Gene regulation in embryonic development

Berg, P.R. van den

2 SINGLE-CELL TRANSCRIPTOMICS REVEALS GENE EXPRESSION DYNAMICS OF HUMAN FETAL KIDNEY DEVELOPMENT

THIS CHAPTER IS BASED ON:
Mazène Hochane, Patrick R van den Berg, Xueying Fan, Noémie Bérenger-Currias, Esmée Adegeest, Monika Bialecka, Maaike Nieveen, Maarten Menschaart, Susana M Chuva de Sousa Lopes, Stefan Semrau. "Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development". In: *PLOS Biology* 17.2 (Feb. 2019), e3000152. DOI: 10.1371/journal.pbio.3000152

Abstract

Regenerative medicine offers an exciting avenue to potential cures of kidney disease. However, a detailed knowledge of the structure and embryonic development of the kidney is crucial, both for stimulating regeneration in the body and for growing healthy kidney tissue in a dish. Most of such knowledge has been obtained from mice, whose development differs in crucial ways from that of humans. We therefore studied the composition of human fetal kidney tissue from five developmental ages, using a technique that can measure gene expression in individual cells. Our measurements revealed 22 distinguishable cell types, some of which we localized in the tissue by fluorescence microscopy. We found several subpopulations of nephron progenitor-cells that give rise to the nephron, the functional unit of the kidney. Our study also focused on the development of podocytes, a cell type that is crucial for the filtration function of the kidney, and our results might inform attempts to recreate these cells in a dish. We hope that our dataset, made conveniently accessible through a web application, will help scientists develop new regenerative medicine approaches to kidney disease.

2.1 Introduction

Mammalian kidney development initiates in the intermediate mesoderm through crosstalk between the metanephric mesenchyme (MM) and the ureteric bud (UB). The UB originates from the nephric duct, invades the MM, and starts to subdivide progressively into multiple ramifications. The UB tip cells, which make the first contact with the MM, become enveloped by an assembly of mesenchymal cells, the cap mesenchyme (CM) (Fig 1A-B). The CM contains nephron progenitor cells (NPCs), which give rise to the whole nephron epithelium through tightly regulated morphogenic transformations [2]. Self-renewal of (mouse) NPCs is governed by key transcription factors, such as *Six2* and *Meox1*, which mark the nephrogenic zone of the kidney [3]. Signaling between UB tip cells and NPCs regulates the balance between self-renewal and differentiation of the NPCs [4]. In humans, about 1 million nephrons are produced before the NPC population is irrevocably exhausted a few weeks before birth [5]. During nephrogenesis, NPCs undergo mesenchymal-epithelial transition and differentiate into a succession of intermediate structures: the pretubular aggregate (PTA), renal vesicle (RV), comma-shaped body (CSB) and s-shaped body (SSB). Then, via the capillary loop stage, mature and functional glomerular and tubular structures are eventually formed. In contrast to the nephron epithelium, the collecting duct system originates from the UB. GDNF/RET signaling between the UB and CM critically regulates proliferation of UB tip cells and branching morphogenesis of the UB [6]. Stromal cells-such as interstitial cells (ICs), mesangial cells, juxtaglomerular cells, smooth muscle cells, fibroblasts, and pericytes-derive from a common interstitial progenitor [7, 8]. Finally, vascular endothelial cells and the highly specified glomerular endothelium originate from the MM [9], and leukocytes and erythrocytes enter with the blood stream. The current understanding of mammalian kidney development is largely based on mouse studies, although it is clear that human and mouse kidneys are morphologically different. Three recent landmark studies have revealed, in great detail, a significant divergence between mouse and human renal embryogenesis in terms of morphology as well as gene expression [10, 11, 12]. These studies underline that the prevailing lack of data on human kidney development severely hinders the detailed understanding of human kidney development and possible developmental origins of kidney disease.

In the study described here, we used scRNA-seq to study gene expression dynamics in human fetal kidney development. Analysis of a fetal kidney from week 16 (w16) of gestation revealed 22 cell types, which we identified by known marker genes. Pseudotime analysis clarified their temporal relationship. We further defined specifically expressed cell type marker genes of which many have not been implied in kidney development. Comparison to four additional samples (from w9, w11, w13, and w18) suggested that most cell types have a constant expression pattern, with the notable exception of podocytes. To highlight two ways in which our dataset can be interrogated, we then explored the nephrogenic niche and the development of podocytes. Gene expression differences between four NPC clusters were related to spatial heterogeneity by immunostaining and single-molecule FISH (smFISH). Expression of the disease-associated gene *UNCX* was localized to NPCs and their early derivatives. Fi-

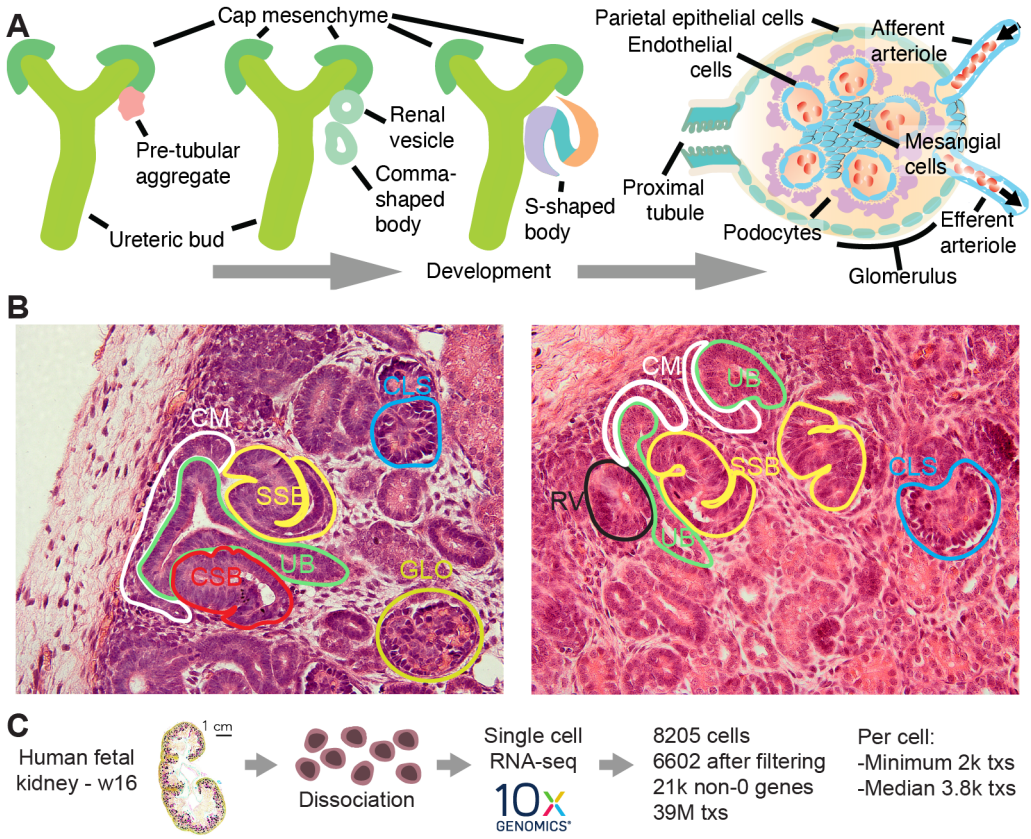


Figure 1. Overview of kidney development and the experimental design. (A) Schematic of kidney epithelium development. (B) Several morphologically distinct stages of nephrogenesis are highlighted by colored lines in images of human fetal kidney sections stained with HE. (C) Overview of the scRNA-seq experiment. Ureteric bud (UB), cap mesenchyme (CM), pretubular aggregate (PTA), renal vesicle (RV) (Stage I), comma-shaped body (CSB) (Stage II), s-shaped body (SSB) (Stage II), capillary-loop stage (CLS) (Stage III), glomerulus (GLO) (Stage IV)

nally, we focused on podocyte development, which proceeds via a distinct precursor state. By immunostaining and smFISH, we localized these precursors in situ and confirmed the disease-associated gene *OLFM3* as a marker.

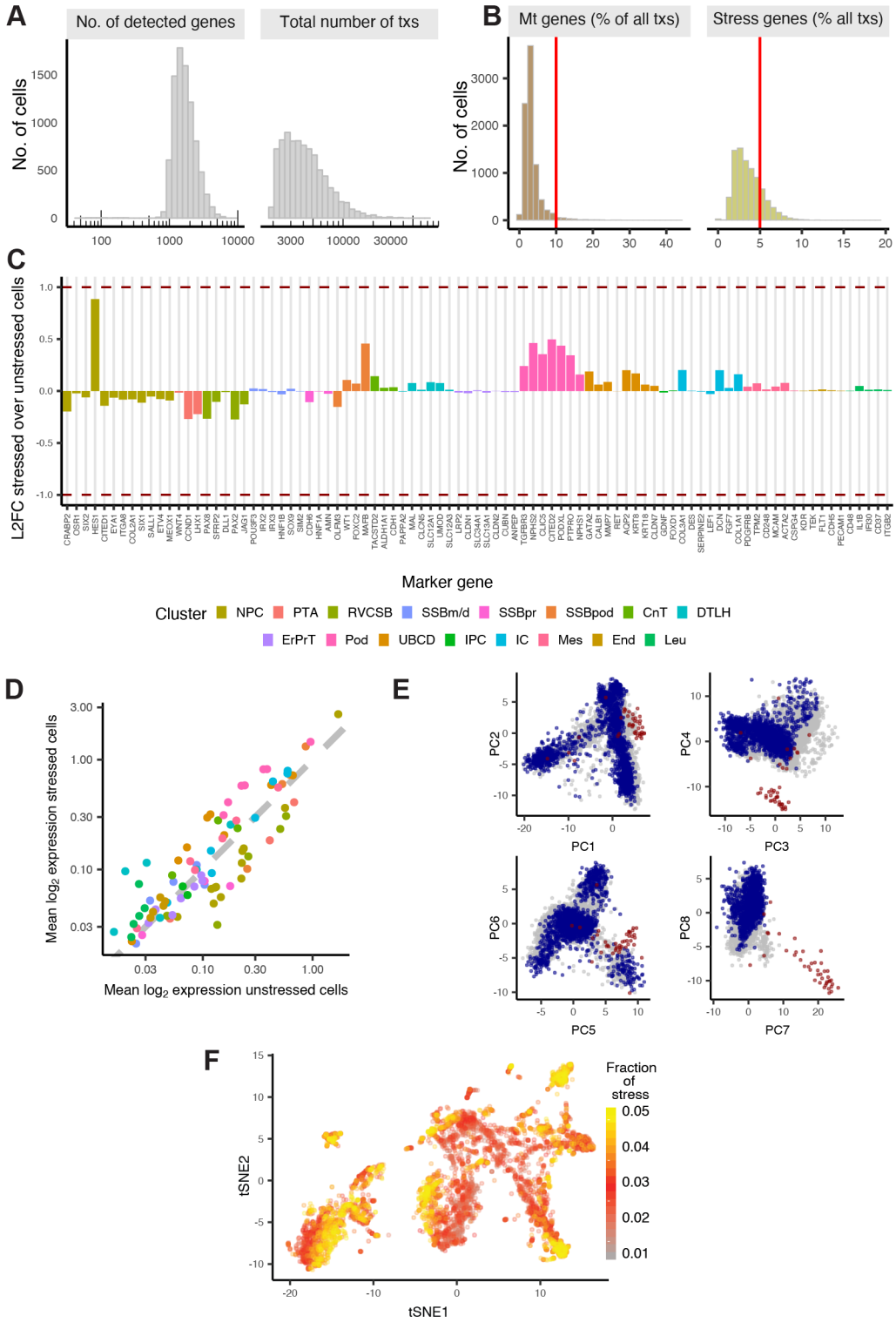
2.2 Results

2.2.1 Clustering and identification of cell types

We performed single-cell transcriptomics on a human fetal kidney from w16 of gestation, equivalent to 14 weeks of development (Fig 1C). After data pruning and stringent removal of

cells affected by stress (Fig 2, Materials and methods), 6,602 cells were retained for further analysis. Clusters of cells were identified by hierarchical clustering after k-nearest neighbors (knn) smoothing [13]. We assigned cell types (see Cell types) to these clusters by expression of marker genes from the literature on mouse kidney development. The studies that linked the genes of this literature set to particular cell types are referenced in Table 2.1. After merging similar clusters (Fig 4, Fig 5, Materials and methods), we obtained 22 cell types (Hochane et al. [1] S4 Fig) and visualized the single-cell transcriptomes in a two-dimensional t-distributed stochastic neighbor embedding (tSNE) map [14] (Fig 3).

Figure 2 ***(following page).*** **Removing stressed cells did not bias the scRNA-seq results in the w16 sample.** (A) Number of detected genes and total number of transcripts per cell. tx = transcript (B) Relative expression of mitochondrial and stress marker genes per cell. Red line indicates the threshold used to define stressed cells. See Materials and methods for the list of mitochondrial genes and Hochane et al. [1] S2 Table for the list of stress markers. (C and D) $\log_2$ fold change (L2FC) and scatter plot of the literature set genes (Table 2.1). Red dashed lines indicate fold-change of 0.5 and 2. (E) PCs one to eight of the top 5% most HVGs for all cells. Blue and red points indicate stressed cells and red blood cells, respectively. Expression values in C-E are normalized to library size and log-transformed with a pseudocount of 1. (F) Fraction of stress markers in the 6,602 remaining cells. tSNE map corresponds to Fig 3.


A
No. of detected genes
Total number of txs
No. of cells
B
Mt genes (% of all txs)
Stress genes (% all txs)
No. of cells
C
L2FC stressed over unstressed cells
Marker gene
Cluster NPC PTA RVCSB SSBm/d SSBpr SSBpod CnT DTLH ErPrT Pod UBCD IPC IC Mes End Leu
D
Mean log2 expression stressed cells
Mean log2 expression unstressed cells
E
PC1
PC2
PC3
PC4
PC5
PC6
PC7
PC8
F
tSNE1
tSNE2
Fraction of stress
0.05
0.04
0.03
0.02
0.01

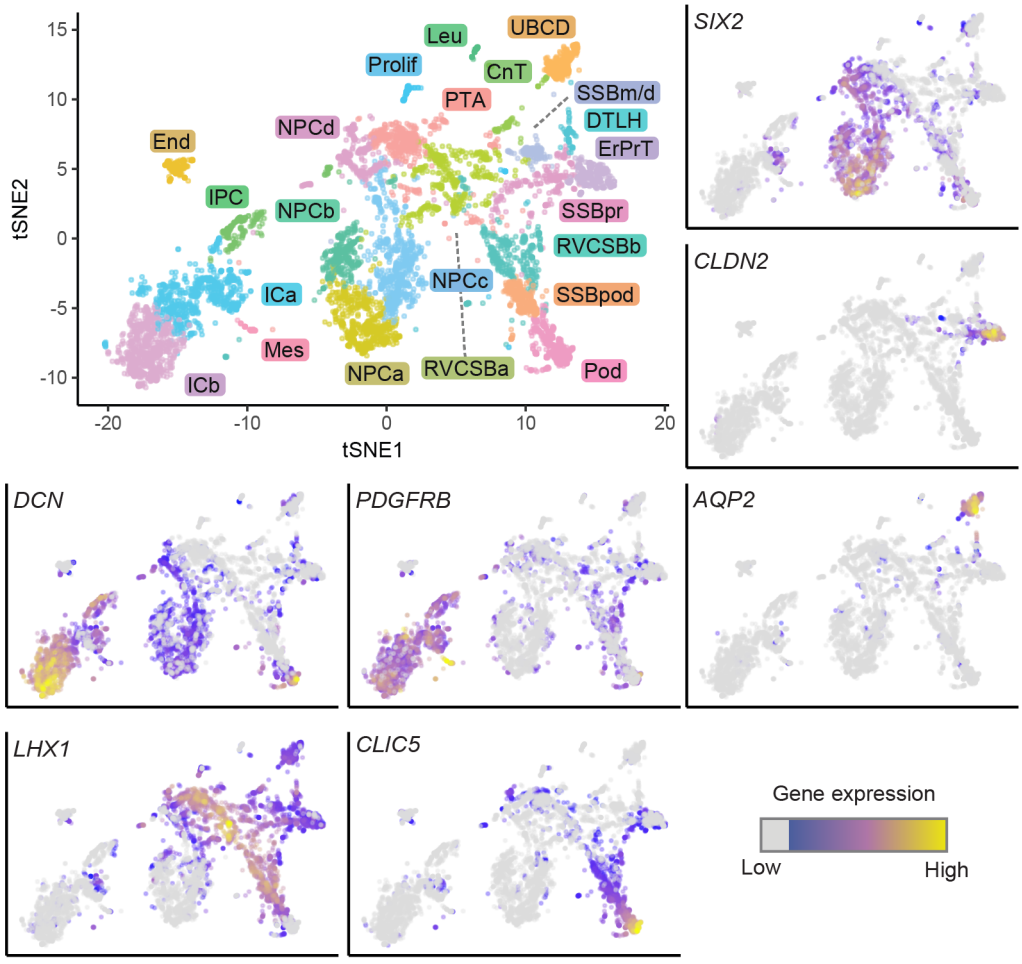
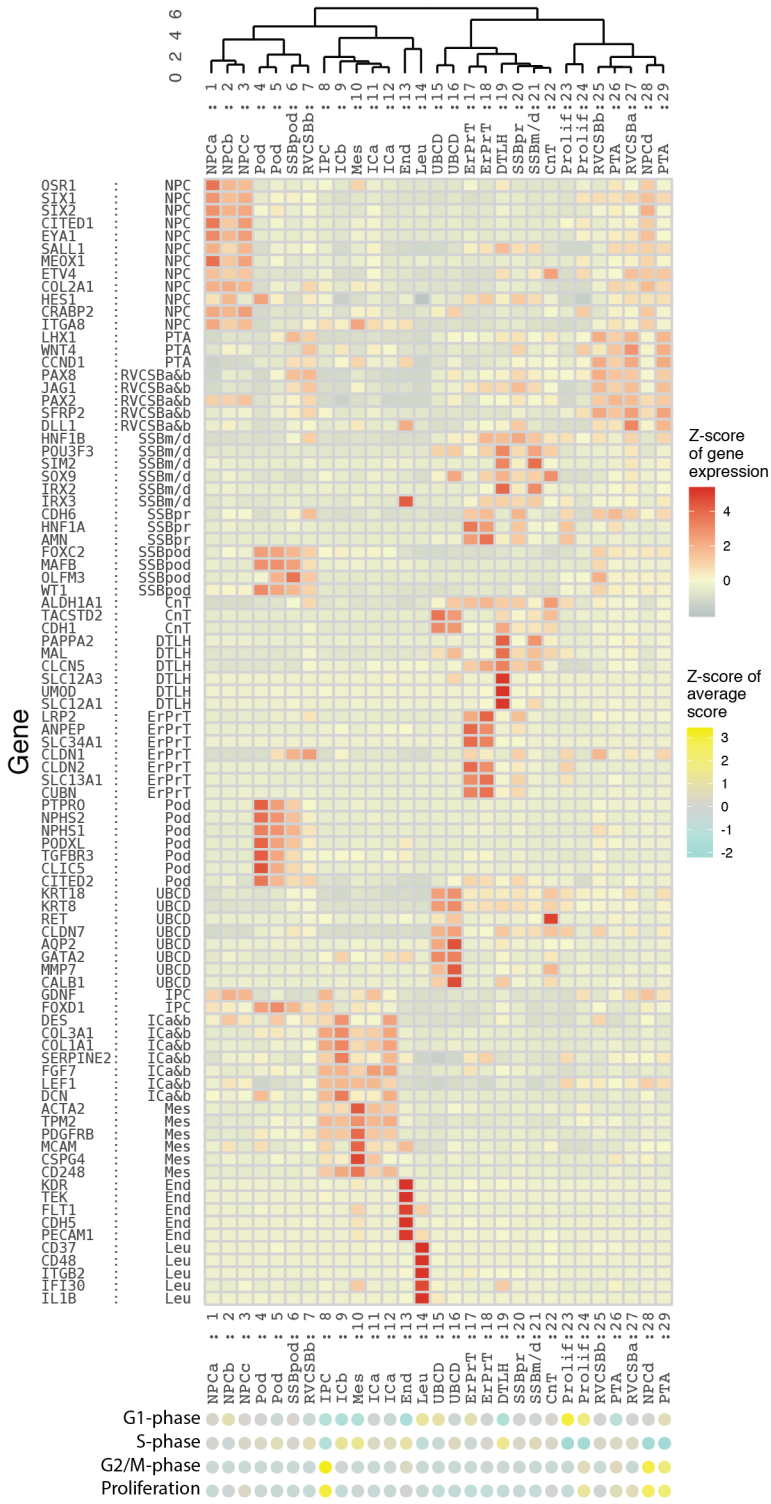
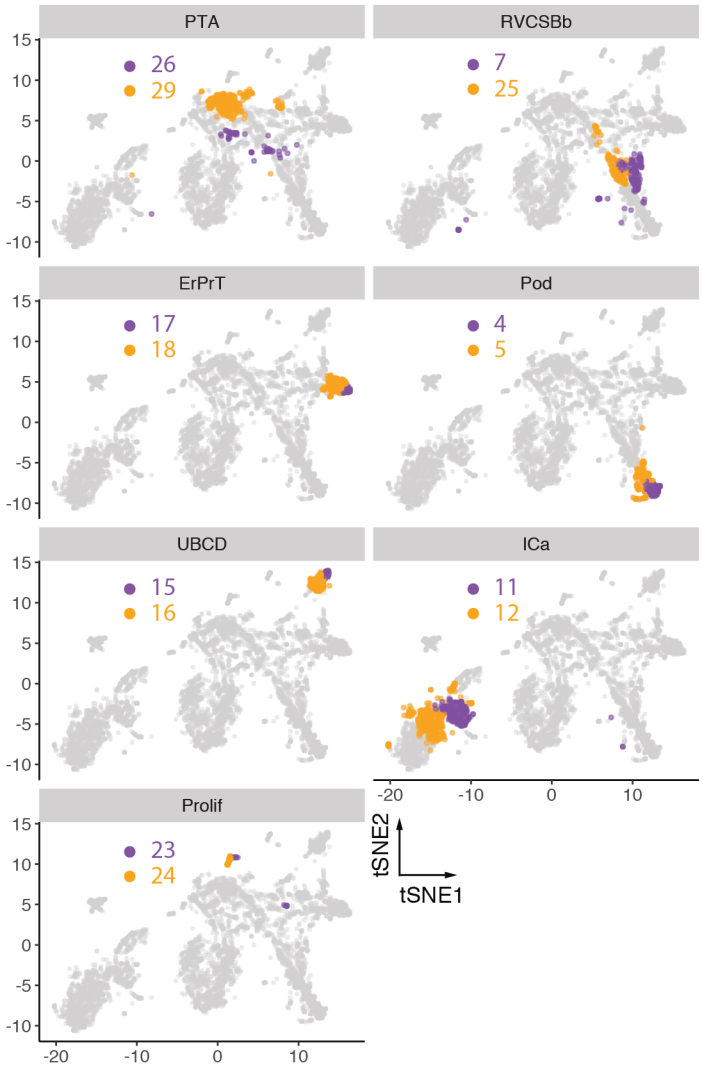


Figure 3. Single-cell transcriptomics identified 22 unique cell types in the human fetal kidney. Top left: 2D tSNE map of 6,602 human fetal kidney cells. Colors and labels indicate the assigned cell type. (Other panels) tSNE maps indicating expression of *SIX2*, *LHX1*, *CLDN2*, *CLIC5*, *DCN*, *PDGFRB*, and *AQP2*. Expression is indicated by color; expression values of 1 are plotted in gray.

Figure 4 *(following page)*. Adjacent clusters were merged based on similarity in literature set gene expression. Heat map of literature set gene expression. Expression was Freeman-Tukey (FT) transformed, averaged over all cells in the 29 clusters found by hierarchical clustering (indicated by the dendrogram on top of the heat map) and standardized gene-wise. Cluster average cell cycle scores, calculated by Cyclone [15] as well as average expression of proliferation markers [16], are indicated by colored circles below each cluster (Z-score of the mean score or mean expression).

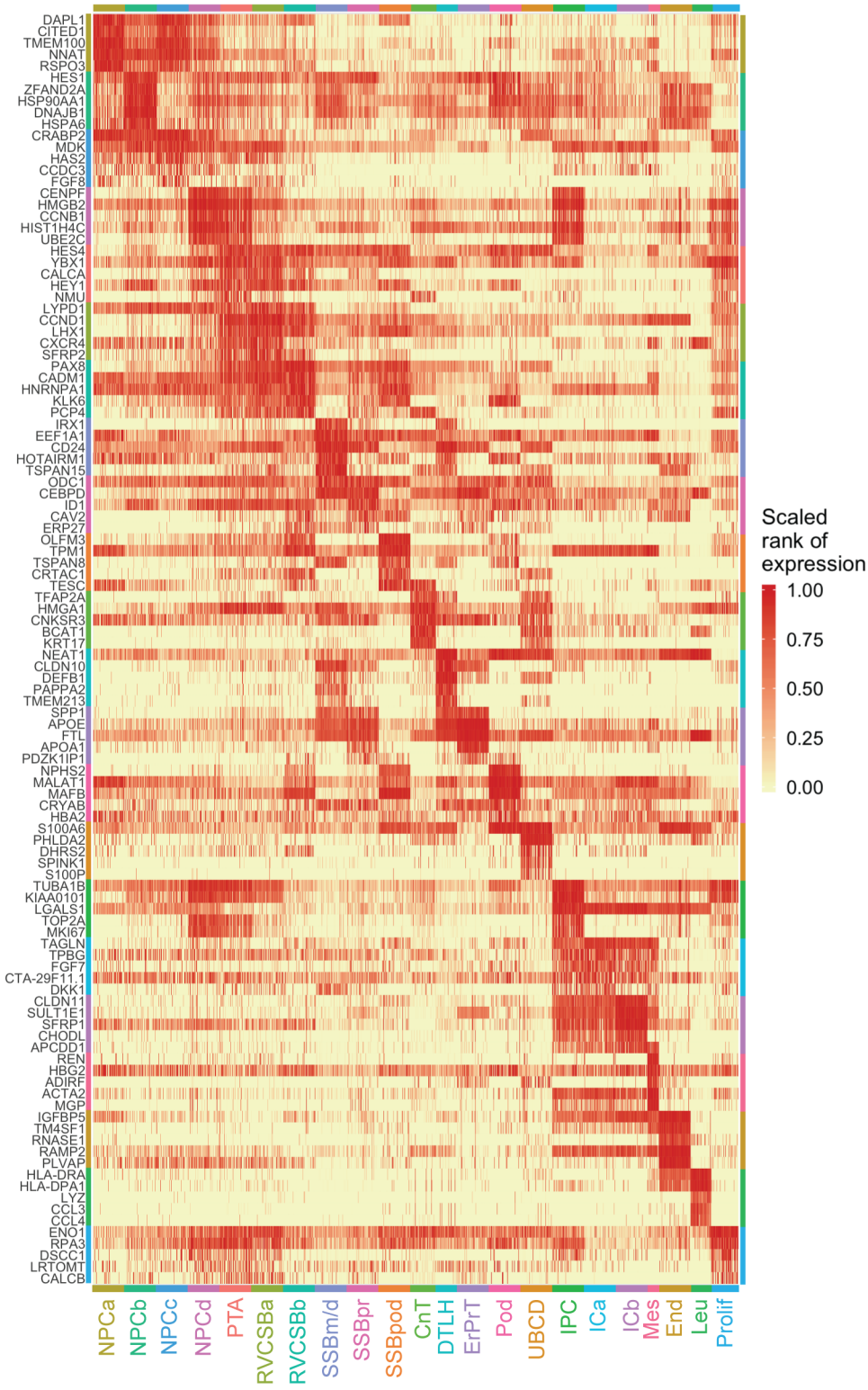

Gene
Z-score of gene expression
4
2
0
Z-score of average score
3
2
1
0
-1
-2
G1-phase
S-phase
G2/M-phase
Proliferation



Cell type	Clusters	No of cells
NPCa	1	522
NPCb	2	296
NPCc	3	596
NPCd	28	280
PTA	26&29	56&403
RVCSBa	27	367
RVCSBb	7&25	181&131
SSBm/d	21	151
SSBpr	20	213
SSBpod	6	257
CnT	22	80
DTLH	19	67
ErPrT	17&18	46&292
Pod	4&5	224&206
UBCD	15&16	55&262
IPC	8	177
ICa	11&12	272&381
ICb	9	677
Mes	10	37
End	13	223
Leu	14	65
Prolif	23&24	16&69

Figure 5. Merged hierarchical clusters were similar on the tSNE map. tSNE maps highlighting the clusters that were merged to give the cell types indicated in the titles of each map. (Inset right) Table listing the numbers of cells in each of the 29 original clusters. tSNE map corresponds to Fig 3.

Figure 6 *(following page)*. HVGs adequately described all cell clusters. Heat map of 2,034 randomly chosen cells (maximum 100 per cluster) and the five most HVGs with a minimum mean expression of 0.01 excluding stress markers (Hochane et al. [1] S2 Table) and ribosomal genes. Genes were assigned to clusters based on highest mean expression within that cluster. Values shown are the ranks of nonzero cells (cells with no expression receive rank 0) divided by the highest rank per gene.


DAPL1
CITED1
TMEM100
NNAT
RSPO3
HES1
ZFAND2A
HSP90AA1
DNAJB1
HSPA6
CRABP2
MDK
HAS2
CCDC3
FGF8
CENPF
HMGB2
CCNB1
HIST1H4C
UBE2C
HES4
YBX1
CALCA
HEY1
NMU
LYPD1
CCND1
LHX1
CXCR4
SFRP2
PAX8
CADM1
HNRNPA1
KLK6
PCP4
IRX1
EEF1A1
CD24
HOTAIRM1
TSPAN15
ODC1
CEBPD
ID1
CAV2
ERP27
OLFM3
TPM1
TSPAN8
CRTAC1
TESC
TFAP2A
HMGA1
CNKSR3
BCAT1
KRT17
NEAT1
CLDN10
DEFB1
PAPPA2
TMEM213
SPP1
APOE
FTL
APOA1
PDZK1IP1
NPHS2
MALAT1
MAFB
CRYAB
HBA2
S100A6
PHLDA2
DHRS2
SPINK1
S100P
TUBA1B
KIAA0101
LGALS1
TOP2A
MKI67
TAGLN
TPBG
FGF7
CTA-29F11.1
DKK1
CLDN11
SULT1E1
SFRP1
CHODL
APCDD1
REN
HBG2
ADIRF
ACTA2
MGP
IGFBP5
TM4SF1
RNASE1
RAMP2
PLVAP
HLA-DRA
HLA-DPA1
LYZ
CCL3
CCL4
ENO1
RPA3
DSCC1
LRTOMT
CALCB
NPCa
NPCb
NPCc
NPCd
PTA
RVCSBa
RVCSBb
SSBm/d
SSBpr
SSBpod
CnT
DTLH
ErPrT
Pod
UBCD
IPC
ICa
ICb
Mes
End
Leu
Prolif
Scaled rank of expression
1.00
0.75
0.50
0.25
0.00

The mean expression levels of the literature set genes showed clear differences between cell types (Fig 7). NPCs, which were distributed over four distinct clusters (NPCa-d), were marked by the established markers *SIX2* (Fig 3), *CITED1*, *MEOX1*, and *EYA1*. Expression of these progenitor markers was highest in NPCa, which we hence considered *bona fide* self-renewing NPCs. NPCb showed lower levels of *CITED1* and *SALL1* and higher levels of *GDNF* and *HES1* compared to the other NPC clusters. *HES1*, a transcription factor downstream of Notch signaling, is important for further renal cell differentiation. Compared to the other NPC clusters, NPCc showed higher expression of *CRABP2*, which is related to retinoic acid (RA) signaling [17]. NPCd exhibited low *OSR1*, *CITED1*, and *MEOX1* expression and increased levels of *LEF1*, a known indicator of NPC induction towards differentiation. Compared to the other NPC subtypes, NPCd was also marked by a larger fraction of cells in G2/M-phase (Fig 8A-B) and a higher expression of proliferation markers (Fig 8C), which indicated faster proliferation. We will discuss the relationship between the various NPC clusters in more detail below (see Heterogeneity in the nephrogenic niche).

Nephrogenesis continues with the creation of PTA cells, which in turn develop into the RV and CSB cells. In our data, PTA cells were identified based on high expression of *LHX1* (Fig 3), *JAG1*, *WNT4*, and *CCND1*. Because RV and CSB are mainly distinguishable by morphology, cells belonging to these two structures were grouped in our analysis (RVCSB). RVCSBs were marked by the same genes as PTA cells but they appeared to proliferate less (Fig 8). Furthermore, they expressed markers reflecting more advanced regional patterning, which allowed us to discriminate between two subtypes (a and b). RVCSBa had a higher expression of genes that were recently associated with the distal RV (*SFRP2*, *DLL1*, *LHX1*), whereas RVCSBb expressed genes that indicate the proximal RV (*CDH6*, *FOXC2*, *MAFB*, *CLDN1*, *WT1*).

The next step in development is the formation of the SSB. In our dataset, this structure was represented by three clusters, named according to the part of tubule and glomerular epithelium they are known to give rise to SSBpr, proximal tubule; SSB medial/distal (SSBm/d); SSB podocyte precursor cell (SSBpod), podocytes. SSBpr were identified in our data by markers of the early proximal tubule (ErPrT) (such as *NF1A*, *CDH6*, *AMN*), as well as low expression of SLC3A1, *LRP2*, and *SLC13A1*, which are known to be found in more mature proximal tubule cells. Therefore, SSBpr were likely precursors of the ErPrT cells, which expressed higher levels of early proximal markers together with CLDN2 (Fig 3), *ANPEP*, and *SLC34A1*. Another cluster accounted for the precursor cells of the loop of Henle (LOH) and the distal tubule in the SSB (SSBm/d). This cluster could be identified by the presence of *IRX1*, *IRX2*, *SIM2*, *SOX9*, *POU3F3*, and *HNF1B*, together with low expression of *PAPPA2* and *MAL* and the absence of *CDH6* and *HNF1A*. Cell types that are known to develop from the SSBm/d were found together in one cluster (DTLH). This cluster showed high expression of the distal markers *MAL*, *CLCN5*, *SLC12A3*, and *POU3F3*, which are specific to the distal tubule, as well as *SLC12A1*, *PAPPA2*, and *UMOD*, which are found in the LOH. Finally, cells that likely gave rise to podocytes, SSBpod, clustered separately. These cells showed high expression of *MAFB* and *FOXC2*, both transcription factors necessary for the development of podocyte identity, and low lev-

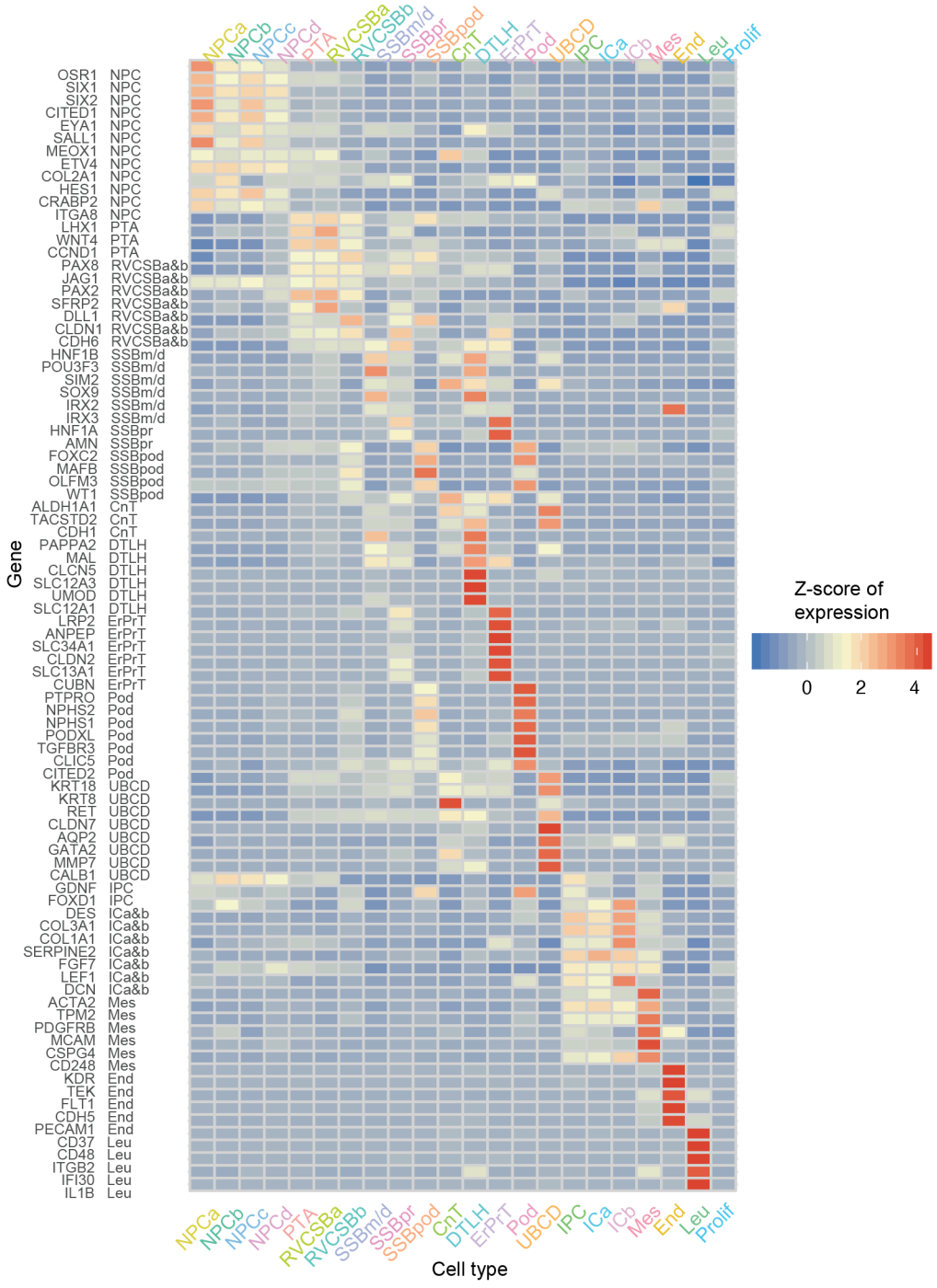


Figure 7. Known markers elucidated the cell types corresponding to each cluster. Heatmap of literature set gene expression in the 22 identified cell types. Expression was FT transformed, averaged over all cells in a cluster and standardized gene-wise.

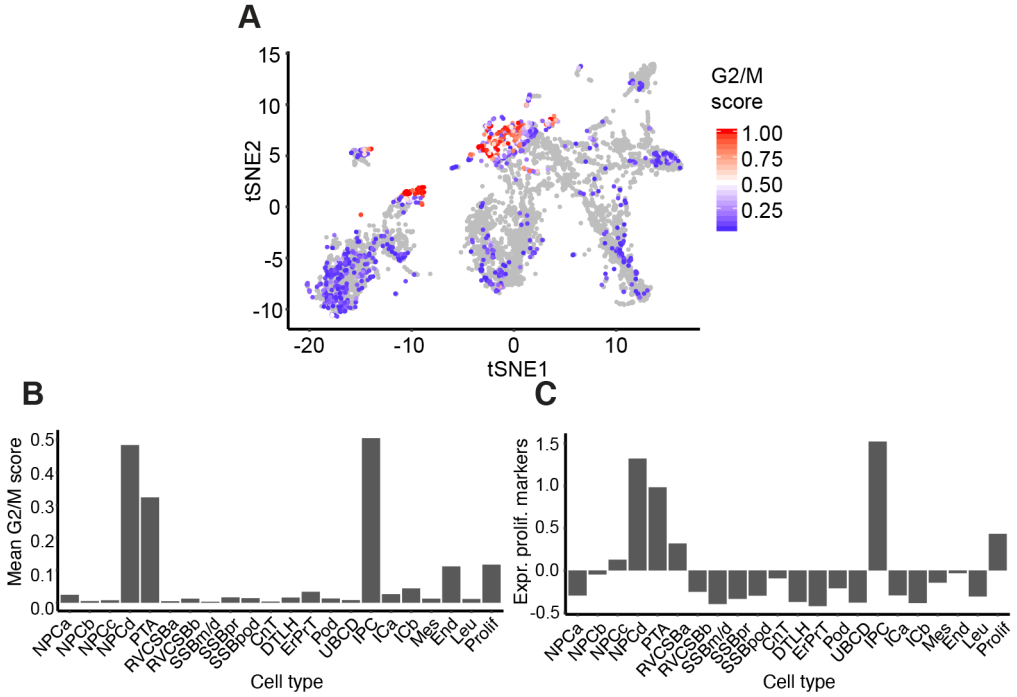


Figure 8. Proliferation states varied per cell type. (A) tSNE map of all cells with color indicating the G2/M score (calculated by the Cyclone tool [15]. This score reflects the likelihood that a cell is in G2/M phase. (B) G2/M scores from panel A averaged over the cells in each cell type. (C) Mean expression of proliferation markers [16] (Z-scores) per cell type.

els of the mature podocytes markers *CLIC5* (Fig 3), *PTPRO*, *NPHS1*, and *NPHS2*. This cluster also showed the highest expression of *OLFM3*, previously identified as a specific marker of podocyte precursors residing in the visceral part of the proximal segment of the SSB. In contrast to SSBpod, podocytes showed higher expression of mature podocyte markers and lower levels of *MAFB*. Differences between SSBpod and podocytes will be studied in more detail below (see Podocyte development). Because of the high similarity in gene expression between SSB and capillary loop stage, we could not exclude that the SSB clusters also contained cells from the capillary loop stage.

Cells of the connecting tubule (CnT), which connects the distal tubule to the collecting duct, could also be identified in the data. They shared markers with the collecting duct (such as *ALDH1A1*, *TACSTD2*, and *CDH1*), distal tubule (*SOX9*, *POU3F3*), and UB (*RET*, *KRT8*, *KRT18*, *MMP7*). Cells of the ureteric bud/collecting duct (UBCD) were strongly marked by well-known genes like *AQP2* (Fig 3), *CALB1*, *KRT8*, *KRT18*, *RET*, and *GATA2*, found in the collecting duct as well as the stalk and tip of the UB.

The developing nephrons are surrounded by interstitial tissue, a separate lineage that originates in interstitial progenitor cells (IPCs). We identified IPCs by coexpression of *FOXD1*

and *GDNF*. These cells also expressed lower levels of markers known to be found in more mature cells like *PDGFRA* for ICs or *PDGFRB* (Fig 3) and ACTA2 for mesangial cells. We identified two subtypes of IC (a and b), which were similar in their marker gene profile. Compared to IPCs, they lacked *FOXD1* and expressed less (ICa) or no (ICb) GDNF. ICa showed high levels of *FGF7*, which has been localized to the renal fibroblasts or stroma surrounding the ureter and the collecting system. ICa also showed high levels of *TPM2* and *ACTA2*, markers of smooth muscle-like cells. ICb, on the other hand, expressed genes like DCN (Fig 3), *DES*, *SERPINE2*, and *COL3A1*, which are known to mark cortical stromal cells. Endothelial cells were identified by markers such as *KDR* and *TEK*, whereas leukocytes showed many specifically expressed genes, such as *CD37* or *CD48*. Finally, one cluster of cells (proliferating cells) had a higher expression of proliferation markers compared to most other cell types (Fig 8C) but lacked discernible cell type markers.

2.2.2 Developmental flow

The literature-based analysis of the found clusters seemed to suggest that cells cluster by developmental progression (e.g., NPCs versus PTA cells), as well as location (e.g., RVCSBa, distal, versus RVCSBb, proximal) Because the interpretation of clusters is sometimes based on genes that are expressed in multiple developmental stages, we wanted to retrieve the developmental flow with an independent method We used *Monocle 2* [18] to learn a graph that represents the developmental hierarchy of the cell types from the PTA on (Fig 9A) Subsequently, cells were placed on a pseudotime scale rooted in the PTA (Fig 9B-C) This analysis showed that PTA cells were followed by RVCSBa and RVCSBb, the SSB clusters, and finally the clusters identified as more mature types (DTLH, ErPrT, podocytes) Therefore, the clustering was strongly driven by developmental progression RVCSBa cells were distributed over a fairly broad period of pseudotime and already occurred before branch point 1, which separates proximal from distal cell fates (Fig 9A) This might indicate that some of these cells preceded the RVCSBb, whereas others were primed to develop into distal fates RVCSBb cells, however, only appeared after branch point 1, which confirmed that they were likely progenitors of proximal cell fates On three separate branches, SSBm/d preceded DTLH, SSBpr preceded ErPrT, and SSBpod preceded podocytes, which confirmed the identity of the SSB clusters The temporal relationship of the NPC subtypes will be discussed in detail below (see Heterogeneity in the nephrogenic niche)

2.2.3 Comparison with existing single-cell transcriptomics data

To further confirm the interpretation of the cell clusters, we wanted to compare our data with an existing single-cell transcriptomics study of a w17 fetal kidney by Lindström et al. [20]. To that end, we first corrected for batch effects, using a method based on matching mutual nearest neighbors in the two datasets [21]. After correction, the two datasets showed a large degree of overlap (Fig 10, Fig 11). This allowed us to use the cell types found by Lind-

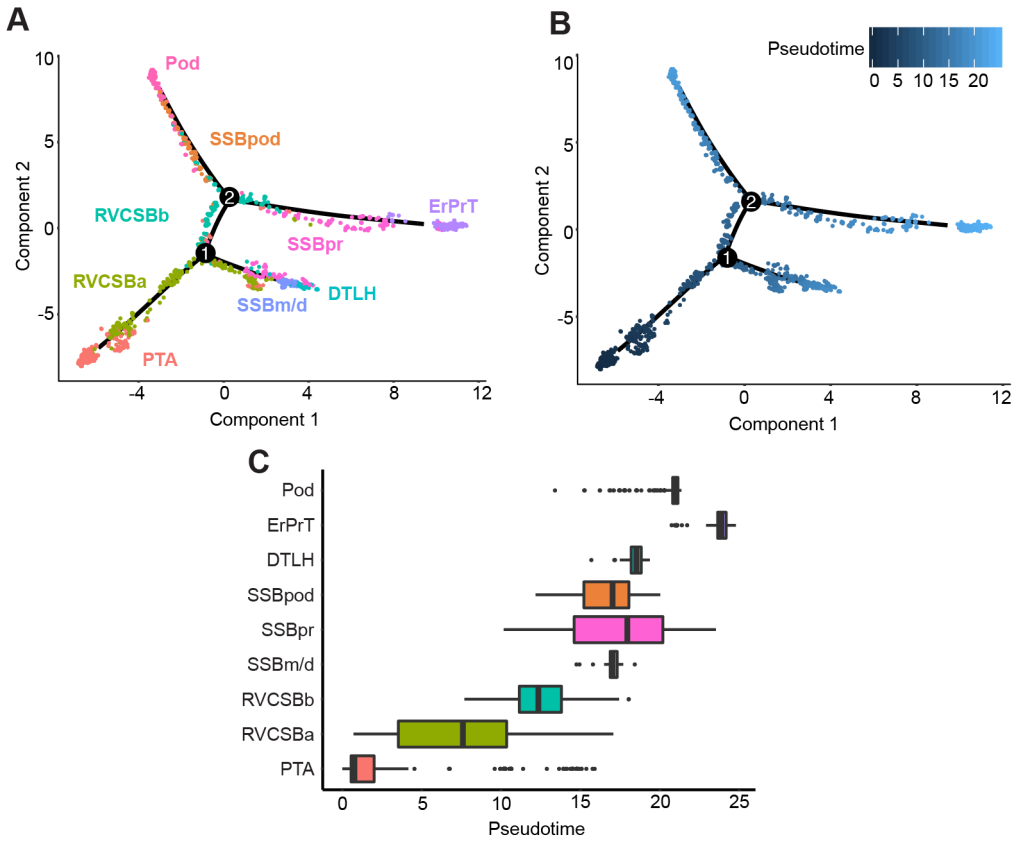


Figure 9. Pseudotime analysis clarified the developmental relationship of the cell clusters. (A) Two-dimensional embedding (with the *DDRTree* algorithm [19] of all w16 kidney cells, calculated by *Monocle 2* [18]. The graph learned by the algorithm is shown as a black line. Colors and labels indicate cell types. (B) Same embedding as in panel A. Color indicates pseudotime calculated by *Monocle 2*. (C) Box plots of cell type distribution over pseudotime.

ström et al. [20] to classify the cell clusters found here, using a knn approach (see Materials and methods). NPCa-c were also classified as NPC by Lindström et al. [20], whereas NPCb were considered *primed NPC*, which supports the notion that NPCb were primed to differentiate. The NPCd cluster was classified as *proliferating cells* This classification is in agreement with our observation that NPCd seemed to proliferate more than other NPC subtypes (Fig 8). Because NPCd cells expressed low levels of NPC markers (such as *SIX2* and *CITED1*), these cells were likely in a transition state between NPCs and PTA cells. Although the majority of PTA cells identified here were considered *PTA/RV I* by Lindström et al. [20], RVCSBa cells were spread over multiple cell types. This spread was likely due to the fact that transitory cell types are transcriptionally similar, and their clustering is therefore less robust. Nevertheless, the *PTA/RV II* cluster received most of the RVCSBa cells. RVCSBb cells were called *podocyte*

precursors in the Lindström et al. [20], whereas SSBpod as well as podocytes were classified as *podocytes*. In our dataset, RVCSBb directly preceded SSBpod (Fig 9A), so they could indeed be considered podocyte progenitors. Below, we will show that SSBpod did form a cell state separate from podocytes and should not be grouped with them (see Podocyte development). In agreement with our analysis, the majority of SSBpr were classified as *proximal precursor* or *proximal tubule*, and all ErPrT were considered *proximal tubule* by Lindström et al. [20]. CnT and DTLH were both classified as *distal/LOH precursor*. The fact that two cell types in the study by Lindström et al. [20] (*podocytes* and *distal/LOH precursor*) were split in multiple subclusters in our study likely reflects differences in sample preparation. Whereas Lindström et al. [20] preferentially released single cells from the nephrogenic niche, here, the whole kidney was used. Consequently, the Lindström et al. [20] dataset has a finer resolution of NPCs and early, proliferating cell types, whereas our dataset allowed us to resolve more mature cell types. The two datasets therefore complement each other.

2.2.4 Marker identification

To confirm the inferred cell types and also identify novel markers, we pursued two complementary strategies. First, we determined a set of marker genes based on their usefulness as classifiers for individual cell types: for each gene, the performance of a binary classifier was evaluated by the area under the ROC (AUROC) and combined with expression level filtering (see Materials and methods). This resulted in 88 marker genes (See Fig 12, marker set, Hochane et al. [1] S3 Table). Only 11 of these markers overlapped with the 89 genes in the literature set (See Fig 13C). To our knowledge, many of the remaining markers had not been associated with kidney development in previous studies. As an independent approach, we used the KeyGenes algorithm [22] to identify classifier genes among the 500 most HVGs, using two-thirds of all cells as a training set. Based on the classifier genes determined by KeyGenes, we next predicted the cell types of the remaining one-third of the cells (test set). Cell types could be predicted with an average certainty (id score) of 0.59; 24% of the cells in the test set obtained an id score higher than 0.8. Of the 95 classifier genes (See Fig 12A, KeyGenes set, Hochane et al. [1] S3 Table), 24 were the same as in the marker set, and 14 were common with the literature set (See Fig 13B).

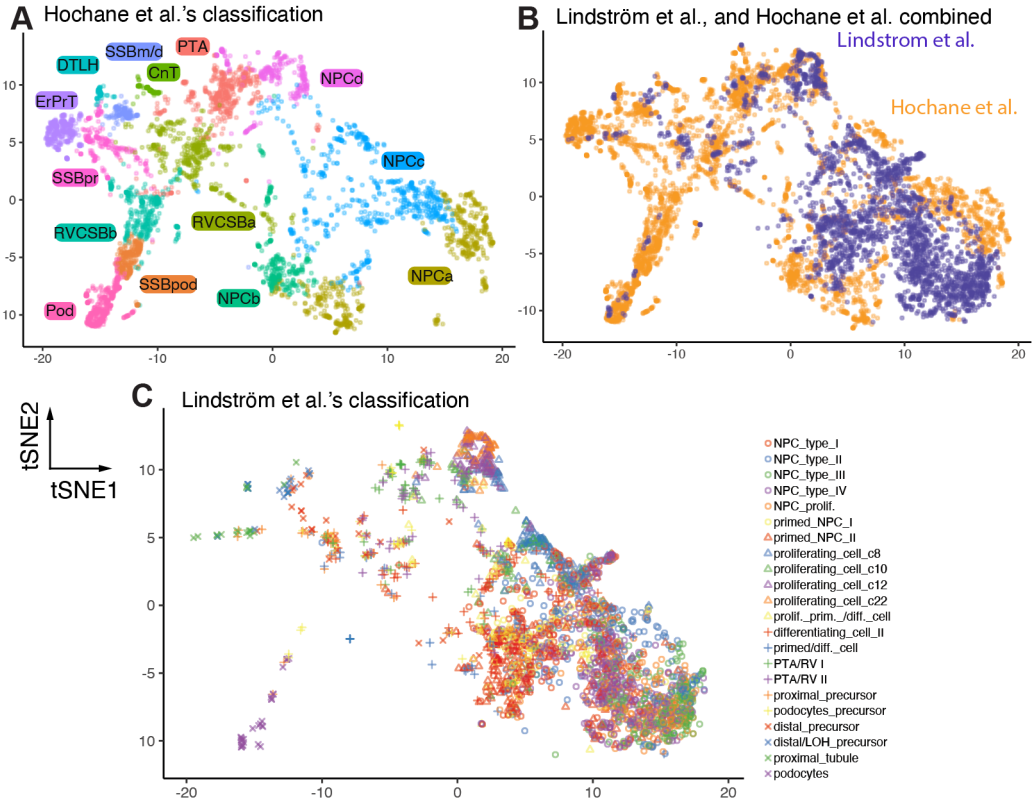


Figure 10. Comparison with an existing single-cell transcriptomics dataset showed congruent expression profiles. Two-dimensional tSNE maps comparing the data presented here with the data from Lindström et al. [20] both restricted to the nephrogenic niche by their own classification. The map was calculated using both data sets after batch correction [21]. (A) Only cells measured in this study are shown. Color and labels indicate the classification developed in this study. (B) Same tSNE map as above. Color indicates the data set. (C) Same tSNE map as (B). Only cells measured by Lindström et al. are shown. Color and labels indicate the classification by Lindström et al. See Fig 11.

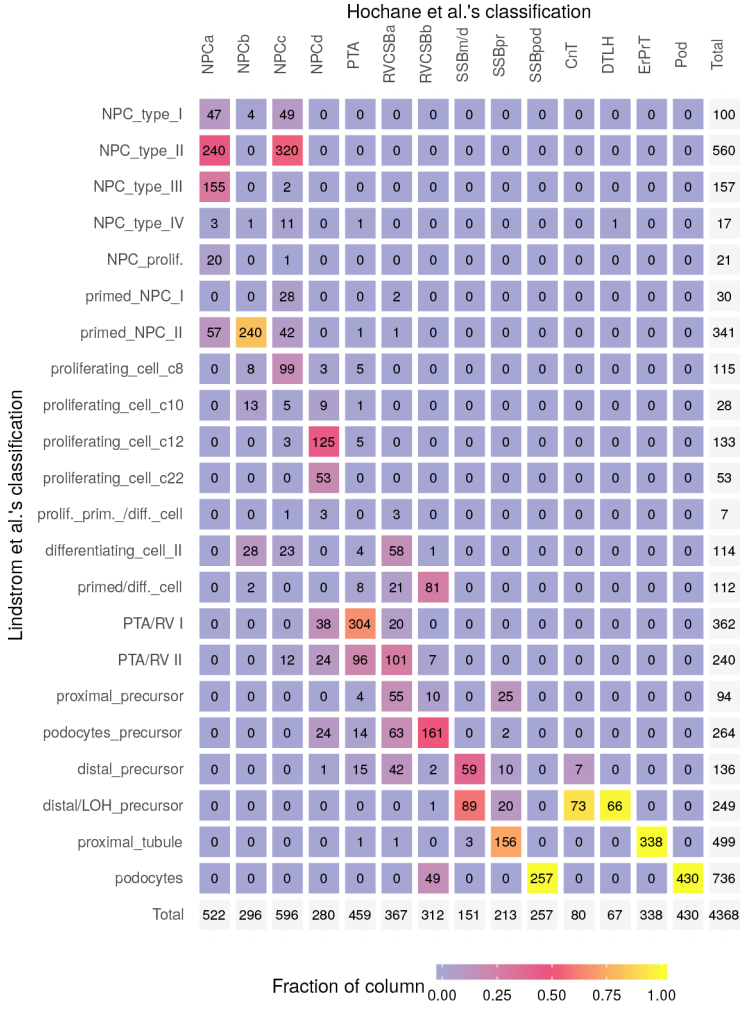


Figure 11. Comparison with an existing single-cell transcriptomics dataset showed differences in cell type distribution. Confusion matrix relating the cells measured in this study to the classification by Lindström et al. [20]. After batch correction, cells measured here were mapped on the cells in the Lindström et al. data set using a nearest neighbors-based approach (see Materials and methods). See Fig 10.

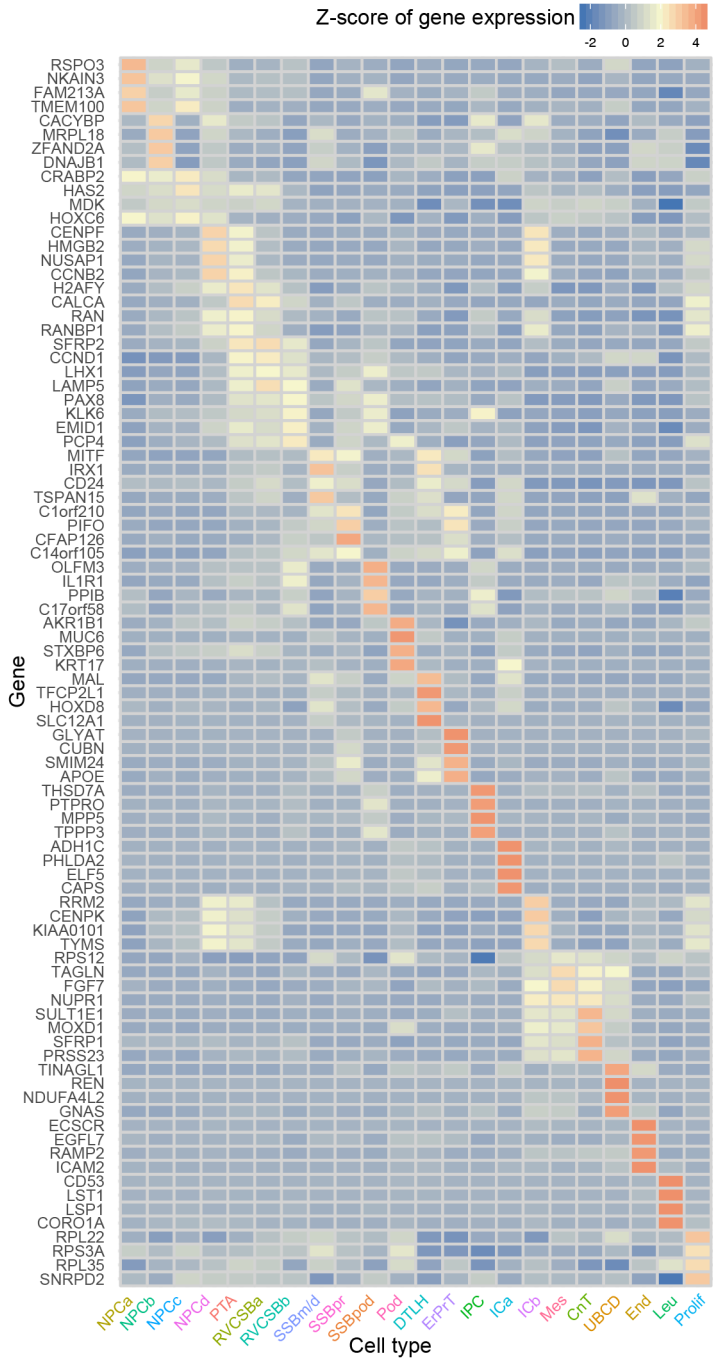


Figure 12. An ROC-based method identified novel marker genes. Expression heat map of the 88 genes identified by a method that evaluates the ROC for each gene (marker set, Hochane et al. [1] S3 Table). Expression was FT transformed, averaged over all cells in a cluster, and standardized gene-wise. See Fig 13.

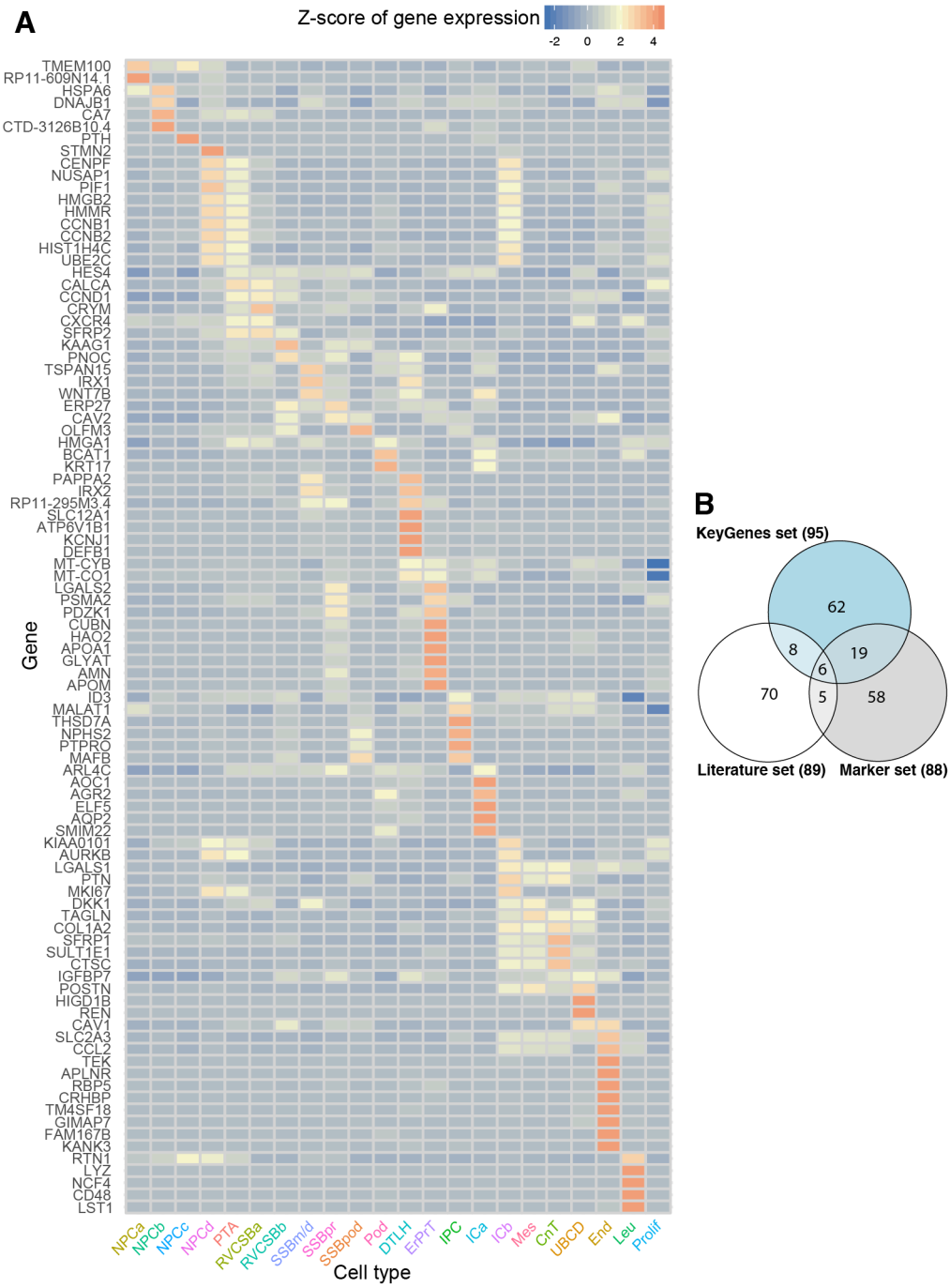


Figure 13. The KeyGenes method identified novel marker genes. (A) Expression heat map of the 95 genes identified by the KeyGenes algorithm (KeyGenes set, Hochane et al. [1] S3 Table). Expression was FT transformed, averaged over all cells in a cluster, and standardized gene-wise. (B) Euler diagram of the literature set, marker set, and KeyGenes set (Hochane et al. [1] S3 Table). See Fig 12.

Because the interpretation of the found cell clusters was largely based on markers identified in mouse development, we were wondering whether the new markers identified here were informative for the classification of cell types in the mouse kidney. Using a scRNA-seq measurement of cells from a whole P1 mouse kidney [23], we plotted the expression of the newly identified marker genes in single cells (Hochane et al. [1] S7 Fig). In many cases, markers that were found to label a particular cell type in the human fetal kidney were coexpressed in the same subset of mouse cells. A few markers, however, were either ubiquitously expressed or almost completely absent. This might be due to interspecies differences.

2.2.5 Comparison of different developmental ages

By establishing the identity of cell clusters at w16, we obtained a snapshot of cell type diversity in the fetal kidney. To explore whether the identified expression patterns change dynamically throughout development, we analyzed four additional samples from different developmental ages (w9, w11, w13, and w18), which together contained 11,359 usable cells. Using, again, batch correction based on mutual nearest neighbors [21], we visualized all samples in a common tSNE map (Fig 15, Fig 14). Overall, gene expression in the different samples was largely overlapping for the majority of cell types. For example, proximal tubules cells (ErPrT) appeared at the same positions in the tSNE map in all samples (Fig 15B). The position of podocytes shifted systematically across different ages, which corresponds to a continuing change in expression pattern (Fig 15C). This observation might suggest that podocytes further matured in terms of their expression pattern after being specified.

Differential expression analysis of podocytes of different ages revealed 109 differentially expressed genes (fold change > 2 in any comparison, FDR < 0.05, Hochane et al. [1] S4 Table). Functional annotation analysis of these genes showed significant enrichment of two gene ontology (GO) terms: *proteinaceous extracellular matrix* (adjusted p-value $= 1.9 \cdot 10^{-3}$, including *SPON2*, *BGN*, *COL1A2*, and *CTGF*) and *extracellular exosomes* (adjusted p-value $= 1.4 \cdot 10^{-3}$, including *NPNT*, *S100A10*, *ANXA1*, and *EPCAM*). Some of the differentially expressed genes have been shown to be important for kidney development. For example, *NPNT* and *DCN* showed increasing expression from w11 to w18. Knockout of the extracellular matrix protein NPNT in mice decreases the invasion of the UB and causes agenesis or hypoplasia [24]. *NPNT* was further shown to be expressed in the glomerular basement membrane and to be necessary for podocyte adhesion in mice [25]. Ablation of this gene in mice causes podocyte effacement. As in the case of *NPNT*, *DCN* has been reported to be part of the glomerular basement membrane proteins [26]. This gene appeared strongly up-regulated in podocytes between w11 and w13 or w18 (fold changes of 3.25 and 4.6, respectively). The increase of *NPNT* and *DCN* expression over time in our data set could reflect an increase in adhesion between podocytes and glomerular basement membrane. Podocytes further showed significant differential expressions of genes related to stress, like *HSPA1A* and *HSPA1B* or *NFKB* genes (*NFKB2*, *NFKBIA*, and *REL*), with the highest levels at w18. This might suggest that dissociation-related stress increases with age for podocytes, maybe related to stronger

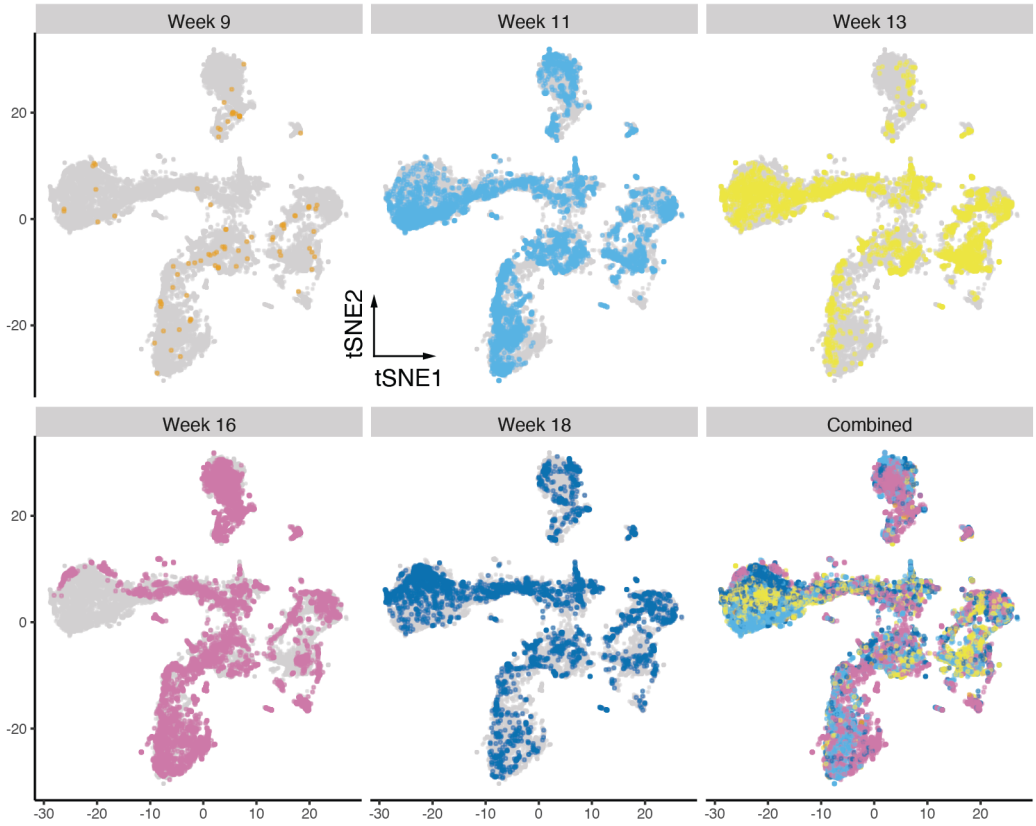


Figure 14. Samples of different developmental ages had a similar cell type diversity. tSNE map calculated for all five samples (w9, w11, w13, w16, w18) combined after batch correction [21]. Developmental age is indicated by color.

adhesion of the cells, or that stress-related genes have another, physiological role in development.

We would like to emphasize that the observed gene expression changes with age should be considered with caution because they might be related to the differences in genotype between the samples. A much larger number of samples would be necessary to rule out such interindividual differences as a cause.

Having established the identity of the cell clusters, we next wanted to demonstrate how the dataset can be used to explore different aspects of kidney development. We specifically focused on the nephrogenic niche, which showed pronounced heterogeneity, and the development of podocytes, which progressed via a distinct, intermediate cell state (SSBpod).

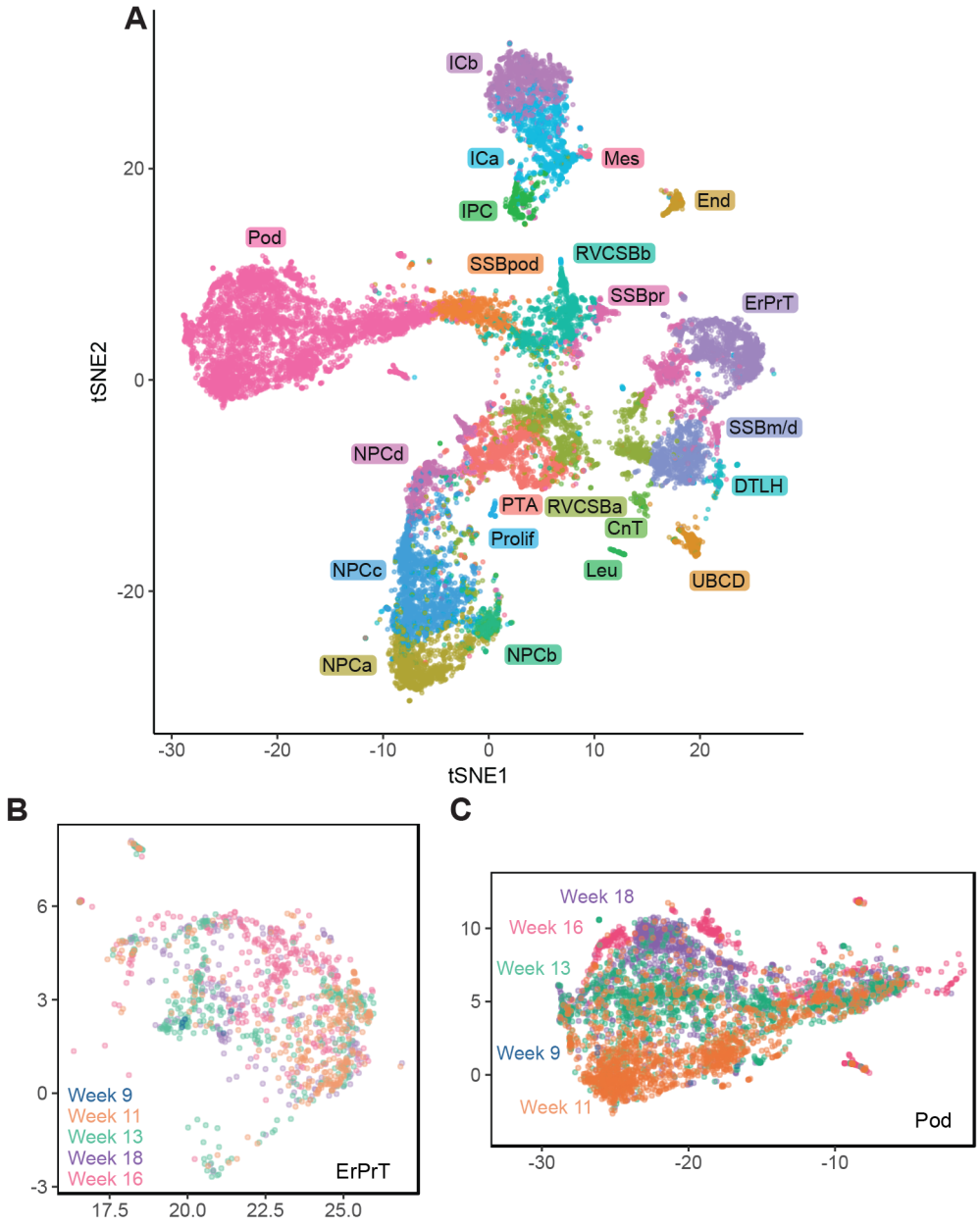


Figure 15. Comparison of different developmental ages suggested continued expression changes in podocytes. (A) tSNE map combining all five samples (w9, w11, w13, w16, w18). Samples were corrected for batch effects by matching mutual nearest neighbors [21]. Cells in the w9, w11, w13, and w18 samples were classified by comparing to the w16 sample using a knn-based approach (see Materials and methods). (B) tSNE map of all ages restricted to ErPrT. Labels and colors indicate ages. Six outlier cells were omitted from this plot to improve visualization. (C) tSNE map of all ages restricted to podocytes. Labels and colors indicate ages.

2.2.6 Heterogeneity in the nephrogenic niche

The formation of the nephron epithelium starts with the NPCs that differentiate and form the PTA, RV, and CSB. Studies in the mouse suggest that cells in the NPC compartment are not biased towards a particular lineage and patterning is first detectable in the PTA [27]. Nevertheless, the w16 scRNA-seq indicated the presence of several nephron progenitor subpopulations, NPCa-d. To clarify the temporal relationship of these clusters, we employed *Monocle 2* again to arrange them together with the PTA cells on a pseudotime scale (Fig 16A). NPCa clearly preceded NPCb and c, which seemed to appear around the same pseudotime. NPCd cells followed NPCb and c and preceded PTA. This analysis suggested that NPCa are the *bona fide* NPCs and give rise to NPCb and c. NPCd, which were likely more proliferative than the other NPCs (Fig 8), seemed to be an intermediate state between (slowly cycling) NPCa-c and the PTA.

To localize the NPC clusters in the tissue, we made use of the fact that they expressed various levels of *CITED1* and *SIX2* (Fig 7): although NPCa and NPCc exhibited roughly similar levels of these markers, NPCb and NPCd had lost *CITED1* almost completely, while retaining some *SIX2* expression. In an immunostaining of a w15 kidney, CITED1 and SIX2 appeared overlapping in a subset of cells (Fig 16C). Quantification of the fluorescence signal (see Materials and methods) revealed clear differences between their expression patterns. Although SIX2 expression was approximately constant throughout the CM, CITED1 expression decreased, relative to SIX2, with increasing (radial) distance from the UB (Fig 16D-E). A marked drop of CITED1 was visible between 10 and 20 μm from the UB, which approximately corresponds to the first layer of cells. To exclude that the observed difference between SIX2 and CITED1 expression was due to the different fluorophores on the secondary antibodies, we repeated the experiment with swapped fluorophores. This measurement produced a very similar expression gradient (Fig 16F). To exclude that the observed effect was influenced by PTA found in the CM towards the stalk of the UB, the analysis was repeated, taking only the 20% of CM cells closest to the edge of the cortex into account. A similar expression gradient was observed (Fig 16H). This result implies the existence of a CITED1 low/SIX2 high subpopulation of cells, which are not in contact with the UB. Secondly, we observed that CITED1 decreased relative to SIX2 towards the interface with the PTA and the stalk of the UB (Fig 16D-E). A similar observation was made when the experiment was repeated with swapped fluorophores (Fig 16G). Taken together, these results suggested that NPCa and NPCc were located closer to the surface of the UB and closer to the tip of the UB compared to the other NPC subtypes. Additionally, we also observed differences in subcellular localization of CITED1 protein within the CITED1 high compartment. Although for the majority of cells CITED1 was found in the cytoplasm, in several cells it was concentrated in the nucleus (right inset in Fig 16C). In contrast, SIX2 was always found restricted to the nucleus (left inset in Fig 16C). This observation might indicate that CITED1 was only active in a small population of cells, which would constitute another layer of cell-cell heterogeneity.

In addition to the observed heterogeneity in *CITED1* and *SIX2*, differences between the

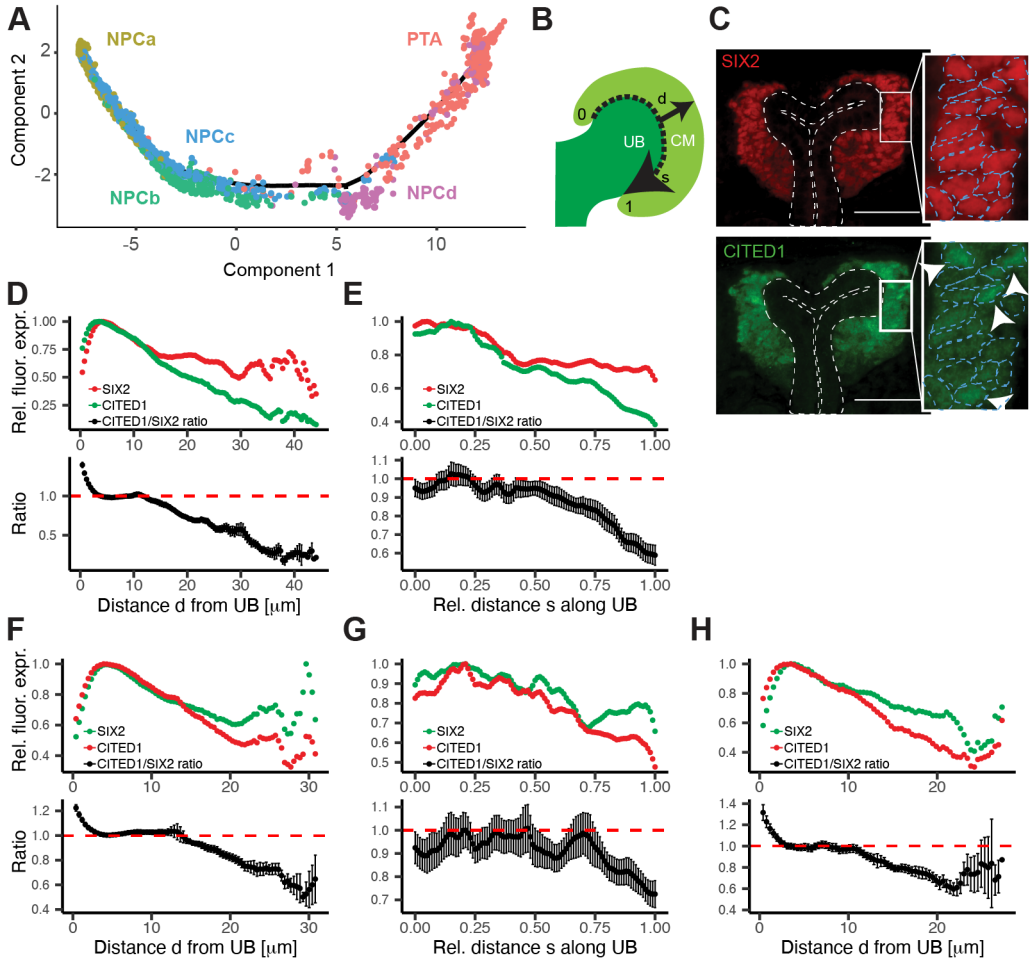


Figure 16. The nephrogenic niche exhibited a complex spatial organization. (A) Pseudotime analysis of the nephrogenic niche and the PTA. Two-dimensional *DDRTree* [19] embedding and the learned graph (shown as a black line) were calculated with *Monocle 2* [18]. Labels and colors indicate cell types. (B) Schematic sketch of the CM indicating the distance d from the UB to the edge of the CM (solid arrow) and the relative distance s along the UB (dashed arrow), in which 0 and 1 represent the top and bottom of the CM, respectively. (C) Representative image of SIX2 and CITED1 immunostaining in a w15 human fetal kidney. Dashed lines in the insets indicate the outline of the nuclei, based on DAPI signal. Arrows in the inset point to cells in which CITED1 is concentrated in the nucleus. Scale bar = 50μm. (D and E) Quantification of SIX2 and CITED1 immunostaining with respect to the distance d from UB or distance s along the UB; see panel A. Error bars indicate the SEM calculated over all evaluated profiles ($n = 24$). (F and G) Same as D and E, but the fluorophores on the secondary antibodies were swapped. ($n = 19$). (H) Quantification of SIX2 and CITED1 immunostaining with respect to the distance d from UB in which only cells with a relative distance s (along the UB) < 0.2 were taken into account. Error bars indicate the SEM calculated over all evaluated profiles ($n = 19$).

NPC clusters could also be gleaned from the set of novel markers (marker set, Hochane et al. [1] S3 Table).

TMEM100 and *RSPO3* specifically marked NPCa. *RSPO3* is an activator of the canonical WNT signaling pathway [28], suggesting a role of WNT either in NPC self-renewal or UB branching morphogenesis. Notably, all markers of NPCb (*CACYBP*, *MRPL18*, *ZFAND2A*, *DNAJB1*) were related to the stress response in some form. The markers of NPCc (*CRABP2*, *HAS2*, *MDK*, *HOXC6*), which were also expressed in the other NPC types, are all either targets of RA or binding it [17, 29, 30]. *MDK* has been shown to be expressed in the CM of the developing rat kidney, and its neutralization reduced the number of formed nephrons in vitro [31, 32]. Finally, the NPCd markers *CENPF*, *HMGB2*, *CCNB2*, and *NUSAP1* all have a role in cell cycle regulation or proliferation [16]. *HMGB2* was recently implied in the activation of quiescent adult neural stem cells [33].

The observation that markers of NPCb were related to the stress response seemed to suggest that this cluster was created as an artifact of cell dissociation [34], despite our best efforts to remove stressed cells (see Materials and methods). On the other hand, the vast majority of NPCb cells were classified as *primed NPC II* in the Lindström et al. [20] dataset (Fig 10, Fig 11). The fact that NPCb cells were only detected in the w16 and w18 kidneys is consistent with single-cell dissociation becoming increasingly difficult with fetal age, or alternatively, with a progenitor cell aging phenomenon. To explore the differences between NPCb and the other NPC clusters further, we immunostained HSPA1A and NR4A1, both known stress-response genes, in w15 kidney sections (Fig 17). *HSPA1A* was identified as a marker of NPCb (Fig 18A, Hochane et al. [1] S3 Table), whereas *NR4A1* was expressed in multiple NPC clusters but highest in NPCb (Fig 18A). Furthermore, *NR4A1* was also identified in the study by Adam et al. [23] to be up-regulated in response to elevated temperatures during enzymatic dissociation. HSPA1A and NR4A1 were both observable in the nephrogenic niche at the level of the stalk of the UB and at the transition to the PTA or RV. Additionally, we studied the expression of *EGR1*, another stress-related gene that marked NPCb, with smFISH (Fig 18B). *EGR1* was mainly found toward the stalk of the UB and in a few cells around the tip of the UB, whereas *SIX2* and *CITED1* transcripts were visible throughout the CM (Fig 19A). Because results obtained in fixed tissue sections are not confounded by dissociation-related artifacts, these immunostainings and smFISH measurements supported the existence of NPCb cells in the fetal kidney.

The fourth NPC cluster, NPCd, was clearly distinguished by proliferation markers (Fig 8). To locate NPCd in the tissue, we immunostained the cell cycle regulator and NPCd marker *CKS2* (Fig 18C). *CKS2* signal could be observed around the stalk of the UB and in RV (Fig 17). This result supported the interpretation that NPCd were a proliferating transitory state between NPCa-c and PTA.

Given the crucial role of the nephrogenic niche in the development of nephrons, it is likely that misexpression or mutation of genes that are specifically expressed in NPC affect kidney function. Mining a database of genome-wide association studies (GWAS) revealed that genes

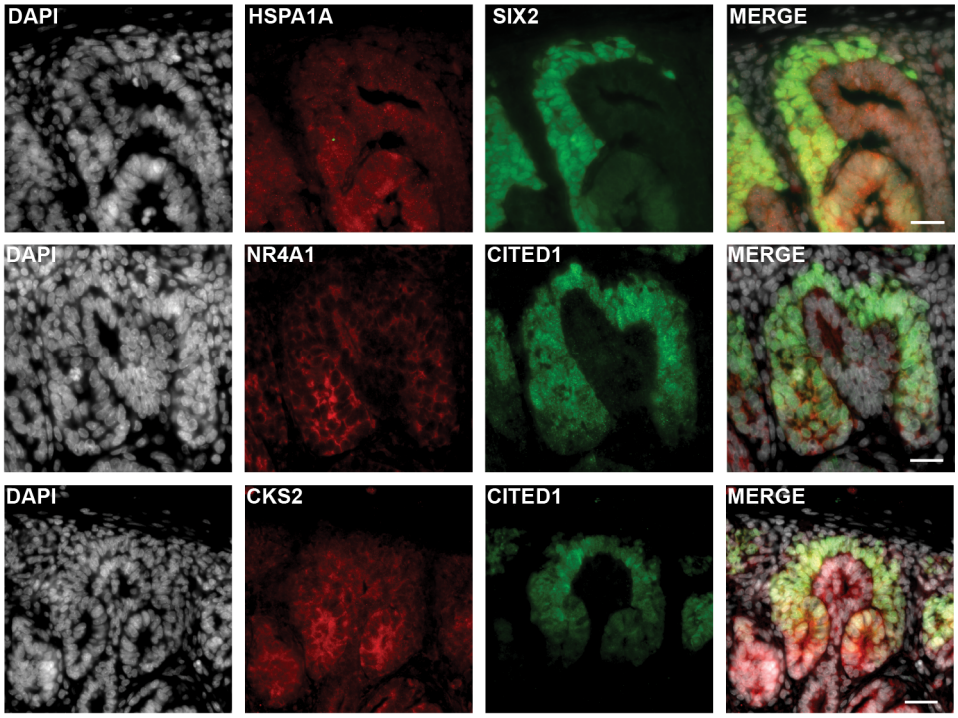


Figure 17. Expression variability in the nephrogenic niche. (Representative image of HSPA1A, NR4A1, and CKS2 immunostaining in a w15 human fetal kidney. Scale bar = 20μm.

that were differentially expressed in NPCs were significantly enriched for association with kidney disease (p-value = $1.7 \cdot 10^{-3}$, one-sided Fisher's exact test). No enrichment was found for lung diseases (p-value = 0.21), one-sided Fisher's exact test) (see Materials and methods, Fig 20A, Hochane et al. [1] S4 Table). Unsurprisingly, several of the disease-associated genes are known regulators of kidney development, such as *SALL1*, *SOX11*, and *HAS2* [35, 36, 37, 38]. The other identified genes had not been previously associated with kidney development. For example, *DDX1*, which was differentially expressed in NPCs as well as SSBpod, is an RNA helicase that promotes microRNA maturation [39]. *UNCX*, which was broadly expressed in all NPC clusters, is a homeobox transcription factor involved in somitogenesis and neurogenesis [40] and has also been found to be up-regulated in the induced mouse nephrogenic mesenchyme in culture [41]. It was recently associated with renal-function-related traits [42] as well as glomerular filtration rate [43, 44, 45]. In our data, the expression profile of *UNCX* was similar to that of *CITED1* (Fig 18D). Immunostaining of UNCX confirmed the scRNA-seq results and showed expression of UNCX in the nephrogenic zone, as marked by CITED1 (Fig 19C-D). These findings suggested *UNCX* as a novel potential regulator of early nephrogenesis.

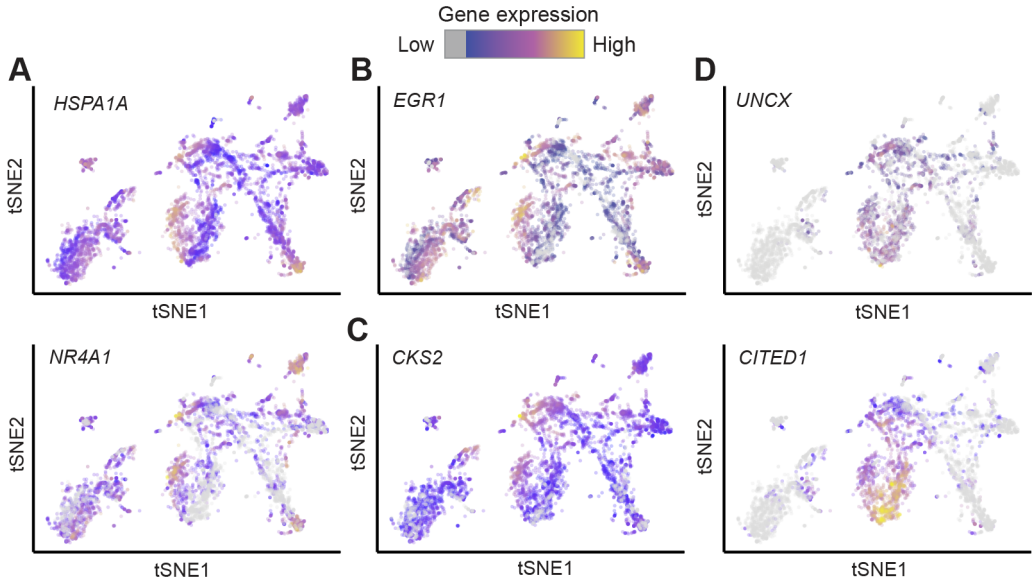


Figure 18. Markers of heterogeneity in the nephrogenic niche. (A [top], A [bottom], B, C, D [top] and D [bottom]) tSNE maps showing expression of *HSPA1A*, *NR4A1*, *EGR1*, *CKS2*, *CITED1* and *UNCX* respectively. Expression is indicated by color; expression values of 1 are plotted in gray.

2.2.7 Podocyte development

Another cell type of high relevance for kidney function is the podocyte. This cell type is critical for filtration and is implied in several forms of kidney disease [25]. podocyte (Pod)ocytes wrap around the glomerular basement membrane (capillary bed) inside Bowman's capsule (Fig 21A). Clustering (Fig 3) and pseudotime analysis (Fig 9) of the w16 kidney dataset had indicated that development into podocytes occurs via a distinct intermediate state that we dubbed SSBpod here. This cell state was likely related or even identical to previously discovered podocyte precursors [20, 46]. In the Lindström et al. [20] dataset, SSBpod and podocytes were both classified as *podocytes*, and the RVCSB were considered *podocyte precursors* (Fig 11). To show that SSBpod cells were indeed a localizable cell state distinct from podocytes, we further investigated their expression pattern (Fig 21B), focusing on known literature markers (literature set) and the marker set (Hochane et al. [1] S3 Table). Compared to RVCSB, SSBpod showed higher expression of *MAFB* and *FOXC2*, which are necessary for the determination of podocyte identity [47, 48]. On the other hand, compared to podocyte cells, they exhibited lower expression of genes typically associated with more mature podocytes, like *CLIC5*, *PODXL*, and *PTPRO*. Filtration function-related genes like *NPHS1*, *NPHS2*, and *PTPRO* were expressed at intermediate levels in SSBpod compared to RVCSBb, where they were absent, and podocytes, where they are highly expressed. A similar pattern could be

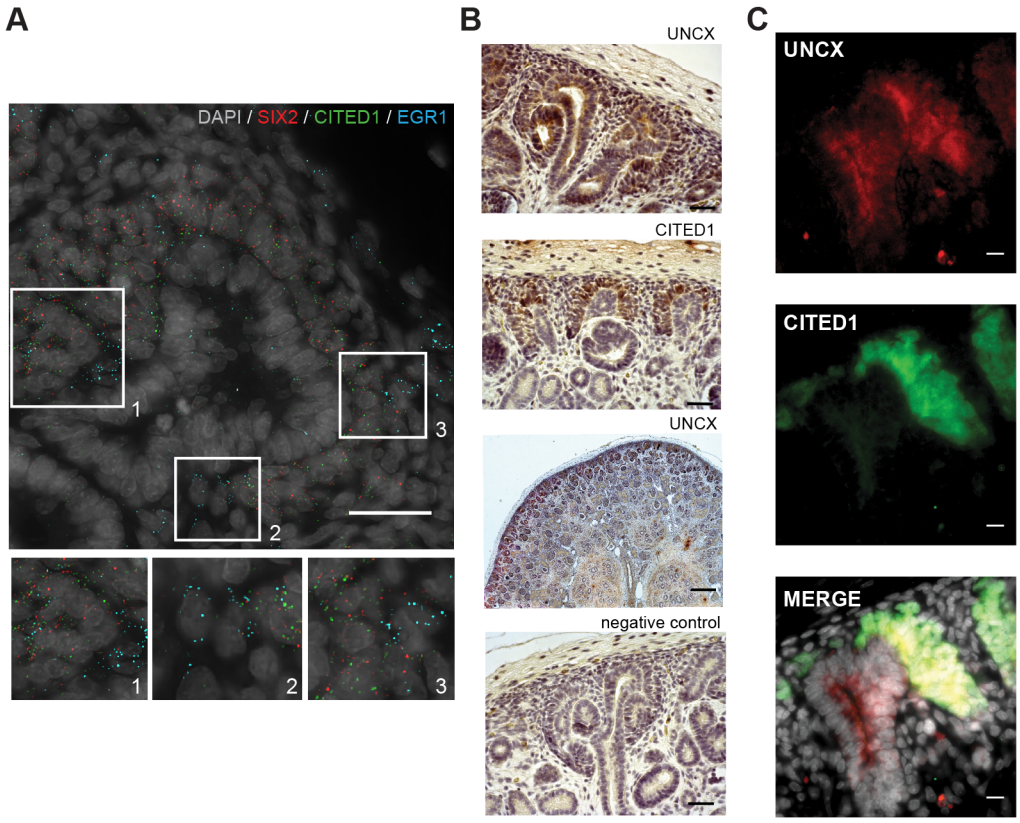


Figure 19. In situ expression of NPC markers. (A)) smFISH of *SIX2*, *CITED1*, and *EGR1*. The three insets at the bottom correspond to the three areas marked by solid boxes in the main image. Scale bar = 25μm. (B) Representative image of UNCX and CITED1 immunostaining. Arrowheads indicate the presence of immunostaining signal. Scale bar = 100μm. (C) Immunostaining of CITED1 and UNCX. Scale bar = 10μm.

observed for genes associated with podocyte polarization or structural organization as well as pedicel growth and patterning. Finally, podocytes showed the expression of genes that negatively regulate the cell cycle and support long term survival, consistent with their post-mitotic nature [49]. In contrast, SSBpod specifically expressed *ORC4*, which has a function in DNA replication. However, proliferation markers were lowly expressed in both SSBpod and podocytes (Fig 8C), which suggested low proliferative potential in both cell types. In contrast to NPCs, association with kidney disease was not significantly enriched among genes differentially expressed in SSBpod (p-value = 0.1, one-sided Fisher's exact test). One of the disease-associated genes was *OLFM3*, which has been associated with glomerular filtration rate (Fig 20B) [50]. *OLFM3*, a secreted glycoprotein, has a known function in brain and retina development [51] and has been identified as a marker for podocyte precursors in two independent studies [20, 46]. In our dataset, it was specifically expressed in SSBpod (Fig 21C)

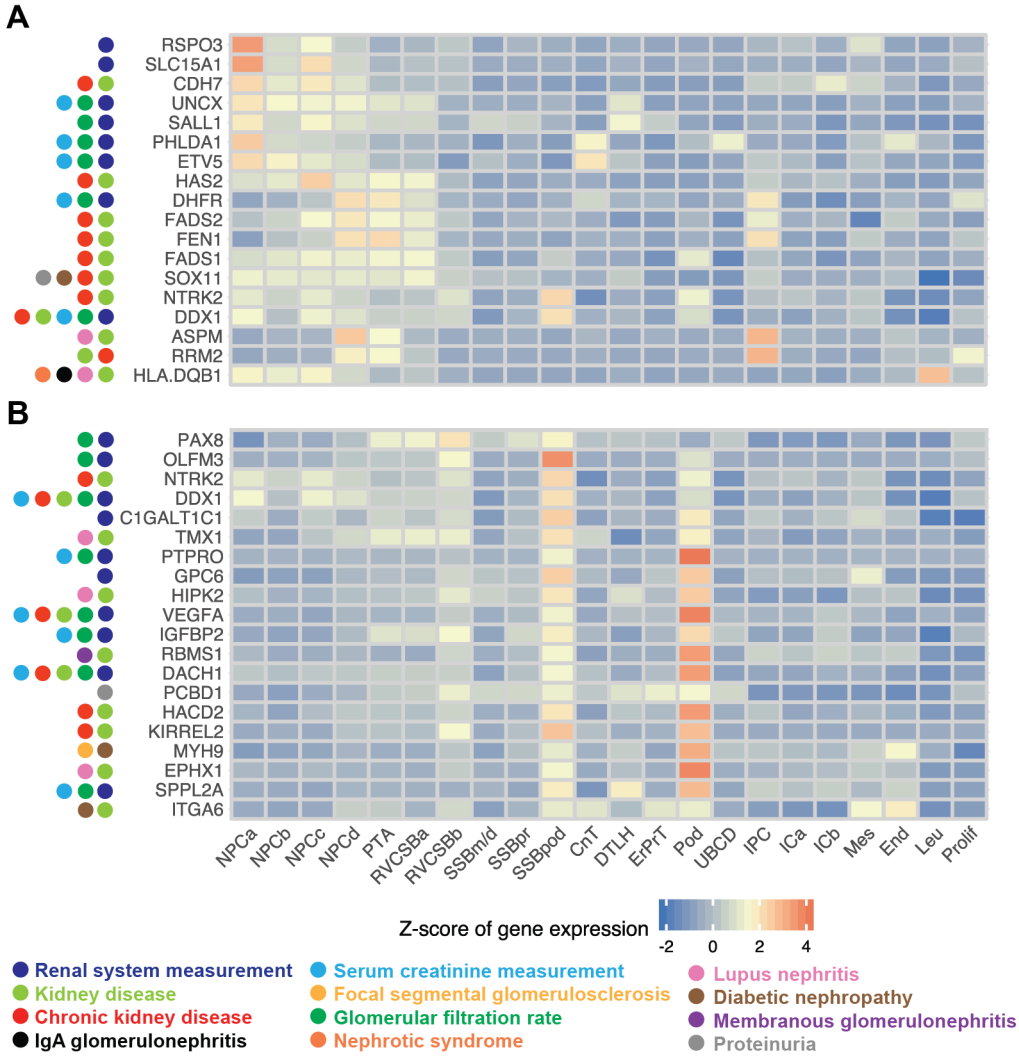


Figure 20. Disease-associated genes were specifically expressed in transient cell types. Expression of genes from GWAS traits related to kidney disease. Disease phenotypes associated with these genes are indicated by color; genes were filtered for high expression in cluster(s) of interest relative to all other cell types. Expression was FT transformed, averaged over all cells in a cluster, and standardized gene-wise. (A) Disease-associated genes expressed in early nephron progenitor states (NPC to PTA). (B) Disease-associated genes expressed in SSBpod.

and was a marker for this cell type in the marker set and KeyGenes set (Hochane et al. [1] S3 Table).

In order to localize SSBpod, podocytes and mesangial cells in situ, we immunostained w15 kidney sections with antibodies for MAFB, PODXL, and ACTA2 (Fig 21D). As expected, PODXL and MAFB were found in podocytes at the capillary loop stage and in more mature glomeruli. MAFB staining extended to the proximal segment of the SSB, which indicated that SSBpod may be part of this structure. To locate the SSBpod cells more precisely, we performed smFISH on *CLIC5* and *MAFB*, expressed both in podocytes and SSBpod (Fig 22A). We observed a subpopulation of *MAFB+/CLIC5-* cells outside the glomeruli, which we identified as the SSBpod. These cells could be found predominantly in the visceral part of the proximal segment of the SSB but also at the capillary loop stage. This result supported the notion that SSBpod were transient cells that preceded (mature) podocytes. Having localized the SSBpod, we next wanted to confirm *OLFM3* as a marker of this cell type. smFISH of *OLFM3*, *MAFB*, and *CLIC5* showed *OLFM3* to be coexpressed with *MAFB* but absent in cells that were positive for *CLIC5* (Fig 22A), a marker that persists in podocytes in the adult kidney. Quantification of the density of smFISH signals (Fig 22B) showed that *OLFM3* was absent in glomeruli but could be detected in the subpopulation we identified as SSBpod (*MAFB+/CLIC5-*). In summary, these results supported *OLFM3* as a robust marker of podocyte precursors.

Finally, we were wondering whether our dataset would also allow us to identify candidate mechanisms that drive development from SSBpod to podocytes. Differential expression analysis revealed 228 genes that had a significant, bigger than 2-fold changes between SSBpod and podocytes (Fig 23, Hochane et al. [1] S4 Table). Among these we found factors belonging to multiple signaling pathways, such as *FGF1*, *VEGFA*, *HES1*, and *EGF1*. *Vegfa* and *Fgf1* are known to have a homeostatic function in podocytes [52, 53, 54], whereas *Hes1*, a target of the Notch signaling pathway, seems to be necessary for the synthesis of extracellular matrix proteins in these cells [55]. Binding sites for the transcription factor AP-1 were strongly enriched in this set of genes (145 out of 228 genes, adjusted p-value $= 1.3 \cdot 10^{-5}$). *AP-1* would therefore be an interesting target for perturbation studies in mouse models.

All in all, the results presented here complement other, recent, single-cell transcriptomics studies of the fetal kidney. We demonstrated how the data can be interrogated to find expression patterns that will improve our understanding of human kidney development.

2.3 Discussion

2.3.1 The nephrogenic niche is heterogeneous

Building an organ during development requires the careful balance between two fundamental processes—growth and the creation of structure. In many organs, these two functions are reconciled by self-renewing progenitor cells that can be induced to differentiate. In kidney development, NPCs give rise to the epithelium of the nephron, the functional unit of the kid-

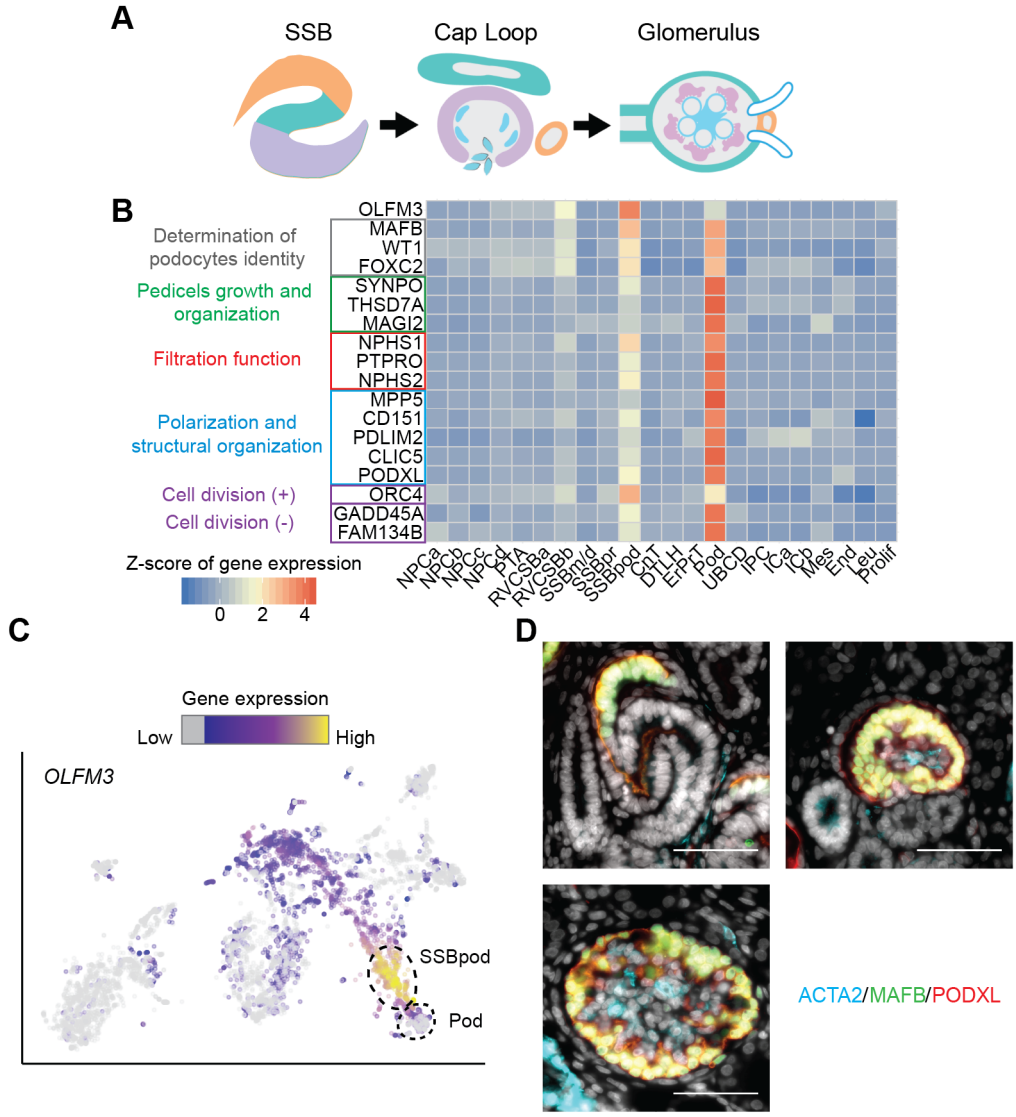


Figure 21. Pod develop via a precursor state. (A) Schematic of development from the SSB to the glomerulus. Regional patterning is shown in color—proximal (purple), medial (green), distal (orange). (B) Expression heat map of the subset of marker set genes that are markers for SSBpod and Pod. Expression was FT transformed, averaged over all cells in a cluster and standardized gene-wise. (C) Two-dimensional tSNE map showing the expression of *OLFM3*. Expression is indicated by color; expression values of 1 are plotted in gray. (D) Typical images of MAFB, PODXL, and ACTA2 immunostaining in SSB and glomeruli. Scale bar = 50μm.

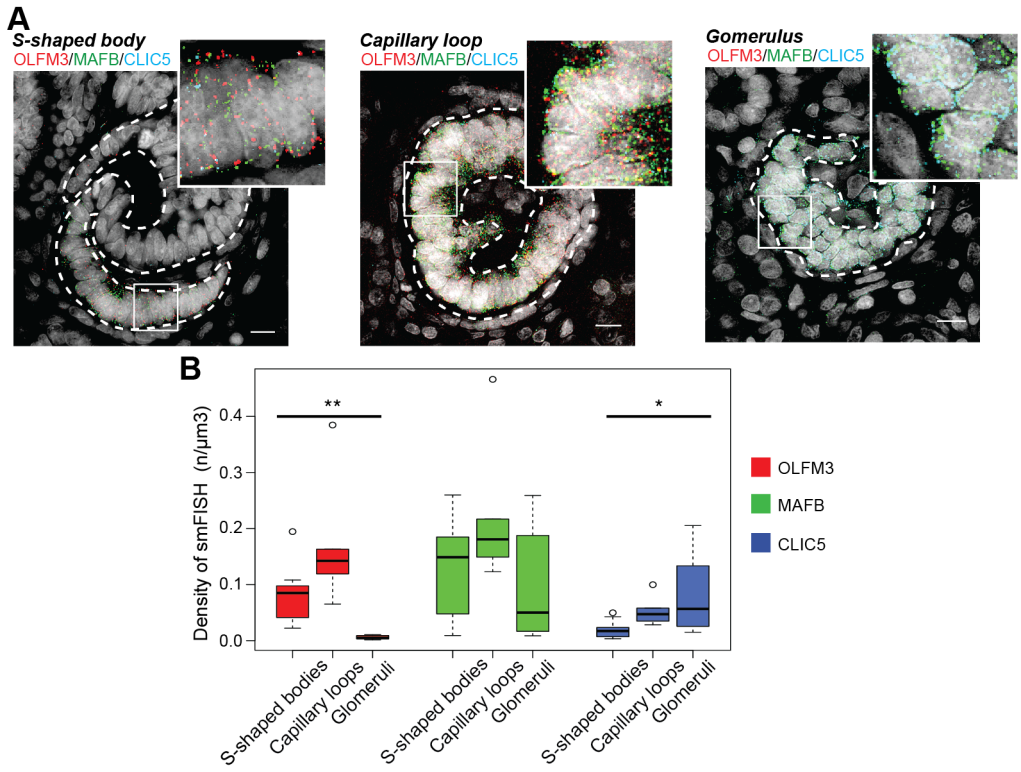


Figure 22. SSBpod is localized in the visceral proximal SSB. (A) Representative images of smFISH of *OLFM3*, *MAFB*, and *CLIC5* in SSBpod and Pod. w15 female kidney. Scale bar = 10µm. (B) Box plots of smFISH signal densities in SSB ($n = 10$), capillary loop ($n = 4$), and glomeruli ($n = 8$), for *OLFM3*, *MAFB* and *CLIC5* (* adjusted $p < 0.05$, ** adjusted $p < 0.0005$).

ney. To balance growth with patterning, self-renewal and differentiation of NPCs have to be tightly controlled. It is well established that the niche of the NPC plays an important role in this control, but the precise mechanisms are not well understood. In particular, it is not clear how the position and movement of NPCs in the niche might impact the induction towards differentiation.

Heterogeneity in the nephrogenic niche was brought to light first by Mugford et al. [27] in 2009 and has been confirmed by multiple recent studies [11, 20, 46, 56, 57]. Mugford et al. [27] used in situ hybridization to study the localization of transcriptional regulators in E15.5 mouse kidney. Three distinct compartments were defined in the CM—inner capping mesenchyme (which lies closest to the cleft of the UB), outer capping mesenchyme (at the tip of the UB), and induced mesenchyme (at the level of the stalk of the UB). Although all compartments express *Six2*, only inner and outer capping mesenchyme express *Cited1*. The induced CM was distinguished by Wnt pathway activity, as evidenced by *Wnt4* expression. Several recent scRNA-seq studies confirmed heterogeneity in the nephrogenic niche. Brunskill et al.

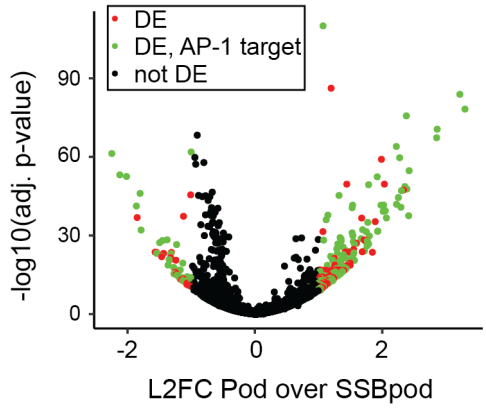


Figure 23. AP-1 targets are enriched in the DEG of Pod versus SSBpod. Volcano plot of differential gene expression between SSBpod and Pod. L2FC Pod over SSBpod versus $-\log_{10}$(adjusted p-value). Genes with an adjusted $p < 0.05$ and L2FC > 1 were considered significant (colored data points). Genes with an AP-1 binding site are shown in red.

[56] studied an E12.5 mouse kidney and found two subpopulations in the CM, which they classified as uninduced (*Six2* positive, *Cited1* positive) and induced (*Six2* positive, *Cited1* negative). Among the hundreds of genes that were differentially expressed between these two populations, they found genes related to the Wnt signaling pathway as well as protein vesicular trafficking and degradation. Wang et al. [57] also found two subclusters in the CM of the human fetal kidney. They interpreted one subcluster as the self-renewing compartment due to higher expression of markers for cell division. The other subcluster, which showed gene expression related to Notch signaling (*HES1*, *HEY1*), was considered induced. Two studies by [11, 20] also explored NPC heterogeneity. The first study [11] identified four NPC clusters (self-renewing, primed, differentiating, and proliferating), whereas the second [20] revealed four clusters of NPCs (I-IV), two clusters of primed NPCs (I-II), as well as several clusters of proliferating cells.

In the dataset presented here, we identified four clusters of NPCs. Among these, NPCa are most likely the self-renewing compartment. In agreement with the studies by Lindström et al. [20] [11, 20], they expressed the highest levels of *CITED1* and *TMEM100* compared with the other NPCs. Furthermore, they preceded all other NPC clusters in pseudotime analysis. NPCb showed expression of several genes that modulate Notch, BMP, and TGF-β pathway activity, as well as low levels of *LEF1*, which has been shown to indicate induction towards differentiation [11, 20]. The classification of NPCb as *primed NPC* by comparison to the Lindström et al. [20] dataset supported the interpretation of NPCb as a state distinct from NPCa that is primed to differentiate. The fact that we detected NPCb only at w16 and w18 leads us to speculate that NPCb could be the result of continuous changes in the nephrogenic niche over the course of development. The third NPC cluster, NPCc, appeared together with NPCb

in pseudotime and was distinguished from the other NPCs by higher expression of genes involved in or regulated by RA signaling. The RA binding protein CRABP2 has been identified as an NPC marker in other reports [11, 57]. We speculate that NPCc are the result of spatially varying concentrations of RA, which is produced in the cortical interstitium [58]. Finally, NPCd appeared between NPCb-c and PTA in pseudotime and were clearly distinguished from the other NPC clusters by increased proliferation, at least as far as that can be inferred from gene expression data. In agreement with our analysis, NPCd were classified as *proliferating cells* by comparison with the Lindström et al. [20] dataset. NPCd cells also lowly expressed markers of induction towards differentiation (such as *LEF1*, *LHX1*, *WNT4*), which indicates a transitory state between induced and/or primed NPCs and PTA. The suggested developmental flow from NPCa via NPCb-c to NPCd was supported by a gradual decrease of *OSR1*, which is a well-known marker of the early CM.

By in situ detection of CITED1, SIX2, and other genes, we also explored the spatial localization of the different NPC clusters. NPCa seemed to reside closest to the tip of the UB, the induced and/or primed NPCb and c were situated closer to the stalk, and NPCd were closest to the PTA. This finding is consistent with the recent report of NPCs streaming from their niche at the UB tip towards the UB branch point to form the PTA and RV [20]. On this path, the cells gradually lose the NPC transcriptional program, and differentiation is induced. In the mouse, trajectories of NPCs also seem to have a large stochastic component: NPCs repeatedly detach from the UB and attach again and also shuttle back and forth between the *uninduced region* at the UB tip and the *committed region* around the stalk of the UB [59]. This observation could indicate that varying expression levels of genes such as *CITED1* occur as a consequence of cell migration and are not necessarily functionally relevant. Indeed, *Cited1* knockout has no adverse effects on kidney development in the mouse [60]. Taken together, our results support a model in which (self-renewing) NPCa reside at the tip of the UB, probably in close proximity or even in contact with the UB. Movement away from the UB tip, toward the stalk, is accompanied by decreased *CITED1* expression and transformation to the (induced and/or primed) NPCb-c states. Arrival at the stalk of the UB is characterized by the NPCd expression state, increased proliferation, and eventually transformation to the PTA. It is conceivable that cells sometimes visit the different NPC states in reverse order, which would reconcile this model with the observed high, multidirectional motility of NPCs [59].

2.3.2 Proximal-distal patterning

In the prevailing model of mammalian kidney development, self-renewing NPCs are not prepatterned to develop into a certain lineage. When the developing nephron first displays signs of proximal-distal patterning is an important, outstanding question. Mugford et al. [27] have found evidence that the PTA, which succeeds the NPC, is already polarized. A recent study by Lindström et al. [20] proposed an intriguing mechanism that couples temporal and spatial cues: although NPCs that are recruited to the PTA early develop into distal cell types, NPCs that are integrated later contribute to the proximal compartment. In our study,

we identified the PTA by known marker genes (*CCND1*, *LHX1*, and *WNT4*) and high proliferation. We were unable to detect any substructure within the PTA, which might be due to the limited resolution of our scRNA-seq method. RV/CSB, the next developmental stage, however, was split in two clusters (a and b). Pseudotime analysis suggested that RVCSBa was a heterogeneous cluster comprising early RV cells (which appeared before RVCSBb) and the distal segment of the RVCSB. This observation is consistent with the time-dependent recruitment model by Lindström et al. [20] in the sense that in that model, distal specification precedes proximal patterning.

2.3.3 Fetal podocytes may have varying degree of maturation

Single-cell transcriptomics studies of various organs have brought to light many new, intermediate cell states. This has provoked the question of whether we should consider gene expression in complex tissues as a continuum rather than a collection of distinct expression profiles. In developmental systems, it is certainly useful to think about gene expression change as a continuous process. Nevertheless, there are clearly distinct intermediate cell states even within linear developmental paths. In our study, we observed that RVCSBb gave rise to podocytes via an intermediate state, the SSBpod, which directly preceded the podocytes in pseudotime. Specific expression of *OLFM3* made it likely that this cluster is identical to previously identified podocyte precursors, which were marked by this gene [20, 46]. Menon et al. [46] defined a cluster of *immature* or *early* podocytes, characterized by high *OLFM3* and low *MAFB* expression. In that study, podocytes showed increased *MAFB* expression but loss of *OLFM3*. Lindström et al. [20] located *OLFM3* positive cells to the proximal part of the SSB. In our study, we confirmed all of these observations: *OLFM3* was localized to the visceral part of the proximal segment of the SSB, and *OLFM3* negative podocytes showed higher expression of mature podocyte markers compared to the *OLFM3* positive SSBpod. Functional annotation analysis of genes that were differentially expressed between SSBpod and podocytes revealed enrichment of a binding site for AP-1. This transcription factor has been found to be important for the development of the skin [61], neural precursor cells [62], and the heart valve [63] in mice. A role of *AP-1* in kidney development has not been described yet, and further research is needed to elucidate its potential function. The analysis of the kidneys from different gestational ages showed high similarity of cell types across different ages with the exception of podocytes. These displayed a systematic change in expression pattern, which might indicate the continued maturation of podocytes over time. This observation is in agreement with a study by Brunskill et al. [64] in the mouse, which compared embryonic (E13.5 and E15.5) with adult podocytes (defined as *Mafb* positive cells). That study found hundreds of genes that were differentially expressed between embryonic and adult podocytes. Furthermore, targeted experiments are needed to demonstrate the possible maturation of podocytes in human kidney development.

In summary, we have leveraged a combination of single-cell transcriptomics and in situ imaging to study the intricate structure of the developing human kidney. The transcrip-

tomics data, accessible via a web application http://www.semraulab.com/kidney, will be a valuable starting point for discovering gene regulatory mechanisms or finding new disease mechanisms.

2.4 Materials and methods

2.4.1 Ethics statement

The collection and use of human material in this study was approved by the Medical Ethics Committee from the Leiden University Medical Center (P08.087). The gestational age was determined by ultrasonography, and the tissue was obtained by vacuum aspiration from women undergoing elective abortion. The material from six embryos (w9, male; w11, male; w13, female; w15, female; w16, male; and w18, female) was donated with written informed consent. Questions about the human material should be directed to S. M. Chuva de Sousa Lopes (Lopes@lumc.nl).

2.4.2 Experimental methods

Single-cell dissociation of human fetal kidney

One human embryo of w16 (male) was isolated and the kidney dissected in cold saline solution (0.9% NaCl, Versylene Fresenius). For sex genotyping, polymerase chain reaction (PCR) for AMELX/Y was used as previously described [65]. The obtained kidney was decapsulated and kept on ice in dissociation buffer (DPBS + Penicillin 100U/mL + Streptomycin 0.1mg/mL; all from Life Technologies) before cutting it into 1-2mm pieces. The pieces were washed three times with washing solution (Advanced DMEM F12 supplemented with ITS commercial solution (Insulin-Transferrin-Selenium; Thermofisher), Glutamax, Penicillin 100U/mL, and Streptomycin 0.1mg/mL) with brief centrifugation (160g) in order to remove as many red blood cells as possible. The washed kidney tissue was then incubated with digestion solution (Trypsin/EDTA solution 0.25% and Collagenase-II 280U/ml) and incubated overnight at 4°C. The next day, the digestion solution was removed, and the kidney was rinsed with washing solution and incubated with washing solution for 30min at 37°C with agitation. Subsequently, the sample was sequentially passed through sterile cell strainers of 100, 70, and 40μm pore size with the help of washing solution. The cells were then centrifuged and counted, and viability was measured to be 78% (trypan blue assay) before proceeding with scRNA-seq library preparation. Four additional human fetal kidneys (w9, male; w11, male; w13, female; and w18, female) were dissected as described above, but, additionally, live cells were purified by FACS before library preparation [66].

scRNA-seq library preparation and sequencing

scRNA-seq libraries were prepared using the Chromium Single Cell 3' Reagent Kit, Version 2 Chemistry (10X Genomics) according to the manufacturer's protocol. Libraries were sequenced on a NextSeq500 in Mid Output mode using a version 2, 150-cycle kit (Illumina).

Immunostaining

A paraffin-embedded w15 human kidney (female) was sectioned (5μm) using a RM2255 microtome (Leica Microsystems GmbH) and mounted on StarFrost slides (Waldemar Knittel).

For immunofluorescence, sections were deparaffinized and rehydrated by standard procedures, starting with xylene (twice for 20min), followed by ethanol with sequential dilution and ending with distilled water, all at room temperature (RT). Antigen retrieval was performed by a double treatment of 10min in a microwave (97°C) with 0.01M sodium citrate buffer (pH 6.0). The sample was then allowed to cool down, rinsed three times with PBS, and blocked for 1h at RT in blocking buffer (1% BSA, 0.05% Tween-20 in PBS). Subsequently, sections were incubated overnight with the following antibodies diluted in blocking buffer—rabbit anti-SIX2 (1:100, 11562-1-AP;Proteintech), mouse anti-CITED1 (1:500, H00004435-M03; Novus Biologicals), mouse anti-MAFB (1:200, LS-C336952; LifeSpan Biosciences), rabbit anti-ACTA2 (1:200, ab5694; Abcam), goat anti-PODXL (1:200, AF1658; R&D Systems), and rabbit anti-UNCX (1:10, PA5-69485; Thermo Fisher Scientific), rabbit anti-CKS2 (HPA003424, 1:100; Sigma Aldrich), rabbit anti-NURR77 (NR4A1) (ab13851, 1:50; Abcam Biochemicals), and mouse anti-HSP70 (HSPA1A) (ab2787, 1:50; Abcam Biochemicals). The secondary antibodies were diluted in blocking buffer and applied at RT for 1h followed by nuclear counterstaining with 4′,6-diamidino-2-phenylindole (DAPI; Life Technologies). The secondary antibodies used were Alexa Fluor 647 donkey anti-rabbit (1:500, A-31573; Life Technologies), Alexa Fluor 594 donkey anti-mouse (1:500, A-21203; Life Technologies), and Alexa Fluor 555 donkey anti-goat (1:500, A32727; Life Technologies). The sections were then mounted using ProLong Gold (Life Technologies).

For immunohistochemistry, sections were deparaffinized and blocked as above. After overnight incubation with primary antibodies rabbit anti-UNCX (1:10, PA5-69485; Thermo Fisher Scientific) and mouse anti-CITED1 (1:500, H00004435-M03; Novus Biologicals) in blocking buffer, 0.3% H_2O_2 was used to quench endogenous peroxidase activity for 20min. Next, the sections were incubated with biotin-labeled goat anti-rabbit IgG (1:200, BA-1000; Vector Laboratories) diluted in normal goat serum (1:66, S-1000; Vector Laboratories) or biotin-labeled horse anti-mouse (1:200, BA-2000; Vector Laboratories) diluted in normal horse serum (1:66, S-2000; Vector Laboratories) for 40min. Sections were then treated for 40min with avidin-biotin-peroxidase complex (VECTASTAIN Elite ABC HRP Kit, #PK-6100; Vector Laboratories) following the manufacturer's instructions, followed by DAB (D5637; Sigma-Aldrich) and hematoxylin (1043020025; Merck) and were mounted with Entellan (1079610100; Merck).

Single-molecule FISH

Paraffin embedded sections from the w15 human fetal kidney (female) used for immunostaining were also used for smFISH experiments. Paraffin was removed by immersion in xylene twice for 10min at RT. The sections were then rehydrated by sequential immersion in ethanol solutions—100% (2x, 10min), 85% (2x, 5min), and 70% (2x, 3min). Subsequently, sections were permeabilized in 70% ethanol for 5h before incubation with proteinase-K (P4850; Sigma Aldrich) for 15min at 37°C (23μg/mL in TE buffer at pH = 8) and a wash in RNAse-free water (3x, 5min). smFISH was performed as described previously [67]. Briefly, custom designed smFISH probes (BioCat,Hochane et al. [1] S5 Table), labeled with Quasar 570, CAL FLuor Red 610, or Quasar 670, were incubated with the samples for 16h at 30°C in hybridization buffer (100 mg/mL dextran sulfate, 25% formamide, 2X SSC, 1mg/mL E.coli tRNA, 1mM vanadyl ribonucleoside complex, 0.25mg/mL BSA). Samples were washed twice for 30min at 30°C with wash buffer (25% formamide, 2x SSC) containing DAPI (1μg/mL, D9542; Sigma). All solutions were prepared with RNAse-free water. Finally, the sections were mounted using ProlongGold (P36930; Life Technologies) and imaged the next day.

Imaging

Immunostained and smFISH-treated kidney sections were imaged on a Nikon Ti-Eclipse epifluorescence microscope equipped with an Andor iXON Ultra 888 EMCCD camera, using a 100x /1.45 Plan Apo Lambda oil objective (Nikon) and dedicated, custom-made fluorescence filter sets (Nikon). To cover large areas of the sectioned kidney, images of multiple adjacent areas were taken and combined using the tiling feature of the NIS Elements software (Nikon). For imaging of smFISH signals, z-stacks were collected with distances of 0.3-0.5μm between planes in four fluorescence channels (GFP, Quasar 570, CAL FLuor Red 610, Quasar 670).

2.4.3 Quantification and statistical analysis

scRNA-seq data pruning and normalization

Single-cell expression for the w16 sample was quantified using unique molecular identifiers (UMIs) by 10X Genomics' *Cell Ranger* software. After removing cells with less than 2,000 transcripts per cell, 8,503 cells were retained for further analysis. On average, 1,789 genes were detected per cell and a median of 4,805 transcripts per cell (Fig 2A). Given the recent report that dissociation can have a significant influence on the single-cell transcriptome [34] and that the kidney is notoriously difficult to dissociate, special attention was paid to dissociation-related artifacts. The amount of 1,859 cells with signs of stress were removed from the dataset (Fig 2B). These cells had more than 10% of their expression come from mitochondrial genes (*MT-ND1*, *MT-ND2*, *MT-CO1*, *MT-CO2*, *MT-ATP8*, *MT-ATP6*, *MT-CO3*, *MT-ND3*, *MT-ND4L*, *MT-ND4*, *MT-ND5*, *MT-ND6*, *MT-CYB*) or more than 5% from stress markers. Stress markers were defined as those genes that were significantly up-regulated upon prolonged enzy-

matic incubation of mouse kidney tissue in the study by Adam et al. [23] (Hochane et al. [1] S2 Table, Fig 2B). Mouse genes from this list were converted to human genes using biomart [68]. Genes of the literature set (Table 2.1) only showed small differences between stressed and nonstressed cells (Fig 2C-D), and stressed cells did not form a separate cluster (Fig 2E-F). Therefore, removing stressed cells did not reduce the cell type diversity in the sample. Additionally, 42 cells had more than 1% of their expression coming from *HBB*, *HBA1*, and *HBA2* and were therefore classified as red blood cells and discarded from any further analysis (Fig 2E). Sporadic expression of hemoglobin genes in other cells was likely due to red blood cells that burst before isolation. The same filtering approach was applied to the samples from the other developmental ages as well as the data from Lindström et al. [20] [20]. Raw UMI counts were smoothened by k-nearest neighbors smoothing version 2.1 [13]. This procedure reduces technical noise by sharing information between transcriptionally similar cells, which likely belong to the same cell type. Briefly, the expression profiles of each cell and its knn were summed (k = 10; distance metric: Euclidean distance of the first 10 PCs with a dither of 0.05). The resulting smoothened count matrix had a higher total count than the original and was therefore scaled back to the original matrix by a global factor. Expression was normalized by the method developed by Lun and colleages [69] (as implemented in the scran (version 1.10.1)R package using the functions quickCluster and computeSumFactors). Normalized gene expression was FT transformed in further analyses unless stated otherwise.

Reduction of dimensionality

Variability of gene expression was calculated using the improvedCV2-function from the scran R package. Intercell distances were calculated using the 5% most HVGs excluding stress markers [23] (Hochane et al. [1] S2 Table) and ribosomal genes (obtained from the HGNC website) without any filter for minimum mean expression. For maps of individual samples, we used (1–Pearson correlation) as distance measure. For maps of combined samples, we used Euclidean distance in the MNN-corrected PC space. All tSNE maps used a perplexity setting of 500. For the DDRTree embedding used with pseudotime analysis, see Pseudotime analysis.

Clustering

Hierarchical cluster analysis was performed using Ward linkage and the same intercell distances as for the reduction of dimensionality. The dendrogram of this clustering was cut at height 0.6 to yield 29 clusters of cells (Fig 4). The cut off was chosen such that the number of resulting clusters was comparable to the number of cell types expected from the literature on mouse development [70] and other scRNA-seq studies of the human fetal kidney [11, 20, 46, 56, 57]. We estimated the number of cell types to be around 20 but created slightly more as a starting point to allow for the discovery of new cell types. On the other hand, we did not want to use a much higher number to avoid overclustering (i.e., creating many clusters that

are merely driven by noise, which would then have to be merged manually). The presence of known markers of the different cell types in the kidney (literature set, Table 2.1) was then used to identify cell types (Fig 4). Based on this analysis, some adjacent clusters (clusters 4 and 5, 11 and 12, 15 and 16, and 17 and 18) showed very similar expression of known marker genes of podocytes, ICs, UB, and collecting duct and proximal tubule cells, respectively. In addition, the aforementioned clusters were in close proximity, both in the clustering dendrogram as well as in tSNE space (Fig 4). Consequently, these clusters were merged. For example, clusters 4 and 5 had similar expression of genes known to be expressed in mature podocytes (*NPHS2*, *PTPRO*, *PODXL*) compared to cluster 6, which showed very weak expression of these genes and had distinctive expression of *OLFM3*, which has been shown to be specifically expressed in podocytes precursors [20, 46]. Furthermore, we also merged clusters 7 and 25, which were more distant in the dendrogram of the hierarchical clustering but had very similar literature marker profiles (e.g., *LHX1*, *WNT4*, *CCND1*, *JAG1*, *PAX2*, and *PAX8*) and appeared in close proximity in tSNE space. Finally, clusters 26 and 29 were also merged. Cluster 26 was a heterogeneous cluster of only 56 cells that were spread in tSNE space between multiple other clusters. This cluster was closest to PTA (cluster 29) in terms of literature marker expression (*WNT4*, *LHX1*, and *CCND1*) and differed from it with respect to proliferative state, which may account for the heterogeneous distribution.

Combining different datasets

To compare cells from multiple scRNA-seq datasets, we used the fastMNN function [21] implemented in scran (version 1.10.1) on the first 50 PCs of the 5% HVGs without stress markers or ribosomal genes. We used a knn approach to infer the cell types of unclassified cells from already classified cells. For each unclassified cell, the 20 nearest neighbors in batch-corrected PC space (Euclidean distance) were determined. The most common cell type among these neighbors was then assigned to the unclassified cell. For the comparison with the dataset from Lindström et al. [20], we restricted our dataset to the nephrogenic niche. The cluster identities for the Lindström et al. [20] dataset were kindly provided to us by the group of Andrew D. Smith.

Cell cycle and proliferation

Cell cycle scores were calculated using the Cyclone tool [15] from the scran (version 1.10.1) R package. A list of proliferation markers was adopted from a publication by Whitfield et al. [16].

Pseudotime analysis

We used the *Monocle 2* algorithm [18] to perform embedding and pseudotime analyses on the 2,594 cells of the nephron epithelium, starting from the PTA (cells classified as PTA, RVCSBa, RVCSBb, SSBm/d, SSBpr, SSBpod, DTLH, ErPrT, or podocytes), and separately on the 2,153

cells of the nephrogenic niche (NPC) and the PTA. The 5% HVGs (without stress or ribosomal genes) were used as input to the algorithm. We used the reduceDimension function (max_components = 3 for Fig 9; max_components = 2 for Fig 16A) to run the DDRTree algorithm [19]. The root of the graph learned by DDRTree was placed on the branch that starts with the PTA to obtain the pseudotime shown in Fig 9B.

Marker genes and KeyGenes

For each gene, the cluster of interest (COI) was defined as the cluster that had the highest mean expression of the gene. Then, a binary classifier based on an expression threshold was defined: cells with expression above that threshold were considered to be part of the COI. We systematically varied this threshold to create a ROC based on the cells' true cluster identities. The AUROC was then used to determine the usefulness of this gene as a marker (rather than the specificity or sensitivity at a specific threshold). Genes that had an AUROC exceeding 0.8 were detected in at least 80% of the cells in the COI, had a minimum mean expression of 1.5 in the COI, and those for which maximally 25% of the cells outside the COI had significant expression, were defined as marker set candidates (Hochane et al. [1] S3 Table). Significant expression was defined here as an expression level higher than the 25th percentile of expression in the COI. Subsequently, the top four candidate marker genes per cluster, as ranked by the AUROC, resulted in a final set of 88 marker genes (marker set, Hochane et al. [1] S3 Table).

To apply the KeyGenes prediction algorithm [22], two-thirds of the cells were assigned to the training set and one-third to the test set. A multinomial logistic regression model was trained on the training set with LASSO shrinkage using the 500 most HVGs, filtered for stress markers and ribosomal genes. The shrinkage parameter was determined by 20-fold cross validation. To apply the KeyGenes method to single cells, each cell was treated as a sample, and cross validation was used to control for overfitting. The model obtained a list of 95 classifier genes with nonzero weights (KeyGenes set, Hochane et al. [1] S3 Table). Thereafter, the cells in the test set were assigned to the cell type with the highest identity (id) score; 84% of the cells in the test set were classified correctly (16% test error). On average, the id score was 0.59, and 24% of the cells in the test set obtained an id score higher than 0.8.

Differential expression analysis

For all differential expression analyses, we used EdgeR (version 3.24.0) [71] on raw counts. Normalization and dispersion estimates were calculated by calcNormFactors and estimateDisp, respectively. We modeled gene expression with a negative binomial generalized linear model with glmQLFit. Besides the conditions to be compared, a detection rate for each gene was added to the design matrix. The detection rate is defined as the fraction of cells with nonzero expression. In the comparison of different ages, we excluded the w9 and w16 samples. The w9 sample contained only a few cells, which results in high uncertainty for average gene expression levels. The w16 sample was created separately from the other samples.

Therefore, to avoid batch effects, which are not corrected for in the differential expression analysis, we therefore also excluded the w16 sample.

GWAS analysis

The NHGRI-EBI GWAS catalog was used to retrieve genes associated with traits related to kidney diseases. Specifically, we selected the following kidney traits: *kidney stone*, *kidney disease*, *rapid kidney function decline*, *chronic kidney disease*, *kidney amyloid deposition measurement*, *acute kidney injury*, *type 1 diabetes nephropathy*, *nephrolithiasis*, *diabetic nephropathy*, *proteinuria*, *GFR change measurement*, *renal cell carcinoma*, *serum creatinine measurement*, *cystatin C measurement*, *type 2 diabetes nephropathy*, *immunosuppressive agent*, *tacrolimus measurement*, *focal segmental glomerulosclerosis*, *nephrotic syndrome*, *membranous glomerulonephritis*, *lupus nephritis*, *IgA glomerulonephritis*, *renal system measurement*, and *Wegener's granulomatosis*. This selection resulted in a list of 560 genes (Hochane et al. [1] S4 Table, Kidney GWAS genes). As a negative control, we also obtained a list of 1,508 genes associated with traits related to lung diseases (Hochane et al. [1] S4 Table, Lung GWAS genes) in which we selected the following traits: lung adenocarcinoma, lung carcinoma, interstitial lung disease, squamous cell lung carcinoma, lung disease severity measurement, family history of lung cancer, non-small cell lung carcinoma, diffusing capacity of the lung for carbon monoxide, pulmonary function measurement, vital capacity, emphysema, idiopathic pulmonary fibrosis, chronic bronchitis, chronic obstructive pulmonary disease, pneumonia, and asthma. We performed a one-sided Fisher's exact test to determine whether the genes in the GWAS lists were significantly enriched in the genes that were differentially expressed in our clusters of interest.

Multiple hypothesis testing

In all cases in which significance is reported, p-values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg method.

Functional annotation enrichment

To look for enrichment of gene ontology (GO) terms or transcription factor binding sites we use the DAVID Functional Annotation tool [72], version 6.8 https://david.ncifcrf.gov/ with all genes in the human genome as background gene set. For enrichment of transcription factor binding sites, we used the *UCSC_TFBS* category.

Image analysis

smFISH image stacks were processed with 3D deconvolution and background correction (rolling ball, radius: 3 pixels = 0.39μm), using the NIS Elements software. Subsequently,

maximum projection was used to create a 2D representation of the image stack. The resulting smFISH images were analyzed with homemade MATLAB scripts. First, autofluorescent background was removed by subtracting the appropriately scaled signal of the GFP channel from each of the other channels. Then a region of interest (ROI) containing the structure of interest was defined manually, and candidate smFISH signals were detected by binarizing the image using a global threshold. Connected components were then counted as smFISH signals if they fulfilled two criteria: their average intensity was bigger than the third quartile of individual pixel intensities and they had an area of three pixels or bigger. The density of smFISH signals in the ROI was calculated as the number of retained connected components divided by the area of the ROI.

Images of immunostaining were pre-processed by background subtraction (rolling ball, radius: 100 pixels = 13μm) using ImageJ [73]. Quantification of the immunostaining signal was carried out using homemade MATLAB scripts. First, the CM region was segmented manually. Then, cross-sections of the CM, roughly perpendicular to the outline of the UB, were drawn by hand, approximately 30 pixels apart. For the starting point of each cross-section, the contour length s along the UB starting from the top of the CM (close to the edge of the cortex) was determined (see Fig 16B). The distance s was expressed relative to the total contour length (from top to bottom of the UB). The distance from the UB along the cross-sections was defined to be the distance d (see Fig 16B). Fluorescence intensities were then averaged over lines of 30 pixels length perpendicular to the drawn cross-section. The resulting intensity profiles (which depend on d and s) were then averaged over multiple images and either s or d to get average intensity profiles depending only on d or s. Normalization to the maximum intensity of each profile resulted in the intensity profiles reported in Fig 16D-E. Division of the CITED1 intensity profile by the SIX2 intensity profile gave the ratio plotted in Fig 16D-E. Accuracy, indicated by error bars in the plots, was quantified as the standard error of the mean calculated over all evaluated profiles.

Data availability

The scRNA-seq data have been deposited in the GEO database under accession number GSE114530. An interactive web application accompanying this paper, which provides convenient access to the data, can be found here: http://www.semraulab.com/kidney.

Acknowledgments

We are thankful to Gynaikon Clinic in Rotterdam for their efforts in collecting and providing the fetal material; Susan Kloet and Emile Meijer from the Leiden Genome Technology Center for cell encapsulation, library preparation, single-cell sequencing, primary data mapping, and quality control; Vanessa Torrens-Juaneda and Ioannis Moustakas for primary and secondary data analysis and discussions; GenomseScan for technical support; and the group of Andrew D. Smith for providing the cluster identities of the Lindström et al. dataset.

2.4.4 Author contributions

Conceptualization: Mazène Hochane, Susana M. Chuva de Sousa Lopes, Stefan Semrau.

Data curation: Mazène Hochane, Patrick R. van den Berg.

Formal analysis: Patrick R. van den Berg.

Funding acquisition: Susana M. Chuva de Sousa Lopes, Stefan Semrau.

Investigation: Mazène Hochane, Patrick R. van den Berg, Xueying Fan, Noémie Bérenger-Currias, Esmée Adegeest, Monika Bialecka, Maaike Nieveen.

Project administration: Mazène Hochane, Susana M. Chuva de Sousa Lopes, Stefan Semrau.

Resources: Xueying Fan, Monika Bialecka, Maaike Nieveen, Susana M. Chuva de Sousa Lopes..

Software: Patrick R. van den Berg, Maarten Menschaart.

Supervision: Susana M. Chuva de Sousa Lopes, Stefan Semrau.

Visualization: Patrick R. van den Berg, Noémie Bérenger-Currias, Esmée Adegeest, Maarten Menschaart.

Writing - original draft: Mazène Hochane, Patrick R. van den Berg, Noémie Bérenger-Currias, Esmée Adegeest, Stefan Semrau.

Writing - review & editing: Mazène Hochane, Patrick R. van den Berg, Xueying Fan, Noémie Bérenger-Currias, Esmée Adegeest, Monika Bialecka, Maaike Nieveen, Maarten Menschaart, Susana M. Chuva de Sousa Lopes, Stefan Semrau.

2.5 Supplementary information

Abbreviation	Cell type	Gene	Reference
NPCs	nephron progenitor cells	OSR1	[35]
		SIX1	[74]
		SIX2	[75, 76]
		CITED1	[77, 78]
		EYA1	[79]
		SALL1	[80, 60]
		MEOX1	[3]
		GDNF	[81]
		ETV4	[82]
		COL2A1	[83]
		HES1	[84, 78]
		CRABP2	[20]
		LEF1	[20]
		ITGA8	[46]
PTA	pretubular aggregate	LHX1	[20]
		WNT4	[85]
		CCND1	[86]
RVCSB a and b	renal vesicle/comma-shaped body	LHX1	[87]
		PAX8	[85]
		JAG1	[88]
		PAX2	[41]
		WNT4	[89]
		SFRP2	[90]
		DLL1	[4]
SSBm/d	SSB medial/distal	HNF1B	[91]
		POU3F3	[92]
		SIM2	[93]
		SOX9	[20]
		IRX2	[46]
		IRX3	[46]
SSBpr	SSB proximal precursor cell	CDH6	[94, 95]
		HNF1A	[96, 97]
		AMN	[93]
SSBpod	SSB podocyte precursor cell	FOXC2	[98]
		MAFB	[48]
		OLFM3	[20]
		WT1	[99]
Pod	podocytes	PTPRO	[100]
		NPHS2	[101]
		NPHS1	[102]
		PODXL	[103]
		TGFBR3	[103]

Continued on next page

Continued from previous page

Abbreviation	Cell type	Gene	Reference
		CLIC5	[104]
		CITED2	[98]
DTLH	distal tubule/loop of Henle	PAPPA2	[105]
		MAL	[106]
		CLCN5	[107]
		SLC12A3	[108]
		UMOD	[20]
		SLC12A1	[109]
ErPrT	early proximal tubule	LRP2	[110]
		ANPEP	[111]
		SLC34A1	[112]
		CLDN1	[113]
		CLDN2	[114]
		SLC13A1	[115]
		CUBN	[116]
CnT	connecting tubule	ALDH1A1	[20]
		TACSTD2	[20]
		CDH1	[117]
IPC	interstitial progenitor cell	GDNF	[118]
		FOXD1	[119]
ICs a and b	interstitial cells a and b	DES	[120]
		COL3A1	[121]
		COL1A1	[122]
		SERPINE2	[123]
		FGF7	[41, 124]
		LEF1	[125]
		DCN	[126]
Mes	mesangial cells	ACTA2	[123]
		TPM2	[127]
		PDGFRB	[128]
		MCAM	[123]
		CSPG4	[123]
		CD248	[129]
UBCD	ureteric bud/collecting duct	KRT18	[130]
		KRT8	[130]
		RET	[131]
		CLDN7	[132]
		AQP2	[133]
		GATA2	[133]
		MMP7	[134]
		CALB1	[117]
End	endothelial cell	KDR	[135]
		TEK	[136]
		FLT1	[135]
		CDH5	[137]

Continued on next page

Continued from previous page			
Abbreviation	Cell type	Gene	Reference
		PECAM1	[138]
Leu	leukocytes	CD37	[139]
		CD48	[140]
		ITGB2	[141]
		IFI30	[142]
		IL1B	[143]

Table 2.1. Literature marker set and references.

Acronyms

AUROC area under the ROC

COI cluster of interest

FDR false discovery rate

FISH fluorescence in situ hybridization

GDNF glial cell-derived neurotrophic factor

GO gene ontology

GWAS genome-wide association studies

knn k-nearest neighbors

L2FC $\log_2$ fold change

PCR polymerase chain reaction

RA retinoic acid

RET ret proto-oncogene

ROC receiver operating characteristic

ROI region of interest

RT room temperature

smFISH single-molecule FISH

tSNE t-distributed stochastic neighbor embedding

UMI unique molecular identifier

w16 week 16

Cell types

CM cap mesenchyme

CnT connecting tubule

CSB comma-shaped body

DTLH distal tubule/loop of Henle

End endothelial cell

ErPrT early proximal tubule

IC interstitial cell

IPC interstitial progenitor cell

Leu leukocytes

LOH loop of Henle

Mes mesangial cells

MM metanephric mesenchyme

NPC nephron progenitor cell

Pod podocyte

PTA pretubular aggregate

RV renal vesicle

RVCSB renal vesicle/comma-shaped body

SSB s-shaped body

SSBm/d SSB medial/distal

SSBpod SSB podocyte precursor cell

SSBpr SSB proximal precursor cell

UB ureteric bud

UBCD ureteric bud/collecting duct

2.6 References

[1] Mazène Hochane et al. “Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development”. In: *PLOS Biology* 17.2 (Feb. 2019), e3000152. DOI: 10.1371/journal.pbio.3000152.

[2] Frank Costantini et al. “Patterning a Complex Organ: Branching Morphogenesis and Nephron Segmentation in Kidney Development”. In: *Developmental Cell* 18.5 (2010), pp. 698–712. DOI: 10.1016/j.devcel.2010.04.008.

[3] Alexander N. Combes et al. “Haploinsufficiency for the Six2 gene increases nephron progenitor proliferation promoting branching and nephron number”. In: *Kidney International* 93.3 (2018), pp. 589–598. DOI: 10.1016/j.kint.2017.09.015.

[4] Melissa H. Little et al. “Mammalian kidney development: principles, progress, and projections”. In: *Cold Spring Harb Perspect Biol* 4.5 (2012), a008300–a008300. DOI: 10.1101/cshperspect.a008300.

[5] John F. Bertram et al. “Why and how we determine nephron number”. In: *Pediatr. Nephrol.* 29.4 (2014), pp. 575–580. DOI: 10.1007/s00467-013-2600-y.

[6] Frank Costantini. “Genetic controls and cellular behaviors in branching morphogenesis of the renal collecting system”. In: *Wiley Interdiscip Rev Dev Biol* 1.5 (2012), pp. 693–713. DOI: 10.1002/wdev.52.

[7] Akio Kobayashi et al. “Identification of a multipotent self-renewing stromal progenitor population during mammalian kidney organogenesis”. In: *Stem Cell Reports* 3.4 (2014), pp. 650–662. DOI: 10.1016/j.stemcr.2014.08.008.

[8] Maria Luisa S. Sequeira-Lopez et al. “The earliest metanephric arteriolar progenitors and their role in kidney vascular development”. In: *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 308.2 (2015), R138–149. DOI: 10.1152/ajpregu.00428.2014.

[9] B. Robert et al. “Evidence that embryonic kidney cells expressing flk-1 are intrinsic, vasculogenic angioblasts”. In: *Am. J. Physiol.* 271.3 Pt 2 (1996), F744–753. DOI: 10.1152/ajprenal.1996.271.3.F744.

[10] Nils O. Lindström et al. “Conserved and Divergent Features of Human and Mouse Kidney Organogenesis”. In: *JASN* 29.3 (2018), pp. 785–805. DOI: 10.1681/ASN.2017080887.

[11] Nils O. Lindström et al. “Conserved and Divergent Features of Mesenchymal Progenitor Cell Types within the Cortical Nephrogenic Niche of the Human and Mouse Kidney”. In: *JASN* 29.3 (2018), pp. 806–824. DOI: 10.1681/ASN.2017080890.

[12] Nils O. Lindström et al. “Conserved and Divergent Molecular and Anatomic Features of Human and Mouse Nephron Patterning”. In: *JASN* 29.3 (2018), pp. 825–840. DOI: 10.1681/ASN.2017091036.

[13] Florian Wagner et al. “K-nearest neighbor smoothing for high-throughput single-cell RNA-Seq data”. In: *bioRxiv* (2018), p. 217737. DOI: 10.1101/217737.

[14] Laurens van der Maaten et al. “Visualizing Data using t-SNE”. In: *Journal of Machine Learning Research* 9.86 (2008), pp. 2579–2605.

[15] Antonio Scialdone et al. “Computational assignment of cell-cycle stage from single-cell transcriptome data”. In: *Methods* 85 (2015), pp. 54–61. DOI: 10.1016/j.ymeth.2015.06.021.

[16] Michael L. Whitfield et al. “Common markers of proliferation”. In: *Nat. Rev. Cancer* 6.2 (2006), pp. 99–106. DOI: 10.1038/nrc1802.

[17] Joseph L. Napoli. “Cellular retinoid binding-proteins, CRBP, CRABP, FABP5: Effects on retinoid metabolism, function and related diseases”. In: *Pharmacol. Ther.* 173 (2017), pp. 19–33. DOI: 10.1016/j.pharmthera.2017.01.004.

[18] Xiaojie Qiu et al. “Reversed graph embedding resolves complex single-cell trajectories”. In: *Nat. Methods* 14.10 (2017), pp. 979–982. DOI: 10.1038/nmeth.4402.

[19] Qi Mao et al. “Dimensionality Reduction Via Graph Structure Learning”. In: 2015, pp. 765–774. DOI: 10.1145/2783258.2783309.

[20] Nils O Lindström et al. “Progressive Recruitment of Mesenchymal Progenitors Reveals a Time-Dependent Process of Cell Fate Acquisition in Mouse and Human Nephrogenesis”. In: *Developmental Cell* 45.5 (2018), 651–660.e4. DOI: 10.1016/j.devcel.2018.05.010.

[21] Laleh Haghverdi et al. “Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors”. In: *Nat. Biotechnol.* 36.5 (2018), pp. 421–427. DOI: 10.1038/nbt.4091.

[22] Matthias S. Roost et al. “KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas”. In: *Stem Cell Reports* 4.6 (2015), pp. 1112–1124. DOI: 10.1016/j.stemcr.2015.05.002.

[23] Mike Adam et al. “Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development”. In: *Development* 144.19 (2017), pp. 3625–3632. DOI: 10.1242/dev.151142.

[24] J M Linton et al. “The ECM protein nephronectin promotes kidney development via integrin 8 1-mediated stimulation of Gdnf expression”. In: *Development* 134.13 (2007), pp. 2501–2509. DOI: 10.1242/dev.005033.

[25] Janina Müller-Deile et al. “Podocytes from the diagnostic and therapeutic point of view”. In: *Pflugers Arch.* 469.7-8 (2017), pp. 1007–1015. DOI: 10.1007/s00424-017-1993-z.

[26] Liliana Schaefer et al. “Small proteoglycans of normal adult human kidney: Distinct expression patterns of decorin, biglycan, fibromodulin, and lumican”. In: *Kidney International* 58.4 (2000), pp. 1557–1568. DOI: 10.1046/j.1523-1755.2000.00317.x.

[27] Joshua W. Mugford et al. “High-resolution gene expression analysis of the developing mouse kidney defines novel cellular compartments within the nephron progenitor population”. In: *Dev. Biol.* 333.2 (2009), pp. 312–323. DOI: 10.1016/j.ydbio.2009.06.043.

[28] Kyung-Ah Kim et al. “R-Spondin Family Members Regulate the Wnt Pathway by a Common Mechanism”. In: *MBoC* 19.6 (2008), pp. 2588–2596. DOI: 10.1091/mbc.e08-02-0187.

[29] C. Pedraza et al. “A retinoic acid-responsive element in human midkine gene”. In: *J. Biochem.* 117.4 (1995), pp. 845–849. DOI: 10.1093/oxfordjournals.jbchem.a124785.

[30] Katri M. Makkonen et al. “Regulation of the hyaluronan synthase 2 gene by convergence in cyclic AMP response element-binding protein and retinoid acid receptor signaling”. In: *J. Biol. Chem.* 284.27 (2009), pp. 18270–18281. DOI: 10.1074/jbc.M109.012492.

[31] José Vilar et al. “Midkine is involved in kidney development and in its regulation by retinoids”. In: *J. Am. Soc. Nephrol.* 13.3 (2002), pp. 668–676.

[32] Waichi Sato et al. “Midkine expression in the course of nephrogenesis and its role in ischaemic reperfusion injury”. In: *Nephrol. Dial. Transplant.* 17 Suppl 9 (2002), pp. 52–54. DOI: 10.1093/ndt/17.suppl_9.52.

[33] Ayaka Kimura et al. “HMGB2 expression is associated with transition from a quiescent to an activated state of adult neural stem cells”. In: *Dev. Dyn.* 247.1 (2018), pp. 229–238. DOI: 10.1002/dvdy.24559.

[34] Susanne C. van den Brink et al. “Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations”. In: *Nature Methods* 14.10 (2017), pp. 935–936. DOI: 10.1038/nmeth.4437.

[35] Joshua W. Mugford et al. “Osr1 expression demarcates a multi-potent population of intermediate mesoderm that undergoes progressive restriction to an Osr1-dependent nephron progenitor compartment within the mammalian kidney”. In: *Dev. Biol.* 324.1 (2008), pp. 88–98. DOI: 10.1016/j.ydbio.2008.09.010.

[36] Yasmine Neirijnck et al. “Sox11 gene disruption causes congenital anomalies of the kidney and urinary tract (CAKUT)”. In: *Kidney Int.* 93.5 (2018), pp. 1142–1153. DOI: 10.1016/j.kint.2017.11.026.

[37] Priscilla Soulié et al. “Spatially restricted hyaluronan production by Has2 drives epithelial tubulogenesis in vitro”. In: *Am. J. Physiol., Cell Physiol.* 307.8 (2014), pp. C745–759. DOI: 10.1152/ajpcell.00047.2014.

[38] Tomoko Ohmori et al. “Sall1 in renal stromal progenitors non-cell autonomously restricts the excessive expansion of nephron progenitors”. In: *Sci Rep* 5.1 (2015), p. 15676. DOI: 10.1038/srep15676.

[39] Cecil Han et al. “The RNA-binding protein DDX1 promotes primary microRNA maturation and inhibits ovarian tumor progression”. In: *Cell Rep* 8.5 (2014), pp. 1447–1460. DOI: 10.1016/j.celrep.2014.07.058.

[40] Neeraja Sammeta et al. “Uncx regulates proliferation of neural progenitor cells and neuronal survival in the olfactory epithelium”. In: *Mol. Cell. Neurosci.* 45.4 (2010), pp. 398–407. DOI: 10.1016/j.mcn.2010.07.013.

[41] Minoru Takasato et al. “Identification of kidney mesenchymal genes by a combination of microarray analysis and Sall1-GFP knockin mice”. In: *Mech. Dev.* 121.6 (2004), pp. 547–557. DOI: 10.1016/j.mod.2004.04.007.

[42] Yukinori Okada et al. “Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations”. In: *Nat. Genet.* 44.8 (2012), pp. 904–909. DOI: 10.1038/ng.2352.

[43] Mathias Gorski et al. “1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function”. In: *Sci Rep* 7 (2017), p. 45040. DOI: 10.1038/srep45040.

[44] Anubha Mahajan et al. “Trans-ethnic Fine Mapping Highlights Kidney-Function Genes Linked to Salt Sensitivity”. In: *Am. J. Hum. Genet.* 99.3 (2016), pp. 636–646. DOI: 10.1016/j.ajhg.2016.07.012.

[45] Cristian Pattaro et al. “Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function”. In: *Nat Commun* 7 (2016), p. 10023. DOI: 10.1038/ncomms10023.

[46] Rajasree Menon et al. “Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney”. In: *Development* 145.16 (2018). DOI: 10.1242/dev.164038.

[47] Masaru Motojima et al. “Foxc1 and Foxc2 are necessary to maintain glomerular podocytes”. In: *Exp. Cell Res.* 352.2 (2017), pp. 265–272. DOI: 10.1016/j.yexcr.2017.02.016.

[48] Virginia Sadl et al. “The mouse Kreisler (Krml1/MafB) segmentation gene is required for differentiation of glomerular visceral epithelial cells”. In: *Dev. Biol.* 249.1 (2002), pp. 16–29. DOI: 10.1006/dbio.2002.0751.

[49] Kevin V. Lemley. “Mechanical challenges to the glomerulus and podocyte loss: evolution of a paradigm”. In: *Pflugers Arch.* 469.7-8 (2017), pp. 959–963. DOI: 10.1007/s00424-017-2012-0.

[50] S. Nanayakkara et al. “An integrative study of the genetic, social and environmental determinants of chronic kidney disease characterized by tubulointerstitial damages in the North Central Region of Sri Lanka.” In: *J Occup Health* 56.1 (2014), pp. 28–38. DOI: 10.1539/joh.13-0172-OA.

[51] T. A. Moreno et al. “The secreted glycoprotein Noelin-1 promotes neurogenesis in Xenopus”. In: *Dev. Biol.* 240.2 (2001), pp. 340–360. DOI: 10.1006/dbio.2001.0472.

[52] Diangeng Li et al. “Mesenchymal stem cells protect podocytes from apoptosis induced by high glucose via secretion of epithelial growth factor”. In: *Stem Cell Res Ther* 4.5 (2013), p. 103. DOI: 10.1186/scrt314.

[53] Guang Liang et al. “Fibroblast growth factor 1 ameliorates diabetic nephropathy by an anti-inflammatory mechanism”. In: *Kidney Int.* 93.1 (2018), pp. 95–109. DOI: 10.1016/j.kint.2017.05.013.

[54] Christina S. Bartlett et al. “Vascular Growth Factors and Glomerular Disease”. In: *Annu. Rev. Physiol.* 78 (2016), pp. 437–461. DOI: 10.1146/annurev-physiol-021115-105412.

[55] Min Yao et al. “The Notch pathway mediates the angiotensin II-induced synthesis of extracellular matrix components in podocytes”. In: *Int. J. Mol. Med.* 36.1 (2015), pp. 294–300. DOI: 10.3892/ijmm.2015.2193.

[56] Eric W. Brunskill et al. “Single cell dissection of early kidney development: multilineage priming”. In: *Development* 141.15 (2014), pp. 3093–3101. DOI: 10.1242/dev.110601.

[57] Ping Wang et al. “Dissecting the Global Dynamic Molecular Profiles of Human Fetal Kidney Development by Single-Cell RNA Sequencing”. In: *Cell Rep* 24.13 (2018), 3554–3567.e3. DOI: 10.1016/j.celrep.2018.08.056.

[58] Carolina Rosselot et al. “Non-cell-autonomous retinoid signaling is crucial for renal development”. In: *Development* 137.2 (2010), pp. 283–292. DOI: 10.1242/dev.040287.

[59] Alexander N. Combes et al. “Cap mesenchyme cell swarming during kidney development is influenced by attraction, repulsion, and adhesion to the ureteric tip”. In: *Dev. Biol.* 418.2 (2016), pp. 297–306. DOI: 10.1016/j.ydbio.2016.06.028.

[60] Scott Boyle et al. “Cited1 and Cited2 are differentially expressed in the developing kidney but are not required for nephrogenesis”. In: *Dev. Dyn.* 236.8 (2007), pp. 2321–2330. DOI: 10.1002/dvdy.21242.

[61] Christina A. Young et al. “Embryonic AP1 Transcription Factor Deficiency Causes a Collodion Baby-Like Phenotype”. In: *J. Invest. Dermatol.* 137.9 (2017), pp. 1868–1877. DOI: 10.1016/j.jid.2017.04.032.

[62] Fumiaki Kawashima et al. “c-jun is differentially expressed in embryonic and adult neural precursor cells”. In: *Histochem. Cell Biol.* 147.6 (2017), pp. 721–731. DOI: 10.1007/s00418-016-1536-2.

[63] Victoria C. Garside et al. “SOX9 modulates the expression of key transcription factors required for heart valve development”. In: *Development* 142.24 (2015), pp. 4340–4350. DOI: 10.1242/dev.125252.

[64] Eric W. Brunskill et al. “Defining the molecular character of the developing and adult kidney podocyte”. In: *PLoS ONE* 6.9 (2011), e24640. DOI: 10.1371/journal.pone.0024640.

[65] A Heeren et al. “Development of the follicular basement membrane during human gametogenesis and early folliculogenesis”. In: *BMC Developmental Biology* 15.1 (2015), p. 4. DOI: 10.1186/s12861-015-0054-0.

[66] Ábel Vértesy et al. “Parental haplotype-specific single-cell transcriptomics reveal incomplete epigenetic reprogramming in human female germ cells”. In: *Nat Commun* 9.1 (2018), p. 1873. DOI: 10.1038/s41467-018-04215-7.

[67] Stefan Semrau et al. “FuseFISH: robust detection of transcribed gene fusions in single cells”. In: *Cell Rep* 6.1 (2014), pp. 18–23. DOI: 10.1016/j.celrep.2013.12.002.

[68] Steffen Durinck et al. “Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt”. In: *Nat Protoc* 4.8 (2009), pp. 1184–1191. DOI: 10.1038/nprot.2009.97.

[69] Aaron T. L. Lun et al. “Pooling across cells to normalize single-cell RNA sequencing data with many zero counts”. In: *Genome Biol.* 17 (2016), p. 75. DOI: 10.1186/s13059-016-0947-7.

[70] Andrew P. McMahon. “Development of the Mammalian Kidney”. In: *Curr. Top. Dev. Biol.* 117 (2016), pp. 31–64. DOI: 10.1016/bs.ctdb.2015.10.010.

[71] Mark D. Robinson et al. “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data”. In: *Bioinformatics* 26.1 (2010), pp. 139–140. DOI: 10.1093/bioinformatics/btp616.

[72] Da Wei Huang et al. “Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources”. In: *Nature Protocols* 4.1 (2009), pp. 44–57. DOI: 10.1038/nprot.2008.211.

[73] Caroline A. Schneider et al. “NIH Image to ImageJ: 25 years of image analysis”. In: *Nat. Methods* 9.7 (2012), pp. 671–675. DOI: 10.1038/nmeth.2089.

[74] Pin-Xian Xu et al. “Six1 is required for the early organogenesis of mammalian kidney”. In: *Development* 130.14 (2003), pp. 3085–3094. DOI: 10.1242/dev.00536.

[75] Michael Marcotte et al. “Gene regulatory network of renal primordium development”. In: *Pediatr Nephrol* 29.4 (2014), pp. 637–644. DOI: 10.1007/s00467-013-2635-0.

[76] Akio Kobayashi et al. “Six2 defines and regulates a multipotent self-renewing nephron progenitor population throughout mammalian kidney development”. In: *Cell Stem Cell* 3.2 (2008), pp. 169–181. DOI: 10.1016/j.stem.2008.05.020.

[77] Scott Boyle et al. “Fate mapping using Cited1-CreERT2 mice demonstrates that the cap mesenchyme contains self-renewing progenitor cells and gives rise exclusively to nephronic epithelia”. In: *Dev. Biol.* 313.1 (2008), pp. 234–245. DOI: 10.1016/j.ydbio.2007.10.014.

[78] Sayoko Fujimura et al. “Notch2 activation in the embryonic kidney depletes nephron progenitors”. In: *J. Am. Soc. Nephrol.* 21.5 (2010), pp. 803–810. DOI: 10.1681/ASN.2009040353.

[79] P. X. Xu et al. “Eya1-deficient mice lack ears and kidneys and show abnormal apoptosis of organ primordia”. In: *Nat. Genet.* 23.1 (1999), pp. 113–117. DOI: 10.1038/12722.

[80] Ryuichi Nishinakamura et al. “Murine homolog of SALL1 is essential for ureteric bud invasion in kidney development”. In: *Development* 128.16 (2001), pp. 3105–3115.

[81] Alexander N. Combes et al. “Cell-cell interactions driving kidney morphogenesis”. In: *Curr. Top. Dev. Biol.* 112 (2015), pp. 467–508. DOI: 10.1016/bs.ctdb.2014.12.002.

[82] Benson C. Lu et al. “Etv4 and Etv5 are required downstream of GDNF and Ret for kidney branching morphogenesis”. In: *Nat. Genet.* 41.12 (2009), pp. 1295–1302. DOI: 10.1038/ng.476.

[83] Masahito Miyamoto et al. “In-depth proteomic profiling of the normal human kidney glomerulus using two-dimensional protein prefractionation in combination with liquid chromatography-tandem mass spectrometry”. In: *J. Proteome Res.* 6.9 (2007), pp. 3680–3690. DOI: 10.1021/pr070203n.

[84] Sophie Jarriault et al. “Signalling downstream of activated mammalian Notch”. In: *Nature* 377.6547 (1995), pp. 355–358. DOI: 10.1038/377355a0.

[85] K. Stark et al. “Epithelial transformation of metanephric mesenchyme in the developing kidney regulated by Wnt-4”. In: *Nature* 372.6507 (1994), pp. 679–683. DOI: 10.1038/372679a0.

[86] Thomas F. Gallegos et al. “A Protein Kinase A and Wnt-dependent network regulating an intermediate stage in epithelial tubulogenesis during kidney development”. In: *Dev Biol* 364.1 (2012), pp. 11–21. DOI: 10.1016/j.ydbio.2012.01.014.

[87] Kimmo Halt et al. “Coordination of kidney organogenesis by Wnt signaling”. In: *Pediatr Nephrol* 29.4 (2014), pp. 737–744. DOI: 10.1007/s00467-013-2733-z.

[88] Madhulika Sharma et al. “Coexpression of Cux-1 and Notch signaling pathway components during kidney development”. In: *Dev. Dyn.* 231.4 (2004), pp. 828–838. DOI: 10.1002/dvdy.20175.

[89] S. Steven Potter et al. “Laser capture-microarray analysis of Lim1 mutant kidney development”. In: *Genesis* 45.7 (2007), pp. 432–439. DOI: 10.1002/dvg.20309.

[90] Cornelia Leimeister et al. “Developmental expression patterns of mouse sFRP genes encoding members of the secreted frizzled related protein family”. In: *Mechanisms of Development* 75.1 (1998), pp. 29–42. DOI: 10.1016/S0925-4773(98)00072-0.

[91] Claire Heliot et al. “HNF1B controls proximal-intermediate nephron segment identity in vertebrates by regulating Notch signalling components and Irx1/2”. In: *Development* 140.4 (2013), pp. 873–885. DOI: 10.1242/dev.086538.

[92] Alexandra Rieger et al. “Missense Mutation of POU Domain Class 3 Transcription Factor 3 in Pou3f3L423P Mice Causes Reduced Nephron Number and Impaired Development of the Thick Ascending Limb of the Loop of Henle”. In: *PLOS ONE* 11.7 (2016), e0158977. DOI: 10.1371/journal.pone.0158977.

[93] Eric W. Brunskill et al. “Atlas of Gene Expression in the Developing Kidney at Microanatomic Resolution”. In: *Dev Cell* 15.5 (2008), pp. 781–791. DOI: 10.1016/j.devcel.2008.09.007.

[94] Steven P. Mah et al. “Kidney Development in Cadherin-6 Mutants: Delayed Mesenchyme-to-Epithelial Conversion and Loss of Nephrons”. In: *Developmental Biology* 223.1 (2000), pp. 38–53. DOI: 10.1006/dbio.2000.9738.

[95] Liwei Jia et al. “Distinct roles of cadherin-6 and E-cadherin in tubulogenesis and lumen formation”. In: *Mol Biol Cell* 22.12 (2011), pp. 2031–2041. DOI: 10.1091/mbc.E11-01-0038.

[96] Filippo Massa et al. “Hepatocyte nuclear factor 1β controls nephron tubular development”. In: *Development* 140.4 (2013), pp. 886–896. DOI: 10.1242/dev.086546.

[97] M. Pontoglio et al. “Hepatocyte nuclear factor 1 inactivation results in hepatic dysfunction, phenylketonuria, and renal Fanconi syndrome”. In: *Cell* 84.4 (1996), pp. 575–585. DOI: 10.1016/s0092-8674(00)81033-8.

[98] Minoru Takemoto et al. “Large-scale identification of genes implicated in kidney glomerulus development and function”. In: *EMBO J.* 25.5 (2006), pp. 1160–1174. DOI: 10.1038/sj.emboj.7601014.

[99] A. J. Buckler et al. “Isolation, characterization, and expression of the murine Wilms’ tumor gene (WT1) during kidney development”. In: *Mol. Cell. Biol.* 11.3 (1991), pp. 1707–1712. DOI: 10.1128/mcb.11.3.1707.

[100] Pedro J. Beltran et al. “Expression of PTPRO during mouse development suggests involvement in axonogenesis and differentiation of NT-3 and NGF-dependent neurons”. In: *J. Comp. Neurol.* 456.4 (2003), pp. 384–395. DOI: 10.1002/cne.10532.

[101] Natalya V. Kaverina et al. “Partial podocyte replenishment in experimental FSGS derives from nonpodocyte sources”. In: *Am. J. Physiol. Renal Physiol.* 310.11 (2016), F1397–1413. DOI: 10.1152/ajprenal.00369.2015.

[102] Ellen F. Carney. “Podocytes: ShcA regulates nephrin turnover”. In: *Nat Rev Nephrol* 13.12 (2017), p. 722. DOI: 10.1038/nrneph.2017.153.

[103] C. Schell et al. “Glomerular development–shaping the multi-cellular filtration unit”. In: *Semin. Cell Dev. Biol.* 36 (2014), pp. 39–49. DOI: 10.1016/j.semcdb.2014.07.016.

[104] Brian A. Pierchala et al. “Proteomic analysis of the slit diaphragm complex: CLIC5 is a protein critical for podocyte morphology and function”. In: *Kidney International* 78.9 (2010), pp. 868–882. DOI: 10.1038/ki.2010.212.

[105] Allen W. Cowley et al. “Pappa2 is linked to salt-sensitive hypertension in Dahl S rats”. In: *Physiol Genomics* 48.1 (2016), pp. 62–72. DOI: 10.1152/physiolgenomics.00097.2015.

[106] Monica Carmosino et al. “MAL/VIP17, a New Player in the Regulation of NKCC2 in the Kidney”. In: *Mol. Biol. Cell* 21.22 (2010), pp. 3985–3997. DOI: 10.1091/mbc.E10-05-0456.

[107] J. Christopher Hennings et al. “The ClC-K2 Chloride Channel Is Critical for Salt Handling in the Distal Nephron”. In: *J Am Soc Nephrol* 28.1 (2017), pp. 209–217. DOI: 10.1681/ASN.2016010085.

[108] Catherina A. Cuevas et al. “Potassium Sensing by Renal Distal Tubules Requires Kir4.1”. In: *J. Am. Soc. Nephrol.* 28.6 (2017), pp. 1814–1825. DOI: 10.1681/ASN.2016090935.

[109] Hayo Castrop et al. “Physiology and pathophysiology of the renal Na-K-2Cl cotransporter (NKCC2)”. In: *Am. J. Physiol. Renal Physiol.* 307.9 (2014), F991–F1002. DOI: 10.1152/ajprenal.00432.2014.

[110] Christopher P. Larsen et al. “LDL Receptor-Related Protein 2 (Megalin) as a Target Antigen in Human Kidney Anti-Brush Border Antibody Disease”. In: *J. Am. Soc. Nephrol.* 29.2 (2017), pp. 644–653. DOI: 10.1681/ASN.2017060664.

[111] Kumar Kotlo et al. “Aminopeptidase N reduces basolateral Na+ -K+ -ATPase in proximal tubule cells”. In: *Am. J. Physiol. Renal Physiol.* 293.4 (2007), F1047–1053. DOI: 10.1152/ajprenal.00074.2007.

[112] Daniel Caballero et al. “Intraperitoneal pyrophosphate treatment reduces renal calcifications in Npt2a null mice”. In: *PLoS ONE* 12.7 (2017), e0180098. DOI: 10.1371/journal.pone.0180098.

[113] Yumiko Kiuchi-Saishin et al. “Differential Expression Patterns of Claudins, Tight Junction Membrane Proteins, in Mouse Nephron Segments”. In: *JASN* 13.4 (2002), pp. 875–886.

[114] Takamoto Ohse et al. “Establishment of Conditionally Immortalized Mouse Glomerular Parietal Epithelial Cells in Culture”. In: *J Am Soc Nephrol* 19.10 (2008), pp. 1879–1890. DOI: 10.1681/ASN.2007101087.

[115] Daniel Markovich. “Na+-sulfate cotransporter SLC13A1”. In: *Pflugers Arch - Eur J Physiol* 466.1 (2014), pp. 131–137. DOI: 10.1007/s00424-013-1388-8.

[116] Erik Ilsø Christensen et al. “Receptor-mediated endocytosis in renal proximal tubule”. In: *Pflugers Arch.* 458.6 (2009), pp. 1039–1048. DOI: 10.1007/s00424-009-0685-8.

[117] Kylie Georgas et al. “Analysis of early nephron patterning reveals a role for distal RV proliferation in fusion to the ureteric tip via a cap mesenchyme-derived connecting segment”. In: *Dev. Biol.* 332.2 (2009), pp. 273–286. DOI: 10.1016/j.ydbio.2009.05.578.

[118] Bliss Magella et al. “Cross-platform single cell analysis of kidney development shows stromal cells express Gdnf”. In: *Dev. Biol.* 434.1 (2018), pp. 36–47. DOI: 10.1016/j.ydbio.2017.11.006.

[119] Benjamin D. Humphreys et al. “Fate tracing reveals the pericyte and not epithelial origin of myofibroblasts in kidney fibrosis”. In: *Am. J. Pathol.* 176.1 (2010), pp. 85–97. DOI: 10.2353/ajpath.2010.090517.

[120] Scott C. Boyle et al. “Notch signaling is required for the formation of mesangial cells from a stromal mesenchyme precursor during kidney development”. In: *Development* 141.2 (2014), pp. 346–354. DOI: 10.1242/dev.100271.

[121] Lino Muñoz Cuellar et al. “Identification and localization of novel genes preferentially expressed in human kidney glomerulus”. In: *Nephrology (Carlton)* 14.1 (2009), pp. 94–104. DOI: 10.1111/j.1440-1797.2008.01009.x.

[122] Prashant S. Patole et al. “Toll-like receptor-4: renal cells and bone marrow cells signal for neutrophil recruitment during pyelonephritis”. In: *Kidney Int.* 68.6 (2005), pp. 2582–2587. DOI: 10.1111/j.1523-1755.2005.00729.x.

[123] Yuqiu Lu et al. “Single-cell RNA-sequence analysis of mouse glomerular mesangial cells uncovers mesangial cell essential genes”. In: *Kidney Int.* 92.2 (2017), pp. 504–513. DOI: 10.1016/j.kint.2017.01.016.

[124] J. Qiao et al. “FGF-7 modulates ureteric bud growth and nephron number in the developing kidney”. In: *Development* 126.3 (1999), pp. 547–554.

[125] Naoki Nakagawa et al. “Dicer1 activity in the stromal compartment regulates nephron differentiation and vascular patterning during mammalian kidney organogenesis”. In: *Kidney Int.* 87.6 (2015), pp. 1125–1140. DOI: 10.1038/ki.2014.406.

[126] Jennifer L. Fetting et al. “FOXD1 promotes nephron progenitor differentiation by repressing decorin in the embryonic kidney”. In: *Development* 141.1 (2014), pp. 17–27. DOI: 10.1242/dev.089078.

[127] Konstantinos Stamatakis et al. “Identification of novel protein targets for modification by 15-deoxy-Delta12,14-prostaglandin J2 in mesangial cells reveals multiple interactions with the cytoskeleton”. In: *J. Am. Soc. Nephrol.* 17.1 (2006), pp. 89–98. DOI: 10.1681/ASN.2005030329.

[128] Taizo Nakagawa et al. “Roles of PDGF receptor-beta in the structure and function of postnatal kidney glomerulus”. In: *Nephrol. Dial. Transplant.* 26.2 (2011), pp. 458–468. DOI: 10.1093/ndt/gfq468.

[129] Stuart W. Smith et al. “CD248+ stromal cells are associated with progressive chronic kidney disease”. In: *Kidney Int.* 80.2 (2011), pp. 199–207. DOI: 10.1038/ki.2011.103.

[130] Sonja Djudjaj et al. “Keratins are novel markers of renal epithelial cell injury”. In: *Kidney Int.* 89.4 (2016), pp. 792–808. DOI: 10.1016/j.kint.2015.10.015.

[131] Frank Costantini. “GDNF/Ret signaling and renal branching morphogenesis: From mesenchymal signals to epithelial cell behaviors”. In: *Organogenesis* 6.4 (2010), pp. 252–262. DOI: 10.4161/org.6.4.12680.

[132] Halim Khairallah et al. “Claudin-7, -16, and -19 during mouse kidney development”. In: *Tissue Barriers* 2.4 (2014), e964547. DOI: 10.4161/21688362.2014.964547.

[133] Lei Yu et al. “GATA2 Regulates Body Water Homeostasis through Maintaining Aquaporin 2 Expression in Renal Collecting Ducts”. In: *Mol Cell Biol* 34.11 (2014), pp. 1929–1941. DOI: 10.1128/MCB.01659-13.

[134] John K. McGuire et al. “Matrilysin (MMP-7) inhibition of BMP-7 induced renal tubular branching morphogenesis suggests a role in the pathogenesis of human renal dysplasia”. In: *J. Histochem. Cytochem.* 60.3 (2012), pp. 243–253. DOI: 10.1369/0022155411435152.

[135] M. Shibuya. “Structure and dual function of vascular endothelial growth factor receptor-1 (Flt-1)”. In: *Int. J. Biochem. Cell Biol.* 33.4 (2001), pp. 409–420. DOI: 10.1016/s1357-2725(01)00026-7.

[136] S. Davis et al. “Isolation of angiopoietin-1, a ligand for the TIE2 receptor, by secretion-trap expression cloning”. In: *Cell* 87.7 (1996), pp. 1161–1169. DOI: 10.1016/s0092-8674(00)81812-7.

[137] P. Carmeliet et al. “Targeted deficiency or cytosolic truncation of the VE-cadherin gene in mice impairs VEGF-mediated endothelial survival and angiogenesis”. In: *Cell* 98.2 (1999), pp. 147–157. DOI: 10.1016/s0092-8674(00)81010-7.

[138] Qi Ren et al. “Platelet endothelial cell adhesion molecule-1 (PECAM1) plays a critical role in the maintenance of human vascular endothelial barrier function”. In: *Cell Biochem. Funct.* 33.8 (2015), pp. 560–565. DOI: 10.1002/cbf.3155.

[139] R. Schwartz-Albiez et al. “The B cell-associated CD37 antigen (gp40-52). Structure and subcellular expression of an extensively glycosylated glycoprotein”. In: *J. Immunol.* 140.3 (1988), pp. 905–914.

[140] S. J. Davis et al. “The structure and ligand interactions of CD2: implications for T-cell function”. In: *Immunol. Today* 17.4 (1996), pp. 177–187. DOI: 10.1016/0167-5699(96)80617-7.

[141] Sam W. Moore et al. “The ITGB2 immunomodulatory gene (CD18), enterocolitis, and Hirschsprung’s disease”. In: *J. Pediatr. Surg.* 43.8 (2008), pp. 1439–1444. DOI: 10.1016/j.jpedsurg.2007.12.057.

[142] Priya Srinivasan et al. “Signal transducer and activator of transcription 1 negatively regulates constitutive gamma interferon-inducible lysosomal thiol reductase expression”. In: *Immunology* 132.2 (2011), pp. 209–216. DOI: 10.1111/j.1365-2567.2010.03355.x.

[143] øyvind Salvesen et al. “Activation of innate immune genes in caprine blood leukocytes after systemic endotoxin challenge”. In: *BMC Vet Res* 12.1 (2016), p. 241. DOI: 10.1186/s12917-016-0870-x.