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
[#]Contributed equally
*Corresponding author

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

Abstract

Determining the epigenetic program that generates unique cell types in the kidney is critical for understanding cell-type heterogeneity during tissue homeostasis and injury response.

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 interactive public websites.

Keywords

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

55 **Introduction**

56

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

70

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

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

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

204 To study the open chromatin heterogeneity in different cell types, we derived a cell type-specific
205 accessible chromatin landscape by conducting pairwise Fisher's exact test for each peak between
206 every cluster (Benjamini-Hochberg adjusted q value ≤ 0.05, **Methods**). In total, we identified
207 60,684 differentially accessible open chromatin peaks (DAPs) across the 13 cell types
208 (**Supplemental Table 2, Figure 2a**). Among these peaks, most showed high specificity for a single
209 cluster. However, we noticed overlaps between the S1 and S3 proximal tubule segments-specific
210 peaks, as well as between the loop of Henle and distal convoluted tubule segments, which is
211 consistent with their biological similarities. In addition to the cell type-specific peaks, we also
212 found some cell-type independent open chromatin areas (present across nephron progenitors,
213 podocytes, proximal tubule and loop of Henle cells), likely consist of basal housekeeping genes
214 and regulatory elements. (**Figures 2a, S2c**).
215
216 We noticed that many genes had strong cell type-specific DAPs at their TSS. Other genes, however,
217 had accessible chromatin at their TSS in multiple cell types. For example, *Umod*, the loop of
218 Henle-specific marker gene, showed accessible chromatin at its TSS at multiple tubule cell types
219 (**Figure S2d, S1h**)**.** Rather than with its TSS, cell type-specific chromatin accessibility of *Umod*
220 strongly correlated with an upstream open chromatin peak, which is likely an enhancer region,
221 indicated by the H3K27Ac ChIP-seq signal (**Figure S2d**). In addition, we noticed the enrichment
222 of intronic regions and distal elements (**Figures S2e-f**) in cell type-specific DAPs, indicating their
223 role in cell type-specific gene regulation.
224
225 These observations motivated us to study cis-regulatory elements using the snATAC-seq data and
226 scRNA-seq data. We reasoned that a subset of the cell type-specific cis-regulatory elements should
227 regulate cell type-specific gene expression in cis. Inspired by Zhu et al. [23], we aligned DAPs and
228 differentially expressed genes from our snATAC-seq and scRNA-seq datasets, and inferred the
229 putative regulatory peaks by their proximity (**Methods**). Such cis-regulatory elements predictions
230 were confirmed by comparing with cis-regulatory elements inferred previously [24], as we
231 recapitulated roughly 20% of elements from their analysis. In addition, our analysis was able to
232 identify several known enhancers such as for *Six2* and *Slc6a18* [24,25] (**Figure S2g**).
233

234 To quantify the contribution of cis-regulatory elements, we analyzed peak co-accessibility patterns
235 using Cicero [26]. By using a heuristic co-accessible score 0.4 as a cutoff, we identified 232,380 and
236 206,701 cis-regulatory element links in the P0 and adult data, respectively. Some of these are likely
237 promoter-enhancer regulatory units. Among these co-accessible elements, only 74,694 were
238 common in P0 and adult kidneys, while most were developmental stage-dependent. While this
239 observation needs further experimental validation, it highlights dynamic changes in gene
240 regulation during development.
241
242 Another long-standing question has been to define the range of distances between interacting cis
243 regulatory elements (such as enhancer-enhancer or enhancer-promoter). To this end, we explored
244 the distances between co-accessible peaks using the Cicero output. We found that the number of
245 co-accessible peaks decreased with increasing distance between open chromatin regions (**Figure**
246 **2b**). 43% of the co-accessible peaks were within 100 kb distance, however, the strength of
247 association did not diminish with increasing distance. Even peaks that were as far as 500 kb apart
248 showed high (0.8) co-accessibility scores (**Figure 2c**). Overall, the median distance between cis
249 regulatory elements was relatively large (125.4 kb) and the number of interactions decreased with
250 increasing distance, however, the strength of association did not change with increasing distance.
251
252 Given the complex interaction between genomic regions, we next looked into identifying key TFs
253 that occupy the cell type-specific open chromatin regions. Until now, information on cell type-
254 specific TFs in the kidney has been scarce. Therefore, we performed motif enrichment analysis on
255 the cell type-specific open chromatin regions using HOMER [27]. HOMER was designed as a
256 differential motif discovery algorithm that scores motifs by computing enrichment of motif
257 sequences in target compared to a reference set. To reduce false discovery, we focused on the
258 known motifs. The full list of cell type-specific TF binding motifs is shown in **Supplemental**
259 **Table 3**. Since several TFs have identical or similar binding sequences, we next correlated motif
260 enrichment with scRNA-seq TF expression. Using this combined motif enrichment and gene
261 expression approach, we have defined the mouse kidney cell type-specific TF landscape. Examples
262 include *Six2* and *Hoxc9* in nephron progenitors, *Wt1* and *Mafb* in podocytes, *Hnf4a*, *Ppara*, and
263 *Bhle41* in proximal tubules, *Esrrb* and *Foxa1* in loop of Henle, *Vdr* in distal convoluted tubule,

264 *Elf5* in principal cells, *Tcfcp2l1* in intercalated cells, *Erg* and *Sox17* in endothelial cells, *Spi1* and
265 *Batf* in immune cells, and *Twist1* and *Nr2f2* in stromal cells (**Figures 2a, S2h**).
266
267 In order to study the putative target genes of TFs, we examined TF regulon activity using Single-
268 Cell rEgulatory Network Inference and Clustering (SCENIC) [28]. SCENIC was designed to reveal
269 TF-centered gene co-expression networks. By inferring a gene correlation network followed by
270 motif-based filtration, SCENIC keeps only potential direct targets of each TF as modules
271 (regulons). The activity of each regulon in each cell was quantified and then binarized to "on" or
272 "off" based on activity distribution across cells (**Methods**). SCENIC was also able to conduct
273 clustering based on the regulon states of each cell. SCENIC results (**Figures 2d-f**) indicated strong
274 enrichment in *Trps1*, *Hnf1b*, *Maf*, *Hnf1a*, and *Hnf4a* regulon activity in proximal tubules, *Hmga2*,
275 *Hoxc6*, *Hoxd11*, *Meox1*, *Six2*, *Tcf4*, and *Uncx* in nephron progenitors, *Esrrg*, and *Ppargc1a* in loop
276 of Henle, *Hmgb3* in proliferating cells and *Foxc1*, *Foxc2*, *Foxd1*, *Lef1*, and *Mafb* in podocytes,
277 respectively. While the expression of several of these TFs was relatively low and was further
278 exacerbated by transcript drop-outs, many TFs did not show strong cell type enrichment. The
279 regulon-based analysis, however, showed a very clear enrichment. SCENIC also successfully
280 reported multiple downstream target genes. The full list of regulons and their respective target
281 genes can be found in **Supplemental Table 4,** scaled and binarized regulon activity is also
282 available in **Supplemental Table 5.** Examples of regulon activity, corresponding TF expression,
283 and target gene expression are depicted in **Figures 2f, S2i**. For example, TFs such as *Eya1*, *Hoxc8*,
284 *Hoxc9*, *Pax2*, *Spock2* and *Wnt4* are important downstream targets within the regulon of nephron
285 progenitor-specific TF *Uncx*, indicating an important transcriptional hierarchy of nephron
286 development [29]. Finally, comparing the number of cell type-specific TFs reported by HOMER and
287 SCENIC to the number of cell type-specific TFs among DEGs from RNA expression data, it
288 became evident that cis-regulatory analysis in both snATAC-seq and scRNA-seq datasets yielded
289 significant benefits in discovering the TF-regulatory network over analyzing transcript data alone
290 (**Figures S2j-k**).
291
292 In summary, we generated a comprehensive atlas for the cell type-specific regulatory elements and
293 TF-centered regulatory network.
294

295 **The regulatory trajectory of nephron progenitor differentiation**

296

297 All cells in the body differentiate from the same genetic template. Cell type-specific chromatin
298 opening and closing events associated with TF binding changes set up the cell type-specific
299 regulatory landscape resulting in cell type specification and development. We found that closing
300 of open chromatin regions was the predominant event during the nephron progenitor
301 differentiation (**Figure S2a**). We then evaluated the cellular differentiation trajectory in the
302 snATAC-seq and scRNA-seq datasets (**Methods**). We identified multiple nephron progenitor sub-
303 groups (**Figures 3a-b**), which will need to be carefully mapped to prior gene expression- and
304 anatomical location-driven nephron progenitor sub-classification. Consistently, across both data
305 modalities, we identified that the podocyte precursors differentiated early from the nephron
306 progenitor pool (**Figures 3a-b**). The tubule cell trajectory was more complex with a shared
307 intermediate stage and later differentiation into proximal tubules and distal tubules/loop of Henle
308 (**Figures 3a-b, S3a-c**). We also integrated snATAC-seq and scRNA-seq data to obtain a single
309 trajectory (**Methods**). The cell types in this dataset were correctly mapped and the trajectory
310 resembled the path observed in individual analyses of the scRNA and snATAC datasets (**Figures**
311 **S3e-g**). The robustness of developmental trajectories was further supported by obtaining similar
312 results when performing RNA velocity analysis using Velocyto [30] (**Figure S3d**) and by comparing
313 with previous human and mouse kidney developmental studies [9,13,14].

314

315 Building on both the SCENIC-generated gene regulatory network and the robust differentiation
316 trajectories of the snATAC-seq and scRNA-seq datasets, we next aimed to understand chromatin
317 dynamics, identify TFs and driver pathways for cell type specification and differentiation. To this
318 end, we first determined variation in chromatin accessibility along the 3 differentiation trajectories
319 using ChromVAR [31]. ChromVAR estimates the accessibility dynamics of motifs in snATAC-seq
320 data (**Methods**). We observed three different patterns when analyzing genes of interest (**Figure**
321 **3c**): *1) Decrease of TF motif accessibility in all lineages.* For example, *Sox11* motif enrichment
322 score was high in nephron progenitor cells at the beginning of all 3 trajectories. It then decreased
323 in all 3 lineages in parallel, underlining the role of *Sox11* in early kidney development. Several
324 other TFs followed this pattern such as *Six2* and *Sox9. 2) Cell type-specific maintenance of*
325 *chromatin accessibility with advancing differentiation.* We observed that chromatin accessibility

326 for the *Wt1* motif was high initially but declined in proximal tubule and loop of Henle lineages,
327 while its expression increased in the podocyte lineage. This is consistent with the important role
328 of *Wt1* in nephron progenitors and podocytes [32,33]. Other TFs that followed this pattern include
329 *Foxc2* and *Foxl1*. *3) A de novo increase in chromatin accessibility with cell type commitment and*
330 *advancing differentiation.* For example, the chromatin accessibility of *Hnf4a* and *Pou3f3* motif
331 increased in proximal tubule and loop of Henle trajectories, respectively, coinciding with the
332 cellular differentiation program [34]. A large number of TFs followed this pattern such as *Mafb* (in
333 podocytes), *Hnf4a* and *Hnf1a* (in proximal tubule), *Hnf1b* (in both proximal tubule and loop of
334 Henle) as well as *Esrrb* and *Tfap2b* (in loop of Henle).
335
336 Next, we correlated changes in chromatin accessibility-based TF motif enrichment with TF
337 expression and their respective target genes along Monocle-generated trajectories. To this end, we
338 used the scRNA-seq differentiation trajectories to find TFs and target genes differentially
339 expressed over pseudotime (**Supplemental Table 6)**. We also noticed a good concordance of time-
340 dependent changes of TF and target gene expression along with TF motif enrichment, including
341 the lineages for podocytes (e.g., *Foxc2*, *Foxl1*, *Mafb*, *Magi2*, *Nphs1*, *Nphs2*, *Plat*, *Synpo*, *Thsd7a*,
342 *Wt1*, and *Zbtb7c*), proximal tubule (e.g., *Ace2*, *Atp1a1*, *Dab2*, *Hnf1a*, Hnf4a, *Hsd17b2*, *Lrp2*, *Maf*,
343 *Slc12a3*, *Slc22a12*, *Slc34a1*, and *Wnt9b*), loop of Henle (e.g.,*Cyfip2*, *Cytip*, *Esrrb*, *Esrrg*, *Irx1*,
344 *Irx2*, *Mecom*, *Pla2g4a*, *Pou3f3*, *Ppargc1a*, *Stat3*, *Sytl2*, *Tfap2b*, *Thsd4*, and *Umod*), as well as for
345 both proximal tubule and loop of Henle (e.g., *Bhlhe40*, *Hnf1b*, and *Tmprss2*), respectively (**Figures**
346 **3c, S3i**). Most interestingly, we noticed two distinct patterns of how gene expression was related
347 to chromatin accessibility. While gene expression of TFs increased over pseudotime, its
348 corresponding motif accessibility either increased in parallel (such as *Hnf4a* and *Pou3f3*) or
349 maintained in a lineage-specific manner (such as *Wt1*). This might indicate different regulatory
350 mechanisms during differentiation.
351
352 We next aimed to interrogate the stage-dependent chromatin dynamics along the identified
353 differentiation trajectory. The differentiation trajectory was binned into 15 developmental steps
354 based on the lineage specification (**Figures S3b-c**). These stages were labeled as NP (nephron
355 progenitor), IM (intermediate cells), Podo (podocytes), PT (proximal tubule), LOH (loop of Henle),
356 and DCT (distal convoluted tubule), however, this designation will need to be matched with prior

357 cell marker-based annotations. To study the chromatin opening and closing, we conducted
358 differential chromatin accessibility analysis between subsequent stages. To understand the
359 biological processes controlled by the epigenetic changes, we examined the nearest genes and
360 performed functional annotation (**Methods**). We found that open chromatin profiles were
361 relatively stable in the early precursor stages such as NP1 to NP3, with fewer than 70 DAPs
362 identified (**Supplemental Table 7, Figure S4a**). The podocyte differentiation branch was
363 associated with marked increase in the number of DAPs, (796 DAPs between NP3 and Podo1).
364 This mainly represented the closing of chromatin areas around nephron progenitor-specific genes
365 such as *Osr1*, *Gdnf*, *Sall1*, *Pax2* and opening of areas around podocyte-specific genes and key TFs
366 such as *Foxc2* and *Efnb2*, both of which are validated to be important for early podocyte
367 differentiation [35,36]. At later stages, there was a strong increase in expression of actin filament-
368 based processes and a significant decrease in *Notch* and *Ctnnb1* in the podocyte lineages
369 (**Supplemental Table 8**). Fewer chromatin closing events were observed (234 DAPs) between
370 NP3 and intermediate cells 1 (IM1), mainly associated with closing of the chromatin around *Osr1*
371 and opening around tubule cell-specific TFs such as *Lhx1* and *Pax3* (**Figure 4**). The decrease in
372 *Six2* expression only occurred at the IM2 stage, at which we also observed an increase in tubule
373 specification genes such as *Hnf1a*. Gene ontology results from the 820 up-regulated peaks between
374 PT1 and IM2 showed enrichment associated with typical proximal tubule functions including
375 sodium-dependent phosphate transport, maintenance of osmotic response in the loop of Henle and
376 active sodium transport in the distal convoluted tubule (**Figure S4a**, the full list can be found in
377 **Supplemental Tables 7, 8 and 9)**.
378
379 In addition to analyzing changes along the trajectory, we also specifically examined cell-fate
380 decision events. We studied the chromatin opening and closing during the first cell commitment
381 event. We found that podocyte specification from nephron progenitors was associated with
382 differential opening of *Foxl1*, *Zbt7c,* and *Smad2* in the podocyte lineage and *Lhx1*, *Sall1*, *Dll1*,
383 *Jag1*, *Cxcr3* and *Pax3* in the other lineage, respectively. While the role of several TFs has been
384 established for podocyte specification, the expression of *Foxl1* has not been described in the kidney
385 until now (**Figure 4**). Our analysis pinpointed that four peaks in the vicinity of *Foxl1* were
386 accessible only in podocyte lineage, which locate in +53,381 bp, +152,832 bp, +237,019 bp, and
387 +268,550 bp of the *Foxl1* TSS, respectively. To confirm the expression of *Foxl1* in nephron

388 progenitors and podocytes, we performed immunofluorescence studies on developing kidneys
389 (E13.5, P0 and P6). Consistent with the computational analysis, we found strong expression of
390 FOXL1 in nephron progenitors (E13.5). At later stages, it was present in comma and S shape body
391 and finally in the glomerular podocytes (**Figure S4b**). While there was no expression within cells
392 destined to become proximal tubule or loop of Henle cells, gene expression of *Foxl1* increased in
393 cells along the podocyte trajectory (**Figure S4c**). While further experimental validation will be
394 important, our study has illustrated the critical role of open chromatin state information and
395 dynamics in cellular differentiation.
396
397 The intermediate cells (IM) gave rise to proximal and distal branches, representing the proximal
398 tubules and the loop of Henle as well as distal convoluted tubule segments. The proximal tubule
399 region was characterized by chromatin opening around *Hnf4a, Maf*, *Tprkb*, and *Gpat2*. The loop
400 of Henle and distal convoluted tubule segments were remarkable for multiple DAPs in the vicinity
401 of *Tfap2a*, *Tfap2b*, *Cited4*, *Ephb2*, *Ephb3*, *Hoxd8*, *Mecom*, and *Prmd16*, indicating a critical novel
402 role for these TFs in distal tubule differentiation (**Figures 3c, 4, S4c**). Consistently, we saw a
403 reduction in chromatin accessibility of *Six2* promoter and enhancers along all three trajectories
404 (podocyte, proximal tubule and loop of Henle) (**Figure S4d**). There was also a decrease in
405 expression of *Jag1* and *Heyl* in the distal loop of Henle segment, concordant with the putative role
406 of Notch driving the proximal tubule fate [37] **(Supplemental Table 11)**. Another striking
407 observation was that tubule segmentation and specification occurred early by an increase in
408 chromatin accessibility around *Lhx1*, *Hnf1a* and *Hnf4a* and *Maf* for proximal tubule and *Tfap2b*
409 for loop of Henle. Terminal differentiation of proximal tubule and loop of Henle cells was strongly
410 linked to nuclear receptors that regulate metabolism, such as *Esrra* and *Ppara* in proximal tubules
411 and *Esrra* and *Ppargc1a* in the loop of Henle segment, once more indicating the critical role of
412 metabolism of driving gene expression and differentiation [38].
413
414 In summary, we reconstructed the developmental and differentiation trajectories of podocytes,
415 proximal tubule and loop of Henle cells. We defined chromatin and gene expression dynamics and
416 identified numerous putative TFs for kidney cell specification and differentiation.
417

418 **Stromal-to-epithelial communication is critical in the developing and adult kidneys**
419
420 Previous studies indicated that the survival, renewal, and differentiation of nephron progenitors is
421 largely regulated through its cross-talk with the adjacent ureteric bud [39]. To investigate the complex
422 cellular communication network, we used CellPhoneDB [40] to systematically infer potential cell-
423 cell communication in the developing and adult kidney. CellPhoneDB provides a comprehensive
424 database and a statistical method for the identification of ligand-receptor interactions in scRNA-
425 seq data. Analysis of our scRNA-seq dataset indicated that the number of cell-cell interaction pairs
426 was larger in developing kidney compared to the adult kidney (**Figure 5a**). In the developing
427 kidney, the stroma showed the greatest number of interactions among all cell types, coinciding the
428 well-known role of epithelial-stromal interactions in driving kidney development. Of the identified
429 interactions, many were related to stroma-secreted molecules such as collagen 1, 3, 4, 6, and 14
430 (**Figure 5b**). Furthermore, the stroma seemed to interact with most cell types, such as podocytes
431 and different tubule cells. Interestingly, the nephron progenitor cluster showed important ligand-
432 receptor interaction between *Fgf1*, *Fgf8* as well as *Fgf9* and the corresponding receptor *Fgfr1*,
433 which is consistent with the well-known role of FGF signaling in kidney development [41]. Of the
434 manifold identified interactions in the fetal kidney, stromal interaction and the VEGF-involving
435 interaction remained significant in the adult data set, underscoring the importance of endothelial-
436 to-epithelial communication.
437
438 We next individually examined the expression of several key pathways known to play important
439 roles in kidney development, such as Gdnf-Ret, sonic hedgehog, FGF, Bmp, Wnt and others [9].
440 Expression of these key ligand-receptor pairs showed strong cell type specificity (**Figure 5c**). For
441 example, *Robo2* of the Gdnf-Ret pathway was expressed in nephron progenitors and in podocytes
442 of P0 and adult kidney. Gdnf signaling through the Ret receptor is required for normal growth of
443 the ureteric bud during kidney development [42] and the Slit2/Robo2 pathway is implicated with
444 congenital kidney anomalies [43] and important for maintenance of podocyte foot process integrity
445 [44]. *Eya1*, however, is genetically upstream of *Gdnf* and acts as a positive regulator for its activation
446 [45]. Consistently, we noted distinct cell type specificity of *Eya1* expression only in nephron
447 progenitors, which was also true for other important signaling molecules such as *Ptch1, Smo* and
448 *Gli3* of the sonic hedgehog pathway. *Fgfr1* showed the highest expression in nephron progenitors

449 as well as in fetal and adult stroma, underscoring the importance of FGF signaling for cell-cell
450 interactions in both the developing and developed kidney. Most interestingly, some cell-cell
451 interactions between specific cell types that were observed in fetal kidney were abrogated in adult
452 kidney because of the loss of expression of either ligand or receptor, such as *Pdgfc* in nephron
453 progenitor signaling to its receptor *Pdgfra* in stroma, *Npnt* from several epithelial cells signaling
454 to *Itga8* in nephron progenitors, *Tnc-Itga9* signaling from nephron progenitors to stroma and
455 *Rspo3* in stroma signaling to *Sdc4* in several epithelial cells. Because not much is known about
456 some of these markers, the significance of these putative interactions requires further investigation.
457 For example, *Rspo3* has been implicated in nephron progenitor-associated interactions during
458 nephrogenesis [10]. Mutations in the *Itga8* gene are known to cause isolated congenital anomalies of
459 kidney and urinary tract in humans [43] and *Pdgfra* has been regarded as a commitment marker in
460 kidney differentiation [10].
461
462 In summary, we inferred cell-cell interactions in the developing and adult kidneys and found the
463 critical role of stromal-epithelial interactions in the developing kidney.
464
465 **Single cell chromatin accessibility identified human kidney GWAS target regulatory regions,**
466 **genes and cell types**
467
468 Finally, we examined whether single cell level chromatin accessibility data can help identify cell
469 and gene targets for human kidney disease development. GWAS have been exceedingly successful
470 in identifying nucleotide variations associated with specific diseases or traits. However, more than
471 90% of the identified genetic variants are in the non-coding region of the genome. Initial
472 epigenome annotation studies indicated that GWAS hits are enriched in tissue-specific enhancer
473 regions. As there are many different cell types in the kidney with differing function, understanding
474 the true cell type specificity of these enhancers is critically important. Here, we reasoned that
475 single cell accessible chromatin information could be extremely useful to identify the cell type-
476 specific enhancer regions and thereby the target cell type for the GWAS hits, however, such maps
477 have not been generated for the human kidney. We combined three recent kidney disease GWAS
478 [46-48], and obtained 26,637 single nucleotide polymorphisms (SNPs) that passed genome-wide
479 significance level of which we retained 7,923 after lift-over from human to mouse.

480

481 By overlapping the kidney disease-associated SNPs with peaks called by snATAC-seq data, we
482 found highly specific accessibility among different cell types (**Figure S5**). We found that most of
483 the peaks overlapped with nephron lineages, especially proximal tubules. The full table including
484 nearest genes is provided in the **Supplemental Table 12**. However, as the causal GWAS variant
485 is unknown, we conducted further investigation of kidney disease target genes and cell types using
486 multi-omics mouse kidney data.

487

488 Specifically, we examined loci where functional validation studies reported conflicting results on
489 target cell types and target genes (**Figure 6).** The *SHROOM3* locus has shown a reproducible
490 association with kidney function in multiple GWAS [48]. However, previous functional follow-up
491 studies have reported confusing and somewhat contradictory results. While one study indicated
492 that the genetic variants were associated with an increase in SHROOM3 levels in tubule cells
493 inducing kidney fibrosis [49], the other suggested that the variant was associated with lower
494 SHROOM3 levels in podocytes resulting in chronic kidney disease development [50]. We found an
495 open chromatin (likely promoter) area in multiple cell types such as nephron progenitors,
496 podocytes, loop of Henle, distal convoluted tubule, principal cells and intercalated cells (**Figure**
497 **6a**). We also identified intronic open chromatin areas only in nephron progenitors and podocytes
498 that overlapped with the GWAS significant variants (**Figure 6a**). Consistent with the cis-
499 regulatory open chromatin, the strongest expression of *Shroom3* was observed in podocytes and
500 nephron progenitor cells. Expression of *Shroom3* in the adult bulk kidney was below our detection
501 limit. To further understand the regulatory dynamics of this locus in the developing mouse kidneys,
502 we examined gene expression and epigenome annotation data generated from bulk mouse kidney
503 samples at different stages of development for H3K27ac and H3K4me1 in adult and fetal samples
504 (**Figure 6b**). Interesting to note that the GWAS-significant SNP that showed strong nephron
505 progenitor-specific enrichment also coincided with the *Six2* binding area. Finally, the Cicero-based
506 co-accessible analysis connected the GWAS top variants, which located in an intronic enhancer
507 region of *Shroom3*, with *Shroom3* exons, indicating that *Shroom3* is the likely target gene of the
508 variant (**Figure 6a**).

509

510 Next, we analyzed the chromosome 15 GWAS region, where we identified some open chromatin
511 regions that were uniformly open in all examined cell types. *Dab2* expression, on the other hand,
512 strongly correlated with open distal enhancer regions in proximal tubule cells (**Figure 6c**). This is
513 consistent with earlier publications indicating the role of proximal tubule-specific DAB2 playing
514 a role in kidney disease development [51]. Interestingly, while single cell analysis indicated an
515 additional distal enhancer in intercalated cells, the GWAS-significant region coincided with the
516 proximal tubule-specific enhancer region and showed strong coregulation (**Figure 6d**). Regulatory
517 annotation of the developing kidney indicated strong enhancer marks in the adult but not in the
518 fetal kidney.
519
520 Lastly, we examined the region around *Uncx*, for which reproducible association with kidney
521 function was shown in multiple GWAS [46,47]. Interestingly, the GWAS locus demonstrated a strong
522 open chromatin region in nephron progenitors but not in any other differentiated cell types (**Figure**
523 **6e**). Consistently, in bulk chromatin accessibility data we only observed regulatory activity such
524 as H3K27ac, H3Kme1 and we show *Six2*-binding at this locus in fetal kidneys. The locus did not
525 show H3K27ac enrichment in the adult kidney, while H3K4me1 remained positive (**Figure 6f**).
526 *Uncx* expression was strong in the fetal kidney samples, but we could not detect its expression in
527 the adult kidney (**Figure 6e**). A closer view of these loci is shown in **Figure S6**.
528
529 These results indicate that variants associated with kidney disease development are located in
530 regions with cell type- and developmental stage-specific regulatory activity and illustrate the
531 critical role of snATAC-seq in defining target genes and target cell types for GWAS variants.

532 **Discussion**

533

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

540

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

549

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

558

559 We identified critical cell type-specific TFs by integrating multiple computational analyses. TF
560 identification is challenging in scRNA-seq data since the expression of several cell type-specific
561 TFs is low and some of them do not show a high degree of cell type-specificity [52]. By extracting
562 motif information, snATAC-seq data provides additional information for TF identification.

563 Together with regulon analysis, as implemented in SCENIC, we have identified several TFs as
564 well as their target genes that are important for kidney development. Leveraging this newly
565 identified cell type-specific regulatory network will be essential for future studies of cellular
566 reprogramming of precursors into specific kidney cell types and for better understanding
567 homeostatic and maladaptive regeneration.
568
569 Our studies revealed dynamic chromatin accessibility that tracks with renal cell differentiation.
570 These states may reveal mechanisms governing the establishment of cell fate during development,
571 in particular those underlying the emergence of specific cell types. We found a consistent and
572 coherent pattern between gene expression and open chromatin information, where the nephron
573 progenitors differentiated into two branches representing podocytes and tubule cells [53]. We found
574 that podocytes commitment occurred earlier, while tubule differentiation and segmentation
575 appeared to be more complex. This podocyte specification correlated with the maintenance of
576 expression of *Foxc2* and *Foxl1* expression in podocytes. While *Foxc2* has been known to play a
577 role in nephron progenitors and podocytes, this is the first description of *Foxl1* in kidney and
578 podocyte development. Our studies are consistent with recent observations from organoid models
579 that recapitulated podocyte differentiation better than tubule cell differentiation [54]. Our study also
580 sheds light on tubule differentiation and segmentation. We confirmed the key role of *Hnf4a* in
581 proximal tubules. We have identified a large number of new transcriptional regulators such as
582 *Tfap2a* that seem to be critical for the distal portion of the nephron. Our data indicate that distal
583 tubule differentiation is linked to the loop of Henle, a critically important observation needing
584 further confirmation. Furthermore, the terminal differentiation of proximal tubule cells correlated
585 with the increase in *Ppara* and *Esrra* expression, both of which are known regulators of oxidative
586 phosphorylation and fatty acid oxidation [38]. Loop of Henle differentiation strongly correlated with
587 *Essrb* and *Ppargc1a* expression. These studies potentially indicate that cell specification events
588 occur early and metabolism controls terminal differentiation of tubule cells [55]. Impaired metabolic
589 fitness of proximal tubules has been a key contributor to kidney dysfunction, explaining the critical
590 association with tubule metabolism and function.
591
592 Furthermore, we show that single cell and stage level epigenome annotation is critical for the
593 annotation of human GWAS. Most identified GWAS signals are in the non-coding region of the

594 genome. Due to the linkage disequilibrium structure of the human genome, each GWAS locus
595 contains a large number of variants, each passing genome-wide significance level [56]. Furthermore,
596 as these signals are often non-coding, the target gene and the target cell type remain unknown.
597 While molecular quantitative trait locus studies and bulk epigenome annotation experiments have
598 been important to define the molecular pathways leading to disease development from the
599 identified signals, these methods have limited resolution, as cell type-specific enhancer regions
600 cannot be identified by bulk analysis [57]. Additionally, bulk molecular quantitative trait locus
601 studies suffer from the same linkage disequilibrium problems as GWAS analyses [58]. Our results
602 indicate that multiple GWAS regions are conserved between mice and humans. Single cell open
603 chromatin information enables not only the identification of affected cell types, but also the
604 understanding of co-regulation of the open chromatin area. It is also able to highlight critical target
605 genes. Performing single cell open chromatin analysis on human kidney tissue samples will be
606 essential to further understand molecular pathways altered by genetic variants. Here we showed
607 three important examples. We confirmed the role of *Dab2* and its specific expression in the
608 proximal tubule during kidney disease development, as its implication therein has been shown in
609 previous expression quantitative trait locus and bulk epigenome analysis experiments [51].
610 Furthermore, we showed that the GWAS variants map only to those regions where chromatin is
611 open exclusively in nephron progenitors, whereas chromatin becomes inaccessible as
612 differentiation progresses during later stages, such as *Shroom3* and *Uncx*. This is an interesting
613 and important novel mechanism, indicating that the altered expression of this gene might play a
614 role in the development rewiring of the kidney. This mechanism is similar to genes associated with
615 autism that are known to be expressed in the fetal but not in the adult stages [59] and highlights the
616 critical role of understanding chromatin accessibility at multiple stages of differentiation.
617
618 While we have generated a large amount of high-quality data, this information will need further
619 experimental validation, which is beyond the scope of the current manuscript. In addition, one
620 needs to be aware of the limitations when interpreting different computational analyses, for
621 example, the motif enrichment analyses such as implemented by HOMER, SCENIC, and
622 chromVAR, are not able to distinguish between TFs with similar binding sites. Future high-
623 throughput studies that analyze open chromatin and gene expression information from the same

624 cells will be exceedingly helpful to correlate open chromatin and gene expression information
625 along the differentiation trajectory [24,60,61].
626
627 In summary, our dataset provides critical novel insight into the cell type-specific gene regulatory
628 network, cell differentiation program, and disease development.
629
630
631 **Acknowledgements**
632 Work in the Susztak lab is supported by the NIH DK076077, DK087635, and DK105821. MSB is
633 supported by a German Research Foundation grant (BA 6205/2-1).
634
635 **Author Contributions**
636 KS and ZM designed and conceived the experiment. ZYM, JW, RS, and TA conducted the
637 experiment. ZM conducted snATAC-seq bioinformatics analysis with advice from KS, HL, ML,
638 and JK. MSB and ZM conducted scRNA-seq bioinformatics analysis with advice from KS. AMK
639 and AYK conducted immunofluorescence staining with supervision from KHK. KS, ZM, and
640 MSB wrote the manuscript and all authors edited and approved of the final manuscript.
641
642 **Declaration of Interests**
643 Authors declare no competing interests.

644 **Figure Legends**

645

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

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

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

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

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

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

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

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

672

673 **Figure 2. Cell type-specific gene regulatory landscape of the mouse kidney.**
674 (A) Left panel: Heatmap showing examples of the cell type-specific differentially accessible peaks
675 (DAPs) (yellow: open chromatin, blue: closed chromatin) (full results are shown in **Supplemental**
676 **Table 2**). Middle panel: Examples of cell type-specific motif enrichment analysis using Homer.
677 (full results are shown in **Supplemental Table 3**). Right panel: TF expression z score heatmap
678 that corresponds to the motif enrichment in each cell type.
679 (B) Histogram showing the distribution of distances between Cicero-inferred correlated peaks in
680 the P0 samples. The number of co-accessible peaks decreases with increasing distance.
681 (C) Density plot showing the distribution of peak co-accessibility scores and distance between
682 peaks. Although the number of peaks decreases with increasing distances, the co-accessibility
683 score distribution remains relatively stable.
684 (D) Regulon activity heatmap. Each column represents a single cell, colored by cluster assignment
685 and ordered by hierarchical clustering; each row represents binarized regulon activities ("on-
686 black", "off-white") and ordered by hierarchical clustering.
687 (E) tSNE representation of regulon density as a surrogate for stability of regulon states, as inferred
688 by SCENIC algorithm.
689 (F) tSNE depiction of regulon activity ("on-blue", "off-grey") and TF gene expression (red scale)
690 of exemplary regulons for proximal tubule (*Hnf1a*), nephron progenitors (*Uncx*), loop of Henle
691 (*Ppargc1a*), proliferating cells (*Hmgb3*) and podocytes (*Mafb*). Examples of target gene
692 expression of the *Uncx* regulon (*Eye1*, *Hoxc8*, *Pax2*, *Spock2* and *Wnt4*) are shown in purple scale.
693 Expression of target genes of *Hnf1a*, *Ppargc1a*, *Hmgb3* and *Mafb* is shown in **Figure S3d**.
694
695 **Figure 3. The cellular trajectory of nephron progenitor differentiation.**
696 (A) UMAP representation of snATAC-seq nephron progenitor differentiation trajectory towards
697 podocytes, proximal tubule, loop of Henle and distal convoluted tubule, respectively, as inferred
698 by Cicero. Cells are colored by pseudotime.
699 (B) UMAP representation of scRNA-seq nephron progenitor differentiation trajectory towards
700 podocytes, proximal tubule and loop of Henle, respectively, as inferred by Monocle3. Cells are
701 colored by pseudotime.
702 (C) Pseudotime-dependent chromatin accessibility and gene expression changes along the
703 proximal tubule (red), podocyte (green) and loop of Henle (blue) cell lineages. The first column

704 shows the dynamics of chromVAR TF enrichment score, the second column shows the dynamics
705 of TF gene expression values and the third and fourth column represent the dynamics of SCENIC-
706 reported target gene expression values of corresponding TFs, respectively. Additional examples
707 are given in **Figure S3e**.
708
709 **Figure 4. Chromatin dynamics of nephron progenitor differentiation.**
710 Di-graph representing cell type and lineage divergence, as derived from Cicero trajectory inference.
711 Nephron progenitors (NP), podocytes (Podo), intermediate stage (IM), proximal tubule (PT), loop
712 of Henle (LOH) and distal convoluted tubule (DCT) are connected with their developmental
713 precursor stages and represented by ascending numbering. Arrows represent cell differentiation
714 along respective trajectories. Genes listed next to the trajectories were derived from analyzing gene
715 enrichment of differentially assessible peaks (DAPs) between two stages. Genes colored red were
716 derived from the opening DAPs between two stages, genes colored blue were derived from the
717 closing DAPs between two stages, and genes colored green were derived from opening DAPs
718 between two branches. Three important genes, *Foxl1*, *Hnf4a* and *Tfap2b* are shown along with
719 their cell type-specific accessibility peaks and immunostaining results. Peaks that were open
720 during the development of specific cell types are shown in red boxes. Immunofluorescence
721 staining of fetal mouse kidney shows FOXL1 in red along cellular differentiation (from right to
722 left) from early progenitor stage (asterisk) over comma-shaped (+) and S shaped bodies (cross)
723 towards podocytes within primitive glomeruli (#). HNF4A and TFAP2B in human adult kidney
724 samples (taken from the Human Protein Atlas, http://www.proteinatlas.org [62]) are visualized by
725 immunohistochemistry in brown.
726
727 **Figure 5. Cell-cell communication analysis in the developing and adult mice highlighted the**
728 **critical role of stroma in driving cell differentiation.**
729 (A) Heatmaps showing the number of cell-cell interactions in the scRNA-seq dataset of P0 (top)
730 and adult (bottom) kidneys, as inferred by CellPhoneDB. Dark blue and dark red colors denote
731 low and high numbers of cell-cell interactions, respectively.
732 (B) CellPhoneDB-derived measures of cell-cell interaction scores and p values. Each row shows
733 a ligand-receptor pair, and each column shows the 2 interacting cell types, which is binned by cell
734 type. Columns are sub-ordered by first interacting cell type into stroma, podocytes, endothelial

735 cells, proximal tubule, loop of Henle and nephron progenitors. Color scale denotes the mean values
736 for all the interacting partners, where mean value refers to the total mean of the individual partner
737 average expression values in the interacting cell type pairs. Orange scale denotes P0, blue scale
738 denotes adult. Dot size denotes corresponding p values of the permutation test.
739 (C) Dot plots of RNA expression of important cell-cell communication candidates within the Gdnf-
740 Ret, Sonic hedgehog, Fgf, Bmp, Wnt and other pathways in both P0 (top) and adult (bottom)
741 kidney. Dot size denotes percentage of cells expressing the marker. Color scale represents average
742 gene expression values, orange denotes P0, blue denotes adult. Arrows indicate ligand-receptor
743 pairs.
744
745 **Figure 6. Single cell level chromatin accessibility highlighted human kidney GWAS target**
746 **genes and cell types.**
747 (A, C, E) From top to bottom: Cicero-inferred co-accessibility of open chromatin regions in mouse
748 orthologues of human *Shroom3*, *Dab2* and *Uncx* loci; Gene browser view of the single nucleotide
749 polymorphisms within the regions; gene browser view of chromatin accessibility for nephron
750 progenitors (NP), collecting duct intercalated cells (IC), collecting duct principal cell types (PC),
751 proximal tubules segment 1 and 3 (PT S1 and PT S3), loop of Henle (LOH), distal convoluted
752 tubule (DCT), stromal cells (stroma), podocytes (Podo), endothelial cells (Endo) and immune cells
753 (Immune). Right subpanel shows violin plots of scRNA-seq gene expression in P0 (orange) and
754 adult (blue) kidneys.
755 (B, D, F) Whole kidney H3K27ac, H3K4me1 and Six2 ChIP-seq, whole genome bisulfate
756 sequencing (WGBS) and RNA-seq data in E15.5, P0 and adult kidney samples.
757
758

759 **Methods**

760

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

762 1-day-old mouse neonate was decapitated with surgical scissors, 2 kidneys were harvested and
763 minced into 1 mm^3 pieces and incubated with digestion solution containing Enzyme D, Enzyme R
764 and Enzyme A from Multi Tissue Dissociation Kit (Miltenyi, 130-110-201) at 37 °C for 15 min
765 with agitation. Reaction was deactivated by adding 10% FBS, then solution was passed through a
766 40 μm cell strainer. After centrifugation at 1,000 RPM for 5 min, cell pellet was incubated with
767 500 μL of RBC lysis buffer on ice for 3 min. We centrifuged the cells at 1,000 RPM for 5 min at
768 4 °C and resuspended the cells in the buffer for further steps. Cell number and viability were
769 analyzed using Countess AutoCounter (Invitrogen, C10227). The cell concentration was 2.2
770 million cells/mL with 92% viability. 10,000 cells were loaded into the Chromium Controller (10X
771 Genomics, PN-120223) on a Chromium Single Cell B Chip (10X Genomics, PN-120262) and
772 processed to generate single cell gel beads in the emulsion (GEM) according to the manufacturer's
773 protocol (10X Genomics, CG000183). The library was generated using the Chromium Single Cell
774 3' Reagent Kits v3 (10X Genomics, PN-1000092) and Chromium i7 Multiplex Kit (10X Genomics,
775 PN-120262) according to the manufacturer's manual. Quality control for constructed library was
776 performed by Agilent Bioanalyzer High Sensitivity DNA kit (Agilent Technologies, 5067-4626)
777 for qualitative analysis. Quantification analysis was performed by Illumina Library Quantification
778 Kit (KAPA Biosystems, KK4824). The library was sequenced on an Illumina HiSeq or NextSeq
779 2x150 paired-end kits using the following read length: 28 bp Read1 for cell barcode and UMI, 8
780 bp I7 index for sample index and 91 bp Read2 for transcript.

781

782 **Single cell ATAC sequencing**

783 3-week-old and 8-week-old mice were euthanized and perfused with chilled 1x PBS via left
784 ventricle. Kidneys (0.25 g) were harvested, minced and lysed in 5 mL lysis buffer for 15 min. 1-
785 day-old mice were decapitated with surgical scissors, and both kidneys were harvested. Kidneys
786 were minced and lysed in 2 mL lysis buffer for 15 min. Tissue lysis reaction was then blocked by
787 adding 10 mL 1x PBS into each tube, and solution was passed through a 40 μm cell strainer. Cell
788 debris and cytoplasmic contaminants were removed by Nuclei PURE Prep Nuclei Isolation Kit
789 (Sigma, NUC-201) after centrifugation at 13,000 RPM for 45 min. Nuclei concentration was

790 calculated with Countess AutoCounter (Invitrogen, C10227). Diluted nuclei suspension was
791 loaded and incubated in transposition mix from Chromium Single Cell ATAC Library & Gel Bead
792 Kit (10X Genomics, PN-1000110) by targeting 10,000 nuclei recovery. GEMs were then captured
793 on the Chromium Chip E (10x Genomics, PN-1000082) in the Chromium Controller according to
794 the manufacturer's protocol (10X Genomics, CG000168). Libraries were generated using the
795 Chromium Single Cell ATAC Library & Gel Bead Kit and Chromium i7 Multiplex Kit N (10X
796 Genomics, PN-1000084) according to the manufacturer's manual. Quality control for constructed
797 library was perform by Agilent Bioanalyzer High Sensitivity DNA kit. The library was sequenced
798 on an Illumina HiSeq 2x50 paired-end kits using the following read length: 50 bp Read1 for DNA
799 fragments, 8 bp i7 index for sample index, 16 bp i5 index for cell barcodes and 50 bp Read2 for
800 DNA fragments.
801
802 **Bulk ATAC sequencing**
803 Bulk ATAC-seq was performed as described earlier [63,64]. Briefly, 50,000 nuclei/sample were
804 tagmented with Tn5 transposase (Illumina) in 50 µl reaction volume including Tween-20 (0.1%)
805 (Sigma) and digitonin (0.01%) (Promega). The reaction was carried out at 37 °C for 30 min in a
806 thermomixer at 1,000 RPM. After purification of DNA with Qiagen Minelute Reaction Cleanup
807 kit (Qiagen), samples were subjected to library amplification (8-10 cycles). Libraries were purified
808 with AmpureXP beads (Beckman Coulter) and their quality was assessed by Agilent High
809 sensitivity DNA Bioanalysis chip (Agilent). Libraries were submitted to 150 bp PE sequencing.
810
811 **snATAC-seq data analysis**
812 *Data processing and quality control*
813 Raw fastq files were aligned to the mm10 (GRCm38) reference genome and quantified using Cell
814 Ranger ATAC (v. 1.1.0). We only kept valid barcodes with number of fragments ranging from
815 1,000 to 40,000 and mitochondria ratio less than 10%. One of the important indicators for ATAC-
816 seq data quality is the fraction of peaks in promoter regions, so we did further filtration based on
817 promoter ratio. We noticed the promoter ratio seemed to follow a binary distribution, with most of
818 cells either having a promoter ratio around 5% (background) or more than 20% (valid cells)
819 (**Figure S1d**). We therefore filtered out cells with a promoter ratio <20%. After this stringent
820 quality control, we obtained 11,429 P0 single cells (5,993 in P0_batch_1 and 5,436 in P0_batch_2)

821 and 16,887 adult single cells (7,129 in P56_batch_3, 6,397 in P56_batch_4, and 3,361 in
822 P21_batch_5).
823
824 *Preprocessing*
825 Since snATAC-seq data are very sparse, previous methods either conducted peak calling or
826 binarization before clustering. Here, we chose to do binarization instead of peak calling for two
827 reasons: 1) Peak calling is time consuming; 2) Many peaks are cell type-specific, open chromatin
828 regions in rare populations are more likely to be treated as background. After binarizing fragments
829 into 5 kb bins and removing the fragments not matched to chromosomes or aligned to the
830 mitochondria, we binarized the cell-bin matrix. In order to only keep bins that were informative
831 for clustering, we removed the top 5% most accessible bins and bins overlapping with ENCODE
832 blacklist. The 484,606 remaining bins were used as input for clustering.
833
834 *Dimension reduction, batch effect correction and clustering*
835 Clustering was conducted using snapATAC [16], a single-cell ATAC-seq algorithm scalable to large
836 dataset. Previous benchmarking evaluation has shown that snapATAC was one of the best-
837 performing methods for snATAC-seq clustering [65]. Diffusion map was applied as a dimension
838 reduction method using function *runDiffusionMaps*. To remove batch effect, we used Harmony [17],
839 in which the low dimensional embeddings obtained from the diffusion map were used as input.
840 Harmony iteratively pulled batch-specific centroid to cluster centroid until convergence to remove
841 the variability across batches. After batch correction, a graph was constructed using k Nearest
842 Neighbor (kNN) algorithm with k=15, which was then used as input for Louvain clustering. We
843 used the first 20 dimensions for the Louvain algorithm. The number of dimensions was chosen
844 using a method recommended by snapATAC, although we noticed that the clustering results were
845 similar among a series of dimensions from 18 to 30.
846
847 *Cell type annotation*
848 We used a published list of marker genes [9,19] to annotate kidney cell types. In order to infer gene
849 expression of each cell type, we built a cell-gene activity score matrix by integrating all fragments
850 that overlapped with gene transcript. We used GENCODE Mouse release VM16 [22] as reference
851 annotation.

852

853 *Peak calling and visualization*

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

860

861 We also visualized public chromatin ChIP-seq data and RNA-seq data obtained from ENCODE
862 Encyclopedia with the following identifiers: ENCFF338WZP, ENCFF872MVE, ENCFF455HPY,
863 ENCFF049LRQ, ENCFF179NTO, ENCFF071PID, ENCFF746MFH, ENCFF563LOO,
864 ENCFF184AYF, ENCFF107NQP, ENCFF465THI, ENCFF769XWI, ENCFF591DAX. The Six2
865 ChIP-seq data were obtained from [66] and the WGBS data were obtained from [67].

866

867 *Genomic elements stratification*

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

877

878 *Identification of differentially accessible regions*

879 Peaks identified in each cell type were combined to build a union peak set. Overlapping peaks
880 were then merged to one peak using *reduce* function from the GenomicRanges package. This
881 resulted in 300,755 peaks, which were used to build binarized cell-by-peak matrix. Differentially
882 accessible peaks (DAPs) for each cell type were identified by pairwise peak comparison.

883 Specifically, for each peak, we conducted a Fisher's exact test between a cell type and each of the
884 other cell types. To address multiple testing problem, we used the Benjamini-Hochberg approach
885 (BH correction) to correct p values. Peaks with corrected p values below significance level (0.05)
886 in all pairwise tests were defined as DAPs. In total, we obtained 60,683 DAPs, which were used
887 for motif enrichment analysis.
888
889 *Motif enrichment analysis*
890 Motif enrichment analysis was conducted using DAPs by HOMER v4.10.4 [27] with parameters
891 background="automatic" and scan.size=300. We noticed that *de novo* motif identification only
892 generated few significant results, so we focused on known motifs for our following study. We used
893 the significance level of 0.05 for BH corrected p value to determine the enriched results. The motif
894 enrichment results are provided in **Supplemental Table 3**.
895
896 *Peak-peak correlation analysis*
897 Peak-peak correlation analysis was conducted using Cicero [26]. In order to find developmental
898 stage-specific peak-peak correlations, the analysis was conducted for P0 and adult separately.
899 Cicero uses Graphic Lasso with distance penalty to assess the co-accessibility between different
900 peaks. Cicero analysis was conducted using the *run_cicero* function with default parameters. A
901 heuristic cutoff of 0.25 score of co-accessibility was used to determine the connections between
902 two peaks.
903
904 *snATAC-seq trajectory analysis*
905 snATAC-seq trajectory was conducted using Cicero, which extended Monocle3 to the snATAC-
906 seq analysis. We obtained the preprocessed P0 snATAC-seq cell-peak matrix from snapATAC as
907 input for Cicero and conducted dimension reduction using Latent Semantic Indexing (LSI) and
908 visualized using UMAP. Trajectory graph was built using the function *learn_graph*. Batch effect
909 was not observed between the two P0 batches, and the trajectory graph was consistent with cell
910 type assignment with clustering analysis (**Figures S3 a-b**).
911
912 In order to study how open chromatin changes are associated with the cell fate decision, we first
913 binned the cells into 15 groups based on their pseudotime and cell type assignment. Next, we

914 studied the DAPs between each group and its ancestral group using the same methods described
915 above. The number of newly open and closed chromatins were reported using pie charts. The exact
916 peak locations are provided in the **Supplemental Table 7**.
917
918 *Genes and gene ontology terms associated with snATAC-seq trajectory*
919 Based on the binned trajectory graphs and DAPs between each group and its ancestral group, we
920 next used GREAT tool [68] to study the enrichment of associated genes and gene ontology (GO)
921 terms along the trajectory. We used the newly open or closed peaks as test regions and all the peaks
922 from peak-calling output as the background regions for the analysis. The output can be found in
923 the **Supplemental Table 8 and 9**.
924
925 *Predict cis-regulatory elements*
926 We implemented two methods to study cis-regulatory elements in the snATAC-seq data. The first
927 method was inspired by Zhu et al. [23], which was based on the observation that there was co-
928 enrichment in the genome between the snATAC-seq cell type-specific peaks and scRNA-seq cell
929 type-specific genes. This method links a gene with a peak if 1) they were both specific in the same
930 cell type, 2) they were in *cis*, meaning that the peak is in ±100 kb region of the TSS of the
931 corresponding gene, and 3) the peak did not directly overlap with the TSS of the gene. This method
932 successfully inferred several known distal elements such as for *Six2* and *Slc6a18* (**Figure S2g**).
933
934 Alternatively, we assessed the co-accessibility of two peaks. We implemented Cicero [26], which
935 aggregates similar cells to obtain a set of "meta-cells" and address the issue with sparsity in the
936 snATAC-seq data. We used *run_cicero* function with default parameters to predict cis-regulatory
937 elements (CREs). Although it is recommended to use 0.25 as a cutoff for co-accessibility score,
938 we noticed that this resulted in a great amount of CREs, which could contain many false positives.
939 Thus, we used a more stringent score of 0.4 for the cutoff and retained 232,380 and 206,701 CRE
940 links in the P0 and adult data, respectively.
941
942 **Bulk ATAC sequencing analysis**
943 Bulk ATAC-seq raw fastq files were processed using the end-to-end tool ENCODE ATAC-seq
944 pipeline (**Software and Algorithms**). This tool provided a standard workflow for ATAC-seq data

945 quality control, adaptor removal, alignment, and peak calling. To obtain high quality ATAC-seq
946 peaks, peak calling results from two biological replicates were compared and only those peaks that
947 were present in both replicates were kept, which were further used to compare with snATAC-seq
948 peaks.
949
950 **Correlation of bulk and single nuclei ATAC sequencing data**
951 snATAC-seq reads were aggregated to a pseudo-bulk data for the comparison purpose. To prevent
952 the effect of sex chromosome and mitochondria chromosome, reads from chromosome X, Y and
953 M were excluded from our analysis. We used multiBigwigSummary tool from deeptools [69] to study
954 the correlation between different samples. Specifically, the whole genome was binned into equally
955 sized (10 kb) windows, and the reads in each bin were aggregated, generating a bin-read count
956 vector for each of the sample. The correlation of these vectors was computed as a measure of
957 pairwise similarity between samples.
958
959 To compare the number of peaks in these two datasets, we used as input the narrowpeak files from
960 the snATAC-seq and bulk ATAC-seq analysis. We filtered out bulk ATAC-seq peaks with q value >
961 0.01 to be consistent with the snATAC-seq setting. Since the snATAC peaks were called after
962 merging different time points, we also took the union set of bulk ATAC-seq peaks from different
963 time points. We then used *findoverlap* function in GenomicRanges package [70] to find and report
964 overlapped peaks.
965
966 **Comparison between single nuclei ATAC sequencing data and single cell RNA sequencing**
967 **data**
968 In order to compare the cluster assignment between snATAC-seq data and scRNA-seq data, we
969 obtained the average gene expression values and peak accessibility in each cluster for P0 and adult
970 samples separately. We next transformed snATAC-seq data by summing up the reads within gene
971 body and 2 kb upstream regions to build gene activity score matrix, as suggested in Seurat [20]. Then,
972 we normalized the data and computed the mean expression and mean gene activity scores in each
973 cell type, and calculated z scores of each gene. Pearson's correlation coefficient was then
974 calculated among top 3,000 highly variable genes between snATAC-seq data and scRNA-seq data.

975 We found high concordance between these two datasets in terms of cell type assignment (**Figure**
976 **S4**).
977
978 **Single cell RNA sequencing data analysis**
979 *Alignment and quality control*
980 Raw fastq files were aligned to the mm10 (Ensembl GRCm38.93) reference genome and
981 quantified using CellRanger v3.1.0. Seurat R package v3.0 was used for data quality control,
982 preprocessing and dimensional reduction analysis. After gene-cell data matrix generation of both
983 P0 and adult datasets, matrices were merged and poor-quality cells with <200 or >3,000 expressed
984 genes and mitochondrial gene percentages >50 were excluded, leaving 25,138 P0 and 18,498 adult
985 cells for further analytical processing, respectively (**Figures S1j-k**).
986
987 *Pre-processing, batch effect correction and dimension reduction*
988 Data were normalized by RPM following log transformation and 3,000 highly variable genes were
989 selected for scaling and principal component analysis (PCA). Harmony R package v1.0 [17] was
990 used to correct batch effects. The top 20 dimensions of Harmony embeddings were used for
991 downstream uniform manifold approximation and projection (UMAP) visualization and clustering
992 (**Figures S1l-m**).
993
994 *Cell clustering, identification of marker genes and differentially expressed genes*
995 Louvain algorithm with resolution 0.4 was used to cluster cells, which resulted in 18 distinct cell
996 clusters. A gene was considered to be differentially expressed if it was detected in at least 25% of
997 one group and with at least 0.25 log fold change between two groups and the significant level of
998 BH-adjusted p value <0.05 in Wilcoxon rank sum test was used. We used a list of marker genes
999 [9,19] to manually annotate cell types. 2 distal convoluted tubule clusters were merged based on the
1000 marker gene expression, resulting in a total of 17 clusters (**Figures S1i, n, o**).
1001
1002 *scRNA-seq trajectory analysis*
1003 <u>Monocle3</u>
1004 To construct single cell pseudotime trajectory and to identify genes whose expression changed as
1005 the cells underwent transition, Monocle3 v0.1.3 [71] was applied to P0 cells of the following Seurat

1006 cell clusters: nephron progenitors (NP), proliferating cells, stroma-like cells, podocytes, loop of
1007 Henle (LOH), early proximal tubule (PT), proximal tubule S1, proximal tubule S3 cells.
1008
1009 To show cell trajectories from both small (nephron progenitors) and large cell populations
1010 (proximal tubule), an equal number of 450 cells per cluster was randomly subsampled. Cells were
1011 re-clustered by Monocle3 using a resolution of 0.0005 with k-nearest neighbor (kNN) k=29.
1012 Highly variable genes along pseudotime were identified using differential *GeneTest* function and
1013 cells were ordered along pseudotime trajectory. NP cluster was defined as earliest principal node.
1014 In order to find genes differentially expressed along pseudotime, trajectories for podocytes, loop
1015 of Henle, and proximal tubule clusters were analyzed separately with the *fit_models* function of
1016 Monocle3. Genes with a q value <0.05 in the differential *GeneTest* analysis were kept. In an
1017 alternate approach, *graph_test* function of Monocle3 was used and trajectory-variable genes were
1018 collected into modules at a resolution of 0.01.
1019
1020 RNA velocity
1021 To calculate RNA velocity, Python-based Velocyto command-line tool as well as Velocyto.R
1022 package were used as instructed [30]. We used Velocyto to calculate the single-cell
1023 trajectory/directionality using spliced and unspliced reads. From loom files produced by the
1024 command-line tool, we subset the exact same cells that were previously selected randomly for
1025 Monocle trajectory analysis. This subset was loaded into R using the SeuratWrappers v0.1.0
1026 package. RNA velocity was estimated using gene-relative model with k-nearest neighbor (kNN)
1027 cell pooling (k = 25). The parameter n was set at 200, when visualizing RNA velocity on the
1028 UMAP embedding.
1029
1030 *Gene regulatory network inference*
1031 In order to identify TFs and characterize cell states, we employed *cis*-regulatory analysis using the
1032 R package SCENIC v1.1.2.2 [28], which infers the gene regulatory network based on co-expression
1033 and DNA motif analysis. The network activity is then analyzed in each cell to identify recurrent
1034 cellular states. In short, TFs were identified using GENIE3 and compiled into modules (regulons),
1035 which were subsequently subjected to *cis*-regulatory motif analysis using RcisTarget with two
1036 gene-motif rankings: 10 kb around the TSS and 500 bp upstream. Regulon activity in every cell

1037 was then scored using AUCell. Finally, binarized regulon activity was projected onto Monocle3-
1038 created UMAP trajectories.
1039
1040 *Ligand-receptor interactions*
1041 To assess cellular crosstalk between different cell types, we used the CellPhoneDB repository to
1042 infer cell-cell communication networks from single cell transcriptome data [40]. We used the Python
1043 package CellPhoneDB v2.1.2 together with the database v2.0.0 to predict cell type-specific ligand-
1044 receptor complexes as per the authors' instructions. Only receptors and ligands expressed in more
1045 than 5% of the cells in the specific cluster were considered. 1,000 iterations were used for pairwise
1046 comparison between cell types and considered for further statistical analysis.
1047
1048 **Immunofluorescence staining**
1049 Mouse kidneys were fixed with 4% paraformaldehyde overnight, rinsed in PBS, and dehydrated
1050 for paraffin embedding. Antigen retrieval was performed using Tris-EDTA buffer pH 9.0 with a
1051 pressure cooker (PickCell Laboratories, Agoura Hills, CA) and antibody staining performed as
1052 described [72]. Antibodies used were as follows: guinea pig FOXL1 (1:1,500) [73], mouse E-cadherin
1053 (1:250; BD Transducton 610182, Franklin Lakes, NJ). Cy2-, Cy3-, and Cy5-conjugated donkey
1054 secondary antibodies were purchased from Jackson ImmunoResearch Laboratories, Inc.
1055 Fluoresecence images were collected on a Keyence microscope.
1056
1057 **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		
Mouse adult kidney WGBS data	(ref. 67)	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. 66)	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

Lead Contact and Data 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 susztaklab.com/developing_adult_kidney/snATAC/, susztaklab.com/developing_adult_kidney/scRNA/, and susztaklab.com/developing_adult_kidney/igv/ . Further information and requests for resources and

1066 reagents should be directed to and will be fulfilled by the lead contact: Katalin Susztak. Email:
1067 ksusztak@pennmedicine.upenn.edu.
1068
1069

1070 **Supplemental Information**

1071

1072 **Supplemental Tables**

1073 **Supplemental Table 1.** Cell type marker genes derived from scRNA-seq analysis.

1074 **Supplemental Table 2.** Cell type-specific open chromatin derived from snATAC-seq analysis.

1075 **Supplemental Table 3.** Cell type-specific motif enrichment.

1076 **Supplemental Table 4.** Regulons and respective target genes inferred by SCENIC.

1077 **Supplemental Table 5.** Binarized regulon activities in each cell type inferred by SCENIC.

1078 **Supplemental Table 6.** Differentially expressed genes along pseudotime in distinct lineages in
1079 scRNA-seq data.

1080 **Supplemental Table 7.** Differentially accessible peaks along pseudotime in distinct lineages in
1081 snATAC-seq data.

1082 **Supplemental Table 8**. Nearest genes of differentially accessible peaks along pseudotime in
1083 distinct lineages in snATAC-seq data.

1084 **Supplemental Table 9**. GO enrichment of differentially accessible peaks along pseudotime in
1085 distinct lineages inferred by GREAT analysis.

1086 **Supplemental Table 10.** ChromVAR cell-TF enrichment score matrix.

1087 **Supplemental Table 11.** Nearest genes of differentially accessible peaks at bifurcation events
1088 along pseudotime in distinct lineages inferred by GREAT analysis.

1089 **Supplemental Table 12.** Proportion of cells in each cell type with accessible chromatin
1090 overlapped with kidney disease associated SNPs.

1091

1092

1093 **Supplemental Figures**

1094 **Figure S1. Quality control and data processing methods for snATAC-seq and scRNA-seq**
1095 **data analysis.**

1096 (A) Insert size distribution of the 5 snATAC-seq samples showing periodic patterns.

1097 (B) Transcription start sites (TSS) signal enrichment of the 5 snATAC-seq samples.

1098 (C) Spearman correlation between snATAC-seq datasets and bulk ATAC-seq of binned genomic
1099 regions.

1100 (D) Distribution of number of unique molecular identifiers (UMIs, x axis) and promoter ratio (y
1101 axis) in 5 samples shown by dot plot.
1102 (E) Violin plots representing the number of accessible peaks across different clusters in the
1103 snATAC-seq dataset indicating similar distributions.
1104 (F) UMAP representation of the snATAC-seq dataset colored by batches.
1105 (G) Stacked bar graphs representing absolute numbers and percentages of identified cell types
1106 across snATAC-seq batches.
1107 (H) Genome browser view of cell type-specific peaks at the TSS of marker genes for 13 cell types
1108 in the snATAC-seq dataset.
1109 (I) From left to right: Stacked bar graphs showing the percentage of different cell types in the P0
1110 and adult scRNA-seq datasets, tables showing the number of cells in each cell type (nCells) and
1111 corresponding percentage. NP, nephron progenitor; Podo, podocyte; PT, proximal tubule; S1,
1112 segment 1; S3, segment 3; LOH, loop of Henle; DCT, distal convoluted tubule cells; PC, collecting
1113 duct principal cells; IC, collecting duct intercalated cells; Endo, endothelial cells; Macro,
1114 macrophages; Neutro, neutrophils.
1115 (J) UMAP representation of scRNA-seq data colored by the mitochondrial gene ratio (Mt %).
1116 (K) Violin plots showing number of informative genes per single cell and unique molecular
1117 identifiers (UMIs) per single cell. Blue denotes adult kidney, orange denotes P0 kidney.
1118 (L, M) Principal component (PC) representation of combined adult and P0 scRNA-seq dataset (left
1119 panel) and violin plots of corresponding embeddings values (right panel) before (L) and after (M)
1120 batch correction using Harmony.
1121 (N) Dot plot of cell type-specific marker genes. Dot size denotes percentage of cells expressing
1122 the marker. Color scale represents average expression, orange denotes P0, blue denotes adult
1123 kidney.
1124 (O) Feature plots of representative marker genes projected on UMAP dimension.
1125 (P) Correlation between snATAC-seq gene activity scores and gene expression values in adult data,
1126 which is complementary to **Figure 1g**.
1127 (Q) We provide the processed chromatin accessibility dataset via a searchable, interactive website
1128 (susztaklab.com/developing_adult_kidney/igv/). Ace2 was used as an example, and we show
1129 proximal tubule-specific enrichment of peaks at transcription start sites of the *Ace2* (Angiotensin-
1130 converting enzyme 2) gene (red boxes).

1131

1132 **Figure S2. Characterization of the cell type-specific regulatory landscape**

1133 (A) Bar graph representing the number of accessible peaks in distal elements, promoters, introns,
1134 5'-UTR, 3'UTR and exons, as distributed across samples of snATAC-seq data and bulk ATAC-
1135 seq data.

1136 (B) Overlap of scATAC-seq differentially accessible peaks among cell types with H3K27Ac ChIP-
1137 seq data.

1138 (C) Number of shared and unique peaks among snATAC-seq cell types. Cell types include nephron
1139 progenitors and cells differentiated from nephron progenitors.

1140 (D) Genome browser view of Umod as an example for distal open chromatin region and its target
1141 promoter region.

1142 (E) Distribution of different open chromatin elements in snATAC-seq cell types.

1143 (F) Distribution of different open chromatin elements among differentially accessible peaks (DAPs)
1144 in snATAC-seq cell types.

1145 (G) Genome browser representations of single cell open chromatin data for individual cell types
1146 at chromosomal loci around Six2 and Slc6a18, along with their known distal elements (red boxes).
1147 Corresponding chromosomal interaction of open chromatin regions, as inferred by Cicero
1148 (**Methods**), is depicted at the top.

1149 (H) Genome browser views of representative marker genes demonstrating cell type-specific
1150 chromatin accessibility for proximal tubule (*Hnf4a* and *Hmgb3*), several tubular segments (*Hnf1b*),
1151 loop of Henle and distal convoluted tubule (*Esrrb* and *Ppargc1a*) as well as nephron progenitors
1152 and podocytes (Wt1).

1153 (I) UMAP depiction of regulon activity ("on-blue", "off-grey") and RNA expression (red scale) of
1154 exemplary regulons of proximal tubule (*Hnf1a*), nephron progenitors (*Six2*), loop of Henle
1155 (*Ppargc1a*), proliferating cells (*Hmgb3*) and podocytes (*Mafb*), respectively. Exemplary target
1156 gene expression for the respective TF is shown in purple scale.

1157 (J, K) Bar graphs depicting the absolute number of cell type-specific TFs reported by HOMER (J)
1158 and SCENIC (K) cis-regulatory analyses, respectively, as well as the number of TFs among DEGs
1159 from RNA expression data alone.

1160

1161 **Figure S3. snATAC-seq and scRNA-seq cell differentiation trajectories.**

1162 (A) UMAP representation of snATAC-seq trajectory lineages of podocytes, proximal tubule and
1163 loop of Henle cells from nephron progenitors colored by 2 P0 batches.

1164 (B) UMAP representation of snATAC-seq trajectory lineages of podocytes, proximal tubule and
1165 loop of Henle cells from nephron progenitors colored by original cell type assignment as in **Figure**
1166 **1b**.

1167 (C) UMAP representation of scRNA-seq trajectory lineages of podocytes, proximal tubule and
1168 loop of Henle cells from nephron progenitors colored by original cell type assignment as in **Figure**
1169 **1b**.

1170 (D) UMAP representation of RNA velocity of scRNA-seq trajectory inferred by VelocytoR,
1171 colored by original cell type assignment. Each dot is one cell and each arrow represents the time
1172 derivative of the gene expression state.

1173 (E) UMAP representation of snATAC-scRNA integration results colored by cell type assignment.

1174 (F) UMAP representation of snATAC-scRNA integration results colored by technologies
1175 (snATAC=red, scRNA=grey). Podo: podocytes, PT: proximal tubule, LOH: loop of Henle, DCT:
1176 distal convoluted tubule, NP: nephron progenitors, IM: intermediate stage cells.

1177 (G) Dot plot showing snATAC-scRNA integration cell type assignment confusion matrix. Each
1178 column represents the original cell type assignment of snATAC-seq data, and each row represents
1179 the predicted cell type assignment by the integration analysis scRNA-seq data. Each dot represents
1180 the number of cells that were matched in the integrated data.

1181 (H) Heatmap of chromVAR enrichment results. The original data matrix is given in **Supplemental**
1182 **Table 10**.

1183 (I) Pseudotime-dependent chromatin accessibility and gene expression changes along the proximal
1184 tubule (red), podocytes (green) and loop of Henle (blue) cell lineages. The first column represents
1185 the dynamics of chromVAR TF enrichment score, the second column represents the dynamics of
1186 TF gene expression values, and the third and fourth column represent the dynamics of SCENIC-
1187 reported target gene expression values.

1188

1189 **Figure S4. Chromatin dynamics of nephron progenitor differentiation.**
1190 (A) Di-graph representing cell type and lineage divergence, as derived from Cicero trajectory
1191 inference. Nephron progenitors (NP), podocytes (Podo), intermediate stage (IM), proximal tubule
1192 (PT), loop of Henle (LOH) and distal convoluted tubule (DCT) are connected with their
1193 developmental precursor stages and ordered by ascending numbering. Pie charts represent
1194 differentially assessible peaks (DAPs) between two stages, where the size of pie charts is
1195 proportional to the number of DAPs, orange color represents the number of open peaks, grey color
1196 the number of closed peaks. Bar graphs depict gene ontology (GO) term analysis of genes nearby
1197 DAPs derived from GREAT analysis (full list in **Supplemental Table 9**).
1198 (B) Immunofluorescence staining of fetal mouse kidney. Upper panel and insert denote E13.5 stage,
1199 lower panel denotes P6 mouse. Blue staining represents nuclei (DAPI), green staining represents
1200 tubular epithelium (E-Cadherin) and red staining represents progenitor cells (FOXL1) along a
1201 developmental trajectory from early progenitor stage (asterisk) over comma-shaped (+) and S
1202 shaped bodies (cross) towards podocytes within primitive glomeruli (#).
1203 (C) Pseudotime-dependent chromatin accessibility and gene expression changes along the
1204 proximal tubule (red), podocytes (green) and loop of Henle (LOH, blue) cell lineages for important
1205 bifurcation TFs in the podocyte (*Foxl1*) and distal tubule (*Tfap2b*) lineage.
1206 (D) Bar graphs denote the percentage of cells with accessible chromatin of several *Six2* promoters
1207 and enhancers (gene loci numbered 1-3) as well as putative *Foxl1* enhancers (gene loci numbered
1208 4-7) along pseudotime. Exact gene loci of enhancers and promoters are given above each
1209 respective graph. Changes along pseudotime are depicted for 3 lineages from nephron progenitors
1210 (NP) to podocytes, proximal tubule (PT) and loop of Henle (LOH) cells, respectively. The right
1211 upper subpanel depicts the genome browser overview of chromatin accessibility for the NP and
1212 therefore corresponds to the first bar in graphs on the left. The right lower subpanel depicts zoom-
1213 in versions of the 7 loci for all 3 lineages.
1214
1215 **Figure S5. Accessibility of peaks overlap with kidney disease SNPs.**
1216 Heatmap showing the proportion of cells in each cell type with accessibility of SNP-overlapped
1217 peaks. The SNP IDs as well as nearest genes are provided in **Supplemental Table 12**.
1218

1219 **Figure S6. Single cell level chromatin accessibility highlighted human kidney GWAS target**
1220 **genes and cell types.**
1221 Open chromatin and co-accessibility view at alternative scales to those shown in **Figure 6**.
1222 (A, C, E) From top to bottom: Cicero-inferred co-accessibility of open chromatin regions in mouse
1223 orthologues of human *Shroom3*, *Dab2* and *Uncx* loci; Gene browser view of the single nucleotide
1224 polymorphisms within the regions; gene browser view of chromatin accessibility for nephron
1225 progenitors (NP), collecting duct intercalated cells (IC), collecting duct principal cell types (PC),
1226 proximal tubules segment 1 and 3 (PT S1 and PT S3), loop of Henle (LOH), distal convoluted
1227 tubule (DCT), stromal cells (stroma), podocytes (Podo), endothelial cells (Endo) and immune cells
1228 (Immune). Right subpanel shows violin plots of scRNA-seq gene expression in P0 (orange) and
1229 adult (blue) kidneys.
1230 (B, D, F) Whole kidney H3K27ac, H3K4me1 and Six2 ChIP-seq, whole genome bisulfate
1231 sequencing (WGBS) and RNA-seq data in E15.5, P0 and adult kidney samples.
1232
1233
1234 **References**
1235

1236 1 Reidy, K., Kang, H. M., Hostetter, T. & Susztak, K. Molecular mechanisms of diabetic
1237 kidney disease. *J Clin Invest* **124**, 2333-2340, doi:10.1172/JCI72271 (2014).

1238 2 Costantini, F. & Kopan, R. Patterning a complex organ: branching morphogenesis and
1239 nephron segmentation in kidney development. *Dev Cell* **18**, 698-712,
1240 doi:10.1016/j.devcel.2010.04.008 (2010).

1241 3 Park, J. S. *et al.* Six2 and Wnt regulate self-renewal and commitment of nephron
1242 progenitors through shared gene regulatory networks. *Dev Cell* **23**, 637-651,
1243 doi:10.1016/j.devcel.2012.07.008 (2012).

1244 4 Harding, S. D. *et al.* The GUDMAP database--an online resource for genitourinary
1245 research. *Development (Cambridge, England)* **138**, 2845-2853, doi:10.1242/dev.063594
1246 (2011).

1247 5 Wu, H. *et al.* Comparative Analysis and Refinement of Human PSC-Derived Kidney
1248 Organoid Differentiation with Single-Cell Transcriptomics. *Cell Stem Cell* **23**, 869-881
1249 e868, doi:10.1016/j.stem.2018.10.010 (2018).

1250 6 Takasato, M. *et al.* Kidney organoids from human iPS cells contain multiple lineages and
1251 model human nephrogenesis. *Nature* **526**, 564-568, doi:10.1038/nature15695 (2015).

1252 7 Morizane, R. & Bonventre, J. V. Generation of nephron progenitor cells and kidney
1253 organoids from human pluripotent stem cells. *Nat Protoc* **12**, 195-207,
1254 doi:10.1038/nprot.2016.170 (2017).

1255 8 Nishinakamura, R. Human kidney organoids: progress and remaining challenges. *Nature*
1256 *Reviews Nephrology* **15**, 613-624, doi:10.1038/s41581-019-0176-x (2019).

1257 9 Combes, A. N. *et al.* Correction: Single cell analysis of the developing mouse kidney
1258 provides deeper insight into marker gene expression and ligand-receptor crosstalk
1259 (doi:10.1242/dev.178673). *Development* **146**, doi:10.1242/dev.182162 (2019).

1260 10 Adam, M., Potter, A. S. & Potter, S. S. Psychrophilic proteases dramatically reduce
1261 single-cell RNA-seq artifacts: a molecular atlas of kidney development. *Development*
1262 *(Cambridge, England)* **144**, 3625-3632, doi:10.1242/dev.151142 (2017).

1263 11 Schmidt-Ott, K. M. How to grow a kidney: patient-specific kidney organoids come of
1264 age. *Nephrol Dial Transplant* **32**, 17-23, doi:10.1093/ndt/gfw256 (2017).

1265 12 Menon, R. *et al.* Single-cell analysis of progenitor cell dynamics and lineage
1266 specification in the human fetal kidney. *Development (Cambridge, England)* **145**,
1267 doi:10.1242/dev.164038 (2018).

1268 13 Lindstrom, N. O. *et al.* Progressive Recruitment of Mesenchymal Progenitors Reveals a
1269 Time-Dependent Process of Cell Fate Acquisition in Mouse and Human Nephrogenesis.
1270 *Dev Cell* **45**, 651-660 e654, doi:10.1016/j.devcel.2018.05.010 (2018).

1271 14 Hochane, M. *et al.* Single-cell transcriptomics reveals gene expression dynamics of
1272 human fetal kidney development. *PLOS Biology* **17**, e3000152,
1273 doi:10.1371/journal.pbio.3000152 (2019).

1274 15 Park, J., Liu, C. L., Kim, J. & Susztak, K. Understanding the kidney one cell at a time.
1275 *Kidney Int* **96**, 862-870, doi:10.1016/j.kint.2019.03.035 (2019).

1276 16 Fang, R. *et al.* Fast and Accurate Clustering of Single Cell Epigenomes Reveals Cis-
1277 Regulatory Elements in Rare Cell Types. *bioRxiv*, 615179, doi:10.1101/615179 (2019).

1278 17 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with
1279 Harmony. *Nat Methods* **16**, 1289-1296, doi:10.1038/s41592-019-0619-0 (2019).

1280 18 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137,
1281 doi:10.1186/gb-2008-9-9-r137 (2008).

1282 19 Park, J. *et al.* Single-cell transcriptomics of the mouse kidney reveals potential cellular
1283 targets of kidney disease. *Science* **360**, 758-763, doi:10.1126/science.aar2131 (2018).

1284 20 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-
1285 1902.e1821, doi:10.1016/j.cell.2019.05.031 (2019).

1286 21 Wu, H., Kirita, Y., Donnelly, E. L. & Humphreys, B. D. Advantages of Single-Nucleus
1287 over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell
1288 States Revealed in Fibrosis. *J Am Soc Nephrol* **30**, 23-32, doi:10.1681/ASN.2018090912
1289 (2019).

1290 22 Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes.
1291 *Nucleic Acids Res* **47**, D766-D773, doi:10.1093/nar/gky955 (2019).

1292 23 Zhu, Q. *et al.* Developmental trajectory of pre-hematopoietic stem cell formation from
1293 endothelium. *bioRxiv*, 848846, doi:10.1101/848846 (2019).

1294 24 Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands
1295 of single cells. *Science* **361**, 1380-1385, doi:10.1126/science.aau0730 (2018).

1296 25 O'Brien, L. L. *et al.* Transcriptional regulatory control of mammalian nephron
1297 progenitors revealed by multi-factor cistromic analysis and genetic studies. *PLoS Genet*
1298 **14**, e1007181, doi:10.1371/journal.pgen.1007181 (2018).

1299 26 Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell
1300 Chromatin Accessibility Data. *Mol Cell* **71**, 858-871 e858,
1301 doi:10.1016/j.molcel.2018.06.044 (2018).

1302 27 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime
1303 cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-
1304 589, doi:10.1016/j.molcel.2010.05.004 (2010).

1305 28 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat*
1306 *Methods* **14**, 1083-1086, doi:10.1038/nmeth.4463 (2017).

1307 29 Nittoli, V. *et al.* Characterization of paralogous uncx transcription factor encoding genes
1308 in zebrafish. *Gene X* **2**, 100011, doi:10.1016/j.gene.2019.100011 (2019).

1309 30 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498,
1310 doi:10.1038/s41586-018-0414-6 (2018).

1311 31 Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring
1312 transcription-factor-associated accessibility from single-cell epigenomic data. *Nat*
1313 *Methods* **14**, 975-978, doi:10.1038/nmeth.4401 (2017).

1314 32 Guo, J. K. *et al.* WT1 is a key regulator of podocyte function: reduced expression levels
1315 cause crescentic glomerulonephritis and mesangial sclerosis. *Hum Mol Genet* **11**, 651-
1316 659, doi:10.1093/hmg/11.6.651 (2002).

1317 33 Kann, M. *et al.* WT1 targets Gas1 to maintain nephron progenitor cells by modulating
1318 FGF signals. *Development* **142**, 1254-1266, doi:10.1242/dev.119735 (2015).

1319 34 Nakai, S. *et al.* Crucial roles of Brn1 in distal tubule formation and function in mouse
1320 kidney. *Development* **130**, 4751-4759, doi:10.1242/dev.00666 (2003).

1321 35 Nilsson, D., Heglind, M., Arani, Z. & Enerbäck, S. Foxc2 is essential for podocyte
1322 function. *Physiol Rep* **7**, e14083-e14083, doi:10.14814/phy2.14083 (2019).

1323 36 Takahashi, T. *et al.* Temporally compartmentalized expression of ephrin-B2 during renal
1324 glomerular development. *J Am Soc Nephrol* **12**, 2673-2682 (2001).

1325 37 Cheng, H. T. & Kopan, R. The role of Notch signaling in specification of podocyte and
1326 proximal tubules within the developing mouse kidney. *Kidney Int* **68**, 1951-1952,
1327 doi:10.1111/j.1523-1755.2005.00627.x (2005).

1328 38 Kang, H. M. *et al.* Defective fatty acid oxidation in renal tubular epithelial cells has a key
1329 role in kidney fibrosis development. *Nat Med* **21**, 37-46, doi:10.1038/nm.3762 (2015).

1330 39 Das, A. *et al.* Stromal–epithelial crosstalk regulates kidney progenitor cell differentiation.
1331 *Nature Cell Biology* **15**, 1035-1044, doi:10.1038/ncb2828 (2013).

1332 40 Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal-fetal interface in
1333 humans. *Nature* **563**, 347-353, doi:10.1038/s41586-018-0698-6 (2018).

1334 41 Trueb, B., Amann, R. & Gerber, S. D. Role of FGFRL1 and other FGF signaling proteins
1335 in early kidney development. *Cell Mol Life Sci* **70**, 2505-2518, doi:10.1007/s00018-012-
1336 1189-9 (2013).

1337 42 Costantini, F. & Shakya, R. GDNF/Ret signaling and the development of the kidney.
1338 *Bioessays* **28**, 117-127, doi:10.1002/bies.20357 (2006).

1339 43 Hwang, D. Y. *et al.* Mutations of the SLIT2-ROBO2 pathway genes SLIT2 and SRGAP1
1340 confer risk for congenital anomalies of the kidney and urinary tract. *Hum Genet* **134**, 905-
1341 916, doi:10.1007/s00439-015-1570-5 (2015).

1342 44 Fan, X. *et al.* SLIT2/ROBO2 signaling pathway inhibits nonmuscle myosin IIA activity
1343 and destabilizes kidney podocyte adhesion. *JCI Insight* **1**, e86934,
1344 doi:10.1172/jci.insight.86934 (2016).

1345 45 Sajithlal, G., Zou, D., Silvius, D. & Xu, P. X. Eya 1 acts as a critical regulator for
1346 specifying the metanephric mesenchyme. *Dev Biol* **284**, 323-336,
1347 doi:10.1016/j.ydbio.2005.05.029 (2005).

1348 46 Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses
1349 of a million individuals. *Nat Genet* **51**, 957-972, doi:10.1038/s41588-019-0407-x (2019).

1350 47 Hellwege, J. N. *et al.* Mapping eGFR loci to the renal transcriptome and phenome in the
1351 VA Million Veteran Program. *Nat Commun* **10**, 3842, doi:10.1038/s41467-019-11704-w
1352 (2019).

1353 48 Pattaro, C. *et al.* Genetic associations at 53 loci highlight cell types and biological
1354 pathways relevant for kidney function. *Nat Commun* **7**, 10023,
1355 doi:10.1038/ncomms10023 (2016).

1356 49 Menon, M. C. *et al.* Intronic locus determines SHROOM3 expression and potentiates
1357 renal allograft fibrosis. *J Clin Invest* **125**, 208-221, doi:10.1172/JCI76902 (2015).

1358 50 Khalili, H. *et al.* Developmental Origins for Kidney Disease Due to Shroom3 Deficiency.
1359 *J Am Soc Nephrol* **27**, 2965-2973, doi:10.1681/ASN.2015060621 (2016).

1360 51 Qiu, C. *et al.* Renal compartment-specific genetic variation analyses identify new
1361 pathways in chronic kidney disease. *Nat Med* **24**, 1721-1731, doi:10.1038/s41591-018-
1362 0194-4 (2018).

1363 52 Ransick, A. *et al.* Single-Cell Profiling Reveals Sex, Lineage, and Regional Diversity in
1364 the Mouse Kidney. *Developmental Cell* **51**, 399-413.e397,
1365 doi:10.1016/j.devcel.2019.10.005 (2019).

1366 53 Kobayashi, A. *et al.* Six2 defines and regulates a multipotent self-renewing nephron
1367 progenitor population throughout mammalian kidney development. *Cell Stem Cell* **3**,
1368 169-181, doi:10.1016/j.stem.2008.05.020 (2008).

1369 54 Subramanian, A. *et al.* Single cell census of human kidney organoids shows
1370 reproducibility and diminished off-target cells after transplantation. *Nat Commun* **10**,
1371 5462, doi:10.1038/s41467-019-13382-0 (2019).

1372 55 Han, S. H. *et al.* PGC-1alpha Protects from Notch-Induced Kidney Fibrosis
1373 Development. *J Am Soc Nephrol* **28**, 3312-3322, doi:10.1681/ASN.2017020130 (2017).

1374 56 Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate
1375 causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491-504,
1376 doi:10.1038/s41576-018-0016-z (2018).

1377 57 Cusanovich, D. A. *et al.* The cis-regulatory dynamics of embryonic development at
1378 single-cell resolution. *Nature* **555**, 538-542, doi:10.1038/nature25981 (2018).

1379 58 Calderon, D. *et al.* Inferring Relevant Cell Types for Complex Traits by Using Single-
1380 Cell Gene Expression. *Am J Hum Genet* **101**, 686-699, doi:10.1016/j.ajhg.2017.09.009
1381 (2017).

1382 59 Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific
1383 molecular pathways and circuits in autism. *Cell* **155**, 1008-1021,
1384 doi:10.1016/j.cell.2013.10.031 (2013).

1385 60 Zhu, C. *et al.* An ultra high-throughput method for single-cell joint analysis of open
1386 chromatin and transcriptome. *Nature Structural & Molecular Biology* **26**, 1063-1070,
1387 doi:10.1038/s41594-019-0323-x (2019).

1388 61 Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and
1389 chromatin accessibility in the same cell. *Nature Biotechnology* **37**, 1452-1457,
1390 doi:10.1038/s41587-019-0290-0 (2019).

1391 62 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**,
1392 1260419, doi:10.1126/science.1260419 (2015).

1393 63 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for
1394 Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 21-
1395 21 29 29, doi:10.1002/0471142727.mb2129s109 (2015).

1396 64 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables
1397 interrogation of frozen tissues. *Nat Methods* **14**, 959-962, doi:10.1038/nmeth.4396
1398 (2017).

1399 65 Chen, H. *et al.* Assessment of computational methods for the analysis of single-cell
1400 ATAC-seq data. *Genome Biol* **20**, 241, doi:10.1186/s13059-019-1854-5 (2019).

1401 66 O'Brien, L. L. *et al.* Differential regulation of mouse and human nephron progenitors by
1402 the Six family of transcriptional regulators. *Development* **143**, 595-608,
1403 doi:10.1242/dev.127175 (2016).

1404 67 Hon, G. C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA
1405 methylation maps from adult mouse tissues. *Nature Genetics* **45**, 1198-1206,
1406 doi:10.1038/ng.2746 (2013).

1407 68 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory
1408 regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).

1409 69 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data
1410 analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).

1411 70 Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS*
1412 *Computational Biology* **9**, e1003118, doi:10.1371/journal.pcbi.1003118 (2013).

1413 71 Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis.
1414 *Nature* **566**, 496-502, doi:10.1038/s41586-019-0969-x (2019).

1415 72 Zhao, J., Hu, Z. Z., Zheng, X. G. & Ng, S. W. [2-(Tetra-zol-1-yl)acetato-kappaO]tris-(tri-
1416 phenyl-phosphine-kappaP)silver(I) mono-hydrate. *Acta Crystallogr Sect E Struct Rep*
1417 *Online* **65**, m1601, doi:10.1107/S1600536809048144 (2009).

1418 73 Aoki, R. *et al.* Foxl1-expressing mesenchymal cells constitute the intestinal stem cell
1419 niche. *Cell Mol Gastroenterol Hepatol* **2**, 175-188, doi:10.1016/j.jcmgh.2015.12.004
1420 (2016).
1421

Figure 1

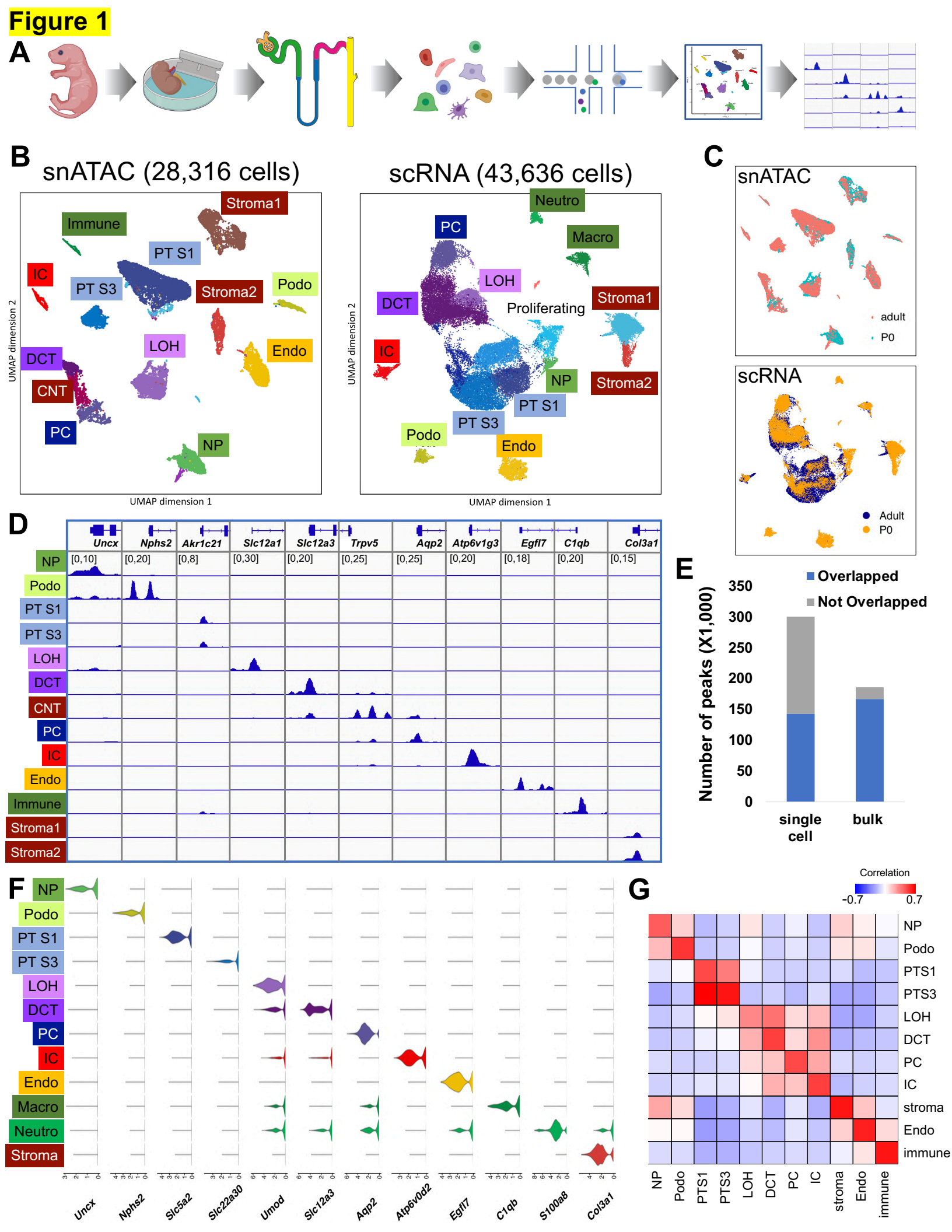

A
B
snATAC (28,316 cells)
scRNA (43,636 cells)
Stroma1
Immune
PT S1
IC
PT S3
Stroma2
Podo
LOH
Endo
DCT
CNT
PC
NP
UMAP dimension 1
UMAP dimension 2
Neutro
Macro
Proliferating
C
snATAC
adult
P0
scRNA
Adult
P0
D
Uncx
Nphs2
Akr1c21
Slc12a1
Slc12a3
Trpv5
Aqp2
Atp6v1g3
Egfl7
C1qb
Col3a1
[0,10]
[0,20]
[0,8]
[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

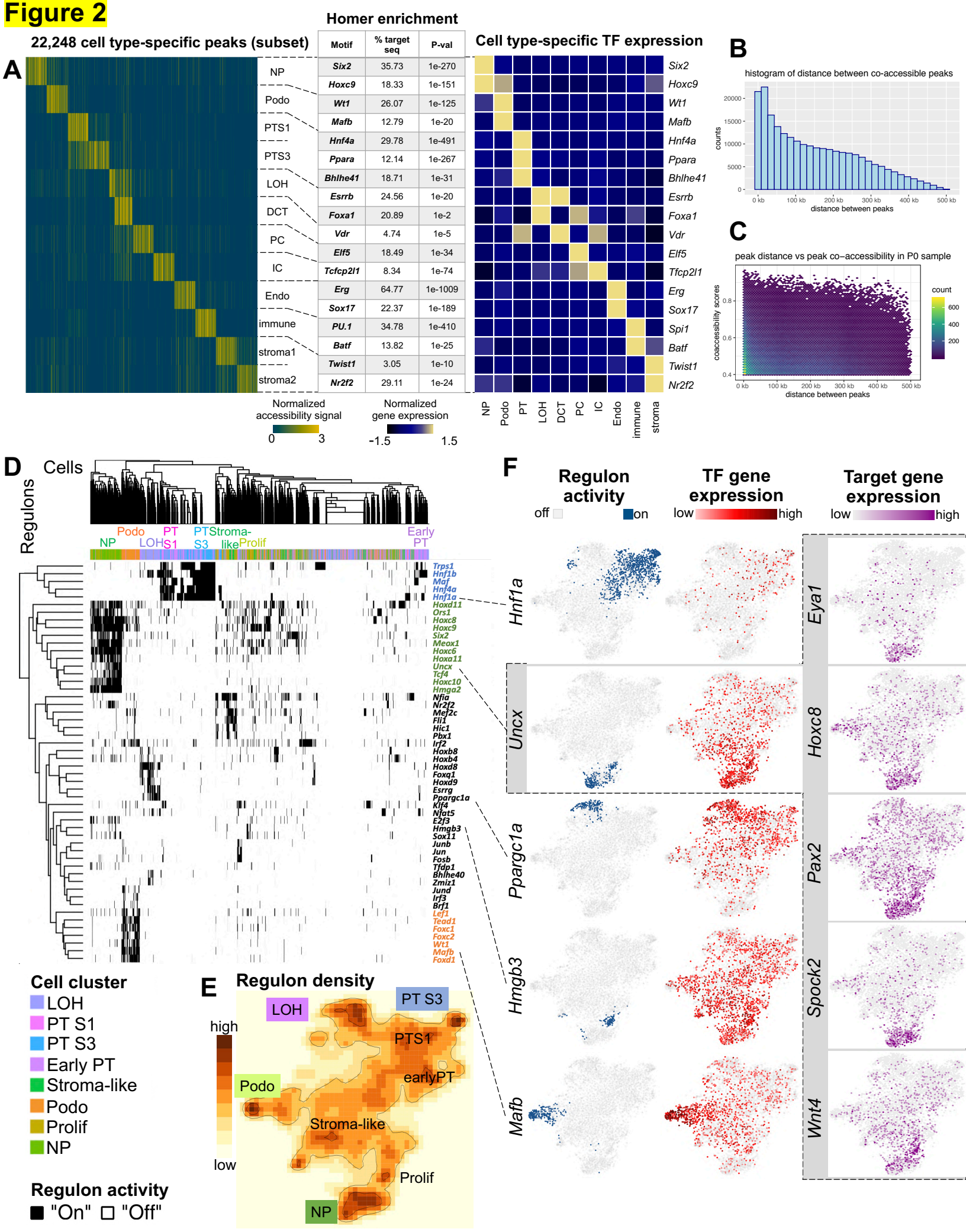

A
22,248 cell type-specific peaks (subset)
NP
Podo
PTS1
PTS3
LOH
DCT
PC
IC
Endo
immune
stroma1
stroma2
Homer enrichment
Motif | % target seq | P-val
Six2 | 35.73 | 1e-270
Hoxc9 | 18.33 | 1e-151
Wt1 | 26.07 | 1e-125
Mafb | 12.79 | 1e-20
Hnf4a | 29.78 | 1e-491
Ppara | 12.14 | 1e-267
Bhlhe41 | 18.71 | 1e-31
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
Cell type-specific TF expression
Six2
Hoxc9
Wt1
Mafb
Hnf4a
Ppara
Bhlhe41
Esrrb
Foxa1
Vdr
Elf5
Tfcp2l1
Erg
Sox17
Spi1
Batf
Twist1
Nr2f2
NP
Podo
PT
LOH
DCT
PC
IC
Endo
immune
stroma
Normalized accessibility signal
0
3
Normalized gene expression
-1.5
1.5
B
histogram of distance between co-accessible peaks
counts
distance between peaks
C
peak distance vs peak co-accessibility in P0 sample
coaccessibility scores
distance between peaks
count
D
Cells
Regulons
NP
Podo
LOH
PT S1
PT S3
Stroma-like
Prolif
Early PT
Trps1
Hnf1b
Maf
Hnf4a
Hnf1a
Hoxd11
Ors1
Hoxc8
Hoxc9
Six2
Meox1
Hoxc6
Hoxa11
Uncx
Tcf4
Hoxc10
Hmga2
Nfia
Nr2f2
Mef2c
Fli1
Hic1
Pbx1
Irf2
Hoxb8
Hoxb4
Hoxd8
Foxq1
Hoxd9
Esrrg
Ppargc1a
Klf4
Nfat5
E2f3
Hmgb3
Sox11
Junb
Jun
Fosb
Tfdp1
Bhlhe40
Zmiz1
Jund
Irf3
Brf1
Lef1
Tead1
Foxc1
Foxc2
Wt1
Mafb
Foxd1
Cell cluster
LOH
PT S1
PT S3
Early PT
Stroma-like
Podo
Prolif
NP
Regulon activity
"On"
"Off"
E
Regulon density
high
low
LOH
PT S3
PTS1
earlyPT
Podo
Stroma-like
Prolif
NP
F
Regulon activity
off
on
TF gene expression
low
high
Target gene expression
low
high
Hnf1a
Uncx
Ppargc1a
Hmgb3
Mafb
Eya1
Hoxc8
Pax2
Spock2
Wnt4

Figure 3

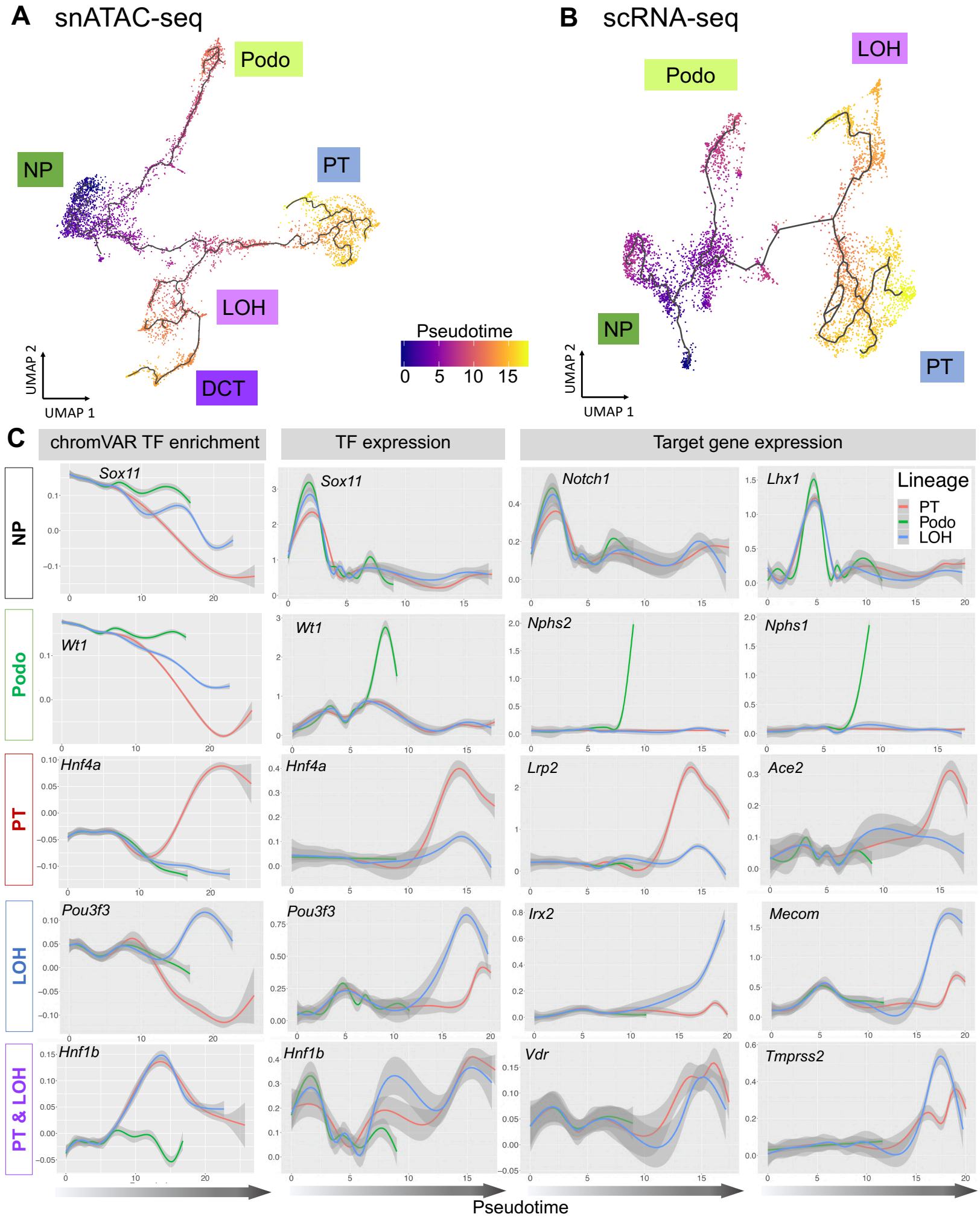

A snATAC-seq
Podo
NP
PT
LOH
DCT
UMAP 2
UMAP 1
Pseudotime
0 5 10 15
B scRNA-seq
LOH
Podo
NP
PT
UMAP 2
UMAP 1
C
chromVAR TF enrichment
TF expression
Target gene expression
NP
Podo
PT
LOH
PT & LOH
Sox11
Notch1
Lhx1
Lineage
PT
Podo
LOH
Wt1
Nphs2
Nphs1
Hnf4a
Lrp2
Ace2
Pou3f3
Irx2
Mecom
Hnf1b
Vdr
Tmprss2
Pseudotime

Figure 4

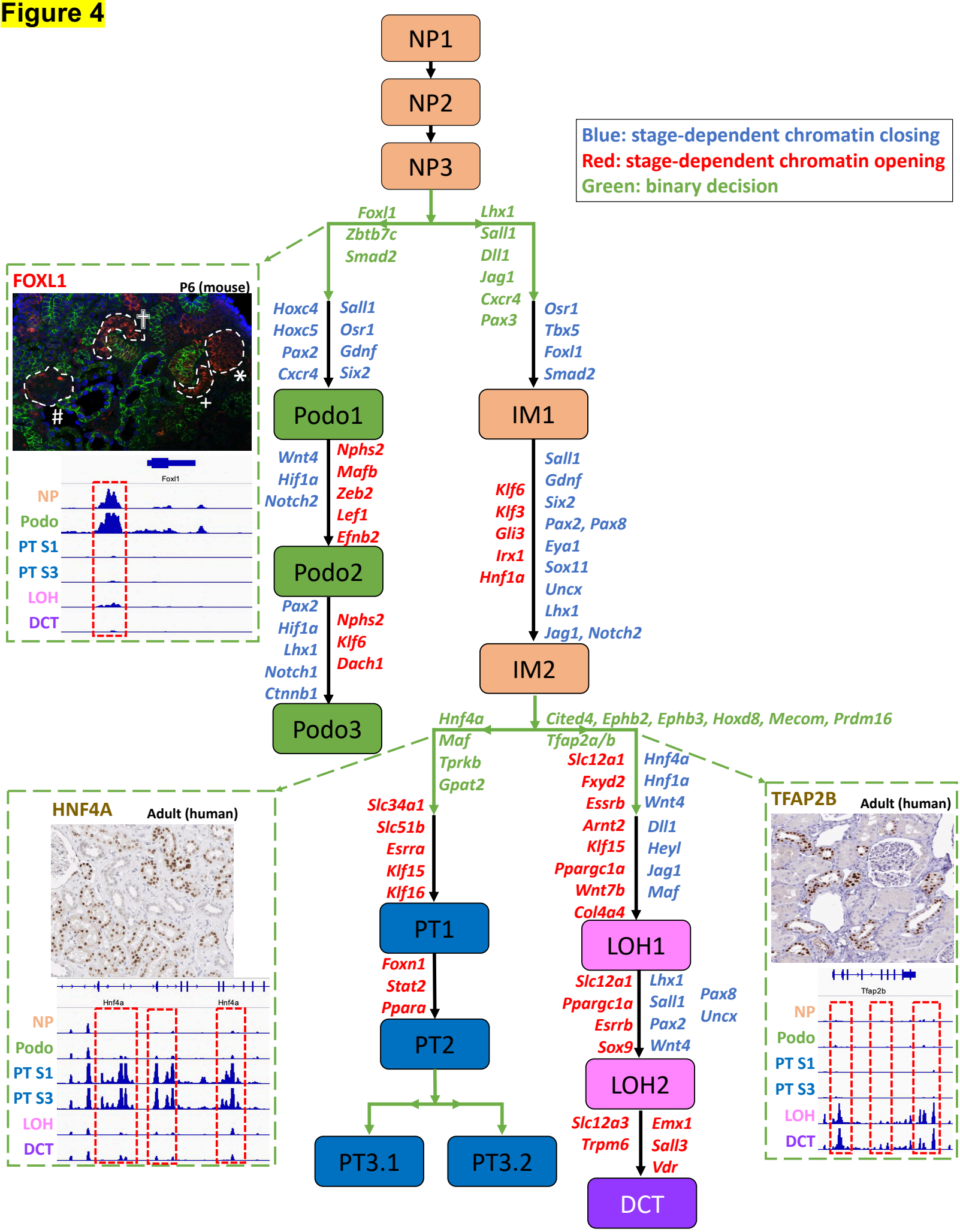

NP1
NP2
NP3
Blue: stage-dependent chromatin closing
Red: stage-dependent chromatin opening
Green: binary decision
Foxl1
Zbtb7c
Smad2
Lhx1
Sall1
Dll1
Jag1
Cxcr4
Pax3
FOXL1
P6 (mouse)
Foxl1
NP
Podo
PT S1
PT S3
LOH
DCT
Hoxc4
Hoxc5
Pax2
Cxcr4
Sall1
Osr1
Gdnf
Six2
Podo1
Wnt4
Hif1a
Notch2
Nphs2
Mafb
Zeb2
Lef1
Efnb2
Podo2
Pax2
Hif1a
Lhx1
Notch1
Ctnnb1
Nphs2
Klf6
Dach1
Podo3
Osr1
Tbx5
Foxl1
Smad2
IM1
Klf6
Klf3
Gli3
Irx1
Hnf1a
Sall1
Gdnf
Six2
Pax2, Pax8
Eya1
Sox11
Uncx
Lhx1
Jag1, Notch2
IM2
Hnf4a
Maf
Tprkb
Gpat2
Cited4, Ephb2, Ephb3, Hoxd8, Mecom, Prdm16
Tfap2a/b
Slc12a1
Fxyd2
Essrb
Arnt2
Klf15
Ppargc1a
Wnt7b
Col4a4
Hnf4a
Hnf1a
Wnt4
Dll1
Heyl
Jag1
Maf
HNF4A
Adult (human)
Hnf4a
Hnf4a
NP
Podo
PT S1
PT S3
LOH
DCT
Slc34a1
Slc51b
Esrra
Klf15
Klf16
PT1
Foxn1
Stat2
Ppara
PT2
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

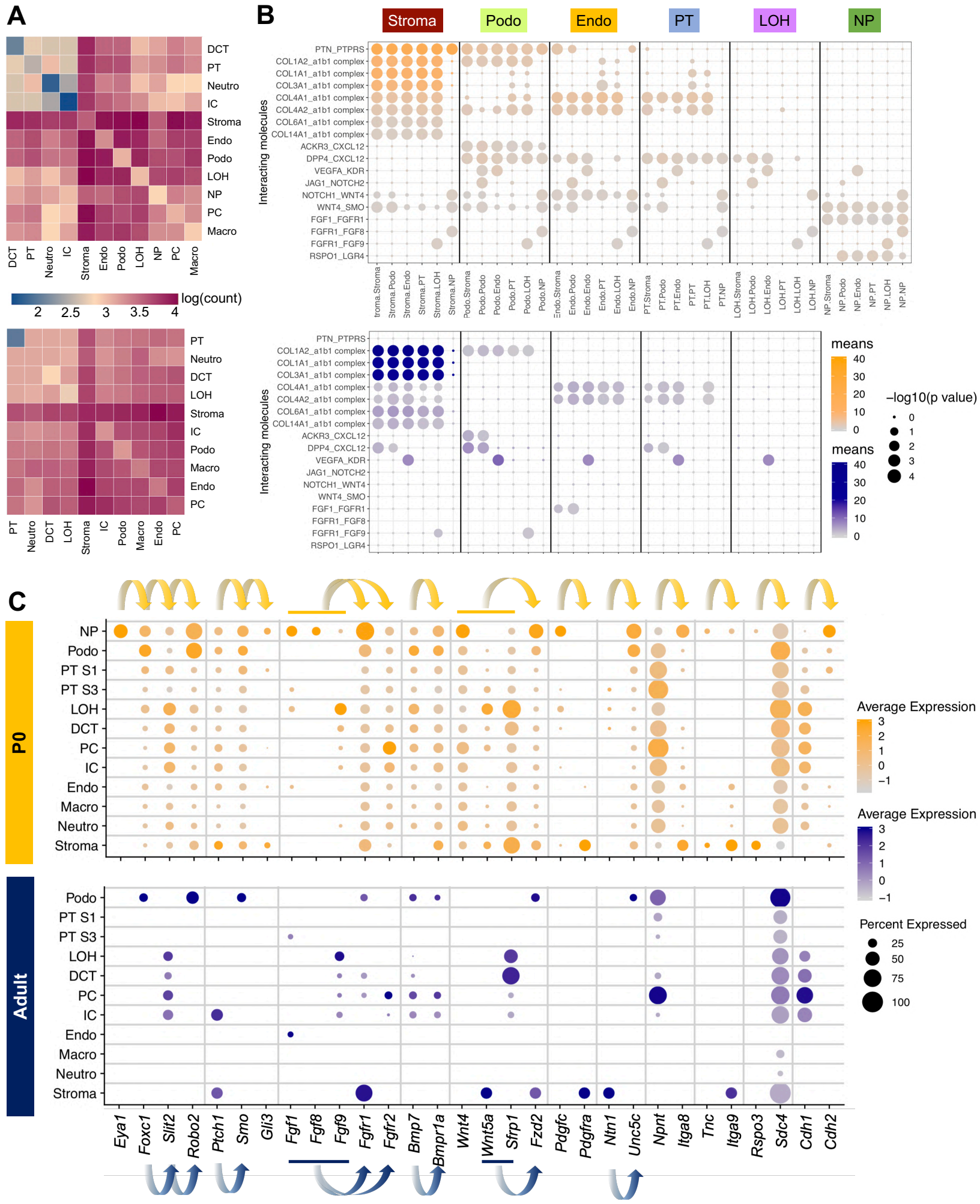
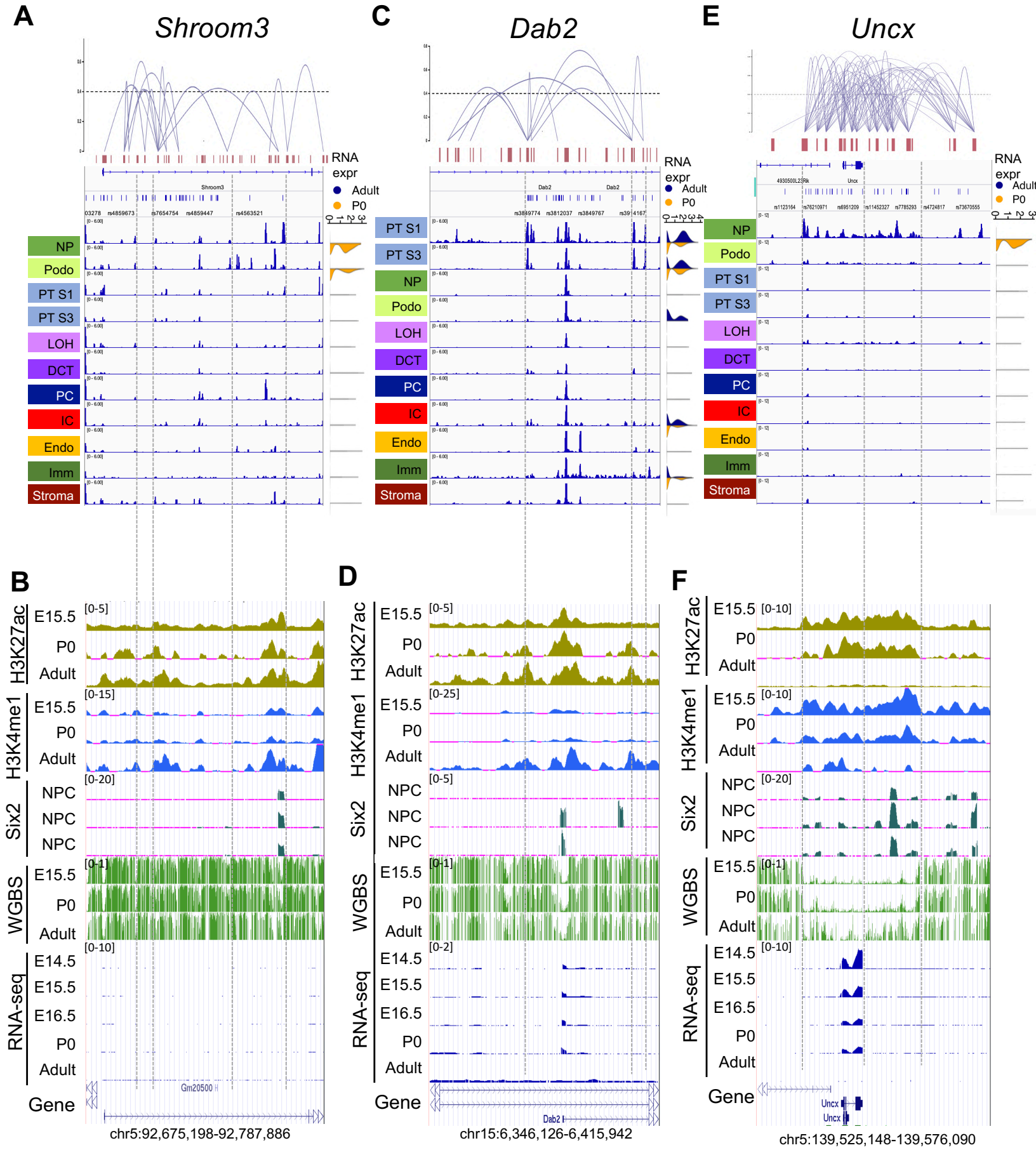

A
DCT
PT
Neutro
IC
Stroma
Endo
Podo
LOH
NP
PC
Macro
log(count)
2
2.5
3
3.5
4
B
Stroma
Podo
Endo
PT
LOH
NP
Interacting molecules
PTN_PTPRS
COL1A2_a1b1 complex
COL1A1_a1b1 complex
COL3A1_a1b1 complex
COL4A1_a1b1 complex
COL4A2_a1b1 complex
COL6A1_a1b1 complex
COL14A1_a1b1 complex
ACKR3_CXCL12
DPP4_CXCL12
VEGFA_KDR
JAG1_NOTCH2
NOTCH1_WNT4
WNT4_SMO
FGF1_FGFR1
FGFR1_FGF8
FGFR1_FGF9
RSPO1_LGR4
means
–log10(p value)
C
P0
Adult
NP
Podo
PT S1
PT S3
LOH
DCT
PC
IC
Endo
Macro
Neutro
Stroma
Average Expression
Percent Expressed
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