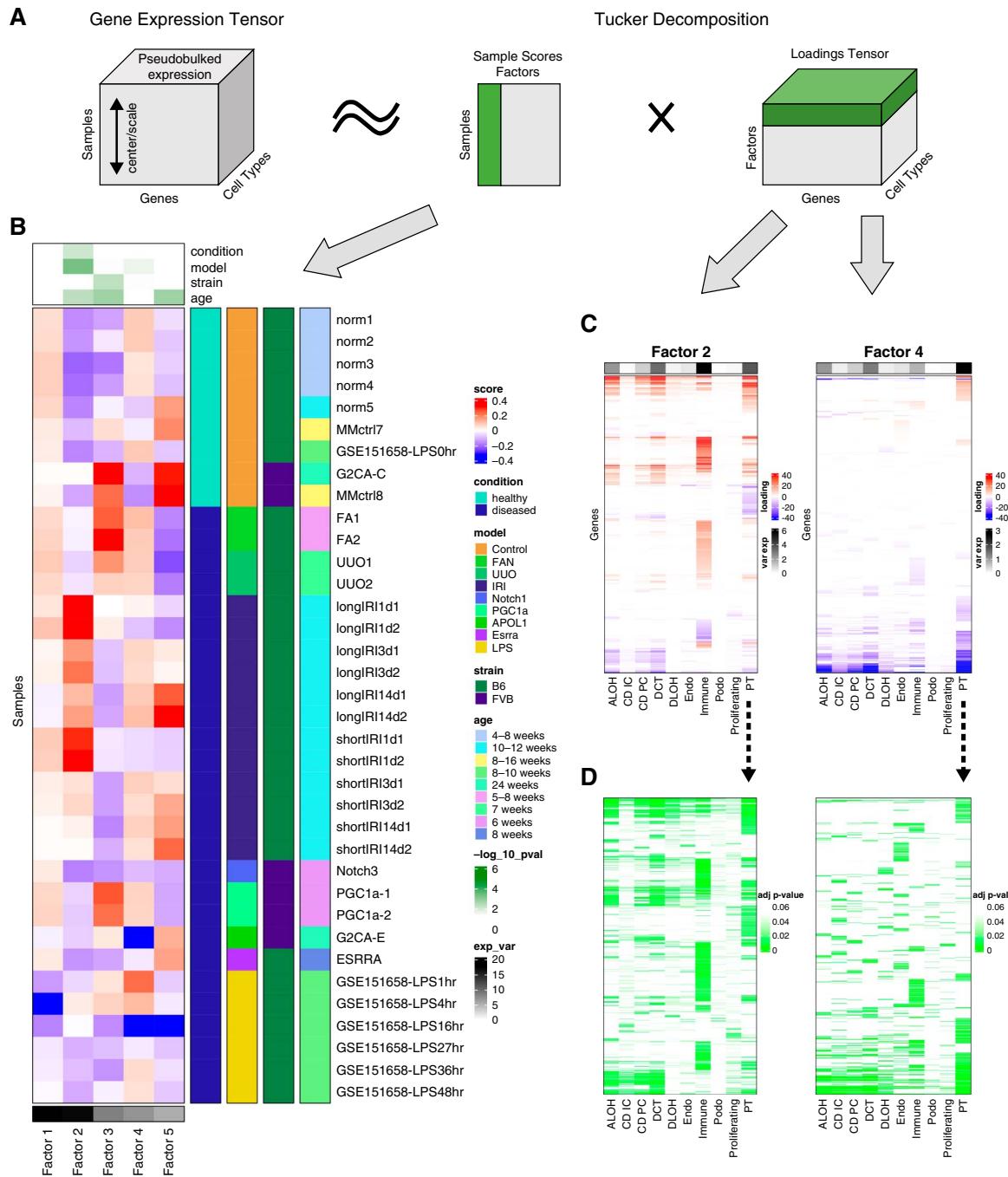
show observable association with sample phenotype. Factor 2 had the strongest association with the analyzed kidney disease model information. IRI showed the greatest enrichment for factor 2, and immune cells explained the most variation among all the cell types, highlighting the key role of immune cells in IRI. Most disease models showed enrichment for Factor 4, and PT cells explained most of the variation. This indicated that PT cells could explain the variation in the model information. Factor 3 was associated with the strain information but was not associated with PT cells. Factor 5 was strongly associated with the age information. Within factor 5, only a few samples had positive sample scores, and PT cells, as well as DCT and immune cells, explained the most variation, albeit this was much smaller than factor 4. In summary, unbiased tensor decomposition analysis indicated the key role of the PT cells (and factor 4) in mouse kidney disease development across different disease models.

Analysis of Proximal Tubule Cells Confirms Conserved Pathway Activities across Different Mouse Disease Models

As unsupervised tensor decomposition analysis highlighted PT as the key cell type associated with phenotypic outcome in mouse kidney disease models, we next decided to focus on PT cells. First, we subclustered the 70,501 PT cells (Figure 4A and Supplemental Figure 6, Methods). Among the resultant cell clusters, we identified the three key PT segments (Figure 4B, Supplemental Dataset 6): S1 (featured by the expression of *Slc5a2*), S2 (featured by the expression of *Slc22a6*), S3 (featured by the expression of *Atp11a*), and a cluster of S2 and S3 cells (featured by the coexpression of *Slc22a6* and *Atp11a*). The remaining clusters were injured PT cells (*i.e.*, Injured1 and Injured2), featured by the coexpression of *Havcr1* and *Krt20* and lower expression of canonical PT markers (*e.g.*, *Lrp2*). The identified PT subtypes showed consistent gene expression changes with previous publication[9] (Figure 4C). We also investigated the fractions of PT subtypes in each studied mouse model (Figure 4D). Notably, most Injured1 cells came from the IRI models. The Injured1 PT cells were mainly associated with the AKI.[5,9] The longIRI1d model

Figure 2. Cell fraction changes account for most bulk kidney gene expression differences in mouse disease models. (A) Upset plots showing the numbers of upregulated (left) and downregulated (right) DEGs (against the control) of each mouse kidney disease model in the bulk RNA-seq data. The black bar represents the DEG count for a single model, while the blue bar shows the number of DEGs conserved between two models. (B) Bar plot showing the average cell fraction of each mouse kidney model in the bulk RNA-seq data predicted by CIBERSORTx deconvolution[44] with a published mouse kidney scRNA-seq dataset[3] as the reference. (C) Upset plots showing the numbers of upregulated (left) and downregulated (right) DEGs (against the control) of each mouse kidney disease model in the bulk RNA-seq data after adjusting for cell fractions. The black bar represents the DEG count for a single model, while the blue bar shows the number of DEGs conserved between two models. Figure 2 can be viewed in color online at www.jasn.org.

Figure 3. Tensor decomposition identifies PT as key disease kidney driving cell type across animal models. (A) A schematic diagram illustrating the tensor decomposition using scITD adapted from the Kharchenko Lab.[28] First, a pseudobulked tensor is created from cell populations of multiple samples (left). Then, tucker decomposition extracted the most informative factors, each comprising a vector of sample scores (middle) and a loadings matrix (right). Sample scores and loadings for one factor are highlighted in green. (B) Sample score heatmap for the decomposition of the mouse kidney scRNA-seq data. At the top, the *P*-values for associations between the factor scores and mouse kidney condition, model, strain, and age information are shown. The *P*-values were calculated using univariate linear model F-tests. Rows are grouped by mouse conditions and models, shown as annotations on the right side. Name, strain, and age of each mouse sample were also shown as annotations on the right side. Columns are ordered by explained variance for each factor, shown as a bottom annotation. (C) Loading matrices for factors 2 and 4 limited to significant genes. The top annotation shows the percentage of overall explained variance for each cell type of the factor. Rows are hierarchically clustered. (D) The same matrices for factors 2 and 4 as those in (C) except that each entry shows the association significance *P*-value of each gene in each cell type of the factor. Figure 3 can be viewed in color online at www.jasn.org.

had a higher Injured1 fraction than the shortIRI1d model, consistent with the more severe injury in the former.[5] The injured1 fraction was lower on day 3 after the IRI compared with day 1 (longIRI3d versus longIRI1d and shortIRI3d versus shortIRI1d). This was concordant with our publication showing the highest serum creatinine on day 1 after IRI.[5] After 14 days, the number of injured cells was minima in both IRI groups.[5] Most Injured2 cells originated from the LPS models.[15]

To understand continuous changes in gene expression, we performed cell trajectory analysis on PT cells (Methods). The analysis indicated that the PT cells branched into two directions (*i.e.*, Trajectories 1 and 2) (Figure 4E and Supplemental Figure 7), where Trajectory 1 headed toward an AKI phenotype (featured by the increasing expression of *Havcr1* and the decreasing expression of *Slc22a30* along the trajectory and dominated by longIRI1d, shortIRI1d, APOL1,[6] and LPS16hr,[15] namely the AKI models), and Trajectory 2 headed toward a CKD phenotype (featured by the increasing expression of *Slc5a2* and the decreasing expression of *Havcr1* along the trajectory and dominated by the CKD models, such FAN and UUO). To validate cell trajectories, we performed RNA velocity analysis on the same involved PT cells (Methods). We found two main trajectories (*i.e.*, red and yellow) (Figure 4F). The red trajectory headed toward AKI (dominated by longIRI1d, shortIRI1d, APOL1,[6] and LPS16hr,[15] namely the AKI models), and the yellow trajectory headed toward CKD (dominated by the CKD models, such as FAN and UUO) (Supplemental Figure 8). Thus, the red trajectory was corresponding to Trajectory 1, and the yellow trajectory was corresponding to Trajectory 2. We next identified genes regulated along the trajectories in each analyzed mouse model and performed gene set enrichment analysis (GSEA) accordingly (Supplemental Datasets 7 and 8, Methods). Interestingly, although very few genes were regulated simultaneously in all the models (Supplemental Figure 9), we found many enriched pathways were quite conserved over these mouse models (Figure 4G and Supplemental Figure 10). In particular, metabolic pathways and glutathione metabolism were the most conserved pathways enriched in Trajectory 1 while PPAR signaling pathway and peroxisome were the most conserved ones enriched in Trajectory 2. These results indicated important potential consistency in PT cell states in different disease models.

To understand gene pathways and networks associated with disease states, we performed weighted gene coexpression network analysis (WGCNA) on the PT cells used for the trajectory analysis above[33,34,50] (Supplemental Figure 11A, Methods). We retrieved nine gene modules (*i.e.*, black, blue, brown, green, magenta, pink, red, turquoise, and yellow). We then performed gene set enrichment analysis for each module (Figure 5 and Supplemental Figure 11B, Supplemental Datasets 9 and 10, Methods). Remarkably, the gene modules were conserved in multiple mouse models. The pink module, for example, was enriched in FAN, UUO, LPS1hr, LPS4hr, and LPS48hr, indicating its conservation
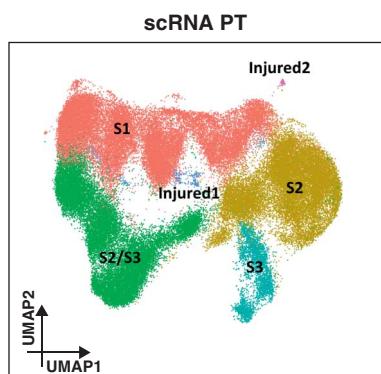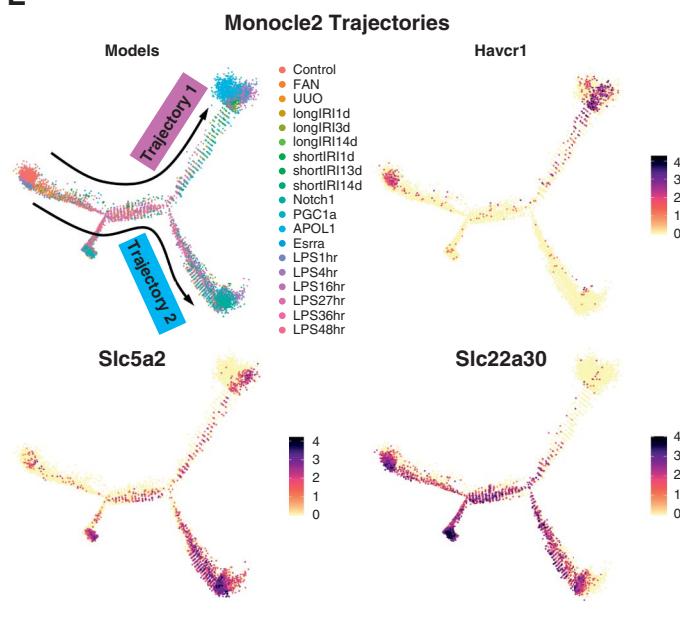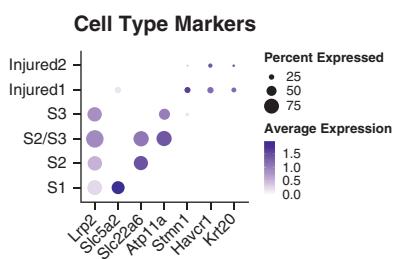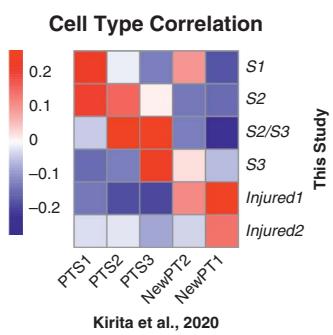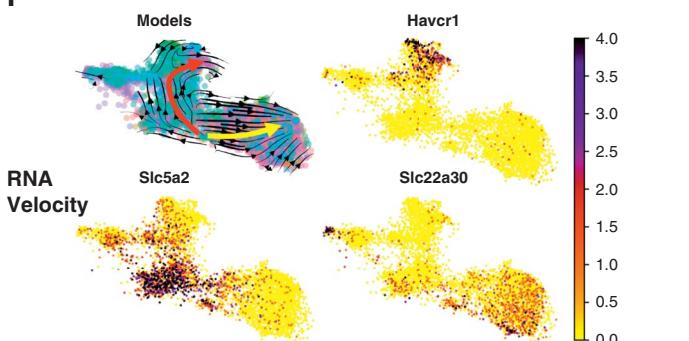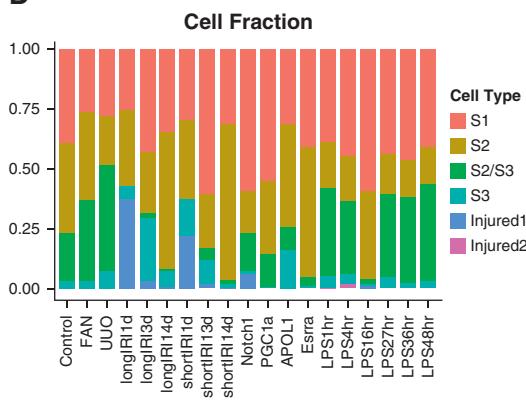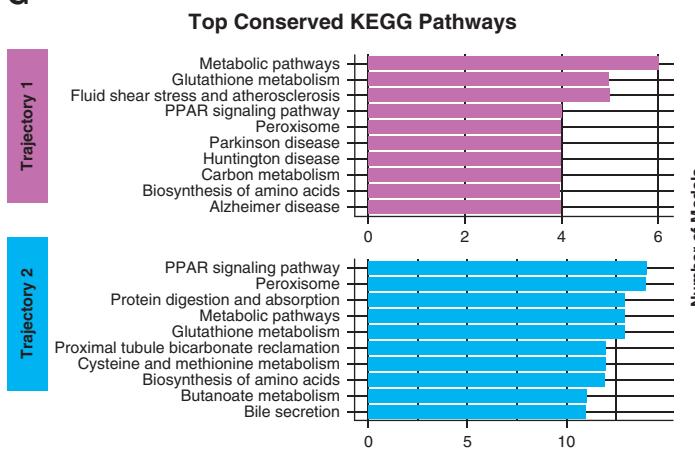
between different CKD models (*e.g.*, FAN and UUO). GSEA showed that peroxisome, lipid metabolism, and cholesterol metabolism were the top pathways enriched in the pink module, which was concordant with the conserved pathways we identified in Trajectory 2. The yellow module was primarily enriched in longIRI1d, shortIRI1d, LPS4hr, and LPS16hr. GSEA showed that TNF signaling pathway was the top pathway enriched in the yellow module.

In summary, despite the differences we observed in gene expression changes between different disease models, we observed consistency in patterns of cell trajectories, states, and pathways between different kidney disease models.

Conserved Pathway Activities in Mouse and Human Diabetic Kidney PT Cells

Finally, we aimed to compare gene expression changes observed in kidneys of mice with patients with similar kidney disease. As a case study, we used a single-nucleus RNA sequencing (snRNA-seq) dataset of human and mouse kidneys previously published by our and other laboratories,[12,49] including diabetic kidney disease (DKD) and healthy control samples (Figure 6A, Supplemental Dataset 1). We applied the same quality control steps to the count matrix of each snRNA-seq sample as those of the mouse scRNA-seq samples (Supplemental Figure 12, Methods), which resulted in 54,945 human and 123,704 mouse high-quality kidney nuclei. We then corrected for batch effects and integrated human and mouse snRNA-seq samples separately (Supplemental Figures 13, A–B and 14, A–B, Methods). We performed unsupervised clustering and identified all previously published major cell types in human nuclei,[11,47,49,51,52] based on the expression of typical marker genes[11,47–49] (Supplemental Figure 13, C and D, Supplemental Dataset 11, Methods). For the mouse samples, we directly used the cell type annotations from their original publication[12] (Supplemental Figure 14, C and D, Supplemental Dataset 12). Next, we integrated the transcriptomes of the annotated human and mouse kidney nuclei (Figure 6B and Supplemental Figure 15, A–C, Methods). We found that human cell types were automatically grouped together with their respective mouse counterparts (Supplemental Figure 15D). To verify the human–mouse cell type annotation consistency, we computed the Pearson correlation coefficients of averaged cell type gene expression between human and mouse (Methods). To further investigate whether human and mouse cell types share similar gene expression patterns, we performed differential expression analysis of each cell type in the human and mouse data and compared the resulting DEGs (Methods). Although the cell annotations aligned well between the human and mouse data (Figure 6C), human and mouse DKD samples shared very few cell type–specific DEGs (Supplemental Figure 16, Supplemental Datasets 13 and 14). These results indicated important differences between mouse and human DKD.

Because PT was the key cell type related to the disease phenotypes in the mouse models, we next investigated the PT

Figure 4. Conserved pathway activities in PT cells across different mouse disease models. (A) UMAP of 70,501 PT cells from the mouse scRNA-seq data. Among the resultant cell clusters, we identified the three key PT segments: S1 (featured by the expression of *Slc5a2*), S2 (featured by the expression of *Slc22a6*), S3 (featured by the expression of *Atp11a*), and a cluster of S2 and S3 cells

Figure 4. *(Continued)* (featured by the coexpression of *Slc22a6* and *Atp11a*). The remaining clusters were injured PT cells (*i.e.*, Injured1 and Injured2), featured by the coexpression of *Havcr1* and *Krt20* and lower expression of canonical PT markers (*e.g.*, *Lrp2*). (B) Dot plot of cell type–specific marker genes (dot size denotes percentage of cells expressing the marker, and color scale represents average gene expression values). (C) Heatmap showing Pearson correlation coefficients of averaged PT cell subtype gene expression between mouse kidney scRNA-seq atlas generated in this study and a published mouse snRNA-seq dataset with control and IRI kidney samples.[9] (D) Bar plot showing the PT subtype cell fractions in each mouse kidney model of the scRNA-seq data. (E) Trajectory analysis of PT cells. The trajectories were calculated using Monocle 2.[29] All four panels show the same trajectories. The top left panel indicates the location of the cells along the trajectories for mouse kidney models. The remaining panels show the expression of PT cell subtype marker genes along the trajectories. (F) RNA velocity analysis of PT cells. The RNA velocity was predicted using scVelo.[31] The top left panel identifies two main trajectories, indicated by the red and yellow arrows, respectively. It also shows the cell location along the trajectories for mouse kidney models, using the same color scheme as (E). The remaining panels illustrate the expression of PT cell subtype marker genes along the trajectories. (G) Bar plots showing the top conserved KEGG pathways among the mouse kidney models in each identified Monocle 2 trajectory. Figure 4 can be viewed in color online at www.jasn.org.

nuclei. First, we integrated the human and mouse PT nuclei (Figure 6D and Supplemental Figure 17, A–C, Methods). We found that human and mouse PT nuclei grouped together nicely (Supplemental Figure 17D). Then, we performed trajectory analysis of the human and mouse PT nuclei (Methods). The analysis indicated that both human and mouse shared a common trajectory from PT to injured PT (Figure 6E), which was featured by the increasing expression of injured PT markers (*i.e.*, *HAVCR1* and *VCAM1* in human and *Havcr1* and *Vcam1* in mouse) and the decreasing expression of pan-PT markers (*i.e.*, *SLC27A2* and *LRP2* in human and *Slc27a2* and *Lrp2* in mouse) along the trajectory. We further identified genes regulated along the human and mouse trajectories (Supplemental Dataset 15, Methods). Although only a limited number of genes were regulated in both human and mouse PT nuclei along the trajectory (Figure 6F), we found multiple conserved pathways enriched in both human and mouse PT nuclei (Figure 6G and Supplemental Figure 18, Supplemental Dataset 16, Methods). Particularly, fibrosis-associated adherens junction and ECM-receptor interaction were among the most conserved pathways showing consistency between mouse and human DKD.

Finally, to study gene pathways and networks related to human and mouse DKD, we performed WGCNA of human and mouse PT nuclei[33,34,50] (Supplemental Figure 19, A and B, Methods). We retrieved three gene modules (*i.e.*, turquoise, blue, and brown) in the human data and ten gene modules (*i.e.*, yellow, brown, purple, red, blue, pink, black, green, turquoise, and magenta) in the mouse data (Supplemental Figure 19, C and D, Supplemental Datasets 17 and 18). Gene set enrichment analysis of each module highlighted important gene expression changes in DKD (Supplemental Figure 20, Supplemental Datasets 19 and 20, Methods). We next analyzed the module preservation between human and mouse DKD (Methods). Surprisingly, we found that two of three human modules showed strong preservation in mouse DKD (*i.e.*, turquoise and blue), and four mouse modules showed strong preservation in human DKD (*i.e.*, yellow, brown, purple, and red) (Figure 6H). GSEA of the human turquoise and mouse yellow modules indicated conserved changes in metabolic pathways between human and mouse DKD. GSEA
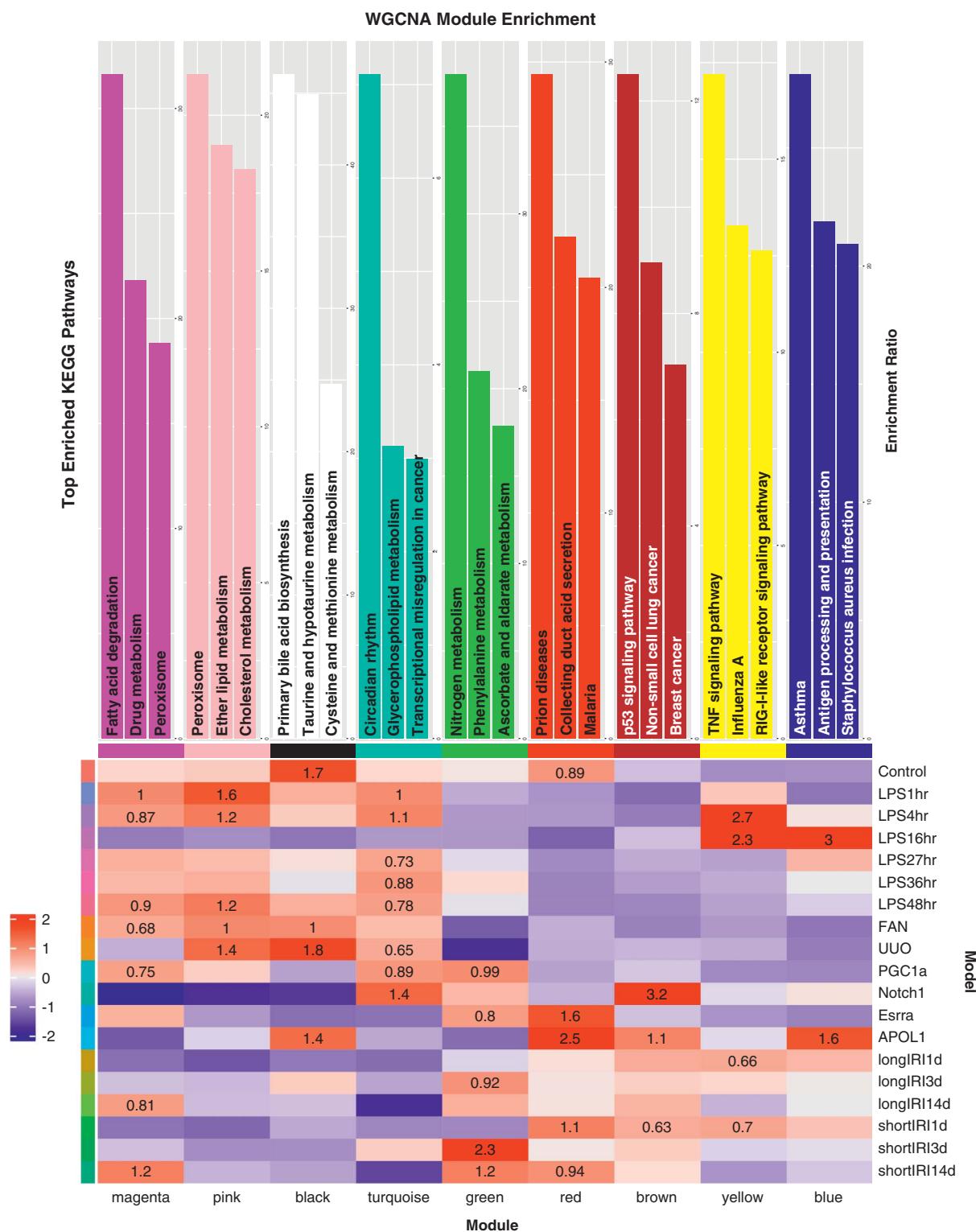
of the human blue module showed that fibrosis-related adherens junction, ECM-receptor interaction, and focal adhesion pathways were also conserved between human and mouse DKD. GSEA of the mouse brown module revealed the enrichment for TNF signaling pathway, NF-k B signaling pathway, AGE-RAGE signaling pathway, and apoptosis was conserved between the human and mouse DKD as well.

In summary, the mouse and human DKD single-nucleus data identified consistent cell types between human and mouse kidneys. Human and mouse DKD samples exhibited important and unique cell type–specific gene expression changes. Despite the differences in cell-type gene expression patterns, gene expression changes were conserved at a pathway level.
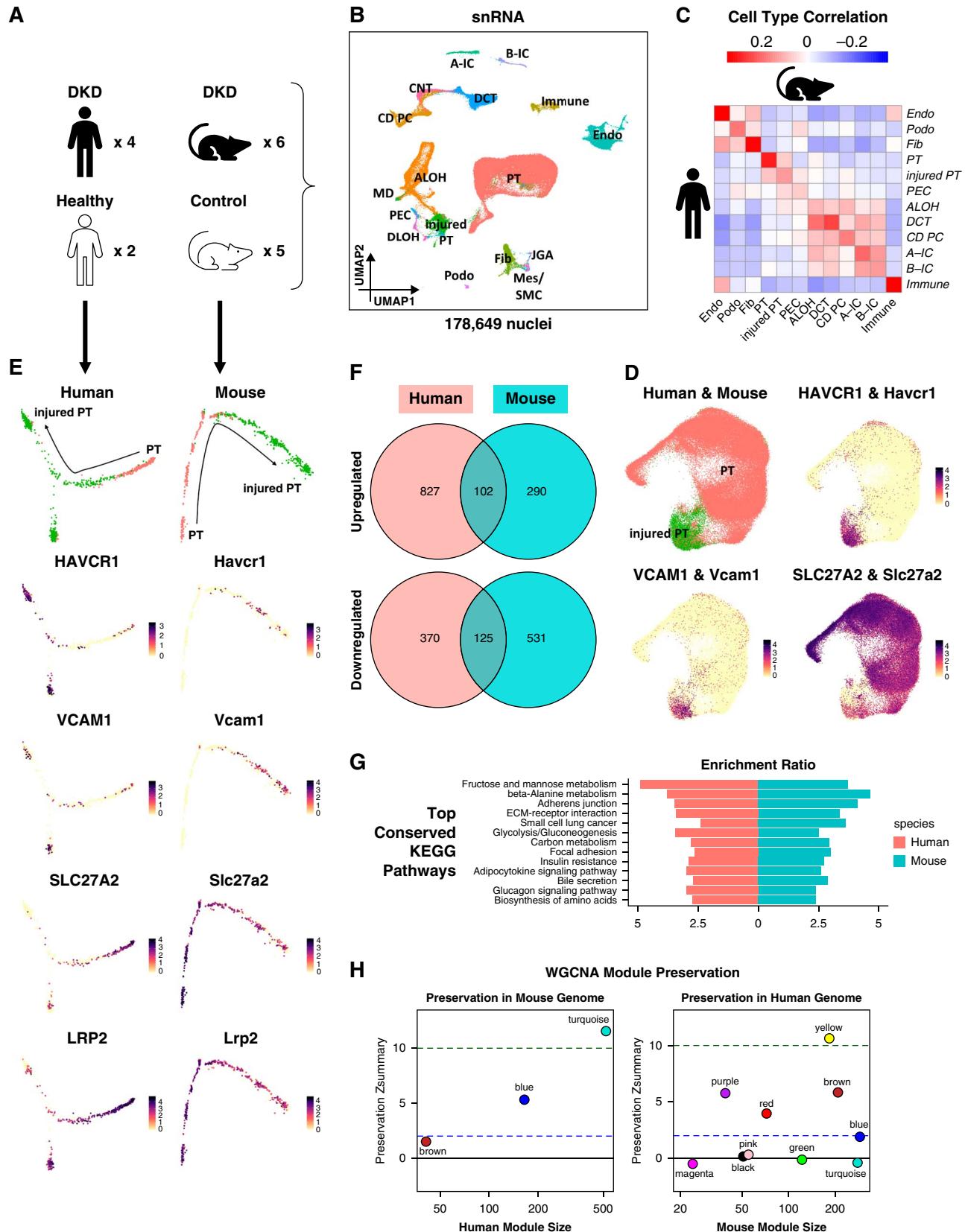
DISCUSSION

Here we present a comprehensive mouse kidney single-cell RNA-seq atlas that offers a uniformly processed dataset for multiple disease states and various conditions, including strain and age. By integrating different mouse models, we identified consistent cell-type markers, which provide insights into the changes across different conditions. Our study highlights the significant role of cell fraction changes in driving bulk gene expression changes in mouse kidney disease models, and we show that different models present with unique cell type–specific changes, which may be model or disease stage-specific, but we also observed important pathway level conservation.

When comparing mouse kidney disease models to healthy controls at the whole kidney level, we found major differences in gene expression. Specifically, we observed changes in over 1500 genes in any single disease model compared with healthy strain-matched, age-matched, and sex-matched controls. Many of these genes showed differential expressions across multiple or all disease models, indicating consistencies between the different conditions. However, after correcting for cell fraction changes, we found dramatically reduced genes that showed differential expressions in these models. While we acknowledge that computationally estimated cell fraction

Figure 5. Weighted gene coexpression network analysis of PT cells. The heatmap (bottom) demonstrates high WGCNA module association and conservation among the mouse kidney models. The bar plots (top) show the top KEGG pathways enriched in each identified WGCNA module. Figure 5 can be viewed in color online at www.jasn.org.

Figure 6. Human and mouse DKD snRNA-seq atlas. (A) Human and mouse samples used to generate the human and mouse DKD snRNA-seq atlas. (B) UMAP of 54,945 human and 123,704 mouse kidney single nuclei. This was generated by integrating human and

Figure 6. *(Continued)* mouse snRNA-seq data. Eighteen cell types were identified: A-IC, alpha intercalated cell; ALOH, ascending loop of Henle; B-IC, beta intercalated cell; CD PC, collecting duct principal cell; CNT, connecting tubule; DCT, distal convoluted tubule; DLOH, descending loop of Henle; Endo, endothelial cell; Fib, fibroblast; Immune, immune cell; injured PT, injured proximal tubule; JGA, juxtaglomerular apparatus; MD, macular densa; Mes, mesangial cell; PEC, parietal epithelial cell; Podo, podocyte; PT, proximal tubule; SMC, smooth muscle cell. (C) Heatmap showing Pearson correlation coefficients of averaged cell type gene expression between human and mouse kidney snRNA-seq data. Each row represents a cell type in the human data, and each column represents a cell type in the mouse data. (D) Integration of 19,319 human and 70,125 mouse PT and injured PT nuclei. The top left panel shows the resultant UMAP. The remaining panels are feature plots showing the expression of PT and injured PT marker genes. (E) Trajectory analysis of human and mouse PT nuclei. The trajectories were calculated using Monocle 2.[29] Panels on the left column were generated using the human data, while panels on the right column were generated using the mouse data. Panels on the top row indicate the location of PT and injured PT nuclei along the trajectories, while the remaining panels illustrate the expression of PT and injured PT marker genes along the trajectories. (F) Venn diagrams showing the numbers of upregulated (top) and downregulated (bottom) genes along the trajectories that were conserved between the human and mouse data. (G) Bar plot showing the top conserved KEGG pathways between the human and mouse data along the trajectories. (H) Preservation $Z_{summary}$ statistics of human WGCNA modules in the mouse data (left) and mouse WGCNA modules in the human data (right). Each point represents a module. Point color reflects the module color. Points are labeled by their respective colors. Blue and green lines depict the rough thresholds for weak ($Z=2$) and strong ($Z=10$) evidence of module preservation. Figure 6 can be viewed in color online at www.jasn.org.

changes could be biased, our findings suggest that changes in cell fraction play a key role in driving bulk gene expression changes in these models.

Our analysis of PT cells revealed a limited number of disease patterns, and we identified two main distinct trajectories. Interestingly, despite the lack of consistency between cell type–specific DEGs in the single-cell data, we observed common pathway activation patterns in disease states. These effects were consistently observed across multiple disease models and considered a conserved pathway in kidney disease and fibrosis.

We were surprised to see that cell type–specific DEGs were not conserved in different mouse kidney disease models. The observed differences could be attributed to several factors, such as different batches or disease states. Although PT cells had a large number of cell-type DEGs, even the cell types that showed the most DEGs in each model were not consistent across models. However, we acknowledge that the number of analyzed cell types will influence the number of identified significant DEGs. Notably, the tensor decomposition analysis highlighted the potential key role of PT cells in the examined model, which is consistent with previous publications.[3,5]

We also investigated patterns of injury and cell state changes in different disease conditions and found that PT cells follow a relatively limited number of disease patterns, with only two identified. Trajectory 1 headed toward AKI while Trajectory 2 headed toward CKD.

Despite the lack of consistency between cell type–specific DEGs in the single-cell data, we observed common pathway activation patterns in disease states. We found that the WGCNA analysis of PT cells detected gene modules that were enriched in multiple models, most associated with metabolic processes. The trajectory analysis of PT cells found conserved metabolic pathways among multiple models as well. PT metabolism has received considerable attention in recent years,[3,5,10] and it is known that PT cells are highly metabolically active and are one of the cell types with the highest number of mitochondria. Our previous studies have shown that the key cell identity and metabolic transcriptional machinery are linked in PT cells including HNF1b, HNF4a, PPARA, and ESRRA.[3] We found that genetic loss of PPARA and ESRRA was associated with more severe kidney disease, while pharmacological activation ameliorated disease in disease models.[3] Consistent with our single-cell dataset, these effects were consistently observed in multiple disease models and considered a conserved pathway in kidney disease and fibrosis.

In this study, we conducted a comparison of mouse and human diabetic kidney single-nucleus datasets and found a high level of consistency in cell types. However, we also found limited overlap in gene expression changes between mouse and human samples. Our WGCNA and trajectory analysis identified many conserved pathways between mice and patients. For example, fibrosis-associated adherens junction and ECM-receptor interaction, TNF signaling pathway, and NF-κB signaling pathway were all conserved between human and mouse DKD.

We recognize that our study has several limitations. First, we only examined 18 disease models as representatives. The inclusion of additional disease models and disease stages as well as sex differences would enhance the strength of our conclusions. Second, we focused only on PT cells. The analysis of immune cells could be particularly interesting given the well-described differences in immune cells. Finally, the inclusion of additional human samples will also be important to better understand consistencies and differences between mice and patients.

In summary, our study generated a comprehensive and standardized scRNA-seq atlas of mouse kidney. We demonstrated the key role of cell fraction changes in driving bulk expression differences, observed variations in cell type–specific gene expression, and identified consistencies in injury patterns and pathways in disease conditions. By comparing human and mouse DKD snRNA-seq data, we highlighted conserved

pathways between mouse disease models and patient samples. These findings provide important insights into kidney disease and lay the foundation for future research aimed at understanding and treating kidney disorders.

DISCLOSURES

FUNDING

AUTHOR CONTRIBUTIONS

Data curation: Jianfu Zhou.
Formal analysis: Jianfu Zhou.
Funding acquisition: Katalin Susztak.
Investigation: Jianfu Zhou.
Project administration: Jianfu Zhou.
Resources: Amin Abedini, Michael S. Balzer, Poonam Dhillon, Hailong Hu, Hongbo Liu, Rojesh Shrestha.
Supervision: Katalin Susztak.
Visualization: Jianfu Zhou.
Writing – original draft: Katalin Susztak, Jianfu Zhou.
Writing – review & editing: Katalin Susztak, Jianfu Zhou.

SUPPLEMENTAL MATERIAL

This article contains the following supplemental material online at http://links.lww.com/JSN/E502, http://links.lww.com/JSN/E503, http://links.lww.com/JSN/E504, http://links.lww.com/JSN/E505, http://links.lww.com/JSN/E506, http://links.lww.com/JSN/E507, http://links.lww.com/JSN/E508, http://links.lww.com/JSN/E509, http://links.lww.com/JSN/E510, http://links.lww.com/JSN/E511, http://links.lww.com/JSN/E512, http://links.lww.com/JSN/E513, http://links.lww.com/JSN/E514, http://links.lww.com/JSN/E515, http://links.lww.com/JSN/E516, http://links.lww.com/JSN/E517, http://links.lww.com/JSN/E518, http://links.lww.com/JSN/E519, http://links.lww.com/JSN/E520, http://links.lww.com/JSN/E521, and http://links.lww.com/JSN/E522.

Supplemental Figure 1. Quality control of the mouse scRNA-seq data.
Supplemental Figure 2. Integration of the mouse scRNA-seq samples.
Supplemental Figure 3. Cell type correlation.
Supplemental Figure 4. Cell type–specific DEG conservation.
Supplemental Figure 5. Loadings matrices produced by tensor decomposition.
Supplemental Figure 6. Integration of PT cells from the mouse scRNA-seq data.
Supplemental Figure 7. Monocle 2 trajectory analysis of PT cells from the mouse scRNA-seq data.
Supplemental Figure 8. RNA velocity analysis of PT cells from the mouse scRNA-seq data.
Supplemental Figure 9. Conservation of genes regulated along Monocle 2 trajectories of PT cells from the mouse scRNA-seq data.
Supplemental Figure 10. Gene set enrichment analysis of Monocle 2 trajectories of PT cells from the mouse scRNA-seq data.
Supplemental Figure 11. WGCNA of PT cells from the mouse scRNA-seq data.
Supplemental Figure 12. Quality control of the human and mouse snRNA-seq data.
Supplemental Figure 13. Integration of the human snRNA-seq samples.
Supplemental Figure 14. Integration of the mouse snRNA-seq samples.
Supplemental Figure 15. Integration of the human and mouse snRNA-seq samples.
Supplemental Figure 16. Human–mouse cell type–specific DEG conservation.
Supplemental Figure 17. Integration of PT nuclei from the human and mouse snRNA-seq data.
Supplemental Figure 18. Gene set enrichment analysis of Monocle 2 trajectory of PT nuclei from the human and mouse snRNA-seq data.
Supplemental Figure 19. WGCNA of PT nuclei from the human and mouse snRNA-seq data.
Supplemental Figure 20. Gene set enrichment analysis of WGCNA gene modules of PT nuclei from the human and mouse snRNA-seq data.
Supplemental Dataset 1. Information of mouse and human samples.
Supplemental Dataset 2. DEGs of each cell type against all the other cell types in the mouse kidney scRNA-seq data.
Supplemental Dataset 3. DEGs of each disease model against the control in each identified cell type of the mouse kidney scRNA-seq data.
Supplemental Dataset 4. DEGs of each disease model against the control in the mouse kidney bulk RNA-seq data.
Supplemental Dataset 5. Cell fraction–adjusted DEGs of each disease model against the control in the mouse kidney bulk RNA-seq data.
Supplemental Dataset 6. DEGs of each PT cell subtype against all the other PT cell subtypes in the mouse kidney scRNA-seq data.
Supplemental Dataset 7. Genes regulated along each Monocle 2 trajectory of PT cells in each mouse model of the mouse kidney scRNA-seq data.
Supplemental Dataset 8. Enriched KEGG pathways and GO BP terms along each Monocle 2 trajectory of PT cells in each mouse model of the mouse kidney scRNA-seq data.
Supplemental Dataset 9. Gene modules identified by WGCNA of PT cells in the mouse kidney scRNA-seq data.
Supplemental Dataset 10. KEGG pathways and GO BP terms enriched in each identified WGCNA gene module of PT cells in the mouse kidney scRNA-seq data.
Supplemental Dataset 11. DEGs of each cell type against all the other cell types in the human DKD snRNA-seq data.
Supplemental Dataset 12. DEGs of each cell type against all the other cell types in the mouse DKD snRNA-seq data.
Supplemental Dataset 13. DEGs of DKD samples against healthy samples in each identified cell type of the human DKD snRNA-seq data.
Supplemental Dataset 14. DEGs of DKD samples against control samples in each identified cell type of the mouse DKD snRNA-seq data.
Supplemental Dataset 15. Genes regulated along the Monocle 2 trajectory of PT nuclei in the human and mouse DKD snRNA-seq data.
Supplemental Dataset 16. Enriched KEGG pathways and GO BP terms along the Monocle 2 trajectory of PT nuclei in the human and mouse DKD snRNA-seq data.

Supplemental Dataset 17. Gene modules identified by WGCNA of PT nuclei in the human DKD snRNA-seq data.

Supplemental Dataset 18. Gene modules identified by WGCNA of PT nuclei in the mouse DKD snRNA-seq data.

Supplemental Dataset 19. KEGG pathways and GO BP terms enriched in each identified WGCNA gene module of PT nuclei in the human DKD snRNA-seq data.

Supplemental Dataset 20. KEGG pathways and GO BP terms enriched in each identified WGCNA gene module of PT nuclei in the mouse DKD snRNA-seq data.

REFERENCES

1. Levin A, Tonelli M, Bonventre J, Coresh J, Donner JA, Fogo AB. Global kidney health 2017 and beyond: a roadmap for closing gaps in care, research, and policy. *Lancet.* 2017;390(10105):1888–1917. doi:10.1016/S0140-6736(17)30788-2

2. National Kidney Foundation Research Roundtable Work Group on behalf of the National Kidney Foundation. Research priorities for kidney-related research-an agenda to advance kidney care: a position statement from the National kidney foundation. *Am J Kidney Dis.* 2022;79(2):141–152. doi:10.1053/j.ajkd.2021.08.018

3. Dhillon P, Park J, Hurtado Del Pozo C, Li L, Doke T, Huang S. The nuclear receptor ESRRA protects from kidney disease by coupling metabolism and differentiation. *Cell Metab.* 2021;33(2):379–394.e8. doi:10.1016/j.cmet.2020.11.011

4. Doke T, Abedini A, Aldridge DL, Yang YW, Park J, Hernandez CM. Single-cell analysis identifies the interaction of altered renal tubules with basophils orchestrating kidney fibrosis. *Nat Immunol.* 2022;23(6):947–959. doi:10.1038/s41590-022-01200-7

5. Balzer MS, Doke T, Yang YW, Aldridge DL, Hu H, Mai H. Single-cell analysis highlights differences in druggable pathways underlying adaptive or fibrotic kidney regeneration. *Nat Commun.* 2022;13(1):4018. doi:10.1038/s41467-022-31772-9

6. Beckerman P, Bi-Karchin J, Park AS, Qiu C, Dummer PD, Soomro I. Transgenic expression of human APOL1 risk variants in podocytes induces kidney disease in mice. *Nat Med.* 2017;23(4):429–438. doi:10.1038/nm.4287

7. Li SY, Park J, Qiu C, Han SH, Palmer MB, Arany Z. Increasing the level of peroxisome proliferator-activated receptor γ coactivator-1α in podocytes results in collapsing glomerulopathy. *JCI Insight.* 2017;2(14):e92930. doi:10.1172/jci.insight.92930

8. Balzer MS, Ma Z, Zhou J, Abedini A, Susztak K. How to get started with single cell RNA sequencing data analysis. *J Am Soc Nephrol.* 2021;32(6):1279–1292. doi:10.1681/ASN.2020121742

9. 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 U S A.* 2020;117(27):15874–15883. doi:10.1073/pnas.2005477117

10. Li H, Dixon EE, Wu H, Humphreys BD. Comprehensive single-cell transcriptional profiling defines shared and unique epithelial injury responses during kidney fibrosis. *Cell Metab.* 2022;34(12):1977–1998.e9. doi:10.1016/j.cmet.2022.09.026

11. Muto Y, Wilson PC, Ledru N, Wu H, Dimke H, Waikar SS. Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. *Nat Commun.* 2021;12(1):2190. doi:10.1038/s41467-021-22368-w

12. Wu H, Gonzalez Villalobos R, Yao X, Reilly D, Chen T, Rankin M. Mapping the single-cell transcriptomic response of murine diabetic kidney disease to therapies. *Cell Metab.* 2022;34(7):1064–1078.e6. doi:10.1016/j.cmet.2022.05.010

13. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep.* 2019;9(1):5233. doi:10.1038/s41598-019-41695-z

14. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir el AD, Tadmor MD. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell.* 2015;162(1):184–197. doi:10.1016/j.cell.2015.05.047

15. Janosevic D, Myslinski J, McCarthy TW, Zollman A, Syed F, Xuei X. The orchestrated cellular and molecular responses of the kidney to endotoxin define a precise sepsis timeline. *Elife.* 2021;10:e62270. doi:10.7554/eLife.62270

16. Wu JN, Ma ZY, Raman A, Beckerman P, Dhillon P, Mukhi D. APOL1 risk variants in individuals of African genetic ancestry drive endothelial cell defects that exacerbate sepsis. *Immunity.* 2021;54(11):2632–2649.e6. doi:10.1016/j.immuni.2021.10.004

17. Doke T, Mukherjee S, Mukhi D, Dhillon P, Abedini A, Davis JG. NAD(+) precursor supplementation prevents mtRNA/RIG-I-dependent inflammation during kidney injury. *Nat Metab.* 2023;5(3):414–430. doi:10.1038/s42255-023-00761-7

18. Bielesz B, Sirin Y, Si H, Niranjan T, Gruenwald A, Ahn S. Epithelial Notch signaling regulates interstitial fibrosis development in the kidneys of mice and humans. *J Clin Invest.* 2010;120(11):4040–4054. doi:10.1172/JCI43025

19. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 2012;41(D1):D991–D995. doi:10.1093/nar/gks1193

20. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002;30(1):207–210. doi:10.1093/nar/30.1.207

21. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R. Massively parallel digital transcriptional profiling of single cells. *Nat Commun.* 2017;8(1):14049. doi:10.1038/ncomms14049

22. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience.* 2020;9(12):giaa151. doi:10.1093/gigascience/giaa151

23. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM III. Comprehensive integration of single-cell data. *Cell.* 2019;177(7):1888–1902.e21. doi:10.1016/j.cell.2019.05.031

24. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* 2019;8(4):329–337.e4. doi:10.1016/j.cels.2019.03.003

25. *Chromium Single Cell 3' Reagent Kits User Guide (V3 Chemistry), Document Number CG000183 Rev C, 10x Genomics.* 2020.

26. Hao Y, Hao S, Andersen-Nissen E, Mauck WM III, Zheng S, Butler A. Integrated analysis of multimodal single-cell data. *Cell.* 2021;184(13):3573–3587.e29. doi:10.1016/j.cell.2021.04.048

27. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods.* 2019;16(12):1289–1296. doi:10.1038/s41592-019-0619-0

28. Mitchel J, Gordon MG, Perez RK, Biederstedt E, Bueno R, Ye CJ. Tensor decomposition reveals coordinated multicellular patterns of transcriptional variation that distinguish and stratify disease individuals. *bioRxiv.* 2022:480703. doi:10.1101/2022.02.16.480703

29. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods.* 2017;14(10):979–982. doi:10.1038/nmeth.4402

30. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V. RNA velocity of single cells. *Nature.* 2018;560(7719):494–498. doi:10.1038/s41586-018-0414-6

31. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol.* 2020;38(12):1408–1414. doi:10.1038/s41587-020-0591-3

32. Morabito S, Reese F, Rahimzadeh N, Miyoshi E, Swarup V. High dimensional co-expression networks enable discovery of transcriptomic drivers in complex biological systems. *bioRxiv.* 2022:509094. doi:10.1101/2022.09.22.509094

33. Morabito S, Miyoshi E, Michael N, Shahin S, Martini AC, Head E. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat Genet.* 2021;53(8):1143–1155. doi:10.1038/s41588-021-00894-z

34. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9(1):559. doi:10.1186/1471-2105-9-559

35. Balzer MS, Rohacs T, Susztak K. How many cell types are in the kidney and what do they do? *Annu Rev Physiol.* 2022;84(1):507–531. doi:10.1146/annurev-physiol-052521-121841

36. Abedini A, Zhu YO, Chatterjee S, Halasz G, Devalaraja-Narashimha K, Shrestha R. Urinary single-cell profiling captures the cellular diversity of the kidney. *J Am Soc Nephrol.* 2021;32(3):614–627. doi:10.1681/ASN.2020050757

37. Miao Z, Balzer MS, Ma Z, Liu H, Wu J, Shrestha R. Single cell regulatory landscape of the mouse kidney highlights cellular differentiation programs and disease targets. *Nat Commun.* 2021;12(1):2277. doi:10.1038/s41467-021-22266-1

38. Park J, Shrestha R, Qiu C, Kondo A, Huang S, Werth M. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science.* 2018;360(6390):758–763. doi:10.1126/science.aar2131

39. Chung JJ, Goldstein L, Chen YJ, Lee J, Webster JD, Roose-Girma M. Single-cell transcriptome profiling of the kidney glomerulus identifies key cell types and reactions to injury. *J Am Soc Nephrol.* 2020;31(10):2341–2354. doi:10.1681/ASN.2020020220

40. Barry DM, McMillan EA, Kunar B, Lis R, Zhang T, Lu T. Molecular determinants of nephron vascular specialization in the kidney. *Nat Commun.* 2019;10(1):5705. doi:10.1038/s41467-019-12872-5

41. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 2019;47(W1):W199–W205. doi:10.1093/nar/gkz401

42. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21. doi:10.1093/bioinformatics/bts635

43. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12(1):323. doi:10.1186/1471-2105-12-323

44. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol.* 2019;37(7):773–782. doi:10.1038/s41587-019-0114-2

45. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. doi:10.1186/s13059-014-0550-8

46. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods.* 2020;17(2):159–162. doi:10.1038/s41592-019-0667-5

47. Wilson PC, Muto Y, Wu H, Karihaloo A, Waikar SS, Humphreys BD. Multimodal single cell sequencing implicates chromatin accessibility and genetic background in diabetic kidney disease progression. *Nat Commun.* 2022;13(1):5253. doi:10.1038/s41467-022-32972-z

48. Young MD, Mitchell TJ, Vieira Braga FA, Tran MGB, Stewart BJ, Ferdinand JR. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science.* 2018;361(6402):594–599. doi:10.1126/science.aat1699

49. Abedini A, Ma Z, Frederick J, Dhillon P, Balzer MS, Shrestha R. Spatially resolved human kidney multi-omics single cell atlas highlights the key role of the fibrotic microenvironment in kidney disease progression. *bioRxiv.* 2022:513598. doi:10.1101/2022.10.24.513598

50. Morabito S, Reese F, Rahimzadeh N, Miyoshi E, Swarup V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Rep Methods.* 2023;3(6):100498. doi:10.1016/j.crmeth.2023.100498

51. Lake BB, Menon R, Winfree S, et al. An atlas of healthy and injured cell states and niches in the human kidney. *Nature.* 2023;619:585–594. doi:10.1038/s41586-023-05769-3

52. Wilson PC, Wu H, Kirita Y, Uchimura K, Ledru N, Rennke HG. The single-cell transcriptomic landscape of early human diabetic nephropathy. *Proc Natl Acad Sci U S A.* 2019;116(39):19619–19625. doi:10.1073/pnas.1908706116