Single cell resolution regulatory landscape of the mouse kidney highlights cellular differentiation programs and renal disease targets

Zhen Miao[1,2,3,*], Michael S. Balzer[1,2,*], Ziyuan Ma[1,2], Hongbo Liu[1,2], Junnan Wu[1,2], Rojesh Shrestha[1,2], Tamas Aranyi[1,2], Amy Kwan[4], Ayano Kondo[4], Marco Pontoglio[5], Junhyong Kim[6], Mingyao Li[7], Klaus H. Kaestner[2,4] and Katalin Susztak[1,2,4]

[1]Renal, Electrolyte, and Hypertension Division, Department of Medicine, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA
[2]Institute for Diabetes, Obesity, and Metabolism, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA
[3]Graduate Group in Genomics and Computational Biology, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA
[4]Department of Genetics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA
[5]Epigenetics and Development Laboratory, Université de Paris Inserm U1151/CNRS UMR 8253, Institut Necker Enfants Malades, Paris, France
[6]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA
[7]Department of Epidemiology and Biostatistics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA
[#]These two authors contributed equally to this work

Correspondence:
Katalin Susztak, MD, PhD
Professor of Medicine
University of Pennsylvania, Perelman School of Medicine
3400 Civic Center Blvd,
Smilow Translational building 12-123,
Philadelphia, PA 19104
Phone: (215) 898-2009
ksusztak@pennmedicine.upenn.edu

Grant support:
Work in the Susztak lab is supported by the NIH DK076077, DK087635, and DK105821. MSB is supported by a German Research Foundation grant (BA 6205/2-1).

Abstract

The kidney has a very limited capacity to repair. Defining cellular differentiation during development therefore could aid the understanding of homeostatic and maladaptive regeneration.

Here, we profiled open chromatin and gene expression in developing and adult mouse kidneys at single cell resolution. We show critical reliance of gene expression on distal regulatory elements (enhancers). We define key cell type-specific transcription factors and major gene-regulatory circuits for kidney cells. Dynamic chromatin and expression changes during nephron progenitor differentiation demonstrated that podocyte commitment occurs early and is associated with sustained *Foxl1* expression. Renal tubule cells followed a more complex differentiation, where *Hfn4a* was associated with proximal and *Tfap2b* with distal fate. Mapping single nucleotide variants associated with human kidney disease identified critical cell types, developmental stages, genes, and regulatory mechanisms.

We provide a global single cell resolution view of chromatin accessibility of kidney development. The dataset is available via an interactive public website.

Keywords

Kidney development; differentiation; single cell; chromatin accessibility; transcription factor; gene regulatory network; cis-regulatory elements; enhancer; kidney disease.

58 **Introduction**

59

60 The mammalian kidney maintains fluid, electrolyte, and metabolite balance of the body and plays
61 an essential role in blood pressure regulation and red blood cell homeostasis. The human kidney
62 makes roughly 180 liters of primary filtrate each day that is then reabsorbed and modified by a
63 long tubule segment. To perform this highly choreographed and sophisticated function, the kidney
64 contains close to 20 highly specialized epithelial cells. The renal glomerulus acts as a 60 kD size-
65 selective filter. The proximal part of the tubules is responsible for reclaiming more than 70% of
66 the primary filtrate, which is done via unregulated active and passive paracellular transport [1], while
67 the loop of Henle plays an important role in concentrating the urine. The distal convoluted tubule
68 is critical for regulated electrogenic sodium reabsorption and potassium secretion. The last
69 segment of kidney tubules is the collecting duct where the final concentration of the urine is
70 determined via regulation of water channels as well as acid or base secretion. Understanding the
71 development of these diverse cell types in the kidney is essential to understand kidney homeostasis,
72 disease, and regeneration.

73

74 The mammalian kidney develops from the intermediate mesoderm via a complex interaction
75 between the ureteric bud and the metanephric mesenchyme [2]. In the mouse kidney, Six2 marks the
76 self-renewing nephron progenitor population [3]. The nephron progenitors commit and undergo a
77 mesenchymal-to-epithelial transformation giving rise to the renal vesicle [3]. The renal vesicle then
78 undergoes segmentation and elongation, giving rise to epithelia from the podocytes to the distal
79 convoluted tubules, while the ureteric bud becomes the collecting duct. Unbiased and hypothesis-
80 driven studies have highlighted critical stages and drivers of early kidney development [4], that have
81 been essential for the development of *in vitro* kidney organoid differentiation protocols [5-7].
82 However, cells in organoids are still poorly differentiated, improving cellular differentiation and
83 maturation of these structures remains a major challenge [8]. Thus, the understanding of late kidney
84 development, especially the cell type-specific driver transcription factors (TFs) is of great
85 importance [9-11]. Alteration in Wnt, Notch, Bmp, and Egf signaling significantly impacts cellular
86 differentiation, but only a handful of TFs that directly drive the differentiation of distinct segments
87 have been identified, such as *Pou3f3*, *Lhx1*, *Irx2*, *Foxc2* and *Mafb* [12]. Further understanding of the
88 terminal differentiation program could aid the understanding of kidney disease development.

While single cell RNA sequencing (scRNA-seq) has improved our understanding of kidney development in mice and humans [9,10,13,14], it provides limited information of TFs, which are usually lowly expressed. Equally difficult is to understand how genes are regulated from scRNA-seq data alone. Chromatin state profiles, on the other hand, provide valuable insight to gene regulation mechanisms during cell differentiation, since they show not only the accessibility of the gene transcription start site (TSS), but also of distal regulatory regions such as enhancers. It is believed that enhancers are critical for establishing the cell type-specific gene expression pattern, but it has not been shown conclusively on a single cell level. Together with gene expression, open chromatin profiles can define the gene regulatory logic, which is the fundamental element of cell identity. However, there is a scarcity of open chromatin information by Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) or chromatin immunoprecipitation (ChIP) data by ChIP-seq related to kidney development. In addition, epigenetic changes observed in bulk analyses mostly represent changes in cell composition, rather than cell type-specific changes [15], making it challenging to interpret bulk ATAC-seq data.

To this end, here we generated a single cell open chromatin and corresponding expression survey for the developing and adult mouse kidney, which will be available for the community via a searchable website (**http://susztaklab.com/VisCello_snATAC/** and **http://susztaklab.com/igv/**). Using this atlas, we have produced a new epigenome-based classification of developing and mature cells and defined cell type-specific regulatory networks. We also investigated key TFs and cell-cell interactions associated with developmental cellular transitions. Finally, we used the single cell open chromatin information to pinpoint putative target genes and cell types of several chronic kidney disease noncoding genome-wide association study (GWAS) loci.

Results

Single cell accessible chromatin landscape of the developing and adult mouse kidneys

To characterize the accessible chromatin landscape of the developing and adult mouse kidneys at single cell resolution, we performed single nuclei ATAC-seq (snATAC-seq) on kidneys of mice on postnatal day 0 (P0) at 3 and 8 weeks of age (**Figure 1a, Methods**). In parallel, we also performed bulk (whole kidney) ATAC-seq analysis at matched developmental stages. Following sequencing, we aggregated all high-quality mapped reads in each sample irrespective of barcode. The combined snATAC-seq dataset from all samples showed the expected insert size periodicity (**Figure S1a**) with a strong enrichment of signal at TSSs (**Figure S1b**), indicating high data quality. The snATAC-seq data showed high concordance with the bulk ATAC data (Spearman correlation coefficient >0.84, **Methods**, **Figure S1c**).

We next revealed cell type annotations from the open chromatin information. After conducting stringent filtering of the number of barcodes, promoter ratio and mitochondria ratio (**Methods, Figure S1d**), we kept 28,316 cells across the samples (**Figures 1b, S1f-g**). Cells were then clustered using snapATAC [16], which binned the whole genome into 5 kb regions and used diffusion map and principal component analysis for dimension reduction (**Methods**). Prior to clustering, we used Harmony [17], an iterative batch correction method, to correct for variability across samples. Using batch-corrected low dimensional embeddings, we clustered all cells together and retained 13 clusters, all of which had consistent representation across the number of peaks, samples and read depth profiles (**Figures 1b, S1e-g**). As expected, some clusters such as nephron progenitors and stromal cells were enriched in the developing kidney (P0).

In order to identify the cell type-specific open chromatin regions, we conducted peak calling using MACS2 [18] on each cell type separately. The peaks were then merged to obtain a comprehensive open chromatin set. We found that the single nuclei open chromatin set showed good concordance with bulk ATAC-seq samples, with most of the peaks in bulk ATAC-seq data captured by the single nuclei data. On the other hand, single nuclei chromatin accessibility data showed roughly 50% more accessible chromatin peaks (total of 300,693 peaks) than the bulk ATAC-seq data

144 (**Figure 1e, Methods**), indicating that the snATAC-seq data was particularly powerful in
145 identifying open chromatin areas that are accessible in single cell types.
146
147 To determine the cell types represented by each cluster, we examined chromatin accessibility
148 around the TSS and gene body regions of the cognate known cell type-specific marker genes [19].
149 Based on the accessibility of the known marker genes, we identified clusters representing nephron
150 progenitors, endothelial cells, podocytes, proximal tubule segment 1 and segment 3 cells, loop of
151 Henle, distal convoluted tubule, connecting tubule, collecting duct principal cells, collecting duct
152 intercalated cells, stromal and immune cells (**Figure 1b**). **Figures 1d and S1h** show chromatin
153 accessibility information for key cell type marker genes, such as *Uncx* and *Cited1* for nephron
154 progenitors, *Nphs1* and *Nphs2* for podocytes, *Akr1c21* for both segments of proximal tubules,
155 *Slc34a1* and *Slc5a2* for segment 1 of proximal tubules, *Kap* for segment 3 of proximal tubules,
156 *Slc12a1* and *Umod* for loop of Henle, *Scl12a3* and *Pvalb* for distal convoluted tubule, *Trpv5* for
157 connecting tubule, *Aqp2* and *Fxyd4* for principal cells, *Atp6v1g3* and *Atp6v0d2* for intercalated
158 cells, *Egfl7* for endothelial cells, *C1qb* for immune cells and *Col3a1* for different types of stromal
159 cells, respectively [19].
160
161 To understand cell type-specific gene expression changes, we also generated a single cell RNA
162 sequencing (scRNA-seq) atlas for mouse kidney samples at the same developmental stages. The
163 single cell transcriptome profiles of P0 and adult mouse kidneys were derived and processed as
164 described in **Methods**. Rigorous quality control yielded a set of 43,636 single cells (**Figures 1b,**
165 **S1i**). Quality control metrics such as gene counts, UMI counts and mitochondrial gene percentage
166 along with batch correction results are shown in **Figures S1j-m**. By unbiased clustering [20] we
167 obtained 17 distinct cell populations in the combined P0 and adult mouse datasets (**Figure S1i**).
168 On the basis of marker gene expression, we identified kidney epithelial, immune and endothelial
169 cells (**Figures 1f, S1n-o**), closely resembling the clustering obtained from snATAC-seq analysis.
170 We then conducted differential expression analysis on the clusters and identified key marker genes
171 for each cell type (**Supplemental Table 1**).
172
173 To compare the consistency between cluster assignment in the snATAC-seq data and the scRNA-
174 seq data, we next derived a gene activity score for the top 3,000 highly variable genes in each

175 snATAC-seq cluster and computed the Pearson's correlation coefficient between each snATAC
176 cluster and scRNA cluster (**Methods**). This analysis indicated good concordance between the two
177 datasets (**Figures 1g, S1p**). While the correlation between gene expression and inferred gene
178 activity score was high, we noted some differences in cell proportions, which was mostly related
179 to the sample preparation-induced cell drop-out (**Figures S1g, i**). Consistent with previous
180 observations that single cell preparations better capture immune cells than single nuclear
181 preparations [21], we noted that the immune cell repertoire was limited in the snATAC-seq dataset;
182 on the other hand, stromal cells were better captured by the nuclear preparation.
183
184 Finally, to allow the interactive use of this dataset by the community, we not only made the raw
185 data available but also the processed dataset via our searchable website
186 (**http://susztaklab.com/VisCello_snATAC/** and **http://susztaklab.com/igv/**). For example, here
187 we show the chromatin accessibility landscape of *Ace2*, which is of major interest currently due to
188 the COVID19 epidemic. We can observe an open chromatin region around the transcription start
189 site of *Ace2* only in proximal tubules, which is consistent with its expression in proximal tubules
190 (**Figure S1q**).
191
192 **Characterization of the cell type-specific regulatory landscape**
193
194 To characterize different genomic elements captured by snATAC-seq data, we first stratified the
195 genome into promoters, exons, 5' and 3' untranslated regions, introns, and distal regions using the
196 GENCODE annotation [22] (**Methods**). We noticed that consistent with bulk ATAC-seq data, most
197 peaks in snATAC-seq data were in regions characterized as distal elements or introns, (**Figure**
198 **S2a**) and relatively small portions (<10%) were in promoter or 5' untranslated regions. The
199 genomic elements proportion was stable across developmental stages. In addition, almost half of
200 the open chromatin peaks overlapped with P0 or adult H3K27Ac ChIP-seq signals (**Figure S2b**),
201 indicating the contribution of enhancer regions to accessible chromatins.
202
203 To study the open chromatin heterogeneity in different cell types, we derived a cell type-specific
204 accessible chromatin landscape by conducting pairwise Fisher's exact test for each peak between
205 every cluster (Benjamini-Hochberg adjusted q value ≤ 0.05, **Methods**). In total, we identified

206 60,684 differentially accessible open chromatin peaks (DAPs) across the 13 cell types
207 (**Supplemental Table 2, Figure 2a**). Among these peaks, most showed high specificity for a single
208 cluster. However, we noticed overlaps between the S1 and S3 proximal tubule segments-specific
209 peaks, as well as between the loop of Henle and distal convoluted tubule segments, which is
210 consistent with their biological similarities. In addition to the cell type-specific peaks, we also
211 found some cell-type independent open chromatin areas (present across nephron progenitors,
212 podocytes, proximal tubule and loop of Henle cells), likely consistent with basal housekeeping
213 genes and regulatory elements. (**Figures 2a, S2c**).
214
215 We noticed that many genes had strong cell type-specific DAPs at their TSS. Other genes, however,
216 had accessible chromatin at their TSS in multiple cell types. For example, *Umod*, the loop of
217 Henle-specific marker gene, showed accessible chromatin at its TSS at multiple tubule cell types
218 (**Figure S2d, S1h**)**.** Rather than with its TSS, cell type-specific chromatin accessibility of *Umod*
219 strongly correlated with an upstream open chromatin peak, which is likely an enhancer region,
220 indicated by the H3K27Ac ChIP-seq signal (**Figure S2d**). Consistently, we noticed the enrichment
221 of intronic regions and distal elements (**Figures S2e-f**) in cell type-specific DAPs, indicating their
222 role in cell type-specific gene regulation.
223
224 These observations motivated us to study cis-regulatory elements using the snATAC-seq data and
225 scRNA-seq data. We reasoned that a subset of the cell type-specific cis-regulatory elements should
226 regulate cell type-specific gene expression in cis. Inspired by [23], we aligned DAPs and
227 differentially expressed genes from our snATAC-seq and scRNA-seq datasets, and inferred the
228 putative regulatory peaks by their proximity (**Methods**). Such cis-regulatory elements predictions
229 were confirmed by comparing with cis-regulatory elements inferred previously [24], as we
230 recapitulated roughly 20% of elements from their analysis. In addition, our analysis was able to
231 identify several known enhancers such as for *Six2* and *Slc6a18* [24,25] (**Figure S2g**). To quantify the
232 contribution of cis-regulatory elements, we analyzed peak co-accessibility patterns using Cicero
233 [26]. By using a heuristic co-accessible score 0.4 as a cutoff, we identified 1,214,638 and 926,288
234 cis-regulatory element links in the P0 and adult data, respectively. Some of these are likely
235 promoter-enhancer regulatory units. Among these co-accessible elements, only 149,389 were
236 common in P0 and adult kidneys, while most were developmental stage-dependent. While this

observation needs further experimental validation, it highlights dynamic changes in gene regulation during development, which includes both chromatin opening and looping.

Given the large contribution of distal regulatory elements to chromatin accessibility, we next looked into identifying key TFs that occupy the cell type-specific open chromatin regions. Until now, information on cell type-specific TFs in the kidney has been scarce. Therefore, we performed motif enrichment analysis on the cell type-specific open chromatin regions using HOMER [27]. HOMER was designed as a differential motif discovery algorithm that scores motifs by computing enrichment of motif sequences in target compared to a reference set. To reduce false discovery, we focused on the known motifs. The full list of cell type-specific TF binding motifs is shown in **Supplemental Table 3**. Since several TFs have identical or similar binding sequences, we next correlated motif enrichment with scRNA-seq TF expression. Using this combined motif enrichment and gene expression approach, we have defined the mouse kidney cell type-specific TF landscape. Examples include *Six2* and *Hoxc9* in nephron progenitors, *Wt1* and *Mafb* in podocytes, *Hnf4a*, *Ppara*, and *Bhle41* in proximal tubules, *Esrrb* and *Foxa1* in loop of Henle, *Vdr* in distal convoluted tubule, *Elf5* in principal cells, *Tcfcp2l1* in intercalated cells, *Erg* and *Sox17* in endothelial cells, *Spi1* and *Batf* in immune cells, and *Twist1* and *Nr2f2* in stromal cells (**Figures 2a, S2h**).

In order to study the putative target genes of TFs, we examined TF regulon activity using Single-Cell rEgulatory Network Inference and Clustering (SCENIC) [28]. SCENIC was designed to reveal TF-centered gene co-expression networks. By inferring a gene correlation network followed by motif-based filtration, SCENIC keeps only potential direct targets of each TF as modules (regulons). The activity of each regulon in each cell was quantified and then binarized to "on" or "off" based on activity distribution across cells (**Methods**). SCENIC was also able to conduct clustering based on the regulon states of each cell. SCENIC results (**Figure 2b**) indicated strong enrichment in *Trps1*, *Hnf1b*, *Maf*, *Hnf1a*, and *Hnf4a* regulon activity in proximal tubules, *Hmga2*, *Hoxc6*, *Hoxd11*, *Meox1*, *Six2*, *Tcf4*, and *Uncx* in nephron progenitors, *Esrrg*, and *Ppargc1a* in loop of Henle, *Hmgb3* in proliferating cells and *Foxc1*, *Foxc2*, *Foxd1*, *Lef1*, and *Mafb* in podocytes, respectively. While the expression of several of these TFs was relatively low and was further exacerbated by transcript drop-outs, many TFs did not show strong cell type enrichment. The

268 regulon-based analysis, however, showed a very clear enrichment. SCENIC also successfully
269 reported multiple downstream target genes. The full list of regulons and their respective target
270 genes can be found in **Supplemental Table 4,** scaled and binarized regulon activity is also
271 available in **Supplemental Table 5.** Examples of regulon activity, corresponding TF expression,
272 and target gene expression are depicted in **Figures 2c, S2i**. For example, TFs such as *Eya1*, *Hoxc8*,
273 *Hoxc9*, *Pax2*, *Spock2* and *Wnt4* are important downstream targets within the regulon of nephron
274 progenitor-specific TF *Uncx*, indicating an important transcriptional hierarchy of nephron
275 development [29].
276
277 In summary, we generated a comprehensive atlas for the cell type-specific regulatory elements and
278 TF-centered regulatory network.
279
280 **The regulatory trajectory of nephron progenitor differentiation**
281
282 All cells in the body differentiate from the same genetic template. Cell type-specific chromatin
283 opening and closing events associated with TF binding changes set up the cell type-specific
284 regulatory landscape resulting in cell type specification and development. We found that closing
285 of open chromatin regions was the predominant event during the nephron progenitor
286 differentiation (**Figure S2a**). We then evaluated the cellular differentiation trajectory in the
287 snATAC-seq and scRNA-seq datasets (**Methods**). We identified multiple nephron progenitor sub-
288 groups (**Figures 3a-b**), which will need to be carefully mapped to prior gene expression- and
289 anatomical location-driven nephron progenitor sub-classification. Consistently, across both data
290 modalities, we identified that the podocyte precursors differentiated early from the nephron
291 progenitor pool (**Figures 3a-b**). The tubule cell trajectory was more complex with a shared
292 intermediate stage and later differentiation into proximal tubules and distal tubules/loop of Henle
293 (**Figures 3a-b, S3a-c**). We also integrated snATAC-seq and scRNA-seq data to obtain a single
294 trajectory (**Methods**). The cell types in this dataset were correctly mapped and the trajectory
295 resembled the path observed in individual analyses of the scRNA and snATAC datasets (**Figures**
296 **S3e-g**). The robustness of developmental trajectories was further supported by obtaining similar
297 results when performing RNA velocity analysis using Velocyto [30] (**Figure S3d**) and by comparing
298 with previous human and mouse kidney developmental studies [9,13,14].

299

300 Building on both the SCENIC-generated gene regulatory network and the robust differentiation
301 trajectories of the snATAC-seq and scRNA-seq datasets, we next aimed to understand chromatin
302 dynamics, identify TFs and driver pathways for cell type specification and differentiation. To this
303 end, we first determined variation in chromatin accessibility along the 3 differentiation trajectories
304 using ChromVAR [31]. ChromVAR estimates the accessibility dynamics of motifs in snATAC-seq
305 data (**Methods**). We observed three different patterns when analyzing genes of interest (**Figure**
306 **3c**): *1) Decrease of TF motif accessibility in all lineages.* For example, *Sox11* motif enrichment
307 score was high in nephron progenitor cells at the beginning of all 3 trajectories. It then decreased
308 in all 3 lineages in parallel, underlining the role of *Sox11* in early kidney development. Several
309 other TFs followed this pattern such as *Six2* and *Sox9. 2) Cell type-specific maintenance of*
310 *chromatin accessibility with advancing differentiation.* We observed that chromatin accessibility
311 for the *Wt1* motif was high initially, but declined in proximal tubule and loop of Henle lineages,
312 while its expression increased in the podocyte lineage. This is consistent with the important role
313 of *Wt1* in nephron progenitors and podocytes [32,33]. Other TFs that followed this pattern include
314 *Foxc2* and *Foxl1*. *3) A de novo increase in chromatin accessibility with cell type commitment and*
315 *advancing differentiation.* For example, the chromatin accessibility of *Hnf4a* and *Pou3f3* motif
316 increased in proximal tubule and loop of Henle trajectories, respectively, coinciding with the
317 cellular differentiation program [34]. A large number of TFs followed this pattern such as *Mafb* (in
318 podocytes), *Hnf4a* and *Hnf1a* (in proximal tubule), *Hnf1b* (in both proximal tubule and loop of
319 Henle) as well as *Esrrb* and *Tfap2b* (in loop of Henle).

320

321 Next, we correlated changes in chromatin accessibility-based TF motif enrichment with TF
322 expression and their respective target genes along Monocle-generated trajectories. To this end, we
323 used the scRNA-seq differentiation trajectories to find TFs and target genes differentially
324 expressed over pseudotime (**Supplemental Table 6)**. We also noticed a good concordance of time-
325 dependent changes of TF and target gene expression along with TF motif enrichment, including
326 the lineages for podocytes (e.g., *Foxc2*, *Foxl1*, *Mafb*, *Magi2*, *Nphs1*, *Nphs2*, *Plat*, *Synpo*, *Thsd7a*,
327 *Wt1*, and *Zbtb7c*), proximal tubule (e.g., *Ace2*, *Atp1a1*, *Dab2*, *Hnf1a*, Hnf4a, *Hsd17b2*, *Lrp2*, *Maf*,
328 *Slc12a3*, *Slc22a12*, *Slc34a1*, and *Wnt9b*), loop of Henle (e.g.,*Cyfip2*, *Cytip*, *Esrrb*, *Esrrg*, *Irx1*,
329 *Irx2*, *Mecom*, *Pla2g4a*, *Pou3f3*, *Ppargc1a*, *Stat3*, *Sytl2*, *Tfap2b*, *Thsd4*, and *Umod*), as well as for

330 both proximal tubule and loop of Henle (e.g., *Bhlhe40*, *Hnf1b*, and *Tmprss2*), respectively (**Figures**
331 **3c, S3i**). Most interestingly, we noticed two distinct patterns of how gene expression was related
332 to chromatin accessibility. While gene expression of TFs increased over pseudotime, its
333 corresponding motif accessibility either increased in parallel (such as *Hnf4a* and *Pou3f3*) or
334 maintained in a lineage-specific manner (such as *Wt1*). This might indicate different regulatory
335 mechanisms during differentiation.
336
337 We next aimed to interrogate the stage-dependent chromatin dynamics along the identified
338 differentiation trajectory. The differentiation trajectory was binned into 15 developmental steps
339 based on the lineage specification (**Figures S3b-c**). These stages were labeled as NP (nephron
340 progenitor), IM (intermediate cells), Podo (podocytes), PT (proximal tubule), LOH (loop of Henle),
341 and DCT (distal convoluted tubule), however, this designation will need to be matched with prior
342 cell marker-based annotations. To study the chromatin opening and closing, we conducted
343 differential chromatin accessibility analysis between subsequent stages. To understand the
344 biological processes controlled by the epigenetic changes, we examined the nearest genes and
345 performed functional annotation (**Methods**). We found that open chromatin profiles were
346 relatively stable in the early precursor stages such as NP1 to NP3, with fewer than 70 DAPs
347 identified (**Supplemental Table 7, Figure S4a**). The podocyte differentiation branch was
348 associated with marked increase in the number of DAPs, (796 DAPs between NP3 and Podo1).
349 This mainly represented the closing of chromatin areas around nephron progenitor-specific genes
350 such as *Osr1*, *Gdnf*, *Sall1*, *Pax2* and opening of areas around podocyte-specific genes and key TFs
351 such as *Foxc2* and *Efnb2*, both of which are validated to be important for early podocyte
352 differentiation [35,36]. At later stages, there was a strong increase in expression of actin filament-
353 based processes and a significant decrease in *Notch* and *Ctnnb1* in the podocyte lineages
354 (**Supplemental Table 8**). Fewer chromatin closing events were observed (234 DAPs) between
355 NP3 and intermediate cells 1 (IM1), mainly associated with closing of the chromatin around *Osr1*
356 and opening around tubule cell-specific TFs such as *Lhx1* and *Pax3* (**Figure 4**). The decrease in
357 *Six2* expression only occurred at the IM2 stage, at which we also observed an increase in tubule
358 specification genes such as *Hnf1a*. Gene ontology results from the 820 up-regulated peaks between
359 PT1 and IM2 showed enrichment associated with typical proximal tubule functions including
360 sodium-dependent phosphate transport, maintenance of osmotic response in the loop of Henle and

361 active sodium transport in the distal convoluted tubule (**Figure S4a**, the full list can be found in
362 **Supplemental Tables 7, 8 and 9)**.
363
364 In addition to analyzing changes along the trajectory, we also specifically examined cell-fate
365 decision events. We studied the chromatin opening and closing during the first cell commitment
366 event. We found that podocyte specification from nephron progenitors was associated with
367 differential opening of *Foxl1*, *Zbt7c,* and *Smad2* in the podocyte lineage and *Lhx1*, *Sall1*, *Dll1*,
368 *Jag1*, *Cxcr3* and *Pax3* in the other lineage, respectively. While the role of several TFs has been
369 established for podocyte specification, the expression of *Foxl1* has not been described in the kidney
370 until now (**Figure 4**). Our analysis pinpointed that four peaks in the vicinity of *Foxl1* were
371 accessible only in podocyte lineage, which locate in +53,381 bp, +152,832 bp, +237,019 bp, and
372 +268,550 bp of the *Foxl1* TSS, respectively. To confirm the expression of *Foxl1* in nephron
373 progenitors and podocytes, we performed immunofluorescence studies on developing kidneys
374 (E13.5, P0 and P6). Consistent with the computational analysis, we found strong expression of
375 FOXL1 in nephron progenitors (E13.5). At later stages, it was present in comma and S shape body
376 and finally in the glomerular podocytes (**Figure S4b**). Consistently, while there was no expression
377 within cells destined to become proximal tubule or loop of Henle cells, gene expression of *Foxl1*
378 increased in cells along the podocyte trajectory (**Figure S4c**). While further experimental
379 validation will be important, our study has illustrated the critical role of open chromatin state
380 information and dynamics in cellular differentiation.
381
382 The intermediate cells (IM) gave rise to proximal and distal branches, representing the proximal
383 tubules and the loop of Henle as well as distal convoluted tubule segments. The proximal tubule
384 region was characterized by chromatin opening around *Hnf4a, Maf*, *Tprkb*, and *Gpat2*. The loop
385 of Henle and distal convoluted tubule segments were remarkable for multiple DAPs in the vicinity
386 of *Tfap2a*, *Tfap2b*, *Cited4*, *Ephb2*, *Ephb3*, *Hoxd8*, *Mecom*, and *Prmd16*, indicating a critical novel
387 role for these TFs in distal tubule differentiation (**Figures 3c, 4, S4c**). Consistently, we saw a
388 reduction in chromatin accessibility of *Six2* promoter and enhancers along all three trajectories
389 (podocyte, proximal tubule and loop of Henle) (**Figure S4d**). There was also a decrease in
390 expression of *Jag1* and *Heyl* in the distal loop of Henle segment, consistent with the putative role
391 of Notch driving the proximal tubule fate [37] **(Supplemental Table 11)**. Another striking

392 observation was that tubule segmentation and specification occurred early by an increase in
393 chromatin accessibility around *Lhx1*, *Hnf1a* and *Hnf4a* and *Maf* for proximal tubule and *Tfap2b*
394 for loop of Henle. Terminal differentiation of proximal tubule and loop of Henle cells was strongly
395 linked to nuclear receptors that regulate metabolism, such as *Esrra* and *Ppara* in proximal tubules
396 and *Esrra* and *Ppargc1a* in the loop of Henle segment, once more indicating the critical role of
397 metabolism of driving gene expression and differentiation [38].
398
399 In summary, we reconstructed the developmental and differentiation trajectories of podocytes,
400 proximal tubule and loop of Henle cells. We defined chromatin and gene expression dynamics and
401 identified numerous putative TFs for kidney cell specification and differentiation.
402
403 **Stromal-to-epithelial communication is critical in the developing and adult kidneys**
404
405 Previous studies indicated that the survival, renewal, and differentiation of nephron progenitors is
406 largely regulated through its cross-talk with the adjacent ureteric bud [39]. To investigate the complex
407 cellular communication network, we used CellPhoneDB [40] to systematically infer potential cell-
408 cell communication in the developing and adult kidney. CellPhoneDB provides a comprehensive
409 database and a statistical method for the identification of ligand-receptor interactions in scRNA-
410 seq data. Analysis of our scRNA-seq dataset indicated that the number of cell-cell interaction pairs
411 was larger in developing kidney compared to the adult kidney (**Figure 5a**). In the developing
412 kidney, the stroma showed the greatest number of interactions among all cell types. This is
413 consistent with the well-known role of epithelial-stromal interactions in driving kidney
414 development. Of the identified interactions, many were related to stroma-secreted molecules such
415 as collagen 1, 3, 4, 6, and 14 (**Figure 5b**). Furthermore, the stroma seemed to interact with most
416 cell types, such as podocytes and different tubule cells. Interestingly, the nephron progenitor
417 cluster showed important ligand-receptor interaction between *Fgf1*, *Fgf8* as well as *Fgf9* and the
418 corresponding receptor *Fgfr1*, which is consistent with the well-known role of FGF signaling in
419 kidney development [41]. Of the manifold identified interactions in the fetal kidney, stromal
420 interaction and the VEGF-involving interaction remained significant in the adult data set,
421 underscoring the importance of endothelial-to-epithelial communication.
422

423 We next individually examined the expression of several key pathways known to play important
424 roles in kidney development, such as Gdnf-Ret, sonic hedgehog, FGF, Bmp, Wnt and others [9].
425 Expression of these key ligand-receptor pairs showed strong cell type specificity (**Figure 5c**). For
426 example, *Robo2* of the Gdnf-Ret pathway was expressed in nephron progenitors and in podocytes
427 of P0 and adult kidney. Gdnf signaling through the Ret receptor is required for normal growth of
428 the ureteric bud during kidney development [42] and the Slit2/Robo2 pathway is implicated with
429 congenital kidney anomalies (Hwang et al., Hum Genet 2015) and important for maintenance of
430 podocyte foot process integrity [43]. *Eya1*, however, is genetically upstream of *Gdnf* and acts as a
431 positive regulator for its activation [44]. Consistently, we noted distinct cell type specificity of *Eya1*
432 expression only in nephron progenitors, which was also true for other important signaling
433 molecules such as *Ptch1, Smo* and *Gli3* of the sonic hedgehog pathway. *Fgfr1* showed the highest
434 expression in nephron progenitors as well as in fetal and adult stroma, underscoring the importance
435 of FGF signaling for cell-cell interactions in both the developing and developed kidney. Most
436 interestingly, some cell-cell interactions between specific cell types that were observed in fetal
437 kidney were abrogated in adult kidney because of the loss of expression of either ligand or receptor,
438 such as *Pdgfc* in nephron progenitor signaling to its receptor *Pdgfra* in stroma, *Npnt* from several
439 epithelial cells signaling to *Itga8* in nephron progenitors, *Tnc-Itga9* signaling from nephron
440 progenitors to stroma and *Rspo3* in stroma signaling to *Sdc4* in several epithelial cells. Because
441 not much is known about some of these markers, the significance of these putative interactions
442 requires further investigation. For example, *Rspo3* has been implicated in nephron progenitor-
443 associated interactions during nephrogenesis [10]. Mutations in the *Itga8* gene are known to cause
444 isolated congenital anomalies of kidney and urinary tract in humans [45] and *Pdgfra* has been
445 regarded as a commitment marker in kidney differentiation [10].
446
447 In summary, we inferred cell-cell interactions in the developing and adult kidneys and found the
448 critical role of stromal-epithelial interactions in the developing kidney.
449

450 **Single cell chromatin accessibility identified human kidney GWAS target regulatory regions,**
451 **genes and cell types**
452
453 Finally, we examined whether single cell level chromatin accessibility data can help identify cell
454 and gene targets for human kidney disease development. GWAS have been exceedingly successful
455 in identifying nucleotide variations associated with specific diseases or traits. However, more than
456 90% of the identified genetic variants are in the non-coding region of the genome. Initial
457 epigenome annotation studies indicated that GWAS hits are enriched in tissue-specific enhancer
458 regions. As there are many different cell types in the kidney with differing function, understanding
459 the true cell type specificity of these enhancers is critically important. Here, we reasoned that
460 single cell accessible chromatin information could be extremely useful to identify the cell type-
461 specific enhancer regions and thereby the target cell type for the GWAS hits, however, such maps
462 have not been generated for the human kidney. We combined three recent kidney disease GWAS
463 [46-48], and obtained 26,637 single nucleotide polymorphisms (SNPs) that passed genome-wide
464 significance level of which we retained 7,923 after lift-over from human to mouse.
465
466 Specifically, we examined loci where functional validation studies reported conflicting results on
467 target cell types and target genes (**Figure 6).** The *SHROOM3* locus has shown a reproducible
468 association with kidney function in multiple GWAS [48]. However, previous functional follow-up
469 studies have reported confusing and somewhat contradictory results. While one study indicated
470 that the genetic variants were associated with an increase in SHROOM3 levels in tubule cells
471 inducing kidney fibrosis [49], the other suggested that the variant was associated with lower
472 SHROOM3 levels in podocytes resulting in chronic kidney disease development [50]. We found an
473 open chromatin (likely promoter) area in multiple cell types such as nephron progenitors,
474 podocytes, loop of Henle, distal convoluted tubule, principal cells and intercalated cells (**Figure**
475 **6a**). We also identified intronic open chromatin areas only in nephron progenitors and podocytes
476 that overlapped with the GWAS significant variants (**Figure 6a**). Consistent with the cis-
477 regulatory open chromatin, the strongest expression of *Shroom3* was observed in podocytes and
478 nephron progenitor cells. Expression of *Shroom3* in the adult bulk kidney was below our detection
479 limit. To further understand the regulatory dynamics of this locus in the developing mouse kidneys,
480 we examined gene expression and epigenome annotation data generated from bulk mouse kidney

481 samples at different stages of development for H3K27ac and H3K4me1 in adult and fetal samples
482 (**Figure 6b**). Interesting to note that the GWAS-significant SNP that showed strong nephron
483 progenitor-specific enrichment also coincided with the *Six2* binding area. Finally, the Cicero-based
484 co-accessible analysis connected the GWAS top variants, which located in an intronic enhancer
485 region of *Shroom3*, with *Shroom3* exons, indicating that *Shroom3* is the likely target gene of the
486 variant (**Figure 6a**).
487
488 Next, we analyzed the chromosome 15 GWAS region, where we identified some open chromatin
489 regions that were uniformly open in all examined cell types. *Dab2* expression, on the other hand,
490 strongly correlated with open distal enhancer regions in proximal tubule cells (**Figure 6c**). This is
491 consistent with earlier publications indicating the role of proximal tubule-specific DAB2 playing
492 a role in kidney disease development [51]. Interestingly, while single cell analysis indicated an
493 additional distal enhancer in intercalated cells, the GWAS-significant region coincided with the
494 proximal tubule-specific enhancer region and showed strong coregulation (**Figure 6d**). Regulatory
495 annotation of the developing kidney indicated strong enhancer marks in the adult but not in the
496 fetal kidney.
497
498 Lastly, we examined the region around *Uncx*, for which reproducible association with kidney
499 function was shown in multiple GWAS [46,47]. Interestingly, the GWAS locus demonstrated a strong
500 open chromatin region in nephron progenitors but not in any other differentiated cell types (**Figure**
501 **6e**). Consistently, in bulk chromatin accessibility data we only observed regulatory activity such
502 as H3K27ac, H3Kme1 and we show *Six2*-binding at this locus in fetal kidneys. The locus did not
503 show H3K27ac enrichment in the adult kidney, while H3K4me1 remained positive (**Figure 6f**).
504 *Uncx* expression was strong in the fetal kidney samples, but we could not detect its expression in
505 the adult kidney (**Figure 6e**). A closer view of these loci is shown in **Figure S6**.
506
507 These results indicate that variants associated with kidney disease development are located in
508 regions with cell type- and developmental stage-specific regulatory activity and illustrate the
509 critical role of snATAC-seq in defining target genes and target cell types for GWAS variants.

510 **Discussion**

511

512 In summary, here we present the first cellular resolution open chromatin map for the developing
513 and adult mouse kidney. Using this dataset, we identified key cell type-specific regulatory
514 networks for kidney cells, defined the cellular differentiation trajectory, characterized regulatory
515 dynamics and identified key driving TFs for nephron development, especially for the terminal
516 differentiation of epithelial cells. Furthermore, our results shed light on the cell types and target
517 genes for genetic variants associated with kidney disease development.

518

519 By performing massively parallel single cell profiling of chromatin state, we were able to define
520 the key regulatory logic for each kidney cell type by investigating cis-regulatory elements and TF-
521 target gene interaction. We found that most cell type-specific open chromatin regions are within
522 distal regulatory elements and intronic regions. Our studies identified a massive amount of highly
523 dynamic co-regulated peaks indicating the important correlation between distal regulatory
524 elements and gene expression. Future studies will examine the relative contribution of promoters
525 and enhancer openness in gene expression regulation. However, these studies highlight that both
526 chromatin opening and looping are critical for gene regulation.

527

528 We also observed that the single cell open chromatin atlas was able to define more distinct cell
529 types even in the developing kidney compared to scRNA-seq analysis. Given the continuous nature
530 of RNA expression, it has been exceedingly difficult to dissect specific cell types in the developing
531 kidney [9,10,13]. In addition, it has been difficult to resolve the cell type origin of lowly expressed
532 transcripts in scRNA-seq data. However, this is not the case for snATAC-seq data, which were
533 able to capture the chromatin state irrespective of gene expression magnitude. There were several
534 examples where accessible peaks were identified in specific cell types even for lowly expressed
535 genes such as Shroom3.

536

537 We identified critical cell type-specific TFs by integrating multiple computational analyses. TF
538 identification is challenging in scRNA-seq data since the expression of several cell type-specific
539 TFs is low and some of them do not show a high degree of cell type-specificity. By extracting
540 motif information, snATAC-seq data provides additional information for TF identification.

541 Together with regulon analysis, as implemented in SCENIC, we have identified several TFs as
542 well as their target genes that are important for kidney development. Leveraging this newly
543 identified cell type-specific regulatory network will be essential for future studies of cellular
544 reprogramming of precursors into specific kidney cell types and for better understanding
545 homeostatic and maladaptive regeneration.
546
547 Our studies revealed dynamic chromatin accessibility that tracks with renal cell differentiation.
548 These states may reveal mechanisms governing the establishment of cell fate during development,
549 in particular those underlying the emergence of specific cell types. We found a consistent and
550 coherent pattern between gene expression and open chromatin information, where the nephron
551 progenitors differentiated into two branches representing podocytes and tubule cells [52]. We found
552 that podocytes commitment occurred earlier, while tubule differentiation and segmentation
553 appeared to be more complex. This podocyte specification correlated with the maintenance of
554 expression of *Foxc2* and *Foxl1* expression in podocytes. While *Foxc2* has been known to play a
555 role in nephron progenitors and podocytes, this is the first description of *Foxl1* in kidney and
556 podocyte development. Our studies are consistent with recent observations from organoid models
557 that recapitulated podocyte differentiation better than tubule cell differentiation [53]. Our study also
558 sheds light on tubule differentiation and segmentation. We confirmed the key role of *Hnf4a* in
559 proximal tubules. We have identified a large number of new transcriptional regulators such as
560 *Tfap2a* that seem to be critical for the distal portion of the nephron. Our data indicate that distal
561 tubule differentiation is linked to the loop of Henle, a critically important observation needing
562 further confirmation. Furthermore, the terminal differentiation of proximal tubule cells correlated
563 with the increase in *Ppara* and *Esrra* expression, both of which are known regulators of oxidative
564 phosphorylation and fatty acid oxidation [38]. Loop of Henle differentiation strongly correlated with
565 *Essrb* and *Ppargc1a* expression. These studies potentially indicate that cell specification events
566 occur early and metabolism controls terminal differentiation of tubule cells [54]. Impaired metabolic
567 fitness of proximal tubules has been a key contributor to kidney dysfunction, explaining the critical
568 association with tubule metabolism and function.
569
570 Furthermore, we show that single cell and stage level epigenome annotation is critical for the
571 annotation of human GWAS. Most identified GWAS signals are in the non-coding region of the

572 genome. Due to the linkage disequilibrium structure of the human genome, each GWAS locus
573 contains a large number of variants, each passing genome-wide significance level [55]. Furthermore,
574 as these signals are often non-coding, the target gene and the target cell type remain unknown.
575 While molecular quantitative trait locus studies and bulk epigenome annotation experiments have
576 been important to define the molecular pathways leading to disease development from the
577 identified signals, these methods have limited resolution, as cell type-specific enhancer regions
578 cannot be identified by bulk analysis [56]. Additionally, bulk molecular quantitative trait locus
579 studies suffer from the same linkage disequilibrium problems as GWAS analyses [57]. Our results
580 indicate that multiple GWAS regions are conserved between mice and humans. Single cell open
581 chromatin information enables not only the identification of affected cell types, but also the
582 understanding of co-regulation of the open chromatin area. It is also able to highlight critical target
583 genes. Performing single cell open chromatin analysis on human kidney tissue samples will be
584 essential to further understand molecular pathways altered by genetic variants. Here we showed
585 three important examples. We confirmed the role of *Dab2* and its specific expression in the
586 proximal tubule during kidney disease development, as its implication therein has been shown in
587 previous expression quantitative trait locus and bulk epigenome analysis experiments [51].
588 Furthermore, we showed that the GWAS variants map only to those regions where chromatin is
589 open exclusively in nephron progenitors, whereas chromatin becomes inaccessible as
590 differentiation progresses during later stages, such as *Shroom3* and *Uncx*. This is an interesting
591 and important novel mechanism, indicating that the altered expression of this gene might play a
592 role in the development rewiring of the kidney. This mechanism is similar to genes associated with
593 autism that are known to be expressed in the fetal but not in the adult stages [58] and highlights the
594 critical role of understanding chromatin accessibility at multiple stages of differentiation.
595
596 While we have generated a large amount of high-quality data, this information will need further
597 experimental validation, which is beyond the scope of the current manuscript. In addition, one
598 needs to be aware of the limitations when interpreting different computational analyses, for
599 example, the motif enrichment analyses such as implemented by HOMER, SCENIC, and
600 chromVAR, are not able to distinguish between TFs with similar binding sites. Future high-
601 throughput studies that analyze open chromatin and gene expression information from the same

602 cells will be exceedingly helpful to correlate open chromatin and gene expression information
603 along the differentiation trajectory [24,59,60].
604
605 In summary, our dataset provides critical novel insight into the cell type-specific gene regulatory
606 network, cell differentiation program, and disease development.

607 **Materials and Methods**

608

609 **Single cell RNA sequencing of P0 mice**

610 1-day-old mouse neonate was decapitated with surgical scissors, 2 kidneys were harvested and
611 minced into 1 mm^3 pieces and incubated with digestion solution containing Enzyme D, Enzyme R
612 and Enzyme A from Multi Tissue Dissociation Kit (Miltenyi, 130-110-201) at 37 °C for 15 min
613 with agitation. Reaction was deactivated by adding 10% FBS, then solution was passed through a
614 40 μm cell strainer. After centrifugation at 1,000 RPM for 5 min, cell pellet was incubated with
615 500 μL of RBC lysis buffer on ice for 3 min. We centrifuged the cells at 1,000 RPM for 5 min at
616 4 °C and resuspended the cells in the buffer for further steps. Cell number and viability were
617 analyzed using Countess AutoCounter (Invitrogen, C10227). The cell concentration was 2.2
618 million cells/mL with 92% viability. 10,000 cells were loaded into the Chromium Controller (10X
619 Genomics, PN-120223) on a Chromium Single Cell B Chip (10X Genomics, PN-120262) and
620 processed to generate single cell gel beads in the emulsion (GEM) according to the manufacturer's
621 protocol (10X Genomics, CG000183). The library was generated using the Chromium Single Cell
622 3' Reagent Kits v3 (10X Genomics, PN-1000092) and Chromium i7 Multiplex Kit (10X Genomics,
623 PN-120262) according to the manufacturer's manual. Quality control for constructed library was
624 performed by Agilent Bioanalyzer High Sensitivity DNA kit (Agilent Technologies, 5067-4626)
625 for qualitative analysis. Quantification analysis was performed by Illumina Library Quantification
626 Kit (KAPA Biosystems, KK4824). The library was sequenced on an Illumina HiSeq or NextSeq
627 2x150 paired-end kits using the following read length: 28 bp Read1 for cell barcode and UMI, 8
628 bp I7 index for sample index and 91 bp Read2 for transcript.

629

630 **Single cell ATAC sequencing**

631 3-week-old and 8-week-old mice were euthanized and perfused with chilled 1x PBS via left
632 ventricle. Kidneys (0.25 g) were harvested, minced and lysed in 5 mL lysis buffer for 15 min. 1-
633 day-old mice were decapitated with surgical scissors, and both kidneys were harvested. Kidneys
634 were minced and lysed in 2 mL lysis buffer for 15 min. Tissue lysis reaction was then blocked by
635 adding 10 mL 1x PBS into each tube, and solution was passed through a 40 μm cell strainer. Cell
636 debris and cytoplasmic contaminants were removed by Nuclei PURE Prep Nuclei Isolation Kit
637 (Sigma, NUC-201) after centrifugation at 13,000 RPM for 45 min. Nuclei concentration was

638 calculated with Countess AutoCounter (Invitrogen, C10227). Diluted nuclei suspension was
639 loaded and incubated in transposition mix from Chromium Single Cell ATAC Library & Gel Bead
640 Kit (10X Genomics, PN-1000110) by targeting 10,000 nuclei recovery. GEMs were then captured
641 on the Chromium Chip E (10x Genomics, PN-1000082) in the Chromium Controller according to
642 the manufacturer's protocol (10X Genomics, CG000168). Libraries were generated using the
643 Chromium Single Cell ATAC Library & Gel Bead Kit and Chromium i7 Multiplex Kit N (10X
644 Genomics, PN-1000084) according to the manufacturer's manual. Quality control for constructed
645 library was perform by Agilent Bioanalyzer High Sensitivity DNA kit. The library was sequenced
646 on an Illumina HiSeq 2x50 paired-end kits using the following read length: 50 bp Read1 for DNA
647 fragments, 8 bp i7 index for sample index, 16 bp i5 index for cell barcodes and 50 bp Read2 for
648 DNA fragments.
649
650 **Bulk ATAC sequencing**
651 Bulk ATAC-seq was performed as described earlier [61,62]. Briefly, 50,000 nuclei/sample were
652 tagmented with Tn5 transposase (Illumina) in 50 µl reaction volume including Tween-20 (0.1%)
653 (Sigma) and digitonin (0.01%) (Promega). The reaction was carried out at 37 °C for 30 min in a
654 thermomixer at 1,000 RPM. After purification of DNA with Qiagen Minelute Reaction Cleanup
655 kit (Qiagen), samples were subjected to library amplification (8-10 cycles). Libraries were purified
656 with AmpureXP beads (Beckman Coulter) and their quality was assessed by Agilent High
657 sensitivity DNA Bioanalysis chip (Agilent). Libraries were submitted to 150 bp PE sequencing.
658
659 **snATAC-seq data analysis**
660 *Data processing and quality control*
661 Raw fastq files were aligned to the mm10 (GRCm38) reference genome and quantified using Cell
662 Ranger ATAC (v. 1.1.0). We only kept valid barcodes with number of fragments ranging from
663 1,000 to 40,000 and mitochondria ratio less than 10%. One of the important indicators for ATAC-
664 seq data quality is the fraction of peaks in promoter regions, so we did further filtration based on
665 promoter ratio. We noticed the promoter ratio seemed to follow a binary distribution, with most of
666 cells either having a promoter ratio around 5% (background) or more than 20% (valid cells)
667 (**Figure S1d**). We therefore filtered out cells with a promoter ratio <20%. After this stringent
668 quality control, we obtained 11,429 P0 single cells (5,993 in P0_batch_1 and 5,436 in P0_batch_2)

669 and 16,887 adult single cells (7,129 in P56_batch_3, 6,397 in P56_batch_4, and 3,361 in
670 P21_batch_5).
671
672 *Preprocessing*
673 Since snATAC-seq data are very sparse, previous methods either conducted peak calling or
674 binarization before clustering. Here, we chose to do binarization instead of peak calling for two
675 reasons: 1) Peak calling is time consuming; 2) Many peaks are cell type-specific, open chromatin
676 regions in rare populations are more likely to be treated as background. After binarizing fragments
677 into 5 kb bins and removing the fragments not matched to chromosomes or aligned to the
678 mitochondria, we binarized the cell-bin matrix. In order to only keep bins that were informative
679 for clustering, we removed the top 5% most accessible bins and bins overlapping with ENCODE
680 blacklist. The 484,606 remaining bins were used as input for clustering.
681
682 *Dimension reduction, batch effect correction and clustering*
683 Clustering was conducted using snapATAC [16], a single-cell ATAC-seq algorithm scalable to large
684 dataset. Previous benchmarking evaluation has shown that snapATAC was one of the best-
685 performing methods for snATAC-seq clustering [63]. Diffusion map was applied as a dimension
686 reduction method using function *runDiffusionMaps*. To remove batch effect, we used Harmony [17],
687 in which the low dimensional embeddings obtained from the diffusion map were used as input.
688 Harmony iteratively pulled batch-specific centroid to cluster centroid until convergence to remove
689 the variability across batches. After batch correction, a graph was constructed using k Nearest
690 Neighbor (kNN) algorithm with k=15, which was then used as input for Louvain clustering. We
691 used the first 20 dimensions for the Louvain algorithm. The number of dimensions was chosen
692 using a method recommended by snapATAC, although we noticed that the clustering results were
693 similar among a series of dimensions from 18 to 30.
694
695 *Cell type annotation*
696 We used a published list of marker genes [9,19] to annotate kidney cell types. In order to infer gene
697 expression of each cell type, we built a cell-gene activity score matrix by integrating all fragments
698 that overlapped with gene transcript. We used GENCODE Mouse release VM16 [22] as reference
699 annotation.

Peak calling and visualization

Peak calling was conducted for each cell type separately using MACS2 [18]. We aggregated all fragments obtained from the same cell types to build a pseudo-bulk ATAC data and conducted peak calling with parameters "--nomodel --keep-dup all --shift 100 --ext 200 --qval 1e-2 -B --SPMR --call-summits". By specifying "--SPMR", MACS2 generated "fragment pileup per million reads" pileup files, which were converted to bigwig format for visualization using UCSC bedGraphToBigWig tool.

We also visualized public chromatin ChIP-seq data and RNA-seq data obtained from ENCODE Encyclopedia (https://www.encodeproject.org/) with the following identifiers: ENCFF338WZP, ENCFF872MVE, ENCFF455HPY, ENCFF049LRQ, ENCFF179NTO, ENCFF071PID, ENCFF746MFH, ENCFF563LOO, ENCFF184AYF, ENCFF107NQP, ENCFF465THI, ENCFF769XWI, ENCFF591DAX. The Six2 ChIP-seq data were obtained from [64] and the WGBS data were obtained from [65].

Genomic elements stratification

Mouse mm10 genome annotation files were download from UCSC Table Browser (https://genome.ucsc.edu/cgi-bin/hgTables) using GENCODE VM23. TSS upstream 5 kb regions were included as promoter regions, but the results were similar when using 2 kb upstream regions as promoters. We then studied the number of overlapped regions between open chromatin regions identified from the snATAC-seq and bulk ATAC-seq dataset and genome annotations. Since one open chromatin region could overlap with multiple genomic elements, we defined an order of genomic elements as exon > 5'-UTR > 3'-UTR > intron > promoter > distal elements. To be more specific, if one peak overlapped with both exon and 5'-UTR, the algorithm would count it as an exon-region peak.

Identification of differentially accessible regions

Peaks identified in each cell type were combined to build a union peak set. Overlapping peaks were then merged to one peak using *reduce* function from the GenomicRanges package. This resulted in 300,755 peaks, which were used to build binarized cell-by-peak matrix. Differentially

accessible peaks (DAPs) for each cell type were identified by pairwise peak comparison. Specifically, for each peak, we conducted a Fisher's exact test between a cell type and each of the other cell types. To address multiple testing problem, we used the Benjamini-Hochberg approach (BH correction) to correct p values. Peaks with corrected p values below significance level (0.05) in all pairwise tests were defined as DAPs. In total, we obtained 60,683 DAPs, which were used for motif enrichment analysis.

Motif enrichment analysis

Motif enrichment analysis was conducted using DAPs by HOMER v4.10.4 [27] with parameters background="automatic" and scan.size=300. We noticed that *de novo* motif identification only generated few significant results, so we focused on known motifs for our following study. We used the significance level of 0.05 for BH corrected p value to determine the enriched results. The motif enrichment results are provided in **Supplemental Table 3**.

Peak-peak correlation analysis

Peak-peak correlation analysis was conducted using Cicero [26]. In order to find developmental stage-specific peak-peak correlations, the analysis was conducted for P0 and adult separately. Cicero uses Graphic Lasso with distance penalty to assess the co-accessibility between different peaks. Cicero analysis was conducted using the *run_cicero* function with default parameters. A heuristic cutoff of 0.25 score of co-accessibility was used to determine the connections between two peaks.

snATAC-seq trajectory analysis

snATAC-seq trajectory was conducted using Cicero, which extended Monocle3 to the snATAC-seq analysis. We obtained the preprocessed P0 snATAC-seq cell-peak matrix from snapATAC as input for Cicero and conducted dimension reduction using Latent Semantic Indexing (LSI) and visualized using UMAP. Trajectory graph was built using the function *learn_graph*. Batch effect was not observed between the two P0 batches, and the trajectory graph was consistent with cell type assignment with clustering analysis (**Figures S3 a-b**).

761 In order to study how open chromatin changes are associated with the cell fate decision, we first
762 binned the cells into 15 groups based on their pseudotime and cell type assignment. Next, we
763 studied the DAPs between each group and its ancestral group using the same methods described
764 above. The number of newly open and closed chromatins were reported using pie charts. The exact
765 peak locations are provided in the **Supplemental Table 7**.
766
767 *Genes and gene ontology terms associated with snATAC-seq trajectory*
768 Based on the binned trajectory graphs and DAPs between each group and its ancestral group, we
769 next used GREAT tool [66] to study the enrichment of associated genes and gene ontology (GO)
770 terms along the trajectory. We used the newly open or closed peaks as test regions and all the peaks
771 from peak-calling output as the background regions for the analysis. The output can be found in
772 the **Supplemental Table 8 and 9**.
773
774 *Predict cis-regulatory elements*
775 We implemented two methods to study cis-regulatory elements in the snATAC-seq data. The first
776 method was inspired by [23], which was based on the observation that there was co-enrichment in
777 the genome between the snATAC-seq cell type-specific peaks and scRNA-seq cell type-specific
778 genes. This method links a gene with a peak if 1) they were both specific in the same cell type, 2)
779 they were in *cis*, meaning that the peak is in ±100 kb region of the TSS of the corresponding gene,
780 and 3) the peak did not directly overlap with the TSS of the gene. This method successfully inferred
781 several known distal elements such as for *Six2* and *Slc6a18* (**Figure S2g**).
782
783 Alternatively, we assessed the co-accessibility of two peaks. We implemented Cicero [26], which
784 aggregates similar cells to obtain a set of "meta-cells" and address the issue with sparsity in the
785 snATAC-seq data. We used *run_cicero* function with default parameters to predict cis-regulatory
786 elements (CREs). Although it is recommended to use 0.25 as a cutoff for co-accessibility score,
787 we noticed that this resulted in a great amount of CREs, which could contain many false positives.
788 Thus, we used a more stringent score of 0.4 for the cutoff and retained 1,214,638 and 926,288
789 CRE links in the P0 and adult data, respectively.
790

791 **Bulk ATAC sequencing analysis**

792 Bulk ATAC-seq raw fastq files were processed using the end-to-end tool ENCODE ATAC-seq
793 pipeline (**Software and Algorithms**). This tool provided a standard workflow for ATAC-seq data
794 quality control, adaptor removal, alignment, and peak calling. To obtain high quality ATAC-seq
795 peaks, peak calling results from two biological replicates were compared and only those peaks that
796 were present in both replicates were kept, which were further used to compare with snATAC-seq
797 peaks.

798

799 **Correlation of bulk and single nuclei ATAC sequencing data**

800 snATAC-seq reads were aggregated to a pseudo-bulk data for the comparison purpose. To prevent
801 the effect of sex chromosome and mitochondria chromosome, reads from chromosome X, Y and
802 M were excluded from our analysis. We used multiBigwigSummary tool from deeptools [67] to study
803 the correlation between different samples. Specifically, the whole genome was binned into equally
804 sized (10 kb) windows, and the reads in each bin were aggregated, generating a bin-read count
805 vector for each of the sample. The correlation of these vectors was computed as a measure of
806 pairwise similarity between samples.

807

808 To compare the number of peaks in these two datasets, we used as input the narrowpeak files from
809 the snATAC-seq and bulk ATAC-seq analysis. We filtered out bulk ATAC-seq peaks with q value >
810 0.01 to be consistent with the snATAC-seq setting. Since the snATAC peaks were called after
811 merging different time points, we also took the union set of bulk ATAC-seq peaks from different
812 time points. We then used *findoverlap* function in GenomicRanges package [68] to find and report
813 overlapped peaks.

814

815 **Comparison between single nuclei ATAC sequencing data and single cell RNA sequencing**
816 **data**

817 In order to compare the cluster assignment between snATAC-seq data and scRNA-seq data, we
818 obtained the average gene expression values and peak accessibility in each cluster for P0 and adult
819 samples separately. We next transformed snATAC-seq data by summing up the reads within gene
820 body and 2 kb upstream regions to build gene activity score matrix, as suggested in Seurat [20]. Then,
821 we normalized the data and computed the mean expression and mean gene activity scores in each

822 cell type, and calculated z scores of each gene. Pearson's correlation coefficient was then
823 calculated among top 3,000 highly variable genes between snATAC-seq data and scRNA-seq data.
824 We found high concordance between these two datasets in terms of cell type assignment (**Figure**
825 **S4**).
826
827 **Single cell RNA sequencing data analysis**
828 *Alignment and quality control*
829 Raw fastq files were aligned to the mm10 (Ensembl GRCm38.93) reference genome and
830 quantified using CellRanger v3.1.0 (http://10xgenomics.com). Seurat R package v3.0 [69] was used
831 for data quality control, preprocessing and dimensional reduction analysis. After gene-cell data
832 matrix generation of both P0 and adult datasets, matrices were merged and poor-quality cells with
833 <200 or >3,000 expressed genes and mitochondrial gene percentages >50 were excluded, leaving
834 25,138 P0 and 18,498 adult cells for further analytical processing, respectively (**Figures S1j-k**).
835
836 *Pre-processing, batch effect correction and dimension reduction*
837 Data were normalized by RPM following log transformation and 3,000 highly variable genes were
838 selected for scaling and principal component analysis (PCA). Harmony R package v1.0 [17] was
839 used to correct batch effects. The top 20 dimensions of Harmony embeddings were used for
840 downstream uniform manifold approximation and projection (UMAP) visualization and clustering
841 (**Figures S1l-m**).
842
843 *Cell clustering, identification of marker genes and differentially expressed genes*
844 Louvain algorithm with resolution 0.4 was used to cluster cells, which resulted in 18 distinct cell
845 clusters. A gene was considered to be differentially expressed if it was detected in at least 25% of
846 one group and with at least 0.25 log fold change between two groups and the significant level of
847 BH-adjusted p value <0.05 in Wilcoxon rank sum test was used. We used a list of marker genes
848 [9,19] to manually annotate cell types. 2 distal convoluted tubule clusters were merged based on the
849 marker gene expression, resulting in a total of 17 clusters (**Figures S1i, n, o**).
850

851 *scRNA-seq trajectory analysis*

852 Monocle3

853 To construct single cell pseudotime trajectory and to identify genes whose expression changed as
854 the cells underwent transition, Monocle3 v0.1.3 [70] was applied to P0 cells of the following Seurat
855 cell clusters: nephron progenitors (NP), proliferating cells, stroma-like cells, podocytes, loop of
856 Henle (LOH), early proximal tubule (PT), proximal tubule S1, proximal tubule S3 cells.

857

858 To show cell trajectories from both small (nephron progenitors) and large cell populations
859 (proximal tubule), an equal number of 450 cells per cluster was randomly subsampled. Cells were
860 re-clustered by Monocle3 using a resolution of 0.0005 with k-nearest neighbor (kNN) k=29.
861 Highly variable genes along pseudotime were identified using differential *GeneTest* function and
862 cells were ordered along pseudotime trajectory. NP cluster was defined as earliest principal node.
863 In order to find genes differentially expressed along pseudotime, trajectories for podocytes, loop
864 of Henle, and proximal tubule clusters were analyzed separately with the *fit_models* function of
865 Monocle3. Genes with a q value <0.05 in the differential *GeneTest* analysis were kept. In an
866 alternate approach, *graph_test* function of Monocle3 was used and trajectory-variable genes were
867 collected into modules at a resolution of 0.01.

868

869 RNA velocity

870 To calculate RNA velocity, Python-based Velocyto command-line tool as well as Velocyto.R
871 package were used as instructed [30]. We used Velocyto to calculate the single-cell
872 trajectory/directionality using spliced and unspliced reads. From loom files produced by the
873 command-line tool, we subset the exact same cells that were previously selected randomly for
874 Monocle trajectory analysis. This subset was loaded into R using the SeuratWrappers v0.1.0
875 package. RNA velocity was estimated using gene-relative model with k-nearest neighbor (kNN)
876 cell pooling (k = 25). The parameter n was set at 200, when visualizing RNA velocity on the
877 UMAP embedding.

878

879 *Gene regulatory network inference*

880 In order to identify TFs and characterize cell states, we employed *cis*-regulatory analysis using the
881 R package SCENIC v1.1.2.2 [71], which infers the gene regulatory network based on co-expression

882 and DNA motif analysis. The network activity is then analyzed in each cell to identify recurrent
883 cellular states. In short, TFs were identified using GENIE3 and compiled into modules (regulons),
884 which were subsequently subjected to *cis*-regulatory motif analysis using RcisTarget with two
885 gene-motif rankings: 10 kb around the TSS and 500 bp upstream. Regulon activity in every cell
886 was then scored using AUCell. Finally, binarized regulon activity was projected onto Monocle3-
887 created UMAP trajectories.
888
889 *Ligand-receptor interactions*
890 To assess cellular crosstalk between different cell types, we used the CellPhoneDB repository to
891 infer cell-cell communication networks from single cell transcriptome data [40]. We used the Python
892 package CellPhoneDB v2.1.2 together with the database v2.0.0 to predict cell type-specific ligand-
893 receptor complexes as per the authors' instructions. Only receptors and ligands expressed in more
894 than 5% of the cells in the specific cluster were considered. 1,000 iterations were used for pairwise
895 comparison between cell types and considered for further statistical analysis.
896
897 **Immunofluorescence staining**
898 Mouse kidneys were fixed with 4% paraformaldehyde overnight, rinsed in PBS, and dehydrated
899 for paraffin embedding. Antigen retrieval was performed using Tris-EDTA buffer pH 9.0 with a
900 pressure cooker (PickCell Laboratories, Agoura Hills, CA) and antibody staining performed as
901 described [72]. Antibodies used were as follows: guinea pig FOXL1 (1:1,500) [73], mouse E-cadherin
902 (1:250; BD Transducton 610182, Franklin Lakes, NJ). Cy2-, Cy3-, and Cy5-conjugated donkey
903 secondary antibodies were purchased from Jackson ImmunoResearch Laboratories, Inc.
904 Fluoresecence images were collected on a Keyence microscope.
905
906 **Material Table**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Antibodies, Peptides, and Recombinant Proteins		
Guinea pig anti-Foxl1	own production	(ref. 73)
Mouse anti-E-cadherin	BD Transducton	Cat#610182

Cy2-conjugated donkey secondary antibody	Jackson ImmunoResearch Laboratories	Cat#715-225-150
Cy3-conjugated donkey secondary antibody	Jackson ImmunoResearch Laboratories	Cat#715-165-150
Cy5-conjugated donkey secondary antibody	Jackson ImmunoResearch Laboratories	Cat#715-175-150
DPBS	Corning	Cat# 21-031-CV
Tet System Approved FBS	Clontech	Cat# 631106
Nonidet™ P40 Substitute	Sigma	Cat# 74385
Magnesium Chloride Solution	Sigma	Cat# M1028
Ultrapure BSA (50 mg/ml)	Thermo Fisher	Cat# AM2616
RNAse inhibitor	Applied Biosystems	Cat# 100021540
Critical Commercial Assays		
Bioanalyzer High Sensitivity DNA kit	Agilent Technologies	5067-4626
Chromium Cell B Chip	10X Genomics	PN-120262
Chromium Chip E	10X Genomics	PN-1000082
Chromium Controller	10X Genomics	PN-120223
Chromium i7 Multiplex Kit	10X Genomics	PN-120262
Chromium Single Cell 3' Reagent Kits v3	10X Genomics	PN-1000092
Chromium Single Cell ATAC Library & Gel Bead Kit	10X Genomics	PN-1000110
Chromium Single Cell ATAC Library & Gel Bead Kit and Chromium i7 Multiplex Kit N	10X Genomics	PN-1000084
Countess AutoCounter	Invitrogen	C10227
Illumina Library Quantification Kit	KAPA Biosystems	KK4824
Multi Tissue dissociation kit	Miltenyi	130-110-201
Nuclei PURE Prep Nuclei Isolation Kit	Sigma	NUC-201
Deposited Data		
sci-CAR seq data	(ref. 24)	GSE117089
Mouse adult kidney WGBS data	(ref. 65)	GSM1051156
Mouse kidney H3K27ac and H3K4me1 CHIP-seq, WGBS, and RNA-seq	ENCODE project	ENCFF338WZP, ENCFF872MVE, ENCFF455HPY, ENCFF049LRQ, ENCFF179NTO, ENCFF071PID, ENCFF746MFH, ENCFF563LOO, ENCFF184AYF, ENCFF107NQP,

		ENCFF465THI, ENCFF769XWI, ENCFF591DAX
Six2 ChIP-seq data in nephron progenitor cells	(ref. 64)	GUDMAP database (RID:Q-Y4CY)
Software and Algorithms		
bedtools v. 2.29.2	open source	https://bedtools.readthedocs.io/en/latest/
Cell Ranger ATAC v. 1.1.0	10X Genomics	https://support.10xgenomics.com/single-cell-atac/software/downloads/latest
Cell Ranger v. 3.1.0	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest
CellPhoneDB v. 2.1.2	open source	https://www.cellphonedb.org
ChromVAR v. 3.1.0	open source	http://bioconductor.org/packages/release/bioc/html/chromVAR.html
Cicero v. 1.5.5	open source	https://github.com/cole-trapnell-lab/cicero-release
deeptools v. 2.0	open source	https://deeptools.readthedocs.io/en/develop/
ENCODE ATAC-seq pipeline	open source	https://github.com/kundajelab/atac_dnase_pipelines
HOMER v. 4.10.4	open source	http://homer.ucsd.edu/homer/motif/
IGV	open source	http://software.broadinstitute.org/software/igv/
MACS2 v. 2.2.6	open source	https://github.com/taoliu/MACS
Monocle3 v. 0.1.3	open source	http://cole-trapnell-lab.github.io/monocle-release/
Harmony	open source	https://github.com/immunogenomics/harmony
SCENIC v. 1.1.2.2	open source	https://aertslab.org/#scenic
Seurat R package v. 3.0	open source	https://satijalab.org/seurat/
snapATAC	open source	https://github.com/r3fang/SnapATAC
UCSC bedgraphtobigwig	open source	http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64.v385/
UCSC liftOver	open source	https://genome.ucsc.edu/cgi-bin/hgLiftOver
VelocytoR	open source	https://github.com/velocyto-team/velocyto.R
VisCello	open source	https://github.com/qinzhu/VisCello

Acknowledgements

Work in the Susztak lab is supported by the NIH DK076077, DK087635, and DK105821. MSB is supported by a German Research Foundation grant (BA 6205/2-1).

Author Contributions

KS and ZM designed and conceived the experiment. ZYM, JW, RS, and TA conducted the experiment. ZM conducted snATAC-seq bioinformatics analysis with advice from KS, HL, ML, and JK. MSB and ZM conducted scRNA-seq bioinformatics analysis with advice from KS. AMK and AYK conducted immunofluorescence staining with supervision from KHK. KS, ZM, and MSB wrote the manuscript and all authors edited and approved of the final manuscript.

Conflict of Interest

Authors declare no competing interests.

Lead Contact and Materials Availability

Raw data files and data matrix are being uploaded onto GEO and an accession number will be provided when it becomes available. The annotated and analyzed data can be viewed at **http://susztaklab.com/VisCello_snATAC/** and **http://susztaklab.com/igv/**. Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact: Katalin Susztak. Email: ksusztak@pennmedicine.upenn.edu.

930 **References**

931

932 1 Reidy, K., Kang, H. M., Hostetter, T. & Susztak, K. Molecular mechanisms of diabetic
933 kidney disease. *J Clin Invest* **124**, 2333-2340, doi:10.1172/JCI72271 (2014).

934 2 Costantini, F. & Kopan, R. Patterning a complex organ: branching morphogenesis and
935 nephron segmentation in kidney development. *Dev Cell* **18**, 698-712,
936 doi:10.1016/j.devcel.2010.04.008 (2010).

937 3 Park, J. S. *et al.* Six2 and Wnt regulate self-renewal and commitment of nephron
938 progenitors through shared gene regulatory networks. *Dev Cell* **23**, 637-651,
939 doi:10.1016/j.devcel.2012.07.008 (2012).

940 4 Harding, S. D. *et al.* The GUDMAP database--an online resource for genitourinary
941 research. *Development (Cambridge, England)* **138**, 2845-2853, doi:10.1242/dev.063594
942 (2011).

943 5 Wu, H. *et al.* Comparative Analysis and Refinement of Human PSC-Derived Kidney
944 Organoid Differentiation with Single-Cell Transcriptomics. *Cell Stem Cell* **23**, 869-881
945 e868, doi:10.1016/j.stem.2018.10.010 (2018).

946 6 Takasato, M. *et al.* Kidney organoids from human iPS cells contain multiple lineages and
947 model human nephrogenesis. *Nature* **526**, 564-568, doi:10.1038/nature15695 (2015).

948 7 Morizane, R. & Bonventre, J. V. Generation of nephron progenitor cells and kidney
949 organoids from human pluripotent stem cells. *Nat Protoc* **12**, 195-207,
950 doi:10.1038/nprot.2016.170 (2017).

951 8 Nishinakamura, R. Human kidney organoids: progress and remaining challenges. *Nature*
952 *Reviews Nephrology* **15**, 613-624, doi:10.1038/s41581-019-0176-x (2019).

953 9 Combes, A. N. *et al.* Correction: Single cell analysis of the developing mouse kidney
954 provides deeper insight into marker gene expression and ligand-receptor crosstalk
955 (doi:10.1242/dev.178673). *Development* **146**, doi:10.1242/dev.182162 (2019).

956 10 Adam, M., Potter, A. S. & Potter, S. S. Psychrophilic proteases dramatically reduce
957 single-cell RNA-seq artifacts: a molecular atlas of kidney development. *Development*
958 **144**, 3625-3632, doi:10.1242/dev.151142 (2017).

959 11 Schmidt-Ott, K. M. How to grow a kidney: patient-specific kidney organoids come of
960 age. *Nephrol Dial Transplant* **32**, 17-23, doi:10.1093/ndt/gfw256 (2017).

961 12 Menon, R. *et al.* Single-cell analysis of progenitor cell dynamics and lineage
962 specification in the human fetal kidney. *Development (Cambridge, England)* **145**,
963 doi:10.1242/dev.164038 (2018).

964 13 Lindstrom, N. O. *et al.* Progressive Recruitment of Mesenchymal Progenitors Reveals a
965 Time-Dependent Process of Cell Fate Acquisition in Mouse and Human Nephrogenesis.
966 *Dev Cell* **45**, 651-660 e654, doi:10.1016/j.devcel.2018.05.010 (2018).

967 14 Hochane, M. *et al.* Single-cell transcriptomics reveals gene expression dynamics of
968 human fetal kidney development. *PLOS Biology* **17**, e3000152,
969 doi:10.1371/journal.pbio.3000152 (2019).

970 15 Park, J., Liu, C. L., Kim, J. & Susztak, K. Understanding the kidney one cell at a time.
971 *Kidney Int* **96**, 862-870, doi:10.1016/j.kint.2019.03.035 (2019).

972 16 Fang, R. *et al.* Fast and Accurate Clustering of Single Cell Epigenomes Reveals
973 Cis-Regulatory Elements in Rare Cell Types. *bioRxiv*, 615179,
974 doi:10.1101/615179 (2019).

975 17 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with
976 Harmony. *Nat Methods* **16**, 1289-1296, doi:10.1038/s41592-019-0619-0 (2019).

977 18 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137,
978 doi:10.1186/gb-2008-9-9-r137 (2008).

979 19 Park, J. *et al.* Single-cell transcriptomics of the mouse kidney reveals potential cellular
980 targets of kidney disease. *Science* **360**, 758-763, doi:10.1126/science.aar2131 (2018).

981 20 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-
982 1902.e1821, doi:10.1016/j.cell.2019.05.031 (2019).

983 21 Wu, H., Kirita, Y., Donnelly, E. L. & Humphreys, B. D. Advantages of Single-Nucleus
984 over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell
985 States Revealed in Fibrosis. *J Am Soc Nephrol* **30**, 23-32, doi:10.1681/ASN.2018090912
986 (2019).

987 22 Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes.
988 *Nucleic Acids Res* **47**, D766-D773, doi:10.1093/nar/gky955 (2019).

989 23 Zhu, Q. *et al.*, doi:10.1101/848846 (2019).

990 24 Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands
991 of single cells. *Science* **361**, 1380-1385, doi:10.1126/science.aau0730 (2018).

992 25 O'Brien, L. L. *et al.* Transcriptional regulatory control of mammalian nephron
993 progenitors revealed by multi-factor cistromic analysis and genetic studies. *PLoS Genet*
994 **14**, e1007181, doi:10.1371/journal.pgen.1007181 (2018).

995 26 Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell
996 Chromatin Accessibility Data. *Mol Cell* **71**, 858-871 e858,
997 doi:10.1016/j.molcel.2018.06.044 (2018).

998 27 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime
999 cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-
1000 589, doi:10.1016/j.molcel.2010.05.004 (2010).

1001 28 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nature*
1002 *Methods* **14**, 1083-1086, doi:10.1038/nmeth.4463 (2017).

1003 29 Nittoli, V. *et al.* Characterization of paralogous uncx transcription factor encoding genes
1004 in zebrafish. *Gene X* **2**, 100011, doi:10.1016/j.gene.2019.100011 (2019).

1005 30 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498,
1006 doi:10.1038/s41586-018-0414-6 (2018).

1007 31 Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring
1008 transcription-factor-associated accessibility from single-cell epigenomic data. *Nat*
1009 *Methods* **14**, 975-978, doi:10.1038/nmeth.4401 (2017).

1010 32 Guo, J. K. *et al.* WT1 is a key regulator of podocyte function: reduced expression levels
1011 cause crescentic glomerulonephritis and mesangial sclerosis. *Hum Mol Genet* **11**, 651-
1012 659, doi:10.1093/hmg/11.6.651 (2002).

1013 33 Kann, M. *et al.* WT1 targets Gas1 to maintain nephron progenitor cells by modulating
1014 FGF signals. *Development* **142**, 1254-1266, doi:10.1242/dev.119735 (2015).

1015 34 Nakai, S. *et al.* Crucial roles of Brn1 in distal tubule formation and function in mouse
1016 kidney. *Development* **130**, 4751-4759, doi:10.1242/dev.00666 (2003).

1017 35 Nilsson, D., Heglind, M., Arani, Z. & Enerbäck, S. Foxc2 is essential for podocyte
1018 function. *Physiol Rep* **7**, e14083-e14083, doi:10.14814/phy2.14083 (2019).

1019 36 Takahashi, T. *et al.* Temporally compartmentalized expression of ephrin-B2 during renal
1020 glomerular development. *J Am Soc Nephrol* **12**, 2673-2682 (2001).

1021 37 Cheng, H. T. & Kopan, R. The role of Notch signaling in specification of podocyte and
1022 proximal tubules within the developing mouse kidney. *Kidney Int* **68**, 1951-1952,
1023 doi:10.1111/j.1523-1755.2005.00627.x (2005).

1024 38 Kang, H. M. *et al.* Defective fatty acid oxidation in renal tubular epithelial cells has a key
1025 role in kidney fibrosis development. *Nat Med* **21**, 37-46, doi:10.1038/nm.3762 (2015).

1026 39 Das, A. *et al.* Stromal–epithelial crosstalk regulates kidney progenitor cell differentiation.
1027 *Nature Cell Biology* **15**, 1035-1044, doi:10.1038/ncb2828 (2013).

1028 40 Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal-fetal interface in
1029 humans. *Nature* **563**, 347-353, doi:10.1038/s41586-018-0698-6 (2018).

1030 41 Trueb, B., Amann, R. & Gerber, S. D. Role of FGFRL1 and other FGF signaling proteins
1031 in early kidney development. *Cell Mol Life Sci* **70**, 2505-2518, doi:10.1007/s00018-012-
1032 1189-9 (2013).

1033 42 Costantini, F. & Shakya, R. GDNF/Ret signaling and the development of the kidney.
1034 *Bioessays* **28**, 117-127, doi:10.1002/bies.20357 (2006).

1035 43 Fan, X. *et al.* SLIT2/ROBO2 signaling pathway inhibits nonmuscle myosin IIA activity
1036 and destabilizes kidney podocyte adhesion. *JCI Insight* **1**, e86934,
1037 doi:10.1172/jci.insight.86934 (2016).

1038 44 Sajithlal, G., Zou, D., Silvius, D. & Xu, P. X. Eya 1 acts as a critical regulator for
1039 specifying the metanephric mesenchyme. *Dev Biol* **284**, 323-336,
1040 doi:10.1016/j.ydbio.2005.05.029 (2005).

1041 45 Hwang, D. Y. *et al.* Mutations of the SLIT2-ROBO2 pathway genes SLIT2 and SRGAP1
1042 confer risk for congenital anomalies of the kidney and urinary tract. *Hum Genet* **134**, 905-
1043 916, doi:10.1007/s00439-015-1570-5 (2015).

1044 46 Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses
1045 of a million individuals. *Nat Genet* **51**, 957-972, doi:10.1038/s41588-019-0407-x (2019).

1046 47 Hellwege, J. N. *et al.* Mapping eGFR loci to the renal transcriptome and phenome in the
1047 VA Million Veteran Program. *Nat Commun* **10**, 3842, doi:10.1038/s41467-019-11704-w
1048 (2019).

1049 48 Pattaro, C. *et al.* Genetic associations at 53 loci highlight cell types and biological
1050 pathways relevant for kidney function. *Nature communications* **7**, 10023,
1051 doi:10.1038/ncomms10023 (2016).

1052 49 Menon, M. C. *et al.* Intronic locus determines SHROOM3 expression and potentiates
1053 renal allograft fibrosis. *J Clin Invest* **125**, 208-221, doi:10.1172/JCI76902 (2015).

1054 50 Khalili, H. *et al.* Developmental Origins for Kidney Disease Due to Shroom3 Deficiency.
1055 *J Am Soc Nephrol* **27**, 2965-2973, doi:10.1681/ASN.2015060621 (2016).

1056 51 Qiu, C. *et al.* Renal compartment-specific genetic variation analyses identify new
1057 pathways in chronic kidney disease. *Nat Med* **24**, 1721-1731, doi:10.1038/s41591-018-
1058 0194-4 (2018).

1059 52 Kobayashi, A. *et al.* Six2 defines and regulates a multipotent self-renewing nephron
1060 progenitor population throughout mammalian kidney development. *Cell Stem Cell* **3**,
1061 169-181, doi:10.1016/j.stem.2008.05.020 (2008).

1062 53 Subramanian, A. *et al.* Single cell census of human kidney organoids shows
1063 reproducibility and diminished off-target cells after transplantation. *Nat Commun* **10**,
1064 5462, doi:10.1038/s41467-019-13382-0 (2019).

1065 54 Han, S. H. *et al.* PGC-1alpha Protects from Notch-Induced Kidney Fibrosis
1066 Development. *J Am Soc Nephrol* **28**, 3312-3322, doi:10.1681/ASN.2017020130 (2017).

1067 55 Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate
1068 causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491-504,
1069 doi:10.1038/s41576-018-0016-z (2018).

1070 56 Cusanovich, D. A. *et al.* The cis-regulatory dynamics of embryonic development at
1071 single-cell resolution. *Nature* **555**, 538-542, doi:10.1038/nature25981 (2018).

1072 57 Calderon, D. *et al.* Inferring Relevant Cell Types for Complex Traits by Using Single-
1073 Cell Gene Expression. *Am J Hum Genet* **101**, 686-699, doi:10.1016/j.ajhg.2017.09.009
1074 (2017).

1075 58 Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific
1076 molecular pathways and circuits in autism. *Cell* **155**, 1008-1021,
1077 doi:10.1016/j.cell.2013.10.031 (2013).

1078 59 Zhu, C. *et al.* An ultra high-throughput method for single-cell joint analysis of open
1079 chromatin and transcriptome. *Nature Structural & Molecular Biology* **26**, 1063-1070,
1080 doi:10.1038/s41594-019-0323-x (2019).

1081 60 Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and
1082 chromatin accessibility in the same cell. *Nature Biotechnology* **37**, 1452-1457,
1083 doi:10.1038/s41587-019-0290-0 (2019).

1084 61 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for
1085 Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 21-
1086 21 29 29, doi:10.1002/0471142727.mb2129s109 (2015).

1087 62 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables
1088 interrogation of frozen tissues. *Nat Methods* **14**, 959-962, doi:10.1038/nmeth.4396
1089 (2017).

1090 63 Chen, H. *et al.* Assessment of computational methods for the analysis of single-cell
1091 ATAC-seq data. *Genome Biol* **20**, 241, doi:10.1186/s13059-019-1854-5 (2019).

1092 64 O'Brien, L. L. *et al.* Differential regulation of mouse and human nephron progenitors by
1093 the Six family of transcriptional regulators. *Development* **143**, 595-608,
1094 doi:10.1242/dev.127175 (2016).

1095 65 Hon, G. C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA
1096 methylation maps from adult mouse tissues. *Nature Genetics* **45**, 1198-1206,
1097 doi:10.1038/ng.2746 (2013).

1098 66 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory
1099 regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).

1100 67 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data
1101 analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).

1102 68 Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS*
1103 *Computational Biology* **9**, e1003118, doi:10.1371/journal.pcbi.1003118 (2013).

1104 69 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902
1105 e1821, doi:10.1016/j.cell.2019.05.031 (2019).

1106 70 Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis.
1107 *Nature* **566**, 496-502, doi:10.1038/s41586-019-0969-x (2019).

1108 71 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat*
1109 *Methods* **14**, 1083-1086, doi:10.1038/nmeth.4463 (2017).

1110 72 Zhao, J., Hu, Z. Z., Zheng, X. G. & Ng, S. W. [2-(Tetra-zol-1-yl)acetato-kappaO]tris-(tri-
1111 phenyl-phosphine-kappaP)silver(I) mono-hydrate. *Acta Crystallogr Sect E Struct Rep*
1112 *Online* **65**, m1601, doi:10.1107/S1600536809048144 (2009).

1113 73 Aoki, R. *et al.* Foxl1-expressing mesenchymal cells constitute the intestinal stem cell
1114 niche. *Cell Mol Gastroenterol Hepatol* **2**, 175-188, doi:10.1016/j.jcmgh.2015.12.004
1115 (2016).

1116 74 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**,
1117 1260419, doi:10.1126/science.1260419 (2015).
1118

1119 **Figure Legends**

1120

1121 **Figure 1. snATAC-seq and scRNA-seq identified major cell types in developing and adult**
1122 **mouse kidney.**

1123 (A) Schematics of the study design. Kidneys from P0 and adult mice were processed for snATAC-
1124 seq and scRNA-seq followed by data processing and analysis including cell type identification and
1125 peak calling.

1126 (B) UMAP embeddings of snATAC-seq data and scRNA-seq data. Using marker genes, cells were
1127 annotated into nephron progenitors (NP), collecting duct intercalated cells (IC), collecting duct
1128 principal cells (PC), proximal tubule segment 1 and 3 (PT S1 and PT S3), loop of Henle (LOH),
1129 distal convoluted tubules (DCT), stromal cells (Stroma), podocytes (Podo), endothelial cells (Endo)
1130 and immune cells (Immune). In scRNA-seq data, the same cell types were identified, with an
1131 additional proliferative population and immune cells were clustered into neutrophils and
1132 macrophages.

1133 (C) UMAP embeddings of snATAC-seq and scRNA-seq data colored by P0 and adult batches.

1134 (D) Genome browser view of read density in each snATAC-seq cluster at cell type marker gene
1135 transcription start sites. We used Uncx for nephron progenitors, *Nphs2* for podocytes, *Akr1c21* for
1136 proximal tubules, *Slc12a1* for loop of Henle, *Slc12a3* for distal convoluted tubule, *Trpv5* for
1137 connecting tubule, *Aqp2* for collecting duct principal cells, *Atp6v1g3* for intercalated cells, *Egfl7*
1138 for endothelial cells, *C1qb* for immune cells and *Col3a1* for stroma. Additional marker gene
1139 examples are shown in **Figure S1h**.

1140 (E) Comparison of peaks identified from snATAC-seq data and bulk ATAC-seq data. Peaks that
1141 are identified in both datasets are colored blue, and peaks that are dataset-specific are grey.

1142 (F) Violin plots showing cell type-specific gene expression in scRNA-seq data. With the exception
1143 of proximal tubule, the same marker genes as in snATAC-seq data were used (*Slc5a2* and *Slc22a30*
1144 for proximal tubule S1 and S3, respectively).

1145 (G) Correlation between snATAC-seq gene activity scores and gene expression values in P0 data.
1146 The correlation of the adult dataset is shown in **Figure S1p**.

1147

1148 **Figure 2. Cell type-specific gene regulatory landscape of the mouse kidney.**
1149 (A) Left panel: Heatmap showing examples of the cell type-specific differentially accessible peaks
1150 (DAPs) (yellow: open chromatin, blue: closed chromatin) (full results are shown in **Supplemental**
1151 **Table 2**). Middle panel: Examples of cell type-specific motif enrichment analysis using Homer.
1152 (full results are shown in **Supplemental Table 3**). Right panel: TF expression z score heatmap
1153 that corresponds to the motif enrichment in each cell type.
1154 (B) Regulon activity heatmap. Each column represents a single cell, colored by cluster assignment
1155 and ordered by hierarchical clustering; each row represents binarized regulon activities ("on-
1156 black", "off-white") and ordered by hierarchical clustering.
1157 (C) UMAP depiction of regulon activity ("on-blue", "off-grey") and TF gene expression (red scale)
1158 of exemplary regulons for proximal tubule (*Hnf1a*), nephron progenitors (*Uncx*), loop of Henle
1159 (*Ppargc1a*), proliferating cells (*Hmgb3*) and podocytes (*Mafb*). Examples of target gene
1160 expression of the *Uncx* regulon (*Eye1*, *Hoxc8*, *Pax2*, *Spock2* and *Wnt4*) are shown in purple scale.
1161 Expression of target genes of *Hnf1a*, *Ppargc1a*, *Hmgb3* and *Mafb* is shown in **Figure S3d**.
1162 (D) tSNE representation of regulon density as a surrogate for stability of regulon states, as inferred
1163 by SCENIC algorithm.
1164
1165 **Figure 3. The cellular trajectory of nephron progenitor differentiation.**
1166 (A) UMAP representation of snATAC-seq nephron progenitor differentiation trajectory towards
1167 podocytes, proximal tubule, loop of Henle and distal convoluted tubule, respectively, as inferred
1168 by Cicero. Cells are colored by pseudotime.
1169 (B) UMAP representation of scRNA-seq nephron progenitor differentiation trajectory towards
1170 podocytes, proximal tubule and loop of Henle, respectively, as inferred by Monocle3. Cells are
1171 colored by pseudotime.
1172 (C) Pseudotime-dependent chromatin accessibility and gene expression changes along the
1173 proximal tubule (red), podocyte (green) and loop of Henle (blue) cell lineages. The first column
1174 shows the dynamics of chromVAR TF enrichment score, the second column shows the dynamics
1175 of TF gene expression values and the third and fourth column represent the dynamics of SCENIC-
1176 reported target gene expression values of corresponding TFs, respectively. Additional examples
1177 are given in **Figure S3e**.
1178

1179 **Figure 4. Chromatin dynamics of nephron progenitor differentiation.**
1180 Di-graph representing cell type and lineage divergence, as derived from Cicero trajectory inference.
1181 Nephron progenitors (NP), podocytes (Podo), intermediate stage (IM), proximal tubule (PT), loop
1182 of Henle (LOH) and distal convoluted tubule (DCT) are connected with their developmental
1183 precursor stages and represented by ascending numbering. Arrows represent cell differentiation
1184 along respective trajectories. Genes listed next to the trajectories were derived from analyzing gene
1185 enrichment of differentially assessible peaks (DAPs) between two stages. Genes colored red were
1186 derived from the opening DAPs between two stages, genes colored blue were derived from the
1187 closing DAPs between two stages, and genes colored green were derived from opening DAPs
1188 between two branches. Three important genes, *Foxl1*, *Hnf4a* and *Tfap2b* are shown along with
1189 their cell type-specific accessibility peaks and immunostaining results. Peaks that were open
1190 during the development of specific cell types are shown in red boxes. Immunofluorescence
1191 staining of fetal mouse kidney shows FOXL1 in red along cellular differentiation (from right to
1192 left) from early progenitor stage (asterisk) over comma-shaped (+) and S shaped bodies (cross)
1193 towards podocytes within primitive glomeruli (#). HNF4A and TFAP2B in human adult kidney
1194 samples (taken from the Human Protein Atlas, http://www.proteinatlas.org [74]) are visualized by
1195 immunohistochemistry in brown.
1196
1197 **Figure 5. Cell-cell communication analysis in the developing and adult mice highlighted the**
1198 **critical role of stroma in driving cell differentiation.**
1199 (A) Heatmaps showing the number of cell-cell interactions in the scRNA-seq dataset of P0 (top)
1200 and adult (bottom) kidneys, as inferred by CellPhoneDB. Dark blue and dark red colors denote
1201 low and high numbers of cell-cell interactions, respectively.
1202 (B) CellPhoneDB-derived measures of cell-cell interaction scores and p values. Each row shows
1203 a ligand-receptor pair, and each column shows the 2 interacting cell types, which is binned by cell
1204 type. Columns are sub-ordered by first interacting cell type into stroma, podocytes, endothelial
1205 cells, proximal tubule, loop of Henle and nephron progenitors. Color scale denotes the mean values
1206 for all the interacting partners, where mean value refers to the total mean of the individual partner
1207 average expression values in the interacting cell type pairs. Orange scale denotes P0, blue scale
1208 denotes adult. Dot size denotes corresponding p values of the permutation test.

1209 (C) Dot plots of RNA expression of important cell-cell communication candidates within the Gdnf-
1210 Ret, Sonic hedgehog, Fgf, Bmp, Wnt and other pathways in both P0 (top) and adult (bottom)
1211 kidney. Dot size denotes percentage of cells expressing the marker. Color scale represents average
1212 gene expression values, orange denotes P0, blue denotes adult. Arrows indicate ligand-receptor
1213 pairs.
1214
1215 **Figure 6. Single cell level chromatin accessibility highlighted human kidney GWAS target**
1216 **genes and cell types.**
1217 (A, C, E) From top to bottom: Cicero-inferred co-accessibility of open chromatin regions in mouse
1218 orthologues of human *Shroom3*, *Dab2* and *Uncx* loci; Gene browser view of the single nucleotide
1219 polymorphisms within the regions; gene browser view of chromatin accessibility for nephron
1220 progenitors (NP), collecting duct intercalated cells (IC), collecting duct principal cell types (PC),
1221 proximal tubules segment 1 and 3 (PT S1 and PT S3), loop of Henle (LOH), distal convoluted
1222 tubule (DCT), stromal cells (stroma), podocytes (Podo), endothelial cells (Endo) and immune cells
1223 (Immune). Right subpanel shows violin plots of scRNA-seq gene expression in P0 (orange) and
1224 adult (blue) kidneys.
1225 (B, D, F) Whole kidney H3K27ac, H3K4me1 and Six2 ChIP-seq, whole genome bisulfate
1226 sequencing (WGBS) and RNA-seq data in E15.5, P0 and adult kidney samples.

1227 **Supplemental Information**

1228

1229 **Supplemental Tables**

1230 **Supplemental Table 1.** Cell type marker genes derived from scRNA-seq analysis.

1231 **Supplemental Table 2.** Cell type-specific open chromatin derived from snATAC-seq analysis.

1232 **Supplemental Table 3.** Cell type-specific motif enrichment.

1233 **Supplemental Table 4.** Regulons and respective target genes inferred by SCENIC.

1234 **Supplemental Table 5.** Binarized regulon activities in each cell type inferred by SCENIC.

1235 **Supplemental Table 6.** Differentially expressed genes along pseudotime in distinct lineages in
1236 scRNA-seq data.

1237 **Supplemental Table 7.** Differentially accessible peaks along pseudotime in distinct lineages in
1238 snATAC-seq data.

1239 **Supplemental Table 8**. Nearest genes of differentially accessible peaks along pseudotime in
1240 distinct lineages in snATAC-seq data.

1241 **Supplemental Table 9**. GO enrichment of differentially accessible peaks along pseudotime in
1242 distinct lineages inferred by GREAT analysis.

1243 **Supplemental Table 10.** ChromVAR cell-TF enrichment score matrix.

1244 **Supplemental Table 11.** Nearest genes of differentially accessible peaks at bifurcation events
1245 along pseudotime in distinct lineages inferred by GREAT analysis.

1246

1247

1248 **Supplemental Figures**

1249 **Figure S1. Quality control and data processing methods for snATAC-seq and scRNA-seq**
1250 **data analysis.**

1251 (A) Insert size distribution of the 5 snATAC-seq samples showing periodic patterns.

1252 (B) Transcription start sites (TSS) signal enrichment of the 5 snATAC-seq samples.

1253 (C) Spearman correlation between snATAC-seq datasets and bulk ATAC-seq of binned genomic
1254 regions.

1255 (D) Distribution of number of unique molecular identifiers (UMIs, x axis) and promoter ratio (y
1256 axis) in 5 samples shown by dot plot.

1257 (E) Violin plots representing the number of accessible peaks across different clusters in the
1258 snATAC-seq dataset indicating similar distributions.
1259 (F) UMAP representation of the snATAC-seq dataset colored by batches.
1260 (G) Stacked bar graphs representing absolute numbers and percentages of identified cell types
1261 across snATAC-seq batches.
1262 (H) Genome browser view of cell type-specific peaks at the TSS of marker genes for 13 cell types
1263 in the snATAC-seq dataset.
1264 (I) From left to right: Stacked bar graphs showing the percentage of different cell types in the P0
1265 and adult scRNA-seq datasets, tables showing the number of cells in each cell type (nCells) and
1266 corresponding percentage. NP, nephron progenitor; Podo, podocyte; PT, proximal tubule; S1,
1267 segment 1; S3, segment 3; LOH, loop of Henle; DCT, distal convoluted tubule cells; PC, collecting
1268 duct principal cells; IC, collecting duct intercalated cells; Endo, endothelial cells; Macro,
1269 macrophages; Neutro, neutrophils.
1270 (J) UMAP representation of scRNA-seq data colored by the mitochondrial gene ratio (Mt %).
1271 (K) Violin plots showing number of informative genes per single cell and unique molecular
1272 identifiers (UMIs) per single cell. Blue denotes adult kidney, orange denotes P0 kidney.
1273 (L, M) Principal component (PC) representation of combined adult and P0 scRNA-seq dataset (left
1274 panel) and violin plots of corresponding embeddings values (right panel) before (L) and after (M)
1275 batch correction using Harmony.
1276 (N) Dot plot of cell type-specific marker genes. Dot size denotes percentage of cells expressing
1277 the marker. Color scale represents average expression, orange denotes P0, blue denotes adult
1278 kidney.
1279 (O) Feature plots of representative marker genes projected on UMAP dimension.
1280 (P) Correlation between snATAC-seq gene activity scores and gene expression values in adult data,
1281 which is complementary to **Figure 1g**.
1282 (Q) We provide the processed chromatin accessibility dataset via a searchable, interactive website
1283 (http://susztaklab.com/igv/). Ace2 was used as an example, and we show proximal tubule-specific
1284 enrichment of peaks at transcription start sites of the *Ace2* (Angiotensin-converting enzyme 2)
1285 gene (red boxes).
1286

1287 **Figure S2. Characterization of the cell type-specific regulatory landscape**

1288 (A) Bar graph representing the number of accessible peaks in distal elements, promoters, introns,
1289 5'-UTR, 3'UTR and exons, as distributed across samples of snATAC-seq data and bulk ATAC-
1290 seq data.

1291 (B) Overlap of scATAC-seq differentially accessible peaks among cell types with H3K27Ac ChIP-
1292 seq data.

1293 (C) Number of shared and unique peaks among snATAC-seq cell types. Cell types include nephron
1294 progenitors and cells differentiated from nephron progenitors.

1295 (D) Genome browser view of Umod as an example for distal open chromatin region and its target
1296 promoter region.

1297 (E) Distribution of different open chromatin elements in snATAC-seq cell types.

1298 (F) Distribution of different open chromatin elements among differentially accessible peaks (DAPs)
1299 in snATAC-seq cell types.

1300 (G) Genome browser representations of single cell open chromatin data for individual cell types
1301 at chromosomal loci around Six2 and Slc6a18, along with their known distal elements (red boxes).
1302 Corresponding chromosomal interaction of open chromatin regions, as inferred by Cicero
1303 (**Methods**), is depicted at the top.

1304 (H) Genome browser views of representative marker genes demonstrating cell type-specific
1305 chromatin accessibility for proximal tubule (*Hnf4a* and *Hmgb3*), several tubular segments (*Hnf1b*),
1306 loop of Henle and distal convoluted tubule (*Esrrb* and *Ppargc1a*) as well as nephron progenitors
1307 and podocytes (Wt1).

1308 (I) UMAP depiction of regulon activity ("on-blue", "off-grey") and RNA expression (red scale) of
1309 exemplary regulons of proximal tubule (*Hnf1a*), nephron progenitors (*Six2*), loop of Henle
1310 (*Ppargc1a*), proliferating cells (*Hmgb3*) and podocytes (*Mafb*), respectively. Exemplary target
1311 gene expression for the respective TF is shown in purple scale.

1312

1313 **Figure S3. snATAC-seq and scRNA-seq cell differentiation trajectories.**
1314 (A) UMAP representation of snATAC-seq trajectory lineages of podocytes, proximal tubule and
1315 loop of Henle cells from nephron progenitors colored by 2 P0 batches.
1316 (B) UMAP representation of snATAC-seq trajectory lineages of podocytes, proximal tubule and
1317 loop of Henle cells from nephron progenitors colored by original cell type assignment as in **Figure**
1318 **1b**.
1319 (C) UMAP representation of scRNA-seq trajectory lineages of podocytes, proximal tubule and
1320 loop of Henle cells from nephron progenitors colored by original cell type assignment as **in Figure**
1321 **1b**.
1322 (D) UMAP representation of RNA velocity of scRNA-seq trajectory inferred by VelocytoR,
1323 colored by original cell type assignment. Each dot is one cell and each arrow represents the time
1324 derivative of the gene expression state.
1325 (E) UMAP representation of snATAC-scRNA integration results colored by cell type assignment.
1326 (F) UMAP representation of snATAC-scRNA integration results colored by technologies
1327 (snATAC=red, scRNA=grey). Podo: podocytes, PT: proximal tubule, LOH: loop of Henle, DCT:
1328 distal convoluted tubule, NP: nephron progenitors, IM: intermediate stage cells.
1329 (G) Dot plot showing snATAC-scRNA integration cell type assignment confusion matrix. Each
1330 column represents the original cell type assignment of snATAC-seq data, and each row represents
1331 the predicted cell type assignment by the integration analysis scRNA-seq data. Each dot represents
1332 the number of cells that were matched in the integrated data.
1333 (H) Heatmap of chromVAR enrichment results. The original data matrix is given in **Supplemental**
1334 **Table 10**.
1335 (I) Pseudotime-dependent chromatin accessibility and gene expression changes along the proximal
1336 tubule (red), podocytes (green) and loop of Henle (blue) cell lineages. The first column represents
1337 the dynamics of chromVAR TF enrichment score, the second column represents the dynamics of
1338 TF gene expression values, and the third and fourth column represent the dynamics of SCENIC-
1339 reported target gene expression values.
1340

1341 **Figure S4. Chromatin dynamics of nephron progenitor differentiation.**
1342 (A) Di-graph representing cell type and lineage divergence, as derived from Cicero trajectory
1343 inference. Nephron progenitors (NP), podocytes (Podo), intermediate stage (IM), proximal tubule
1344 (PT), loop of Henle (LOH) and distal convoluted tubule (DCT) are connected with their
1345 developmental precursor stages and ordered by ascending numbering. Pie charts represent
1346 differentially assessible peaks (DAPs) between two stages, where the size of pie charts is
1347 proportional to the number of DAPs, orange color represents the number of open peaks, grey color
1348 the number of closed peaks. Bar graphs depict gene ontology (GO) term analysis of genes nearby
1349 DAPs derived from GREAT analysis (full list in **Supplemental Table 9**).
1350 (B) Immunofluorescence staining of fetal mouse kidney. Upper panel and insert denote E13.5 stage,
1351 lower panel denotes P6 mouse. Blue staining represents nuclei (DAPI), green staining represents
1352 tubular epithelium (E-Cadherin) and red staining represents progenitor cells (FOXL1) along a
1353 developmental trajectory from early progenitor stage (asterisk) over comma-shaped (+) and S
1354 shaped bodies (cross) towards podocytes within primitive glomeruli (#).
1355 (C) Pseudotime-dependent chromatin accessibility and gene expression changes along the
1356 proximal tubule (red), podocytes (green) and loop of Henle (LOH, blue) cell lineages for important
1357 bifurcation TFs in the podocyte (*Foxl1*) and distal tubule (*Tfap2b*) lineage.
1358 (D) Bar graphs denote the percentage of cells with accessible chromatin of several *Six2* promoters
1359 and enhancers (gene loci numbered 1-3) as well as putative *Foxl1* enhancers (gene loci numbered
1360 4-7) along pseudotime. Exact gene loci of enhancers and promoters are given above each
1361 respective graph. Changes along pseudotime are depicted for 3 lineages from nephron progenitors
1362 (NP) to podocytes, proximal tubule (PT) and loop of Henle (LOH) cells, respectively. The right
1363 upper subpanel depicts the genome browser overview of chromatin accessibility for the NP and
1364 therefore corresponds to the first bar in graphs on the left. The right lower subpanel depicts zoom-
1365 in versions of the 7 loci for all 3 lineages.
1366
1367 **Figure S5. Single cell level chromatin accessibility highlighted human kidney GWAS target**
1368 **genes and cell types.**
1369 Open chromatin and co-accessibility view at alternative scales to those shown in **Figure 6**.
1370 (A, C, E) From top to bottom: Cicero-inferred co-accessibility of open chromatin regions in mouse
1371 orthologues of human *Shroom3*, *Dab2* and *Uncx* loci; Gene browser view of the single nucleotide

1372 polymorphisms within the regions; gene browser view of chromatin accessibility for nephron
1373 progenitors (NP), collecting duct intercalated cells (IC), collecting duct principal cell types (PC),
1374 proximal tubules segment 1 and 3 (PT S1 and PT S3), loop of Henle (LOH), distal convoluted
1375 tubule (DCT), stromal cells (stroma), podocytes (Podo), endothelial cells (Endo) and immune cells
1376 (Immune). Right subpanel shows violin plots of scRNA-seq gene expression in P0 (orange) and
1377 adult (blue) kidneys.
1378 (B, D, F) Whole kidney H3K27ac, H3K4me1 and Six2 ChIP-seq, whole genome bisulfate
1379 sequencing (WGBS) and RNA-seq data in E15.5, P0 and adult kidney samples.

Figure 1

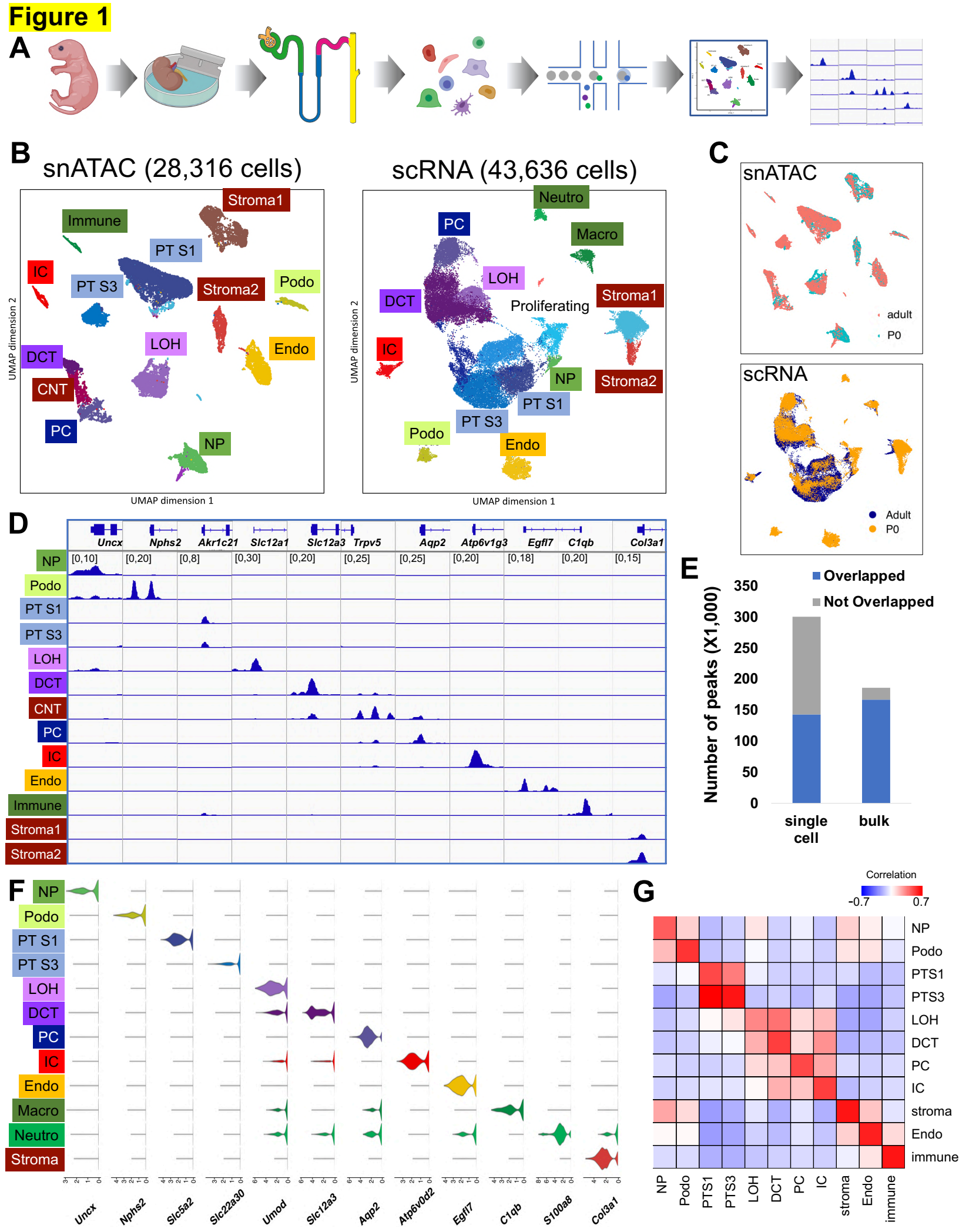
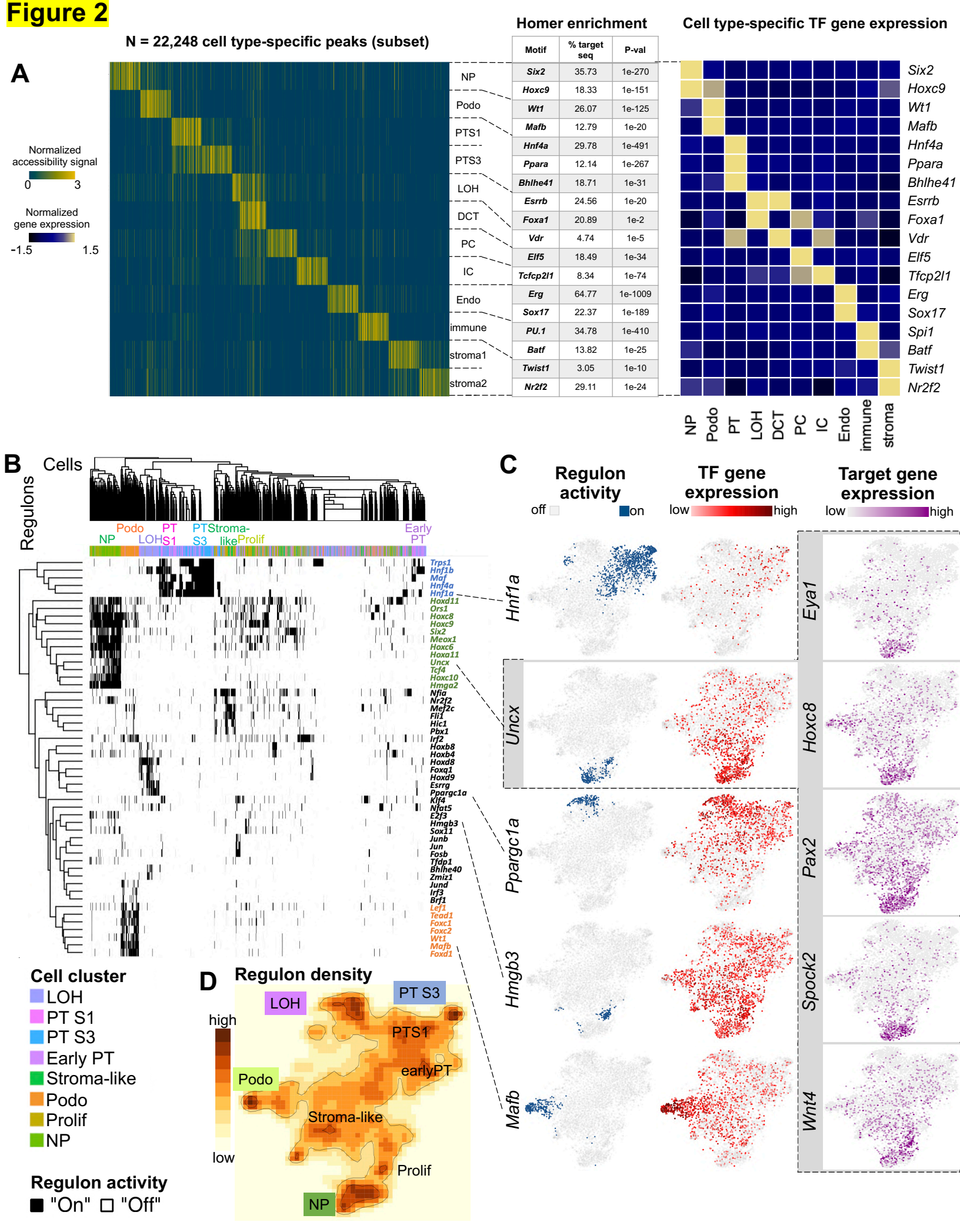

A
B
snATAC (28,316 cells)
scRNA (43,636 cells)
Stroma1
Immune
PT S1
IC
PT S3
Stroma2
Podo
LOH
Endo
DCT
CNT
PC
NP
UMAP dimension 1
UMAP dimension 2
Neutro
Macro
Proliferating
C
snATAC
adult
P0
scRNA
Adult
P0
D
Uncx
Nphs2
Akr1c21
Slc12a1
Slc12a3
Trpv5
Aqp2
Atp6v1g3
Egfl7
C1qb
Col3a1
[0,10]
[0,20]
[0,8]
[0,30]
[0,20]
[0,25]
[0,25]
[0,20]
[0,18]
[0,20]
[0,15]
E
Overlapped
Not Overlapped
Number of peaks (X1,000)
350
300
250
200
150
100
50
0
single cell
bulk
F
Uncx
Nphs2
Slc5a2
Slc22a30
Umod
Slc12a3
Aqp2
Atp6v0d2
Egfl7
C1qb
S100a8
Col3a1
G
Correlation
-0.7
0.7
NP
Podo
PTS1
PTS3
LOH
DCT
PC
IC
stroma
Endo
immune

Figure 2



Motif	% target seq	P-val
Six2	35.73	1e-270
Hoxc9	18.33	1e-151
Wt1	26.07	1e-125
Mafb	12.79	1e-20
Hnf4a	29.78	1e-491
Ppara	12.14	1e-267
Bhlhe41	18.71	1e-31
Esrrb	24.56	1e-20
Foxa1	20.89	1e-2
Vdr	4.74	1e-5
Elf5	18.49	1e-34
Tcfcp2l1	8.34	1e-74
Erg	64.77	1e-1009
Sox17	22.37	1e-189
PU.1	34.78	1e-410
Batf	13.82	1e-25
Twist1	3.05	1e-10
Nr2f2	29.11	1e-24

Figure 3

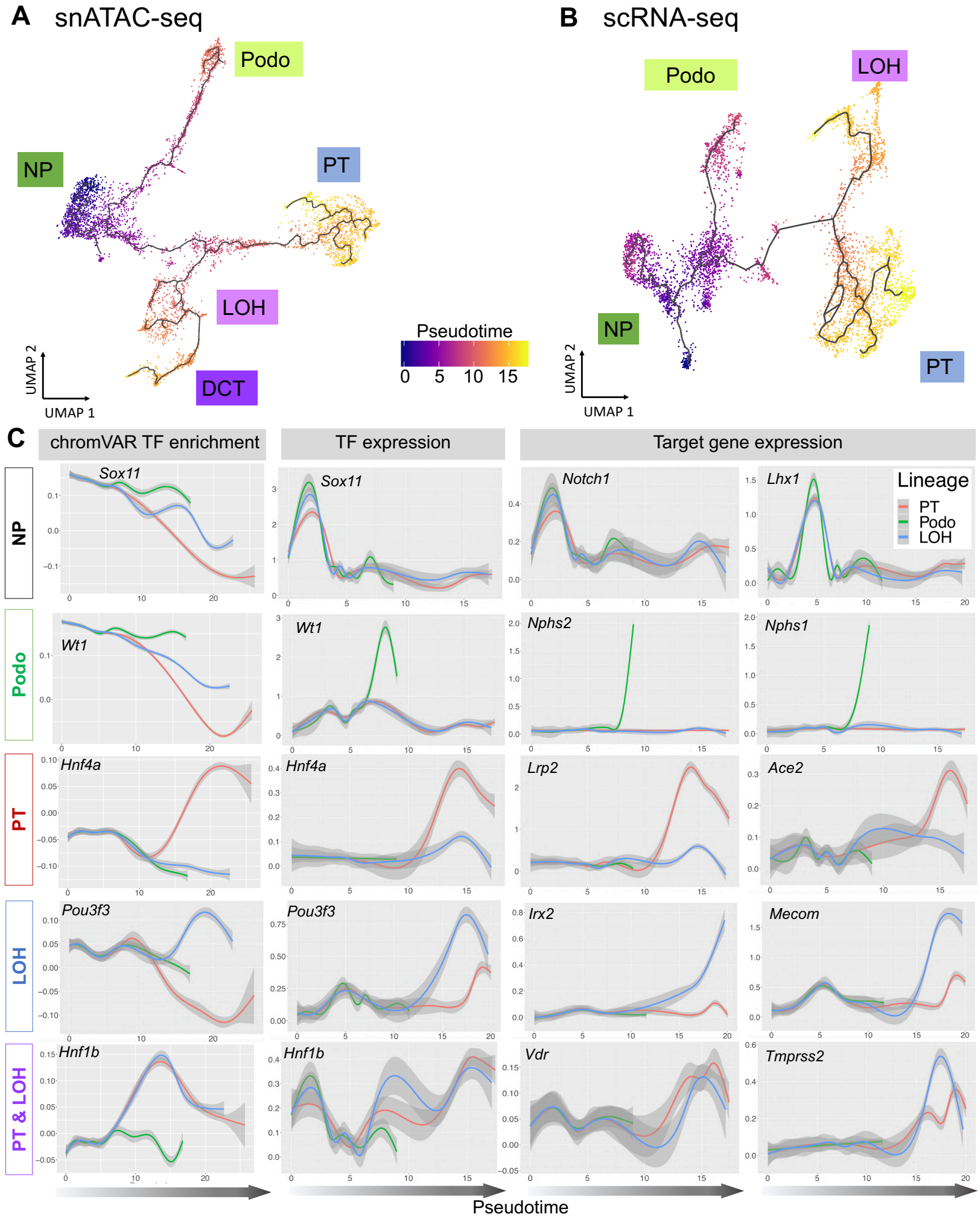

A snATAC-seq
Podo
NP
PT
LOH
DCT
UMAP 2
UMAP 1
Pseudotime
0 5 10 15
B scRNA-seq
Podo
LOH
NP
PT
UMAP 2
UMAP 1
C
chromVAR TF enrichment
TF expression
Target gene expression
NP
Podo
PT
LOH
PT & LOH
Sox11
Sox11
Notch1
Lhx1
Lineage
PT
Podo
LOH
Wt1
Wt1
Nphs2
Nphs1
Hnf4a
Hnf4a
Lrp2
Ace2
Pou3f3
Pou3f3
Irx2
Mecom
Hnf1b
Hnf1b
Vdr
Tmprss2
Pseudotime

Figure 4

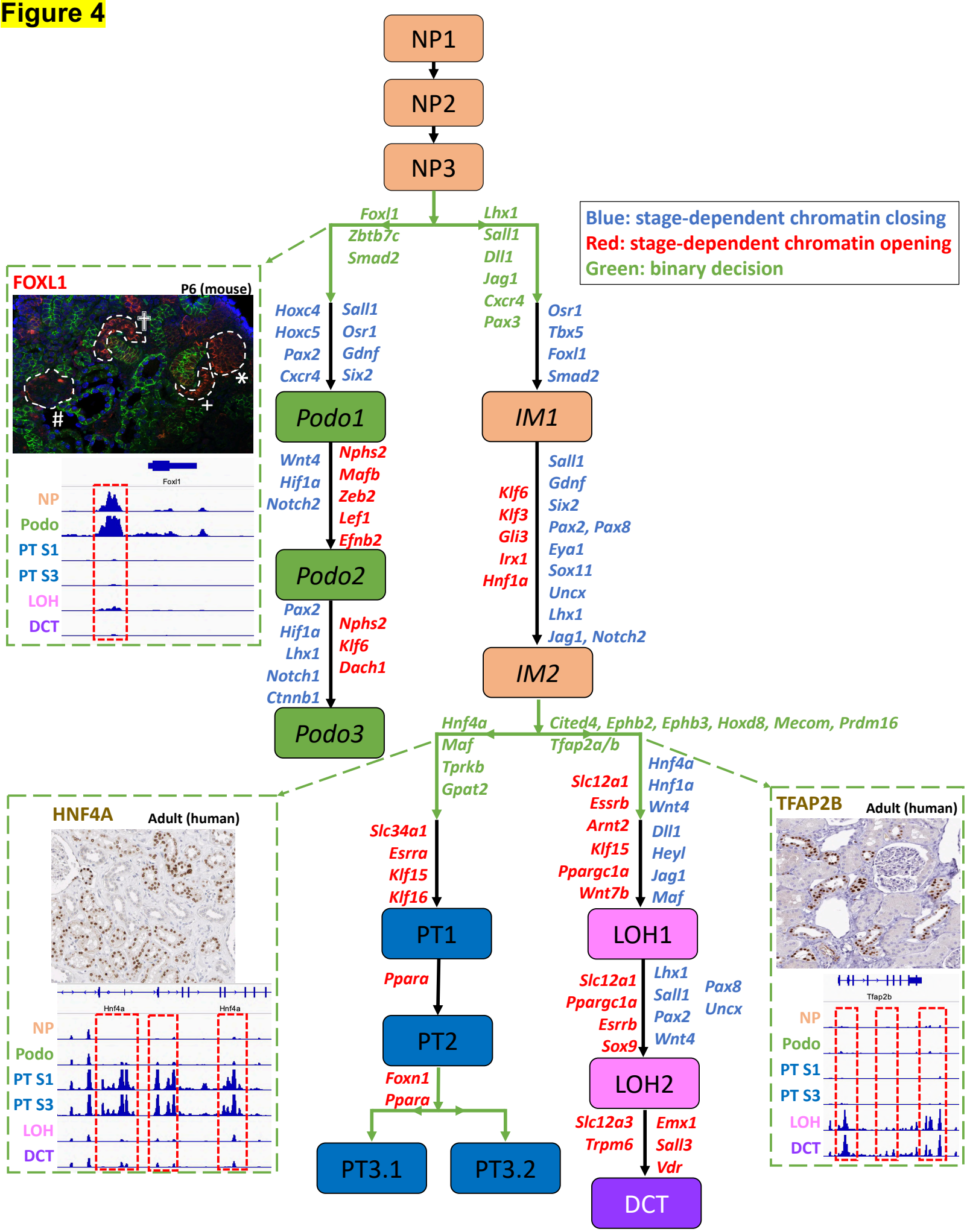

NP1
NP2
NP3
Foxl1
Zbtb7c
Smad2
Lhx1
Sall1
Dll1
Jag1
Cxcr4
Pax3
Blue: stage-dependent chromatin closing
Red: stage-dependent chromatin opening
Green: binary decision
FOXL1
P6 (mouse)
Foxl1
NP
Podo
PT S1
PT S3
LOH
DCT
Hoxc4
Hoxc5
Pax2
Cxcr4
Sall1
Osr1
Gdnf
Six2
Osr1
Tbx5
Foxl1
Smad2
Podo1
IM1
Wnt4
Hif1a
Notch2
Nphs2
Mafb
Zeb2
Lef1
Efnb2
Klf6
Klf3
Gli3
Irx1
Hnf1a
Sall1
Gdnf
Six2
Pax2, Pax8
Eya1
Sox11
Uncx
Lhx1
Jag1, Notch2
Podo2
IM2
Pax2
Hif1a
Lhx1
Notch1
Ctnnb1
Nphs2
Klf6
Dach1
Podo3
Hnf4a
Maf
Tprkb
Gpat2
Cited4, Ephb2, Ephb3, Hoxd8, Mecom, Prdm16
Tfap2a/b
Slc12a1
Essrb
Arnt2
Klf15
Ppargc1a
Wnt7b
Hnf4a
Hnf1a
Wnt4
Dll1
Heyl
Jag1
Maf
HNF4A
Adult (human)
Hnf4a
Hnf4a
NP
Podo
PT S1
PT S3
LOH
DCT
Slc34a1
Esrra
Klf15
Klf16
PT1
Ppara
PT2
Foxn1
Ppara
PT3.1
PT3.2
LOH1
Slc12a1
Ppargc1a
Esrrb
Sox9
Lhx1
Sall1
Pax2
Wnt4
Pax8
Uncx
LOH2
Slc12a3
Trpm6
Emx1
Sall3
Vdr
DCT
TFAP2B
Adult (human)
Tfap2b
NP
Podo
PT S1
PT S3
LOH
DCT

Figure 5

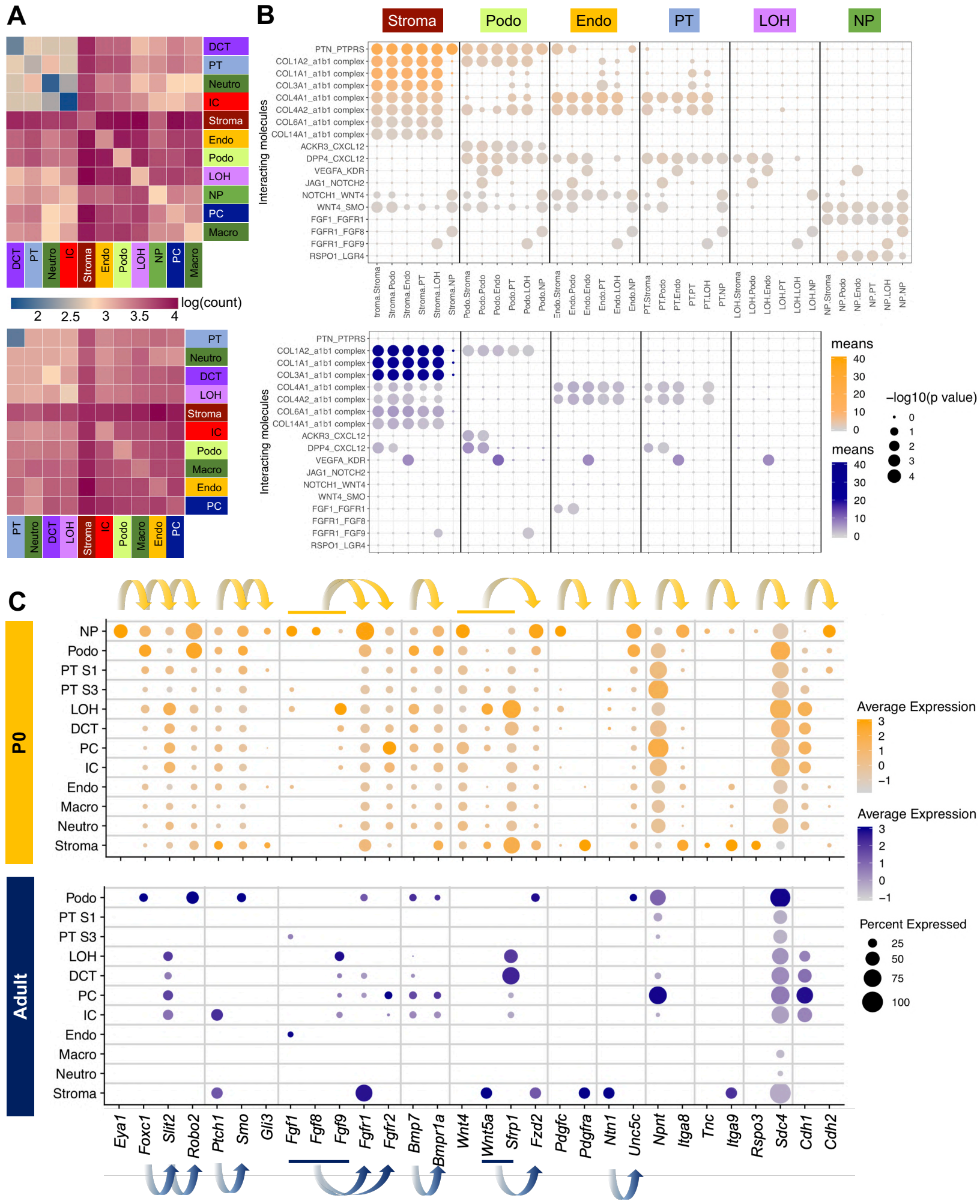
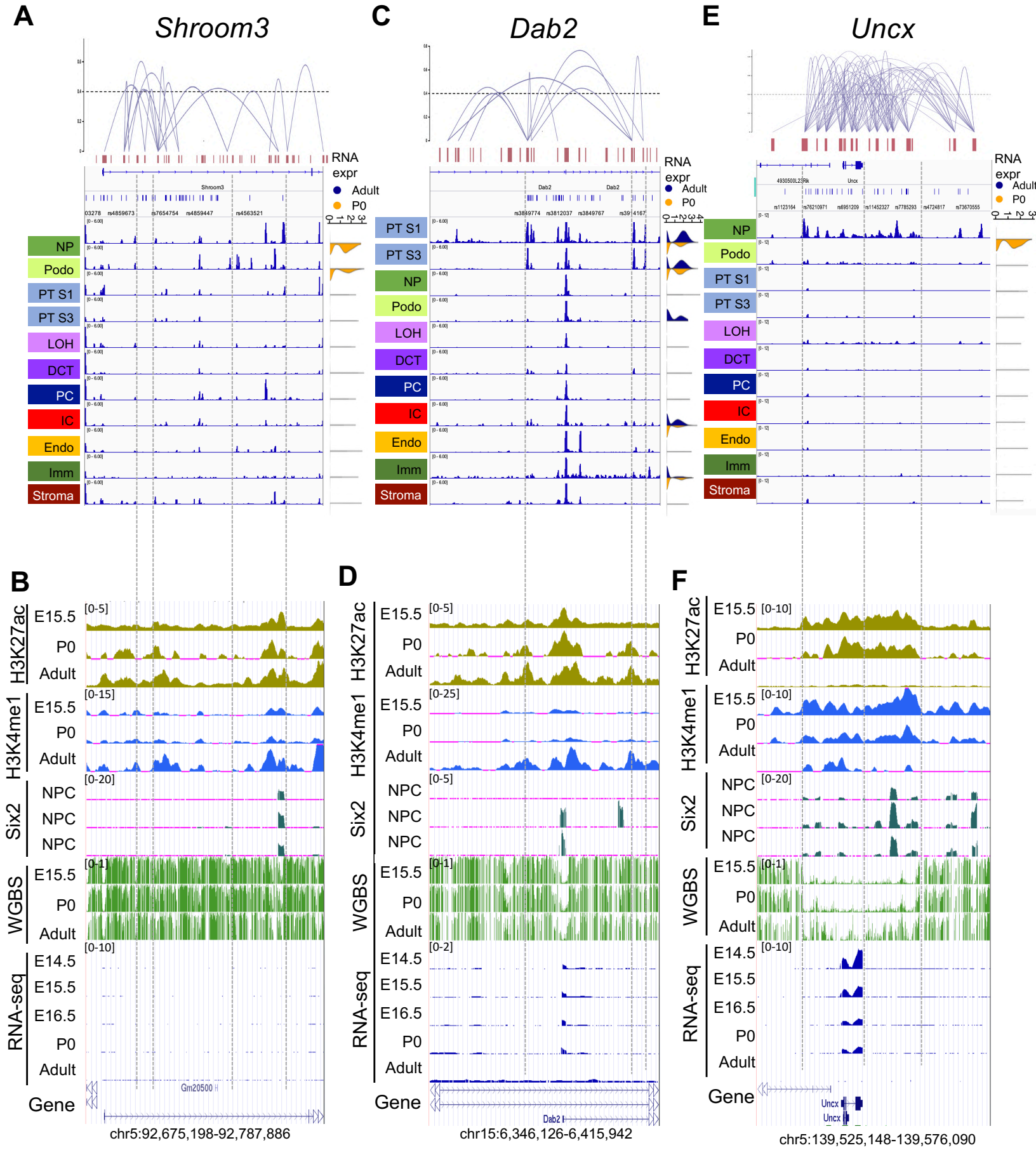

A
DCT
PT
Neutro
IC
Stroma
Endo
Podo
LOH
NP
PC
Macro
log(count)
2
2.5
3
3.5
4
B
Stroma
Podo
Endo
PT
LOH
NP
Interacting molecules
PTN_PTPRS
COL1A2_a1b1 complex
COL1A1_a1b1 complex
COL3A1_a1b1 complex
COL4A1_a1b1 complex
COL4A2_a1b1 complex
COL6A1_a1b1 complex
COL14A1_a1b1 complex
ACKR3_CXCL12
DPP4_CXCL12
VEGFA_KDR
JAG1_NOTCH2
NOTCH1_WNT4
WNT4_SMO
FGF1_FGFR1
FGFR1_FGF8
FGFR1_FGF9
RSPO1_LGR4
means
40
30
20
10
0
−log10(p value)
0
1
2
3
4
C
P0
Adult
NP
Podo
PT S1
PT S3
LOH
DCT
PC
IC
Endo
Macro
Neutro
Stroma
Average Expression
3
2
1
0
−1
Percent Expressed
25
50
75
100
Eya1
Foxc1
Slit2
Robo2
Ptch1
Smo
Gli3
Fgf1
Fgf8
Fgf9
Fgfr1
Fgfr2
Bmp7
Bmpr1a
Wnt4
Wnt5a
Sfrp1
Fzd2
Pdgfc
Pdgfra
Ntn1
Unc5c
Npnt
Itga8
Tnc
Itga9
Rspo3
Sdc4
Cdh1
Cdh2

Figure 6


A
Shroom3
C
Dab2
E
Uncx
RNA expr
Adult
P0
NP
Podo
PT S1
PT S3
LOH
DCT
PC
IC
Endo
Imm
Stroma
B
D
F
H3K27ac
H3K4me1
Six2
WGBS
RNA-seq
E15.5
P0
Adult
NPC
E14.5
E16.5
Gene
Gm20500
Dab2
Uncx
chr5:92,675,198-92,787,886
chr15:6,346,126-6,415,942
chr5:139,525,148-139,576,090