Not all controls are made equal: Definition of human kidney reference samples by single cell gene expression profiles

Rajasree Menon[1]*, Paul L. Kimmel[2], Edgar A. Otto[3], Lalita Subramanian[3], Celine C. Berthier[3], Christopher L. O' Connor[3], Bradley Godfrey[3], Abhijit S. Naik[3], Minnie Sarwal[4], E. Steve Woodle[5], Laura Pyle[6], Ye Ji Choi[6], Patricia Ladd[7], John R. Sedor[8], Syvia E. Rosas[9], Sushrut S. Waikar[10], Markus Bitzer[3], Petter Bjornstad[6], Jeffrey B. Hodgin[11], Matthias Kretzler[1,3]*, for the Kidney Precision Medicine Project (KPMP)

Affiliations:

[1]Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI

[2]Department of Medicine, George Washington University, Washington DC

[3]Department of Internal Medicine, Division of Nephrology, University of Michigan Medical School, Ann Arbor, MI

[4]Division of Transplant Surgery, School of Medicine, University of California, San Francisco, CA

[5]Department of Surgery, University of Cincinnati College of Medicine, Cincinnati, OH

[6]Departments of Pediatrics and Medicine, Division of Metabolism, Endocrinology and Nutrition, University of Washington, Seattle, WA

[7]Department of Radiology, Division of Renal Diseases and Hypertension, University of Colorado School of Medicine, Aurora, CO

[8]Medical Specialties and Lerner Research Institutes, Cleveland Clinic and Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH

[9]Kidney and Hypertension unit, Joslin Diabetes Center and Harvard Medical School, Boston, MA

[10]Section of Nephrology, Department of Medicine, Boston Medical Center and Boston University Chobanian & Avedisian School of Medicine, Boston, MA

[11]Department of Pathology, University of Michigan, Ann Arbor, MI.

***Corresponding authors:**

Rajasree Menon, Ph.D.
Department of Computational Medicine and Bioinformatics
University of Michigan
4544E MSRB2
1150 Medical Center Dr
Ann Arbor, MI 48109-2218
rajmenon@med.umich.edu

Matthias Kretzler, M.D.
Warner-Lambert/Parke-Davis Professor
Medicine/Nephrology and Computational Medicine and Bioinformatics
University of Michigan
4544D MSRB2
1150 Medical Center Dr
Ann Arbor, MI 48109-2218
kretzler@med.umich.edu

ABSTRACT

Identifying kidney disease mechanisms often requires comparing samples from disease states with healthy reference tissues. However, the effect of variations in sample procurement, storage and donor baseline characteristics of reference samples has thus far not been evaluated. Three distinct kidney reference sample types were evaluated for integrity and injury biomarkers and in their ability to define differentially expressed genes (DEGs) when compared to three different diabetic kidney disease (DKD) states. Unaffected parts of tumor nephrectomies (TN), pre-transplant living donor biopsies (LD), and percutaneous kidney research biopsies from healthy volunteers (HC) served as sources for reference tissue. Single cell gene expression profiles showed differences in the expression of injury or disease markers and the proportion of immune and proximal cell states. TN exhibited the highest expression of early stress response genes. A gene set associated with procurement effect in post-operative biopsies (LD and TN) was identified. An age-associated transcriptional signature was extracted from the reference data. Providing these tools to control for age and tissue procurement effects, immune-related pathways were found to be most enriched in DKD when compared to HC. Energy-related processes were enriched in DEGs from DKD versus LD. TN samples exhibited more underlying pathology than LD. The pathway analyses using the DEGs underscore the importance of accounting for appropriate confounding factors in differential expression analyses between disease and reference samples. Comparable controls are essential for appropriate molecular evaluation of pathologic tissues.

TRANSLATIONAL STATEMENT

Integrated single-cell data analysis of three reference sample types—needle biopsy from young healthy kidney tissue, pre-perfusion biopsy from transplant kidneys, and cancer-free tissue from tumor-nephrectomies—revealed distinct transcriptional profiles influenced by the biopsy procurement method and age. These differences impacted findings in diabetes-related kidney disease versus reference comparisons highlighting the need and providing the tools to account for these differences in interpreting analyses and identifying disease mechanisms.

INTRODUCTION

Due to the highly heterogeneous nature of kidney diseases, our understanding of the disease pathogenesis remains incomplete. Advances in genomics technologies have facilitated the unveiling of molecular mechanisms underlying kidney diseases[1-4]. However, these findings have relied heavily on comparisons between disease and healthy reference tissue. Given the challenges of obtaining disease-free kidney tissue from healthy humans, a wide variety of sources for reference tissue have been used[5]. Further, factors such as the sample procurement process, preservation method, tissue processing technology, and the age and underlying physiological condition of the tissue donor could impact the quality of the reference tissue and the corresponding gene expression profiles[6].

With the rapid development of molecular analysis tools, understanding the advantages and limitations of different reference tissue sources is critical[5]. In this study, we compared single-cell gene expression data from three reference tissue sources: Percutaneous kidney research biopsies from healthy controls (HC), unaffected parts of tumor-nephrectomies (TN), and living kidney donor biopsies (LD).

We utilized single-cell gene expression data of kidney biopsy samples from patients with type 2 diabetes with diabetic kidney disease (DKD), patients with type 1 diabetes for > 25 years with no evidence of kidney disease referred henceforth as diabetes mellitus resilient (DMR), and youth with early onset type 2 diabetes and kidney disease (early DKD) as disease comparators to determine how choice of reference type could impact identification of underlying molecular mechanisms in disease progression (Figure 1).

MATERIALS & METHODS

Human data

Key clinical and demographic characteristics of participants who contributed kidney biopsy samples in the three reference and three diabetes-related groups are summarized in Table 1. For this study, we have used data from samples collected by the Kidney Precision Medicine Project (KPMP), which focuses on molecular mechanisms in acute kidney injury (AKI) and chronic kidney disease (CKD)[7]. From the KPMP cohort, nine LD, twenty-one DKD, and six DMR biopsy samples were used in this study. LD and TN samples were obtained through intra- or post-operative surgical biopsies, whereas healthy control samples (HC), early DKD, and DKD samples were collected via percutaneous needle biopsies.

The TN samples comprised tumor-free kidney cortical tissue from nephrectomy specimens (PRECISE cohort, University of Michigan). The 9 samples were selected from donors with high eGFR (>80), low pathology, no diabetes and no hypertension.

The 9 early DKD samples were from the Impact of Metabolic Surgery on Pancreatic, Renal and Cardiovascular Health in Youth with Type 2 Diabetes (IMPROVE-T2D, NCT03620773) study and the Renal Hemodynamics, Energetics and Insulin Resistance in Youth Onset Type 2 Diabetes Study (RENAL HEIR, NCT03584217) and 12 HC from the Control of Renal Oxygen Consumption, Mitochondrial Dysfunction, and Insulin Resistance (CROCODILE) study (NCT04074668).

For identifying the genes associated with age using weighted correlation network analysis, (WGCNA)[8], we used data from LD. In addition to the 9 KPMP LD samples, we included 13 samples from the University of Michigan living donor cohort generated for the Human Kidney Transplant Transcriptomic Atlas.

After procurement with identical protocols (**dx.doi.org/10.17504/protocols.io.7dthi6n**) all samples were processed by the KPMP **Pre**cision **M**edicine Through **I**nt**e**rrogation of **R**NA in the Kidn**e**y (PREMIERE) tissue interrogation site personnel and core facility staff.

Single cell data generation

Single cell dissociation of the samples was performed using Liberase TL at 37˚C. Detailed sample processing protocol can be found in: dx.doi.org/10.17504/protocols.io.7dthi6n. The single cell suspension was immediately transferred to the University of Michigan Advanced Genomics Core facility for further processing. 10X Genomics single cell RNA sequencing (scRNAseq) of the samples was done according to dx.doi.org/10.17504/protocols.io. The raw sequences files were aligned to the reference genome (GRCh38 (hg38)) and the feature-barcode matrices were generated by Cell Ranger program embedded in 10X Genomics analysis software.

Single cell data analysis

The read count matrices were initially processed using SoupX (v1.5.0) to correct for ambient mRNA contamination. Data processing was performed using the Seurat R package (version 5.0). Quality control was enforced with cutoffs of >500 and <5000 genes per cell, and <50% mitochondrial reads per cell. Each sample group underwent normalization, variable gene identification, and principal component analysis independently. Subsequently, the integration of all six sample types was achieved via reciprocal principal component analysis (RPCA). Dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP) and unsupervised clustering (resolution 0.4) were conducted on the integrated dataset. Cell clusters were annotated based on markers of major nephron segments, as well as endothelial, interstitial, and immune cell types. Further analyses were performed on a down-sampled integrated object containing 20,000 randomly selected cells from each sample type (the integrated dataset containing a total of 120,000 cells from all 6 sample types together).

Trajectory analysis

To illustrate the separation of cell states in proximal cells, we utilized the Slingshot R package[9] to perform a trajectory/pseudotime analysis. This method orders cells along a trajectory or lineage, with each cell receiving a specific pseudotime value that indicates its position along that trajectory. The Slingshot R package, designed for modeling developmental trajectories in single-cell RNA sequencing data, incorporates prior knowledge through supervised graph construction[9]. For our analysis, we used Seurat's Uniform Manifold Approximation and Projection (UMAP) embeddings to construct the pseudotime trajectories.

Post-operative tissue procurement effect analysis

The gene set upregulated in post-operative sample types (LD and TN) compared to other groups (HC, DMR, early DKD, and DKD) was identified via differential expression analysis using DESeq2[10] on pseudobulk RNA counts. The final gene set was selected after manually checking for overexpression across all samples in LD and TN groups to avoid sample-to-sample variation. The interaction network of this gene set was constructed using STRING [11] and visualized as a network in Cytoscape[12]. Additionally, gene ontology biological processes enrichment analysis for this gene set was conducted using the ClusterProfiler R package[13].

Age effect

WGCNA[8] on sample-level pseudo bulk read counts from the 21 LD samples were used to generate gene modules. Genes from all modules showing positive or negative correlations (correlation > 0.5) with age were compiled into a positive and negative age correlated gene set. For the positive and negative age correlated gene sets, we calculated an 'Age_gene_Score' using the 'AddModuleScore' function in the integrated Seurat object.

Adjusting for procurement and age effect in differential expression analyses

We conducted differential expression analyses between disease and reference types using the DESeq2 R package on pseudo-bulk mRNA counts to identify disease-associated genes. The pseudo-bulk counts were generated by aggregating raw read counts from each sample. We performed differential expression analyses in three ways: without adjusting for confounding factors, adjusting for age, and adjusting for both age and procurement effect. Each of the three disease groups was compared separately with the three reference groups. To include the procurement effect as a covariate, we calculated a score based on the expression of 25 genes positively associated with post-operative tissue procurement. These scores were generated by creating Z-scores from the RNA expression of the genes and then using the average Z-score as the composite score.

Functional annotation

We utilized Qiagen Ingenuity Pathway Analysis (IPA) [14] software to compare the canonical pathways enriched in the significantly differentially expressed genes (DEGs) (adjusted p-value < 0.05) from disease versus reference type analyses. Additionally, using decoupleR[15] directly on the integrated single cell dataset, we inferred the enrichment of pathway activities in all 6 sample types.

RESULTS

Single cell analysis

In our down-sampled integrated single-cell dataset, comprising 120,000 cells from 67 samples across 6 sample types, we annotated 25 distinct cell types (Figure 2a). Within these, we identified two cell states for proximal tubule cells—healthy and adaptive/maladaptive (PT and aPT). The aPT state was specifically marked by the over-expression of vascular cell adhesion molecule 1 (VCAM1), a known indicator of adaptive/maladaptive proximal cells[1]. For the thick ascending loop of Henle cells (TAL), we identified three states: TAL1, TAL2, and aTAL, with the aTAL state being characterized by high expression of prominin 1 (PROM1) [1]. Importantly, no batch effect was observed in the integrated dataset (Figure 2b).

While the proportions of most cell types were relatively consistent across the 6 sample types, notable exceptions included the ascending thin limb of Henle (ATL), immune cells (B, T, and Myeloid (MYL), and adaptive/maladaptive cell states (aPT and aTAL) (Figure 2c). The disproportion of ATL cells is likely due to the predominantly cortical nature of the biopsies used in this study, as ATL is a nephron segment of the medulla not regularly sampled in percutaneous biopsies. Among the reference groups, HC had the smallest proportion of adaptive/maladaptive cell states. Figure 2d shows the specific expression of markers used to annotate the cell types

Quality control features and disease markers

The proportion of reads mapping to mitochondrial genes within a cell serves as an essential quality control metric to identify potentially damaged or poor-quality cells[7]. Single-cell analysis of kidney tissue typically shows a high mitochondrial read content. In this study, we included only cells with 50% or fewer mitochondrial reads. The density plot (Figure 3a) indicates that HC had the fewest cells with a high percentage of mitochondrial reads. Conversely, LD and TN exhibited a higher percentage of cells with elevated mean mitochondrial read content, surpassing even the disease groups.

Early stress response genes can be activated within minutes after stimulation[16, 17]. Highest expression of early stressors including Activating transcription factor 3 (ATF3), Dual specificity phosphatase 1 (DUSP1), Early growth response 1 (EGR1), Fos proto-oncogene (FOS) and Jun proto-oncogene (JUN) were observed in TN (Figure 3b). Both LD and TN had higher expression of these genes than DMR and early DKD. Notably, relatively few cells expressed these stressor genes in HC (Figure 3b).

Expression levels of epidermal growth factor (EGF) and uromodulin (UMOD) markers of healthy kidneys[18, 19], were the lowest in DKD. Pseudo bulk differential expression analysis of EGF and UMOD between the reference groups did not show any significant differences. The expression levels of kidney injury or disease markers including hepatitis A virus cellular receptor 1 (HAVCR1), interleukin 2 (IL2,) interleukin 6 (IL6), lipocalin 2 (LCN2), matrix metalloproteinase 7 (MMP7), TIMP metallopeptidase inhibitor 1 (TIMP1) and VCAM1[1, 20-23] is shown in Figure 3c and 3d. HC was the only reference group with undetectable or very low expression of these genes. LD and TN had comparable expression of these genes.

Immune cells and proximal cell states

Immune cell enrichment was over five times greater in DKD compared to HC and LD (Figure 4a). Figure 4b shows the row-scaled heatmap of the log fold changes from the differential immune cell composition analyses performed between DKD and the reference groups; all comparisons were significantly different (adjusted p value < 0.05). Low B cell composition in HC was indicated by the high log fold change in DKD versus HC. Similarly, the heatmap indicates high B and T cell composition in TN compared to HC and LD. Supplementary Table 1 provides the number of cells in each cell type for every sample group.

The pseudotime ordering of cells from the proximal cell type along a one-dimensional trajectory (Figures 4c-d) revealed similarities among the three reference tissues. Figure 4c illustrates the trajectory of the two PT cell states across all sample types. In Figure 4d, the PT cell state

trajectory is shown separately for each studied sample type; the cell state ordering in HC, LD, and TN appeared similar. Notably, HC also contained the adaptive/maladaptive cell state. However, the proportion of the adaptive cell states in HC were lower compared to the other two reference sample types (Supplementary Figure 1). As expected, the DKD cell states displayed the most distinct ordering along the trajectory, with a high number of adaptive/maladaptive state cells.

Effect of post-operative tissue procurement

Differential expression analysis between the two different biopsy procurement method groups, along with manual curation, identified a set of 25 genes (Supplementary Table 2) that exhibited significantly elevated expression in the post-operative surgical biopsy group (Figure 5a). Network analysis using STRING[11] revealed direct interactions among 19 of these 25 genes (Figure 5b), with enrichment in the Jun N-terminal kinase (JNK) and p38 mitogen-activated protein kinase (MAPK) pathways (Figure 5d). Notably, every LD and TN sample demonstrated markedly higher expression of this 25 gene set score without much variability between individual samples compared to the percutaneous biopsy tissues from HC, DMR, early DKD, and DKD. (Figure 5c).

Age effect

Figure 6a presents Venn diagrams illustrating the overlap of DEGs between disease and reference groups. The number of unique DEGs in DMR vs HC (1428 genes) that did not overlap with any comparison was higher (> 1.5 times) compared to that of DMR vs LD and DMR vs TN. Similarly, for DKD vs HC, the number of unique DEGs (2004 genes) was more than twice that for DKD vs LD and DKD vs TN. In contrast, for early DKD vs HC, where both groups had an average age below 26, the number of unique DEGs was lower than in early DKD vs LD and early DKD vs TN. The average age in the LD, TN, DMR, and DKD groups was above 37 years (Table 1). These observations point to a potential age effect contributing to the DEGs.

Figure 6b shows the age distribution per sample type. HC and early DKD were considerably younger than all other sample types. DKD were the oldest. We identified two sets of genes that positively (n = 58) and negatively associated (n = 96) with age (correlation >0.5) from the WGCNA analysis of pseudo bulk read counts of LD (Supplementary Table 3).

Based on the mRNA expression levels in single cell data, Age_gene_scores at cell level were calculated for these positively and negatively age correlated genes. Density plot of the gene scores show enrichment of the positively correlated genes in the PT, TAL and principal cells and the negatively correlated genes were enriched in endothelial, podocytes and fibroblast cells (Figure 6c). Figures 6d and 6e show significant correlation ($p<0.0002$, $p<0.001$) between age and the mean Age_gene_scores calculated at sample level. Gene ontology biological processes enriched for positively regulated genes included methylation, macro molecule modification and microtubule polymerization; negatively correlated genes were enriched for protein kinase c signaling, lymph vessel development, endothelial cell proliferation etc. (Figure 6f, 6g).

Disease versus reference groups

Figures 7a and 7b present Venn diagrams illustrating the overlap of DEGs comparing DKD to reference types. Specifically, Figure 7a displays the overlap with adjustments for age alone, while Figure 7b shows the overlap with adjustments for both age and procurement method. Supplementary Figure 2 shows the significantly enriched gene ontology biological processes for the 1792, 320 and 501 significant DEGs that were unique to HC , LD and TN groups in the differential expression analyses with DKD (Figure 7b).

Figure 7c illustrates the top 5 processes from the IPA analysis of DEGs between DKD and three reference groups: HC, LD, and TN, after adjusting for age and procurement effects. The top enriched pathways were all immune-related. The DKD versus HC comparison exhibited the highest enrichment in immune processes, while the DKD versus TN comparison showed the

lowest enrichment. In the DKD versus LD analysis, enrichment in immune-related processes markedly decreased after adjustments for age and procurement effects. Interestingly, some of the energy related pathways were more enriched in DKD versus LD comparison than DKD versus HC or DKD versus TN after adjusting for the confounding factors (Supplementary Figure 3).

Figure 7d displays the pathway enrichment scores for DEGs in the DKD versus HC comparison. The scores remained similar when adjusted for only age and when adjusted for both age and procurement effects, as both DKD and HC samples were obtained via needle biopsies, rendering procurement effects irrelevant.

Figure 7e highlights the top 5 pathways for DEGs from the DMR versus LD comparison, with and without the adjustment for confounding factors. Adjusting for these factors reduced enrichment in all pathways except for the Sirtuin signaling pathway. Figure 7f shows that Sirtuin 1 (SIRT1) expression is more enriched in the DMR group compared to all other groups studied. The enrichment score for neutrophil degranulation significantly decreased after adjusting for age and procurement effects.

Figure 7g highlights the top 5 pathways for DEGs from the early DKD versus LD comparison, both with and without confounding factors. The top 5 enrichment pathways included Rho GTPase cycle, Senescence pathway, CLEAR signaling, super pathway of inositol phosphate components and protein kinase A signaling. Adjusting for just age increased the enrichment score in all processes except inositol phosphate components pathway. The increased effect disappeared when adjusted for both age and procurement effect except for Rho GTPase cycle. The expression of ArfGAP with RhoGAP Domain, ankyrin repeat and pH domain 1 (ARAP1) a key player in Rho GTPase cycle was enriched markedly in early DKD group compared to other groups (Figure 7h).

Pathway activity

Figure 8a presents a heat map illustrating pathway activity enrichment based on scaled activity scores across different sample types. In the reference groups, phosphoinositide 3-kinase (PI3K) was the only pathway enriched in HC. In contrast, multiple pathways were activated in both LD and TN; specifically, tumor protein p53 (P53), trail, and transforming growth factor beta (TGFβ) pathways were enriched in both groups. Additionally, hypoxia was enriched in LD, while mitogen-activated protein kinases (MAPK), nuclear factor kappa-B (NF-κB), PI3K and wingless-related integration site (WNT) were enriched in TN.

Hierarchical clustering of aggregated pseudo read counts at sample group level without adjusting for confounding factors is shown in Figure 8b. After accounting for these confounding factors, the DKD and DMR groups clustered separately from all other sample groups studied (Figure 8c).

DISCUSSION

Reference samples are crucial in disease pathogenesis studies, but obtaining healthy human kidney samples is particularly challenging[5]. As a result, researchers frequently rely on kidney reference sources that may not serve as adequate comparator with disease samples. In this study, we compared the molecular profiles of reference samples from three different kidney tissue sources to address this issue. We integrated and compared transcript profiles from single-cell data obtained from HC, LD, and TN samples. Although the samples originated from different projects, tissue dissociation and single cell RNA sequencing were carried out using the identical protocol and KPMP - Tissue Interrogation Site personnel and core facility staff. This consistency allowed for integration without detectable batch effects.

Kidney tissue procurement, handling, and dissociation can lead to early stress effects and associated transcriptional changes[24]. However, since all samples in this study were stored and processed using an identical protocol, the observed differences are likely attributable to tissue procurement methods or the physiological and clinical conditions of the donors. Samples were obtained via percutaneous needle biopsy (HC, early DKD, DMR, DKD) and peri- (LD) or post-operative tissue biopsy (TN). In HC compared to TN and LD, lower mitochondrial read content per cell and reduced expression of early stress markers were encountered. The mean percentage of mitochondrial read content per cell for LD and TN exceeded that of needle biopsy disease samples (Figure 3a), indicating that kidney cells start losing their viability during the procedure in operative biopsies. Furthermore, we identified a gene set associated with the operative tissue procurement method that was consistently higher in every LD and TN samples compared to percutaneous needle biopsy samples (Figure 5). As expected, the enriched pathways for this gene set were stress associated JNK and p38MAPK cascades.

The TN samples included non-cancerous tissue from donors with high eGFR (>80), minimal pathology, no diabetes, and no hypertension. However, the higher proportion of B, and T cells in

TN samples compared to HC and LD (Figures 4a and 4b) suggests underlying remaining tissue pathology. The low expression of UMOD and EGF in the DMR group may be attributed to having only six samples, along with a lower TAL cell proportion compared to other groups (Figure 2c).

The low or nonexistent expression of disease markers, including IL2, IL6, LCN2, MMP7, and VCAM1, in HC indicates the pristine quality of the tissue (Figure 3c and 3d). The minimal kidney damage marker expression and low B cell infiltration signal in HC can be attributed to both the biopsy procurement and the younger age of this group compared to LD and TN.

Interestingly, LD exhibited higher expression of HAVCR1, which encodes the kidney injury marker 1 (KIM1), in more cells compared to HC and TN (adjusted p-value < 0.05). HAVCR1 is significantly expressed in the adaptive/maladaptive proximal cell state, which is more prevalent in LD and TN than in HC (Supplementary Figure 1).

Across the sample groups a gene set associated with post-operative tissue procurement and age could be characterized. Elevated scores from the gene set linked to operative procurement effects were consistently observed in every sample from HC and TN, compared to all other percutaneous biopsy samples. The genes that were positively associated with age were found to be enriched for methylation. Increase in methylation with age has been previously shown[25]. The enrichment of endothelial associated processes for the genes negatively associated with age is also supported by published studies[26, 27].

Incorporating age and procurement effects as confounding factors in the differential expression analysis minimized their impact on gene expression (Figures 6 and 7a-e). Several immune-related pathways were highly enriched among the DEGs in the DKD versus reference sample analyses, highlighting the complex role of immune cells in disease progression (Figure 7c, Supplementary Figure 2). DEGs from the DKD versus HC comparison showed the highest

enrichment for immune-related processes whereas, the DKD versus TN comparison showed overall lowest enrichment. This suggests that many of these top processes, as shown in Figures 7c, may already be activatd in TN compared to HC and LD.

Adjusting for the procurement effect in comparisons between disease biopsies and healthy controls did not significantly alter the pathway enrichment scores (Figure 7d), indicating the specificity of the procurement effect gene set (both disease and HC were surgical needle biopsies).

Enrichment of oxidative phosphorylation was observed exclusively in DKD versus LD, compared to DKD versus HC and TN suggesting high levels of these activities in HC and TN (Supplementary Figure 2). Energy-related activities are known to be higher in younger individuals compared to older adults[28, 29], which could explain the elevated levels of these processes in HC. Neighboring cells around a tumor often exhibit high levels of oxidative phosphorylation (OXPHOS) due to the 'reverse Warburg effect'[30]. Despite the distance to the tumor of > 3cm in all TN samples, this phenomenon could still account for the increased energy-related processes in TN.

Adjusting for confounding factors did not alter the enrichment score for the Sirtuin signaling process in the DEGs from the DMR versus LD analyses suggesting its role in DMR (Figure 7e). The role of SIRT1 in diabetes and insulin sensitivity have been previously implicated by several studies [31-34]. Moreover, the expression of SIRT1 was most enriched in DMR compared to all other sample groups in our analysis (Figure 7f). Another interesting observation in the comparative analyses was the enrichment of Rho GTPase cycle process in DEGs from early DKD versus LD (Figure 7g). The confounding factor adjustments increased the enrichment score of Rho GTPase cycle in early DKD versus LD analyses. The specific over-expression of ARAP1 a protein involved in Arf and Rho signaling [35] in early DKD (Figure 7h) further supports the observation from the pathway enrichment analyses. ARAP1 has been implicated in kidney

fibrosis, particularly in DKD[36], ARAP1 knockdown attenuated kidney injury and dysfunction in diabetic mice[36].

The absence of enriched activity in 'Pathway Responsive GENes' (PROGENy[37]) except for PI3K in HC indicates that these samples are unperturbed (Figure 8a). The enrichment of PI3K in HC could refelct its crucial role in physiological cell function such as growth and metabolism[38]. Conversely, the enrichment of stress related activities including p53, Trail, and TGFβ activities in LD and TN, suggest a potential response to the effect of operative tissue procurement. Additionally, the enrichment of NFKβ and MAPK activities in TN indicates a possible underlying pathology in these samples.

One limitation of this study is the small number of samples per group, which prevented matching sex distributions, shown to impact health and disease[39]. Another limitation is the lack of surgical disease biopsy samples to clearly demonstrate the procurement effect adjustment in differential expression analyses.

In conclusion, HC tissue sourced from younger donors, appeared to be the most pristine. In contrast, TN samples displayed more underlying molecular pathology compared to LD. Comparative analyses of top enriched pathways for differentially expressed genes highlight the importance of accounting for appropriate confounding factors in differential expression analyses between disease and reference samples to achieve credible results. We hope that the data sets provided in this study can server as a first step towards better and comparable controls to validate findings and identify key pathways and targets of kidney diseases.

References

1. 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.

2. Menon R, Bomback AS, Lake BB*, et al.* Integrated single-cell sequencing and histopathological analyses reveal diverse injury and repair responses in a participant with acute kidney injury: a clinical-molecular-pathologic correlation. *Kidney Int* 2022; **101:** 1116-1125.

3. Menon R, Otto EA, Berthier CC*, et al.* Glomerular endothelial cell-podocyte stresses and crosstalk in structurally normal kidney transplants. *Kidney International* 2022; **101:** 779-792.

4. Menon R, Otto EA, Hoover P*, et al.* Single cell transcriptomics identifies focal segmental glomerulosclerosis remission endothelial biomarker. *JCI insight* 2020; **5**.

5. Kimmel PL, Wendler DS. Biopsies from Healthy Volunteers to Advance Precision Medicine. *N Engl J Med* 2023; **389:** 1834-1837.

6. Jain S. Conundrums of choice of 'normal' kidney tissue for single cell studies. *Curr Opin Nephrol Hypertens* 2023; **32:** 249-256.

7. de Boer IH, Alpers CE, Azeloglu EU*, et al.* Rationale and design of the Kidney Precision Medicine Project. *Kidney International* 2021; **99:** 498-510.

8. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; **9:** 559.

9. Street K, Risso D, Fletcher RB*, et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 2018; **19:** 477.

10. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 2014; **15:** 550.

11. Szklarczyk D, Gable AL, Lyon D*, et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* 2019; **47:** D607-d613.

12. Shannon P, Markiel A, Ozier O*, et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13:** 2498-2504.

13. Yu G, Wang LG, Han Y*, et al.* clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 2012; **16:** 284-287.

14. Krämer A, Green J, Pollard J, Jr*, et al.* Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 2013; **30:** 523-530.

15. Badia IMP, Vélez Santiago J, Braunger J*, et al.* decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinform Adv* 2022; **2:** vbac016.

16. Bahrami S, Drabløs F. Gene regulation in the immediate-early response process. *Advances in Biological Regulation* 2016; **62:** 37-49.

17. Fowler T, Sen R, Roy AL. Regulation of primary response genes. *Mol Cell* 2011; **44:** 348-360.

18. Garimella PS, Biggs ML, Katz R*, et al.* Urinary uromodulin, kidney function, and cardiovascular disease in elderly adults. *Kidney International* 2015; **88:** 1126-1134.

19. Postalcioglu M, Scherzer R, Ix JH*, et al.* Urine Epidermal Growth Factor and Kidney Function Decline in Middle-Aged Adults. *Kidney Medicine* 2024; **6:** 100846.

20. Avello A, Guerrero-Mauvecin J, Sanz AB. Urine MMP7 as a kidney injury biomarker. *Clin Kidney J* 2024; **17:** sfad233.

21. Su H, Lei CT, Zhang C. Interleukin-6 Signaling Pathway and Its Role in Kidney Disease: An Update. *Front Immunol* 2017; **8:** 405.

22. Kozeny GA, Nicolas JD, Creekmore S*, et al.* Effects of interleukin-2 immunotherapy on renal function. *J Clin Oncol* 1988; **6:** 1170-1176.

23. Bienaimé F, Muorah M, Yammine L*, et al.* Stat3 Controls Tubulointerstitial Communication during CKD. *J Am Soc Nephrol* 2016; **27:** 3690-3705.

24. Denisenko E, Guo BB, Jones M*, et al.* Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biology* 2020; **21:** 130.

25. Jones MJ, Goodman SJ, Kobor MS. DNA methylation and healthy human aging. *Aging Cell* 2015; **14:** 924-932.

26. Lähteenvuo J, Rosenzweig A, Sinclair D*, et al.* Effects of Aging on Angiogenesis. *Circulation Research* 2012; **110:** 1252-1264.

27. Mammoto T, Torisawa YS, Muyleart M*, et al.* Effects of age-dependent changes in cell size on endothelial cell proliferation and senescence through YAP1. *Aging (Albany NY)* 2019; **11:** 7051-7069.

28. Manini TM. Energy expenditure and aging. *Ageing Res Rev* 2010; **9:** 1-11.

29. Wilson M-MG, Morley JE. Invited Review: Aging and energy balance. *Journal of Applied Physiology* 2003; **95:** 1728-1736.

30. Fu Y, Liu S, Yin S*, et al.* The reverse Warburg effect is likely to be an Achilles' heel of cancer that can be exploited for cancer therapy. *Oncotarget* 2017; **8**.

31. Kitada M, Koya D. SIRT1 in Type 2 Diabetes: Mechanisms and Therapeutic Potential. *Diabetes Metab J* 2013; **37:** 315-325.

32. Kitada M, Kume S, Kanasaki K*, et al.* Sirtuins as possible drug targets in type 2 diabetes. *Curr Drug Targets* 2013; **14:** 622-636.

33. Bordone L, Motta MC, Picard F*, et al.* Sirt1 regulates insulin secretion by repressing UCP2 in pancreatic beta cells. *PLoS Biol* 2006; **4:** e31.

34. Moynihan KA, Grimm AA, Plueger MM*, et al.* Increased dosage of mammalian Sir2 in pancreatic beta cells enhances glucose-stimulated insulin secretion in mice. *Cell Metab* 2005; **2:** 105-117.

35. Miura K, Jacques KM, Stauffer S*, et al.* ARAP1: a point of convergence for Arf and Rho signaling. *Mol Cell* 2002; **9:** 109-119.

36. Li X, Ma T-K, Wang M*, et al.* YY1-induced upregulation of LncRNA-ARAP1-AS2 and ARAP1 promotes diabetic kidney fibrosis via aberrant glycolysis associated with EGFR/PKM2/HIF-1α pathway. *Frontiers in Pharmacology* 2023; **14**.

37. Schubert M, Klinger B, Klünemann M*, et al.* Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nature Communications* 2018; **9:** 20.

38. Hoxhaj G, Manning BD. The PI3K-AKT network at the interface of oncogenic signalling and cancer metabolism. *Nat Rev Cancer* 2020; **20:** 74-88.

39. Lopes-Ramos CM, Chen CY, Kuijjer ML*, et al.* Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues. *Cell Rep* 2020; **31:** 107795.

Acknowledgement

We are deeply indebted to the generosity of participants volunteering to donate tissue primarily for research purposes despite receiving no direct immediate benefit to their clinical care. We thank the clinical coordinators for their efforts in participant enrollments and biopsy tissue procurement. All the members of the Kidney Precision Medicine Project consortium are listed in supplementary acknowledgement.

Funding

The study is funded by the following grants from the NIH: The KPMP is funded by the following grants from the NIDDK: U01DK133081, U01DK133091, U01DK133092, U01DK133093, U01DK133095, U01DK133097, U01DK114866, U01DK114908, U01DK133090, U01DK133113, U01DK133766, U01DK133768, U01DK114907, U01DK114920, U01DK114923, U01DK114933, U24DK114886, UH3DK114926, UH3DK114861, UH3DK114915, UH3DK114937. Additional funding sources are K23 DK116720, R01 DK132399, UC2 DK114886, and P30 DK116073, JDRF (2-SRA-2019-845-S-B, 3-SRA-2022-1097-M-B), the Boettcher Foundation, and in part by the Intramural Research Program at NIDDK and the Centers for Disease Control and Prevention (CKD Initiative) under inter-agency agreement #21FED2100157DPG (P.B), U54DK083912, P30 DK081943/HCA: Kidney Seed Network (M.K), U2C DK114886, R21 DK126329 and Chan-Zuckerberg Foundation - CZF2019-002447 (M.B.), K23 DK125529 (A.S.N). This work was also supported by the George M. O'Brien Michigan Kidney Translational Resource Center funded by NIH/NIDDK grant U54DK137314.

Dataset availability

The KPMP single cell data used in this study is available from https://atlas.kpmp.org/repository/. The single cell count matrices for the Human Kidney Transplant Transcriptomic Atlas is available from the Gene Expression Omnibus under accession number GSE 169285. Healthy control and early chronic disease sample read count matrices are available at GSE220939, and

GSE279086. The tumor-nephrectomy data can be found in https://cellxgene.cziscience.com/collections/a98b828a-622a-483a-80e0-15703678befd.

Author contributions:

RM and MK conceived and designed the work that led to the submission. ASN, BG, CO'C, EAO, JBH, RM, ESW, MS, YJC, LP, PL, SSW, MB, PB acquired data, (including methods). RM, CCB, LS, EAO, CO'C performed the analyses generating figures and tables. RM, PPK, LS, PB, MK, MB, SER, JS, CCB, EAO, JBH played a role in interpreting the results. RM, LS, MK drafted the manuscript. All authors reviewed and approved the manuscript.

Disclosures:

Dr. Kretzler reports grants and contracts through the University of Michigan with the National Institutes of Health (NIH), Chan Zuckerberg Initiative, Breakthrough T1D, Alport Foundation, amfAR, AstraZeneca, NovoNordisk, Eli Lilly, Gilead, Janssen, Boehringer-Ingelheim, Moderna, European Union Innovative Medicine Initiative, Certa, Chinook, Angion, RenalytixAI, Travere, Regeneron, IONIS, Maze, Sanofi, Dimerix, Roche-Genentech and Vera Therapeutics. He has received consulting fees through the University of Michigan from Janssen, NovoNordisk, Otsuka, Alexion, Variant Bio and Novartis. Dr. Kretzler served on the NIH-NCATS council, is the committee chair for the American Society of Nephrology Program and is on the board of Nephcure Kidney International. In addition, he has a patent PCT/EP2014/073413 "Biomarkers and methods for progression prediction for chronic kidney disease" licensed. Dr. Bjornstad reports serving or having served as a consultant for AstraZeneca, Bayer, Bristol-Myers Squibb, Boehringer Ingelheim, Eli-Lilly, LG Chemistry, Sanofi, Novo Nordisk, and Horizon Pharma. P.B. also serves or has served on the steering committees, advisory boards and/or data safety committees of AstraZeneca, Bayer, Boehringer Ingelheim, Lilly, Novo Nordisk, and XORTX. Dr.

Rosas received research funding paid to her institution from Bayer and the National Institutes of Diabetes and Digestive and Kidney Diseases (NIDDK) and benefiting from an enrichment program (P30 DK03836) during the conduct of the study. She is a member of the steering committee for the FINE-ONE Study and attended advisory board meeting for Fresenius and Bayer. Dr. Sedor has received research funding paid to his institution, outside of this study, from NIDDK and the National Institute of Allergy and Infectious Diseases; for clinical trials from Vertex, Chinook and Maze. He has consulted for Maze and Boehringer Ingelheim; and receives royalties from Sanofi for US Issued Patent US/11,645,753 (ML segmentation Kidney biopsies). Dr. Kimmel received royalties and advances from Elsevier and Mayo Clinic Press. Outside of current work, Dr. Naik reports consulting fees from Vera and a patent pending, Targeting the IGF-1 signaling pathway to extend organ lifespan. Dr. Hodgin received, outside of the current study, research funding from NIH, Janssen, Astra Zeneca, Gilead, Novo Nordisk. Dr. Bitzer has received research funding outside of this study from NIH and honoraria from Wayne State University. Dr. Pyle has received research funding outside of this study from the American Diabetes Association and has been a member of the DSMB for two NIH funded clinical trials. Dr. Waikar reports outside of the current study research funding from Vertex, Pfizer, JNJ, Natera; consultancy with Wolters Kluewer, Bain, BioMarin, Goldfinch, GSK, Ikena, Strataca, Google, CANbridge, NovoNordisk, Ono, PepGen, Sironax, NovoNordisk, Vertex, Mineralys, Motric Bio; expert witness for litigation related to dialysis lab testing (Davita), PPIs (Pfizer), PFAO exposure (Dechert), and a voclosporin patent (Aurinia). All other authors have no disclosures.

Supplementary Materials

1. Supplementary acknowledgement

   The KPMP consortium members

2. Supplementary Tables

   Supplementary Table S1. Reference samples cell type summary

   Supplementary Table S2. Processing effect gene set

   Supplementary Table S3. Age effect gene sets

3. Supplementary Figures

   Supplementary Figure S1. Proportion PT and TAL cell states

   Supplementary Figure S2. Enriched Gene Ontology biological processes

   Supplementary Figure S3: Heatmap of the comparative enrichment of energy related pathways

Table 1: Characteristics of study participants

Sample category	Samples (N)	Average Age in years ± standard deviation	Sex % Female (Male)	Race (%)			
				Asian	Black	White	Other[a]
Healthy volunteer (HC)	12	25.5 ± 3	46 (54)	8	0	91	0
Preperfusion living donor biopsy (LD)	9[b]	37.8 ± 10.4	77 (23)	11	0	78	11
Tumor-nephrectomy (TN)	9	42.7 ± 11.67	33 (67)	0	0	100	0
Diabetic mellitus-resilient (DMR)	6[b]	54.17 ± 19	100 (0)	0	0	83	17
Early chronic kidney disease (early DKD)	9	17.09 ± 2.29	44 (56)	0	22	67	11
Diabetic kidney disease (DKD)	21[b]	60.9 ± 9.79	62 (38)	4	40	56	0

[a]Native American/Alaskan, Pacific Islanders/Hawaiian
[b]KPMP samples

Figure legends

Figure 1. Overview of the analysis approach

Figure 2. Single cell data. a. UMAP showing the annotated cell types from the integrated analysis of a total of 120,000 cells from 6 sample types (20,000 cells each from Healthy control "HC", Living donor "LD", Tumor-nephrectomy "TN", Diabetic mellitus resistor "DMR", early diabetic kidney disease "early DKD", Diabetic kidney disease "DKD"). **b.** UMAP of the integrated dataset showing successful integration of HC, LD, TN, DMR, early DKD and DKD. **c.** stacked barplot showing the proportion of the 6 sample types in each of the cell type identified. **d.** Violin plot showing the specific markers for the annotated cell types. Abbreviations include: Podocyte, POD; Parietal epithelial cell, PEC; Proximal tubule, PT; adaptive/maladaptive, aPT; descending thin loop of Henle, DTL; ascending thin loop of Henle, ATL; thick ascending loop of Henle, TAL; adaptive/maladaptive thick ascending loop of Henle, aTAL; distal convoluted tubule, DCT; connecting tubule, CNT; Principal cell, PC; Intercalated type A, IC-A; Intercalated type B, IC-B; transient between PC and IC, tPC-IC; Endothelial cell EC; Efferent and Afferent arteriolar endothelial cells EC-AEA; Glomerular endothelial cell, EC-GC; Fibroblast, FIB; Vascular smooth muscle cell, VSMC; Pericyte, P; Mesangial cell, MC; Myeloid cell, MYL; B cell, B; T cell, T; Natural killer T cell, NK T/C.

Figure 3. Quality control features and disease markers. a. Violin density plot showing the percentage of mitochondrial reads per cell in the 6 sample types studied. Healthy control (HC) has the lowest number of cells with high percentage of mitochondrial reads. Living donors (LD) and tumor-nephrectomies (TN) have the highest number of cells with high percentage of mitochondrial reads. **b.** Dot plot with the scaled expression levels of early stressors (ATF3, DUSP1, FOS, JUN and EGR1) in the 6 sample types studied. **c.** Dot plots and **d.** violin plots

showing the expression levels of healthy (EGF & UMOD) and injured/disease kidney (IL2, IL6, HAVCR1, LCN2, MMP7, TIMP1 and VCAM1 markers in the 6 sample types**.**

Figure 4. Immune cells and tubular epithelial cell states. a. Proportion of immune cell types found in each of the 6 sample types (HC, LD, TN, DMR, early DKD and DKD). The number on top of each bar is the total number of immune cells identified in each cell type. **b.** Row-scaled heatmap of the log fold changes from the differential immune cell composition analyses performed between DKD and the reference groups; all comparisons were significantly different (adjusted p value < 0.05.). Log fold changes indicate T and B cells were highest in TN samples while HC had fewest B cells. **c.** The two proximal cell states ordered along a trajectory. **d.** The proximal cell states identified in the 6 sample types ordered along the trajectory. Among reference samples, LD had the most adaptive cell state cells.

Figure 5. Effect of post-operative tissue procurement. Using bulk mRNA analysis, a gene set that was upregulated in post-operative tissue procurement method compared to percutaneous needle biopsies was identified. LD and TN samples were acquired by post-operative biopsy procedure and HC, DMR, early DKD and DKD samples were by percutaneous needle biopsies. **a.** Violin density plot showing the level of the score calculated at cell level for the 25 genes that were upregulated in post-operative tissue biopsies. **b.** String interaction network showing direct interaction of 19/25 genes. **c.** Violin plot showing the elevated expression of the score in every LD and TN sample compared to the needle biopsy samples. **d.** Top 5 enriched Gene Ontology biological processes for the gene set.

Figure 6. Effect of age. a. Venn diagrams showing the overlap of significantly (adjusted p value < 0.05) differentially expressed genes (DEGs) in DMR (left panel) versus the three reference

samples (HC, LD and TN); early DKD (middle panel) versus the three reference samples; and DKD (right panel) versus the three reference samples. **b.** Bar plot showing the age distribution in each of the 6 sample types. **c.** Dot plot showing the scaled expression of the scores calculated for the genes that positively and negatively associated with age. The gene scores from the negatively associated genes were enriched in endothelial cells and the gene scores from the positively associated genes were enriched in PC, PT and TAL cell states. **d.** Plot showing the significant correlation between the scores from the positively age-associated genes and age of the living donors. **e.** Plot showing the significant correlation between the scores from the negatively age-associated genes and age of the living donors. **f.** Top enriched Gene Ontology biological processes for the positively age-associated gene set. **g.** Top enriched Gene Ontology biological processes for the negatively age-associated gene set.

Figure 7. Disease versus reference groups (HC, LD and TN). a. Venn diagram illustrating the overlap of age-adjusted, significant differentially expressed genes (DEGs) (adjusted p-value < 0.05) between DKD and the three reference types. **b.** Venn diagram depicting the overlap of DEGs, adjusted for age and procurement effects (adjusted p-value < 0.05), between DKD and the three reference types. **c**. Heat map displaying enrichment scores for the top 5 canonical pathways from IPA for the significant DEGs in DKD versus HC, DKD versus LD and DKD versus TN. Scores are shown after adjusting for age and procurement effects. **d**. Heat map displaying the top 5 enriched canonical pathways from IPA for the significant DEGs between DKD versus HC. Scores are shown for the three types of analyses: without confounding factor adjustment, adjusted for age, and adjusted for age and procurement effects. **e.** Bar plot presenting the top 5 enriched canonical pathways from IPA for the significant DEGs between early DMR and LD. Scores are shown for the three types of analyses: without confounding factor adjustment, adjusted for age, and adjusted for age and procurement effects. **f.** Dot plot showing SIRT1

expression levels across the sample groups. **g.** Bar plot presenting the top 5 enriched canonical pathways from IPA for the significant DEGs between early DKD and LD. Scores are shown for the three types of analyses: without confounding factor adjustment, adjusted for age, and adjusted for age and procurement effects. **h.** Dot plot illustrating ARAP1 expression levels in the sample groups studied.

Figure 8. Pathway activity. a. Heatmap showing the enrichment of top pathway activities inferred by decoupleR analysis on the single cell expression of the 6 sample types. **b.** Hierarchical clustering using the aggregate read counts at the sample type level. **c**. Hierarchical clustering of the age and procurement effect adjusted aggregate read counts.

Figure 1

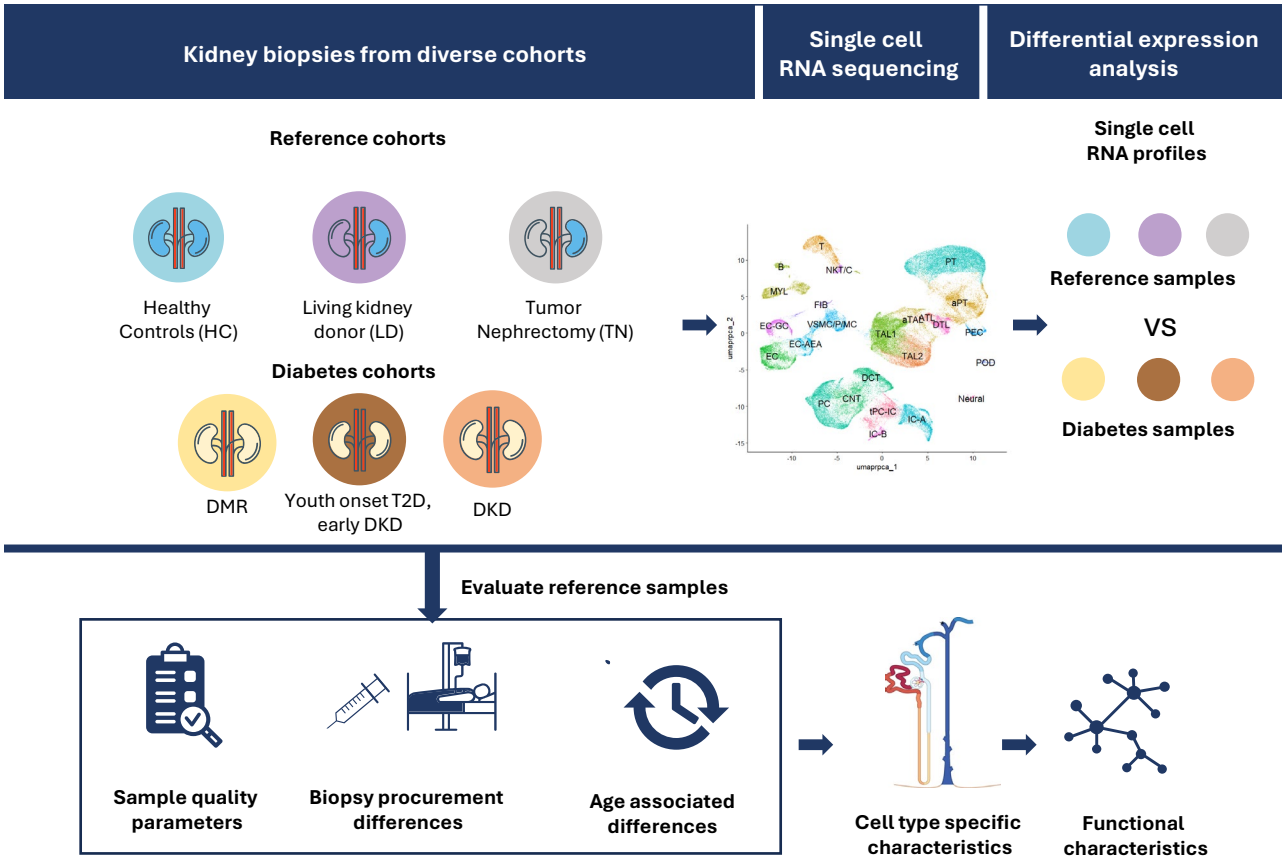

Kidney biopsies from diverse cohorts
Single cell RNA sequencing
Differential expression analysis
Reference cohorts
Healthy Controls (HC)
Living kidney donor (LD)
Tumor Nephrectomy (TN)
Diabetes cohorts
DMR
Youth onset T2D, early DKD
DKD
Single cell RNA profiles
Reference samples
VS
Diabetes samples
Evaluate reference samples
Sample quality parameters
Biopsy procurement differences
Age associated differences
Cell type specific characteristics
Functional characteristics

Figure 2

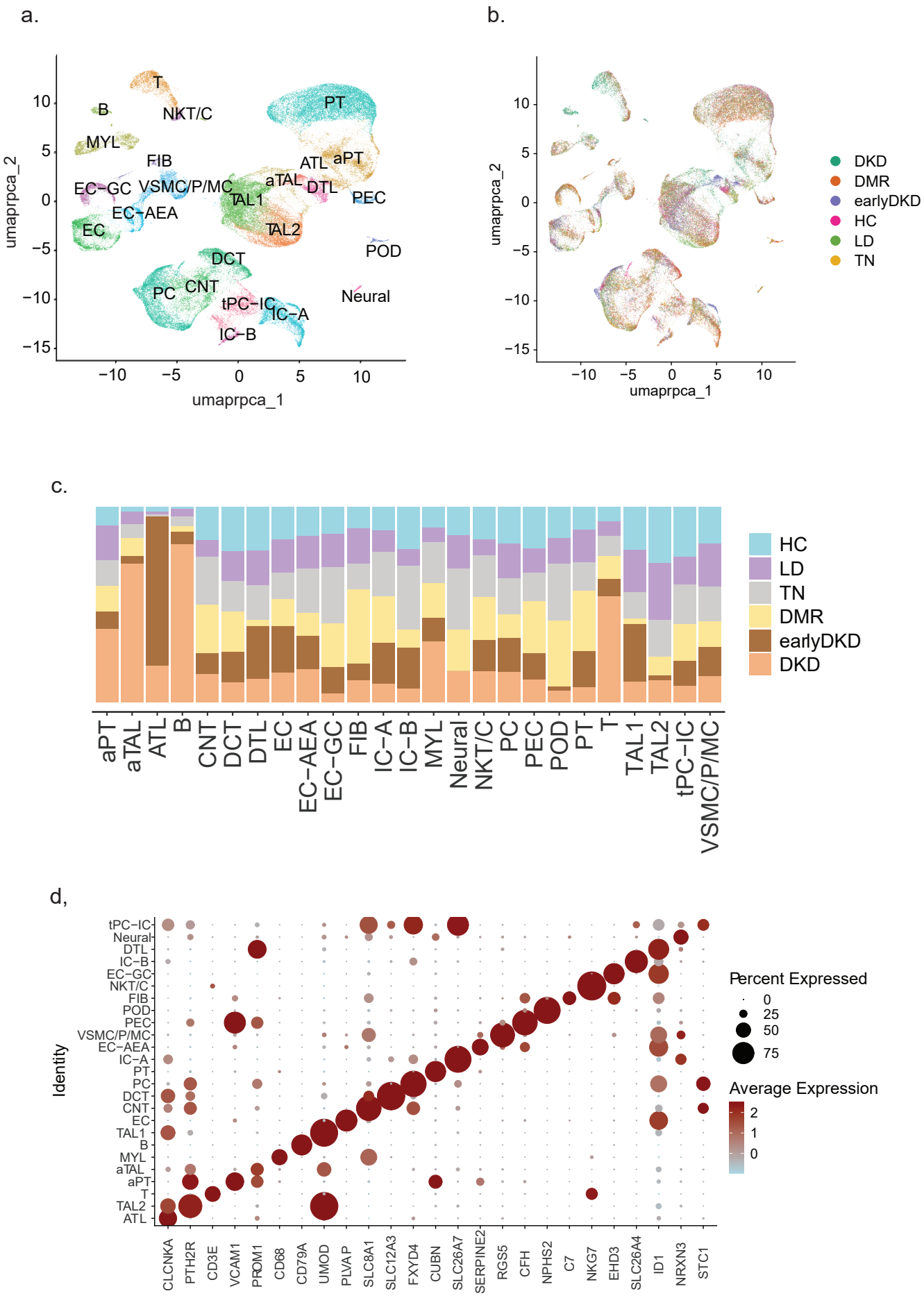

a.
b.
umaprpca_1
umaprpca_2
T
B
NKT/C
PT
MYL
FIB
ATL
aPT
EC−GC
VSMC/P/MC
aTAL
DTL
TAL1
PEC
EC−AEA
EC
TAL2
POD
DCT
PC
CNT
Neural
tPC−IC
IC−A
IC−B
DKD
DMR
earlyDKD
HC
LD
TN
c.
HC
LD
TN
DMR
earlyDKD
DKD
aPT
aTAL
ATL
B
CNT
DCT
DTL
EC
EC−AEA
EC−GC
FIB
IC−A
IC−B
MYL
Neural
NKT/C
PC
PEC
POD
PT
T
TAL1
TAL2
tPC−IC
VSMC/P/MC
d,
Identity
tPC−IC
Neural
DTL
IC−B
EC−GC
NKT/C
FIB
POD
PEC
VSMC/P/MC
EC−AEA
IC−A
PT
PC
DCT
CNT
EC
TAL1
B
MYL
aTAL
aPT
T
TAL2
ATL
CLCNKA
PTH2R
CD3E
VCAM1
PROM1
CD68
CD79A
UMOD
PLVAP
SLC8A1
SLC12A3
FXYD4
CUBN
SLC26A7
SERPINE2
RGS5
CFH
NPHS2
C7
NKG7
EHD3
SLC26A4
ID1
NRXN3
STC1
Percent Expressed
0
25
50
75
Average Expression
2
1
0

Figure 3

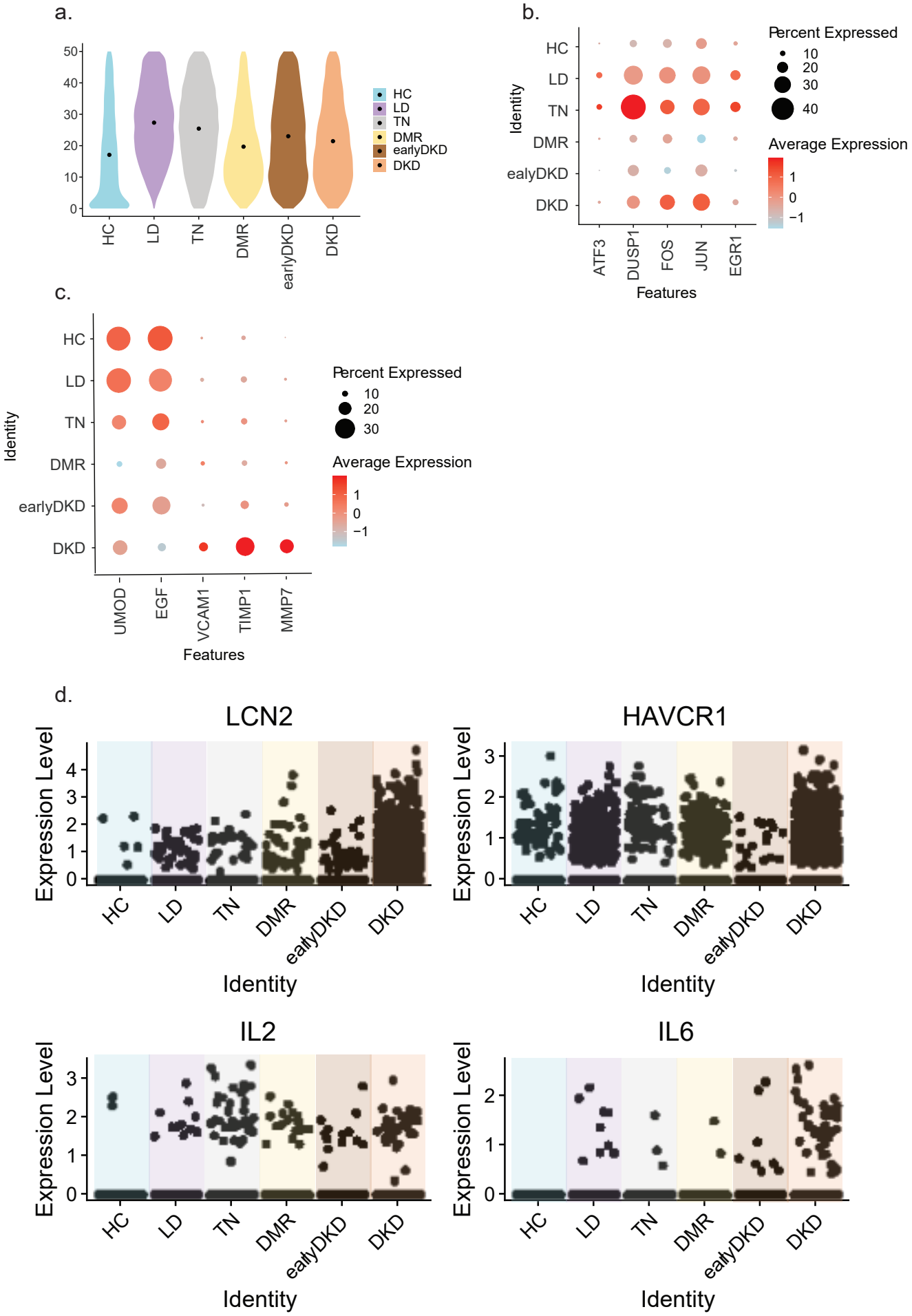

a.
50
40
30
20
10
0
HC
LD
TN
DMR
earlyDKD
DKD
b.
Identity
HC
LD
TN
DMR
ealyDKD
DKD
ATF3
DUSP1
FOS
JUN
EGR1
Features
Percent Expressed
10
20
30
40
Average Expression
1
0
−1
c.
Identity
HC
LD
TN
DMR
earlyDKD
DKD
UMOD
EGF
VCAM1
TIMP1
MMP7
Features
Percent Expressed
10
20
30
Average Expression
1
0
−1
d.
LCN2
HAVCR1
IL2
IL6
Expression Level
Identity
HC
LD
TN
DMR
eatlyDKD
DKD

FIgure 4

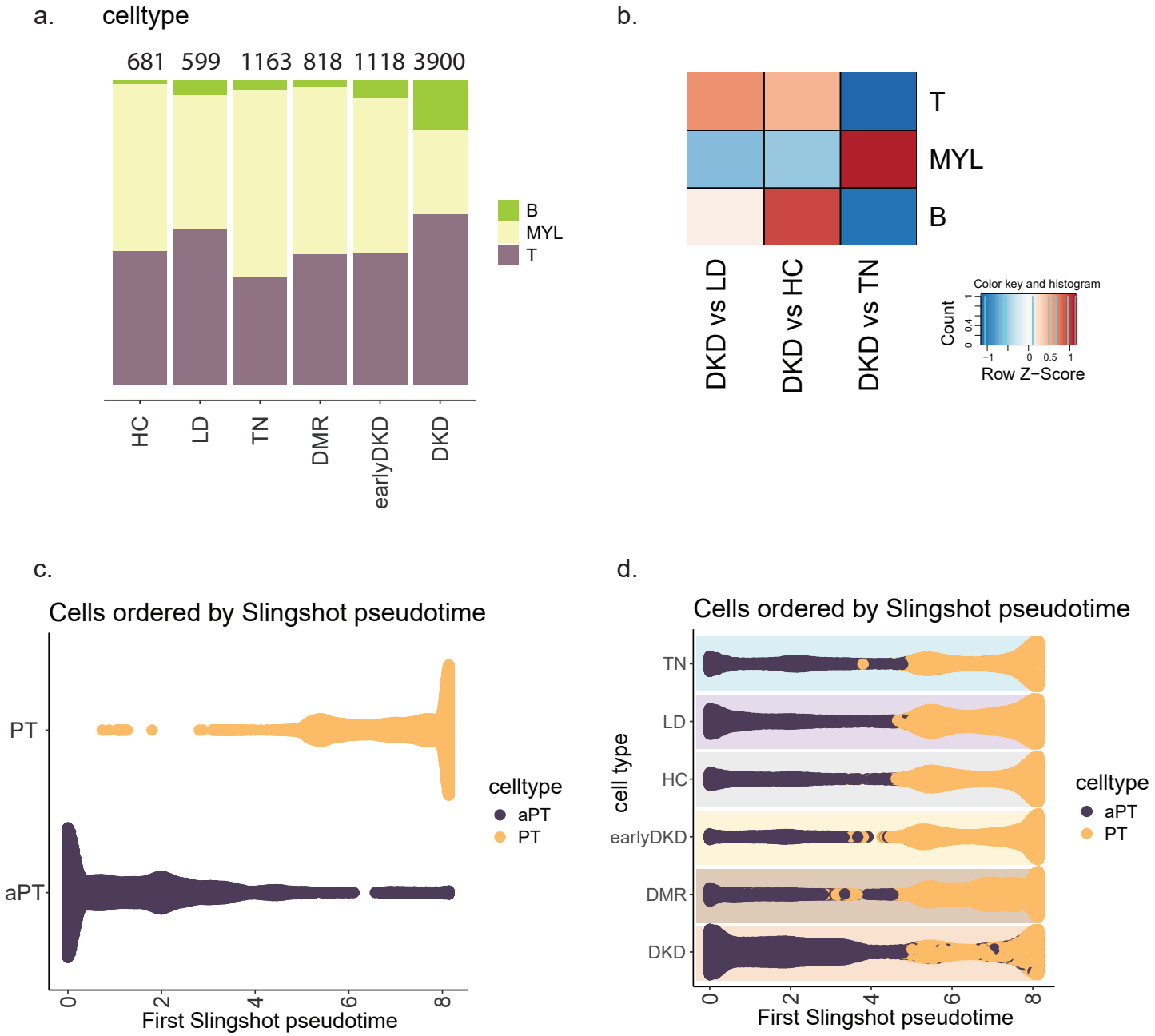

a.
celltype
681 599 1163 818 1118 3900
B
MYL
T
HC
LD
TN
DMR
earlyDKD
DKD
b.
T
MYL
B
DKD vs LD
DKD vs HC
DKD vs TN
Color key and histogram
Count
Row Z−Score
c.
Cells ordered by Slingshot pseudotime
PT
aPT
celltype
aPT
PT
First Slingshot pseudotime
d.
Cells ordered by Slingshot pseudotime
cell type
TN
LD
HC
earlyDKD
DMR
DKD
celltype
aPT
PT
First Slingshot pseudotime

Figure 5

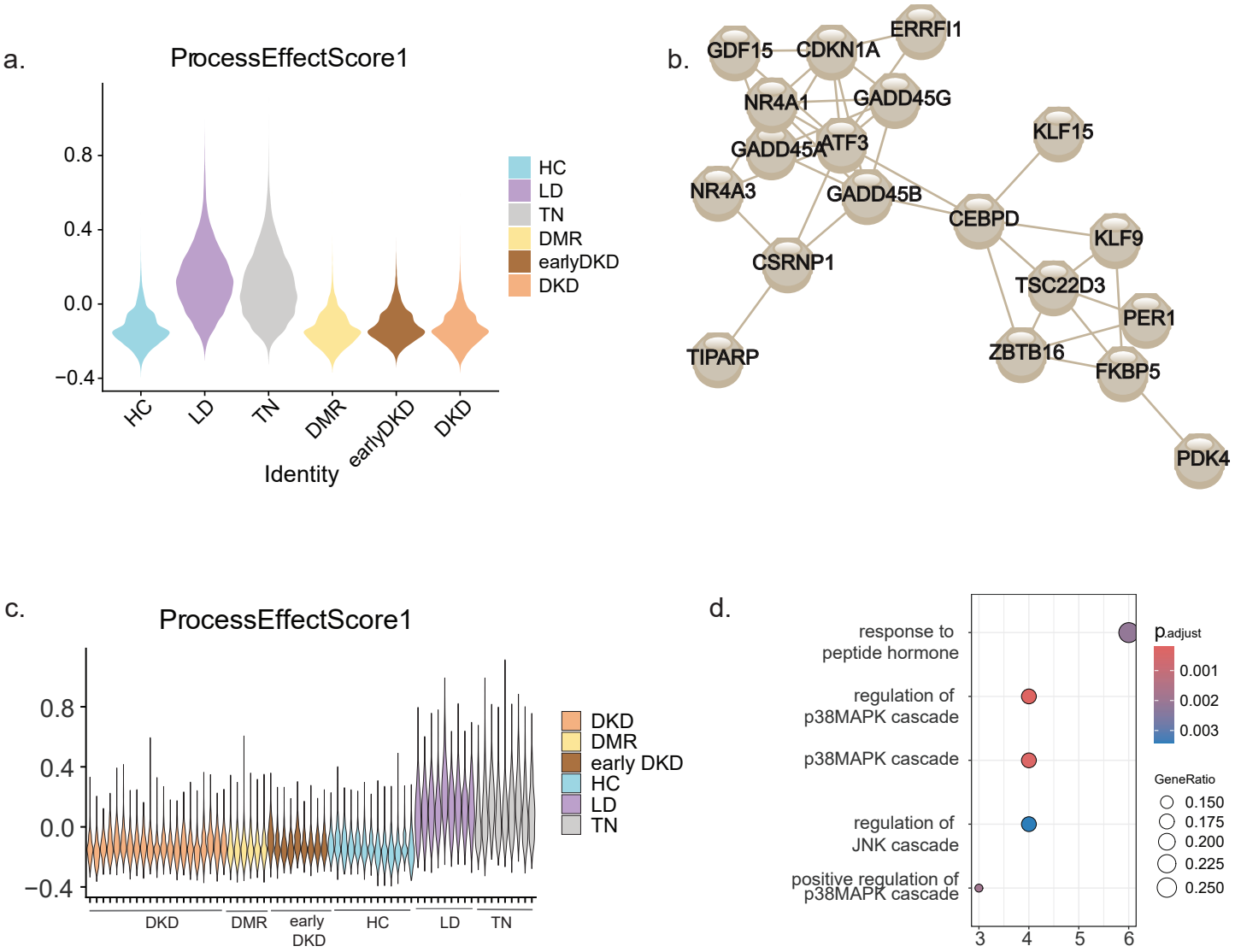

a.
ProcessEffectScore1
0.8
0.4
0.0
−0.4
HC
LD
TN
DMR
earlyDKD
DKD
Identity
HC
LD
TN
DMR
earlyDKD
DKD
b.
GDF15
CDKN1A
ERRFI1
NR4A1
GADD45G
GADD45A
ATF3
KLF15
NR4A3
GADD45B
CEBPD
KLF9
CSRNP1
TSC22D3
PER1
TIPARP
ZBTB16
FKBP5
PDK4
c.
ProcessEffectScore1
0.8
0.4
0.0
−0.4
DKD
DMR
early DKD
HC
LD
TN
DKD
DMR
early DKD
HC
LD
TN
d.
response to peptide hormone
regulation of p38MAPK cascade
p38MAPK cascade
regulation of JNK cascade
positive regulation of p38MAPK cascade
3
4
5
6
p.adjust
0.001
0.002
0.003
GeneRatio
0.150
0.175
0.200
0.225
0.250

Figure 6

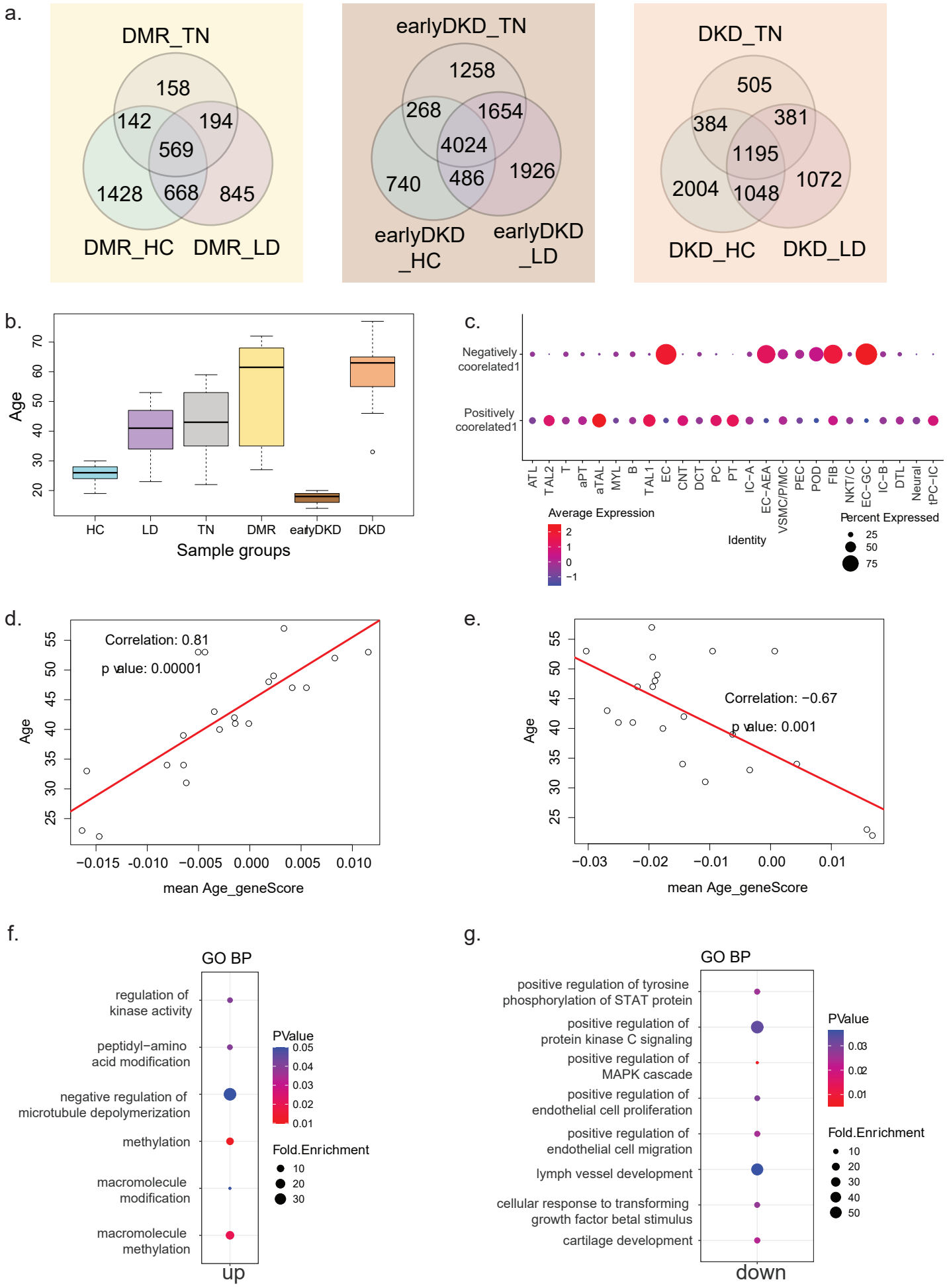
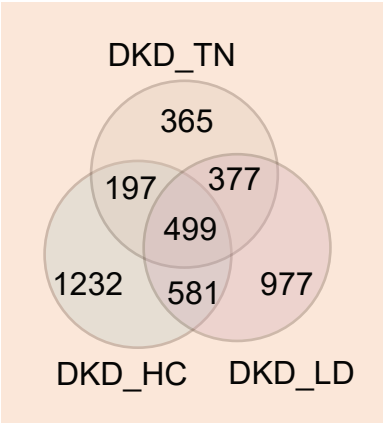
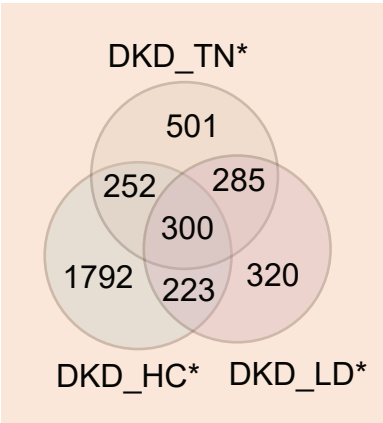

a.
DMR_TN
158
142
194
569
1428
668
845
DMR_HC
DMR_LD
earlyDKD_TN
1258
268
1654
4024
740
486
1926
earlyDKD _HC
earlyDKD _LD
DKD_TN
505
384
381
1195
2004
1048
1072
DKD_HC
DKD_LD
b.
Age
HC
LD
TN
DMR
earlyDKD
DKD
Sample groups
c.
Negatively coorelated1
Positively coorelated1
ATL
TAL2
T
aPT
aTAL
MYL
B
TAL1
EC
CNT
DCT
PC
PT
IC−A
EC−AEA
VSMC/P/MC
PEC
POD
FIB
NKT/C
EC−GC
IC−B
DTL
Neural
tPC−IC
Identity
Average Expression
Percent Expressed
d.
Correlation: 0.81
p value: 0.00001
Age
mean Age_geneScore
e.
Correlation: −0.67
p value: 0.001
Age
mean Age_geneScore
f.
GO BP
regulation of kinase activity
peptidyl−amino acid modification
negative regulation of microtubule depolymerization
methylation
macromolecule modification
macromolecule methylation
up
PValue
Fold.Enrichment
g.
GO BP
positive regulation of tyrosine phosphorylation of STAT protein
positive regulation of protein kinase C signaling
positive regulation of MAPK cascade
positive regulation of endothelial cell proliferation
positive regulation of endothelial cell migration
lymph vessel development
cellular response to transforming growth factor betal stimulus
cartilage development
down
PValue
Fold.Enrichment

Figure 7

a. Age



b. Age and procurement



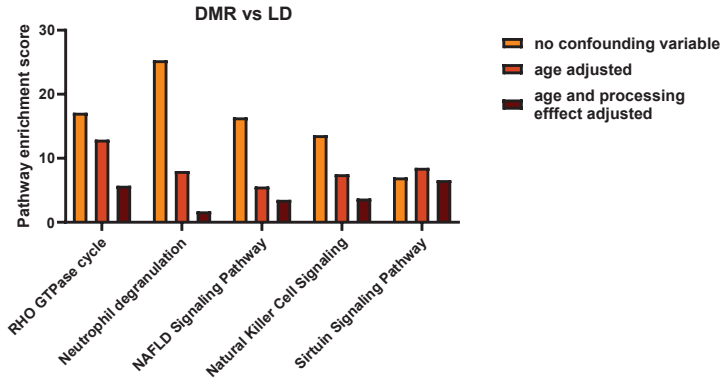
c.

	Age and procurement effect adjusted		
Canonical pathways	DKD_HC	DKD_LD	DKD_TN
Neutrophil degranulation	6.4	2.5	0.5
Interferon gamma signaling	10.9	3.0	1.5
Pathogen Induced Cytokine Storm Signaling	7.4	4.3	0.9
Macrophage Classical Activation Signaling	8.4	5.5	1.2
Th1 and Th2 Activation	7.4	1.9	0.8

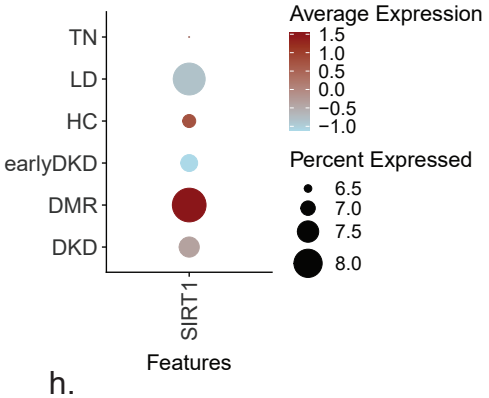
d.

Canonical Pathways	DKD_HC	DKD_HC age effect adjusted	DKD_HC age and procurement effect adjusted
Neutrophil degranulation	19.3	5.8	6.4
Interferon gamma signaling	20.4	11.2	10.9
Pathogen Induced Cytokine Storm Signaling	14.4	7.0	7.4
Macrophage Classical Activation Signaling	15.3	8.7	8.4
Th1 and Th2 Activation	16.1	7.6	7.4

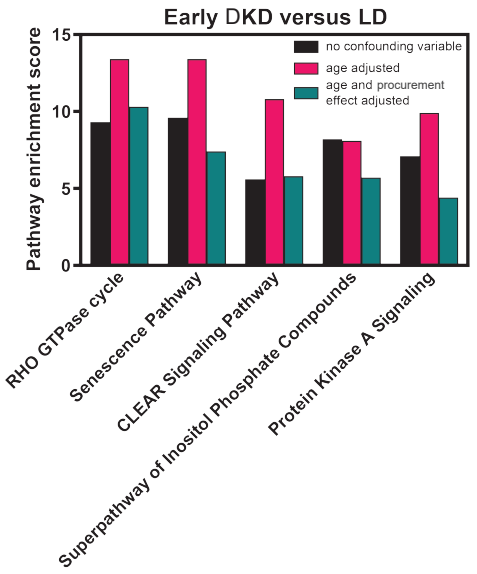
e.



f.



g.



h.

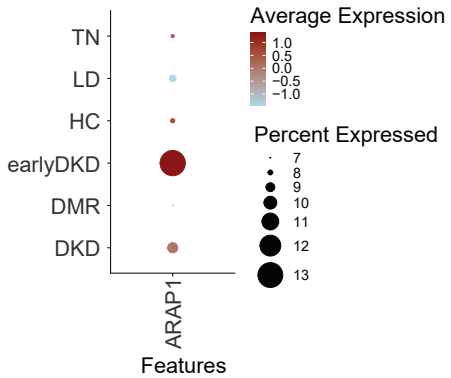
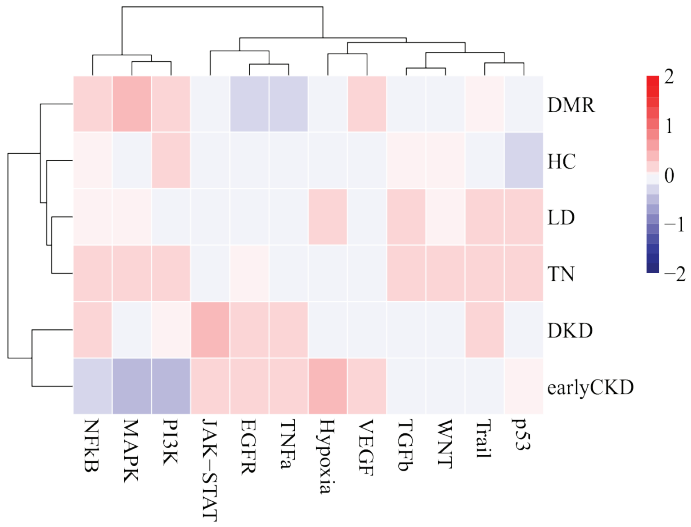
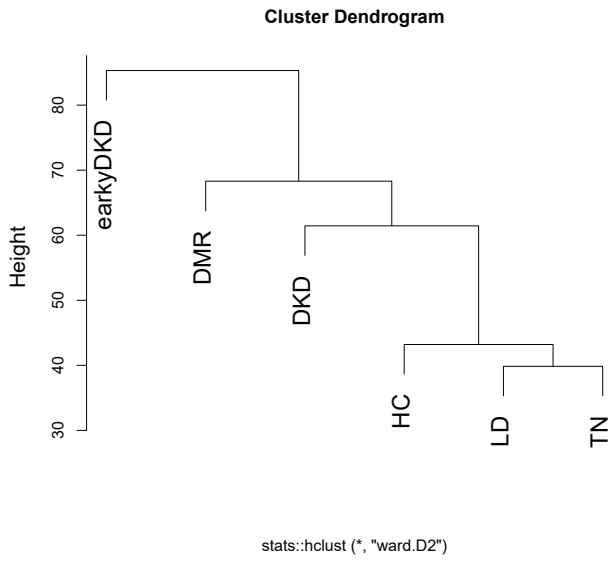


Figure 8

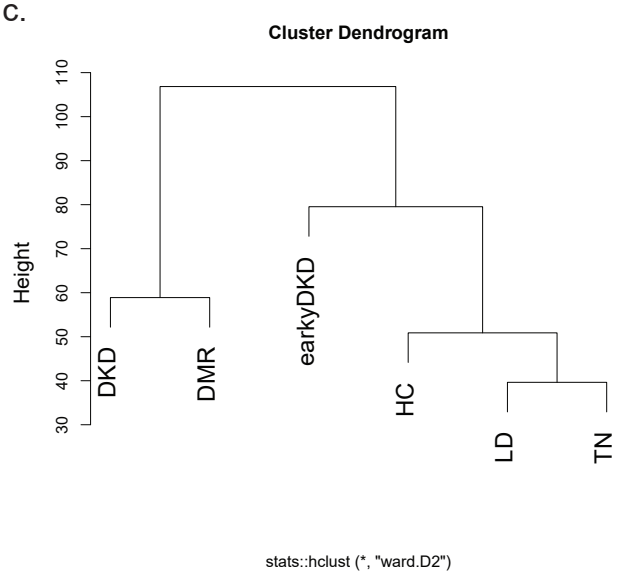
a.


NFKB
MAPK
PI3K
JAK−STAT
EGFR
TNFa
Hypoxia
VEGF
TGFb
WNT
Trail
p53
DMR
HC
LD
TN
DKD
earlyCKD
2
1
0
−1
−2

b.


Cluster Dendrogram
Height
30
40
50
60
70
80
earkyDKD
DMR
DKD
HC
LD
TN
stats::hclust (*, "ward.D2")

c.


Cluster Dendrogram
Height
30
40
50
60
70
80
90
100
110
DKD
DMR
earkyDKD
HC
LD
TN
stats::hclust (*, "ward.D2")