

## Single cell transcriptomics identifies focal segmental glomerulosclerosis remission endothelial biomarker

Rajasree Menon, ... , Kidney Precision Medicine Project (KPMP), Nephrotic Syndrome Study Network (NEPTUNE)

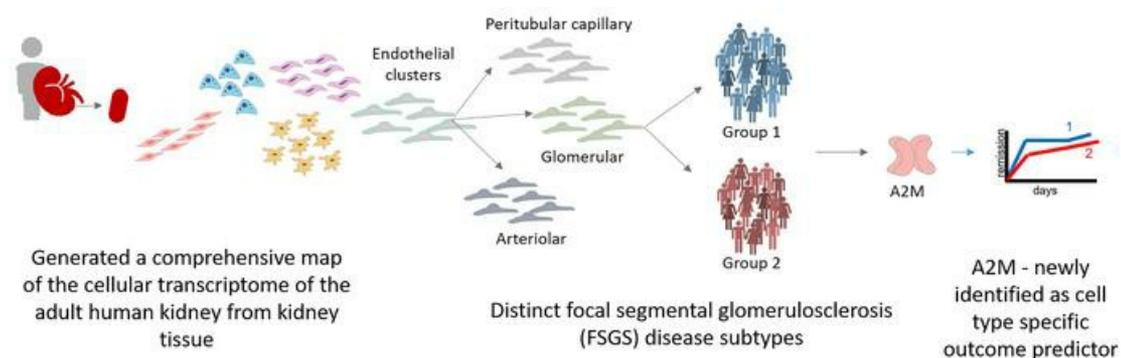
*JCI Insight.* 2020;5(6):e133267. <https://doi.org/10.1172/jci.insight.133267>.

Research Article

Cell biology

Nephrology

### Graphical abstract



Find the latest version:

<https://jci.me/133267/pdf>



# Single cell transcriptomics identifies focal segmental glomerulosclerosis remission endothelial biomarker

Rajasree Menon,<sup>1</sup> Edgar A. Otto,<sup>1</sup> Paul Hoover,<sup>2</sup> Sean Eddy,<sup>1</sup> Laura Mariani,<sup>1</sup> Bradley Godfrey,<sup>1</sup> Celine C. Berthier,<sup>1</sup> Felix Eichinger,<sup>1</sup> Lalita Subramanian,<sup>1</sup> Jennifer Harder,<sup>1</sup> Wenjun Ju,<sup>1</sup> Viji Nair,<sup>1</sup> Maria Larkina,<sup>1</sup> Abhijit S. Naik,<sup>1</sup> Jinghui Luo,<sup>1</sup> Sanjay Jain,<sup>3</sup> Rachel Sealfon,<sup>4</sup> Olga Troyanskaya,<sup>4</sup> Nir Hacohen,<sup>2</sup> Jeffrey B. Hodgin,<sup>1</sup> Matthias Kretzler,<sup>1</sup> Kidney Precision Medicine Project (KPMP),<sup>5</sup> and Nephrotic Syndrome Study Network (NEPTUNE)<sup>5</sup>

<sup>1</sup>Michigan Medicine, Ann Arbor, Michigan, USA. <sup>2</sup>Broad Institute, Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, USA. <sup>3</sup>Renal Division, Department of Medicine, Washington University School of Medicine, St. Louis, Missouri, USA. <sup>4</sup>Flatiron Institute, Simons Foundation, New York, New York, USA. <sup>5</sup>The KPMP and the NEPTUNE are detailed in Supplemental Acknowledgments.

**Authorship note:** RM and EAO are co-first authors. JBH and MK are co-senior authors.

**Conflict of interest:** MK serves on the advisory boards of JDRF, Astra-Zeneca, NovoNordisc, Eli Lilly, Gilead, Goldfinch Bio, Merck, Chan Zuckerberg Initiative, Janssen, Boehringer-Ingelheim, Elpidera, European Union Innovative Medicine Initiative. MK and WJ hold a patent Biomarkers for CKD progression (encompassing urinary EGF as biomarker of CKD progression). OT serves on the Scientific Advisory Boards of Caris Life Sciences and Goldfinch Bio; OT served on Scientific Advisory Boards, received grants or honoraria from UBC, Google, and Renaissance Technologies LLC. LM reports grant funding from Boehringer-Ingelheim and serves as an advisory board member for Reata Pharmaceuticals. WJ has received support for diabetic kidney disease-associated research from the Joint Institute for Translational and Clinical Research (PUUMA). CCB has received support for IgA nephropathy-associated research from the Joint Institute for Translational and Clinical Research (PUUMA). JBH reports grants from Astra-Zeneca, Gilead, and Moderna.

**Copyright:** © 2020, American Society for Clinical Investigation.

**Submitted:** September 6, 2019

**Accepted:** February 19, 2020

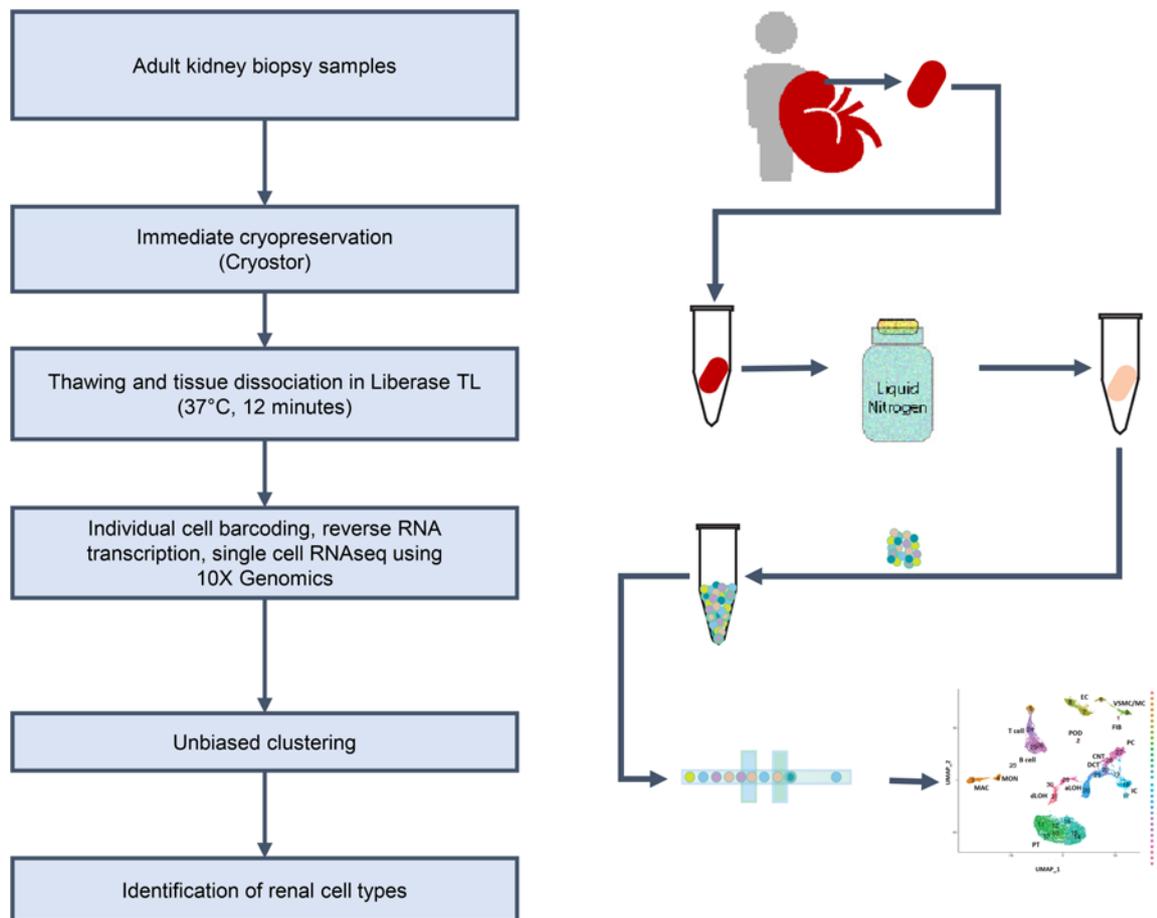
**Published:** February 27, 2020

**Reference information:** *JCI Insight*. 2020;5(6):e133267.  
<https://doi.org/10.1172/jci.insight.133267>.

To define cellular mechanisms underlying kidney function and failure, the KPMP analyzes biopsy tissue in a multicenter research network to build cell-level process maps of the kidney. This study aimed to establish a single cell RNA sequencing strategy to use cell-level transcriptional profiles from kidney biopsies in KPMP to define molecular subtypes in glomerular diseases. Using multiple sources of adult human kidney reference tissue samples, 22,268 single cell profiles passed KPMP quality control parameters. Unbiased clustering resulted in 31 distinct cell clusters that were linked to kidney and immune cell types using specific cell markers. Focusing on endothelial cell phenotypes, *in silico* and *in situ* hybridization methods assigned 3 discrete endothelial cell clusters to distinct renal vascular beds. Transcripts defining glomerular endothelial cells (GEC) were evaluated in biopsies from patients with 10 different glomerular diseases in the NEPTUNE and European Renal cDNA Bank (ERCB) cohort studies. Highest GEC scores were observed in patients with focal segmental glomerulosclerosis (FSGS). Molecular endothelial signatures suggested 2 distinct FSGS patient subgroups with  $\alpha$ -2 macroglobulin (A2M) as a key downstream mediator of the endothelial cell phenotype. Finally, glomerular A2M transcript levels associated with lower proteinuria remission rates, linking endothelial function with long-term outcome in FSGS.

## Introduction

The complexity of kidney structure and function make understanding the mechanisms responsible for kidney disease development and progression challenging. Genome-wide bulk mRNA profiles of kidney tissues provide a starting point to map the molecular underpinning of disease pathogenesis and progression. Successes of this approach include the identification of noninvasive biomarkers from tissue gene expression studies of acute kidney injury (1), chronic kidney disease (2), and targets for potentially novel therapeutic approaches (3, 4). However, bulk mRNA gene expression measurements reflect an average across all captured cell types rather than focusing on cell-specific transcriptional responses. Analytical approaches to identify the most likely cellular source of transcriptional profiles have been used to overcome this limitation (5–7). While these methods have improved our understanding of cellular contribution to bulk mRNA expression profiles, they do not fully capture the transcriptional programs of individual cells. In recent years, RNA sequencing approaches at the single cell transcriptome level (scRNAseq) have been developed and are being refined specifically for kidney tissue (8, 9). scRNAseq allows the measurement and comparison of comprehensive gene sets obtained from individual cells and, thus, enables the capture of previously underappreciated cellular heterogeneity within kidney tissue in health and disease. For example, scRNAseq studies (10–14) have provided comprehensive and dynamic gene expression profiles of the human developing kidney and human organoids at the single cell level (10, 12) and have identified resident immune phenotypes in kidney tissues in lupus nephritis (15).



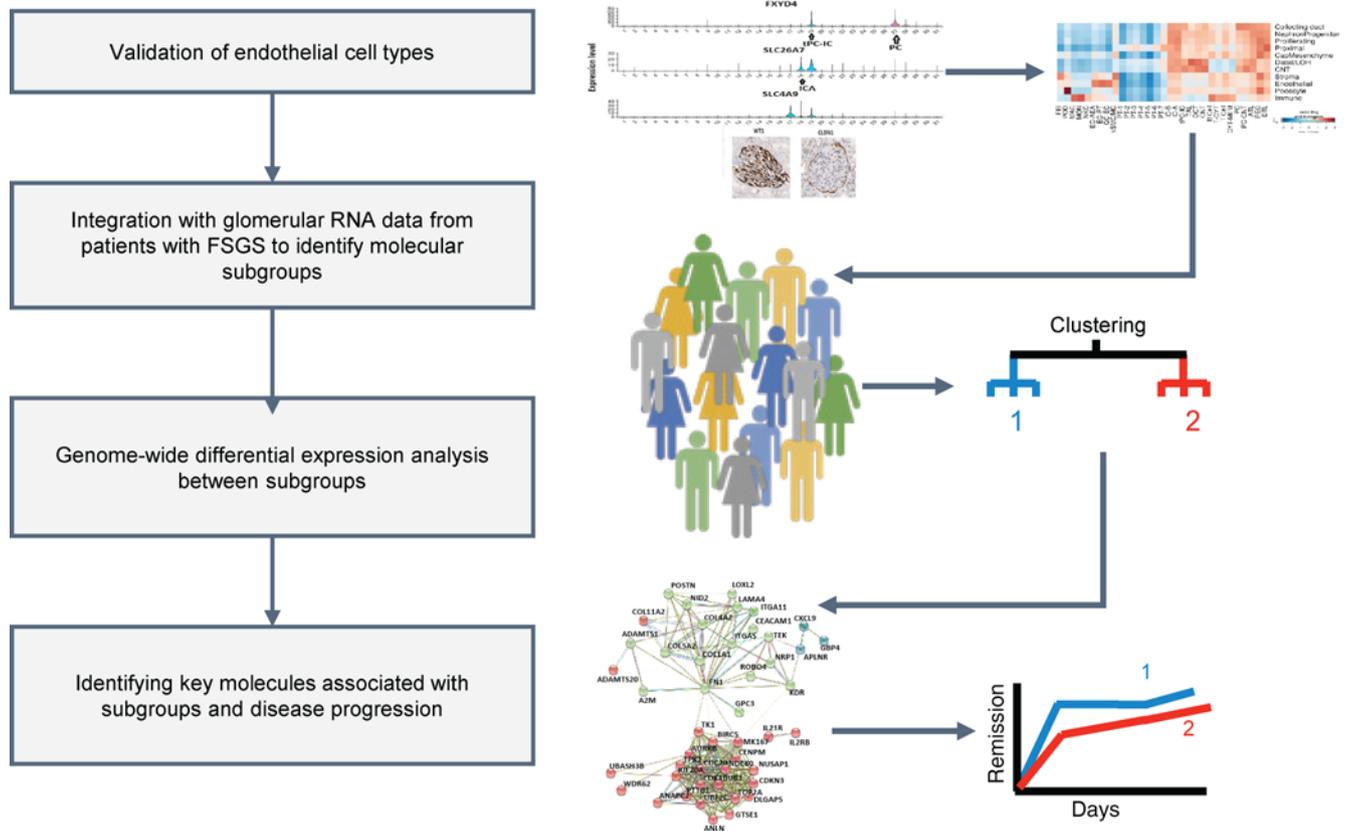
**Figure 1. Analysis workflow showing the generation of adult kidney single cell clusters.** Combined processing of 24 samples including 16 tumor-nephrectomy, 5 surveillance, and 3 preperfusion biopsies using the scRNAseq protocol yielded 22,268 cells (4690 from surveillance biopsies; 14,744 from tumor-nephrectomies; and 2834 cells from preperfusion biopsies) and resulted in 31 distinct clusters. Seurat R package was used for the unsupervised clustering of the single cell data.

The KPMP has been created with a goal to change the way we understand and treat kidney disease. KPMP has identified access to human biopsy tissue as a critical first step in defining disease heterogeneity and determining the precise molecular pathways driving kidney disease. Tissue from a kidney biopsy can provide the opportunity for a variety of molecular analyses of diseases embedded in both the structural organ context and the clinical phenotype (16). Despite collaborative success in collecting biopsy tissue, generating single cell data from adult human nephrons, with their highly differentiated epithelial and endothelial subtypes, has proven particularly challenging due to the rapid decline in cell viability upon dissociation (17). Recently, first cell/nucleus analyses of normal adult human kidneys have been published (9, 18, 19). However, a comprehensive map of single cell types from multiple normal human kidney sources such as nontumor sections of tumor-nephrectomies, surveillance, and preperfusion biopsies is yet to be developed. Such a reference map for the adult human kidney will provide insights into human disease and is critical as a comparator for biopsy-driven single cell studies aiming to identify underlying targetable mechanisms for early detection and treatment.

An overall strategy was developed to address 3 main goals presented in this study: (a) to establish a protocol for kidney biopsy-based scRNAseq analysis capturing all cellular elements of the kidney (Figure 1); (b) to establish a map of the transcripts found in kidney cell types defined by data-driven clustering; and (c) to test the ability of this technique to identify distinct, cell type-specific features seen in bulk RNA profiling from human kidney disease as a means to understand and identify key cellular mechanisms underlying kidney disease phenotypes (Figure 2).

## Results

*KPMP scRNAseq protocol for analysis of adult human kidney.* Intrinsic nephron cell types and tissue-resident immune cells were recovered from kidney biopsy specimens using a scRNAseq protocol for KPMP, designed



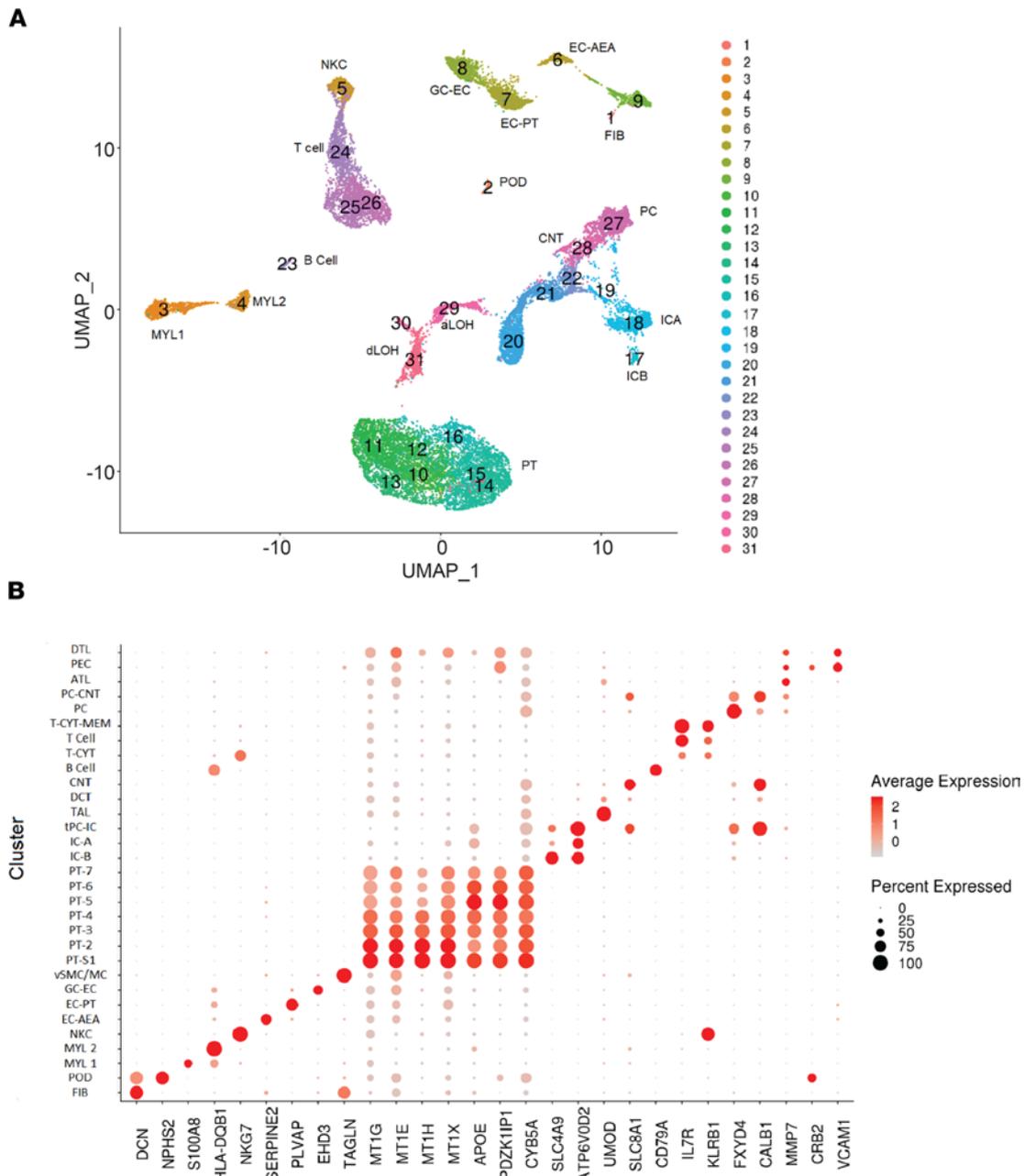
**Figure 2. Analysis workflow showing the integration of single cell with bulk mRNAseq data.** The glomerular endothelial-specific transcripts identified from scRNAseq analyses were used to identify distinct subgroups within FSGS using mRNAseq profiles available from NEPTUNE.

for deployment in a multicenter research network (kpmp.org). Twenty-four adult, nondiseased, human kidney tissue samples were processed using the KPMP scRNAseq protocol, including 16 unaffected specimens from tumor-nephrectomies, 3 preperfusion living donor kidney biopsies, and 5 surveillance kidney transplant biopsies. The combined samples yielded 22,268 cells (4690 from surveillance biopsies, 14,744 from tumor-nephrectomies, and 2834 cells from preperfusion biopsies) that passed the KPMP quality control parameters. Unsupervised clustering of all 22,268 cells resulted in 31 distinct clusters, ranging in size from 1412 cells to 58 cells (Figure 3A and Table 1). The scRNAseq expression data set has been loaded into the scRNAseq viewer Nephrocell and can be explored at Nephrocell.MiKTMC.org.

*Identification of kidney cell types.* Differential expression analysis of the genes expressed in each cluster versus all other clusters generated cell type-specific gene sets (FDR < 0.05) (Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.133267DS1>). Each gene set included known kidney cell type-specific markers, which enabled assignment of clusters to specific cell types. Once cluster identities were established, the gene sets provided potentially novel cell type-specific markers. Figure 3B indicates the cluster (or cell type) specificity of the most differentially expressed genes in each cluster.

Each cluster included cells from all 3 different sample sources used, supporting comparable cellular yields from each source and the absence of sample-specific batch effects (Figure 4). Distribution of cells from all 24 samples in the 31 clusters are listed in Supplemental Table 2. Alignment of cell clusters with kidney cell types (20) using known cell type markers are outlined in Supplemental Table 1.

Proximal tubule (PT) cells populated the most abundant clusters (clusters 10–16, Figure 3A). Cells from at least 21 out of the 24 samples were found across 7 PT clusters, confirming broad contribution to the cellular clusters by the diverse samples (Supplemental Table 2). The top markers for cluster 10–16 were mitochondrial genes (Figure 3B and Supplemental Table 1), suggestive of gradient differences in mitochondrial gene expression in these 7 clusters (Supplemental Figure 1A). Increased mitochondrial activity in the PTs is well known (21). Moreover, single nucleus RNAseq (snRNAseq) studies, including those from KPMP on identical tumor-nephrectomy tissues, have also reported more than the conventional PT cell types (S1, S2, and S3 cells) in their analyses (18, 19).



**Figure 3. Unsupervised clustering of cells derived from adult kidney tissue. (A)** UMAP plot from unsupervised clustering of 22,268 cells and number of cells in each of the 31 clusters. **(B)** Expression levels of the top differentially expressed gene from each cluster across all clusters. Differential gene expression analyses, to identify cell type specific genes, were performed using the nonparametric Wilcoxon rank sum test. The cell type-specific genes with FDR < 0.05 were considered as differentially expressed. UMAP, Uniform Manifold Approximation and Projection.

The cells of the distal nephron constituted the next-largest clusters, were identified based on known markers, and included descending thin limb (DTL, cluster 31), ascending thin limb (ATL, cluster 29), thick ascending limb (TAL, cluster 20) of the loop of Henle (LOH), distal convoluted tubule (DCT, cluster 21), connecting tubule (CNT, cluster 22), and collecting duct cells (principal cells [PC], cluster 27; and intercalated cells [IC] A and B, clusters 17 and 18, respectively). The Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction process of the expression data preserved the global structure (22) of the distal nephron (Figure 3A). In addition to the major distal nephron cell types, we identified 3 cell clusters (clusters 19, 28, and 30) for which the top differentially expressed genes suggested that the clusters consisted of potentially transitioning cell types.

The top differentially expressed genes in cluster 19 included known markers from principal (*FXYD4*) and intercalated (ICA [*SLC26A7*] and ICB [*SLC4A9*]) collecting duct cells (Figure 5A). A similar cell

**Table 1. Cluster cell type and number of cells**

Cluster	Cluster name	Abbreviation	Number of cells
1	Fibroblast	FIB	58
2	Podocyte	POD	170
3	Myeloid cell	MYL1	896
4	Myeloid cell	MYL2	473
5	NK cells	NKC	565
6	Endothelial cell	EC-AEA	461
7	Endothelial cell, peritubular	EC-PT	988
8	Glomerular capillary endothelial cell	GC-EC	900
9	Vascular smooth muscle cell and mesangial	vSMC/MC	655
10	Proximal tubule cell	PT-1	891
11	Proximal tubule cell	PT-2	1212
12	Proximal tubule cell	PT-3	987
13	Proximal tubule cell	PT-4	813
14	Proximal tubule cell	PT-5	891
15	Proximal tubule cell	PT-6	1412
16	Proximal tubule cell	PT-7	525
17	Intercalated $\beta$ cell	IC-B	174
18	Intercalated $\alpha$ cell	IC-A	977
19	Transitioning principal cell/intercalated cell	tPC-IC	165
20	Thick ascending limb cell	TAL	1343
21	Distal convoluted tubule cell	DCT	901
22	Connecting tubule	CNT	451
23	B cell	B cell	141
24	Cytotoxic T cell	T-CYT	1091
25	T cell, activated	T Cell	1095
26	T cell, memory	T-CYT-MEM	898
27	Principal cell	PC	1164
28	Cortical connecting tubule principal cell	PC-CNT	414
29	Ascending thin limb cell	ATL	656
30	Parietal epithelial cell/loop of Henle	PEC	158
31	Descending thin limb cell	DTL	743
			Total 22268 cells

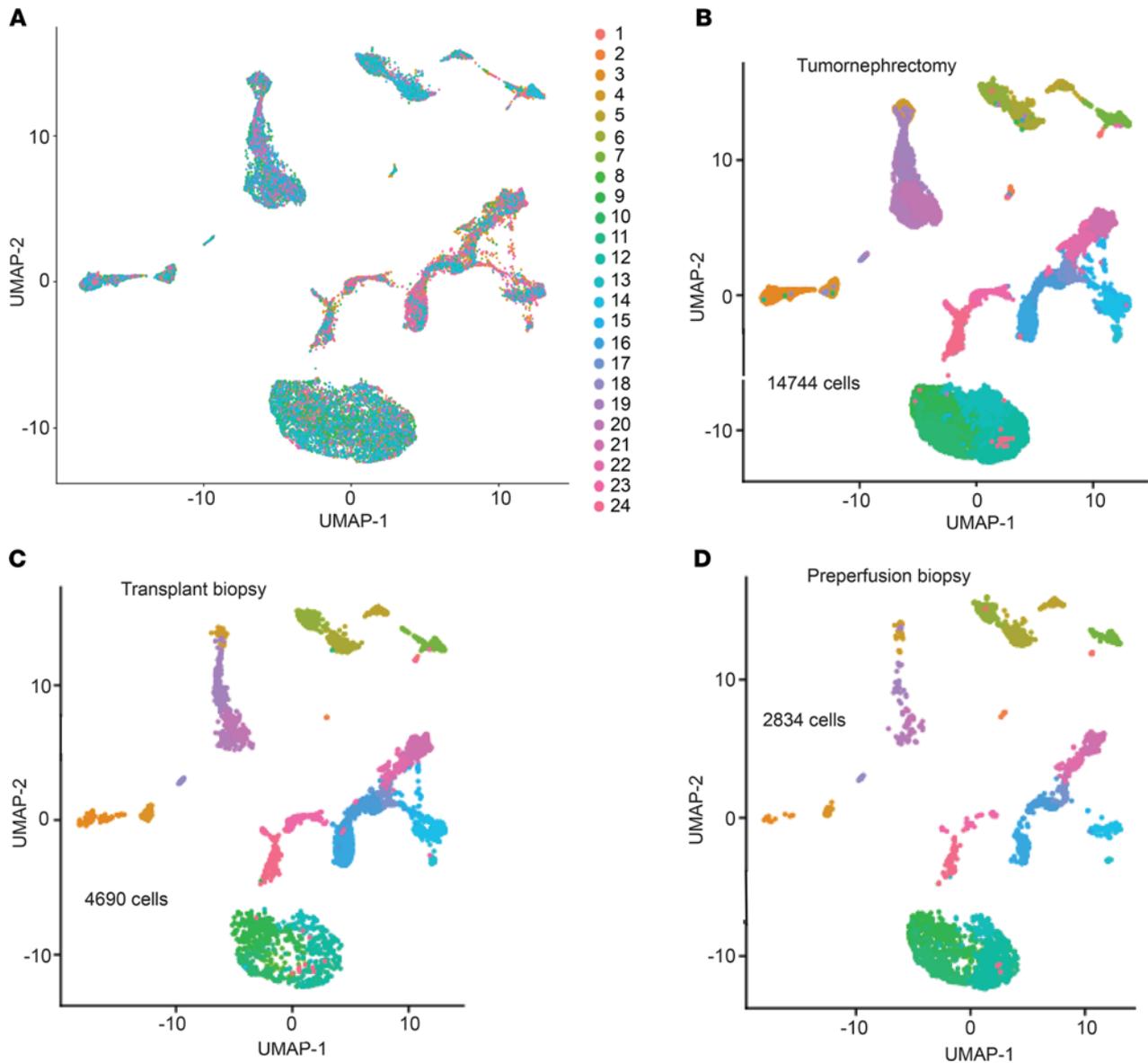
Distribution of single cells isolated from tissue samples in 31 clusters and associated cell types.

cluster was previously reported and validated by a scRNAseq study in mice (14). This cluster had a comparatively higher mRNA content (Supplemental Figure 1B).

Cluster 28, which we annotated to contain CNT-PC (cortical CNT IC) expressed known markers of both PCs (*FXRD4*, *STC1*) and CNT cells (*SLC8A1*) (Supplemental Figure 1C), consistent with transitional states between CNT to cortical collecting duct cells.

The cell type markers of cluster 30 included *CTGF*, *CRB2*, *VCAM1*, *SLC4A11*, and *CLDN1* (Supplemental Table 1). Antibody staining of the coded proteins of those top genes referring to the Human Protein Atlas (HPA; <https://www.proteinatlas.org>), showed that most of these markers were expressed in glomerular parietal epithelial cells (PEC) and LOH cells. On examining this cluster more closely, subclustering showed 2 clusters, subcluster 0 and 1, contained within cluster 30 (Figure 5B). The expression of *WT1*, *CLDN1*, and *CRB2* in subcluster 0 suggests that this subcluster was enriched for PEC, while the expression of *TACSTD2* and *SPP1* in subcluster 1 indicates enrichment for LOH-derived cells.

Cluster 2 contains cells with known podocyte-specific markers, allowing the identification of potentially novel podocyte-specific transcripts (Figure 5C). Cluster 2 is composed of 170 cells identified as podocytes based on expression of known, podocyte-specific markers *NPHS1*, *NPHS2*, and *PODXL*. Podocytes were identified in 19 out of the 24 tissue samples and constituted 0.75% of total cells. *CRB2* was significantly overexpressed in podocytes compared with all other cell types. Additionally, *CRB2* expression was specific to clusters representing podocytes and PEC (Figure 5C).



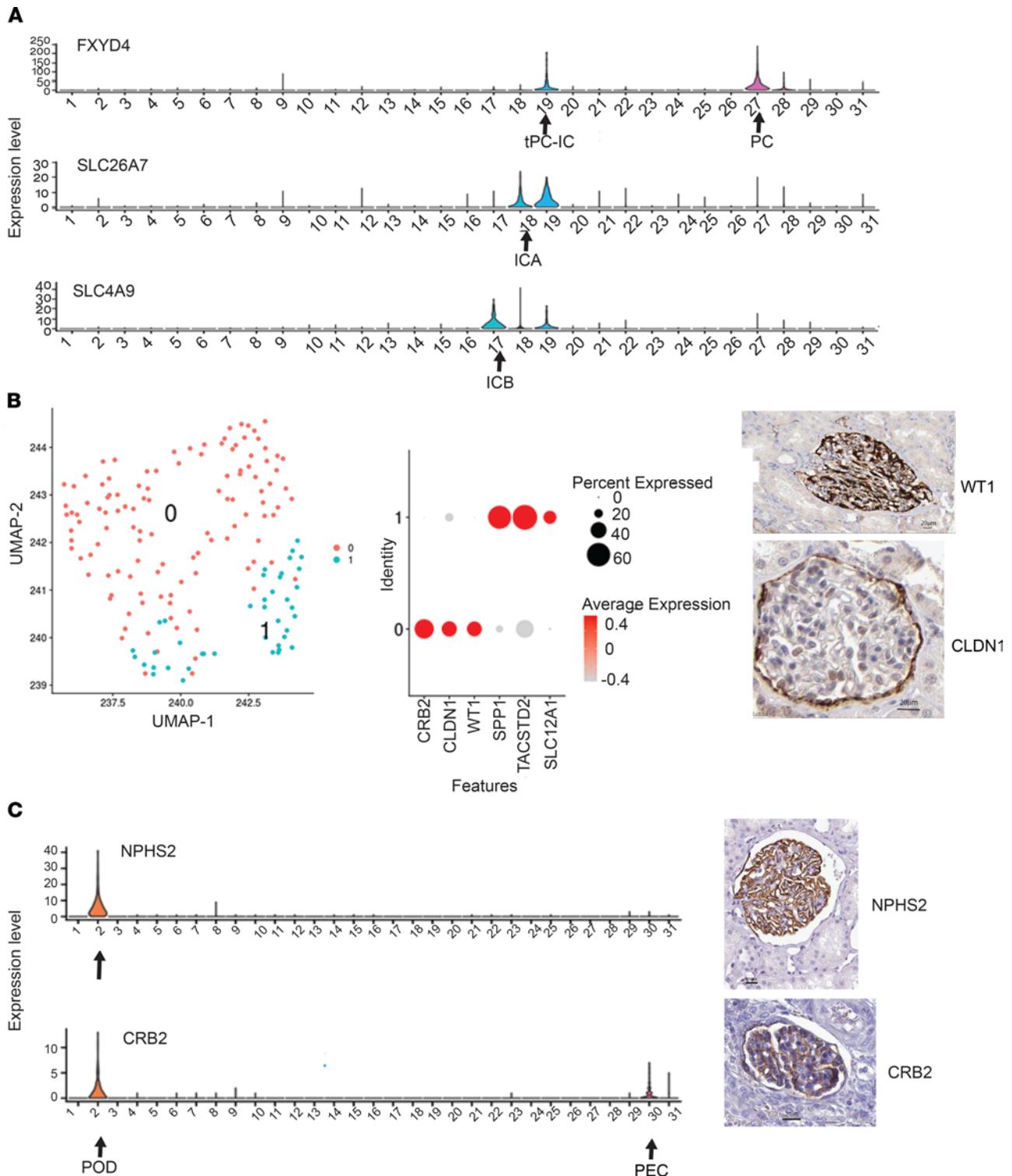
**Figure 4. UMAP plot showing the distribution of the single cells from the 24 samples used in the study.** (A) UMAP plot showing the distribution of the cells in the 31 clusters resulted from the unsupervised clustering of the scRNAseq data from tumor-nephrectomy, surveillance, and preperfusion biopsies. (B) UMAP plot of the 14,744 cells from the 16 tumor-nephrectomy samples. (C) UMAP plot of the 4690 cells from the 5 surveillance transplant biopsy samples. (D) UMAP plot of the 2834 cells from the 3 preperfusion biopsy samples.

Cluster 9 contained cells with known markers of smooth muscle cells, *ACTA2*, and *TAGLN*, consistent with vascular smooth muscle cells and mesangial cells (VSMC/MC). The specific antibody staining of *ITGA8* and *SYNPO2* in the HPA authenticated the presence of MCs in this cluster (Supplemental Figure 2).

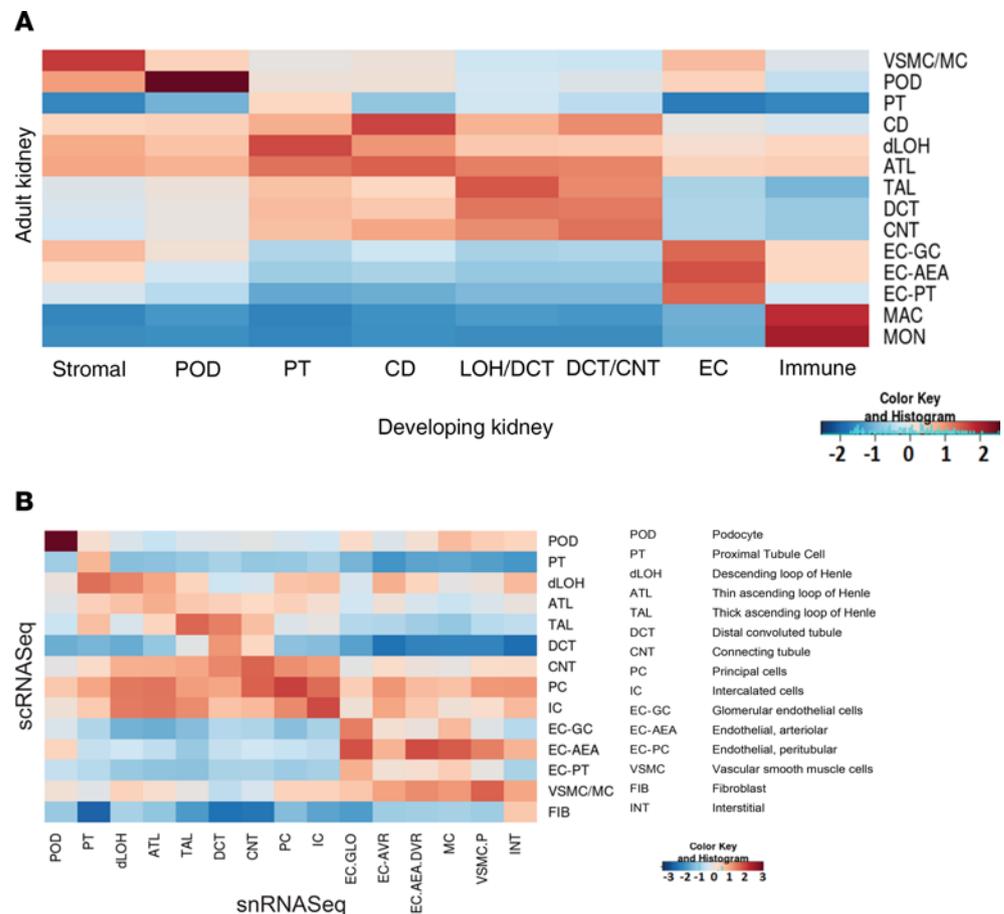
Multiple distinct immune cell types were observed in the scRNAseq data of the reference tissues. Seven distinct immune clusters were identified: myeloid cells (clusters 3 and 4), NK cell (NKC) (cluster 5), B cell (cluster 23), cytotoxic T cell (T-CYT) (cluster 24), activated T cell (T-ACT) (cluster 25), and T memory cell (T-MEM) (cluster 26).

Cluster 1 — the smallest cluster, comprising only 58 cells — was identified as a fibroblast cluster based on its top markers, including decorin (*DCN*), collagen type 1  $\alpha 2$  (*COL1A2*), and lumican (*LMN*).

Based on the known endothelial marker *CD34*, three endothelial clusters were identified (clusters 6, 7, and 8); specific markers for these 3 clusters include *SERPINE2* (cluster 6), *PLVAP* (cluster 7), and *EHD3* (cluster 8) (Figure 3B). These 3 endothelial clusters were the focus of further evaluation of underlying biological processes and their relationship to disease progression.



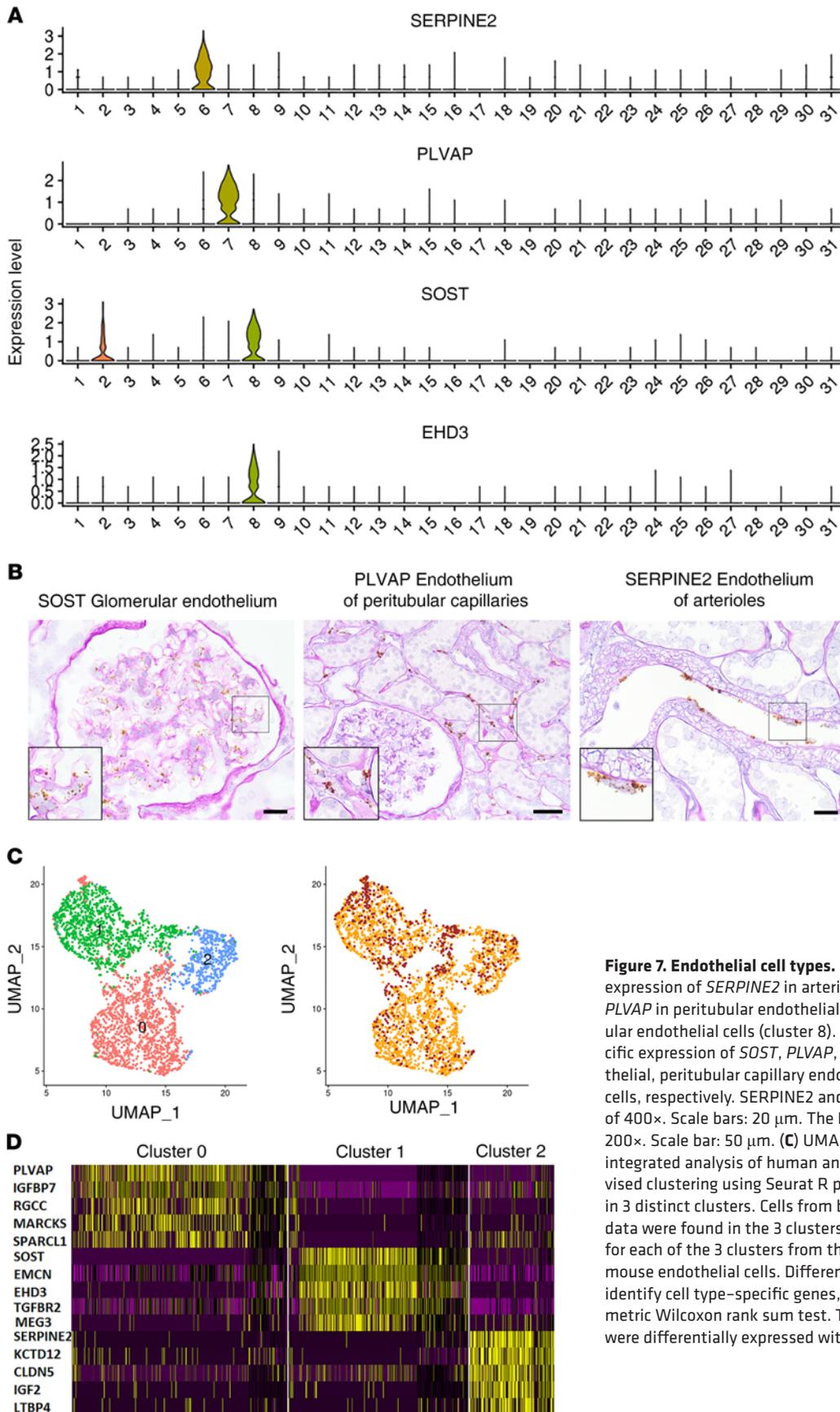
**Figure 5. Cell type markers of clusters 19 and 30.** (A) Violin plots of PC and IC cell markers: *FXYD4* expression in PC (cluster 27); *SLC26A7* in ICA cells (cluster 18); *SLC4A9* in ICB (cluster 17); and all 3 in transitional PC-IC cells (cluster 19). (B) tSNE plot of the 2 subclusters of cluster 30; dot plot of relative expression levels and HPA antibody staining of parietal endothelial cells (*CRB2*, *CLDN1*, and *WT1* in subcluster 0) and LOH cell markers (*SPP1*, *TACSTD2*, and *SLC12A1* in subcluster 1). (C) Violin plot shows the specific *NPHS2* expression in cluster 2 (podocytes). HPA antibody staining of *NPHS2* confirms its specific expression in podocytes. Violin plot of *CRB2* expression shows that this gene is expressed in clusters 2 and 30. HPA antibody staining of *CRB2* shows that it is expressed in podocytes and parietal epithelial cells. The HPA antibody stainings are from normal human kidneys, and figures indicate scale of 20  $\mu$ m; additional details in Supplemental Figure 6). PC, principal cells; IC, intercalated; PEC, parietal epithelial cells; LOH, Loop of Henle; HPA, Human Protein Atlas.



**Figure 6. Validation of cell cluster assignment.** (A) Heatmap of correlation between the average expression of genes in cell types identified in human developing kidney and adult kidney scRNASeq data (columns, developing kidney; rows, adult kidney). (B) Heatmap showing correlation between the average expression of genes in cell types identified in scRNASeq (rows) and snRNASeq (columns) analyses. Row-wise Z-score scaling of gene expression was used for heatmap visualization.

*Comparative assessment of adult kidney single cell data.* We used independently generated kidney single cell or single nucleus datasets to assess the cell type distribution and coverage of the scRNAseq data from adult kidney tissue samples. To compare the average gene expression of genes in major cell types identified in published developing kidney tissue data (10) with that of the cell types in our adult kidney scRNAseq data, heatmaps using Pearson's correlation values were generated. Positive correlation of average expression levels between the corresponding cell types in the 2 data sets — especially between the podocytes, distal tubular, collecting duct, endothelial, stromal, and immune cells — were observed (Figure 6A). In addition, we similarly compared our data with published adult human kidney snRNAseq dataset (19). Average transcript expression of the major cell types in the 2 data sets had correlation values greater than 0.7 (Figure 6B).

*Validation of endothelial cell types.* Unlike bulk mRNA analysis, the single cell technology was successful in characterizing multiple endothelial cell clusters. Confirming the distinct cellular identities of these clusters, we demonstrated the expression of 1 distinct marker for each of the 3 endothelial cell clusters, clusters 6–8 (Figure 7A), using in situ hybridization (ISH; Figure 7B). As predicted from the scRNAseq analysis, the ISH signal for *SOST* was found to be specific to glomerular endothelial cells (GECs) (cluster 6); *PLVAP* probes hybridized to peritubular capillary endothelial cells (cluster 7) and *SERPINE2* were specifically expressed in arteriolar endothelial cells (cluster 8). We were also able to validate these findings by clustering mouse kidney scRNAseq data available at NCBI, Gene Expression Omnibus (GEO) dataset (GSE107585) (14) and identifying 3 similar endothelial cell subclusters (Figure 7C). Furthermore, the top markers in the 3 human kidney endothelial clusters also marked peritubular capillary endothelial, glomerular endothelial, and arteriolar endothelial cells in murine scRNAseq data (Figure 7D).



**Figure 7. Endothelial cell types.** (A) Violin plots with cluster-specific expression of *SERPINE2* in arteriolar endothelial cells (cluster 6), *PLVAP* in peritubular endothelial cells (cluster 7), and *EHD3* in glomerular endothelial cells (cluster 8). (B) Validation using ISH for the specific expression of *SOST*, *PLVAP*, and *SERPINE2* in glomerular endothelial, peritubular capillary endothelial, and arteriolar endothelial cells, respectively. *SERPINE2* and *SOST* images have magnification of 400×. Scale bars: 20 μm. The *PLVAP* image has magnification of 200×. Scale bar: 50 μm. (C) UMAP plot showing the clusters from the integrated analysis of human and mouse endothelial cells. Unsupervised clustering using Seurat R package at a resolution of 0.1 resulted in 3 distinct clusters. Cells from both human and mouse scRNAseq data were found in the 3 clusters. (D) Heatmap of the top 5 markers for each of the 3 clusters from the integrated analysis of human and mouse endothelial cells. Differential gene expression analyses, to identify cell type-specific genes, were performed using the nonparametric Wilcoxon rank sum test. The cell type-specific genes that were differentially expressed with FDR < 0.05 were considered.

*Integrative analysis of single cell endothelial marker with bulk mRNA identifies a distinct FSGS subgroup with activated GECs.* Endothelial dysfunction has been shown to contribute to the development and progression of kidney diseases (23). Focusing on the 3 endothelial cell clusters in our data (Figure 8A), we identified 26 genes selectively expressed in all 3 endothelial clusters and 78 genes specific to GECs (Figure 8A, yellow). The set of GEC-specific genes were used to derive an aggregate GEC score to extract the differential regulation of these transcripts in bulk kidney biopsy transcriptomic data sets. The GEC scores were determined for the glomerular transcripts across 10 glomerular kidney diseases studied in the European Renal cDNA Bank (ERCB, Figure 8B) and the NEPTUNE (Figure 8C) cohorts. In both cohorts, the highest average and median GEC scores were observed in patients with focal segmental glomerulosclerosis (FSGS), and GEC scores were significantly higher in glomerular disease samples compared with biopsies from living kidney donors ( $P < 0.005$ ) (Figure 8, B and C). Interestingly, FSGS patients with exposure to immunosuppressive treatment showed significantly lower GEC scores compared with treatment naive patients at time of biopsy, consistent with an acute inflammatory nature of the endothelial score signature (Figure 8C). Given the functional heterogeneity of patients with FSGS, we evaluated the GEC transcripts for their ability to distinguish groups within FSGS using bulk RNAseq profiles available from NEPTUNE. Two distinct clusters (Figure 9A), group 1 (40 patients; median GEC score,  $-0.26$ ) and group 2 (28 patients; median GEC score,  $0.37$ ) were identified. In a genome-wide comparison, we found 3745 genes significantly differentially expressed (FDR  $< 0.05$ ) between group 1 and group 2 (top 100 differentially expressed genes shown in Supplemental Table 3).

For the top 100 overexpressed genes between groups 1 and 2, we were able to identify 3 significant kidney-specific functional modules: Angiogenesis, extracellular matrix, and cell cycle (Figure 9B). An interaction network generated using STRING (24) for the same top 100 genes is shown in Figure 9C. This network included direct interactions between the top endothelial markers, including *NRPI*, *TEK*, and *KDR*, with *FNI*, a known marker implicated in glomerulosclerosis. Additionally, extracellular matrix- and mitotic cycle-associated proteins were enriched in the interaction network (Figure 9C).

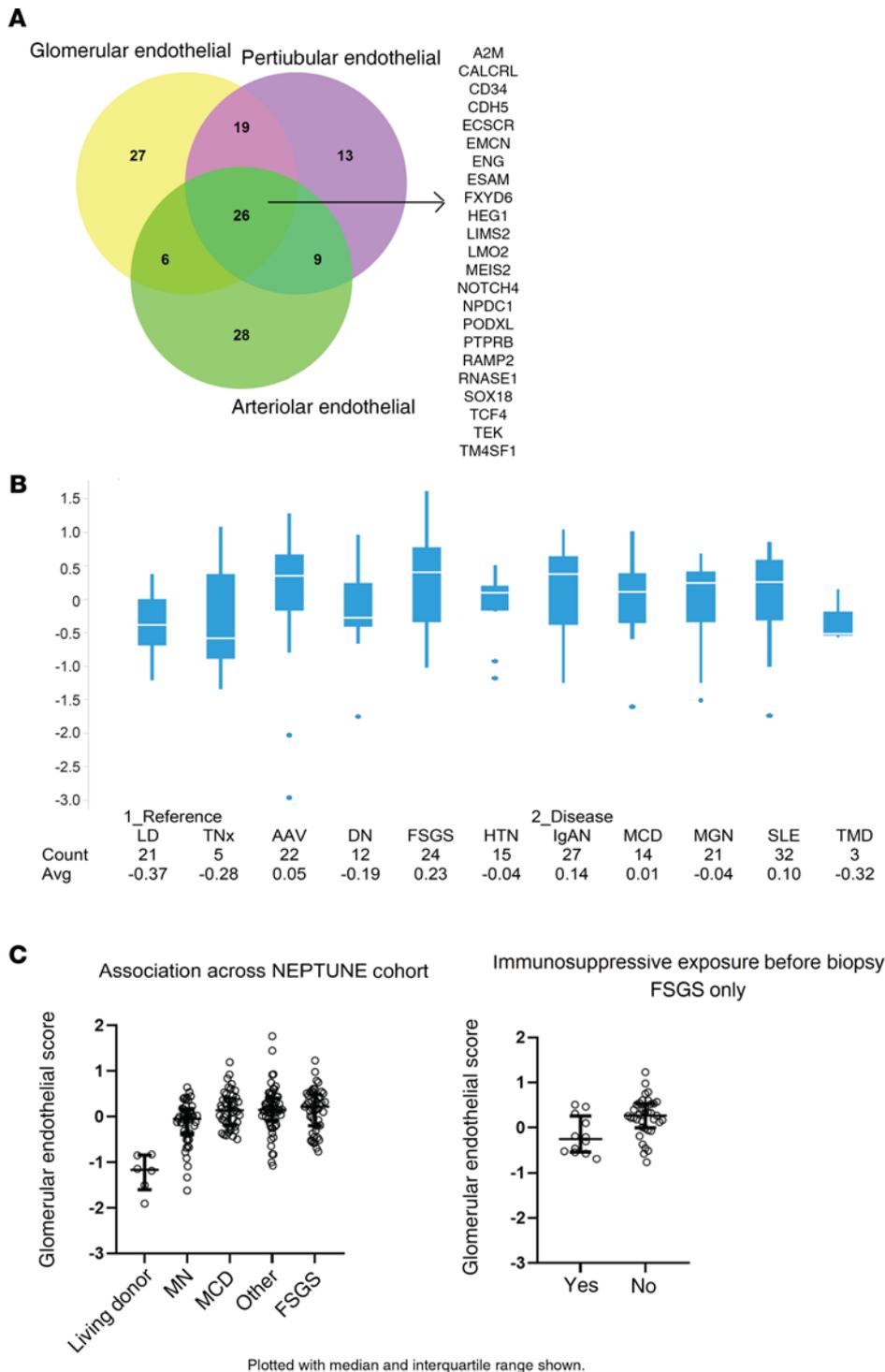
We used weighted gene coexpression network analysis (WGCNA), which identifies coexpression modules that encapsulate the correlation patterns among genes across samples in order to further elucidate the functional context of the endothelial network contained in the 2 FSGS groups. The top 2 gene set modules that significantly correlated with the groups (1 and 2) and GEC scores (for each sample) were brown module, with 517 genes positively correlated with the GEC score and the groups, whereas a blue module with 1106 genes negatively correlated with these traits (Supplemental Figure 3A). The positively correlated genes were enriched for Gene Ontology (GO) biological processes, including glomerulus vasculature, Ras protein signaling, and type 1 IFN response. The negatively correlated genes were mainly enriched for the TGF- $\beta$  signaling pathway (Supplemental Figure 3A). WGCNA using just patient data with no immunosuppressive treatment exposure at time of biopsy from group 1 and group 2 showed the enrichment of similar biological processes (Supplemental Figure 3B).

We then examined predicted upstream regulators of the genes differentially expressed between group 1 and group 2. Evaluating enrichment and direction of differential gene expression changes using known cause-and-effect relationships, the top upstream regulators were predicted using Ingenuity Pathway Analysis (IPA). A positive activation score indicates that the upstream regulator was activated for the observed differential gene expression pattern. STAT1 and VEGF were the top 2 upstream regulators with high activation scores (Figure 9D).

The mechanistic network of both STAT1 and VEGFA shown in Figure 9E identified *A2M* as a relevant downstream target gene. *A2M* was also the top differentially expressed gene in FSGS samples when comparing group 2 with group 1, based on log counts per million (log cpm) and log fold change (logFC).

To summarize, from this multilevel integrative analysis, we identified 2 subgroups of FSGS patients with distinct molecular expression patterns differentiated by the expression of GEC genes. Multiple lines of evidence implicated A2M as a downstream target differentially regulated in those FSGS subgroups.

*Intrarenal A2M mRNA expression associates with FSGS disease progression.* Figure 10A shows the average mRNA expression of A2M in the 2 groups identified based on GEC scores among FSGS participants in NEPTUNE. Significantly higher A2M expression is observed in group 2 compared with group 1. A2M IHC in normal human kidney tissue in HPA indicates robust glomerular protein staining in a pattern following the mRNA expression (Figure 10B). In scRNAseq A2M expression was found in the 3 endothelial cell subtypes and VSMC/MCs in adult human kidney (Figure 10C). A2M ISH in sections

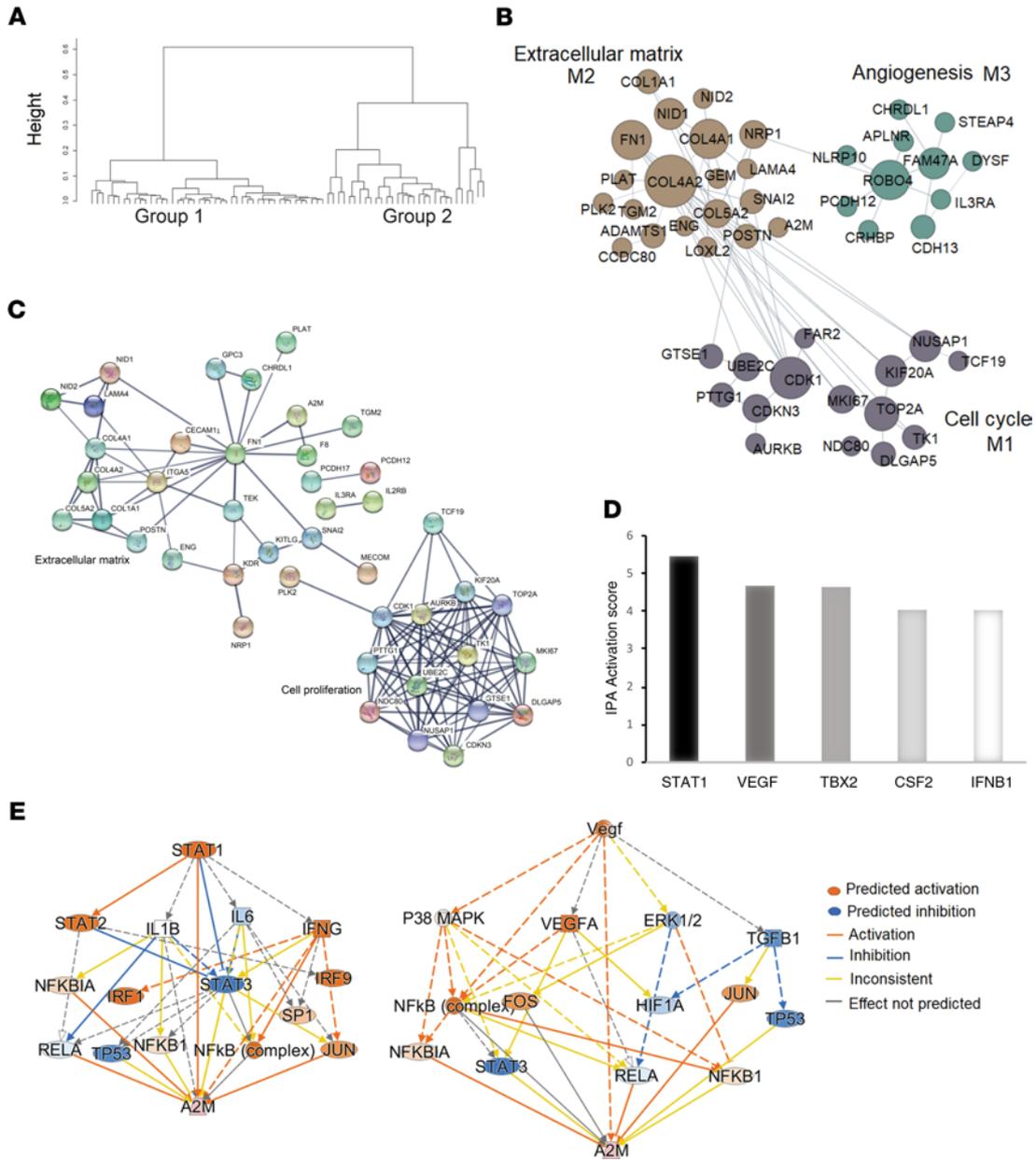


**Figure 8. Analysis of glomerular endothelial marker genes.** (A) Venn diagram intersect of endothelial-specific clusters from scRNAseq data highlighted 30 genes specific to all 3 endothelial clusters. (B and C) Glomerular endothelial cell (GEC) scores, derived based on 78 genes selective for glomerular endothelial cells (yellow), were calculated for patients and compared across the kidney disease spectrum in the European renal cDNA bank cohort (ERCB) (B) and the NEPTUNE cohort (C). GEC scores were elevated in disease groups compared with those of living donors. The GEC scores in FSGS patients with exposure to immunosuppressants prior to time of biopsy was substantially lower compared with those with no prior exposure. AAV, ANCA-associated vasculitis; DN, diabetic nephropathy; FSGS, focal segmental glomerulosclerosis; HTN, hypertensive nephropathy; IgAN, IgA nephropathy; MCD, minimal change disease; MN-MGN, membranous glomerulonephritis; SLE, systemic lupus erythematosus; TMD, thin basement membrane disease; LD, living donor; TNx, unaffected parts of tumor-nephrectomy.

from reference and FSGS kidney tissue showed a pattern consistent with the endothelial and mesangial scRNAseq expression (Figure 10D). Replicating the NEPTUNE results, higher expression levels of A2M transcripts were observed in samples from patients with FSGS and diabetic nephropathy (DN) in the ERCB cohort compared with living donors (Figure 11A and Table 2). Finally, high A2M mRNA expression at the time of initial biopsy were associated with poor prognosis, as evidenced by the longer time to reach complete remission of proteinuria in NEPTUNE study participants (Figure 11B).

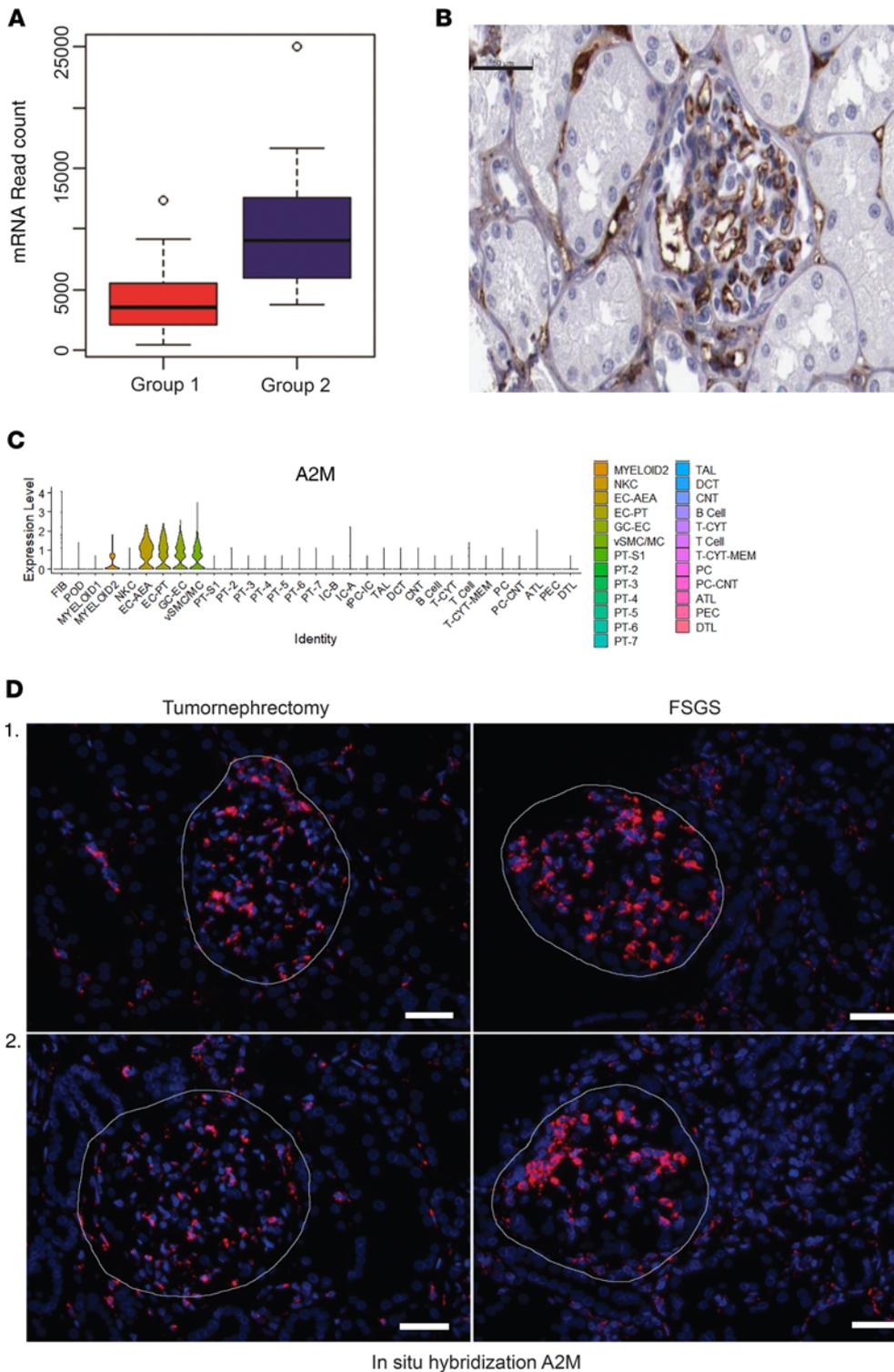
## Discussion

We successfully generated a reference kidney single cell expression atlas with 22,268 cells derived from 31 distinct kidney cell types. The cell clusters covered the entire spectrum of kidney cell types found



**Figure 9. Integrated analysis of scRNAseq and bulk mRNAseq data.** (A) Two distinct groups from the hierarchical clustering of the 78 glomerular endothelial cell defining mRNA transcripts in isolated glomeruli of FSGS biopsy samples from the NEPTUNE cohort. (B) Kidney-specific functional module gene interaction network (<https://hb.flatironinstitute.org>). The figure shows the 3 significant functional modules and their interactions. (C) Protein interaction network generated using STRING for the top 100 genes that were substantially overexpressed in group 2 versus group 1. (D) Top upstream regulators predicted by IPA. The bar plots show the activation scores predicted by IPA for the top 5 upstream regulators ( $P < 0.05$  by Fisher's exact test). (E) The regulatory interaction of the top 2 regulators, VEGF growth factor, and STAT1 show A2M as a downstream target. FSGS, focal segmental glomerulosclerosis.

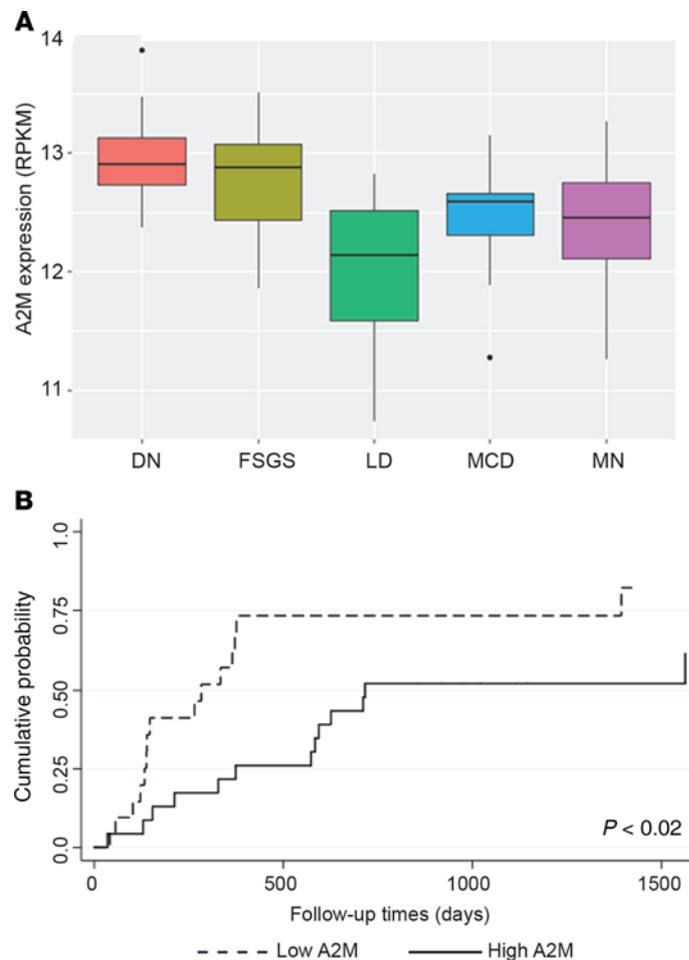
along the nephron, as well as resident immune cells. Clustering of the single cell data from a single sample type may miss a rare cell type, like podocytes, due to lack of power based on number of cells. However, the combined analysis of the 22,268 cells enabled the detection of a distinct podocyte cell cluster. Additionally, generating single cell transcriptomics data from adult human kidney tissue can be challenging due to the rapid decline of cell viability upon tissue dissociation (17). In this study, we were able to analyze single cell transcriptomic data from 16 tumor-nephrectomies, 5 surveillance biopsies derived from transplant patients with normal kidney function and histology, and 3 preperfusion living donor biopsy samples and assign cells to distinct clusters without significant confounders due to the tissue source. Furthermore, by integrating the cell type-specific information of GECs with bulk



**Figure 10.  $\alpha$ -2-Macroglobulin (A2M) expression.** (A) Box plot showing average glomerular RNAseq read count of A2M transcripts in samples from the GEC group 1 and group 2. (B) Human Protein Atlas antibody staining of A2M in normal human kidney. Scale bar: 50  $\mu$ m. (Supplemental Figure 6 contains specific information on HPA immunostaining of A2M). (C) Violin plot showing the cell type scRNAseq expression of A2M transcripts in adult human kidney. (D) In situ hybridization fluorescence images showing the glomerular signal of A2M probes in 2 unaffected parts of tumor-nephrectomy (left panel) and FSGS (right panel) kidney tissue sections. Glomeruli are encircled by thin white lines. The images are at 200 $\times$  magnification. Scale bar: 50  $\mu$ m.

mRNA data, we were able to identify 2 distinct groups of FSGS patients, with GEC activation being most prominent in immunosuppressive naive patients.

The need to generate single cell suspensions from complex tissues can introduce a selection bias by the isolation procedure. In KPMP parallel strategies of cell-level analyses are pursued and allow direct comparison between the different technologies. Comparison of our kidney single cell data with published single nucleus data (19) show a significant overlap with robust cell-to-cell correlation but also are complementary, with certain cell types more enriched in one compared with the other technology. For example, enrichment of immune cells is observed in scRNAseq compared with snRNAseq,



**Figure 11. A2M expression in FSGS.** (A) Box plot showing the expression of A2M mRNA in different kidney diseases in ERCB cohort: DN, FSGS, MCD, MN, LD. (B) Kaplan-Meier curves by A2M mRNA expression levels at baseline biopsy for the first complete remission event in NEPTUNE FSGS patients. Two-sided Student's *t* test showed that high A2M mRNA expression at the time of initial biopsy was associated with poor prognosis, as evidenced by the longer time to reach complete remission of proteinuria in NEPTUNE study participants ( $P < 0.02$ ). DN, diabetic nephropathy; FSGS, focal segmental glomerulosclerosis; MCD, minimal change disease; MN, membranous nephropathy; and LD, living donor.

providing a strong rationale to study the adult human kidney at the single cell level for a thorough understanding of the kidney transcriptome.

One of the key features in our study is the integration of the single cell data, with glomerular mRNA biopsy gene expression data dissecting the contribution of specific cell types to the disease state. In this study, we focused on endothelial cell phenotypes in the kidney, given their known involvement, and the limited mechanistic elucidation of their role in human glomerular disease pathophysiology, thus far. We observed 3 distinct endothelial cell types across all 3 reference tissue sources and confirmed the same cluster pattern by reanalyzing a published murine study (14, 25) with the recent study of murine kidney by Dumas et al., showing extensive renal endothelial cell phenotypic heterogeneity (25).

The scRNAseq endothelial subtype-specific gene sets enabled the development of a GEC score. The GEC score was subsequently applied to bulk gene expression data sets from microdissected glomeruli (Glom) obtained from patients with a wide spectrum of glomerular diseases. We not only could observe differential regulation of the GEC markers across diseases, but we could also identify 2 subgroups of patients with activation of the GEC-specific genes in FSGS, a disease known for its substantial heterogeneity in disease initiation, treatment responsiveness, and rates of progression (26, 27).

In the FSGS patient kidney biopsies, known endothelial marker genes *NRP1*, *KDR*, and *TEK*, encoding the receptors for growth factor ligands including VEGFA, VEGFD, ANGPT1, and ANGPT2,

**Table 2. A2M expression in different kidney diseases in the ERCB cohort**

ANOVA results			FSGS	MN	MCD	DN
Group	A2M	LD	$6 \times 10^{-7}$	0.05	0.04	0
P value	0.006	DN	0.83	0.02	0.11	NA
F value	7.714	MCD	0.4	0.99	NA	0.1

Two-way ANOVA indicates significant difference ( $P < 0.001$ ) in A2M transcript levels in FSGS and DN patients compared with LD. DN, diabetic nephropathy; FSGS, focal segmental glomerulosclerosis; MCD, minimal change disease; MN, membranous nephropathy; and LD, living donor.

were overexpressed in FSGS group 2 compared with group 1 (Supplemental Figure 4). The single cell analysis enabled us to identify the cell type-specific expression of these ligands and receptors. Supplemental Figure 4A illustrates the potential cell-to-cell interactions between the endothelial cells and podocytes, fibroblast, and MCs.

The endothelial cell-driven gene signatures juxtapose patients with an inflammatory GEC activation state from those with profibrotic glomerular transcriptional signatures (28). The 2 FSGS groups were characterized by distinct expression signatures with group-specific functional enrichments. Both kidney-specific functional module analysis and the STRING interaction network illustrated similar functional enrichments. Extracellular matrix functional processes, endothelial development, and cell division were significantly enriched for the genes that were overexpressed between group 2 and group 1. The functional enrichment of the genes in the WGCNA module showed glomerular vasculature development, type 1 IFN, and RAS protein signaling pathways as enriched in group 2 compared with group 1 (Supplemental Figure 3A). TGF- $\beta$  signaling response genes, which have an established profibrotic role in FSGS pathology, were higher in group 1 compared with group 2 (29).

The enrichment of type 1 IFN signaling in group 2 compared with group 1 and the identification of STAT1 as the top upstream regulator for the differentially expressed genes in this group is consistent with established STAT1 activation by IFNs (30, 31). Tao et al. have described activation of the JAK/STAT signaling pathway in FSGS in the NEPTUNE cohort (32), with higher STAT1 transcriptional activity scores and elevated levels of phosphorylated STAT1 in glomerular and tubular compartments from FSGS patients. Together, these findings suggest that type 1 IFN-mediated STAT1 activation could be a key mechanism involved in glomerular disease progression for group 2 patients with endothelial cell activation. In a phase II clinical trial for patients with diabetic kidney disease, treatment with a JAK/JAK2 inhibitor resulted in a reduction in albuminuria (4) and was well tolerated, suggesting that this mechanism may also be amenable to therapeutic intervention in patients with FSGS. Intriguingly, the GEC score was highest in immunosuppressive treatment-naïve NEPTUNE participants and significantly lower in patients with steroid and/or second-line immunosuppression at time of biopsy (Figure 8C). This argues not only for a high inflammatory state of the endothelium in a substantial subset of untreated FSGS patients (Supplemental Figure 3B), but it is also consistent with the ability of immunosuppressive treatments to rapidly impact the GEC activation in FSGS. Crosstalk between endothelial cells with MCs and podocytes supports a highly dynamic role of the glomerular endothelium in FSGS (Supplemental Figure 4).

Analysis of the differences in the gene expression of the 2 GEC-defined groups pointed to A2M as a downstream mediator linked to the FSGS heterogeneity. Upstream regulator analysis predicted A2M as a target gene of STAT1. This observation is further supported by Satoh and Tabunoki (33), who identified A2M as a target gene for the STAT1 transcription factor in their ChIP-seq analysis. A2M protein is a known marker of glomerular permselectivity, a property of the glomerular capillary wall that serves to exclude macromolecules from the urinary space. The ratio of urine A2M protein to albumin has been reported as a predictor of steroid response since as early as 1964 (34), with filtered A2M in the urine as an indicator of severe loss of glomerular permselectivity. The limited immunohistochemical staining of A2M in kidney biopsies have, thus far, been viewed as systemic and incidental to loss of glomerular permselectivity, rather than the consequence of intrarenal A2M synthesis (35). Our mRNA gene expression data sets from scRNAseq, microdissected glomerular RNAseq, and ISH argue for an intrinsic intrarenal regulation of A2M in glomerular disease (Figure 10D). A2M is a nonspecific protease inhibitor able to impact the function of a wide range of proteins (36). Relevant in a glomerular context is the ability of A2M to inhibit matrix-degrading proteases, including MMPs (37). Inhibition of MMP is one mechanism leading to the glomerular matrix accumulation characteristic of FSGS. The

STRING protein interaction and kidney-specific functional module analyses in the glomerular NEPTUNE data sets indicate a close link between A2M and glomerular matrix changes in the FSGS population studied.

Podocyte injury is the earliest observed morphological feature of FSGS (26, 38). A2M is an effective inhibitor of activated protein C (aPC) (39), with recent data indicating aPC as protective against apoptosis of podocytes in DN (40). Thus, inhibition of aPC by A2M may block the protective role of aPC. In addition, repression of aPC is known to enhance thrombin generation (41), and Kerlin and colleagues recently have shown that thrombin injures podocytes (42). A2M-mediated inhibition of aPC-mediated podocyte protection may simultaneously enhance thrombin-mediated podocyte injury.

There are several limitations to be considered with the described study. Multiple tissue sources were used to represent the normal tissue state of the kidney. While each of these sources has its specific limitations, contribution of all 3 tissue sources to the cellular clusters identified argues for the robustness of the cellular differentiation reported.

One intrinsic limitation of single cell studies is the need to generate single cell suspension, which may cause selective loss of some cell populations (17). For example, podocytes are highly sensitive to tissue dissociation processes. However, the proportion of podocytes recovered by our protocol was aligned with the abundance of these cells in adult kidneys reported by others (20, 43). The resolution used in the unsupervised clustering step could have resulted in the overclustering of proximal cells to 7 clusters rather than to the known 3 segments, namely S1, S2, and S3. However, a lower resolution would have missed the identification of other major kidney cell types. Comparison with the snRNAseq data indicates a good overall agreement between clusters and cluster-associated transcripts. Further studies to systematically compare and integrate the different single cell technologies on the same adult human tissue are ongoing in KPMP. Another caveat of this study is the impact of tissue dissociation at 37°C. Our initial bulk mRNA analysis on pre- and postdissociation kidney tissue indicated no effect; however, more detailed analysis is warranted to rule out the impact of tissue dissociation at 37°C (Supplemental Figure 5, A–C).

In this study, we generated a comprehensive single cell transcriptome of the adult human kidney and identified known and potentially novel cell type-specific markers, along with transient cell types between renal segments. Our scRNAseq-driven endothelial cell characterization in FSGS Glom revealed potentially novel molecular heterogeneity within patients with FSGS. We identified 2 distinct FSGS groups with significant differences in the intrarenal *A2M* gene expression levels. The group with the higher *A2M* gene expression associated with poor prognosis. As molecular pathways linked to FSGS are targeted by several emerging therapeutics, this integrative analysis provides avenues for investigation whereby not only insights into differences in normal and diseased adult human kidneys can be gathered, but such explorations can also serve as frameworks for precision medicine in order to identify the appropriate drug for a specific patient diagnosed with a glomerular disease.

## Methods

**Human samples.** For generating the single cell transcriptome, we processed 24 CryoStor preserved human kidney samples: 16 tumor-nephrectomy, 5 human allograft kidney transplant surveillance, and 3 preperfusion transplant biopsy samples. Ten of the 16 tumor-nephrectomy samples were processed at the Broad Institute, and all the remaining samples were processed at the University of Michigan. The surveillance biopsies were obtained from patients at 3–6 months after transplant; all 5 biopsy samples were confirmed as histologically normal. Additionally, 3 preperfusion biopsies procured from donor kidneys before transplant were used as most representative of normal kidney tissue. Only small quantities (2–3 mg) of the biopsy samples were used for the scRNAseq procedure.

As part of the Accelerating Medicines Partnerships lupus network Pathway Exploration and Analysis in Renal disease (AMP-PEARL) network, we have developed and optimized protocols for kidney biopsy sample acquisition in distributed research networks for dissociation of single cell suspension and subsequent RNAseq studies (15). We modified the protocol published by the AMP-PEARL network to increase kidney epithelial cell yield. Briefly, kidney biospecimen were placed in 1 mL HypoThermosol (Stemcell Technologies) and immediately transferred to the laboratory and cryopreserved in DMSO containing CryoStor CS 10 (Stemcell Technologies) solution for long-term storage in liquid nitrogen (Figure 1).

**Single cell isolation.** For tissue dissociation, samples were quickly thawed in a 37°C water bath for 1 minute, transferred to a small dish containing 1 mL DMEM/F12 medium (with L-Glutamine, HEPES, and high glucose) supplemented with 10% heat-inactivated FBS, and incubated for 10 minutes at room

temperature. Tissue/biopsy samples were cut into 1-mm<sup>3</sup> pieces to increase the surface area before they were transferred to a 1.5-mL tube containing 500  $\mu$ L DMEM/F12 digestion media, which contained 125 ng Liberase TL (MilliporeSigma) enzyme. Samples were incubated on an orbital shaker (500 rpm) for 12 minutes at 37°C with a mild trituration step using a wide-bore 1000  $\mu$ L tip applied after 6 minutes. Digestion was stopped by adding 500  $\mu$ L DMEM/F12 medium containing 10% FBS and incubated at room temperature for 1 minute. The cell suspension was then filtered through a 30- $\mu$ m strainer (Milteny Biotec) into a 15-mL conical tube. Remaining solid tissue was passed through the nylon mesh using the rubber end of a 3-mL syringe pestle, and the filter was washed with 10-mL DMEM/F12/10%FBS medium. Cells were pelleted by centrifugation for 10 minutes at 200 g at 4°C. The supernatant was immediately, completely removed by vacuum suction using a glass Pasteur pipet. The cell pellet was then gently resuspended in 55  $\mu$ L DMEM/F12/10%FBS medium, and cell viability was analyzed with a Countess II FL automated cell counter exploiting the trypan blue dye exclusion method.

*10× Genomics and Next Generation Sequencing (NGS) of single cell mRNA.* Single cell samples were immediately transferred to the University of Michigan Sequencing Core facility. About 10,000 viable cells in up to 46  $\mu$ L DMEM/F12/10%FBS per sample were added to each channel of a chip on the droplet-based high-throughput 10× Genomics Chromium platform allowing cell lysis, individual cell barcoding, and reverse RNA transcription. We used the 10× Single Cell 3' GEX – version 2 chemistry for all tumor-nephrectomy samples and the enhanced 10× Single Cell 3' GEX – version 3.1 for the more recently processed transplant surveillance biopsies and preperfusion biopsies. Single cell library preparation was executed by the core facility and included emulsion breakage, PCR amplification, cDNA fragmentation, oligo adapter, and Illumina sample index addition. Libraries have been pooled and sequenced on an Illumina HiSeq4000 platform as asymmetric paired end runs (26  $\times$  115 bases) with a median of 100 million raw sequencing reads per sample. The output from the sequencer was first processed by Cell Ranger, the proprietary 10× Chromium single cell gene expression analysis software (<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger>).

*Tissue dissociation effect analysis.* To account for the potential for genes to be induced during tissue dissociation using Liberase TL at 37°C, we compared bulk gene expression RNAseq data before and after cell dissociation for 4 tumor-nephrectomy samples (Supplemental Figure 5A). We found a stringent correlation between pre- and postdissociation gene expression for each of the 4 samples (Supplemental Figures 5, B and C). Pre- and postdissociation bulk mRNA analysis show that the short duration of 12 minutes of dissociation using Liberase TL did not have a significant impact downstream biological processes.

*Data analysis of single cell data.* Data analyses were performed on the Cell Ranger output data files using Seurat 3 R package (<https://cran.r-project.org/web/packages/Seurat/index.html>). As a quality control step, cells with less than 500 genes per cell were filtered out. The percentage of mitochondrial gene read content has emerged as a quality measure for the cell viability in single cell studies. A high percentage of mitochondrial reads indicate that the cells are less viable, indicative of cytoplasmic mRNA loss due to the degradation of plasma membrane during the dissociation process. For the tumor-nephrectomy data generated applying the Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit version 2, a threshold of < 20% mitochondrial reads per cell was used, whereas a cutoff of < 50% mitochondrial reads per cell was applied for the data from the version 3 kit. It has been observed by us and other groups that, for certain tissues, including kidney, 10× Genomics version 3 chemistry results in a higher percentage of mitochondrial reads, irrespective of the viability status of the cells. This issue was circumvented by deeper sequencing of the samples (200 million reads) to obtain more genes per cell. After the quality control step, the data from the 3 sample types (tumor-nephrectomy, surveillance biopsy, and preperfusion biopsy) were merged and further analyzed. Doublet formation is a concern in single cell technology, where 2 cell types in close proximity can be engulfed by a single droplet. The mRNA content and number of genes of doublets are comparatively higher than mRNA content and number of genes of single cells. In order to reduce doublets or multiplets from the analysis, we used a cutoff of > 500 and < 5000 genes per cell. However, in certain cell types (for example, in the cells transitioning between 2 cell types), the mRNA content can be higher than other cell types (cluster 19, in this study).

The downstream analysis steps included SCTransform, dimensionality reduction principal component analysis (PCA) and UMAP (44), standard unsupervised clustering, and the discovery of differentially expressed cell type-specific markers (45). We used the SCTransform functionality embedded in Seurat R package for normalization and scaling of UMI and mitochondrial content. In addition, we included the

batch information as a variable in the SCTransform to scale out the differences between the batches. For the unsupervised clustering, we chose 1.75 as the resolution parameter. Differential gene expression analyses, to identify cell type-specific genes, were performed using the nonparametric Wilcoxon rank sum test. Harmony, a R package, was used to integrate single cell datasets before the unsupervised clustering (46).

For cell type annotation, we used publicly available resources, including Kidney Interactive Transcriptome (<http://humphreyslab.com/SingleCell/>), HPA (<https://www.proteinatlas.org>), the Epithelial Systems Biology Laboratory (ESBL) (<https://hpcwebapps.cit.nih.gov/ESBL/Database/>), scRNAseq data (19), and Immgen (<https://www.immgen.org/>).

*Bulk mRNA profiling from patients with kidney disease.* Total RNA was isolated from microdissected kidney Glom of biopsy samples from the ERCB cohort and from the NEPTUNE cohort using well-established protocols that demonstrate enrichment for abundance of glomerular mRNAs (47, 48). Microarray profiles were generated on glomerular samples as previously described (49) and were deposited into GEO under accession numbers GSE104948 (ERCB) and GSE104066 (NEPTUNE). Bulk glomerular RNAseq expression data were obtained for 74 participants with biopsy-proven FSGS enrolled in the NEPTUNE study. Briefly, mRNA samples were prepared using the Illumina TruSeq mRNA Sample Prep v2 kit. Multiplex amplification was used to prepare cDNA with a paired-end read length of 100 bases using an Illumina HiSeq2000. RNAseq was performed by the University of Michigan Advanced Genomics Core (<https://brcf.medicine.umich.edu/cores/advanced-genomics/>). Quality of the sequencing data was assessed using the FastQC tool (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Read counts were extracted from the fastq files using HTSeq (version 0.11). The objectives and study design of NEPTUNE can be found in the clinicaltrials.gov database under NCT1209000 (50).

*Development of the GEC transcription score.* Aggregate GEC scores were generated in samples profiled on microarrays from patients across the kidney disease spectrum who were part of the ERCB and NEPTUNE cohorts. We established a glomerular GEC score from 78 glomerular endothelial marker genes identified from single cell data using previously described methods (12, 49). Genes expressed on glomerular microarray profiles from ERCB (60 of 78 genes) and NEPTUNE (77 of 78 genes) were Z-transformed; then, the average Z-score per disease group was used for comparison among diseases.

*WGCNA of bulk mRNA data.* Gene set modules were generated by WGCNA (51) of the top 5000 variable genes between the groups 1 and 2 identified through hierarchical clustering of bulk mRNA expression profiles of the 74 FSGS patients. Enriched GO biological processes were identified for the coexpressed gene sets.

*Kidney-specific functional modules.* Module detection was performed using a kidney tissue-specific gene functional network (52, 53). We consider the subgraph of the kidney-specific network consisting of the top 100 genes significantly overexpressed in group 2 FSGS patients relative to group 1 FSGS patients, as well as all edges between them. Four of the top genes were not present in the kidney-specific functional network, leaving a network of 96 genes. For each pair of these genes, we then counted the number of shared neighbors among the top k edges in the subgraph incident to each gene, and we chose the 5% of pairs with the most shared neighbors as edges in a shared k-nearest neighbors (sKNN) graph. (Here, k is set to 20% of the number of genes). We then performed community clustering using the Louvain algorithm to identify modules in the sKNN graph. Community clustering was repeated 100 times, and a comembership score was computed for each pair of genes. The comembership score was equal to the fraction of runs in which the gene pair was assigned to the same module. To produce output modules, community clustering was performed on the graph with edges between each pair of genes with a comembership score of at least 90%.

*Validation of endothelial cell types using ISH.* In situ detection of sclerostin (*SOST*, RNAscope probe\_Hs-SOST, Advanced Cell Diagnostics [ACD], catalog 452941), plasmalemma vesicle associated protein (*PVLAP*, RNAscope probe\_Hs-PLVAP, ACD, catalog 437461), serpin family E member 2 (*SERPINE2*, RNAscope probe\_Hs-SERPINE2, ACD, catalog 412661) and  $\alpha$ -2-macroglobulin (*A2M*, RNAscope probe\_Hs-A2M-C2, ACD, catalog 532501-C2) mRNA transcripts using the RNAscope kit (Advanced Cell Diagnostics) were performed according to the manufacturer's protocol by kidney pathologists able to identify cell types in kidney tissue. Sections (3  $\mu$ m) sections of deidentified human kidney tissue were cut from formalin-fixed, paraffin-embedded (FFPE) blocks supplied by the University of Michigan Tissue Procurement Service. Housekeeping gene peptidylprolyl isomerase B (PPIB) was used as an internal, mRNA control; if the PPIB score was 0, the sample was regarded as not available for gene expression. A horseradish peroxidase-based signal amplification system (RNAscope 2.0 HD Detection Kit-Brown, ACD, catalog 310035) was used for hybridization to the target probes,

followed by color development with DAB, and the slides were counter-stained with hematoxylin (RICCA chemical company, catalog 3535-16). Positive staining was determined by brown punctate dots in the cytoplasm.

*Data sharing.* The gene matrix files from the scRNAseq data and the raw read counts from the bulk mRNA data used in the above analyses are uploaded to NCBI GEO database under accession number GSE140989. Data can also be viewed at <http://nephrocell.mikmtmc.org>. The single cell data are also available at the KPMP Atlas data repository (<https://atlas.kpmp.org/repository>).

*Statistics.* The workflow of our integrative analysis of single cell and bulk mRNA is described in Figure 1. EdgeR R package (54) was used for genome-wide differential gene expression analysis between the FSGS subgroups resulting from the hierarchical clustering of GEC type gene expression. Model-based normalization function embedded in the edgeR adjusted for the sequencing depth and RNA composition differences between the samples. Next, an empirical Bayes strategy for squeezing the tagwise dispersions toward a global dispersion trend or toward a common dispersion value was applied. After negative binomial models were fitted and dispersion estimates were obtained, testing procedures for determining differential expression using the exact test was performed. We considered only the genes that were differentially expressed with an adjusted  $P < 0.05$  (54). Association between the significantly differentially expressed genes with clinical parameters, including time to complete remission of proteinuria, were performed using STATA, v12.1 with 2-tailed Student's  $t$  tests of hypotheses and  $P < 0.05$  as the criterion for statistical significance. Comparison of upstream regulators related to differentially regulated genes were explored using IPA Software Suite (<https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/>).

*Study approval.* All samples used in this study were obtained with patient consent and with the approval of IRBs of participating institutions, described below. The KPMP study is reviewed and approved through central IRB process at The Human Research Protection Office and Washington University in St. Louis (IRB no. 201902013). Transplant biopsies were obtained after review and approval of the Transplant Transcriptomic Atlas study, HUM00150968, by the University of Michigan IRBMED. NEPTUNE data were obtained based on the multisite, central IRB approval for the study, HUM00158219, by the University of Michigan IRBMED.

## Author contributions

Designing the research study was contributed by RM and MK; sample acquisition was contributed by PH, NH, JBH, ASN, and SJ; sample processing was contributed by PH and EAO; conducting of experiments was contributed by EAO, ASN, JL, JBH, and BG; data analysis was contributed by RM, RS, SE, CCB, FE, LM, and ML; analysis review was contributed by OT, SJ, JBH, NH, and MK; clinical input was contributed by LM, JBH, and MK; scientific input was contributed by JH, WJ, VN, MK, and RM; and manuscript writing was contributed by RM, EAO, SE, LS, and MK. All authors contributed to manuscript revisions and accept accountability for the overall work by ensuring that questions pertaining to the accuracy or integrity of any portion of the work were appropriately investigated and resolved. RM is listed first as the co-first author, as she designed the data analysis strategy and wrote bulk of the manuscript. KPMP contributed the tissue and the methods development of the scRNAseq studies; NEPTUNE contributed the glomerular and tubulointerstitial gene expression data sets and matching clinical phenotypes.

## Acknowledgments

KPMP, UG-DK-114907, is a multiyear project funded by the National Institute of Diabetes, Digestive, and Kidney Diseases (NIDDK) with the purpose of understanding and finding new ways to treat chronic kidney disease and acute kidney injury. See Supplemental Acknowledgments for consortium details. The NEPTUNE consortium, U54-DK-083912, is a part of the NIH Rare Disease Clinical Research Network (RDCRN), supported through collaboration between the Office of Rare Diseases Research, National Center for Advancing Translational Sciences, and the NIDDK. See Supplemental Acknowledgments for consortium details. ERCB members at the time of the study: Clemens David Cohen, Holger Schmid, Michael Fischereder, and Lutz Weber (Munich, Germany); Matthias Kretzler (Ann Arbor, Michigan, USA); Detlef Schlöndorff (New York, New York, USA); Jean Daniel Sraer and Pierre Ronco (Paris, France); Maria Pia Rastaldi and Giuseppe D'Amico (Milano, Italy); Peter Doran and Hugh Brady (Dublin, Ireland); Detlev Mönks and Christoph Wanner (Würzburg, Germany); Andrew Rees (Vienna, Austria); Frank Strutz and Gerhard Anton Müller (Göttingen, Germany); Peter Mertens and Jürgen Floege (Aachen, Germany); Norbert Braun and Teut Risler (Tübingen, Germany); Loreto Gesualdo and Francesco Paolo Schena (Bari, Italy); Gunter Wolf (Jena, Germany); Rainer Oberbauer and Donscho Kerjaschki (Vienna, Austria);

Bernhard Banas and Bernhard Krämer (Regensburg, Germany); Moin Saleem (Bristol, United Kingdom); Rudolf Wüthrich (Zurich, Switzerland); Walter Samtleben (Munich, Germany); Harm Peters and Hans-Hellmut Neumayer (Berlin, Germany); Mohamed Daha (Leiden, Netherlands); Katrin Ivens and Bernd Grabensee (Düsseldorf, Germany); Francisco Mampaso (Madrid, Spain); Jun Oh, Franz Schaefer, Martin Zeier, and Hermann-Joseph Gröne (Heidelberg, Germany); Peter Gross (Dresden, Germany); Giancarlo Tonolo (Sassari, Italy); Vladimir Tesar (Prague, Czech Republic); Harald Rupprecht (Bayreuth, Germany); Hermann Pavenstädt (Münster, Germany); Hans-Peter Marti (Bern, Switzerland); Peter Mertens (Magdeburg, Germany); and Jens Gerth (Zwickau, Germany).

We thank Thomas Giordano and Daffyd Thomas, University of Michigan Tissue Procurement Service, for providing us with deidentified human kidney tissue in FFPE blocks and Yingbao Yang for the PAS staining of slides for the ISH data. This study was supported, in part, by the George M. O'Brien Michigan Kidney Translational Core Center, funded by NIH/NIDDK grant 2P30-DK-081943.

Address correspondence to: Matthias Kretzler, Warner-Lambert/Parke-Davis Professor of Medicine, Nephrology/Internal Medicine and Computational Medicine and Bioinformatics, University of Michigan, MSRB II, 4544-D, 1150 W. Medical Center Drive, Ann Arbor, Michigan 48109, USA. Phone: 734.615.5757; Email: kretzler@umich.edu.

1. Devarajan P. Neutrophil gelatinase-associated lipocalin: a promising biomarker for human acute kidney injury. *Biomark Med.* 2010;4(2):265–280.
2. Ju W, et al. Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. *Sci Transl Med.* 2015;7(316):316ra193.
3. Berthier CC, et al. Enhanced expression of Janus kinase-signal transducer and activator of transcription pathway members in human diabetic nephropathy. *Diabetes.* 2009;58(2):469–477.
4. Tuttle KR, et al. JAK1/JAK2 inhibition by baricitinib in diabetic kidney disease: results from a Phase 2 randomized controlled clinical trial. *Nephrol Dial Transplant.* 2018;33(11):1950–1959.
5. Schröppel B, Huber S, Horster M, Schlöndorff D, Kretzler M. Analysis of mouse glomerular podocyte mRNA by single-cell reverse transcription-polymerase chain reaction. *Kidney Int.* 1998;53(1):119–124.
6. Ju W, et al. Defining cell-type specificity at the transcriptional level in human disease. *Genome Res.* 2013;23(11):1862–1873.
7. Shen-Orr SS, et al. Cell type-specific gene expression differences in complex tissues. *Nat Methods.* 2010;7(4):287–289.
8. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol.* 2018;14(8):479–492.
9. Young MD, et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science.* 2018;361(6402):594–599.
10. Menon R, et al. Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney. *Development.* 2018;145(16):dev164038.
11. Czerniecki SM, et al. High-Throughput Screening Enhances Kidney Organoid Differentiation from Human Pluripotent Stem Cells and Enables Automated Multidimensional Phenotyping. *Cell Stem Cell.* 2018;22(6):929–940.e4.
12. Harder JL, et al. Organoid single cell profiling identifies a transcriptional signature of glomerular disease. *JCI Insight.* 2019;4(1):122697.
13. Wu H, Uchimura K, Donnelly EL, Kirita Y, Morris SA, Humphreys BD. Comparative Analysis and Refinement of Human PSC-Derived Kidney Organoid Differentiation with Single-Cell Transcriptomics. *Cell Stem Cell.* 2018;23(6):869–881.e8.
14. Park J, et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science.* 2018;360(6390):758–763.
15. Arazi A, et al. The immune cell landscape in kidneys of patients with lupus nephritis. *Nat Immunol.* 2019;20(7):902–914.
16. Gluck C, Ko YA, Susztak K. Precision Medicine Approaches to Diabetic Kidney Disease: Tissue as an Issue. *Curr Diab Rep.* 2017;17(5):30.
17. Nguyen QH, Pervolarakis N, Nee K, Kessenbrock K. Experimental Considerations for Single-Cell RNA Sequencing Approaches. *Front Cell Dev Biol.* 2018;6:108.
18. Wu H, Kirita Y, Donnelly EL, Humphreys BD. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *J Am Soc Nephrol.* 2019;30(1):23–32.
19. Lake BB, et al. A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. *Nat Commun.* 2019;10(1):2832.
20. Chen L, Clark JZ, Nelson JW, Kaissling B, Ellison DH, Knepper MA. Renal-Tubule Epithelial Cell Nomenclature for Single-Cell RNA-Sequencing Studies. *J Am Soc Nephrol.* 2019;30(8):1358–1364.
21. Bhargava P, Schnellmann RG. Mitochondrial energetics in the kidney. *Nat Rev Nephrol.* 2017;13(10):629–646.
22. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv. Cornell University. <https://arxiv.org/abs/1802.03426>. Accessed February 25, 2020.
23. Goligorsky MS. Pathogenesis of endothelial cell dysfunction in chronic kidney disease: a retrospective and what the future may hold. *Kidney Res Clin Pract.* 2015;34(2):76–82.
24. Szklarczyk D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607–D613.

25. Dumas SJ, et al. Single-Cell RNA Sequencing Reveals Renal Endothelium Heterogeneity and Metabolic Adaptation to Water Deprivation. *J Am Soc Nephrol*. 2020;31(1):118–138.
26. Jefferson JA, Shankland SJ. The pathogenesis of focal segmental glomerulosclerosis. *Adv Chronic Kidney Dis*. 2014;21(5):408–416.
27. Korbet SM. Treatment of primary FSGS in adults. *J Am Soc Nephrol*. 2012;23(11):1769–1776.
28. Dimke H, Maezawa Y, Quaggin SE. Crosstalk in glomerular injury and repair. *Curr Opin Nephrol Hypertens*. 2015;24(3):231–238.
29. Kim JH, Kim BK, Moon KC, Hong HK, Lee HS. Activation of the TGF-beta/Smad signaling pathway in focal segmental glomerulosclerosis. *Kidney Int*. 2003;64(5):1715–1721.
30. Khodarev NN, Roizman B, Weichselbaum RR. Molecular pathways: interferon/stat1 pathway: role in the tumor resistance to genotoxic stress and aggressive growth. *Clin Cancer Res*. 2012;18(11):3015–3021.
31. Bazzi C, et al. Urinary IgG and  $\alpha$ 2-macroglobulin are powerful predictors of outcome and responsiveness to steroids and cyclophosphamide in idiopathic focal segmental glomerulosclerosis with nephrotic syndrome. *Biomed Res Int*. 2013;2013:941831.
32. Tao J, et al. JAK-STAT signaling is activated in the kidney and peripheral blood cells of patients with focal segmental glomerulosclerosis. *Kidney Int*. 2018;94(4):795–808.
33. Satoh J, Tabunoki H. A Comprehensive Profile of ChIP-Seq-Based STAT1 Target Genes Suggests the Complexity of STAT1-Mediated Gene Regulatory Mechanisms. *Gene Regul Syst Bio*. 2013;7:41–56.
34. Joachim GR, Cameron JS, Schwartz M, Becker EL. Selectivity of protein excretion in patients with the nephrotic syndrome. *J Clin Invest*. 1964;43:2332–2346.
35. Hedfors E, Kistner S, Norberg R. Serum levels and urinary excretion of alpha-2-macroglobulin in patients with renal disease. *Acta Med Scand*. 1971;190(5):347–351.
36. Garcia-Ferrer I, Marrero A, Gomis-Rüth FX, Goulas T.  $\alpha$ 2-Macroglobulins: Structure and Function. In: Harris JR, Marles-Wright J eds. *Macromolecular Protein Complexes: Structure and Function*. Cham, Switzerland: Springer International Publishing; 2017:149–183.
37. Barrett AJ, Starkey PM. The interaction of alpha 2-macroglobulin with proteinases. Characteristics and specificity of the reaction, and a hypothesis concerning its molecular mechanism. *Biochem J*. 1973;133(4):709–724.
38. Greka A, Mundel P. Cell biology and pathology of podocytes. *Annu Rev Physiol*. 2012;74:299–323.
39. Hoogendoorn H, Toh CH, Nesheim ME, Giles AR. Alpha 2-macroglobulin binds and inhibits activated protein C. *Blood*. 1991;78(9):2283–2290.
40. Isermann B, et al. Activated protein C protects against diabetic nephropathy by inhibiting endothelial and podocyte apoptosis. *Nat Med*. 2007;13(11):1349–1358.
41. Madhusudhan T, Kerlin BA, Isermann B. The emerging role of coagulation proteases in kidney disease. *Nat Rev Nephrol*. 2016;12(2):94–109.
42. Sharma R, et al. Thrombin-Induced Podocyte Injury Is Protease-Activated Receptor Dependent. *J Am Soc Nephrol*. 2017;28(9):2618–2630.
43. Clark JZ, Chen L, Chou CL, Jung HJ, Lee JW, Knepper MA. Representation and relative abundance of cell-type selective markers in whole-kidney RNA-Seq data. *Kidney Int*. 2019;95(4):787–796.
44. Becht E, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. 2019;37:38–44.
45. Stuart T, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177(7):1888–1902.e21.
46. Korsunsky I, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;16(12):1289–1296.
47. Cohen CD, Frach K, Schlöndorff D, Kretzler M. Quantitative gene expression analysis in renal biopsies: a novel protocol for a high-throughput multicenter application. *Kidney Int*. 2002;61(1):133–140.
48. Lindenmeyer MT, et al. Systematic analysis of a novel human renal glomerulus-enriched gene expression dataset. *PLoS ONE*. 2010;5(7):e11545.
49. Grayson PC, et al. Metabolic pathways and immunometabolism in rare kidney diseases. *Ann Rheum Dis*. 2018;77(8):1226–1233.
50. Gadegbeku CA, et al. Design of the Nephrotic Syndrome Study Network (NEPTUNE) to evaluate primary glomerular nephropathy by a multidisciplinary approach. *Kidney Int*. 2013;83(4):749–756.
51. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
52. Krishnan A, et al. Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat Neurosci*. 2016;19(11):1454–1462.
53. Greene CS, et al. Understanding multicellular function and disease with human tissue-specific networks. *Nat Genet*. 2015;47(6):569–576.
54. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–140.