

Single cell analysis of the developing mouse kidney provides deeper insight into marker gene expression and ligand-receptor crosstalk

Alexander N Combes^{1,2*+}, Belinda Phipson^{2,3*}, Kynan T Lawlor², Aude Dorison⁴, Ralph Patrick^{4,5}, Luke Zappia^{2,6}, Richard P Harvey^{4,5,7}, Alicia Oshlack^{1,6}, Melissa H Little^{1,2,3}.

1. Department of Anatomy & Neuroscience, University of Melbourne, Victoria, Australia.
2. Murdoch Children's Research Institute, Flemington Rd, Parkville, Victoria, Australia
3. Department of Paediatrics, The University of Melbourne, Victoria, Australia.
4. Developmental and Stem Cell Biology Division, Victor Chang Cardiac Research Institute, Darlinghurst, New South Wales, Australia
5. St. Vincent's Clinical School, University of New South Wales, Kensington, New South Wales, Australia
6. School of Biosciences, University of Melbourne, Victoria, Australia
7. School of Biotechnology and Biomolecular Science, University of New South Wales, Kensington, New South Wales, Australia

* Equal contribution

+ Corresponding authors:

A.N.C.: alexander.combes@unimelb.edu.au

M.H.L.: melissa.little@mcri.edu.au

Keywords: Kidney development, single cell RNA-seq, organogenesis, nephron progenitor, ureteric epithelium

Significance statement

This study provides deeper insight into cell type-specific gene expression and the mechanisms of progenitor maintenance and differentiation in the major lineages of the developing mouse kidney.

Abstract

Recent advances in the generation of kidney organoids and the culture of primary nephron progenitors from mouse and human have been based on knowledge of the molecular basis of kidney development in mice. While gene expression during kidney development has been intensely investigated, single cell profiling provides new opportunities to further subdivide component cell types and the signalling networks at play. Here, we describe the generation and analysis of 6732 single cell transcriptomes from the fetal mouse kidney (E18.5) and 7853 sorted nephron progenitor cells (E14.5). These datasets provide improved resolution of cell types and specific markers, including subdivision of the renal stroma and heterogeneity within the nephron progenitor population. Ligand-receptor interaction and pathway analysis reveals novel crosstalk between cellular compartments and associates new pathways with differentiation of nephron and ureteric epithelium cell types. We identify transcriptional congruence between the distal nephron and ureteric epithelium, showing that most markers previously used to identify ureteric epithelium are not specific. Together, this work improves our understanding of metanephric kidney development and provides a template to guide the regeneration of renal tissue.

Introduction

Mammalian kidney development has been studied using the mouse for over 70 years. The developing mammalian kidney consists of three main cell lineages, all of which derive from multipotent progenitors. *Foxd1*-expressing progenitors give rise to most cell types in the interstitial compartments (Kobayashi et al., 2014) aside from the stroma surrounding the ureter, which derives from *Tbx18*-expressing cells (Bohnenpoll et al., 2013). *Ret*-expressing ureteric tip (UT) cells give rise to the collecting duct and ureter (Chi et al., 2009). Finally, the filtration units of the kidney, the epithelial nephrons, arise from *Six2*-expressing nephron progenitor cells (Kobayashi et al., 2008). During kidney development, these progenitor populations signal to each other to ensure the ongoing expansion of the organ and accumulation of nephrons, with the resulting kidney containing approximately 1,000,000 nephrons in human and 16,000 in mouse (Bertram et al., 2011; Merlet-Benichou et al., 1999).

Our understanding of the molecular identity of cellular components within the mouse kidney is arguably richer than in almost any other organ system. Initial microarray analyses (Challen et al., 2005; Schmidt-Ott et al., 2005; Schwab et al., 2003; Stuart et al., 2003) were followed by some of the earliest profiling of laser-captured and sorted cell populations (Brunskill et al., 2008; Brunskill et al., 2011) with extensive section *in situ* hybridisation studies both validating compartment-enriched gene expression as well as further subsecting cellular domains (Georgas et al., 2009; Harding et al., 2011; Mugford et al., 2009; Thiagarajan et al., 2011). Anatomical and molecular comparisons of kidney development between human and mouse have now identified species-specific cell type markers within a developmental program that is largely conserved (Lindstrom et al., 2018b; Lindstrom et al., 2018c; Lindstrom et al., 2018d). With the advent of single cell RNA sequencing (scRNA-seq), such analyses of the developing mouse (Adam et al., 2017; Brunskill et al., 2014; Magella et al., 2017) and human (Lindstrom et al., 2018a; Lindstrom et al., 2018b; Menon et al., 2018; Wang et al., 2018; Young et al., 2018) kidney have given further insight into the cellular composition and molecular profiles of cells in both species. However, existing datasets have apparently not provided the transcriptional depth to identify the signalling pathways operating within the human fetal kidney and fail to detect several known ligand and receptor expression patterns in mouse. Understanding the signals involved in specifying renal progenitors in mouse has formed the basis of current human kidney organoid protocols (Little et al., 2016), and underpinned advances in culturing primary progenitor cells from both species (Brown et al., 2015; Li et al., 2016; Tanigawa et al., 2016; Yuri et al., 2017). A deeper understanding of

unique cell type-specific marker genes and the local signalling environment for all component cell types during mouse kidney development will provide avenues to optimise the maintenance and differentiation of nephron, stromal, and ureteric epithelium cell types from mouse and human cells for drug screening and disease modelling applications.

In this study, we used single cell profiling to interrogate cell types and gene expression within 6732 cells from three distinct E18.5 developing mouse kidney pairs, and 7853 sorted nephron progenitor cells from E14.5. By combining biological replication with robust clustering algorithms to define cell types, and rigorous statistical testing to determine differentially expressed cluster markers, we have generated an in-depth single cell view of the developing mouse kidney. Global analysis of receptor and ligand interactions within this dataset provides information with which to improve specification, maintenance, and maturation of renal cell types *in vitro*. Importantly, this dataset more deeply subdivides stromal subcompartments as well as better addressing the need for unique markers of specific nephron segments.

Results

Single cell profiling of the developing kidney identifies all major lineages and cell types including a stromal-nephron progenitor cluster

We sought to explore cell types and developmental programs in the late fetal mouse embryonic kidney (E18.5), a time at which all progenitor populations and most mature and maturing cell types co-exist. Using three independent kidney pairs captured in parallel using the 10x Chromium system, our aggregated data set consists of 6752 cells, 5639 of which passed quality control (see Methods), with a median of 2896 unique genes detected per cell. We used Seurat (Butler et al., 2018; Satija et al., 2015), to perform normalisation, variable gene selection and subsequent unsupervised clustering of cells, yielding 16 distinct whole kidney clusters (K0-K15, Fig. 1A). Following initial exploration of the data, we normalised for the effects of the three biological replicates as well as for cell cycle stage. After normalisation, an overlay of the three independent kidney data sets showed an even distribution of cells from each replicate amongst clusters, and visualisation of cell cycle state across the t-SNE projection illustrates no association between cell cycle state and any specific cluster after cell cycle normalisation (Fig. S1). TREAT tests from the edgeR package (McCarthy and Smyth, 2009; Robinson et al., 2010) were used to find genes that were

differentially expressed between cells in each cluster and all other cells (log-fold change>1, FDR<0.05). Genes that were enriched or specifically expressed in each cluster were cross-referenced to validated anchor genes and established markers to identify cell types (Fig. 1B, Table 1) (Georgas et al., 2009; Georgas et al., 2008; Thiagarajan et al., 2011). Entire gene lists were also compared to available kidney cell-type specific profiling using ToppGene (toppgene.cchmc.org) (Chen et al., 2009). This provided a provisional identification for all clusters (Fig. 1A-C). The number of clusters and key markers from our dataset are generally consistent with previous single cell analyses of the developing mouse kidney (Adam et al., 2017; Magella et al., 2017) though our analysis identifies more established marker genes per cluster (Fig. S1). Lists of differentially expressed genes for each cluster are provided in Table S1, and tSNE plots of key marker genes are displayed in Fig. S2. One or more clusters representing each of the major renal lineages - stroma, nephron, and ureteric epithelium - were identified in the data. Vascular endothelial and tissue-resident immune cell populations were also identified (Fig. 1A-C). We note that resident immune cells expressed *Bmp2* and *Tgfb1* while the endothelial cells expressed *Igf1*, *Igf2*, *Tgfb1*, *Notch1*, and *Notch 2*, which may influence cell-cell signalling within the developing kidney. Clusters corresponding to nephron progenitor cells, all major nephron segments, and a nephron progenitor-like cluster that co-expresses stromal markers including *Penk*, and *Col3a1* (Fig. 1A,B) were identified. Four additional populations with a stromal signature were identified, all expressing *Meis1*, *Col3a1*, and *Pdgfra*. These populations correspond to a cortical/nephrogenic zone stroma (cluster K2, *Meis1*⁺*Foxd1*⁺*Wnt4*⁻), medullary stroma (K4, *Meis1*⁺*Foxd1*⁻*Wnt4*⁻), collecting duct-associated stroma (K1, *Meis1*⁺*Wnt4*⁺*Wnt11*⁺), and a population marked by genes known to be expressed in several locations such as the cortical stroma, renal capsule, mesangium, smooth muscle cells and ureteric stroma (K13, *Meis1*⁺*Foxd1*⁺*Tbx18*⁺) indicating further heterogeneity within these clusters (Fig. S1). We next sought to examine potential signalling interactions between cell types.

Global analysis of putative ligand-receptor interactions

Known expression domains for key ligands and receptors from GDNF-RET, TGFB, Wnt, and FGF pathways were observed in expected cell types in our differential expression analysis (Fig. 2A). To investigate cell communication in the entire dataset we screened all cell types for a curated list of 2422 known and inferred receptor-ligand interactions (Ramilowski et al., 2015) adapted for use in single cell data (Farbehi et al., 2019). This identified >12000 potential interactions within and between the whole kidney clusters (Fig. 2B, Table S2). While interactions between some cell populations are implausible due to lack of proximity,

this provides an unbiased analysis of autocrine signalling and interactions between adjacent cell types. As an illustration of these results, we focus on potential paracrine interactions between the cortical stroma (K2 CS), nephron progenitor (K0 NP), and ureteric epithelium (K9 UE) cell clusters (Fig. 2C). This identifies known interactions between NP and UE through *Gdnf-Ret*, *Fgf*, *Bmp*, and *Wnt* ligands and receptors and identifies additional putative interactions through NP-produced *Rspo1* & *Rspo3*, and UE-produced *Nrtn*. We note that NP-produced *Fgf1* was identified to signal back to the UE. *Fgf1* is the most differentially expressed FGF ligand in the NP cluster within this data, though *Fgf20*, and lower levels of *Fgf8*, *Fgf9* and *Fgf10* were also detected. We previously identified *Fgf1* as a candidate driver of increased NP proliferation in a heterozygous knockout of *Six2* (Combes et al., 2018) and exogenous FGF1 promotes NP maintenance in culture (Brown et al., 2011). This analysis relies on a curated list of interacting factors (Ramilowski et al., 2015), which has some notable exceptions including *Wnt9b*, however expression of specific genes can be interrogated in Table S1. Signalling between the CS and NP populations is important for kidney development (Das et al., 2013; Fetting et al., 2014) but our understanding of the pathways underlying these interactions is incomplete. This analysis identifies potential interactions involving *Wnt5a* and *Bmp7*. These genes have kidney phenotypes on knockout and are expressed in the CS (*Wnt5a*) (Nishita et al., 2014) or influence organisation of the CS (*Bmp7*) (Oxburgh et al., 2004). Putative NP-CS interactions involving *Ntn1*, *Sfrp1*, *Fgf1*, *Fgf2*, *Pdgfr*, and *Ntf3* were identified (Fig. 2C). While the significance of putative interactions requires testing, these provide candidate pathways to improve nephron progenitor specification and maintenance *in vitro*.

Subclustering of ureteric epithelium cells identifies known subpopulations and established developmental trajectories

The ureteric epithelium in the developing mouse kidney has distinct zones of gene expression defining the tips, cortical, and medullary domains of this epithelium (Rutledge et al., 2017; Thiagarajan et al., 2011). Whole kidney cluster K9 expressed genes characteristic of the ureteric epithelium, including *Wnt11*, *Ret*, *Gata3*, and *Wnt9b*. Cells belonging to K9 were re-clustered resulting in the identification of three ureteric epithelium (U) sub-populations, with differential expression defining marker genes corresponding to tips (U0), cortical (U1), and medullary (U2) segments of the ureteric epithelium (Fig. 3A-C, Table S3) (Thiagarajan et al., 2011). The cluster enriched for medullary collecting duct marker genes also contained genes expressed in the urothelium of the renal pelvis (Fig. 3C) (Thiagarajan et al., 2011). Testing of Gene Ontology (GO), and Kyoto Encyclopaedia of Genes and Genomes (KEGG) annotations

identified major signalling pathways active in these subpopulations. This indicated the activity of several pathways known to be involved in ureteric tip development such as WNT, retinoic acid, TGF β , FGF, and Hippo signalling (Fig. 3D) (Reginensi et al., 2015; Yuri et al., 2017). This analysis also identified TGFB and PI3K-AKT pathways as active in the cortical collecting duct and phosphatidylinositol, PPAR, and Notch pathways as active in the medullary collecting duct and or urothelium (Fig. 2D). These represent candidate pathways for attempts to direct differentiation of mature collecting duct from progenitors of the ureteric epithelium. Where clustering attempts to group single cell transcriptomes into distinct cell types, pseudotime analysis involves ordering cells along a continuous trajectory that represents progress through differentiation. This is done by maximising the transcriptional similarity between successive pairs of cells, using dimensionality reduction and minimal spanning trees. Branches can occur along the trajectory when precursor cells make cell fate decisions that result in multiple subsequent lineages (Trapnell et al., 2014). Pseudotime analysis of cells from kidney cluster K9 using Monocle 2 (Qiu et al., 2017) replicated the established developmental trajectory from tip progenitor to cortical then medullary collecting duct and identified cohorts of genes that change during this progression (Fig. 3E,F).

Nephron lineage and relationships

Some nephron clusters within the whole kidney analysis represented multiple nephron populations such as cluster K3 (Fig. 1A), which co-expressed markers of the connecting segment (*Calb1*) and distal tubule (*Slc12a1*), which do not overlap in the embryo. We reclustered cells from the nephron lineage to gain further insight into nephron segments and subpopulations. This identified eight nephron (N) clusters representing established early nephron states and mature nephron segments (Fig. 4A) while markers including *Six2*, *Cited1*, and *Meox1*, defined a further five nephron progenitor clusters. All cells of the nephron arise from nephron progenitors via a mesenchyme to epithelial transition in response to WNT9B, produced at highest levels in the tip-stalk junction of the ureteric epithelium (Carroll et al., 2005)(Kobayashi et al, 2008). The first morphological sign of this transition is clustering of progenitors into a pretubular aggregate (PTA), marked by expression of *Wnt4* and *Tmem100* (Rumballe et al., 2011), which then forms a polarised epithelial renal vesicle (RV), marked by elevated levels of *Ccnd1*, *Jag1*, and *Fgf8* (Fig. 4B) (Georgas et al., 2009). Distinct proximal and distal gene expression is seen at RV with distal, medial and proximal segments are evident in the S-shaped body (SSB) (Fig. 4B) (Georgas et al., 2008). Podocytes are located in the proximal segment of the SSB (marked by *Mafb*) (Fig. 4B). The SSB matures into a

capillary loop nephron, which contains precursors for all major nephron segments, including a connecting segment (*Calb1*), distal tubule (*Slc12a1*), loop of Henle (*Umod*), proximal tubule (*Lrp2*, *Fbp1*), and podocyte-enriched glomerulus (*Mafb*, *Podxl*) (Fig. 4B).

Clusters representing all major nephron segments were present in the single cell data (Fig. 4A-C), with RV and SSB markers overlapping in cluster N3, and cluster N8 expressing markers of the distal SSB. Clusters representing the PTA, connecting segment, distal tubule and loop of Henle, early proximal tubule and proximal tubule, and podocytes were also identified (Fig. 4C). We refer to the ‘early proximal tubule’ (K5) cluster as such because it appears to represent a less mature version of the proximal tubule cluster (K8) however, K5 may also represent a distinct proximal tubule segment identity rather than a state of maturation. DE genes enriched were identified for each cluster (Table S4).

Pseudotime analysis was used to further interrogate nephron formation. This identified three main nephron states (Fig. 4D). An initial state combined nephron progenitors with early nephron up to SSB and podocytes. The trajectory subsequently forked into two arms representing the connecting segment and distal tubule on one arm, and the proximal tubule on the other (Fig. 4D). The split between proximal and distal tubule, and the association between distal tubule/connecting segment was anticipated (Georgas et al., 2009; Georgas et al., 2008). This positioning of podocytes between RV/SSB and the branch point between proximal and distal fates is different for the separate trajectory reported in human fetal kidney (Lindstrom et al., 2018a), or that more closely associated with proximal nephron as reported by Hochane et al (Hochane et al., 2019).

Mechanisms regulating nephron formation and maturation

Transcriptional regulation is a critical mechanism for determining and maintaining cell fate during development. Segment specific transcriptional regulators may facilitate direct reprogramming, as previously reported for nephron progenitor and proximal tubule (Hendry et al., 2013; Kaminski et al., 2016). The top differentially expressed transcription factors (TF) within each mouse nephron lineage cluster, including nephron progenitors, were identified (Fig. S3, Table S4), highlighting cell type-specific TFs such as *Six2* (nephron progenitor/early nephron), *Mafb* (podocytes), and *Hnf4a* (proximal tubule) (Kaminski et al., 2016; Thiagarajan et al., 2011). Signalling pathways identified as active within the nephron progenitor cluster include several pathways shown to regulate nephron progenitor fate *in vivo* such as PI3K-AKT, WNT, Hippo and MAPK signalling (Brown et al., 2015; Das et al., 2013; Karner et al.,

2011; Lindstrom et al., 2015; McNeill and Reginensi, 2017). Likewise, signalling pathways capable of triggering nephron formation, including Notch and TGF- β signalling (Brown et al., 2015; Chung et al., 2017), were identified in early nephron cell types (Fig. 4E, Fig. S3). Novel developmentally significant signalling pathways, including Hedgehog and JAK-STAT, were implicated by this analysis and may improve methods for maintaining isolated nephron progenitors (Brown et al., 2015; Li et al., 2016; Tanigawa et al., 2016). While nephron progenitor regulation and early nephron segmentation has been intensely studied, very little is known about the signals active in maturing nephron segments. Table S4 provides candidate signalling pathways that could be used to produce specific states of nephron maturation from primary nephron progenitor cells and in human kidney organoids. cAMP, cGMP-PKG, and insulin signalling associated with the distal and early proximal tubule while PPAR, AMPK, and glucagon signalling associated with proximal tubule.

Identification of Notch2+Spry2+ state and nephron progenitor stromal cluster

Nephrons derive from a self-renewing mesenchymal population (Boyle et al., 2008; Kobayashi et al., 2008). The nephron progenitor population is thought to be divided into a *Six2*+*Cited1*+ uncommitted and *Six2*+*Cited1*⁻ committing state (Brown et al., 2015; Mugford et al., 2008). However, previous timelapse imaging of kidney morphogenesis has revealed substantial cell movement (Combes et al., 2016; Lawlor et al., 2019; O'Brien et al., 2018) and variation in cell cycle length (Short et al., 2014) within the nephron progenitor population suggesting it may be more heterogeneous than previously thought. As noted above (Fig. 4A), five nephron progenitor populations were identified expressing *Six2*, *Cited1* and *Meox1*. Two of these clusters (N7 and N9) appeared to be driven by cell cycle genes (e.g. *Cenpa*, *Cenpf*, *Pclaf*, *Top2a*), with top DE genes also relating to cell cycle. These clusters are likely driven by cycle profile not accounted for in the cell cycle normalisation. Cell cycle clusters N7 and N9 displayed a partially committed phenotype with low expression of *Wnt4* and *Tmem100*. This could associate cell division with priming for commitment or reflect the increase in cell proliferation seen in committing nephron progenitor cells (Short et al., 2014). The three remaining nephron progenitor clusters expressed cluster specific DE markers (Fig. 5A). We define these as 1) 'uncommitted' (cluster N0) with the highest levels of *Cited1* and *Meox1* and little to no expression of *Wnt4* and *Tmem100*. 2) 'primed' (N6) with lower levels of *Six2* and *Cited1*, expression of *Notch2* and *Sprouty2* and low levels of renal vesicle marker *Jag1* and 3) a nephron progenitor-stromal population (N10) with modest expression of commitment markers and stromal characteristics, including expression of *Pdgfra* and *Col3a1* (Fig. 5A). Cluster N6 expressed *Sprouty2*, a negative regulator of FGF signalling, with FGF signalling

associated with nephron progenitor maintenance (Walker et al., 2016). Notch signalling has recently been shown to regulate early commitment to nephron formation (Chung et al., 2016; Chung et al., 2017). Hence, cluster N6 may represent a transitional state between nephron progenitor and PTA, N4 (Fig. 5A).

Pseudotime analysis of all cells within early nephron clusters (N0 NP to N4 PTA, Fig. 5B) reproduced the expected developmental trajectory from nephron progenitor to PTA. Cell cycle-associated cluster NP7 grouped with PTA cluster N4, representing a more committed state (Fig. 5B). Cell cycle-associated cluster N9 and a putative ‘primed’ cluster N6 were distributed along the entire trajectory. As such, N6 may represent a transient state that NP cells cycle through or reflect cells that are positioned adjacent to the stroma or ureteric tip at any point in time. The nephron progenitor / stromal state N10 diverged from this main trajectory after the undifferentiated nephron progenitor state. Of note, cells from the N6 NP cluster were also present with the N10 NP-STR cluster (Fig. 5B)

Sorted nephron progenitors recapitulate NP cell states identified in whole kidney

To gain a deeper insight into nephron progenitor sub populations within the developing mouse kidney, 7853 *Six2*GFP⁺ nephron progenitor cells from three pooled replicates of E14.5 kidney were isolated using fluorescence activated cell sorting from the *Six2*^{GCE} mouse line (Kobayashi et al., 2008) and profiled using scRNAseq. Sorted cells were combined with all nephron progenitor cells from the whole kidney analysis and clustered using Seurat’s dataset integration approach (Butler et al., 2018) (Fig. S4). We refer to the resulting clusters as nephron progenitor (NP) clusters 0-9 (NP0-NP9, Fig. 5C). Clusters that did not relate to cell cycle were marked by the same differentially expressed genes observed in the whole kidney nephron progenitor sub clusters. Cluster NP0 displayed increased expression of uncommitted progenitor genes such as *Cited1*, NP3 had increased levels of *Notch2* and *Spry2*, cluster NP4 represented a committing state with increased levels of *Wnt4* and *Tmem100*, and a nephron progenitor-stromal cluster (NP7) remained. Cells from both whole kidney and sorted NPs were present in all clusters (Fig. 5D, Fig. S4). While NP cells across all clusters expressed stromal markers (*Meis1*, *Lgals1*, and *Meg3*), *Pdgfra* was enriched in NP7 (Fig. 5D).

Trajectory analysis of nephron progenitors from clusters NP0, NP3, NP4, and NP7 (Fig. 5E) reproduced the trajectory analysis of nephron progenitors from the whole kidney data (Fig. 5B). See Table S5 for cluster markers and DE genes for NP0, NP3, NP4 and NP7. This larger dataset reinforced the nephron progenitor sub-populations identified in the whole mouse kidney.

The nephron progenitor-stromal cluster may be the result of a technical artefact

Nephron progenitors and stromal progenitors arise from the same lineage before the onset of nephron formation (Brunskill et al., 2014; Mugford et al., 2008) but are not thought to cross lineage after this time (Kobayashi et al., 2014; Naiman et al., 2017). Stochastic expression of stromal markers in nephron progenitor cells has been reported at the single cell level (Magella et al., 2017), and expression of stromal markers *Foxd1* and *Meis1* within our analysis suggests this may be more than random expression (Fig 5D). However, NP7 represented a small but distinct cell cluster expressing both nephron progenitor and a broad range of stromal markers. This combined profile could represent i) a technical artefact where stromal and nephron progenitor cells are labelled by a single barcode creating a ‘doublet’, ii) the existence of a genuine *in vivo* cell population transitioning between NP and stromal lineages, or iii) an artefactual change in progenitor identity upon dissociation. No other populations were observed with mixed signatures, but doublet finding algorithms (doubletCells and doubletCluster functions in the scan package) did identify this cluster within the whole kidney data (K14) as having an increased probability of containing doublets (Fig. S4).

Lineage tracing was performed to investigate the possibility of these cells representing a transitional state. Using a constitutively active *Six2*Cre (*Six2*TGC) *Six2*-derived cells were observed in the cortical and medullary stroma in all samples (Fig. 5F). However, as this *Six2*-Cre is active from E11.5 or earlier, labelled stromal cells may reflect the early plasticity between stromal and nephron lineages rather than continued transdifferentiation. Using an inducible *Six2*Cre (*Six2*GCE, induced from E12.5) to assess NP contributions to stroma after the establishment of the proposed lineage boundary did result in rare *Six2*-derived cells in the nephrogenic zone that did not express SIX2 protein, but labelled cells were observed at a frequency lower than expected based on NP7 cluster size and most labelled cells were unusually small, suggesting they may be undergoing apoptosis (Fig. 5G). Evidence of lineage transition was also observed deeper in the kidney. Lineage tracing from an inducible *Pdgfra* stromal cre activated at E13.5 and assessed at E18.5 did not label cells within the nephron progenitor population or nephron lineage (Fig. 5H). Hence, stromal cells do not appear to transition to nephron progenitor fate. SIX2 antibody staining did not overlap with a transgenic mouse line expressing nuclear *Pdgfra*-GFP (Fig. 5I) despite transcripts for *Six2* and *Pdgfra* being co-expressed in the scRNA-Seq data. A genuine discrepancy between mRNA and protein expression is improbable, as the reporters used drive inducible Cre expression from the native *Pdgfra* and *Six2* promoters and therefore should evade mechanisms targeted at

preventing production of proteins from the other lineage. Cumulatively, these data affirm the current model of boundaries between nephron and stromal lineages after early kidney development. While this mixed signature could represent transcriptional confusion induced by dissociation, increased library size and a merged signature supports selective doublets between nephron progenitor and stromal cells. It remains unclear why this doublet was enriched however this may suggest differential cell-cell adhesion between these states.

Defining stromal subpopulations within developing kidney

Interrogating the role of stromal subpopulations in kidney development has been hampered by a lack of understanding of specific markers of these populations. Although ontological terms were defined for distinct anatomical regions of the kidney stroma (Little et al., 2007), definitive markers for such regions have been less well defined. Adam et al (2017) and Magella et al (2017) identified 3 stromal clusters and regionally assigned them (cortical, medullary, mesangial) with respect to *in situ* hybridisation data from the Allen Brain Atlas. Stromal cell types and signalling from these are critical to normal kidney development (Li et al., 2014). Reclustering of stromal clusters identified seven stromal lineage (S) clusters (Fig. 6A-C). *In situ* hybridisation data from the Allen Developing Mouse Brain Atlas was mined to map the expression domains of cluster markers (Fig. S5). Stromal clusters S0 and S4 are marked by several genes expressed in the cortical and nephrogenic zone stroma (*Foxd1*, *Ntn1*, *Ighbp5*, *Aldh1a2*, *Gdnf*). Cluster S4 revealed a cell cycle signature, likely representing cells within the same region as S0 that are proliferating. Clusters S2 and S3 (*Alx1*, *Wnt4*, *Nkd1*, *Wnt11*) represent the *Alx1*⁺ collecting duct-associated stroma. S2 may reflect proliferating cells within S3 as the majority of genes that differ between these clusters relate to cell cycle, with the notable exception of *Ren1*, which may identify this cluster as perivascular. Markers within cluster S1 have a heterogeneous expression with overlap in the medullary region. The S5 population expresses markers of vascular associated smooth muscle cells and pericytes (*Angpt1*, *Angpt2*, *Mef2c*, *Pdgfrb*, *Cspg4/Ng2*, *Ren1*, *Gata3*). DE genes for cluster S6 included both genes such as *Tbx18*, with established profiles in the stroma surrounding the ureter (Airik et al., 2006), and *Dlk1*, *Igf1*, and *CD34*, suggesting vascular-associated cells (Fig. 6B-C, Fig. S5). The DE genes from this analysis will aid in characterising stromal populations in the developing kidney (Table S6). Further integration of scRNA-seq datasets such as this one with emerging spatial transcriptomics methods (Stahl et al., 2016) will also aid in defining more precise regions and cell types within the stroma.

Congruence between markers of the ureteric epithelium and distal nephron

In the process of defining cluster identities, a strong congruence between markers of the UE and distal nephron was observed. Most established markers of the ureteric epithelium, such as *Hoxb7*, *Gata3*, *Calb1*, *Krt8*, *Krt18*, *Krt19*, and *Aqp2*, were also expressed in the distal nephron, albeit at lower levels. Likewise, nephron markers such as *Cdh16*, *Mal*, *Spp1*, and *Spint2* were expressed in the ureteric epithelium (Fig. 7A). Indeed, over half of the top 30 DE genes in either distal nephron or UE were expressed in both clusters. This has significant implications in the directed differentiation of pluripotent cells to kidney organoids, which has relied upon many of these markers to identify collecting duct versus distal nephron.

To check that these results were not due to inappropriate clustering of UE cells, we re-examined the presence of GATA3 protein within the distal nephron segments (connecting segment / distal tubule) *in vivo* using antibody staining and lineage tracing driven by a nephron progenitor-specific *Six2*-Cre mouse line (*Six2*TGC) (Kobayashi et al., 2008). As expected, all connecting segments and distal regions of the nephron tubules were derived from the nephron lineage but these structures clearly express GATA3 protein (Fig. 7B). Indeed *Hoxb7*, an established marker of the UE, was most highly expressed in UE but also in distal nephron (Fig. 7C) and some endothelial cells (not shown). Expression of GATA3 and *Hoxb7*-GFP in the distal nephron has likely been previously overlooked as *in situ* hybridisation and immunofluorescence focus on sites of highest expression.

Comparing genes upregulated in the connecting segment (N12) and UE (K9) clusters, and checking these against relevant clusters in the nephron and UE lineage clustering identified 36 genes representing markers specific to the connecting segment and/ or expressed more broadly in the nephron lineage that could be used in combination with *Gata3* expression to distinguish connecting segment from UE. Likewise, 29 UE genes not expressed in the distal nephron were identified (Table S7).

Discussion

The developing mouse kidney represents an invaluable tool with which to understand the formation and maturation of each renal cell type. The single cell data presented here offers a unique opportunity to understand the mechanisms of progenitor maintenance and differentiation in the stroma, the ureteric epithelium, and the nephron lineages. Dynamic changes in gene expression and signalling pathway activity from progenitor to mature cell type provide a roadmap of the signals that regulate progenitor maintenance and differentiation *in vivo*. Likewise, global analysis of receptor-ligand interactions between all cell clusters in the whole kidney identified potential novel interactions and interactions known to play a significant role in kidney development.

Previous scRNA-Seq analyses of developing mouse kidney have been performed at E11.5-E14.5 and postnatal day (P) 1 (Adam et al., 2017; Brunskill et al., 2014; Magella et al., 2017) (Table S8). This study examined E18.5, a developmental stage that contains a broader complement of cell types compared to E11-14.5 but precedes the cessation of nephrogenesis initiating at P1 (Hartman et al., 2007; Rumballe et al., 2011). While the DE genes identified here correlate with these previous studies this dataset provides a deeper insight into cluster-specific gene expression identifying both anticipated receptors / ligand expression patterns and revealing novel relationships. For example, while >20,000 cells were profiled at P1 (Adam et al., 2017), several known signalling molecules with functionally validated roles in the nephrogenic niche such as *Gdnf*, *Fgf20*, *Fgf9*, *Bmp7*, *Wnt4* and *Fgf8*, were not detected in that analysis, precluding further insight into signalling interactions. The cross-platform study conducted at E14.5 (Magella et al., 2017) provided insight into some novel signalling interactions, including *Gdnf* expression in the nephrogenic zone stroma, but expression of genes encoding key ligands such as *Gdnf*, *Fgf9*, or *Bmp7* did not feature in the nephron progenitor population, perhaps favouring detection of ligands with more restricted expression patterns such as *Fgf20*. The improved resolution of gene expression in our study may be due to sequencing depth (~3000 genes detected per cell), biological replication, and differential expression analysis with the edgeR method, which has recently been shown to be a top performer in a comparison of 36 differential expression analysis methods for scRNA-Seq data (Soneson and Robinson, 2018). Critically, we have use *in vivo* gene expression and lineage tracing studies validate or dismiss novel compartments.

This analysis identifies heterogeneity within the nephron progenitor population with a *Six2/Cited1* high undifferentiated state, a moderate *Six2/Cited1* expression cluster co-expressing *Notch2* and *Spry2*, and a *Six2* moderate *Cited1* low/off cluster with upregulated expression of early commitment markers (*Wnt4*, *Tmem100*) potentially representing PTA. These clusters reflect previously described undifferentiated, primed and PTA clusters based on regionally restricted expression of markers such as *Six2*, *Cited1*, *Dpf3*, and *Meox1* (Brown et al., 2015; Georgas et al., 2009; Mugford et al., 2009). In contrast to previous work on nephron progenitor subpopulations, *Cited1* was not absent prior to the upregulation of pretubular aggregate genes though *Cited1* levels were reduced between the ‘undifferentiated’ and ‘primed’ populations. Additional cell cycle associated nephron progenitor clusters were also identified, but pseudotime analysis suggests they are dividing cells within other NP populations. A nephron progenitor-stromal cluster was identified by clustering and pseudotime analyses but not supported by subsequent lineage tracing or protein co-localisation experiments. Again, this may reflect a difference in the range of expression levels detected by this analysis versus those evident by *in situ* hybridisation or immunofluorescence. Changes in expression patterns between nephron progenitor subpopulations were graded rather than sharp, perhaps reflecting smooth transitions between states. Further work will be required to determine whether these subpopulations correlate to distinct anatomical regions within the nephrogenic niche.

Human kidney organoids contain epithelial, stromal, and endothelial cell types with transcriptional congruence to equivalent populations in the human fetal kidney (Combes et al., 2019). However, our ability to interpret the cellular composition and authenticity of engineered renal tissue depends on our understanding of the markers that define a particular cell type or state of maturation *in vivo*. Likewise, our capacity to generate a cell type depends on knowledge of the programs that specify and maintain cellular identity. Here we identify a strong transcriptional congruence between the UE and the distal nephron, validating the expression of *GATA3* and *HOXB7* (*Hoxb7GFP*) in the murine distal nephron, two markers previously thought to be specific to the UE. While expression of UE markers such as *Calb1* have been documented in the distal nephron before (Georgas et al., 2008), the extent of the similarities between these cell types has not been fully appreciated. Emerging scRNAseq studies of human fetal kidney identify *GATA3*, *KRT8*, *KRT18*, *KRT19*, *WFDC2* and *CDH16* as expressed in human distal nephron and UE clusters (Wang et al., 2018). More definitive UE markers, such as *RET* and *WNT11* (when co-expressed with *GATA3*), were not detected despite these genes known to be expressed in human kidneys (Rutledge et al., 2017). We have

previously described the formation of UE within kidney organoids based upon co-staining for PAX2⁺ ECAD⁺ GATA3⁺ KRT8⁺ and DBA⁺ (Takasato et al., 2015). Indeed, recent lineage tracing experiments within such kidney organoids confirmed nephron epithelium as arising from *SIX2*-expressing cells, but not this presumptive GATA3⁺ UE (Howden et al., 2019). In contrast, Taguchi et al propose that the UE is derived from anterior intermediate mesoderm and should not arise simultaneously with the metanephric mesenchyme (Taguchi et al., 2014). We now show that the markers previously used to define UE in our kidney organoids are not specific to UE. While this leaves the identity of this epithelium undefined, it provides the field with specific UE markers with which to improve protocols.

In summary, this study provides the most comprehensive reference of cell-type specific expression within the developing kidney to date, associating known and new signalling molecules and pathways with specific cell types. As such, this data represents a roadmap with which to improve *in vitro* models of the developing kidney.

Materials and Methods

Mouse Strains and Embryo Staging

In mouse experiments, noon of the day on which the mating plug was observed was designated embryonic day (E) 0.5. C57Bl/6 mice were used for the E18.5 embryonic kidney analysis. E14.5 *Six2*GCE mice were used for the sorted NP cell analysis. Sample gender was not determined prior to analysis. Mouse lines used were: *Six2*TGC (Tg(*Six2*-EGFP/cre)1Amc, Jackson labs reference (JAX):009606), *Six2*GCE (*Six2*^{tm3(EGFP/cre/ERT2)} JAX:009600) (Kobayashi et al., 2008); *Pdgfra*MerCreMer (CDB0674K RIKEN Center for Life Science Technologies) (Ding et al., 2013); LSLTdTomo (Gt(*ROSA*)26Sor^{tm9(CAG-tdTomato)} JAX:007909) (Madisen et al., 2010); *Hoxb7*-GFP (Tg(*Hoxb7*-EGFP)33Cos JAX:016251) (Srinivas et al., 1999); and *Pdgfra*GFP (*Pdgfra*^{tm11(EGFP)Sor} JAX_007669) (Hamilton et al., 2003). All animal experiments were approved by the Murdoch Children's Research Institute Animal Ethics Committees and conducted under Australian guidelines for the care and use of animals for scientific purposes.

Immunofluorescence and Microscopy

E18.5 embryonic kidneys were fixed in 4% PFA for 20 minutes, washed in PBS and cleared using the PACT method (Yang et al., 2014) to preserve tdTomato or GFP fluorescence. Cleared samples were stained using rabbit anti-SIX2 (1:600 Proteintech, 11562-1-AP), Goat anti-GATA3 (1:600 R&D Systems, AF2605), or mouse anti-Cytokeratin (1:300 Abcam, ab115959) and Alexa Fluor 488 and/or 647 labelled secondary antibodies (1:600 Thermo Fisher). Antibodies previously used in (Combes et al., 2018; Combes et al., 2019). Samples were blocked in PBST (PBS + 0.1% Triton-X) with 10% normal donkey serum and incubated at room temperature with each antibody solution for at least 48 hours followed by washing for 24 hours in PBST. Nuclei were stained using Draq5 (Abcam). Samples were mounted in

RIMS (88% Histodenz) and imaged using an Andor Dragonfly spinning disk system with a 40 μ m pinhole disk and Nikon 1.15NA 40x water-immersion objective. Images were processed in Fiji (Schindelin et al., 2012).

Single cell sample prep and sequencing

Mouse kidneys were dissected into ice cold PBS then digested over 15 minutes at 37C in Accutase (#A1110501 Life technologies), with manual dissociation via pipetting through a P1000 tip every 5 minutes. Following dissociation, cells were passed through a 30 micron filter and stored on ice in 50% PBS, 50% DMEM with 5% FCS. Three pairs of 18.5 dpc mouse kidneys were run in parallel on a chromium 10x Single Cell Chip (10x Genomics). Kidneys from multiple litters of *Six2*GFP⁺ (*Six2*^{GCE} Kobayashi et al., 2008) mouse embryos at E14.5 were pooled into three replicate tubes and dissociated in parallel by the same protocol. *Six2*GFP⁺ cells were isolated using gates for *Six2*GFP fluorescence, propidium iodide to exclude dead cells, and size to exclude cell debris and doublets. Isolated *Six2*GFP⁺ cells were collected and stored on ice in 50% PBS, 50% DMEM with 5% FCS, then run in parallel on a 10x chip. Libraries were prepared using Chromium Single Cell Library kit V2 (10x Genomics), and sequenced on an Illumina HiSeq using 100bp paired-end sequencing.

Data submission information

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar *et al.*, 2002) and are accessible through GEO Series accession numbers: E18.5 whole kidney data GSE108291; E14.5 sorted nephron progenitor data GSE130606.

Single cell data analysis

For the whole mouse kidney samples, raw sequencing data was processed using Cell Ranger (v1.3.1, 10x Genomics) to produce gene-level counts for each cell in each sample, which were aggregated to form a single matrix of raw counts for 6752 cells. All subsequent analysis was performed in the R statistical programming language. Cells with greater than 95 per cent of genes with zero assigned reads were removed, leaving 5639 cells for further analysis. Genes with zero counts in more than 5589 cells (assuming a minimum cluster size of 50 cells), mitochondrial and ribosomal genes, and genes without annotation were also filtered out. The final dataset used for analysis consisted of 5639 cells and 13116 genes. The Seurat package (v2.0.1) (Macosko et al., 2015; Satija et al., 2015) was used to normalise data, regressing out factors related to biological replicate and cell cycle. For clustering, 1962 highly variable genes were selected and the first 30 principal components based on those genes used to build a graph, which was segmented with a resolution of 0.8. This identified 16 clusters across the 5639 cells. We obtained lists of differentially expressed genes for each cluster by testing for genes that had an absolute log fold change greater than one between cells in each cluster compared with the remaining cells using the glmTreat method in the edgeR package (Robinson et al., 2010). To identify corresponding cell types we focussed on genes that were significantly up-regulated in each cluster. In addition, we used pathway analysis to aid our interpretation, including GO and KEGG analysis, which was performed with limma (Ritchie et al., 2015), as well as pathway analysis using the ToppGene suite (Chen et al., 2009). Trajectory analysis of the various lineages was performed using Monocle (v2.4.0) (Qiu et al., 2017; Trapnell et al., 2014).

For the sorted cap mesenchyme data, raw sequencing data was processed using Cell Ranger as above. The 7853 cells all had fewer than 95 per cent of genes with zero assigned reads. Cells with low diversity were removed, leaving 7844 cells for further analysis. Gene filtering proceeded as described above, leaving 12344 genes for further analysis. To identify clusters within the nephron progenitor population, we performed an integrated analysis of the cells from the sorted cap mesenchyme and the nephron progenitor populations identified in the whole mouse kidney dataset, represented by clusters 0, 4, 6 and 10 in the nephron lineage. This was done using the alignment technique in the Seurat package. For both datasets, biological replicate, cell cycle and the total UMI counts were regressed out using the ScaleData function in Seurat. The two datasets were merged using canonical correlation analysis on 2187 highly variable genes and 20 canonical correlation vectors. Ten clusters were identified using 20 canonical correlation vectors and the resolution parameter set to 0.6. Marker genes for the 10 clusters were defined using Wilcoxon rank sum tests in the Seurat package. Five of the 10 clusters showed strong cell cycle related expression patterns (clusters 1, 2, 5, 6 and 9), while cluster 8 had high immune cell markers. Clusters 0, 3, 4 and 7 corresponded to clusters 0, 6, 4 and 10 respectively in the nephron lineage reclustering of the whole mouse kidney dataset, hence validating these clusters in a much larger dataset. Focusing on these four clusters, differential expression analysis with edgeR and glmTreat (fold-change threshold of 20%) was performed, further refining the marker gene lists for these populations. Trajectory analysis of the cells in these four clusters was performed using Monocle (Qiu et al., 2017), which identified three states.

Ligand-receptor interactions

Ligand-receptor interaction analysis was performed according to the approach described previously (Farbehi et al., 2019). Briefly, a weighted directed graph was built linking ‘source’ cell types, defined by expression of a ligand, to ‘target’ cell types expressing a corresponding receptor, after reference to a curated map of human ligand-receptor pairs (Ramilowski et al., 2015). Source-ligand and receptor-target edges were weighted according to expression fold-change in ligands and receptors, respectively. Ligand-receptor edges were weighted according to mouse-specific protein-protein association scores from STRING (Szklarczyk et al., 2017). Significant cell-cell connections were determined by network permutation testing (100,000 permutations, $P_{adj} < 0.01$).

Doublet analysis

We ran two doublet detection algorithms available in the scran Bioconductor package (Lun et al., 2016) on the whole mouse kidney dataset as the nephron progenitor-stromal cluster (cluster 14) proved difficult to validate with subsequent experiments. First we ran the doubletCluster function in scran which aims to identify clusters that have intermediate expression profiles of two other clusters (Bach et al., 2017). Every possible trio of clusters (the query cluster and its two “parents”) were examined, and a number of statistics computed providing support for the cluster arising from doublet cells. This analysis ranked cluster 14 as the most likely to contain doublets, with the parent clusters identified as clusters 4 (medullary stroma) and 12 (pretubular aggregate). Cluster 14 had very few unique marker genes (N=11), had cells with much larger library sizes compared to cells in clusters 12 and 4, and the proportion of cells belonging to cluster 14 was low (1.4%), providing further evidence for

doublets. In addition we also ran the doubletCells function, which simulates doublets from the single cell expression profiles (Dahlin et al., 2018). Thousands of doublet cells are simulated by adding together two randomly chosen single-cell profiles, ignoring clustering information. Each original cell is then compared to the simulated doublets, as well as the observed cells, and a doublet score is computed for each cell. High scores indicate a greater likelihood that the cells are doublets. Once more, cluster 14 was flagged as comprising of doublet cells as the majority of cells had high doublet scores.

Author contributions

B.P. performed the single cell mapping, normalisation, differential expression and trajectory analysis with some assistance from LZ, under the supervision of AO. RP developed the ligand-receptor analysis, implemented by KTL. KTL and AD performed additional data analysis and experimental validation. RPH provided resources and supervised AD and RP. ANC designed and performed experiments and was primarily responsible for interpretation and presentation of the data, and writing. MHL provided interpretation of data and manuscript writing. All authors contributed to manuscript revisions.

Acknowledgements

The authors thank Dr Hiroshi Kataoka for access to the *Pdgfra*CreER line. Single cell sequencing was performed at the Australian Genome Research Facility Genomics Innovation Hub with the assistance of J. Jabbari. Microscopy was performed at the Murdoch Children's Research Institute.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by seed funding from the Murdoch Children's Research Institute, the University of Melbourne, and grants from the Australian Research Council (DE150100652 to A.N.C.), National Health and Medical Research Council (GNT1156567 to A.N.C, A.O, M.H.L.). M.H.L. is a Senior Principal Research Fellow of the NHMRC (GNT1136085). AO is a Career Development Fellow of the NHMRC (APP1126157). LZ is supported by an Australian Government Research Training Program (RTP) Scholarship. MCRI is supported by the Victorian Government's Operational Infrastructure Support Program.

References

- Adam, M., Potter, A. S. and Potter, S. S. (2017). Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. *Development* **144**, 3625-3632.
- Airik, R., Bussen, M., Singh, M. K., Petry, M. and Kispert, A. (2006). Tbx18 regulates the development of the ureteral mesenchyme. *The Journal of clinical investigation* **116**, 663-674.
- Bach, K., Pensa, S., Grzelak, M., Hadfield, J., Adams, D. J., Marioni, J. C. and Khaled, W. T. (2017). Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. *Nature communications* **8**, 2128.
- Bertram, J. F., Douglas-Denton, R. N., Diouf, B., Hughson, M. D. and Hoy, W. E. (2011). Human nephron number: implications for health and disease. *Pediatric nephrology* **26**, 1529-1533.
- Bohnenpoll, T., Bettenhausen, E., Weiss, A. C., Foik, A. B., Trowe, M. O., Blank, P., Airik, R. and Kispert, A. (2013). Tbx18 expression demarcates multipotent precursor populations in the developing urogenital system but is exclusively required within the ureteric mesenchymal lineage to suppress a renal stromal fate. *Developmental biology* **380**, 25-36.
- Boyle, S., Misfeldt, A., Chandler, K. J., Deal, K. K., Southard-Smith, E. M., Mortlock, D. P., Baldwin, H. S. and de Caestecker, M. (2008). Fate mapping using Cited1-CreERT2 mice demonstrates that the cap mesenchyme contains self-renewing progenitor cells and gives rise exclusively to nephronic epithelia. *Developmental biology* **313**, 234-245.
- Brown, A. C., Adams, D., de Caestecker, M., Yang, X., Friesel, R. and Oxburgh, L. (2011). FGF/EGF signaling regulates the renewal of early nephron progenitors during embryonic development. *Development* **138**, 5099-5112.
- Brown, A. C., Muthukrishnan, S. D. and Oxburgh, L. (2015). A synthetic niche for nephron progenitor cells. *Developmental cell* **34**, 229-241.
- Brunskill, E. W., Aronow, B. J., Georgas, K., Rumballe, B., Valerius, M. T., Aronow, J., Kaimal, V., Jegga, A. G., Yu, J., Grimmond, S., et al. (2008). Atlas of gene expression in the developing kidney at microanatomic resolution. *Developmental cell* **15**, 781-791.
- Brunskill, E. W., Georgas, K., Rumballe, B., Little, M. H. and Potter, S. S. (2011). Defining the molecular character of the developing and adult kidney podocyte. *PloS one* **6**, e24640.
- Brunskill, E. W., Park, J. S., Chung, E., Chen, F., Magella, B. and Potter, S. S. (2014). Single cell dissection of early kidney development: multilineage priming. *Development* **141**, 3093-3101.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology* **36**, 411-420.
- Carroll, T. J., Park, J. S., Hayashi, S., Majumdar, A. and McMahon, A. P. (2005). Wnt9b plays a central role in the regulation of mesenchymal to epithelial transitions underlying organogenesis of the mammalian urogenital system. *Developmental cell* **9**, 283-292.
- Challen, G., Gardiner, B., Caruana, G., Kostoulas, X., Martinez, G., Crowe, M., Taylor, D. F., Bertram, J., Little, M. and Grimmond, S. M. (2005). Temporal and spatial transcriptional programs in murine kidney development. *Physiological genomics* **23**, 159-171.
- Chen, J., Bardes, E. E., Aronow, B. J. and Jegga, A. G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic acids research* **37**, W305-311.
- Chi, X., Michos, O., Shakya, R., Riccio, P., Enomoto, H., Licht, J. D., Asai, N., Takahashi, M., Ohgami, N., Kato, M., et al. (2009). Ret-dependent cell rearrangements in the Wolffian duct epithelium initiate ureteric bud morphogenesis. *Developmental cell* **17**, 199-209.
- Chung, E., Deacon, P., Marable, S., Shin, J. and Park, J. S. (2016). Notch signaling promotes nephrogenesis by downregulating Six2. *Development*.
- Chung, E., Deacon, P. and Park, J. S. (2017). Notch is required for the formation of all nephron segments and primes nephron progenitors for differentiation. *Development*.
- Combes, A. N., Lefevre, J. G., Wilson, S., Hamilton, N. A. and Little, M. H. (2016). Cap mesenchyme cell swarming during kidney development is influenced by attraction, repulsion, and adhesion to the ureteric tip. *Developmental biology* **418**, 297-306.

- Combes, A. N., Wilson, S., Phipson, B., Binnie, B. B., Ju, A., Lawlor, K. T., Cebrian, C., Walton, S. L., Smyth, I. M., Moritz, K. M., et al. (2018). Haploinsufficiency for the *Six2* gene increases nephron progenitor proliferation promoting branching and nephron number. *Kidney Int* **93**, 589-598.
- Combes, A. N., Zappia, L., Er, P. X., Oshlack, A. and Little, M. H. (2019). Single cell analysis reveals congruence between kidney organoids and human fetal kidney. *Genome Medicine* **11**.
- Dahlin, J. S., Hamey, F. K., Pijuan-Sala, B., Shepherd, M., Lau, W. W. Y., Nestorowa, S., Weinreb, C., Wolock, S., Hannah, R., Diamanti, E., et al. (2018). A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. *Blood* **131**, e1-e11.
- Das, A., Tanigawa, S., Karner, C. M., Xin, M., Lum, L., Chen, C., Olson, E. N., Perantoni, A. O. and Carroll, T. J. (2013). Stromal-epithelial crosstalk regulates kidney progenitor cell differentiation. *Nature cell biology* **15**, 1035-1044.
- Ding, G., Tanaka, Y., Hayashi, M., Nishikawa, S. and Kataoka, H. (2013). PDGF receptor alpha+ mesoderm contributes to endothelial and hematopoietic cells in mice. *Developmental dynamics : an official publication of the American Association of Anatomists* **242**, 254-268.
- Farbehi, N., Patrick, R., Dorison, A., Xaymardan, M., Janbandhu, V., Wystub-Lis, K., Ho, J. W. K., Nordon, R. E. and Harvey, R. P. (2019). Single-cell expression profiling reveals dynamic flux of cardiac stromal, vascular and immune cells in health and injury. *eLife* **8**.
- Fetting, J. L., Guay, J. A., Karolak, M. J., Iozzo, R. V., Adams, D. C., Maridas, D. E., Brown, A. C. and Oxburgh, L. (2014). *FOXD1* promotes nephron progenitor differentiation by repressing decorin in the embryonic kidney. *Development* **141**, 17-27.
- Georgas, K., Rumballe, B., Valerius, M. T., Chiu, H. S., Thiagarajan, R. D., Lesieur, E., Aronow, B. J., Brunskill, E. W., Combes, A. N., Tang, D., et al. (2009). 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. *Developmental biology* **332**, 273-286.
- Georgas, K., Rumballe, B., Wilkinson, L., Chiu, H. S., Lesieur, E., Gilbert, T. and Little, M. H. (2008). Use of dual section mRNA in situ hybridisation/immunohistochemistry to clarify gene expression patterns during the early stages of nephron development in the embryo and in the mature nephron of the adult mouse kidney. *Histochemistry and cell biology* **130**, 927-942.
- Hamilton, T. G., Klinghoffer, R. A., Corrin, P. D. and Soriano, P. (2003). Evolutionary divergence of platelet-derived growth factor alpha receptor signaling mechanisms. *Molecular and cellular biology* **23**, 4013-4025.
- Harding, S. D., Armit, C., Armstrong, J., Brennan, J., Cheng, Y., Haggarty, B., Houghton, D., Lloyd-MacGilp, S., Pi, X., Roochun, Y., et al. (2011). The GUDMAP database--an online resource for genitourinary research. *Development* **138**, 2845-2853.
- Hartman, H. A., Lai, H. L. and Patterson, L. T. (2007). Cessation of renal morphogenesis in mice. *Developmental biology* **310**, 379-387.
- Hendry, C. E., Vanslambrouck, J. M., Ineson, J., Suhaimi, N., Takasato, M., Rae, F. and Little, M. H. (2013). Direct transcriptional reprogramming of adult cells to embryonic nephron progenitors. *Journal of the American Society of Nephrology : JASN* **24**, 1424-1434.
- Hochane, M., van den Berg, P. R., Fan, X., Berenger-Currias, N., Adegeest, E., Bialecka, M., Nieveen, M., Menschaart, M., Chuva de Sousa Lopes, S. M. and Semrau, S. (2019). Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development. *PLoS biology* **17**, e3000152.
- Howden, S. E., Vanslambrouck, J. M., Wilson, S. B., Tan, K. S. and Little, M. H. (2019). Reporter-based fate mapping in human kidney organoids confirms nephron lineage relationships and reveals synchronous nephron formation. *EMBO reports*.
- Kaminski, M. M., Tasic, J., Kresbach, C., Engel, H., Klockenbusch, J., Muller, A. L., Pichler, R., Grahammer, F., Kretz, O., Huber, T. B., et al. (2016). Direct reprogramming of fibroblasts into renal tubular epithelial cells by defined transcription factors. *Nature cell biology* **18**, 1269-1280.

- Karner, C. M., Das, A., Ma, Z., Self, M., Chen, C., Lum, L., Oliver, G. and Carroll, T. J. (2011). Canonical Wnt9b signaling balances progenitor cell expansion and differentiation during kidney development. *Development* **138**, 1247-1257.
- Kobayashi, A., Mugford, J. W., Krautzberger, A. M., Naiman, N., Liao, J. and McMahon, A. P. (2014). Identification of a multipotent self-renewing stromal progenitor population during mammalian kidney organogenesis. *Stem cell reports* **3**, 650-662.
- Kobayashi, A., Valerius, M. T., Mugford, J. W., Carroll, T. J., Self, M., Oliver, G. and McMahon, A. P. (2008). Six2 defines and regulates a multipotent self-renewing nephron progenitor population throughout mammalian kidney development. *Cell stem cell* **3**, 169-181.
- Lawlor, K. T., Zappia, L., Lefevre, J., Park, J. S., Hamilton, N. A., Oshlack, A., Little, M. H. and Combes, A. N. (2019). Nephron progenitor commitment is a stochastic process influenced by cell migration. *eLife*.
- Li, W., Hartwig, S. and Rosenblum, N. D. (2014). Developmental origins and functions of stromal cells in the normal and diseased mammalian kidney. *Developmental dynamics : an official publication of the American Association of Anatomists* **243**, 853-863.
- Li, Z., Araoka, T., Wu, J., Liao, H. K., Li, M., Lazo, M., Zhou, B., Sui, Y., Wu, M. Z., Tamura, I., et al. (2016). 3D Culture Supports Long-Term Expansion of Mouse and Human Nephrogenic Progenitors. *Cell stem cell* **19**, 516-529.
- Lindstrom, N. O., Carragher, N. O. and Hohenstein, P. (2015). The PI3K pathway balances self-renewal and differentiation of nephron progenitor cells through beta-catenin signaling. *Stem cell reports* **4**, 551-560.
- Lindstrom, N. O., De Sena Brandine, G., Tran, T., Ransick, A., Suh, G., Guo, J., Kim, A. D., Parvez, R. K., Ruffins, S. W., Rutledge, E. A., et al. (2018a). Progressive Recruitment of Mesenchymal Progenitors Reveals a Time-Dependent Process of Cell Fate Acquisition in Mouse and Human Nephrogenesis. *Developmental cell* **45**, 651-660 e654.
- Lindstrom, N. O., Guo, J., Kim, A. D., Tran, T., Guo, Q., De Sena Brandine, G., Ransick, A., Parvez, R. K., Thornton, M. E., Basking, L., et al. (2018b). Conserved and Divergent Features of Mesenchymal Progenitor Cell Types within the Cortical Nephrogenic Niche of the Human and Mouse Kidney. *Journal of the American Society of Nephrology : JASN* **29**, 806-824.
- Lindstrom, N. O., McMahon, J. A., Guo, J., Tran, T., Guo, Q., Rutledge, E., Parvez, R. K., Saribekyan, G., Schuler, R. E., Liao, C., et al. (2018c). Conserved and Divergent Features of Human and Mouse Kidney Organogenesis. *Journal of the American Society of Nephrology : JASN* **29**, 785-805.
- Lindstrom, N. O., Tran, T., Guo, J., Rutledge, E., Parvez, R. K., Thornton, M. E., Grubbs, B., McMahon, J. A. and McMahon, A. P. (2018d). Conserved and Divergent Molecular and Anatomic Features of Human and Mouse Nephron Patterning. *Journal of the American Society of Nephrology : JASN* **29**, 825-840.
- Little, M. H., Brennan, J., Georgas, K., Davies, J. A., Davidson, D. R., Baldock, R. A., Beverdam, A., Bertram, J. F., Capel, B., Chiu, H. S., et al. (2007). A high-resolution anatomical ontology of the developing murine genitourinary tract. *Gene expression patterns : GEP* **7**, 680-699.
- Little, M. H., Combes, A. N. and Takasato, M. (2016). Understanding kidney morphogenesis to guide renal tissue regeneration. *Nature reviews. Nephrology* **12**, 624-635.
- Lun, A. T., McCarthy, D. J. and Marioni, J. C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* **5**, 2122.
- Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214.
- Madisen, L., Zwingman, T. A., Sunkin, S. M., Oh, S. W., Zariwala, H. A., Gu, H., Ng, L. L., Palmiter, R. D., Hawrylycz, M. J., Jones, A. R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature neuroscience* **13**, 133-140.
- Magella, B., Adam, M., Potter, A. S., Venkatasubramanian, M., Chetal, K., Hay, S. B., Salomonis, N. and Potter, S. S. (2017). Cross-platform single cell analysis of kidney development shows stromal cells express Gdnf. *Developmental biology*.

- McCarthy, D. J. and Smyth, G. K.** (2009). Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* **25**, 765-771.
- McNeill, H. and Reginensi, A.** (2017). Lats1/2 Regulate Yap/Taz to Control Nephron Progenitor Epithelialization and Inhibit Myofibroblast Formation. *Journal of the American Society of Nephrology : JASN* **28**, 852-861.
- Menon, R., Otto, E. A., Kokoruda, A., Zhou, J., Zhang, Z., Yoon, E., Chen, Y. C., Troyanskaya, O., Spence, J. R., Kretzler, M., et al.** (2018). Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney. *Development* **145**.
- Merlet-Benichou, C., Gilbert, T., Vilar, J., Moreau, E., Freund, N. and Lelievre-Pegorier, M.** (1999). Nephron number: variability is the rule. Causes and consequences. *Laboratory investigation; a journal of technical methods and pathology* **79**, 515-527.
- Mugford, J. W., Sipila, P., McMahon, J. A. and McMahon, A. P.** (2008). 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. *Developmental biology* **324**, 88-98.
- Mugford, J. W., Yu, J., Kobayashi, A. and McMahon, A. P.** (2009). High-resolution gene expression analysis of the developing mouse kidney defines novel cellular compartments within the nephron progenitor population. *Developmental biology* **333**, 312-323.
- Naiman, N., Fujioka, K., Fujino, M., Valerius, M. T., Potter, S. S., McMahon, A. P. and Kobayashi, A.** (2017). Repression of Interstitial Identity in Nephron Progenitor Cells by Pax2 Establishes the Nephron-Interstitium Boundary during Kidney Development. *Developmental cell* **41**, 349-365 e343.
- Nishita, M., Qiao, S., Miyamoto, M., Okinaka, Y., Yamada, M., Hashimoto, R., Iijima, K., Otani, H., Hartmann, C., Nishinakamura, R., et al.** (2014). Role of Wnt5a-Ror2 signaling in morphogenesis of the metanephric mesenchyme during ureteric budding. *Molecular and cellular biology* **34**, 3096-3105.
- O'Brien, L. L., Combes, A. N., Short, K. M., Lindstrom, N. O., Whitney, P. H., Cullen-McEwen, L. A., Ju, A., Abdelhalim, A., Michos, O., Bertram, J. F., et al.** (2018). Wnt11 directs nephron progenitor polarity and motile behavior ultimately determining nephron endowment. *eLife* **7**.
- Oxburgh, L., Chu, G. C., Michael, S. K. and Robertson, E. J.** (2004). TGFbeta superfamily signals are required for morphogenesis of the kidney mesenchyme progenitor population. *Development* **131**, 4593-4605.
- Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A. and Trapnell, C.** (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nature methods* **14**, 979-982.
- Ramilowski, J. A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V. P., Itoh, M., Kawaji, H., Carninci, P., Rost, B., et al.** (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. *Nature communications* **6**, 7866.
- Reginensi, A., Hoshi, M., Boualia, S. K., Bouchard, M., Jain, S. and McNeill, H.** (2015). Yap and Taz are required for Ret-dependent urinary tract morphogenesis. *Development* **142**, 2696-2703.
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K.** (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**, e47.
- Robinson, M. D., McCarthy, D. J. and Smyth, G. K.** (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.
- Rumballe, B. A., Georgas, K. M., Combes, A. N., Ju, A. L., Gilbert, T. and Little, M. H.** (2011). Nephron formation adopts a novel spatial topology at cessation of nephrogenesis. *Developmental biology* **360**, 110-122.
- Rutledge, E. A., Benazet, J. D. and McMahon, A. P.** (2017). Cellular heterogeneity in the ureteric progenitor niche and distinct profiles of branching morphogenesis in organ development. *Development* **144**, 3177-3188.
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. and Regev, A.** (2015). Spatial reconstruction of single-cell gene expression data. *Nature biotechnology* **33**, 495-502.

- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676-682.
- Schmidt-Ott, K. M., Yang, J., Chen, X., Wang, H., Paragas, N., Mori, K., Li, J. Y., Lu, B., Costantini, F., Schiffer, M., et al. (2005). Novel regulators of kidney development from the tips of the ureteric bud. *Journal of the American Society of Nephrology : JASN* **16**, 1993-2002.
- Schwab, K., Patterson, L. T., Aronow, B. J., Luckas, R., Liang, H. C. and Potter, S. S. (2003). A catalogue of gene expression in the developing kidney. *Kidney Int* **64**, 1588-1604.
- Short, K. M., Combes, A. N., Lefevre, J., Ju, A. L., Georgas, K. M., Lamberton, T., Cairncross, O., Rumballe, B. A., McMahon, A. P., Hamilton, N. A., et al. (2014). Global quantification of tissue dynamics in the developing mouse kidney. *Developmental cell* **29**, 188-202.
- Soneson, C. and Robinson, M. D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. *Nature methods* **15**, 255-261.
- Srinivas, S., Goldberg, M. R., Watanabe, T., D'Agati, V., al-Awqati, Q. and Costantini, F. (1999). Expression of green fluorescent protein in the ureteric bud of transgenic mice: a new tool for the analysis of ureteric bud morphogenesis. *Developmental genetics* **24**, 241-251.
- Stahl, P. L., Salmen, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J. O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82.
- Stuart, R. O., Bush, K. T. and Nigam, S. K. (2003). Changes in gene expression patterns in the ureteric bud and metanephric mesenchyme in models of kidney development. *Kidney Int* **64**, 1997-2008.
- Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research* **45**, D362-D368.
- Taguchi, A., Kaku, Y., Ohmori, T., Sharmin, S., Ogawa, M., Sasaki, H. and Nishinakamura, R. (2014). Redefining the in vivo origin of metanephric nephron progenitors enables generation of complex kidney structures from pluripotent stem cells. *Cell stem cell* **14**, 53-67.
- Takasato, M., Er, P. X., Chiu, H. S., Maier, B., Baillie, G. J., Ferguson, C., Parton, R. G., Wolvetang, E. J., Roost, M. S., Chuva de Sousa Lopes, S. M., et al. (2015). Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis. *Nature* **526**, 564-568.
- Tanigawa, S., Taguchi, A., Sharma, N., Perantoni, A. O. and Nishinakamura, R. (2016). Selective In Vitro Propagation of Nephron Progenitors Derived from Embryos and Pluripotent Stem Cells. *Cell reports*.
- Thiagarajan, R. D., Georgas, K. M., Rumballe, B. A., Lesieur, E., Chiu, H. S., Taylor, D., Tang, D. T., Grimmond, S. M. and Little, M. H. (2011). Identification of anchor genes during kidney development defines ontological relationships, molecular subcompartments and regulatory pathways. *PloS one* **6**, e17286.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S. and Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology* **32**, 381-386.
- Walker, K. A., Sims-Lucas, S. and Bates, C. M. (2016). Fibroblast growth factor receptor signaling in kidney and lower urinary tract development. *Pediatric nephrology* **31**, 885-895.
- Wang, P., Chen, Y., Yong, J., Cui, Y., Wang, R., Wen, L., Qiao, J. and Tang, F. (2018). Dissecting the Global Dynamic Molecular Profiles of Human Fetal Kidney Development by Single-Cell RNA Sequencing. *Cell reports* **24**, 3554-3567 e3553.
- Yang, B., Treweek, J. B., Kulkarni, R. P., Deverman, B. E., Chen, C. K., Lubeck, E., Shah, S., Cai, L. and Gradinaru, V. (2014). Single-cell phenotyping within transparent intact tissue through whole-body clearing. *Cell* **158**, 945-958.

- Young, M. D., Mitchell, T. J., Vieira Braga, F. A., Tran, M. G. B., Stewart, B. J., Ferdinand, J. R., Collord, G., Botting, R. A., Popescu, D. M., Loudon, K. W., et al. (2018). Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594-599.
- Yuri, S., Nishikawa, M., Yanagawa, N., Jo, O. D. and Yanagawa, N. (2017). In Vitro Propagation and Branching Morphogenesis from Single Ureteric Bud Cells. *Stem cell reports* **8**, 401-416.

Table 1: Top DE and cluster-specific genes from whole kidney (K) clusters

Representative top DE cluster markers	Globally specific or enriched markers*
K0 Nephron progenitor: <i>Cited1, Crym, Meox1, Six2, Traf1, Uncx, Eya1, Spock2</i>	<i>Cited1, Traf1, Elavl4, Ocm, Pnmt, Meox2, Tcap, E030013119Rik, Ripply2, Fgf20, Siglecg, Phf19, Btbd11</i>
K1 Stroma – Collecting duct associated: <i>Penk, Nts, Acta2, Cldn11, Tagln1, Alx1</i>	<i>Cck, Cpxm2, Ndp, Lrfr5, Mansc4, Myh11, Akr1b8, Slc17a8, Mr1</i>
K2 Stroma – Cortical stroma: <i>Ren1, Fibin, Mgp, Hic1, Igfbp5, Rgs5, Fhl2, Ntn1, Lhfp, Foxd1, Gdnf</i>	<i>Lypd1, Far2, Actr3b, Tmem18, Scn9a</i>
K3 Distal nephron: <i>S100g, Tmem52b, Ly6a, Sostdc1, Slc12a1, Wfdc2, Mal, Aqp2</i>	<i>Clcnkb, Kcnj1, Bsnd, Tmem72, Atp6v1b1, Slc12a1, Cldn16, Nrcam</i>
K4 Stroma – Medullary stroma: <i>Igfbp3, Fbln5, Acta2, Mgp, Dcn, Ace2, Cfh, Col3a1, Vegfd, Colla1, Ndufa4l2, Rgs5</i>	<i>Eepd1, Megf11, Colec11, Cpne7, Scn7a</i>
K5 Early proximal tubule: <i>Sult1d1, Spink1, Aldob, Hdc, Pdzk1, Slc34a1, Fut9, Fxyd2, Osr2, Slc39a5, Keg1, Cpn1, Ttc36, Ly6a</i>	EPT & PT specific: <i>Aldob, Spp2, Slc34a1, Fbp1, Gsta2, Sult1d1, Lrp2, Pdzk1, Aadat, Slc22a6, Mep1a, Acsm2, Slc27a2, Gm10639, Defb19, Cyp2j5, Slc5a8, Pck1</i>
K6 S-shaped body: <i>Lhx1, Pcp4, Cldn5, Sfrp2, Osr2, Clec18a, Clu, Uncx, Npy, Ccnd1, Pax8, Wnt4, Mafk, Sox11, Jag1</i>	<i>Nppc, Plpp4, Tcf23, Sh3bgr</i>
K7 Renal vesicle: <i>Fam132a, Wnt4, Tmem100, Bmper, Pax2, Eya1, Fam107a, Wt1, Frzb, Gxylt2, Kazald1, Mycn, Snap91</i>	All genes also expressed in NP, PTA, or SSB
K8 Proximal tubule: <i>Aldob, Ttc36, Spp2, Kap, Fxyd2, Slc34a1, Fbp1, Gsta2, Sult1d1, Spink1, Ass1, Lrp2, Gatm, Pdzk1</i>	Specific to PT not in EPT: <i>Kap, Serpina6, Pah, Acmsd, Gsta1, Gm853, G6pc, Rdh16, Kap, Serpina6, Pah, Acmsd, Gsta1, Gm853, G6pc, Rdh16, Acox2, Slc10a2</i>
K9 Ureteric epithelium: <i>Calb1, Upk3a, Rprm, Aqp2, Trp63, Crlf1, Krt18, Lcn2, Gata3, Wfdc2, Krt19, Krt8, Ret, Mal, Mia</i>	<i>Sprrla, Psca, Upk2, Gm14133, Wnt9b, Cd79a, Aldh3b2, Ret, Anxa9, Gulo, Grik2</i>
K10 Immune system cells: <i>Ly2, Clqc, Clqb, Pf4, S100a8, Fcrlg, Tyrobp, Apoe</i>	<i>Ly2, Clqc, Clqb, Fcrlg, Ctss, Cd52, Ms4a6c, Ccl4, Fcgr3, Ccl3, Corol1a, Aif1, Ly86, Ms4a6b, Cx3cr1</i>
K11 Endothelial: <i>Plvap, Cdh5, Pecam1, Cldn5, Esam, Cd34, Flt1, Kdr, Tie1, Ecsr</i>	<i>Fam167b, Aplnr, Gpihbp1, Sox17, Myct1, Tie1, Adgrl4, Mmrn2, Sox18, Clec1a, Ptprrb, Pecam1</i>
K12 Committing NP: <i>Cited1, Six2, Pclaf, Eya1, Uncx, Spock2, Crym, Wnt4</i>	All genes also expressed in NP or RV populations
K13 Stroma – Ureter + : <i>Dlk1, Dcn, Igf1, Meg3, Colla1, Postn, Col3a1, Lum, Tbx18,</i>	<i>Clec3b, Dpt, Tnxb, Col5a3, Col6a6, Fndc1, S100b, Slc7a10, Fgf7, Wfikkn2, Mst1r, Gfpt2, Wnt9a, Rab33a</i>
K14 Nephron progenitor-stromal: <i>Cited1, Acta2, Six2, Penk, Col3a1, Cfh, Col14a1, Crym, Dcn, Tpm2, Gucy1a3</i>	All genes also expressed in NP or Stromal populations

K15 Podocyte: <i>R3hdml, Mafb, Nphs2, Magi2, Podxl, Cdkn1c, Cldn5, Nphs1, Rasl11a, Dpp4, Synpo, Mapt, Ptpro, Wt1</i>	<i>Nphs2, Rhpn1, 4921504A21Rik, Vcpkmt, Ryr1, Entpd7, Med24, Rab3b, Dhx34, Fbxo3</i>
---	--

* Some ‘specific’ were markers also expressed in closely related populations. For example NP markers are also expressed to some extent in NP-Str, PTA, and RV. Stromal markers are expressed in Np-Str population, EPT genes also expressed in SSB, PT, and elsewhere.

Figures

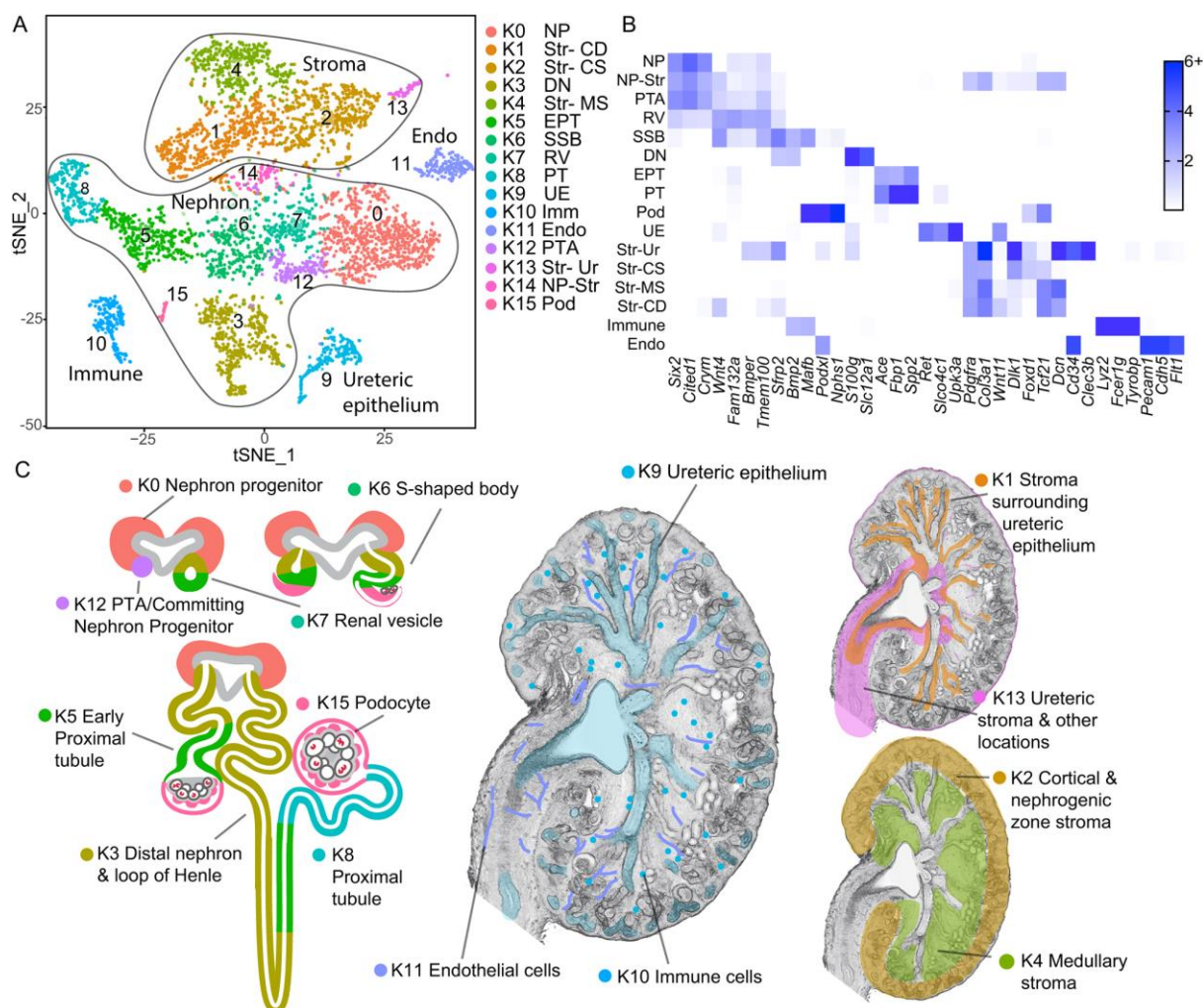


Figure 1. Markers and population map for E18.5 mouse kidney. **A.** tSNE plot revealing 16 cell clusters within the whole developing kidney (K) identified from largest to smallest population as representing nephron progenitor (K0 NP), stroma surrounding collecting duct/ureteric epithelium (K1 Str- CD), cortical stroma (K2 Str- CS), distal nephron (K3 DN), medullary stroma (K4 Str- MS), early proximal tubule (K5 EPT), S-shaped body (K6 SSB), renal vesicle (K7 RV), proximal tubule (K8 PT), ureteric epithelium (K9 UE), Immune cells (K10 Imm), endothelial cells (K11 Endo), committing nephron progenitors/pretubular aggregate (K12 PTA), a stromal cluster with a mixed expression domain including ureteric stroma (K13 Str- Ur), a nephron progenitor – stroma cluster (K14 NP-Str) and podocytes (K15 Pod). **B.** Key cell type markers within whole kidney clusters. Scale indicates log fold change differential expression of cells within cluster relative to all other cells. **C.** Diagram relating single cell clusters to tissue structure or anatomical location. Populations coloured according to key in A aside from K6 and K7, which are coloured to reflect known patterning.

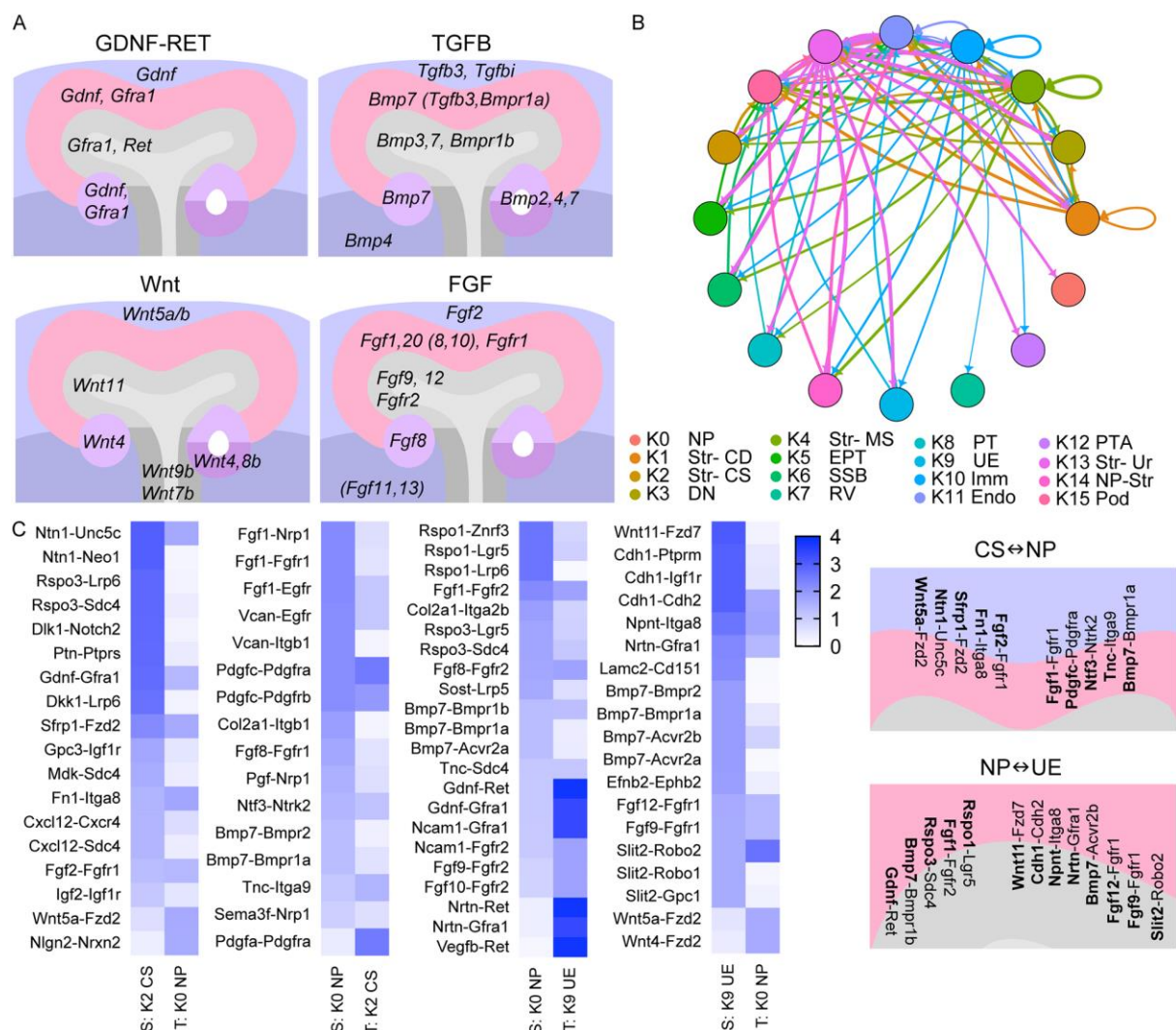


Figure 2. Global analysis of receptor-ligand interactions. **A.** Illustration of expression domains for known ligands and or receptors involved in GDNF-RET, TGFB, Wnt, and FGF signalling pathways from the differential expression analysis. LogFC ≥ 1 for most genes shown aside from those in brackets, which have lower values (available in Table S1). **B.** Plot illustrating potential interactions between all cell types in the whole kidney data. Arrows originate from ligand producing cluster and end in putative target cluster. Line colour indicates cluster of origin, thickness indicates the number of interactions. Note >30 interactions between K0 NP and K9 UE clusters are detailed in Table S2, but this number was not sufficient to produce a line in this chart. Abbreviations for cluster key as per Figure 1. **C.** Specific ligand-receptor interactions predicted by the global analysis between cortical stroma (K2 CS), nephron progenitor (K0 NP), and ureteric epithelium (K9 UE) clusters. Heatmaps indicate LogFC differential expression values within the cluster, listing ligand-receptor pairs on the Y-axis, and the ligand source (S) cluster and receptor target (T) cluster on the X-axis. Cartoons to the right illustrate some of these interactions in context.

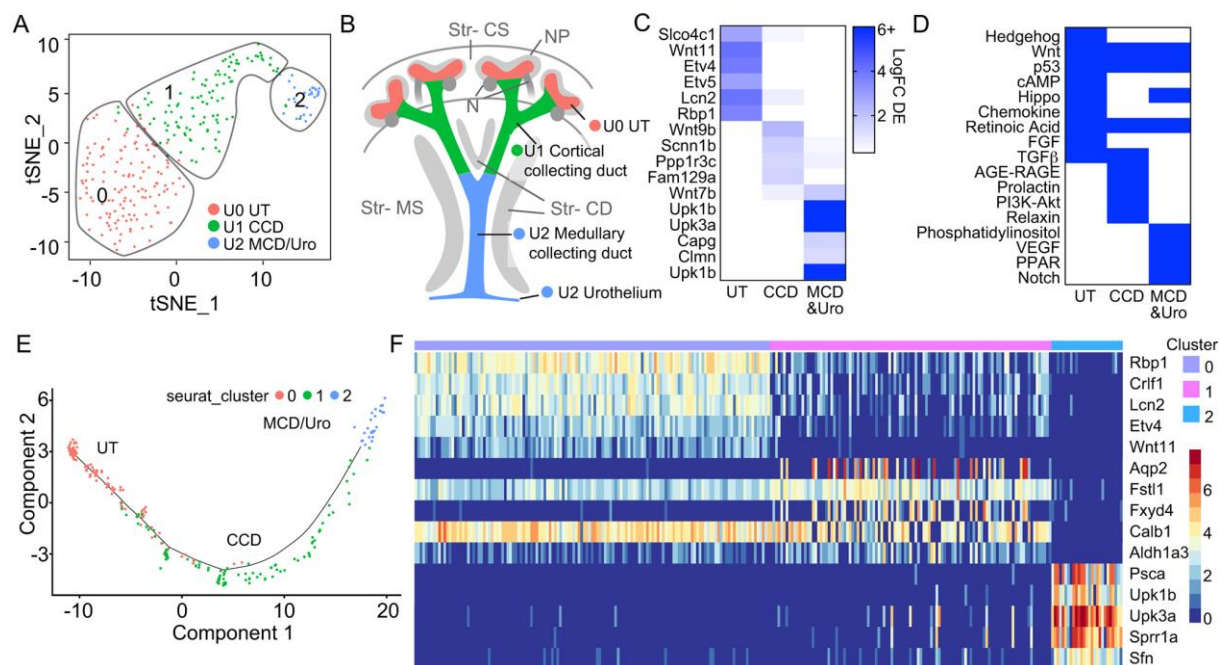


Figure 3. Ureteric epithelium subclustering identifies known subpopulations and established developmental trajectories. **A.** Reclustering of ureteric epithelium cells (K9 UE) identifies three ureteric (U) subclusters representing ureteric tip (U0 UT), cortical collecting duct (U1 CCD) and medullary collecting duct / urothelium (U2 MCD/Uro). **B.** Diagram of the relative location of these three ureteric epithelial cell types with respect to surrounding stromal populations. Str- CS, cortical stroma; NP, nephron progenitor; MS, medullary stroma; Str- CD, collecting duct associated stroma. **C.** Expression of key marker genes in ureteric epithelium subclusters. **D.** Identification of differential signalling pathway activity across these three UE populations. **E.** Pseudotime trajectory of the three UE subclusters reflects a developmental origin of all clusters from the ureteric tip, with cells progressing through the CCD with the final cell type state being MCD/Uro. X and Y axes represent independent component space, where the cells have been iteratively shifted onto the vertices of a spanning tree used to determine the trajectory. **F.** Heat map of marker genes for subpopulations within the ureteric epithelium. Clusters represent UT (mauve, cluster U0), CCD (pink, cluster U1) and MCD (blue, cluster U2).

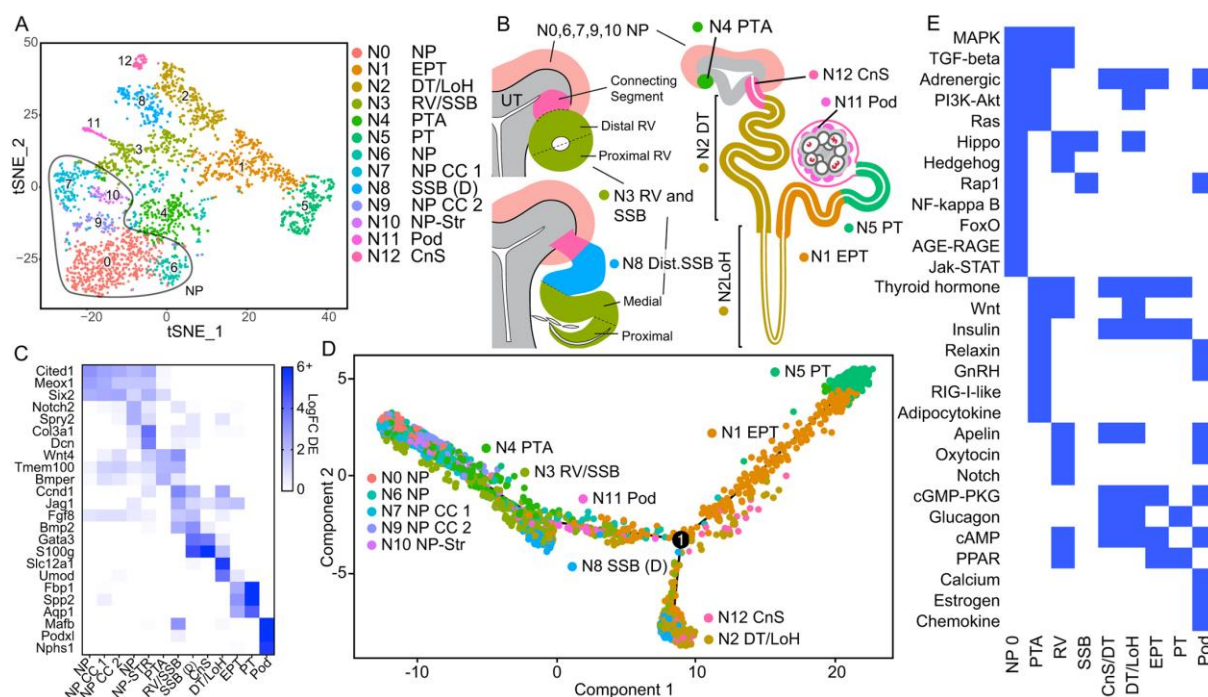


Figure 4. Nephron lineage reclustering. **A.** tSNE plot of 13 nephron lineage clusters from the developing mouse kidney. This includes 8 clusters representing distinct stages or segments of developing nephron (pretubular aggregate PTA, renal vesicle RV, s-shaped body SSB, early proximal tubule EPT, podocyte Pod, proximal tubule PT, distal tubule/loop of Henle DT/LOH, connecting segment CnS) including 5 clusters with nephron progenitor (NP) identity. **B.** Diagram of nephron maturation. Note the connecting segment that links the nephron to the ureteric tip arises at late RV stage, by which time the distal and proximal RV already displays distinct gene expression (Georgas et al., 2009). By SSB, a medial domain of gene expression can be identified. **C.** Heatmap illustrating key differentially expressed markers across the nephron lineage clusters. Log fold change (LogFC) differential expression (DE) shown. **D.** Pseudotime analysis including all nephron lineage cells illustrates an anticipated transition from NP through PTA, RV/SSB, SSB distal (SSB (D)). A branchpoint is observed between distal and proximal arms of nephron development. Of note, podocyte clusters are closer to RV/SSB than either proximal or distal tubule. X and Y axes represent independent component space. **E.** Select signalling pathway activity across major nephron clusters. More shown in Fig. S3.

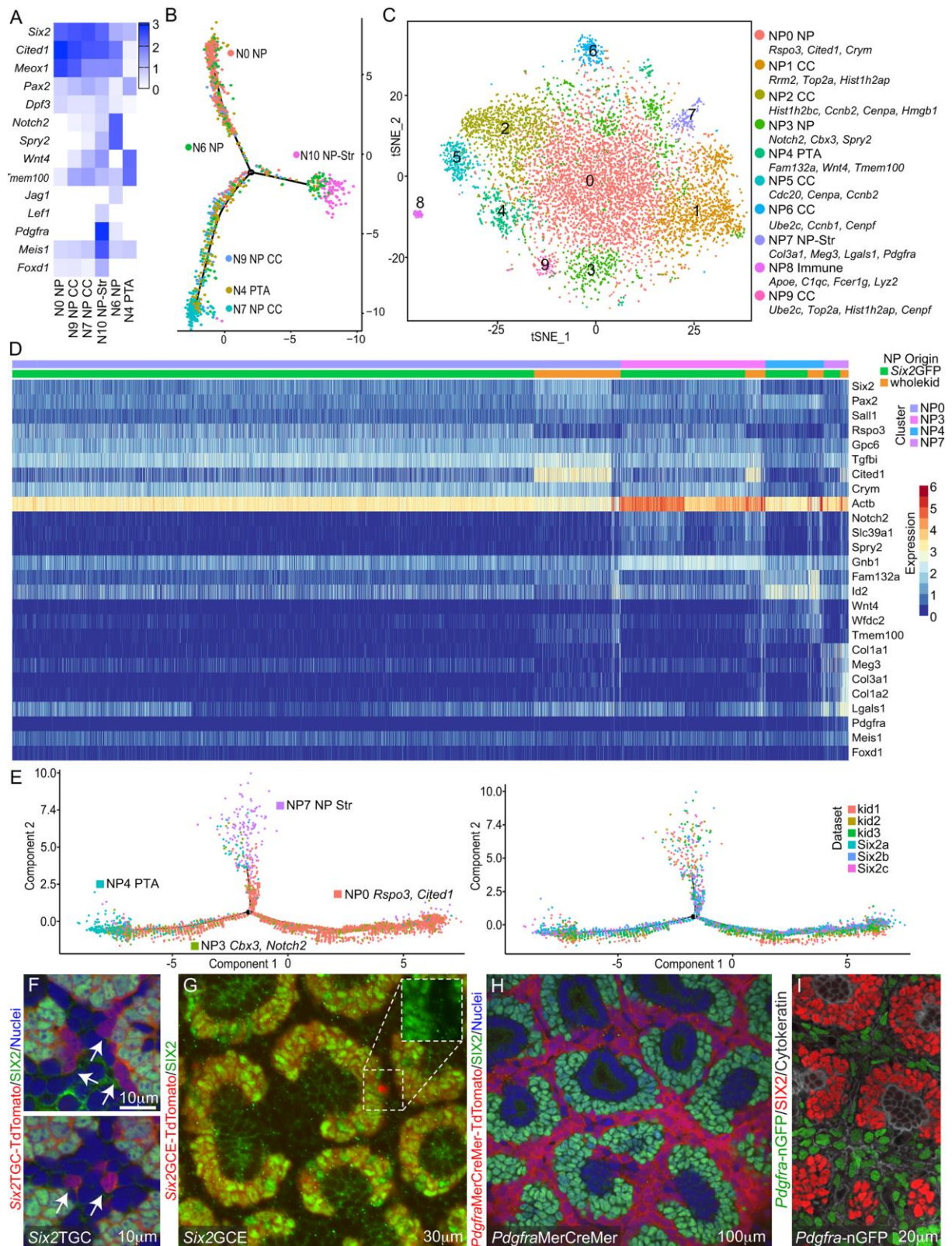


Figure 5. Nephron progenitor subpopulations. **A.** Expression of key nephron progenitor (NP) subpopulation markers across NP clusters in the nephron lineage analysis including two cell cycle (CC)-associated, one stromal (Str)-like, and a pretubular aggregate (PTA) cluster. Scale represents LogFC differential expression within the nephron lineage. **B.** Monocle 2 analysis of early nephron lineage clusters identifies a trajectory for NP-Str cluster cells distinct from the expected NP-PTA trajectory taken by most cells. X and Y axes represent

independent component space **C.** tSNE plot of integrated NP data. Clusters are referred to as NP clusters 0-9 (NP0-NP9). The integrated dataset is composed of non cycling clusters from the e18.5 nephron lineage dataset (N0, N4, N6, N10) and >7800 sorted *Six2*GFP+ cells. A provisional identification and top marker genes are listed next to each cluster ID. Clusters NP1, NP2, NP5, NP6, & NP9 are dominated by cell cycle genes. Cluster NP8 is defined by immune cell markers- an unintended inclusion from the FACS isolation; NP0 features 'uncommitted' progenitor genes such as *Cited1*; NP3 markers include *Cbx3*, *Notch2*, *Spry2*; NP4 features commitment markers such as *Wnt4*; NP7 features nephron progenitor and stromal markers. **D.** Heatmap showing gene expression of NP and stromal markers in four clusters identified from an integrated analysis of sorted *Six2*GFP cells and the non-cycling NP clusters from the nephron lineage dataset. Scale represents log expression. **E.** Trajectory analysis using monocle on cells from clusters NP0, NP3, NP4 and NP7 from the integrated NP analysis. Cells are coloured by Seurat cluster (left) and biological replicate (right). Axes as per B **F-H.** Representative images from lineage tracing and reporter experiments. **F.** Lineage tracing with the *Six2*TGC cre from the start of kidney development identifies *Six2*-derived cells (red), negative for SIX2 protein (green), in the stroma between nephron progenitor niches, and deeper in the kidney (not shown). **G.** Lineage tracing with an inducible *Six2*GCE Cre line did not result in *Six2*-derived stromal labelling aside from infrequent cells with abnormal morphology. Dashed box shows SIX2 channel only, illustrating absence of SIX2 signal. **H.** Lineage tracing with an inducible *Pdgfra*MerCreMer line at E13.5 did not result in co-labelling with SIX2 at E18.5. **I.** Mutually exclusive expression of stromal (*Pdgfra*-nGFP) and nephron progenitor (SIX2) markers in the nephrogenic zone. Ureteric tip marked by Cytokeratin.

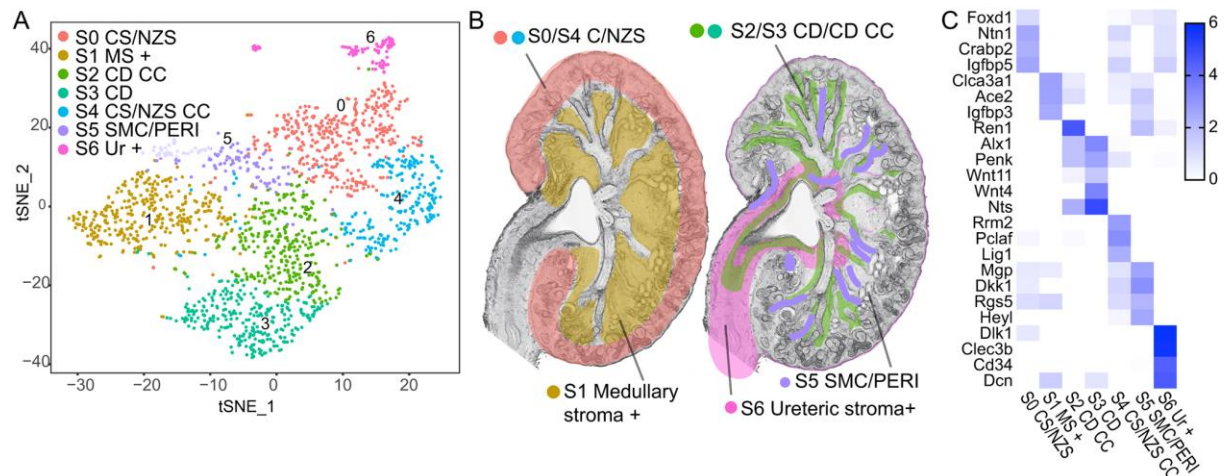


Figure 6. Analysis of stromal clusters within developing mouse kidney. A. Reclustering of all cells from the stromal lineage resulted in seven clusters: S0 Cortical/nephrogenic zone stroma (CS/NZS), S1 Medullary stroma + (MS+), S2 collecting duct-associated stroma cell cycle (CD CC), S3 collecting duct-associated stroma (CD), S4 CS/NZS cell cycle (CS/NZS CC), S5 smooth muscle cell/pericyte-like (SMC/PERI), S6 ureteric stroma + (US). **B.** Analysis of expression patterns for cluster markers (Fig. S3) defined regions of common expression for five of the seven clusters. The remaining two likely represent proliferating subpopulations. **C.** Representative genes differentially expressed between stromal clusters.

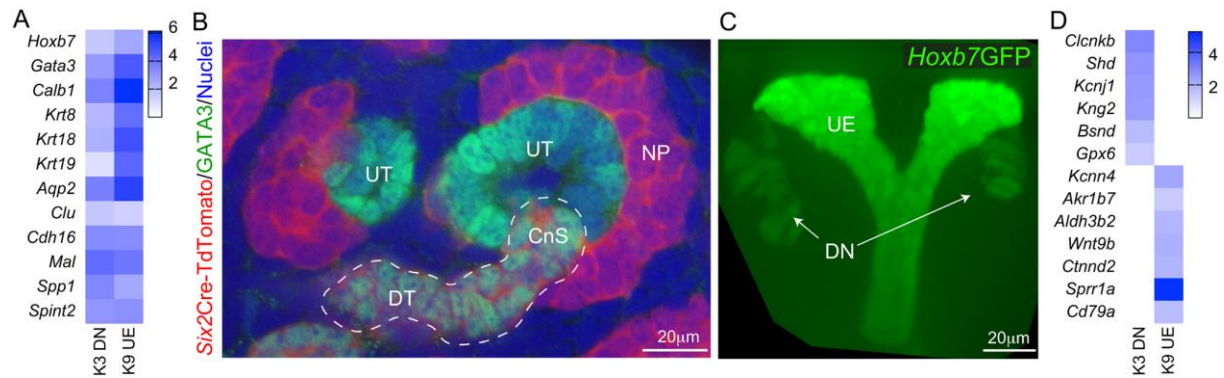


Figure 7. Congruence between markers of the ureteric epithelium and distal nephron. A. Expression of common ureteric epithelium (UE) markers in the distal nephron (DN) and vice versa in the whole kidney clusters. Scale represents LogFC DE. **B.** TdTomato expression (red) activated by *Six2* cre affirms the nephron lineage of the connecting segment (CnS) and distal tubule (DT). GATA3 protein is detected in the ureteric tips (UT) and the distal nephron (CnS/DT, outline). **C.** Expression of GFP driven by the *Hoxb7* promoter is detected in the ureteric epithelium (UE) and distal nephron (DN). **D.** Examples of markers that can be used to distinguish between distal nephron and ureteric epithelium. Full lists detailed in Table S7. Scale represents LogFC DE within whole kidney clusters.

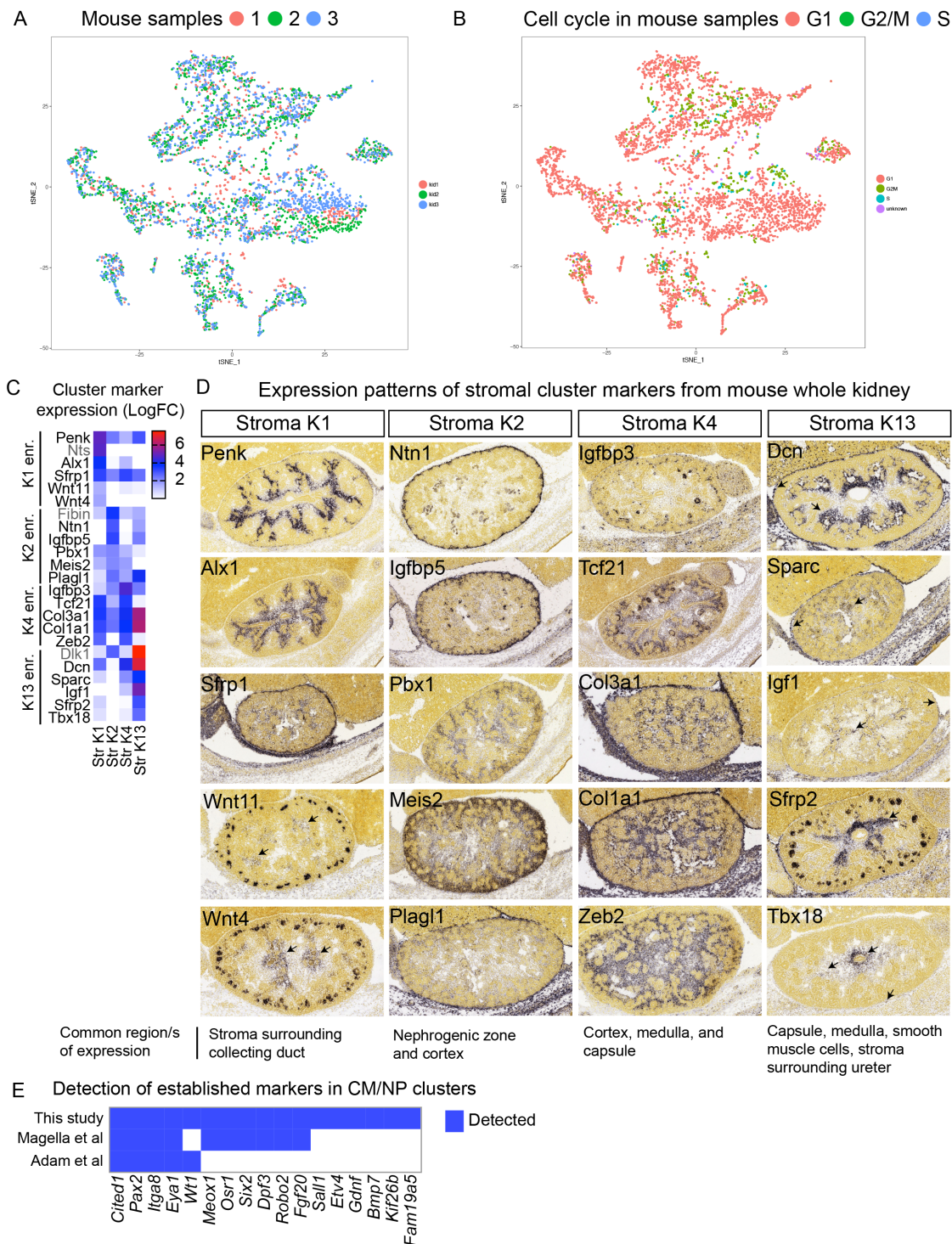


Figure S1. Visualisation of data by sample and cell cycle and supporting information. **A.** tSNE plot of cells identified by sample of origin shows an even distribution of cell types present within all samples. **B.** tSNE plot of all cells from mouse developing kidney identified by stage of cell cycle (G1, G2/M, S). **C.** Heatmap of stromal cluster markers from whole kidney. Markers in black indicate genes shown in D. **D.** *In situ* hybridisation results from the Allen Developing Mouse Brain Atlas (<http://developingmouse.brain-map.org>) for stromal marker genes used to aid in cluster mapping. Note that expression domains within each cluster do not completely overlap indicating further heterogeneity. **E.** Detection of established cap mesenchyme (CM)/nephron progenitor (NP) markers in any corresponding clusters from this study, Adam et al (2017) and Magella et al (2017) shows a 40-70% increase in detection of relevant markers in this dataset. Detection in this dataset = LogFC>0.94; Magella = featured in 'cell-type specific gene lists' reported in SuppTable4 for any cap mesenchyme cluster (at any Pearson.rho value); Adam = featured in TableS6 'compartment specific gene lists' for cap mesenchyme. NP expression of *Etv4* (aka *Pea3*) first demonstrated in Lu et al., Nat. Genet. 2009 and Mugford et al., Dev. Biol. 2009.

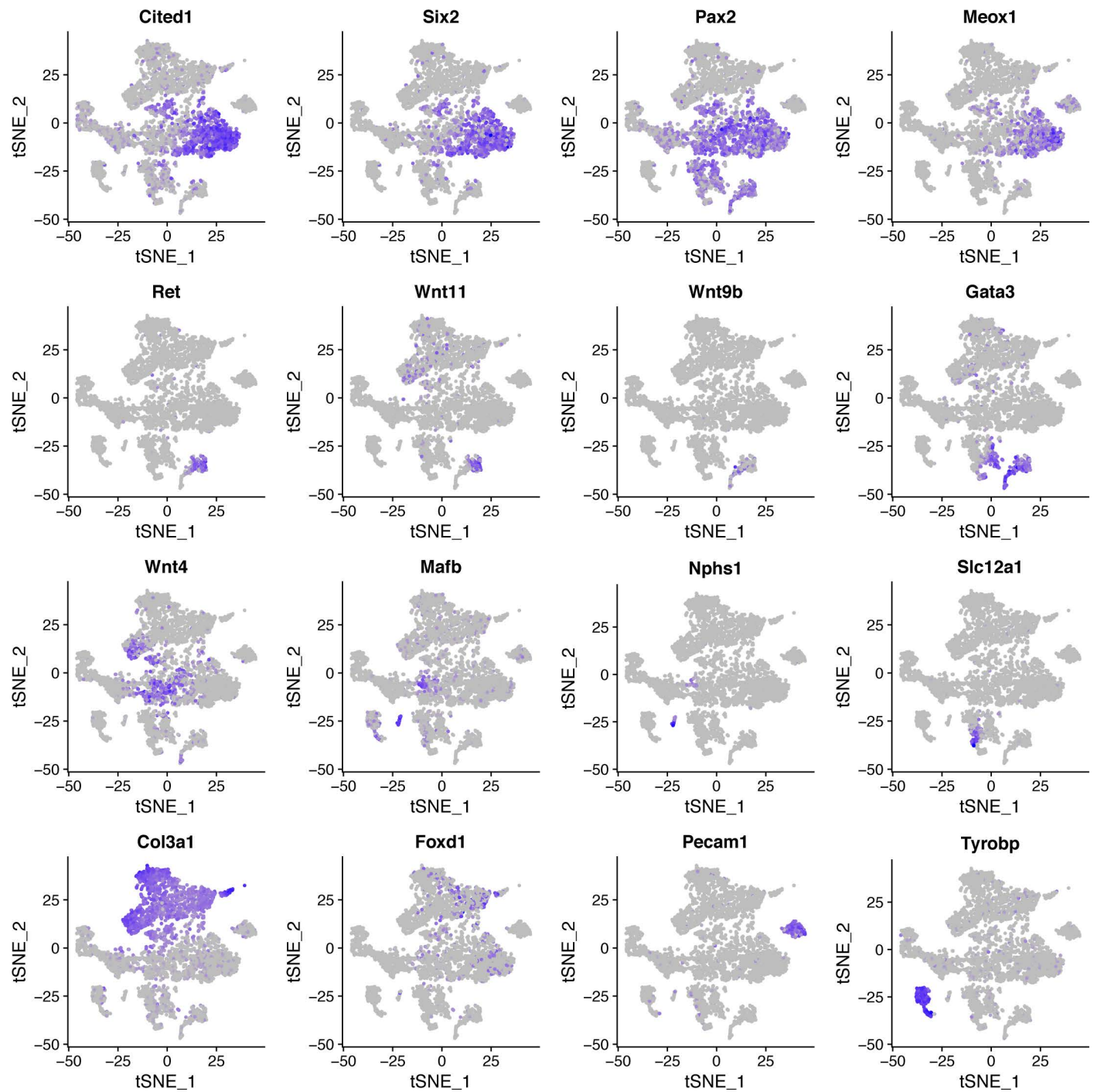


Figure S2. Marker expression across all cells in the developing mouse kidney data. tSNE plot of all cells from E18.5 developing mouse kidney showing the location of cells expressing key marker genes, including markers of nephron progenitor (*Cited1*, *Six2*, *Pax2*, *Meox1*), ureteric epithelium (*Ret*, *Wnt11*, *Wnt9b*, *Gata3*), early nephron (*Wnt4*), podocyte (*Mafk*, *Nphs1*), distal tubule (*Slc12a1*), stroma (*Col3a1*, *Foxd1*), endothelium (*Pecam1*) and immune cells (*Tyrobp*).

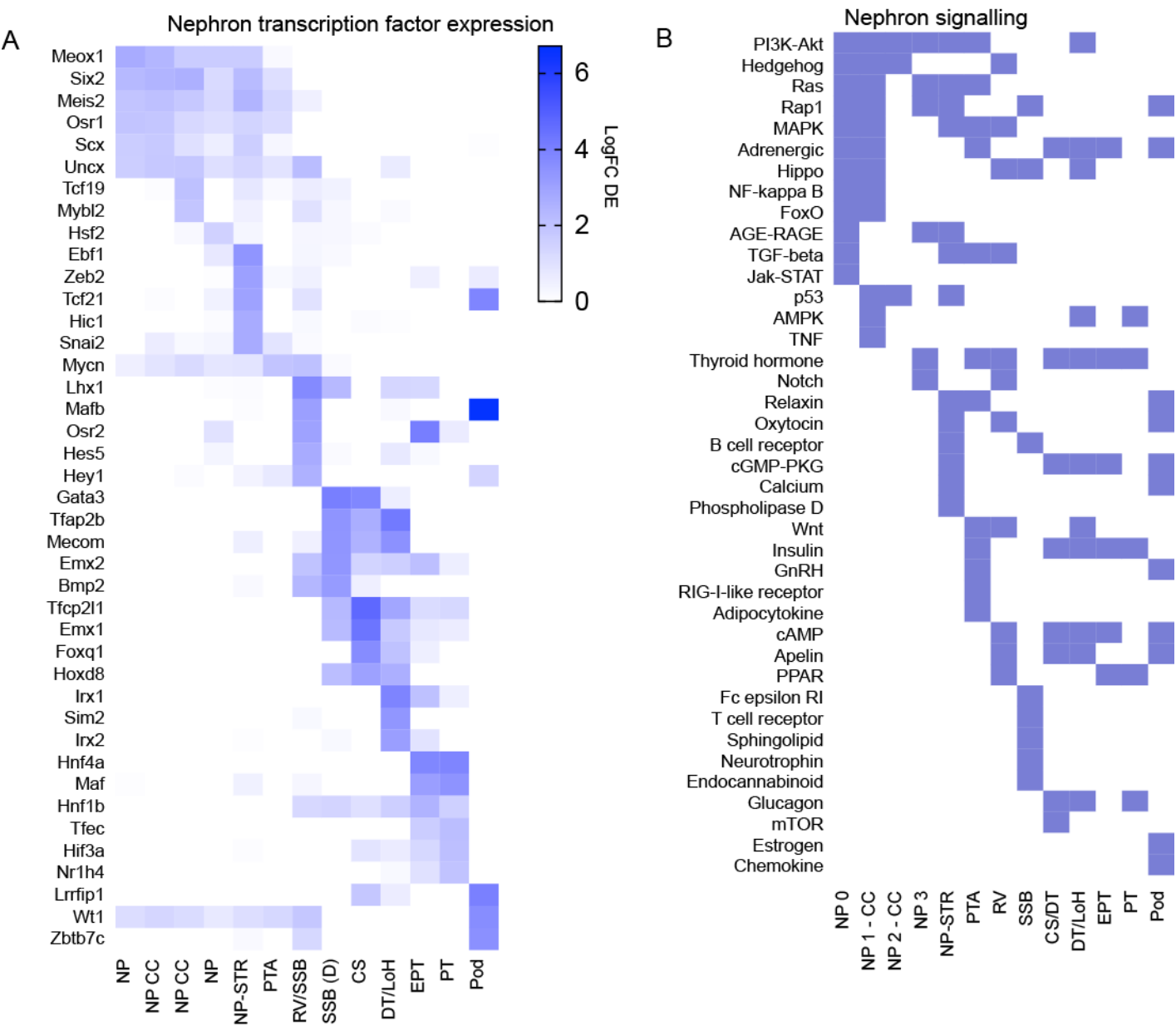
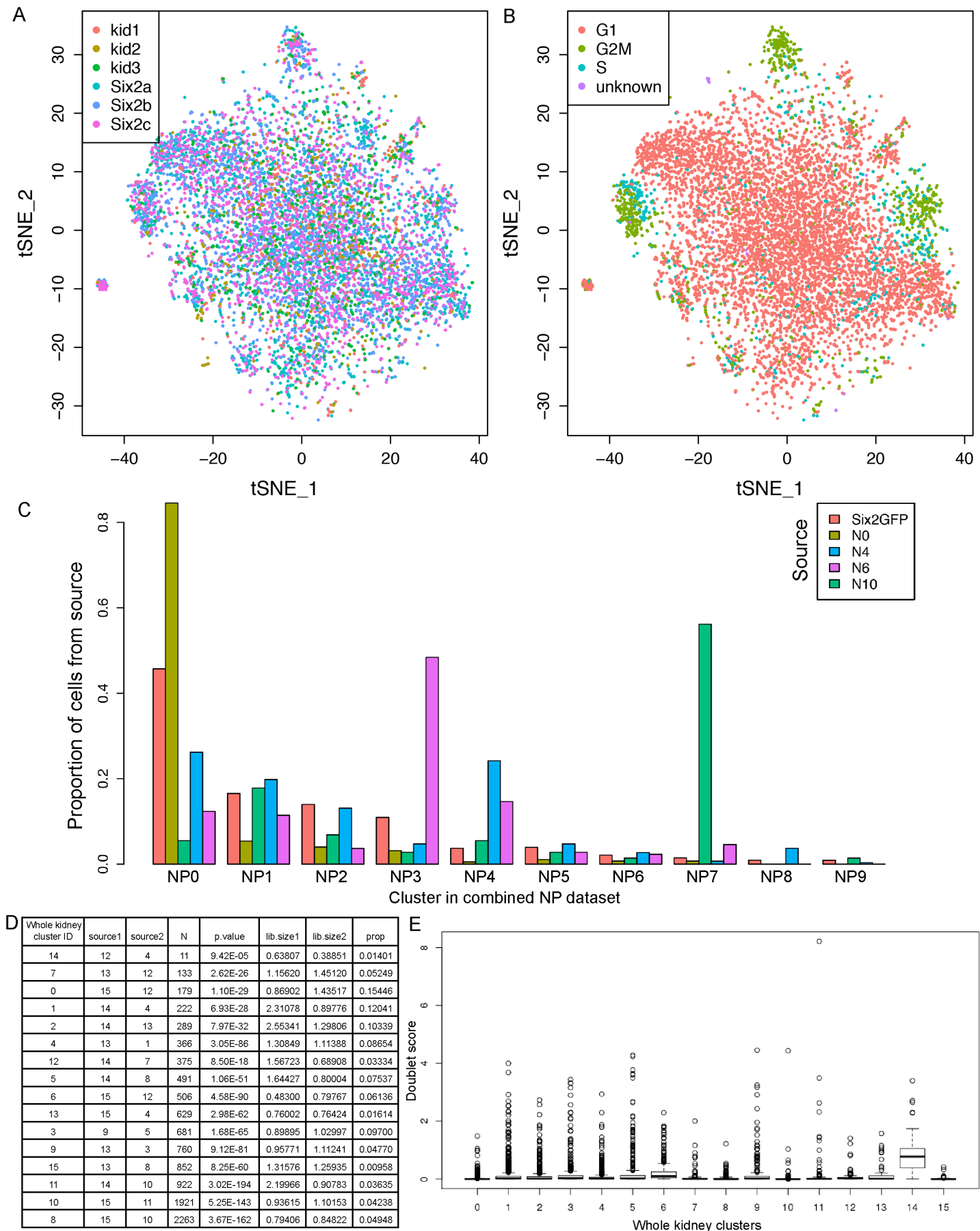


Figure S3. Transcription factors and KEGG analysis of signalling pathways within the nephron lineage. **A.** Top differentially expressed key transcription factors within nephron lineage clusters. **B.** Signalling pathways active within individual nephron lineage clusters identified by GO and KEGG analysis. Information about which ligands, receptors, and effectors are expressed in each cell type can be accessed in Supplementary file 4.



nephron progenitor clusters. Note this does not show the actual number of cells from each source. **D.** Output from doubletCluster algorithm on whole kidney data. Most likely parent clusters are shown in “source1” and “source2” columns. N = number of unique marker genes for each query cluster, p.value = P-value against the doublet hypothesis for query cluster, lib.size1 = ratio of library sizes of parent1 versus query cluster, lib.size2 = ratio of library sizes of parent2 versus query cluster, prop = proportion of cells making up the query cluster compared to the entire dataset. ‘Suspicious’ clusters have low N, lib.size1 and lib.size2 < 1 and prop < 5%. **E.** Boxplot showing the distributions of doublet scores for the cells in each cluster of the whole kidney dataset. The doubletCells algorithm outputs doublet scores based on simulating pseudo-doublets by randomly selecting two cells in the dataset and adding them together, completely independently of the cluster assignment. High scores indicate higher likelihood of the cell being a doublet. Cluster 14 has markedly higher doublet scores compared to the remaining clusters.

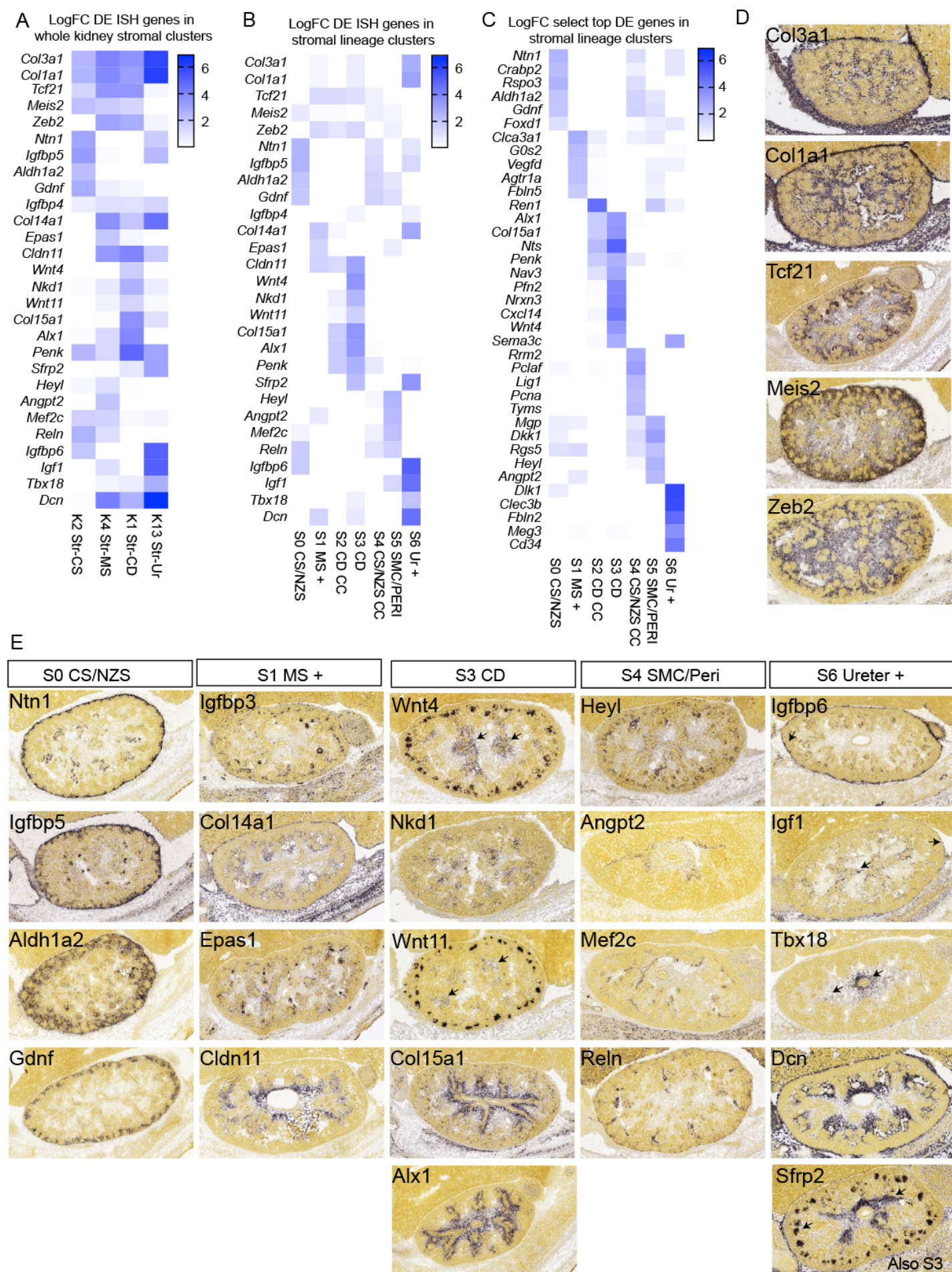


Figure S5. Mapping stromal subpopulations. A. Differential expression of genes with available *in situ* hybridisation (ISH) results* that are also enriched in all stromal clusters or stromal subpopulations in the whole

kidney data. Scale represents log fold change (LogFC) differential expression (DE) within the whole kidney stromal clusters compared to other clusters in the whole kidney. **B.** Differential expression results for the same genes in A within the stromal lineage clustering. Note the low differential expression results for *Col3a1* and *Colla1* indicate a lack of change in expression rather than an absence of expression. **C.** Expression of select top DE genes within each stromal cluster. **D.** ISH results for markers enriched in all or several stromal populations (refer to B for enriched populations). **E.** *ISH results from the Allen Developing Mouse Brain Atlas (<http://developingmouse.brain-map.org>) for stromal cluster enriched genes. Several genes are expressed in other cell types within the developing kidney; their stromal expression domain has been taken into account for this analysis. Some genes pictured are expressed in more than one cluster (refer to B).

Table S1. Differentially expressed genes between 16 clusters within the E18.5 developing mouse kidney. Note: For each spreadsheet, use the “lookup” tab to access an interactive sheet. Input an official gene symbol in the left column to retrieve differential expression results for that gene across all clusters.

[Click here to Download Table S1](#)

Table S2. Putative ligand-receptor interactions within and between all clusters in the whole kidney dataset.

[Click here to Download Table S2](#)

Table S3. Differentially expressed genes between 3 ureteric epithelium subclusters within the E18.5 developing mouse kidney.

[Click here to Download Table S3](#)

Table S4. Differentially expressed genes between 8 nephron and 5 nephron progenitor subclusters within the E18.5 developing mouse kidney.

[Click here to Download Table S4](#)

Table S5. Integrated nephron progenitor cluster markers and differential expression analysis for clusters NP0, NP3, NP4, and NP7.

[Click here to Download Table S5](#)

Table S6. Differentially expressed genes between 6 stromal subclusters within the E18.5 developing mouse kidney.

[Click here to Download Table S6](#)

Table S7. Genes that distinguish distal nephron and ureteric epithelium.

[Click here to Download Table S7](#)

Table S8: Comparison of this scRNA-seq dataset to other developing mouse kidney scRNA-seq data.

Manuscript	Tissue age / stage	Profiling approach	Cell number	Analysis approach	Conclusion
Brunskill et al, 2014, Development	11.5, 12.5, P4 renal vesicle	Fluidigm C1	33	Genespring 12.6.1	Read through of Hox genes, inappropriate expression of presumed lineage markers within CM, partially degraded non-coding RNAs.
Adam et al, 2017, Development	P1	DropSeq	20,000 (in batches of about 4000 cells for each condition of isolation)	Seurat Find All Markers; DEGseq	Use psychrophilic enzymes to avoid c-fos signature, single cell expression profile of the new born mouse kidney.
Magella et al, 2018, Dev Biol	14.5	Drop-Seq, Chromium 10x Genomics and Fluidigm C1	>8000	AltAnalyze	Nephrogenic stroma makes GDNF; stochastic multilineage priming, single cell expression profile of E14.5 kidney.
This study	18.5 whole kidney, 14.5 sorted Six2GFP	Chromium 10x	6732 18.5 kidney, 7853 14.5 sorted Six2GFP.	Seurat, EdgeR, Monocle 2	Expression profile of E18.5 kidney, improved resolution of known cell type markers and signalling pathway component expression. Identification of congruence and new distinct markers for connecting segment and ureteric epithelium. New insight into mouse nephron progenitor heterogeneity.