ORIGINAL ARTICLE

Single-cell transcriptomic profiling of human kidneys uncovers profound principal cell responses to diabetic conditions, enabling targeted drug discovery

Zahra Hamidifarid[1,2], **Yousof Gheisari**[1,2]

[1]Department of Genetics and Molecular Biology, Isfahan University of Medical Sciences, Isfahan, Iran, [2]Regenerative Medicine Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

Background: Ranked as the fifth leading global health burden in 2022, diabetes mellitus is the driver of diabetic kidney disease (DKD), a condition responsible for nearly half of end-stage kidney disease cases. Despite its significant impact, therapeutic progress in DKD is hindered by an incomplete understanding of the underlying cell-specific molecular mechanisms. **Materials and Methods:** Single-cell RNA sequencing data (GSE183276) from the kidney precision medicine project, including kidney samples from 20 healthy individuals and 14 DKD patients, were obtained. After quality control, batch effect correction, and cell type annotation, cell type-specific differentially expressed genes were identified. Enrichment analyses were performed using WebGestalt. Trajectory inference was conducted to characterize dynamic cellular states. Finally, DrugTar, a previously developed deep learning tool, was applied to identify the most druggable genes. **Results:** Among the 22 annotated cell types, collecting duct principal cells exhibited the largest number of differentially expressed genes, even after adjustment for cell abundance. This finding was also observed in an independent human kidney single-cell RNA-seq dataset (GSE211785), suggesting that this cell type experiences a profound transcriptional reprogramming in response to a diabetic milieu. Notably, trajectory analysis revealed distinct transcriptional states within principal cells, with mitochondrial metabolic processes being enriched across these subpopulations. Furthermore, a list of the most druggable genes is proposed as potential candidates for targeted delivery to principal cells. **Conclusion:** This study underscores the significant transcriptional reprogramming of principal cells in DKD pathogenesis, a phenomenon that has rarely received attention in previous works. These cells can be considered as candidates in strategies aimed at targeted drug delivery for DKD.

Key words: Diabetic nephropathy, drug targetintroduction, pathogenesis, principal cell, RNA-Seq

How to cite this article: Hamidifarid Z, Gheisari Y. Single-cell transcriptomic profiling of human kidneys uncovers profound principal cell responses to diabetic conditions, enabling targeted drug discovery. J Res Med Sci 2026;31:23.

INTRODUCTION

Ranked as the fifth most burdensome disease worldwide in 2022 and projected to climb to third place by 2050, diabetes mellitus is at the forefront of a growing global health crisis.[1] Among its severe complications, diabetic kidney disease (DKD) remains the leading cause of chronic kidney disease (CKD), contributing to nearly half of all end-stage kidney disease cases.[2] CKD affects approximately 9% of the global population, ranking as the tenth leading cause of death in 2022, and is expected to become the fifth by 2050.[3] These epidemiological trends underscore the urgent need to elucidate the molecular underpinnings of DKD to facilitate the development of novel therapeutic strategies.[4]

The kidney is a highly complex organ composed of numerous specialized cell types, each with distinct functions.[5] Bulk omics techniques, which provide

Access this article online

Quick Response Code:

Website: https://journals.lww.com/jrms

DOI: 10.4103/jrms.jrms_1331_25

Address for correspondence: Dr. Yousof Gheisari, Regenerative Medicine Research Center, Isfahan University of Medical Sciences, Isfahan, 8174673461, Iran.
E-mail: ygheisari@med.mui.ac.ir
Submitted: 11-Dec-2025; **Revised:** 18-Feb-2026; **Accepted:** 06-May-2026; **Published:** 28-Aug-2026

an averaged signal across diverse cell populations, lack the resolution to reveal cell-specific molecular changes critical to understanding DKD.[6] Single-cell technologies, however, offer an unparalleled opportunity to dissect cellular heterogeneity, enabling a clearer understanding of disease mechanisms.[7] The application of such techniques is particularly important for an organ like the kidney, which harbors a wide variety of specialized cell types.[8,9]

Even when high-dimensional data are available, translating them into meaningful insight remains a key bottleneck.[10] Indeed, the increased availability of omics technologies has led to an explosion in data generation, shifting the critical challenge in biomedical research toward extracting maximal clinical knowledge from high-quality existing resources.[11,12] Despite its potential, the application of single-cell approaches in DKD research, particularly with human samples, remains limited. Among the few available single-cell transcriptional profiling datasets of human kidneys, we obtained the kidney precision medicine project (KPMP) scRNA-seq dataset, which fulfills the quality check parameters and provides an acceptable sample size.[13] Here, this dataset was reanalyzed, revealing that collecting duct principal cells are the most affected cell type in diabetic nephropathy and play a significant role in altered mitochondrial pathways.

Although specific cell types, such as mesangial cells and proximal tubule epithelial cells, have been widely explored in the context of kidney injury, the current study highlights the essential role of principal cells in DKD pathogenesis, which has rarely been investigated so far. Furthermore, these findings were translated into actionable insights by proposing novel therapeutic targets using our previously established druggability scoring pipeline.

MATERIALS AND METHODS

Data collection

This study utilized publicly available single-cell RNA sequencing datasets from the KPMP gene expression omnibus (GEO accession GSE183276) and the Suztak laboratory (GSE211785),[14] both of which were obtained from the NCBI GEO. All analyses were conducted using the Seurat (v5.1.1, Satija Lab, New York Genome Center, New York, NY, USA) package in R (v4.5.0, R Foundation for Statistical Computing, Vienna, Austria).[15,16]

Quality control and filtering

Cells with genes (nFeature_RNA) between 100 and 5000 were retained. A mitochondrial content threshold of ≤ 50% was applied. Doublets were identified and removed using the DoubletFinder package (v 2.0.6, University of California San Francisco, San Francisco, CA, USA).[17]

Normalization and data transformation

Data normalization was performed using the log normalization method with a scale factor of 10,000 to adjust for differences in sequencing depth across cells. Variable features were subsequently identified using the "vst" method, and the top 5000 most variable genes were selected for further analysis. The data were then scaled using the ScaleData function, which standardized gene expression across all cells.

Batch effect correction

Two approaches were employed to mitigate potential batch effects. First, the impact of cell cycle phase differences on clustering was evaluated. Next, batch effects across samples were corrected using the Harmony package (v1.2.3, Broad Institute of MIT and Harvard, Cambridge, MA, USA), which adjusts for technical variability while preserving biological signals.[18]

Dimensionality reduction and feature selection

Linear dimensionality reduction was first performed via principal component analysis. Uniform manifold approximation and projection (UMAP), a nonlinear dimensionality reduction technique, was then applied to the top 30 principal components.

Unsupervised clustering and cell type identification

Clustering was performed using the FindClusters function with a resolution of 3.5 to group the cells into distinct clusters. To identify marker genes for each cluster, the FindAllMarkers function was used, with a minimum percentage of 0.25 and a log fold change threshold of 0.25. Cell type annotation was then performed by referencing the KPMP atlas, assigning cell identities based on marker gene expression profiles.

Differential expression analysis

Differentially Expressed Genes (DEGs) analysis was conducted using the DESeq2 package (v1.48.1) between healthy kidney and DKD samples.[19] DESeq2 uses a negative binomial distribution to model the count data and applies statistical tests to detect genes with significant expression changes. To account for multiple testing, the Benjamini–Hochberg procedure was used to control the false discovery rate (FDR), and genes with an FDR ≤ 0.05 were considered statistically significant.

Gene set enrichment analysis

Gene set enrichment analysis was performed using the WebGestalt tool with the overrepresentation method to identify gene sets that were significantly enriched among the differentially expressed genes (DEGs). The analysis was conducted using default parameters, with the exception of adjusting the significance threshold to the FDR to control for false positives.

Trajectory inference and pseudotime analysis

Trajectory analysis was performed using the Monocle3 package (v1.4.25, Trapnell Laboratory, University of Washington, Seattle, WA, USA).[20] After the principal cells were subsetted and reverted to the raw data layer for this cell type, the standard single-cell RNA sequencing analysis workflow in Seurat (v5.3.0, Satija Lab, New York Genome Center, New York, NY, USA) was applied as described earlier. The processed object was then converted into a "cell_data_set" using the as. cell_data_set function for compatibility with Monocle3. Dimensionality reduction was performed using UMAP, and a trajectory graph was constructed using the learn_graph function with default parameters. Cluster 1, located at one extremity of the UMAP space and composed predominantly of cells from healthy individuals, was considered the root of the trajectory. Pseudotime values were then assigned across the trajectory using the order_cells function.

RESULTS

Twenty-two transcriptionally distinct cell populations were identified in human kidneys through cell type annotation

High-quality, single-cell resolved omics data pave the way for understanding the cellular heterogeneity and molecular mechanisms underlying human kidney diseases.[21] In this study, a publicly available single-cell RNA-seq dataset (GSE183276) from the KPMP consortium was reanalyzed to achieve this goal. This dataset was selected among the few publicly available human kidney datasets due to its high quality, including correction for ambient RNA contamination and high gene coverage per cell. Furthermore, unlike most scRNA-seq pipelines in which cells with more than 15%–20% mitochondrial content are excluded, cells containing up to 50% mitochondrial reads were retained, given the inherently high metabolic activity of the kidney as a highly energy-demanding organ. This dataset contains transcriptomic data from both healthy and diseased human kidneys, including DKD, acute kidney injury (AKI), and hypertensive nephropathy, with a large number of samples and annotated clinical conditions. While the KPMP study is aimed at generating a high-resolution spatial atlas of the human kidney, in the current study, we focused specifically on diabetic nephropathy. Profiles from DKD and normal control samples from this study were retrieved and reanalyzed through a hypothesis-free approach. Disease classification and sample selection were performed based on KPMP metadata annotations linking each cell to its originating specimen and clinical diagnosis, and only profiles labeled as DKD or normal control were retained for downstream analysis.

Using the GSE183276 dataset, 14 DKD and 20 healthy kidney samples were processed through a standardized single-cell RNA-seq analysis workflow [Figure 1]. Doublet cells were identified [Supplementary Figure 1] and removed. Batch effects were assessed at two levels: first, no cell cycle–associated bias was observed, as all phases were represented across clusters [Supplementary Figure 2]; second, sample-level effects were present and subsequently corrected, resulting in clusters containing a mixture of DKD and healthy cells [Supplementary Figure 3]. Unsupervised clusters were then annotated using canonical marker genes from the KPMP atlas, leading to the identification of 22 distinct cell types [Figure 2], which served as the foundation for downstream analyses.

The functional significance of variable expressed genes was identified across different kidney cells using gene set enrichment analysis

To capture molecular differences underlying kidney cell diversity between healthy controls and DKD patients, differentially expressed genes (DEGs) analysis was performed across annotated cell types using DESeq2 [Supplementary Tables 1–22]. Each gene set was subsequently subjected to pathway enrichment analysis, resulting in a total of 735 enriched pathways. Many of these pathways

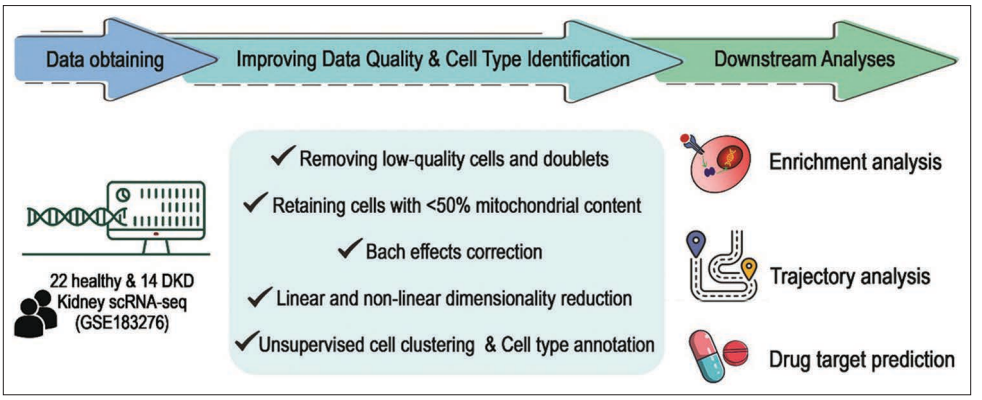


Figure 1: Schematic overview of the study pipeline. Single-cell transcriptomics data from kidney specimens of diabetic nephropathy patients and healthy donors were obtained from the gene expression omnibus database and subjected to upstream and downstream analyses. DKD = Diabetic kidney disease

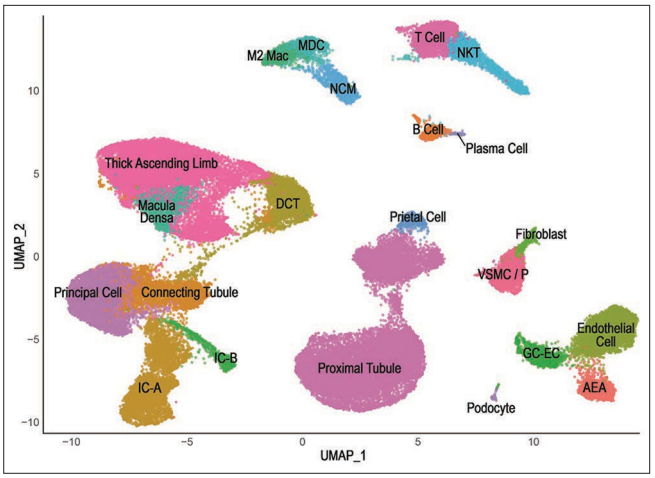


Figure 2: Annotation of cell types in human kidney samples. A total of 60,028 cells from healthy donors and diabetic kidney disease patients were projected using uniform manifold approximation and projection and classified into 22 distinct cell types on the basis of kidney-specific markers curated by the kidney precision medicine project. UMAP = Uniform manifold approximation and projection; KPMP = kidney precision medicine project; DKD = Diabetic kidney disease; IC = Intercalated cell; VSMC/P = Vascular smooth muscle cell/pericyte; DCT = Distal convoluted tubule; GC-EC = Glomerular capillary endothelial cell; M2 Mac = M2 macrophage; NCM = Nonclassical monocyte; MDC = Monocyte-derived cell; NKT = Natural killer T cell; AEA = Afferent/efferent arteriole to descending vasa recta endothelial cell

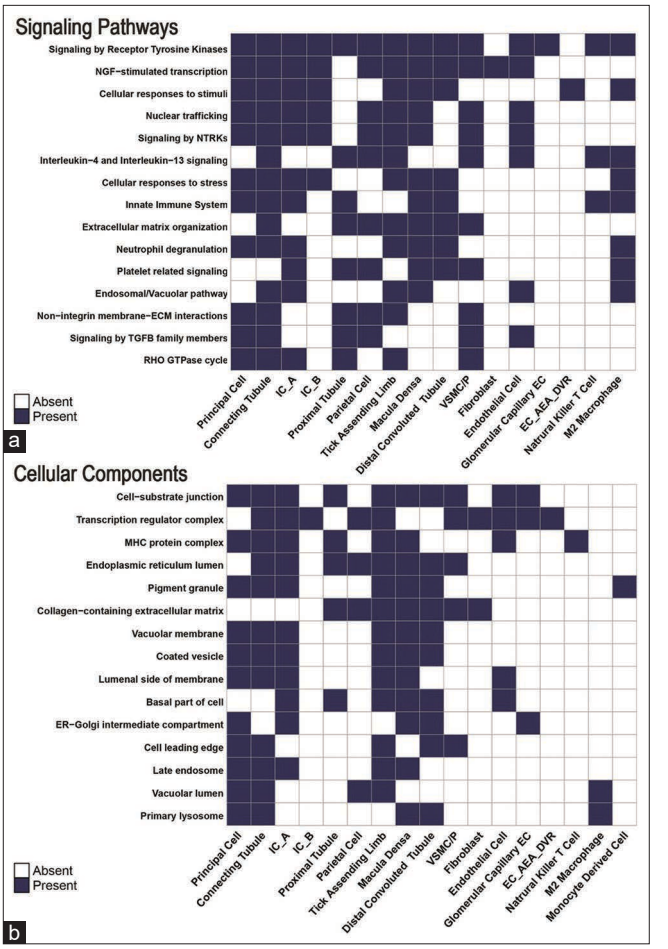


Figure 3: Functional enrichment analysis of human kidney cell types. Binary heatmaps showing the presence (dark blue) or absence (white) of significantly enriched signaling pathways (a) and cellular component GO terms (b) based on DEGs between healthy controls and diabetic kidney disease patients across distinct kidney cell types. Abbreviations: IC, intercalated cell; EC, endothelial cell; VSMC/P, vascular smooth muscle cell/pericyte; EC_AEA_DVR, afferent/efferent arteriole to descending vasa recta endothelial cell

were shared across multiple cell types, whereas 346 were unique to specific cell types [Supplementary Tables 23-39].

To provide a clearer and more informative representation of the molecular programs associated with each cell type, a matrix was constructed with cell types as columns and unique pathways as rows [Supplementary Table 40]. This structure illustrates the frequency and distribution of each pathway across cell types [Figure 3a]. Signaling through nerve growth factor (NGF)-stimulated transcription and its receptors (NTRKs) was enriched in 11 cell types. This widespread enrichment suggests that NGF–NTRK signaling may play a broader role in the diabetic kidney response than previously recognized. Furthermore, pathways involved in extracellular matrix organization, transforming growth factor-β signaling, inflammatory responses, and platelet-related signaling were among the most frequent pathways, consistent with previous studies.[22-25] Taken together, these results highlight both shared and cell-type-specific pathways that may underlie the progression of DKD.

To further investigate the functional relevance of significant transcriptional alterations, gene ontology (GO) cellular component enrichment analysis was performed for each kidney cell type [Supplementary Tables 41-57]. The matrix of the top 15 most frequently enriched terms across cell types [Figure 3b] provided insights into the roles of cell–ECM adhesion, transcriptional regulation, endoplasmic reticulum stress, cell polarization, and secretory signaling in DKD pathogenesis. Notably, most of the frequently enriched terms were observed in the distal parts of the nephron, including the distal convoluted tubule, principal cells, thick ascending limb (TAL), connecting tubule (CT), intercalated cells-A, and macula densa.

Principal cells exhibit the most transcriptional reprogramming in diabetic injury

To identify the most affected cell types in DKD in terms of transcriptional alterations, the number of DEGs was compared across the clusters. Notably, among the annotated kidney-resident cell types, collecting duct principal cells displayed the most extensive transcriptional response, with a markedly greater number of DEGs than other nephron segments. To normalize for differences in cluster size, a DEG-to-cell count ratio was calculated for each cell type. This analysis demonstrated that principal cells had the highest DEG/cell ratio, significantly exceeding those of proximal tubule and TAL cells [Figure 4a].

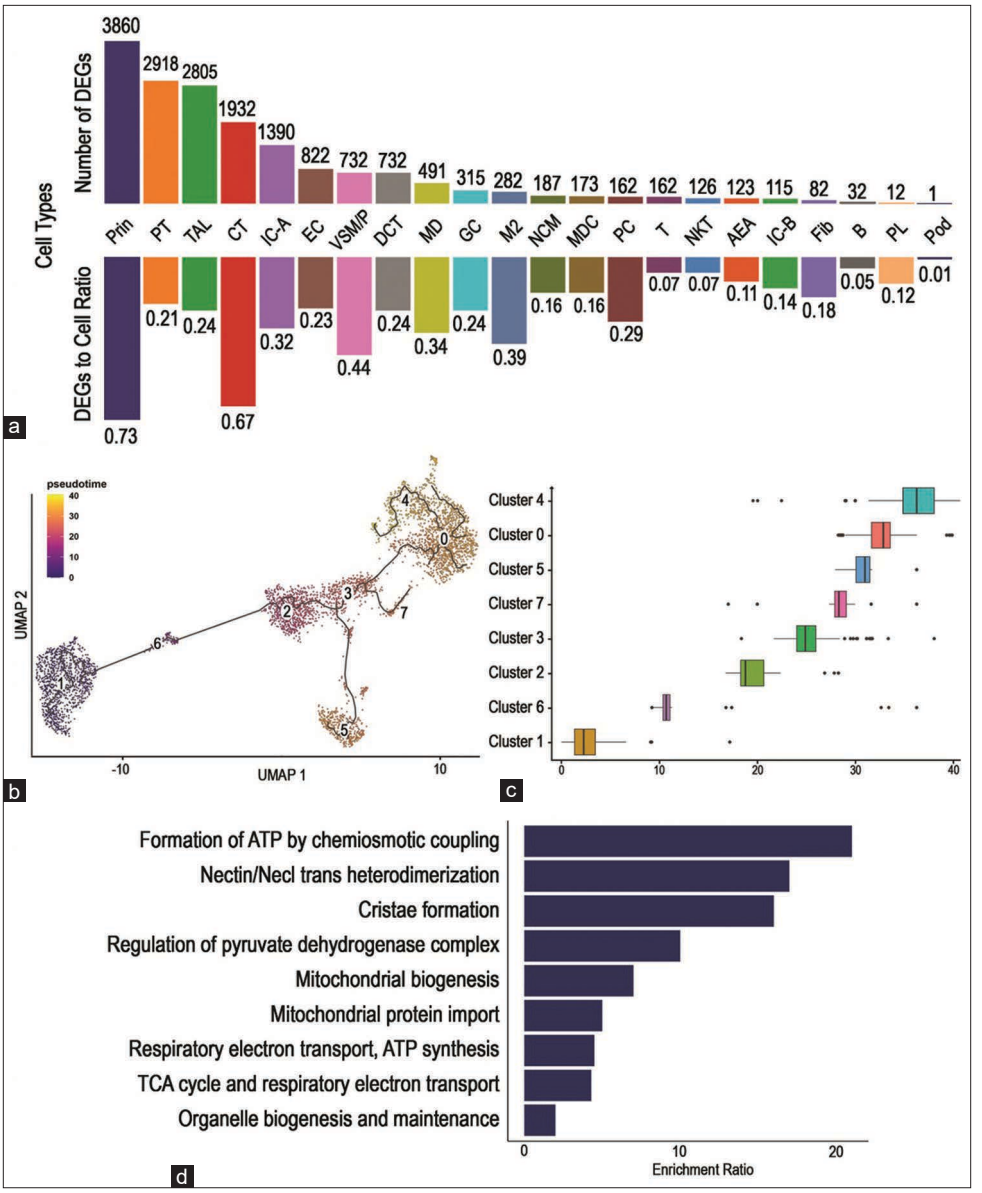


Figure 4: Principal cells undergo the most extensive transcriptional reprogramming in diabetic injury. (a) The number of differentially expressed genes (DEGs, left) and DEG-to-cell ratios (right) across kidney cell types indicate that principal cells exhibit the highest degree of transcriptional change. (b) Pseudotime trajectory of principal cells visualized by UMAP reveals distinct transcriptionally heterogeneous subclusters, with (c) showing the distribution of pseudotime values across these subclusters. (d) Pathway enrichment analysis comparing a disease-associated terminal subpopulation (cluster 5) with the root population (cluster 1) highlights pronounced alterations in metabolic and mitochondrial programs. Abbreviations: Prin, principal cell; PT, proximal tubule; TAL, thick ascending limb; CT, connecting tubule; IC, intercalated cell; EC, endothelial cell; VSM/P, vascular smooth muscle cell/pericyte; DCT, distal convoluted tubule; MD, macula densa; GC, glomerular capillary endothelial cell; M2, M2 macrophage; NCM, nonclassical monocyte; MDC, monocyte-derived cell; PC, parietal cell; T, T cell; NKT, natural killer T cell; AEA, afferent/efferent arteriole to descending vasa recta endothelial cell; Fib, fibroblast; B, B cell; PL, plasma cell; Pod, podocyte; and TCA, tricarboxylic acid

Although pseudobulk methods do not inherently weight the number of individual cells, larger cell numbers increase statistical confidence in aggregate signals, mitigating the risk of missing biological signals. Therefore, the high DEG count observed in principal cells – despite their moderate abundance – suggests a genuine and biologically meaningful transcriptional shift rather than a statistical artifact.

To ensure the reproducibility of the pronounced transcriptional reprogramming observed in principal cells under diabetic conditions, an independent human DKD single-cell RNA-seq dataset (GSE211785) was reanalyzed [Supplementary Figure 5]. Consistent with the original findings, principal cells exhibited the highest DEG burden compared with other nephron segments, such as proximal tubule and TAL cells. Taken together, these results indicate that principal cells may undergo extensive transcriptional reprogramming in response to diabetic injury, a feature that has been underappreciated in prior studies.

To further investigate the heterogeneity of this response, pseudotime trajectory analysis of principal cells was conducted. Prior to trajectory inference, batch effects among principal cells were assessed and corrected [Supplementary Figure 4]. This analysis resolved eight transcriptional subpopulations arranged along a branched trajectory, with pseudotime values ranging from 0 to 40, indicating distinct segregation of the subpopulations [Figure 4b and c]. Notably, one terminal cluster (Cluster 5) was composed exclusively of cells from DKD samples [Supplementary Figure 6], suggesting the emergence of a disease-associated transcriptional state. To shed light on the molecular features underlying this state, pseudobulk DEG analysis was performed between Cluster 1 (located at the trajectory root) and Cluster 5 [Supplementary Table 58-60]. Differential genes with FDR ≤ 0.05 and absolute $\log_2$ (FC) ≥1 were subjected to pathway enrichment analysis, which revealed strong associations with mitochondrial pathways [Figure 4d]. This observation suggests that mitochondrial reprogramming is a fundamental component of the transcriptional response of principal cells to diabetic stress.

Identification of novel therapeutic targets paves the way for targeted drug delivery to principal cells

Although omics data provide global and novel insights into the molecular landscape of diseases, identifying the most appropriate therapeutic interventions remains a significant challenge.[26] To address this gap, a tool named DrugTar was previously developed by our team.[27] This algorithm employs a robust deep learning framework to predict druggability by integrating protein sequence embeddings from a pretrained language model with functional insights derived from GO annotations. On the basis of their potential as drug targets, each gene can be assigned a druggability score.

In this study, DrugTar was applied to identify potential drug targets in principal cells for DKD, yielding a list of highly druggable genes [Supplementary Table 61]. Among the top candidates are IDH1, KCNJ8, LRRK2, SOD1, INSR, and PPARG. The therapeutic potential of targeting PPARG is well established; its agonists, including thiazolidinediones, have demonstrated renoprotective effects in both clinical and preclinical studies.[28,29] Previous studies have also repeatedly reported the contribution of SOD1 in CKD pathogenesis. For instance, renal SOD1-deficient C57BL/6-Akita mice exhibited an impaired ability to eliminate excessive glomerular and renal superoxide produced under hyperglycemic conditions, leading to the development of diabetic kidney injury.[30]

Although IDH1, KCNJ8, LRRK2, and INSR have not yet been targeted in DKD, they appear to be mechanistically relevant through their involvement in biological pathways such as insulin signaling, metabolic regulation, and cellular stress responses.[31-34] The identification of these druggable genes underscores oxidative stress (SOD1), inflammation and lipid metabolism (PPARG), and metabolic dysfunction (INSR) as key actionable processes in DKD pathogenesis.[35-37] While SOD1 and PPARG offer immediate translational potential, the remaining candidates require further investigation in future interventional studies. Therefore, this study provides a basis for further assessment and validation of novel therapeutic targets for targeted drug delivery to principal cells.

DISCUSSION

This study identified collecting duct principal cells as one of the most transcriptionally responsive kidney cell types under diabetic conditions, suggesting a previously underappreciated role in DKD pathogenesis. Most prior studies have focused on endothelial cells, proximal tubule cells, and mesangial cells, while principal cells have received limited attention.[38-40] To our knowledge, direct investigations into their role in DKD are rare, with only one study using mouse scRNA-seq that reported mitochondrial heterogeneity involving principal and B cells.[41] This research gap may not be surprising, as most existing studies on DKD have been designed around predefined hypotheses or have relied on bulk omics approaches, which often mask cell type-specific signals. Single-cell technologies now offer the opportunity to uncover previously hidden layers of biological complexity. As such, these findings may represent a starting point for uncovering the potential contribution of principal cells to the pathogenesis of DKD.

This broad range of transcriptional alterations observed in principal cells may be explained by the multifaceted roles these cells play in kidney physiology. Principal cells play a critical role in fine-tuning renal output by regulating the urine concentration and maintaining electrolyte balance via specific transporters for water and electrolytes, including the aquaporin-2 (AQP2) water channel, the epithelial Na + channel, and the renal outer medullary K + channel.[42] These functions are tightly regulated through sophisticated gene regulatory circuits induced by a variety of hormones, such as vasopressin,[43] renin,[44] aldosterone,[45] insulin,[46] and atrial natriuretic peptide.[47] Principal cells are also responsive to a spectrum of other regulators, including the autonomous nervous system,[48] prostaglandins,[49] adenosine,[50] nitric oxide,[51] endothelin,[52] and hemodynamic factors such as tubular fluid flow and osmolarity.[53,54] This complex integration of hormonal cues, electrolyte transport, and epithelial integrity renders them sensitive to metabolic and osmotic challenges. Notably, these extensive responses and transport activities demand high energy consumption.[55] The substantial dependency of principal cells on mitochondrial stability supports

our finding of enrichment of respiration and adenosine triphosphate-generation pathways, particularly between the DKD-specific branch and the root branch in the trajectory analysis. These physiological characteristics suggest that metabolic and osmotic stress in diabetes may disrupt principal cell homeostasis, potentially contributing to impaired urine concentrating ability, electrolyte imbalance, and tubular remodeling during DKD progression.

While previous studies have reported that the high energetic demand of proximal tubule cells under diabetic conditions may come at a heavy cost, manifested as prolonged mitochondrial stress and associated oxidative damage in CKD,[56-59] our findings extend this understanding to principal cells. The altered mitochondrial activity observed in these cells may contribute to the growing body of evidence suggesting that mitochondrial dysfunction is a shared, although heterogeneous, feature across multiple nephron segments in diabetes as supported by pathway enrichment and trajectory analysis.

The distinct embryonic origin of CT principal cells from the ureteric bud, unlike most other nephron segments, which originate from the metanephric mesenchyme, may partly explain the observed differences in response profiles and transcriptional reprogramming.[60-62] This developmental distinction may also contribute to understanding gender-dependent responses in kidney physiology and pathology.

In addition, the current study identified a set of druggable targets using a deep learning–based tool previously developed by our team to facilitate the translation of big data into clinical knowledge. The high-score targets, which appear to contribute to DKD pathogenesis based on their roles in cell-specific functions and pathways described earlier, could be used for targeted delivery to principal cells.

We appreciate that the conclusions drawn here are based on transcriptomic profiles, which may not fully reflect other layers of gene regulation. Single-cell multiomics approaches, followed by functional studies, will be required to illuminate the unrecognized complexity of principal cells. Similarly, the therapeutic potential of the proposed top-ranked targets depends on validation in relevant DKD models. In addition, DrugTar predictions are derived from transcriptomic features and functional annotations and therefore require protein-level validation and experimental confirmation before therapeutic translation.

CONCLUSION

Collecting duct principal cells were identified as one of the most transcriptionally reprogrammed renal cell populations under diabetic conditions, revealing a previously underrecognized dimension of DKD pathophysiology. Trajectory and pathway analyses indicated pronounced mitochondrial and metabolic alterations, extending the concept of energy stress beyond proximal tubule cells. By integrating single-cell transcriptomics with computational druggability prediction, a set of high-priority therapeutic candidates was delineated, providing a rational framework for cell type–specific intervention strategies. Although transcript-level findings require functional validation, these results position principal cells as potential keystone contributors to DKD progression and as promising targets for precision therapeutic development.

Financial support and sponsorship

This study was supported by Isfahan University of Medical Sciences,Isfahan, Iran (Grant no. 3402457).

Conflicts of interest

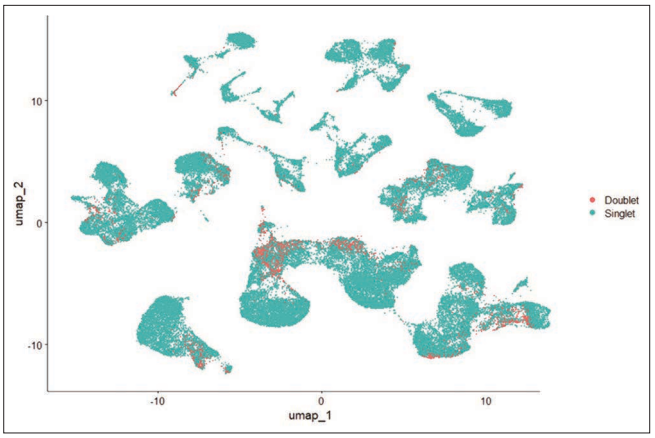
There are no conflicts of interest.

REFERENCES

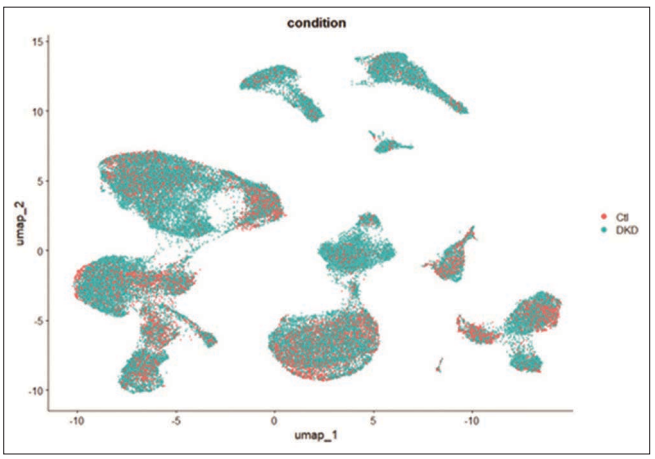
1. Ong KL, Stafford LK, McLaughlin SA, Boyko EJ, Vollset SE, Smith AE, *et al*. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: A systematic analysis for the global burden of disease study 2021. Lancet 2023;402:203-34.
2. Fu H, Liu S, Bastacky SI, Wang X, Tian XJ, Zhou D. Diabetic kidney diseases revisited: A new perspective for a new era. Mol Metab 2019;30:250-63.
3. Hoogeveen EK. The epidemiology of diabetic kidney disease. Kidney Dial 2022;2:433-42.
4. Shetty S, Suvarna R, Awasthi A, Bhojaraja MV, Pappachan JM. Emerging biomarkers and innovative therapeutic strategies in diabetic kidney disease: A pathway to precision medicine. Diagnostics (Basel) 2025;15:973.
5. Balzer MS, Rohacs T, Susztak K. How many cell types are in the kidney and what do they do? Annu Rev Physiol 2022;84:507-31.
6. Lim J, Park C, Kim M, Kim H, Kim J, Lee DS. Advances in single-cell omics and multiomics for high-resolution molecular profiling. Exp Mol Med 2024;56:515-26.
7. Delrue C, Speeckaert MM. Decoding kidney pathophysiology: Omics-driven approaches in precision medicine. J Pers Med 2024;14:1157.
8. Schreibing F, Kramann R. Mapping the human kidney using single-cell genomics. Nat Rev Nephrol 2022;18:347-60.
9. Abedini A, Levinsohn J, Klötzer KA, Dumoulin B, Ma Z, Frederick J, *et al*. Single-cell multi-omic and spatial profiling of human kidneys implicates the fibrotic microenvironment in kidney disease progression. Nat Genet 2024;56:1712-24.
10. Bhadani AK, Jothimani D. Big Data: Challenges, Opportunities, and Realities. In: Singh MK, G DK, editors. Effective Big Data Management and Opportunities for Implementation. Hershey, PA, USA: IGI Global Scientific Publishing; 2016. p. 1-24.
11. Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, *et al*. Big data: Astronomical or genomical? PLoS Biol 2015;13:e1002195.
12. Fan J, Han F, Liu H. Challenges of big data analysis. Natl Sci Rev 2014;1:293-314.
13. Lake BB, Menon R, Winfree S, Hu Q, Melo Ferreira R, Kalhor K,

et al. An atlas of healthy and injured cell states and niches in the human kidney. Nature 2023;619:585-94.
14. Klötzer KA, Abedini A, Li S, Balzer MS, Liang X, Levinsohn J, *et al*. Analysis of individual patient pathway coordination in a cross-species single-cell kidney atlas. Nat Genet 2025;57:1922-34.
15. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol 2015;33:495-502.
16. Chan BK. Data analysis using R programming. Adv Exp Med Biol 2018;1082:47-122.
17. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. Cell Syst 2019;8:329-37.e4.
18. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, *et al*. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 2019;16:1289-96.
19. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
20. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, *et al*. The single-cell transcriptional landscape of mammalian organogenesis. Nature 2019;566:496-502.
21. Schumacher A, Rookmaaker MB, Joles JA, Kramann R, Nguyen TQ, van Griensven M, *et al*. Defining the variety of cell types in developing and adult human kidneys by single-cell RNA sequencing. NPJ Regen Med 2021;6:45.
22. Ziyadeh FN. The extracellular matrix in diabetic nephropathy. Am J Kidney Dis 1993;22:736-44.
23. Wang L, Wang HL, Liu TT, Lan HY. TGF-beta as a master regulator of diabetic nephropathy. Int J Mol Sci 2021;22:7881.
24. Rustiasari UJ, Roelofs JJ. The role of platelets in diabetic kidney disease. Int J Mol Sci 2022;23:8270.
25. Pérez-Morales RE, Del Pino MD, Valdivielso JM, Ortiz A, Mora-Fernández C, Navarro-González JF. Inflammation in diabetic kidney disease. Nephron 2019;143:12-6.
26. Tanoli Z, Alam Z, Ianevski A, Wennerberg K, Vähä-Koskela M, Aittokallio T. Interactive visual analysis of drug-target interaction networks using drug target profiler, with applications to precision medicine and drug repurposing. Brief Bioinform 2020;21:211-20.
27. Borhani N, Izadi I, Motahharynia A, Sheikholeslami M, Gheisari Y. DrugTar improves druggability prediction by integrating large language models and gene ontologies. Bioinformatics 2025;41:btaf360.
28. Sarafidis PA, Bakris GL. Protection of the kidney by thiazolidinediones: An assessment from bench to bedside. Kidney Int 2006;70:1223-33.
29. Zheng F, Guan Y. Thiazolidinediones: A novel class of drugs for the prevention of diabetic nephropathy? Kidney Int 2007;72:1301-3.
30. Fujita H, Fujishima H, Takahashi K, Sato T, Shimizu T, Morii T, *et al*. SOD1, but not SOD3, deficiency accelerates diabetic renal injury in C57BL/6-Ins2(Akita) diabetic mice. Metabolism 2012;61:1714-24.
31. Lai K, Chen Z, Lin S, Ye K, Yuan Y, Li G, *et al*. The IDH1-R132H mutation aggravates cisplatin-induced acute kidney injury by promoting ferroptosis through disrupting NDUFA1 and FSP1 interaction. Cell Death Differ 2025;32:242-55.
32. Guo J, Zhang C, Zhao H, Yan Y, Liu Z. The key mediator of diabetic kidney disease: Potassium channel dysfunction. Genes Dis 2024;11:101119.
33. Wang JP, Yan JP, Xiao RL, Li RS. Leucine-rich repeat kinase 2 is protective during acute kidney injury through its activation of autophagy in podocytes. Environ Toxicol 2022;37:2947-56.
34. Daza-Arnedo R, Rico-Fontalvo J, Aroca-Martínez G, Rodríguez-Yanez T, Martínez-Ávila MC, Almanza-Hurtado A, *et al*. Insulin and the kidneys: A contemporary view on the molecular basis. Front Nephrol 2023;3:1133352.
35. Jha JC, Banal C, Chow BS, Cooper ME, Jandeleit-Dahm K. Diabetes and kidney disease: Role of oxidative stress. Antioxid Redox Signal 2016;25:657-84.
36. Stadler K, Goldberg IJ, Susztak K. The evolving understanding of the contribution of lipid metabolism to diabetic kidney disease. Curr Diab Rep 2015;15:40.
37. Hashimoto S, Maoka T, Kawata T, Mochizuki T, Koike T, Shigematsu T. Roles of insulin receptor substrates (IRS) in renal function and renal hemodynamics. PLoS One 2020;15:e0242332.
38. Maezawa Y, Takemoto M, Yokote K. Cell biology of diabetic nephropathy: Roles of endothelial cells, tubulointerstitial cells and podocytes. J Diabetes Investig 2015;6:3-15.
39. Lay AC, Tran VD, Nair V, Betin V, Hurcombe JA, Barrington AF, *et al*. Profiling of insulin-resistant kidney models and human biopsies reveals common and cell-type-specific mechanisms underpinning Diabetic Kidney Disease. Nat Commun 2024;15:10018.
40. Efiong EE, Maedler K, Effa E, Osuagwu UL, Peters E, Ikebiuro JO, *et al*. Decoding diabetic kidney disease: A comprehensive review of interconnected pathways, molecular mediators, and therapeutic insights. Diabetol Metab Syndr 2025;17:192.
41. Wu C, Song Y, Yu Y, Xu Q, Cui X, Wang Y, *et al*. Single-cell transcriptional landscape reveals the regulatory network and its heterogeneity of renal mitochondrial damages in diabetic kidney disease. Int J Mol Sci 2023;24:13502.
42. Pearce D, Soundararajan R, Trimpert C, Kashlan OB, Deen PM, Kohan DE. Collecting duct principal cell transport processes and their regulation. Clin J Am Soc Nephrol 2015;10:135-46.
43. Bankir L, Bichet DG, Bouby N. Vasopressin V2 receptors, ENaC, and sodium reabsorption: A risk factor for hypertension? Am J Physiol Renal Physiol 2010;299:F917-28.
44. Prieto MC, Gonzalez AA, Visniauskas B, Navar LG. The evolving complexity of the collecting duct renin-angiotensin system in hypertension. Nat Rev Nephrol 2021;17:481-92.
45. Funder JW. Aldosterone and Mineralocorticoid Receptors-Physiology and Pathophysiology. Int J Mol Sci 2017;18.
46. Zaika O, Palygin O, Tomilin V, Mamenko M, Staruschenko A, Pochynyuk O. Insulin and IGF-1 activate Kir4.1/5.1 channels in cortical collecting duct principal cells to control basolateral membrane voltage. Am J Physiol Renal Physiol 2016;310:F311-21.
47. Klokkers J, Langehanenberg P, Kemper B, Kosmeier S, von Bally G, Riethmüller C, *et al*. Atrial natriuretic peptide and nitric oxide signaling antagonizes vasopressin-mediated water permeability in inner medullary collecting duct cells. Am J Physiol Renal Physiol 2009;297:F693-703.
48. Loesch A, Unwin R, Gandhi V, Burnstock G. Sympathetic nerve varicosities in close apposition to basolateral membranes of collecting duct epithelial cells of rat kidney. Nephron Physiol 2009;113:p15-21.
49. Breyer MD, Breyer RM. Prostaglandin E receptors and the kidney. Am J Physiol Renal Physiol 2000;279:F12-23.
50. Rieg T, Vallon V. ATP and adenosine in the local regulation of water transport and homeostasis by the kidney. Am J Physiol Regul Integr Comp Physiol 2009;296:R419-27.
51. Hyndman KA, Boesen EI, Elmarakby AA, Brands MW, Huang P, Kohan DE, *et al*. Renal collecting duct NOS1 maintains fluid-electrolyte homeostasis and blood pressure. Hypertension 2013;62:91-8.
52. Lyon-Roberts B, Strait KA, van Peursem E, Kittikulsuth W, Pollock JS, Pollock DM, *et al*. Flow regulation of collecting duct endothelin-1 production. Am J Physiol Renal Physiol 2011;300:F650-6.

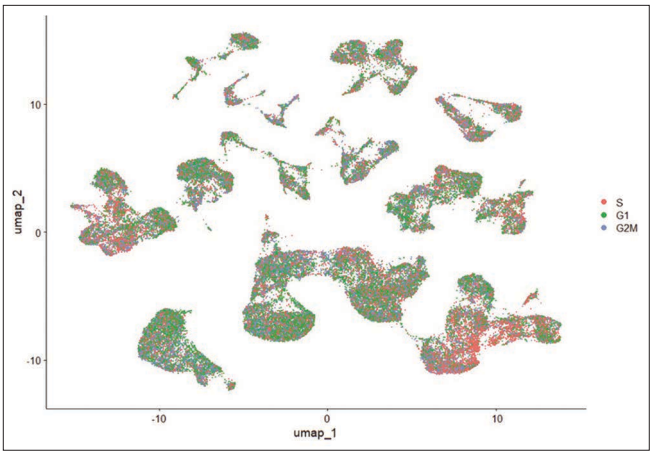
53. Kashlan OB, Kleyman TR. Epithelial Na(+) channel regulation by cytoplasmic and extracellular factors. Exp Cell Res 2012;318:1011-9.
54. Sun P, Liu W, Lin DH, Yue P, Kemp R, Satlin LM, *et al*. Epoxyeicosatrienoic acid activates BK channels in the cortical collecting duct. J Am Soc Nephrol 2009;20:513-23.
55. Carlson BM. Chapter 13 - The Urinary System. In: The Human Body. London: Academic Press; 2019. p. 357–72.
56. Kim K, Lee EY. Excessively enlarged mitochondria in the kidneys of diabetic nephropathy. Antioxidants (Basel) 2021;10:741.
57. Ni L, Yuan C. The mitochondrial-associated endoplasmic reticulum membrane and its role in diabetic nephropathy. Oxid Med Cell Longev 2021;2021:8054817.
58. Mise K, Galvan DL, Danesh FR. Shaping up mitochondria in diabetic nephropathy. Kidney360 2020;1:982-92.
59. Wu W, Jin Y, Teng L, Shao X, Wang Y, Feng S, *et al*. Mitochondria-related reversal of early-stage diabetic nephropathy in donor kidney after transplantation in mice. Ann Transl Med 2019;7:801.
60. Little MH, McMahon AP. Mammalian kidney development: Principles, progress, and projections. Cold Spring Harb Perspect Biol 2012;4:a008300.
61. Clotet-Freixas S, Zaslaver O, Kotlyar M, Pastrello C, Quaile AT, McEvoy CM, *et al*. Sex differences in kidney metabolism may reflect sex-dependent outcomes in human diabetic kidney disease. Sci Transl Med 2024;16:eabm2090.
62. Neugarten J, Golestaneh L. Gender and the prevalence and progression of renal disease. Adv Chronic Kidney Dis 2013;20:390-5.



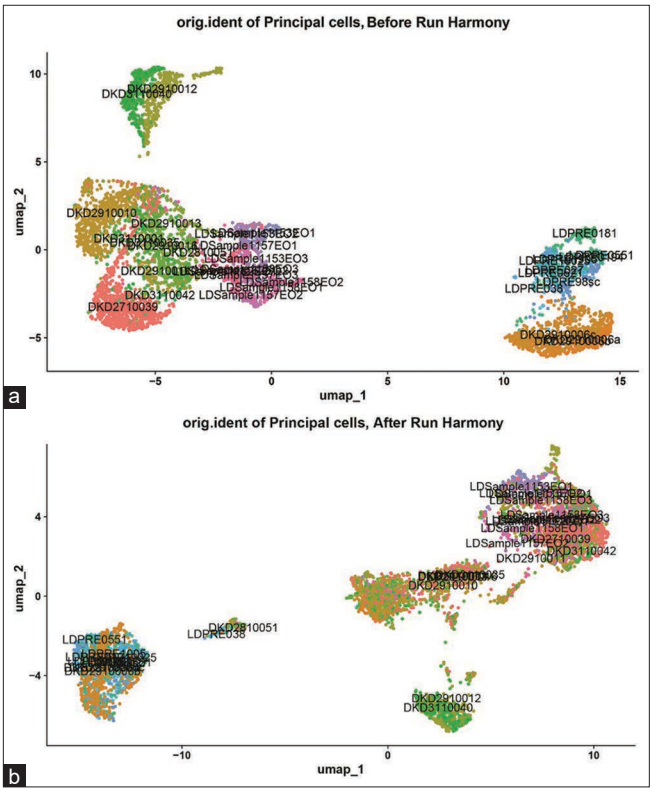
Supplementary Figure 1: Doublet identification and removal. Uniform manifold approximation and projection visualization of single-cell RNA-seq data from human kidney samples, where each point represents an individual cell positioned based on transcriptomic similarity. Singlet cells are shown in cyan, and doublets are highlighted in red. UMAP = Uniform manifold approximation and projection



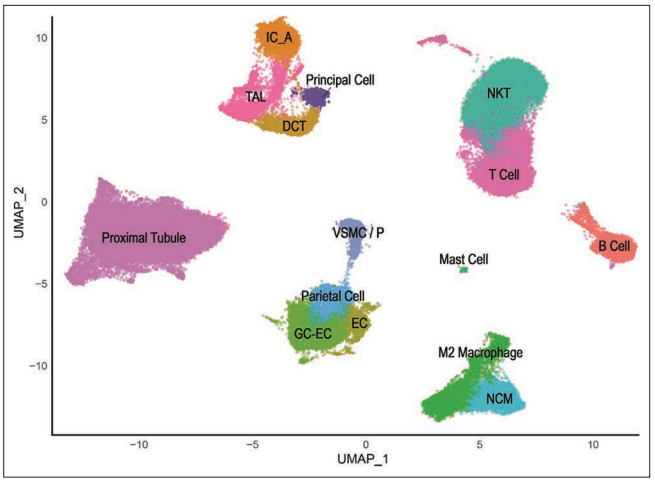
Supplementary Figure 3: Sample-level batch effect correction. The uniform manifold approximation and projection embedding of single-cell transcriptomic data from human kidney samples, the color coding indicates a mixture of cells from diabetic kidney disease (DKD) and healthy controls within the same clusters, confirming that clustering reflects physiological differences rather than disease condition or sample-specific biases. UMAP = Uniform manifold approximation and projection; DKD = Diabetic kidney disease



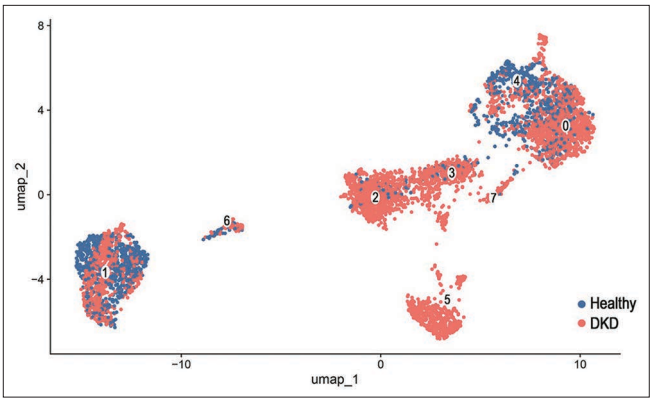
Supplementary Figure 2: Assessment of cell cycle phase effects. Uniform manifold approximation and projection embedding of single-cell transcriptomic data from human kidney samples showing the distribution of cells across cell cycle phases. The color coding reveals a mixture of cells in S, G1, and G2/M phases within the same clusters, confirming that clustering reflects physiological differences rather than biases from cell cycle phase. UMAP = Uniform manifold approximation and projection



Supplementary Figure 4: Batch effect correction among principal cells. Uniform manifold approximation and projection embedding of single-cell transcriptomic data from principal cells, a subset of the 22 transcriptionally distinct populations identified previously.(a) Before batch effect correction, cells are color-coded by sample origin, revealing batch effects as clustering is driven by sample differences rather than physiological state.(b) After batch effect correction, cells are color-coded by sample origin, showing integrated clustering of diabetic kidney disease (DKD) and living donor (LD) samples, indicating successful mitigation of batch effects. UMAP = Uniform manifold approximation and projection; DKD = Diabetic kidney disease; LD = living donor



Supplementary Figure 5: Annotation of cell types in human kidney samples from GSE211785. A total of 71,953 cells from healthy donors and diabetic kidney disease (DKD) patients were projected using uniform manifold approximation and projection (UMAP) and classified into 15 distinct cell types based on kidney-specific markers curated by the Kidney Precision Medicine Project (KPMP). IC = Intercalated cell; VSMC/P = Vascular smooth muscle cell/pericyte; DCT = Distal Convoluted tubule; GC-EC = Glomerular capillary endothelial cell; NCM = Nonclassical monocyte; NKT = Natural killer T cell; UMAP = Uniform manifold approximation and projection



Supplementary Figure 6: Condition distribution among principal cell subpopulations. uniform manifold approximation and projection embedding of principal cells from healthy (blue) and diabetic kidney disease (DKD) (red) samples; points are colored by sample condition and labeled by cluster (0–7). The spatial distribution of conditions across the embedding shows that cluster 5 is composed exclusively of DKD-derived cells, consistent with a disease-associated principal cell subpopulation. UMAP = Uniform manifold approximation and projection; DKD = Diabetic kidney disease