

Organoid single cell profiling identifies a transcriptional signature of glomerular disease

Jennifer L. Harder,¹ Rajasree Menon,² Edgar A. Otto,¹ Jian Zhou,³ Sean Eddy,¹ Noel L. Wys,¹ Christopher O'Connor,¹ Jinghui Luo,⁴ Viji Nair,¹ Cristina Cebrian,⁵ Jason R. Spence,⁵ Markus Bitzer,¹ Olga G. Troyanskaya,^{3,6,7} Jeffrey B. Hodgin,⁴ Roger C. Wiggins,¹ Benjamin S. Freedman,^{8,9,10,11} Matthias Kretzler,^{1,2} European Renal cDNA Bank (ERCB),¹² Nephrotic Syndrome Study Network (NEPTUNE)¹³

¹Department of Internal Medicine, Division of Nephrology, and ²Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan, USA. ³Flatiron Institute, Simons Foundation, New York, New York, USA. ⁴Department of Pathology and ⁵Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan, USA. ⁶Lewis-Sigler Institute for Integrative Genomics and ⁷Department of Computer Science, Princeton University, Princeton, New Jersey, USA. ⁸Department of Medicine, Division of Nephrology, ⁹Kidney Research Institute, ¹⁰Institute for Stem Cell and Regenerative Medicine, and ¹¹Department of Pathology, University of Washington School of Medicine, Seattle, Washington, USA. ¹²ERCB is detailed in the Supplemental Acknowledgments.

¹³The Nephrotic Syndrome Study Network is detailed in the Supplemental Acknowledgments.

Podocyte injury is central to many forms of kidney disease, but transcriptional signatures reflecting podocyte injury and compensation mechanisms are challenging to analyze in vivo. Human kidney organoids derived from pluripotent stem cells (PSCs), a potentially new model for disease and regeneration, present an opportunity to explore the transcriptional plasticity of podocytes. Here, transcriptional profiling of more than 12,000 single cells from human PSC-derived kidney organoid cultures was used to identify robust and reproducible cell lineage gene expression signatures shared with developing human kidneys based on trajectory analysis. Surprisingly, the gene expression signature characteristic of developing glomerular epithelial cells was also observed in glomerular tissue from a kidney disease cohort. This signature correlated with proteinuria and inverse eGFR, and it was confirmed in an independent podocytopathy cohort. Three genes in particular were further characterized as potentially novel components of the glomerular disease signature. We conclude that cells in human PSC-derived kidney organoids reliably recapitulate the developmental transcriptional program of podocytes and other cell lineages in the human kidney and that transcriptional profiles seen in developing podocytes are reactivated in glomerular disease. Our findings demonstrate an approach to identifying potentially novel molecular programs involved in the pathogenesis of glomerulopathies.

Conflict of interest: The authors have declared that no conflict of interest exists.

Authorship note: JLH, RM, EAO, and JZ contributed equally to this work.

License: Copyright 2019, American Society for Clinical Investigation.

Submitted: August 27, 2018

Accepted: November 27, 2018

Published: January 10, 2019

Reference information:

JCI Insight. 2019;4(1):e122697.

<https://doi.org/10.1172/jci.insight.122697>.

insight.122697.

Introduction

Elucidating the molecular events that underlay the evolution of human glomerular disease has been challenging. While animal and 2-dimensional (2-D) cell culture models have added significantly to the understanding of podocytopathies (1), they are limited in their ability to accurately reflect events in human kidney disease (2, 3). A particularly pertinent gene expression network within a specific cell type can be masked in bulk tissue transcriptional profiling of diseased human kidney tissue by diluting the signal of interest below the level of significance, especially when cells are of low abundance, such as podocytes (PCs). Methods to improve detection of glomerulus-specific gene expression alterations have provided some insight, including transcriptional profiling of microdissected glomeruli from individuals with kidney disease (4–7) and enhancing cell-specific signals from kidney tissue transcriptional analysis by in silico nanodissection (deconvolution of expression “signatures” based on known cell-specific markers such as *NPHS2*/podocin) (8). However, decrease in PC-specific gene expression either by loss of PCs or by transcriptional modulation introduces further complexity to the gene expression analysis.

Recent technological advances including pluripotent stem cell–derived (PSC-derived) kidney organoid cultures (9–12) and single cell RNA sequencing (scRNA-seq) (13, 14) present an opportunity to more precisely explore the PC developmental transcriptional program. Human kidney organoids can be derived from PSCs over the course of approximately 15 days of differentiation and represent a simplified cohort of kidney cell types. Cells self-organize in vitro in patterns that resemble nephron subunits and include PCs that can be clearly discerned from neighboring tubular cells. By enabling transcriptional profiling of individual cells within this heterogeneous cell population, scRNA-seq of human PSC–derived (hPSC-derived) kidney organoid cultures may improve the ability to identify and investigate molecular events more specific to the human kidney. Use of such organoids could be expanded through more careful definition of both cell types generated and their relevancy to human kidney (10, 15–17), assessment of reproducibility, and limitation of nonkidney cell types in kidney organoid cultures.

In the present study, we sought to address these concerns by defining the PC transcriptional program generated in hPSC-derived kidney organoids in the context of the developing human kidney. By performing a potentially novel type of trajectory analysis using combined single cell transcriptomic datasets, we were able to define the repertoire of cells present in the organoid cultures and to benchmark them against cell types in the developing human kidney. This led to the discovery of a gene signature of an early PC lineage in organoids, which was used to interrogate gene expression profiles in tissues from diseased human glomeruli. Our results delineate where hPSC-derived kidney organoid cultures can help to define human kidney disease, provide more specific context to the relative transcriptional “age” of cells produced in kidney organoid cultures relative to the developing human kidney, and establish the robustness and reproducibility of this system. In addition to identifying molecular programs involved in the pathogenesis of glomerulopathies, these findings provide a foundation for researchers to explore therapeutics aimed at reversing these alterations.

Results

Single cell transcriptional profiling reveals 2 putative PC cell clusters in kidney organoids. To gain insight into the transcriptional program of PCs, we transcriptionally profiled PCs created in kidney organoids. Organoid cultures were generated using UM77-2 human embryonic stem cells (hESCs) and the matrix sandwich protocol (9, 16). Kidney organoids appeared as discrete clusters of translucent tubular structures, which could be identified by light microscopy (16). We confirmed that organoids contained cells expressing proteins typical of PCs, as well as proximal and distal tubular epithelial cells in nephron-like segments (Figure 1, A–C).

To comprehensively characterize cell types present in the cultures and to establish reproducibility of single cell transcriptional profiling of organoids, all cells in the organoid culture wells (“whole well”) were collected (Figure 1C). Analysis of 10,113 single cell transcriptomes from 7 datasets using an unsupervised cell-clustering algorithm generated 11 separate cell clusters (Figure 2A). Individual clusters contained between 159–1,751 cells and were defined by between 73–272 differentially expressed genes (Supplemental Figure 1A; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.122697DS1>). Cells included in this analysis were of consistent high quality with low mitochondrial content (mean 5.1%) and consistent gene number and transcripts, while cell clusters showed good separation by differential gene expression (Supplemental Figure 1, B and C). The full gene list used to define each cluster is shown in Supplemental Table 1, while expression profiles of characteristic genes used to identify cell types included in clusters are shown in Figure 2B. Several of these clusters contained cells that did not express obvious kidney-associated genes, and gene expression was more consistent with nonkidney stromal, neural, and proliferating cells (18). However, 6 of the 11 clusters (68.6% of transcriptomes analyzed) could be provisionally assigned as kidney-associated lineage by inspection of each cluster’s gene list (outlined by a dotted line in Figure 2A) and were selected for viewing in subsequent images (further referred to as kidney clusters) (Figure 2C).

Within the kidney clusters, expression of characteristic markers of PCs (including *NPHS1*, *PTPRO*, and *NPHS2*) was seen in 2 separate clusters, C1 and C7 (Figure 2, B and G, left). Compared with C7, relatively lower *PTPRO* and *NPHS1*, as well as higher *CDH6*, expression was observed in C1, suggesting that this cell cluster represents cells at an earlier stage of glomerular epithelial cell development. Thus, we designated the C1 cluster (lower *PTPRO/NPHS1*) as early glomerular epithelial (EGE) cells and C7 cluster as maturing podocyte (MPC) cells. The other clusters were consistent with early tubule (*POU3F3*), proximal tubule (*SLC3A1*), loop of Henle/distal tubule (*CLDN10*), and stromal (*COL1A1*) cell lineages. Each of the clusters

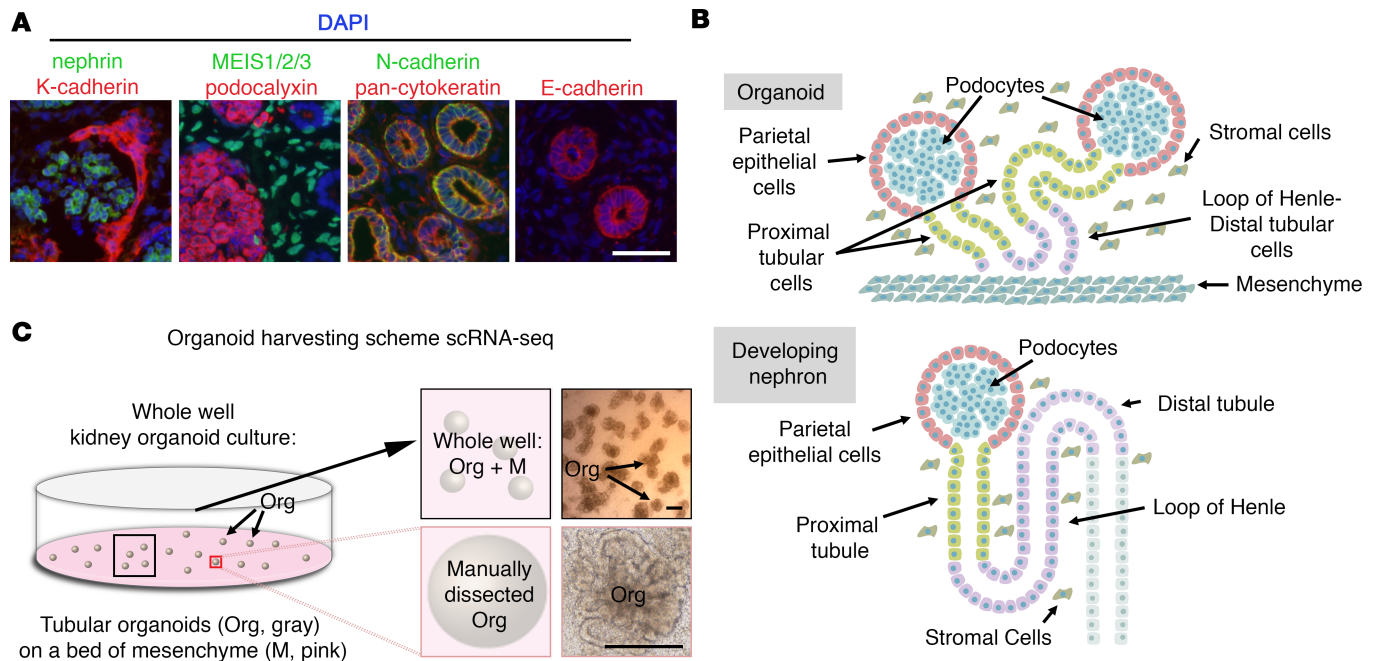


Figure 1. Kidney organoids profiled by single cell transcriptomics contain nascent nephrons. (A) Immunofluorescence imaging of sectioned kidney organoids containing nascent podocytes (nephtrin, podocalyxin), stromal (MEIS1/2), and epithelialized tubular (cytokeratin, K-/N-/E-cadherin) cells. Scale bar: 100 μ m. (B) Cartoon showing organoid cultures in situ and corresponding developing nephron structures. (C) Cartoon depicting organoid harvesting scheme for individual cell analysis by Drop-seq, with corresponding brightfield microscopic images of organoid cultures. Scale bars: 0.5 mm. Either whole well (organoids plus mesenchyme bed) or manually dissected organoids (20–50) were collected and dissociated per experiment.

present was consistent with what we previously observed in developing human kidney (15) based on examination of the top 20 cluster-defining genes.

Overlay of cells from individual transcriptional datasets revealed consistent contribution to each cell cluster by cells from each dataset (Figure 2D). Each dataset exhibited well-dispersed cells within each cluster, confirming the reproducibility of our organoid culture system, transcriptome sequencing, and bioinformatics analysis pipeline. Overall, an average of $70.3\% \pm 9.9\%$ (\pm SD) of cells from each dataset contributed to the kidney cell clusters. (Supplemental Figure 1D). Each cell type was consistently represented, though wider variability was noted in the relative contribution to each cluster among datasets with $SD \pm 1.1\%$ to $\pm 7.8\%$ in C10 and C2, respectively (Supplemental Figure 1E). Taken together, these data indicate that organoid cultures are reproducible and include mostly kidney but also other nonkidney cell types, supporting the use of the single cell transcriptomic approach for analysis of kidney organoids to both identify and quantify kidney cell type gene expression.

To further confirm the kidney cell type content of kidney organoids, a second iteration of scRNA-seq was performed using isolated tubular organoids, which were microdissected from the surrounding mesenchyme (Figure 1C) (9). Unsupervised clustering of transcriptomes representing 1,926 cells revealed 8 identifiable cell clusters, similar to the kidney clusters shown in Figure 2C, but did not include the off-target cell types detected in the whole well cultures, with the exception of a proliferating cell cluster (Figure 2E). A full list of genes used to define each cluster is in Supplemental Table 2, while the top 20 cluster-defining genes are shown in Figure 2G. (Note that a gene can be included in multiple cluster-defining gene lists since significance of gene expression by cells in 1 cluster is compared with expression in cells in all other clusters combined.) Quality assessment of these cells was on par with the prior combined datasets (Supplemental Figure 1, F and G).

Expression of the same characteristic kidney and proliferation genes by the isolated organoid cell clusters supported the earlier assignment of cell type identity in the whole well analysis. Again, 2 discrete *NPHS2/NPHS1*-expressing cell clusters were present, reinforcing the concept that there are 2 transcriptional states of PC lineage cells in the organoids. The possibility that this represented 2 developmental stages of cells was supported by the presence of cells in both clusters expressing *CDK6*, *LHX1*, *FOXC2*, *CCNLI*, and *SIX1*, indicating that these cells were still mitotically and developmentally active (19, 20).

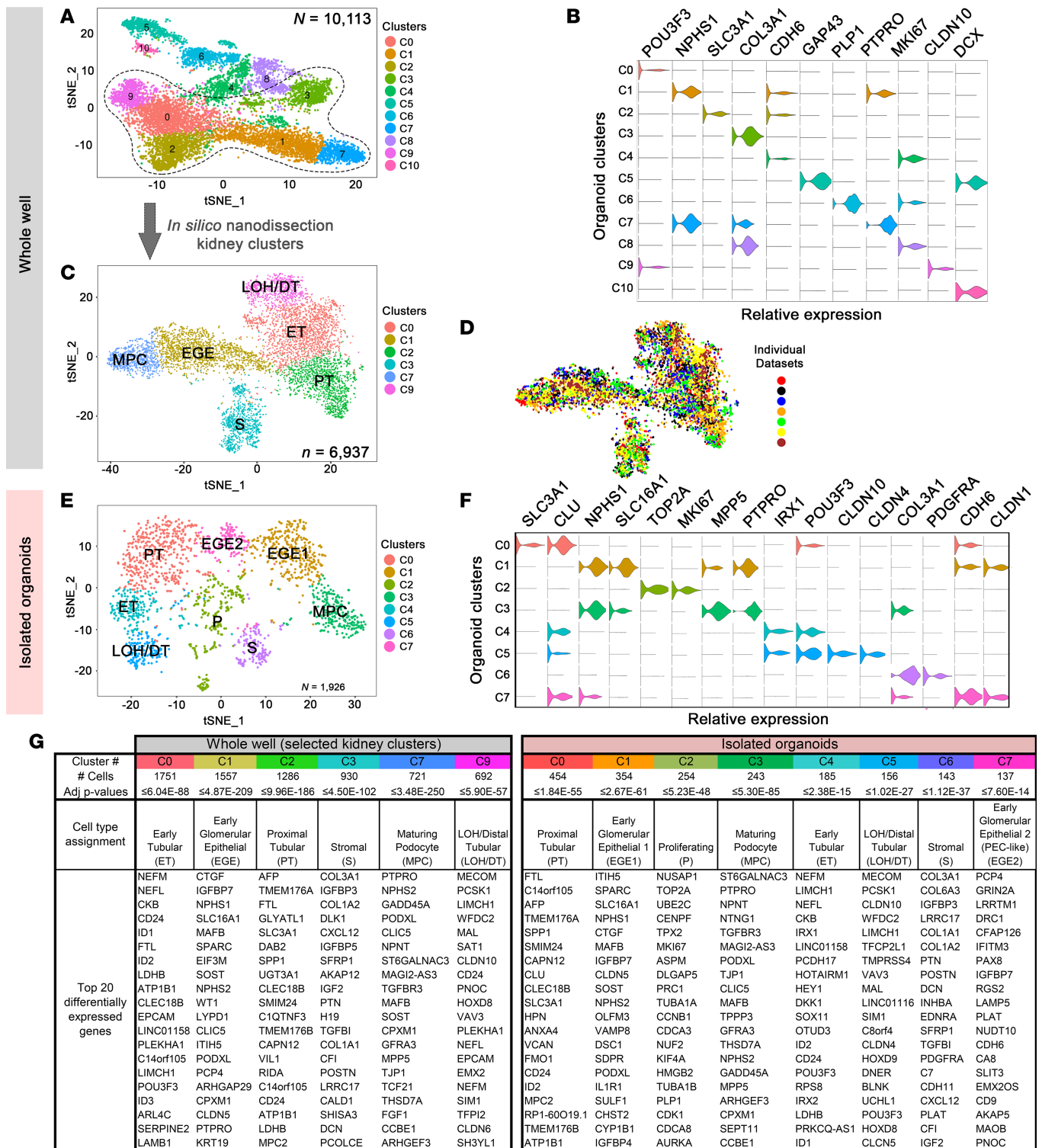


Figure 2. Kidney organoids profiled by single cell transcriptomics reveal spectrum of maturing nephron elements. (A) t-SNE plot including all cell clusters generated from scRNA-seq analysis of whole wells of kidney organoids. n , number of transcriptomes from 7 datasets from 4 individual experiments, 3 technical replicates. Dotted line encircles kidney clusters. (B) Violin plots showing relative expression of selected genes for each cell cluster in A. (C) t-SNE plot showing selected kidney cell populations in whole wells of kidney organoids, identified by clustering similar single cell transcriptomes. (D) Overlay t-SNE plots of 7 individual datasets contributing to C. (E) t-SNE plot showing all cell populations present in isolated organoids. Data representative of 3 separate experiments. (F) Violin plots for cell clusters in F, as in B. (G) Top differentially expressed genes in order of decreasing significance within cells of these clusters compared with other cell clusters from C on left, and F on right. Calculated using Seurat's default parameters for nonparametric Wilcoxon rank sum test, reported with adjusted P values. LOH, loop of Henle. Gene names not italicized for ease of viewing in B, F, and G. See related Supplemental Figure 1, Supplemental Table 1, and Supplemental Table 2.

While 2 *NPHS1*-expressing clusters were detected (in whole well; C1 and C7 in Figure 2C), 3 *NPHS1*-expressing clusters were observed in the isolated organoids (Figure 2, F and G): 1 low (C7) and 2 high (C1 and C3). This low *NPHS1*-expressing cluster was designated as a second EGE (EGE2) cell cluster and expressed markers associated with parietal epithelial cells (PECs) including *CLDN1*, *CDH6*, and *PAX8* (Figure 2, F and G, right) (15). In order to compare EGE1 and EGE2 directly with the original EGE cluster to see if they contain similar cell types, the EGE1 and EGE2 clusters were combined, and subclustering was performed for the EGE and EGE1–EGE2 merged clusters. This revealed 4 subclusters for each with similar gene expression profiles, indicating that these clusters contain similar cells (Supplemental Figure 1, H–J). Closer attention to the gene profiles suggests a spectrum of subcluster cell types, from more tubular epithelial-like in the top rows of the violin plots (*CLU*, *CDH6*, *CLDN1*) to more PC-like in the bottom rows (*PTPRO*, *TJPI*). Indeed, this is consistent with well-described protein expression patterns observed in adult human kidneys (Supplemental Figure 1K) in which EGE1 (nephrin, podocalyxin) and EGE2 (PAX8, K-cadherin) gene products usually localized to the glomerulus and Bowman's capsule, respectively. However, we occasionally found cells on Bowman's capsule expressing EGE1 genes, and vice versa, in the glomerulus expressing EGE2 genes. These intriguing results in adult human kidneys reinforce the transcriptional proximity of developing PEC and PCs during development.

The many disparate cell types present in the whole well study most likely drove aggregation of these cells into a single EGE cluster, despite using the same cluster resolution in both analyses (Seurat clustering parameter resolution = 0.6). The lack of such off-target cell types in the isolated organoids (mesenchymal cells in Figure 1, B and C) allowed a more refined clustering at the same resolution. Further illustration of this difference is reflected in the variation of cluster-defining genes seen between the clustering results from whole well and isolated organoids. For example, the early tubular (ET) cluster in whole well culture shares 10 of the top 20 genes with the ET cluster in the isolated organoids. Of note, cells with gene expression characteristic of collecting ducts (CDs), mesangial, or endothelial cells were not appreciated in either cell clustering iteration. In summary, these results support the findings of prior histologic and IF characterization of kidney organoids, reveal 2 discrete putative PC clusters, and suggest that the majority of off-target cells reside in the mesenchyme surrounding the tubular organoids.

Kidney organoid cells recapitulate developmental transcriptional programming of human kidney cells. To determine the similarity of gene expression in cells from hPSC-derived organoids and developing human kidney, we compared the gene expression profiles of the 6 kidney clusters from the whole well to those generated from developing human kidney (around 15 weeks of gestation) (21). Whole well cell clusters were chosen over the isolated organoid clusters due to the higher number of cell transcriptomes for comparison. Intriguingly, unsupervised cluster analysis of developing human kidney also revealed 2 distinct PC clusters (21), suggesting that what is observed in organoids is reflective of the developing human kidney. Indeed, averaged gene expression in MPC and EGE organoid clusters was most similar (darker red) to the corresponding clusters (mature and early PC clusters, respectively) in the developing kidney (Figure 3A). Additionally, compared with the MPC organoid cell cluster, the EGE cluster shared greater similarity with the proliferating cell cluster of the developing kidney, reinforcing the assertion that these cells were at an earlier developmental state.

Similarly, averaged gene expression of the organoid stromal cell cluster resembled stromally derived cell clusters in the human kidney (cortical and medullary stromal, as well as endothelial) while averaged gene expression of tubular cell clusters of the organoids was most similar to the corresponding tubular cell clusters of the developing kidney; the ET organoid cell cluster shared gene expression with all tubular segments and the proliferating clusters of the developing kidney, suggesting a less differentiated cell cluster. Though gene expression correlation for tubular clusters had a lower Z score than the previously mentioned clusters, this likely reflects variable expression of transporters and other segment-specific genes in tubular segments at different maturation stages (Supplemental Table 1). Overall, the correlation analysis confirms our assignment of cell type identities to organoid kidney clusters and also demonstrates consistency between different stem cell lines (15). Moreover, this analysis further reinforces the notion that the presence of 2 discrete PC cell clusters in both organoids and developing human kidney reflects a fundamental developmental process in PC development.

To better define the cell types represented in organoid cultures, we sought to more directly compare cells from organoids to those from developing human kidneys. Using a second and entirely separate analysis of the single cell transcriptome data, all aforementioned organoid and developing human

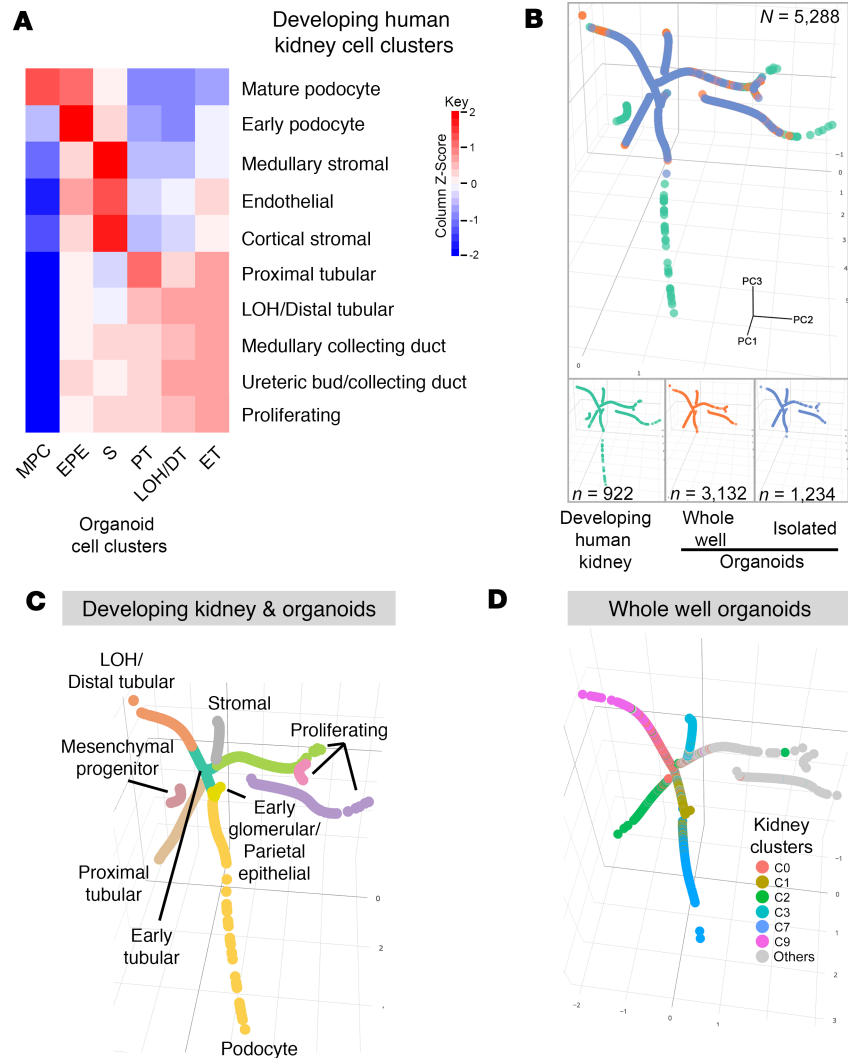


Figure 3. Cells of kidney organoids share the same cell type-specific transcriptional trajectories as developing human kidney. (A) Correlation matrix comparing gene expression profiles of cells in clusters from organoids (whole well) to developing human kidney of estimated gestation days 105–115. The intensity of the overall image was adjusted to maximize the discernable visible range of Z score intensities, color indicates correlation (red, positive; blue, negative). (B) Overlaid trajectory analyses of combined single cell transcriptomes from developing human kidneys, whole well organoid cultures, and isolated organoids. N , cumulative number of transcriptomes used in trajectory analysis, and n , contributing number of transcriptomes from each dataset. See related Supplemental Video 1. (C) Trajectory segment cell type assignment of developing kidney and organoids based on gene expression patterns as in Supplemental Figure 2B. (D) Superimposed t-SNE cluster cell type assignments of selected kidney clusters from whole well organoids only (see Figure 2C) overlaid on cell lineage trajectory. Cells from all other clusters are represented in gray. See related Supplemental Figure 2.

kidney single cell transcriptome datasets were combined, and a developmental cell lineage trajectory was generated using a method that provides statistical inference. All single cell transcriptomes with at least 500 genes/cell were included at the outset of this analysis, but the trajectory algorithm dispensed with cell types not well represented in both organoids and developing human kidneys. This approach allowed the examination of cell lineages of organoids within the context of cells from developing kidneys and within the context of contribution of cells found in isolated organoids relative to the whole well organoid cultures (Figure 3B and Supplemental Video 1). We opted for this initial stringent gene expression threshold to produce a robust trajectory, but when the threshold was decreased to 200 genes/cell and the transcriptomes were overlaid onto the trajectory in Figure 3B, there was very good alignment (Supplemental Figure 2A). Additional cell lineages were also seen and, for the most part, were contributed by the developing human kidney, as expected.

Contribution to each cell lineage trajectory was seen by both developing kidney and organoids. Top differentially expressed genes in each trajectory segment served as the basis for cell lineage assignment (Supplemental Figure 2B). As shown in Figure 3C, multiple cell lineages radiated out from a central hub of precursor cells, and include podocyte, proximal and distal tubular, stromal, and noncontiguous proliferating cell lineages in a parallel fashion between human developing kidney and organoids. The PC lineage further diverged into 2 separate lineages, MPCs and PECs, the latter of which were distinguishable by their different expression profiles of 3 characteristic PEC genes (*CLDN1*, *ANXA3*, and *PAX8*) (Supplemental Figure 2C). Cells from developing kidney cells (green in Figure 3B) contributed to the distal ends of each of the radiating cell lineage trajectories, indicating that developing kidneys contain cells that are relatively more mature than those found in organoids of this vintage.

To determine the interreliability of the 2 independent analytic methods of the transcriptomic data, cluster-assigned cells were overlaid onto the combined trajectory. We specifically used the kidney clusters from the whole well organoid clusters for this analysis to further assess the veracity of our earlier cell type assignments to each cluster and to further explore the basis of the separation of the *NPHS1/NPHS2*-expressing cell clusters. Cluster-assigned cells of the kidney organoid clusters were overlaid onto the trajectory analysis from Figure 3C and shown in Figure 3D. Importantly, cell cluster assignments correlated reliably with cell lineage trajectory. Cells from the EGE cluster (C1) populated the early segment of the combined PEC-PC trajectory, while those in the MPC cluster localized solely to the post-PEC divergence segment (C7). This observation indicated that cells in the EGE cluster included precursor cells common to both PC and PEC lineages, which suggested a basis for the segregation of EGE and MPC cell clusters. Moreover, the early overlap and proximity of the PEC and PC trajectories reinforced the concept that these cell types share significant transcriptional programming during development. As demonstrated in Supplemental Figure 1K, the gene expression patterns revealed in these trajectories persist into adulthood. Further, expression of *WT1* and *PTPRO* expression in PEC and PC lineages (Supplemental Figure 2D) was reflected on a protein expression level in both PECs ($WT1^+/PTPRO^-$ cells lining Bowman's capsule) and PCs (intraglomerular $WT1^+/PTPRO^+$ cells) in adult human kidney (Supplemental Figure 2E). These results add to those in Supplemental Figure 1K and demonstrate that a subset of genes may be expressed across cell types, similar to the description in partial epithelial-to-mesenchymal transition seen in renal fibrosis (22).

The segmentation of early and later developmental stages seen in PCs was repeated in tubular cell lineage trajectories (Figure 3D). Cells from the ET cluster (C0) localized more centrally, while those from the proximal tubular (C2) and loop of Henle (LOH)/distal tubular (DT) (C9) clusters localized more peripherally. To determine which organoid cells the algorithm included in the trajectory analysis, cells were mapped back onto their corresponding t-Distributed Stochastic Neighbor Embedding (t-SNE) plots (Supplemental Figure 2F). This revealed that cells from each cell cluster contributed to the trajectory, with the off-target clusters contributing to the proliferating lineages. Taken together, these data indicate that cells in kidney organoids reliably recapitulate the developmental transcriptional programming observed in homologous cell types of the developing human kidney.

Organoid PC cell clusters demonstrate distinct transcriptional states. We next sought to further characterize the transcriptional program in the 2 PC clusters to understand the nature of their segregation. The EGE and MPC clusters together represented 22.5% of all cells in organoid cultures (Figure 2C and Supplemental Figure 1E), and both were characterized by expression of typical PC genes, including *PTPRO*, *NPHS1*, *NPHS2*, *PODXL*, and *WT1* (Figure 2G and Figure 4A). These 2 cell clusters differed, however, by the relative expression of epithelial polarity genes *MPP5*, *TJPI1*, *PARD3B*, and *IQGAP2*, with MPC exhibiting higher expression than EGE — indicating that MPC cells were more polarized and, thus, at a later stage of maturation. This was confirmed on a protein expression level with the tight junction gene ZO-1 (*TJPI1*), which was found only in a subset of nephrin⁺ cells in organoids (Figure 4B). When these genes were examined in the context of PC lineage trajectory (Figure 3C), expression of these genes was noted to increase in sync with the developmental trajectory, with PC marker genes preceding polarity gene expression (Figure 4C, compare top and bottom rows).

To determine if transcription factors (TFs) were contributing to the transcriptional segregation of PC clusters, the expression pattern of all known 1,639 human TFs (23) was assessed in cells along the PC trajectory (Figure 4D). A subset (7 of 20 total TFs expressed in the PC lineage) of these was included in the differentially expressed genes in EGE and MPC clusters. While regulation of gene expression in specific tissues is much more complex than a reflection of expression levels of TFs (24), coexpression of TFs

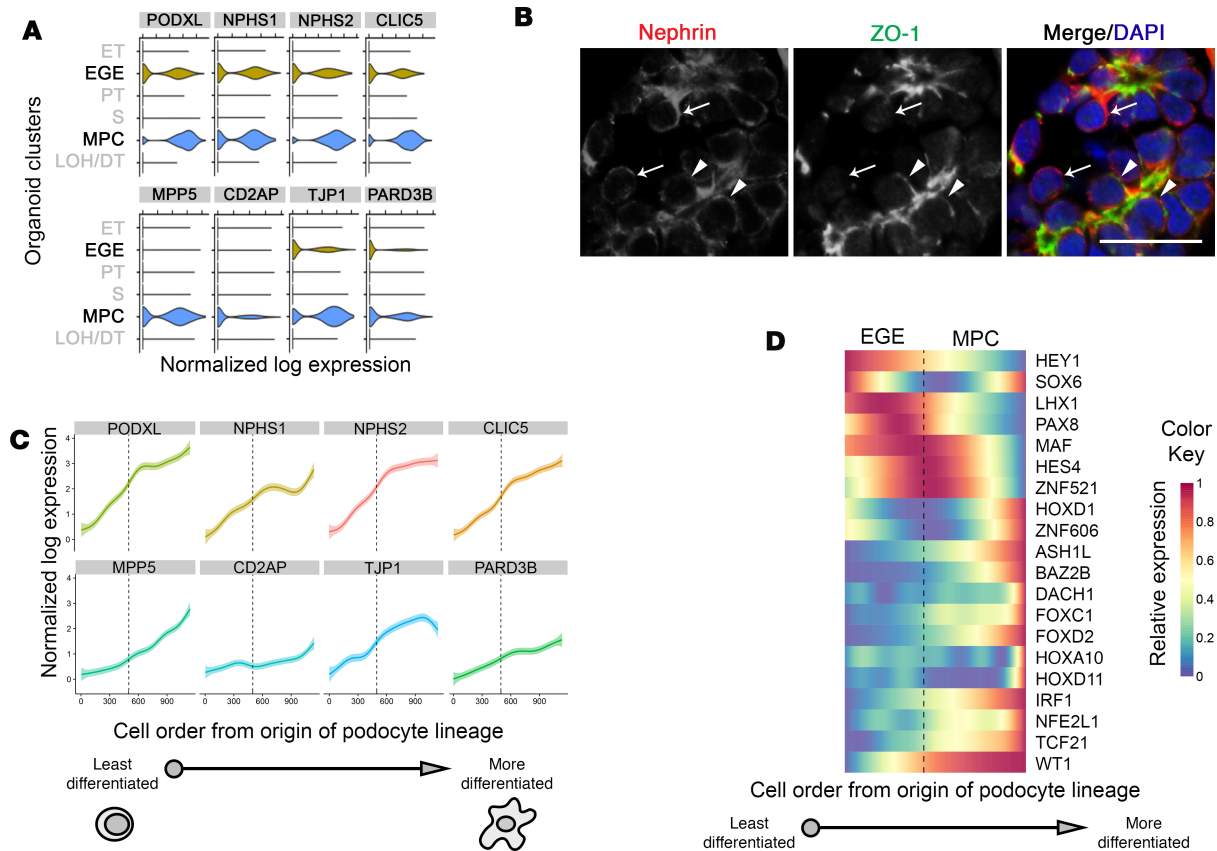


Figure 4. Cells farther along podocyte lineage trajectory are more mature. (A) Violin plots depicting expression of quintessential (top row) podocyte and (bottom row) epithelial cell polarity genes in organoid clusters from Figure 2C. Ticks on x axis indicate whole integers starting from 0 on the left of each plot. (B) Immunofluorescence confocal images showing protein expression in nascent podocytes in day-20 organoids. Arrows highlight nephryn⁺/ZO-1⁻ cells, while arrowheads highlight nephryn⁻/ZO-1⁺ cells. ZO-1 (TJP1). Scale bar: 20 μm. Representative images of 3 independent experiments. (C) Average expression of indicated genes by cells along combined organoid-developing kidney podocyte lineage trajectory from Figure 3C using a curve-fitting method (GAM) with 95% CI represented by shading. (D) Heatmap of gene expression of transcription factors along the podocyte lineage trajectory as in C. Vertical dotted line indicates position at which podocyte and PEC lineages diverge (C and D). Gene names not italicized for ease of viewing (A, C, and D). See Supplemental Figure 2.

and their targets is frequently observed (25, 26) and at least consistent with potential regulatory patterns present. Consistent with established developmental patterns, *LHX1* and *PAX8* expression peaked earlier in PC development, while expression of several TFs described as involved in PC maturation was seen later, including *FOXC1* and *TCF21* (27, 28). An increase in expression of *WT1*, *HES4*, and *MAF* around the divergence of the PC and PEC lineages suggested a possible basis for a regulatory transcriptional switch associated with PC maturation. Together, the trajectory analysis and gene expression characterization indicate that the EGE and MPC cell clusters represent 2 transcriptionally discrete states within the continuum of PC development.

Genes highly expressed in immature glomerular epithelial cells of organoids are dysregulated in human kidney disease. We hypothesized that the gene expression pattern seen in the EGE cluster is reactivated in injured PCs in glomerular disease. To test this hypothesis, genes relatively unique to or shared by both PC lineage clusters (C1 and C7, Supplemental Table 1) were identified. This resulted in 3 sets of genes: EGE (69 genes), shared (104 genes), and MPC (168 genes) (Figure 5A). These gene sets were used to generate aggregate gene expression scores in isolated glomerular tissue from a cohort of individuals with various etiologies of chronic kidney disease (ERCB) (29, 30). Query of the subset of the identified genes available on the microarray platforms revealed that the EGE aggregate score was significantly increased in the kidneys from 170 individuals with chronic kidney disease relative to kidneys from living donors, specifically in individuals with lupus nephritis (LN), diabetic kidney disease (DKD), anti-neutrophil cytoplasmic antibody-associated (ANCA-associated) vasculitis (AAV), and focal segmental glomerulosclerosis (FSGS) (Figure 5B and Supplemental Figure 3A). Conversely, the MPC and shared

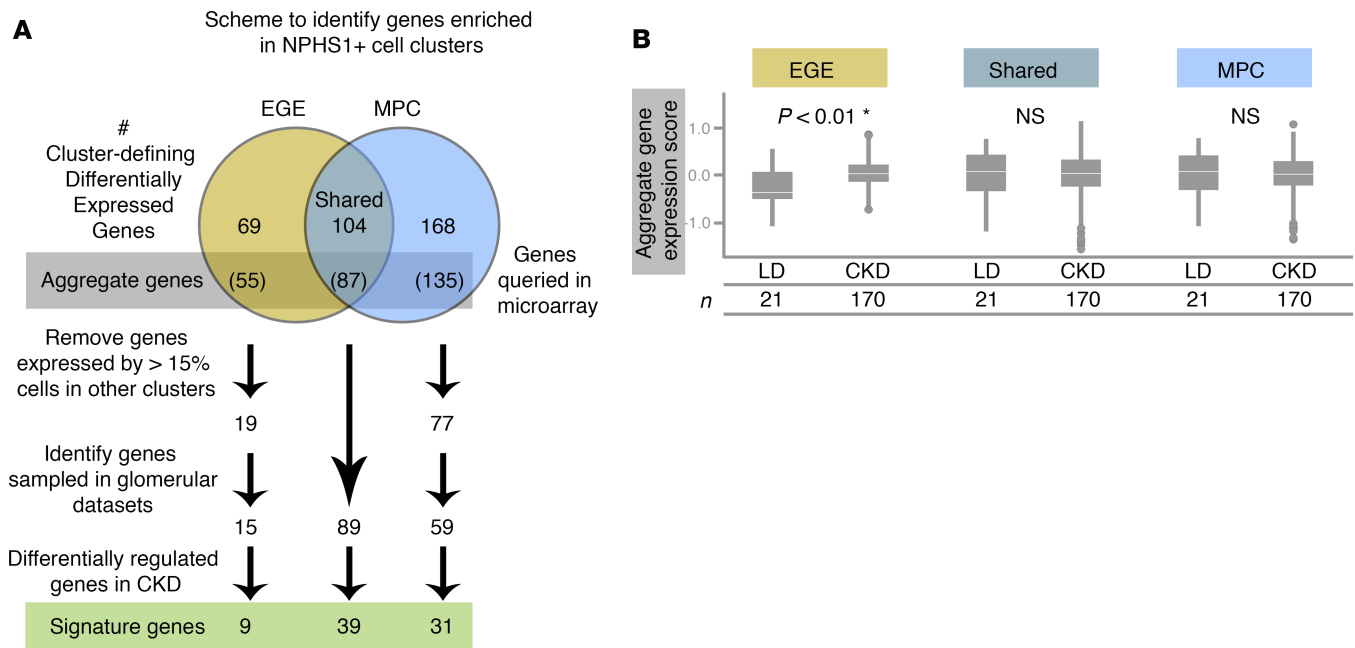


Figure 5. Genes differentially expressed in podocytes in organoids correlate with human kidney disease. (A) Illustration depicting method of identification of genes relatively specific to early glomerular epithelial (EGE) and maturing podocyte (MPC) clusters from Figure 2C. (B) Box plots of aggregate gene expression scores for 3 sets of genes from A (gray box) in microdissected glomerular tissue from individuals with chronic kidney disease (CKD) compared with living donors (LD) from the ERCB cohort. Statistical significance calculated using 2-tailed Student's *t* test. * $P < 0.05$. Whiskers represent the upper and lower adjacent values. See Table 1, Table 2, and Supplemental Figure 3.

aggregate scores were unchanged. These results suggest that the EGE aggregate score was capturing a pathologically relevant transcriptional state in human glomerular disease.

To identify potential mechanistic pathways associated with these gene sets, pathway analysis was performed. Canonical pathways generated using all of the 173 EGE genes modestly favored inhibition of angiogenesis by TSP1, hypoxia signaling in the cardiovascular system, and ephrin receptor signaling, while those generated using all of the 272 MPC genes more evidently favored unfolded protein response, axonal guidance signaling, and tight junction signaling (Supplemental Figure 3B). Only 16 of the unique 69 genes in the EGE subset were found in any known pathway, indicating that these genes are not well represented in known biological networks (Supplemental Table 3). This improved only modestly by expanding to include the full 173 EGE gene set (Supplemental Table 3C). Thus, pathway analysis of EGE genes provided limited information about potential pathogenic mechanisms of disease, pointing to the need to examine the contribution of individual genes in the EGE set.

A gene prioritization approach was used to identify candidate genes driving the EGE signature expression in CKD. Genes were defined as enriched in the EGE cluster when expression in all other clusters combined was below 15%, indicating that expression was relatively unique to the EGE cluster cells. The same threshold was used to identify MPC-enriched and shared-enriched genes (Figure 5A). Application of this algorithm identified 19 genes whose expression was enriched in the EGE cluster, 15 of which were present and analyzed on the microarray platforms used to interrogate the transcriptomes of human kidney glomerular tissues. Of the 15, expression of 8 was significantly increased, while 1 was decreased in CKD (Table 1). Meanwhile, 31 of 59 MPC-enriched genes and 33 of 89 shared-enriched genes were differentially regulated in CKD (Supplemental Table 4). Notably, expression of *MPP5*, *IQGAP2*, *NPHS1*, *PTPRO*, *TJPI1*, and *WT1* was decreased in CKD. Examination of the protein products of the EGE gene signature (Table 1) revealed a paucity of detailed literature especially pertaining to the functions of these proteins within the kidney (Supplemental Table 5); however, a unifying theme for functions of these gene products is development and cell survival.

Thus, by employing a subtractive approach, a small set of EGE genes was identified, the majority of which associated with significantly increased gene expression in diseased glomeruli. Together with the finding that expression of genes characteristic of mature PCs was decreased, these data suggest that PC loss or dedifferentiation was associated with expression of EGE signature genes.

Table 1. Expression of EGE signature genes in CKD

EGE signature gene expression in CKD	Gene	Log ₂ -fold change	q value
	<i>LYPD1</i>	1.51	<0.001
	<i>PRSS23</i>	0.97	<0.001
	<i>IFITM1</i>	0.84	<0.001
	<i>CDH6</i>	0.67	0.001
	<i>GUCY1B3</i>	0.41	<0.001
	<i>VGLL4</i>	0.27	<0.001
	<i>OAZ2</i>	0.25	<0.001
	<i>CYP1B1</i>	0.23	0.031
	<i>FYN</i>	-0.25	0.006

List of EGE signature genes from Figure 5A that are significantly differentially expressed in CKD compared with LD. Differential expression statistics calculated using SAM; *q* values shown. See Figure 5, Supplemental Figure 3, and Table 2.

Increased expression of LYPD1, PRSS23, and CDH6 correlates with human glomerular disease. We focused on 3 significantly differentially regulated genes of the EGE list associated with CKD. The first 2 (*LYPD1* and *PRSS23*) showed the most significant differential expression in disease, while the third (*CDH6*) has a well-described role in kidney development and was therefore of particular interest in a disease context (31). Expression of each was variably increased in subgroups of individuals with different etiologies of CKD (Table 2). As expected, cells expressing these genes in organoids were predominantly located in the EGE cluster (Figure 6A and Supplemental Figure 1K). The pattern of expression of these genes was also examined in the context of the PC-PEC lineage trajectory from Figure 3C and revealed that expression of each of these genes was decreased in MPCs but was increased in common precursor and PEC lineage cells (Figure 6B). These results indicate that the combination of increased gene expression of *LYPD1*, *PRSS23*, and *CDH6* is consistent with a less differentiated PC or PEC.

To explore why these 3 genes were not identified as enriched in early PC lineage cells in our earlier study of single cell transcriptional profiling of the developing human kidney (32), we specifically interrogated gene expression in the context of the developing kidney cell lineage trajectory analysis generated in that study. Examination of the 3 gene expression patterns revealed that each gene was expressed by cells early in PC lineage, as well as other cell type lineages: *LYPD1* expression was found in nephron progenitor (NPG) cells; *PRSS23* expression was found in stromal, CD, and LOH lineage cells; and *CDH6* was found in PT and NPG lineage cells (Figure 6C). Thus, these 3 genes were indeed expressed in developing PCs in human kidney, but expression in other cell types masked their recognition. Therefore, the PC-enriched organoid cultures revealed insights about PC development that were not apparent by solely examining the list of differentially expressed genes in the cell clusters of the developing human kidney (32).

Expression of *LYPD1*, *PRSS23*, and *CDH6* was investigated in diseased kidney tissue. Reiterating Table 1, expression of each was increased in various specific subgroups of individuals with CKD from the ERCB

Table 2. Expression of 3 EGE signature genes by types of kidney diseases in ERCB relative to living donors (LD)

By disease vs. LD	Log ₂ -fold change (<i>q</i> < 0.05)		
	<i>LYPD1</i>	<i>PRSS23</i>	<i>CDH6</i>
ANCA-associated vasculitis	2.2	1.3	1.1
Diabetic nephropathy	1.8	1.5	1.1
Lupus nephritis	1.8	0.8	0.7
Focal segmental glomerulosclerosis	1.2	0.8	0.5
Hypertensive nephropathy	1.3	0.6	-
IgA nephropathy	1.7	1.0	0.8
Minimal change disease	0.8	0.8	-
Membranous nephropathy	0.5	1.1	-

Statistics calculated as in Table 1; differential expression statistics calculated using SAM; *q* values shown. See related Figure 5 and Supplemental Figure 3.

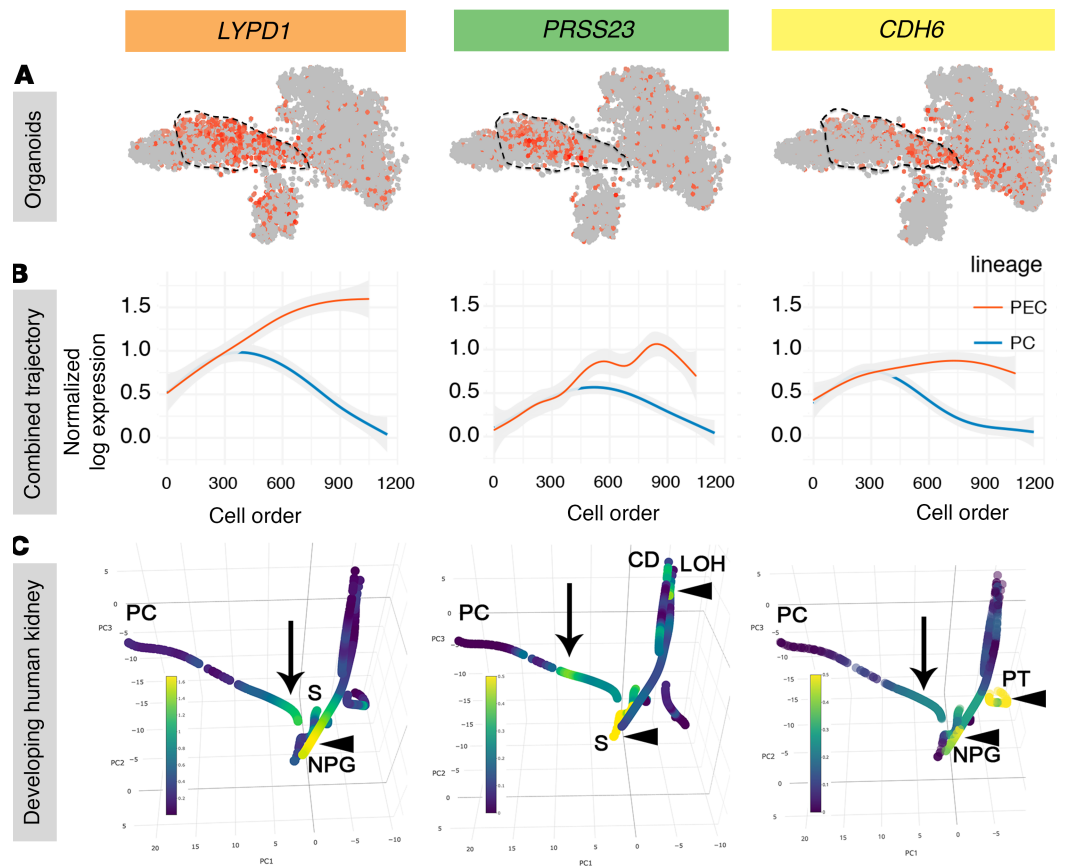


Figure 6. Expression of *LYPD1*, *PRSS23*, and *CDH6* declines during podocyte maturation in organoids and human kidneys. Columns demonstrating *LYPD1* (left), *PRSS23* (middle), and *CDH6* (right) expression in **A–C**. **(A)** Feature plots of overlaid kidney organoid t-SNE plot from Figure 2C. Dashed lines outline cells in EGE cluster. **(B)** Overlaid podocyte ($n = 645$ transcriptomes) and PEC ($n = 546$ transcriptomes) lineage trajectories showing average expression of *LYPD1*, *PRSS23*, and *CDH6*; color key on far right. Gray shading, 95% CI. **(C)** Developing human kidney trajectories as described in ref. 32 representing 6,414 transcriptomes with overlaid gene expression of *LYPD1*, *PRSS23*, and *CDH6* oriented for optimal visualization of podocyte lineage. Arrows, maximal podocyte lineage expression; arrowheads, nonpodocyte lineage maximal expression. PC, podocyte; S, stromal; NPG, nephron progenitor; CD, collecting duct; LOH, Loop of Henle; PT, proximal tubule. Color scales represent normalized log expression.

cohort (Figure 7A). More specifically, expression of each also correlated positively with markers of CKD progression, including loss of eGFR (Figure 7B) and presence of nephrotic range proteinuria in LN (Figure 7C). In contrast, expression of genes from the MPC (*MPP5* and *IQGAP2*) and shared (*NPHS1*, *WT1*, *PTPRO*, and *TJPI*) sets correlated positively with eGFR (Supplemental Figure 4A). These results are especially striking, as this means that increased expression of the 3 signature genes accompanied loss of kidney function and loss of characteristic PC marker expression.

To validate these findings, expression of these 3 genes was also examined in a separate cohort of individuals with glomerulonephritis, including nephrotic syndrome from NEPTUNE ($n = 90$) and antineutrophil cytoplasmic antibody-associated vasculitis ($n = 15$) (29). *LYPD1*, *PRSS23*, and *CDH6* expression again correlated inversely with eGFR at baseline (acquired within 4 weeks of kidney biopsy) (Supplemental Figure 4B), but only *PRSS23* expression significantly correlated with proteinuria in the nephrotic individuals for whom urine protein to creatinine ratios (UPCRs) were available (Supplemental Figure 4C). However, when the nephrotic group was analyzed by etiology, all 3 genes significantly correlated with proteinuria in the membranous nephropathy subgroup ($n = 44$) (Supplemental Table 6). Additionally, expression of *LYPD1*, *PRSS23*, and *CDH6* was examined in isolated glomerular tissue from a rat model of FSGS (Supplemental Figure 4D) (33). Here, gene expression levels are reported relative to podocin (*Nphs2*) and nephrin (*Nphs1*) expression levels, as performed previously. Again, all 3 genes showed increased expression relative to PC genes in the FSGS rat as compared with

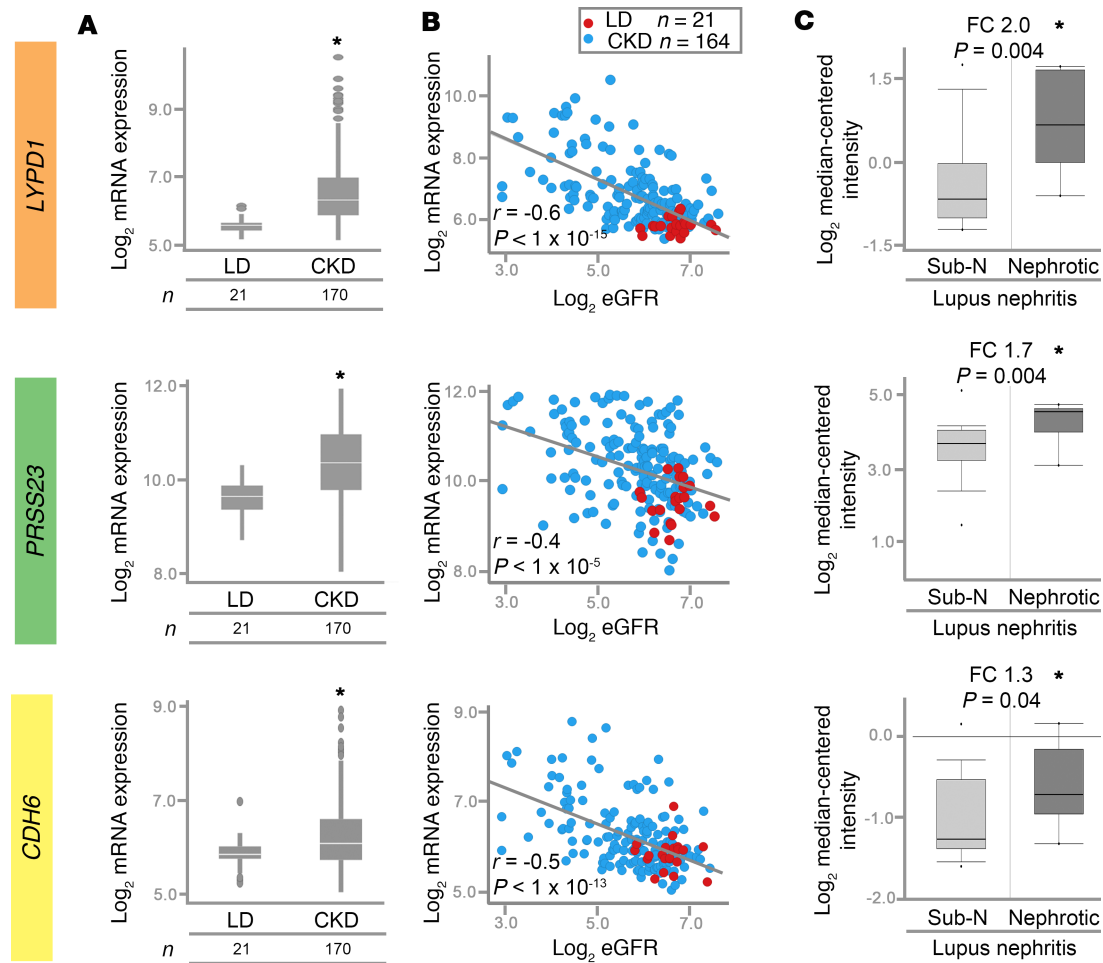


Figure 7. Expression of candidate early glomerular epithelial/PEC genes correlates with chronic kidney disease. (A) Box plots comparing *LYPD1* (top), *PRSS23* (middle), and *CDH6* (bottom) expression in CKD relative to LD, as in Table 1. Whiskers represent the upper and lower adjacent values. (B) Scatter plots showing gene expression from A relative to eGFR in CKD and LD. Statistical significance calculated using Pearson correlation. (C) Box plots as in A comparing individual gene expression levels to proteinuria groups in lupus nephritis samples from ERCB in Nephroseq (www.nephroseq.org). Subnephrotic (Sub-N), $n = 18$; nephrotic, $n = 9$. Statistical significance calculated using 1-tailed t test. Whiskers represent the 10th/90th percentile values as precomputed within Nephroseq. FC, fold change. * $P < 0.05$. See Supplemental Figure 4.

the control rat tissues. Thus, both a second cohort of podocytopathies and a rat model of FSGS also showed that expression of these 3 signature genes was increased in glomerular disease.

To address which cell type within the dissected glomeruli may be contributing to the increase in expression of these genes, *PRSS23* expression was examined in diseased human kidney tissue (Figure 8), as it appeared to have the most robust association with both proteinuria and renal function in our earlier studies. ISH revealed that *PRSS23* transcripts were detected in cells lying along Bowman's capsule in both control and diseased kidney tissue, consistent with PECs. However, in diseased kidney, *PRSS23* transcripts were also detected in cells within the glomerulus present along the urinary surface.

Our results show that a gene expression signature defined in EGE cells in kidney organoid cultures was detected in human glomerular disease. The gene signature included *LYPD1*, *PRSS23*, and *CDH6*, each of whose expression was increased in a cohort of glomerular disease; these findings were recapitulated in a second human glomerular disease cohort, as well as a rat model of FSGS. Taken together, expression of these genes in human glomerular disease is consistent with an activation of a transcriptional program also seen in developing glomerular epithelial cells in organoids and human kidney.

Discussion

PC injury and adaptation are central to progressive decline in kidney function in glomerular disease, but identifying the molecular events underlying this process in human disease has been challenging. We describe

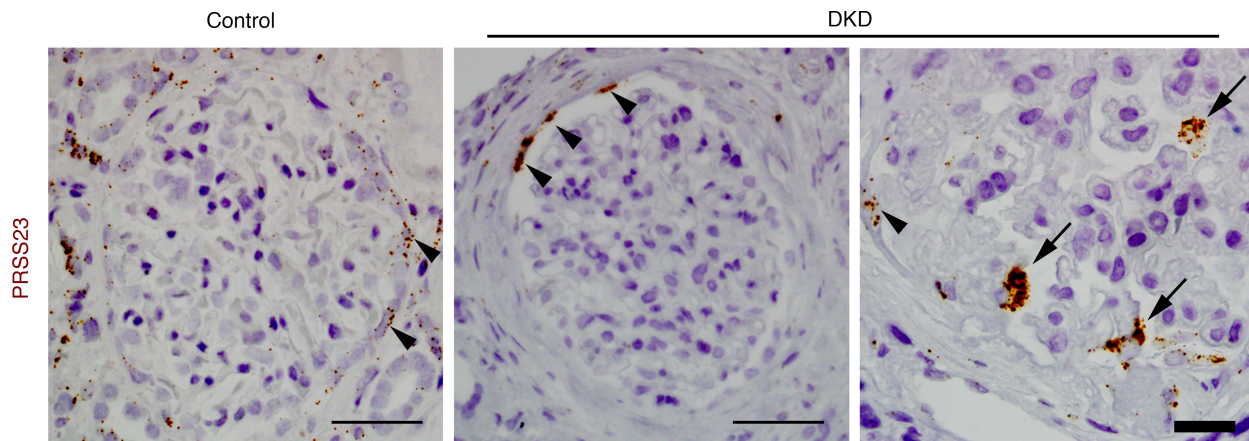


Figure 8. Expression of PRSS23 in diseased human glomeruli. PRSS23 expression as detected by ISH in control and diabetic kidney disease (DKD) sections; arrowheads highlight staining of cells at the glomerular parietal surface, and arrows highlight cells on the visceral surface. Scale bars: 50 μ m (left and middle); 20 μ m (right). Representative images of 5–10 glomeruli from each of 2 separate DKD samples.

an EGE cell gene expression signature generated from single cell transcriptional profiling of hPSC-derived kidney organoids, and we show that this gene set is differentially expressed in diseased human glomeruli. By combining cell clustering and cell lineage trajectory analyses of organoid and developing human kidney single cell transcriptomes, we show that cells in 2 distinct *NPHS1/NPHS2*-expressing organoid clusters shared the same developmental transcriptional program with human PCs in vivo but represented different stages of PC development, revealing that the gene signature arose from EGE cells. Identification of a glomerular disease–relevant gene expression signature corresponding to the early developmental PC stages supports the concept of transcriptional plasticity as a compensatory mechanism in glomerular disease.

Most of the 9 EGE signature genes were not previously appreciated as being involved in glomerular pathology, and several are minimally annotated, revealing new gene targets for investigation. Of the 3 signature genes (*LYPD1*, *PRSS23*, and *CDH6*) that were found to be associated with proteinuria and loss of kidney function, *PRSS23* is of particular interest, given this serine protease’s proposed role in Snail-dependent epithelial-to-mesenchymal transition (34). We found that *PRSS23* expression was increased in developing EGEs and PECs relative to PCs, but only *Cdh6* (of the 9 signature genes) was increased in healthy PEC-enriched rat glomerular isolates (35). This suggests that expression of these genes in our diseased human glomerular tissues does not simply represent inclusion of healthy PECs in the glomerular tissue samples. In DKD glomeruli, cells expressing *PRSS23* were seen both at Bowman’s capsule and the urinary glomerular basement membrane surface. The influx of cells from Bowman’s capsule is consistent with prior descriptions of podocytopathies in humans (36–38) and was shown to be of PEC origin by marker expression in human tissue and lineage tracing in a mouse FSGS model (39). An independent mouse model of FSGS showed that *Prss23* expression was significantly increased in lineage-tagged PCs isolated 6 weeks after disease induction (40), though it is possible that expression may be due to transdifferentiating PECs that express *COL1A1*, which was used in the lineage tag. Together, these observations support the intriguing possibility that *PRSS23* is involved in the development of a migratory PEC-PC phenotype in glomerular disease. Other possible cells responsible for the signature gene expression include transdifferentiating renin lineage cells (41), which can migrate across the glomerular or mesangial basement membrane. However, our description of these 9 genes within the context of developing PCs provides more compelling evidence in support of PEC-PC phenotype plasticity.

In addition to identifying a potentially novel disease signature, our comparative trajectory analysis fills a critical knowledge gap by validating that cells in hPSC-derived organoids faithfully reproduce the genetic developmental program of PCs, as well as other cell types in kidney organoids. The organoids recapitulated gene expression in cells of developing human kidneys and led to the description of a gene expression signature unique to developing PCs. Surprisingly, though developing PCs expressed these genes in vivo, some genes were identified only by differential transcriptional profiling in developing PCs isolated from kidney organoids as opposed to developing kidneys themselves, due to the more targeted analysis in organoid cultures. Our analysis also revealed the complexity of molecular signatures distinguishing PC

dedifferentiation versus PEC transdifferentiation, as illustrated by our trajectory analysis in which PC and PEC developmental lineages emerged from a common glomerular epithelial precursor (42). Divergence of these 2 cell lineages may be due to a transition of transcriptional programming similar to what was recently shown for Notch signaling driving intercalated to principal cell transition (43), a concept which is supported in our current study by the alterations in TF expression patterning along the PC trajectory. Comprehensive analysis of TF patterning in PC-PEC development may help discern these lineages more fully and, when this knowledge is applied to glomerular disease, may aid in identification of upstream regulators of disease-driving transcriptional activity.

Our dual transcriptional characterization of the repertoire of cells generated in PSC-derived kidney organoids also elucidates how each analytic method contributes complementary insights regarding kidney organoid cultures, as well as kidney development and disease. While the trajectory analysis confirmed cell types and vintages of the organoid cells relative to developing human kidney, cell clustering analysis highlighted cell type-specific gene expression profiles and provided intriguing clues regarding cell lineage development (e.g., 2 separate clusters of PCs corresponded to different segments of PC development, as discussed above). Cluster analysis also revealed that kidney cells (including PCs, tubular, and stromal) could be highly enriched by separating the tubular organoids from the surrounding mesenchyme, whereas off-target cell types identified in previous characterizations (15, 43, 44) were successfully depleted. However, both single cell transcriptional analyses showed that kidney cell types generated in organoid cultures were reproducible and robust. Thus, our study demonstrates that single cell transcriptional characterization of kidney organoids can be exploited in multiple ways, including defining and refining the cellular complexity of organoids and discovering transcriptional programmatic themes.

Although this work represents a significant advance in our understanding of kidney organoid transcriptional states, several factors limit interpretation of our results and warrant future investigation. Mesangial and endothelial cells were not described in this analysis, potentially as a result of developmental immaturity. An additional observation was that organoid kidney cells corresponded to earlier developmental stages of human developing kidneys in late first trimester to mid-second trimester (not surprising for cells differentiated less than 3 weeks in culture), meaning late stages of nephron development are not well represented. Indeed, prior studies showed that PSC-derived organoid PCs are similar to mammalian PCs at the capillary loop stage based on cytoskeletal architecture and marker expression (17). Together, identification of such factors provides a framework to direct further optimization of organoid differentiation protocols.

In conclusion, in addition to expanding our knowledge of transcriptional events during PC development, our approach has unveiled several glomerular disease-associated transcriptional programs. Further exploration regarding the basis of this developmental glomerular gene signature may aid in identification of potentially novel glomerular disease biomarkers and treatment strategies. Deciphering such events in diseased human glomeruli, often accompanied by manifestations of advanced disease such as sclerosis and fibrosis, is challenging. In contrast, human kidney organoids present an opportunity to focus on key glomerular cell types such as PCs and PEC in a system that is highly amenable to purification, treatment with small molecules, and omics-scale analysis. Moreover, characterizing kidney organoids on a single cell transcriptomic level provides a priori knowledge of gene sets available for experimental manipulation *ex vivo*. Thus, a beneficial role of PSC-derived kidney organoid cultures to identify potentially novel biomarkers of glomerular specific kidney diseases is revealed. Our results highlight the utility of kidney organoids as a discovery tool to define and investigate pathomechanisms of glomerular disease.

Methods

Generation of kidney organoids. Kidney organoids were generated using UM77-2 hESCs (NIH registration no. 0278) as described (9, 15). Mycoplasma contamination-free status of actively passaged cell lines was confirmed prior to differentiation using the Universal Mycoplasma Detection Kit (ATCC, 30-1012K) per manufacturer's protocol. Organoids for immunocytochemistry were picked 20–27 days after plating, fixed in 4% paraformaldehyde (Electron Microscopy Sciences)/PBS (Thermo Fisher Scientific), infiltrated with a sequential gradient of sucrose in PBS, and embedded in 20% sucrose/OCT (Tissue-Plus, Thermo Fisher Scientific) as previously described (44).

Human kidney tissues. Kidney tissue was obtained from patients undergoing nephrectomy at the University of Michigan (45, 46) and prepared via the tissue procurement service at the University of Michigan Comprehensive Cancer Center. Normal human kidney tissues were obtained from protocol pretransplant

donor biopsy, and human diabetic kidney tissues were obtained from individuals then procured by, routinely processed by, and accessed through the archives of the University of Michigan Department of Pathology. All tissues were formalin fixed, paraffin embedded, and sectioned at 3 μ m thickness.

Immunostaining and imaging of organoids and human kidney tissues. Cryosections were rehydrated with PBS, blocked with 5% normal donkey serum (The Jackson Laboratory) plus 0.1% Triton X-100 (IBI Scientific) in PBS, and immunostained with antibodies in 3% BSA (Cohn Fraction V, MilliporeSigma) in PBS. Sequential WT1 and PTPRO staining of the same section of normal human kidney (donor) protocol biopsy was performed as previously described, except that anti-WT1 was used to identify PC nuclei (47, 48). Species-appropriate fluorescence-tagged or peroxidase conjugated secondary antibodies were used to detect primary antibodies. Human nephrectomy sections stained with PAX8, NPISH1, PODXL, and CDH6 primary antibodies, followed by appropriate HRP conjugated secondary antibodies. Samples were mounted using SlowFade or Prolong Gold with DAPI (Invitrogen). DAB was used to develop peroxidase product and counterstained with hematoxylin and/or periodic acid Schiff. Images were acquired using Olympus SZ61 microscope with TL3 light source and Dino-Eye 5.0MP color microscope camera, Leica DM IRB microscope with Olympus DP70 digital camera, Aperio scanner, or Olympus FV500 confocal microscope.

ISH of human kidney tissue. RNAscope probe for Hs-PRSS23 was obtained from Advanced Cell Diagnostics (catalog 506571), and samples were prepared, probed, and counterstained per the manufacturer's protocol.

Rat FSGS model tissue and glomerular transcriptional data. Frozen kidney tissue from sham nephrectomized WT and unilaterally nephrectomized TG Fischer 344 rats (expressing the AA-4E-BP1 transgene under the control of the human podocin promoter) was previously generated as described (33) and cryosectioned. Microarray data were previously generated as part of the same study from isolated glomeruli using Rat Gene ST 2.1 Affymetrix gene array and reanalyzed for genes of interest using the described method (33).

Antibodies. Alexa-FLUOR-labeled secondary Abs were obtained from Invitrogen. A peroxidase-conjugated mouse IgG kit was obtained from Vectorstain (catalog PK-6102). Primary antibodies were sourced as follows: Nephryn (R&D Systems, AF4269), Cadherin-6 (R&D Systems, MAB2715), MEIS1/2/3 (Active Motif, ATM39795), Podocalyxin (R&D Systems, AF1658), WT-1 (Abcam, ab89901), PTPRO (5C11, in-house; ref. 49), N-cadherin (R&D Systems, AF6426), ZO-1 (Invitrogen, 40-2300), pan-cytokeratin (MilliporeSigma, C2562), and E-cadherin (BD Biosciences, 610181).

Organoid scRNA-seq and bioinformatic analysis. Kidney organoids were harvested between days 18–21 in culture, as either whole well or isolated spheroids (Figure 1B). Cells were dissociated using cold active protease into single cells and droplet sequencing (Drop-seq) was performed as previously described (15). Single cell transcriptomes from 7 datasets generated from 4 separate kidney organoid culture experiments (3 biological replicates) were included in the whole well analysis, while a single dataset was included in the isolated organoids analysis. Organoid scRNA-seq data matrix preprocessing, normalization, log transformation, unsupervised cell clustering, scaling, PCA dimensionality reduction, highly variable gene identification, individual gene expression query and presentation, and correlation with cell clusters previously derived from developing human kidney scRNA-seq (32) were all performed using the Seurat R package as described (15, 32, 50). Cells expressing <500 or >4,000 genes or >25% mitochondrial reads were excluded in the analysis. Seurat ScaleData function was used to regress out technical variables, including mitochondrial read content, number of Unique molecular identifier (UMI) per cell, and batch effect. Unsupervised clustering was performed with a resolution level of 0.6. t-SNE clustering was used for 2-D frame viewing of clusters (51).

Combined developmental trajectory cell lineage from single cell transcriptomes. Trajectory analysis was performed with a method developed by Menon et al. (32). Cells expressing at least 1,000 genes were included for analysis. To control for technical variations within the single cell transcriptomic data, linear modeling in Seurat was used to regress out correlation with the number of genes expressed, percentage of mitochondria reads, and batch variables. To control for confounding effects from cell cycle, the cell cycle signal was regressed out using regularized linear models, as previously described (32). Top principal components were used for trajectory analysis. Joint trajectory analysis combined human developing kidney (GEO GSE109205) and organoid datasets (7 whole well plus 1 isolated organoid) as follows. Each was individually processed and technical variations removed as described above; then, all were aligned using the canonical correlation analysis (CCA) method as implemented in Seurat; the canonical component projections were used instead of the principal component projections for the joint trajectory analysis. In Supplemental

Figure 2A, which includes all cells expressing ≥ 200 genes, we first mapped all cells to the CCA-transformed space based on their nearest neighbors in gene expression; then, cells were projected to the trajectory using the same procedure. For data presentation referring to cell order, cells are aligned sequentially based on the pseudo time order from the estimated trajectory.

Transcriptomic analysis of human kidney tissue. The discovery cohort included previously generated microarray data from microdissected human glomeruli and sourced from individuals with kidney disease ($n = 170$) and healthy donors ($n = 21$) from ERCB (52), which were accessed in GEO by accession GSE104948. The validation cohort ($n = 111$) also included previously generated microarray data from microdissected glomeruli from individuals with FSGS ($n = 30$), MCD ($n = 15$), MN ($n = 44$) (all drawn from NEPTUNE; ref. 53), and AAV ($n = 15$) (drawn from ERCB) (29), accessed in GEO by accession GSE108113. Differential expression analysis was restricted to probe sets ($n = 12,074$) common to both Affymetrix microarray platforms (GeneChip Human Genome U133A 2.0 and U133 Plus 2.0, Thermo Fisher Scientific) by comparing transcriptional profiles from individuals with CKD versus living donors using the significance analysis of microarrays (SAM) method (29).

Aggregate gene expression scores (Figure 5A) were calculated based on the genes from Supplemental Table 1 for EGE (C1) and MPC (C7). Genes shared by these 2 lists were used to generate a shared gene list, with the residual genes in each of the EPC and MPC lists used in order to define those clusters. Gene lists were further distilled to the subset of genes queried on the microarray. An aggregate gene expression score was then generated for each of these gene lists by averaging expression of all genes, after each gene's \log_2 expression was Z score transformed.

GFR in ERCB and NEPTUNE was estimated by the 4-variable Modification of Diet in Renal Disease (MDRD) study equation (54) and \log_2 transformed prior to correlation with expression data, as described (29). eGFRs were available in all 191 individuals in the discovery cohort and 101 individuals in the validation cohort. Correlation analyses between gene expression and \log_2 eGFR were performed using Pearson's correlation coefficient.

Correlation of gene expression with proteinuria was performed in the ERCB LN cohort in which proteinuria was quantified (55) (GEO GSE32591), accessed in Nephroseq (www.nephroseq.org, Berthier Lupus Glomerulonephritis dataset, and analyzed using the inherent analysis mechanism that includes precomputation of differential expression profiles using Student's t test for 2 class differential expression analyses. In the validation cohort, 88 individuals had a UPCr available, and the cohort did not include individuals with AAV.

Data access. Normalized organoid scRNA-seq gene expression data files reported herein were submitted to NCBI Gene Expression Omnibus (GEO GSE115986).

Statistics. For single cell transcriptional analysis by t-SNE clustering, the Seurat program's default non-parametric Wilcoxon rank sum test was used for differential expression analysis, and values were considered significant for adjusted $P \leq 0.05$. For single cell transcriptional trajectory analysis, heatmaps were generated from genes with average \log_2 -scaled expression > 0.1 and P values of differential expression along the trajectory < 0.001 by 2-sided Wald test, while individual gene expression plots were generated from average cell gene expression along the trajectory using a curve fitting method (generalized additive model; GAM) with 95% CI to indicate significance. For human kidney tissue transcriptional analysis, the significance analysis of microarrays (SAM) method (29) was used to define genes as differentially expressed in disease biopsies relative to healthy living donor biopsies if they met $q < 0.05$. Association between \log_2 gene expression values and \log_2 UPCr was precalculated in Nephroseq.org using 1-tailed Student's t test for 2 class differential expression and considered significant if $P < 0.05$. Correlation analyses between \log_2 gene expression values and \log_2 eGFR were performed using Pearson's correlation. For rat kidney tissue transcriptional analysis, statistical significance was calculated using t test analysis and the statistical significance threshold is $P \leq 0.05$.

Box outlines indicate 25th–75th percentile data range, and horizontal white or black lines inside the boxes indicate median values; dots are outliers. For Figure 5 and Figure 7, A and B, whiskers represent the upper adjacent value (UAV) and lower adjacent value (LAV). The LAV is the smallest observation that is greater than or equal to the first quartile minus $1.5 \times$ interquartile range (IQR), while the UAV is the largest observation that is less than or equal to the third quartile plus $1.5 \times$ IQR. For Figure 7C, whiskers represent the 10th/90th percentile values as precomputed within Nephroseq. For Supplemental Figure 1, D and E, whiskers represent the minimal and maximal values.

Study approval. Study approval was granted by the University of Michigan Human Pluripotent Stem Cell Research Oversight compliance program (application no. 1096) and the ethics committees for ERCB and NEPTUNE studies. The University of Michigan IRB approved gene expression studies (HUM0002468) and human kidney tissue procurement (HUM00045864, HUM00083116, HUM00052918). Written informed consent was received from participants prior to inclusion.

Author contributions

All authors critically reviewed the manuscript. JLH planned and performed organoid experiments, supervised NLW, designed and managed the project, critically reviewed data, and wrote the manuscript. NLW and EAO planned and performed organoid and Drop-seq experiments, respectively. NLW, JL, and CO performed immunostaining of human kidney tissue. RM performed single cell transcriptomic analysis for organoids and developing human kidney. JZ performed trajectory analysis of single cell transcriptomes. SE analyzed human tissue transcriptome-disease correlations and identified high-priority genes. VN analyzed rat transcriptomic data. CC and JRS provided developing human kidney data and provided expertise related to stem cell culture protocols. MB supervised CO. OGT supervised JZ. JBH supervised JL and, along with RCW, provided and analyzed human kidney tissue samples. RCW also provided rat tissue transcriptomic data and critically reviewed data. BSF provided organoid and stem cell expertise and critically reviewed data. MK critically reviewed data and supervised the entire project. ERCB and NEPTUNE consortia provided patient data for analysis.

Acknowledgments

The authors extend their gratitude to investigators and participants involved in the ERCB and the NEPTUNE consortium. See Supplemental Acknowledgments for consortium details. This research was supported by NIH grants K08 DK089119 to JLH, U54 DK083912 to MK, UG3 TR002158 to MK and BSF, K01 DK102826 to BSF, and P30 DK081943 Applied Systems Biology Core at the University of Michigan George M. O'Brien Kidney Translational Core Center and P30 CA046592 Rogel Cancer Center Tissue and Molecular Pathology Shared Resource. Additional funding and/or programmatic support was provided by the NephCure Kidney International and the Halpin Foundation through support for NEPTUNE. See Supplemental Acknowledgments for ERCB and Nephrotic Syndrome Study Network details.

Address correspondence to: Matthias Kretzler, University of Michigan, 1150 W. Medical Center Dr., 1570 MSRB2, Ann Arbor, Michigan 48109 USA. Phone: 734.615.5757; Email: kretzler@umich.edu.

CC's present address is: Department of Pediatrics, Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA.

- Liu ZH, He JC, eds. *Podocytopathy*. Basel: Karger; 2014.
- Hodgin JB, et al. Identification of cross-species shared transcriptional networks of diabetic nephropathy in human and mouse glomeruli. *Diabetes*. 2013;62(1):299–308.
- Shankland SJ, Pippin JW, Reiser J, Mundel P. Podocytes in culture: past, present, and future. *Kidney international*. 2007;72(1):26–36.
- Cohen CD, Frach K, Schlondorff D, Kretzler M. Quantitative gene expression analysis in renal biopsies: a novel protocol for a high-throughput multicenter application. *Kidney international*. 2002;61(1):133–140.
- Cohen CD, Grone HJ, Grone EF, Nelson PJ, Schlondorff D, Kretzler M. Laser microdissection and gene expression analysis on formaldehyde-fixed archival tissue. *Kidney international*. 2002;61(1):125–132.
- Bennett MR, Czech KA, Arend LJ, Witte DP, Devarajan P, Potter SS. Laser capture microdissection-microarray analysis of focal segmental glomerulosclerosis glomeruli. *Nephron Exp Nephrol*. 2007;107(1):e30–40.
- Takemoto M, et al. Large-scale identification of genes implicated in kidney glomerulus development and function. *The EMBO journal*. 2006;25(5):1160–1174.
- Ju W, et al. Defining cell-type specificity at the transcriptional level in human disease. *Genome Res*. 2013;23(11):1862–1873.
- Freedman BS, et al. Modelling kidney disease with CRISPR-mutant kidney organoids derived from human pluripotent epiblast spheroids. *Nat Commun*. 2015;6:8715.
- Takasato M, et al. Kidney organoids from human iPSCs contain multiple lineages and model human nephrogenesis. *Nature*. 2015;526(7574):564–568.
- Morizane R, Lam AQ, Freedman BS, Kishi S, Valerius MT, Bonventre JV. Nephron organoids derived from human pluripotent stem cells model kidney development and injury. *Nat Biotechnol*. 2015;33(11):1193–1200.
- Taguchi A, et al. Redefining the in vivo origin of metanephric nephron progenitors enables generation of complex kidney structures from pluripotent stem cells. *Cell stem cell*. 2014;14(1):53–67.

13. Macosko EZ, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*. 2015;161(5):1202–1214.
14. Adam M, Potter AS, Potter SS. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. *Development*. 2017;144(19):3625–3632.
15. 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.
16. Cruz NM, et al. Organoid cystogenesis reveals a critical role of microenvironment in human polycystic kidney disease. *Nature materials*. 2017;16(11):1112–1119.
17. Kim YK, et al. Gene-Edited Human Kidney Organoids Reveal Mechanisms of Disease in Podocyte Development. *Stem cells*. 2017;35(12):2366–2378.
18. Uhlén M, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419.
19. Little MH, ed. *Kidney development, disease, repair, and regeneration*. Amsterdam; Boston: Academic Press; 2016.
20. Whitfield ML, George LK, Grant GD, Perou CM. Common markers of proliferation. *Nat Rev Cancer*. 2006;6(2):99–106.
21. Menon R, et al. Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney. *Development*. 2018;145(16):dev164038.
22. Lovisa S, et al. Epithelial-to-mesenchymal transition induces cell cycle arrest and parenchymal damage in renal fibrosis. *Nat Med*. 2015;21(9):998–1009.
23. Lambert SA, et al. The Human Transcription Factors. *Cell*. 2018;172(4):650–665.
24. Sonawane AR, et al. Understanding Tissue-Specific Gene Regulation. *Cell Rep*. 2017;21(4):1077–1088.
25. Carro MS, et al. The transcriptional network for mesenchymal transformation of brain tumours. *Nature*. 2010;463(7279):318–325.
26. Walsh LA, et al. An Integrated Systems Biology Approach Identifies TRIM25 as a Key Determinant of Breast Cancer Metastasis. *Cell Rep*. 2017;20(7):1623–1640.
27. Motojima M, Kume T, Matsusaka T. Foxc1 and Foxc2 are necessary to maintain glomerular podocytes. *Exp Cell Res*. 2017;352(2):265–272.
28. Maezawa Y, et al. Loss of the podocyte-expressed transcription factor Tcf21/Pod1 results in podocyte differentiation defects and FSGS. *J Am Soc Nephrol*. 2014;25(11):2459–2470.
29. Grayson PC, et al. Metabolic pathways and immunometabolism in rare kidney diseases. *Ann Rheum Dis*. 2018;77(8):1226–1233.
30. 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.
31. Cho EA, Patterson LT, Brookhiser WT, Mah S, Kintner C, Dressler GR. Differential expression and function of cadherin-6 during renal epithelium development. *Development*. 1998;125(5):803–812.
32. Menon R, et al. Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney. *Development*. 2018;145(16):dev164038.
33. Nishizono R, et al. FSGS as an Adaptive Response to Growth-Induced Podocyte Stress. *J Am Soc Nephrol*. 2017;28(10):2931–2945.
34. Chen IH, Wang HH, Hsieh YS, Huang WC, Yeh HI, Chuang YJ. PRSS23 is essential for the Snail-dependent endothelial-to-mesenchymal transition during valvulogenesis in zebrafish. *Cardiovasc Res*. 2013;97(3):443–453.
35. Gharib SA, Pippin JW, Ohse T, Pickering SG, Krofft RD, Shankland SJ. Transcriptional landscape of glomerular parietal epithelial cells. *PLoS ONE*. 2014;9(8):e105289.
36. Smeets B, et al. Renal progenitor cells contribute to hyperplastic lesions of podocytopathies and crescentic glomerulonephritis. *J Am Soc Nephrol*. 2009;20(12):2593–2603.
37. Kuppe C, et al. Common histological patterns in glomerular epithelial cells in secondary focal segmental glomerulosclerosis. *Kidney Int*. 2015;88(5):990–998.
38. Smeets B, et al. Detection of activated parietal epithelial cells on the glomerular tuft distinguishes early focal segmental glomerulosclerosis from minimal change disease. *Am J Pathol*. 2014;184(12):3239–3248.
39. Appel D, et al. Recruitment of podocytes from glomerular parietal epithelial cells. *J Am Soc Nephrol*. 2009;20(2):333–343.
40. Grgic I, et al. Discovery of new glomerular disease-relevant genes by translational profiling of podocytes in vivo. *Kidney Int*. 2014;86(6):1116–1129.
41. Eng DG, et al. Detection of renin lineage cell transdifferentiation to podocytes in the kidney glomerulus with dual lineage tracing. *Kidney Int*. 2018;93(5):1240–1246.
42. Guhr SS, et al. The expression of podocyte-specific proteins in parietal epithelial cells is regulated by protein degradation. *Kidney Int*. 2013;84(3):532–544.
43. Park J, et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science*. 2018;360(6390):758–763.
44. Barthel LK, Raymond PA. Improved method for obtaining 3-microns cryosections for immunocytochemistry. *J Histochem Cytochem*. 1990;38(9):1383–1388.
45. Hodgin JB, et al. Glomerular Aging and Focal Global Glomerulosclerosis: A Podometric Perspective. *J Am Soc Nephrol*. 2015;26(12):3162–3178.
46. Kato M, et al. An endoplasmic reticulum stress-regulated lncRNA hosting a microRNA megacluster induces early features of diabetic nephropathy. *Nat Commun*. 2016;7:12864.
47. Venkatarreddy M, et al. Estimating podocyte number and density using a single histologic section. *J Am Soc Nephrol*. 2014;25(5):1118–1129.
48. Yang Y, et al. The two kidney to one kidney transition and transplant glomerulopathy: a podocyte perspective. *J Am Soc Nephrol*. 2015;26(6):1450–1465.
49. Kim YH, et al. Podocyte depletion and glomerulosclerosis have a direct relationship in the PAN-treated rat. *Kidney Int*. 2001;60(3):957–968.
50. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018;36(5):411–420.
51. Shekhar K, et al. Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell*.

- 2016;166(5):1308–1323.e30.
52. Schmid H, Cohen CD, Henger A, Schlöndorff D, Kretzler M. Gene expression analysis in renal biopsies. *Nephrol Dial Transplant*. 2004;19(6):1347–1351.
53. 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.
54. Levey AS, et al. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Ann Intern Med*. 2006;145(4):247–254.
55. Berthier CC, et al. Cross-species transcriptional network analysis defines shared inflammatory responses in murine and human lupus nephritis. *J Immunol*. 2012;189(2):988–1001.