Single-cell chromatin and gene-regulatory dynamics of mouse nephron progenitors

Sylvia Hilliard,[1] **Giovane Tortelote,**[1] **Hongbing Liu,**[1] **Chao-Hui Chen,**[1] **and Samir S. El-Dahr**[1*]

[1]Section of Pediatric Nephrology, Department of Pediatrics, Tulane University School of Medicine, New Orleans, LA 70112, USA

[*]Correspondence: seldahr@tulane.edu

Running title: Multiome analysis of nephron progenitors

2

Significance Statement

- Nephron progenitor cells (NPCs) are a multipotent population giving rise to all cell types of the nephron. At any given time, the NPC's choice to self-renew or differentiate is determined not only by its transcription factor (TF) repertoire but also by the genome accessibility of the cognate cis-regulatory elements.

- Using single-cell analysis, we demonstrate the heterogeneity of NPCs at the epigenetic level and observe dynamic and cell type-specific changes in chromatin accessibility. Fate-determining TFs harbor domains rich in interactive chromatin that are established prior to gene activation.

- These findings illustrate the importance of chromatin-based mechanisms in the regulation of nephrogenesis and may have implications for nephron regeneration and repair.

STRUCTURED ABSTRACT

Background: Cis-regulatory elements (CREs), such as enhancers and promoters, and their cognate transcription factors play a central role in cell fate specification. Bulk analysis of CREs has provided insights into gene regulation in nephron progenitor cells (NPCs). However, the cellular resolution required to unravel the dynamic changes in regulatory elements associated with cell fate choices remains to be defined.

Methods: We integrated single-cell chromatin accessibility (scATAC-seq) and gene expression (scRNA-seq) in embryonic E16.5 (self-renewing) and postnatal P2 (primed) mouse Six2^{GFP} NPCs. This analysis revealed NPC diversity and identified candidate CREs. To validate these findings and gain additional insights into more differentiated cell types, we performed a multiome analysis of E16.5 and P2 kidneys.

Results: CRE accessibility recovered the diverse states of NPCs and precursors of differentiated cells. Single-cell types such as podocytes, proximal and distal precursors are marked by differentially accessible CREs. Domains of regulatory chromatin as defined by rich CRE-gene associations identified NPC fate-determining transcription factors (TF). Likewise, key TF expression correlates well with its regulon activity. Young NPCs exhibited enrichment in accessible motifs for bHLH, homeobox, and Forkhead TFs, while older NPCs were enriched in AP-1, HNF1, and HNF4 motif activity. A subset of Forkhead factors exhibiting high chromatin activity in podocyte precursors.

Conclusion: Defining the regulatory landscape of nephrogenesis at single-cell resolution informs the basic mechanisms of nephrogenesis and provides a foundation for future studies in disease states characterized by abnormal nephrogenesis.

3

INTRODUCTION

In the developing kidney, nephron progenitor cells (NPCs) reside in a niche called the cap mesenchyme, a crescent-shaped cellular compartment surrounding the adjacent ureteric branch tip. Based on molecular marker analysis, the cap mesenchyme is divided into a self-renewing Cited1$^+$/Six2$^+$ compartment, a transit Cited1$^-$/Six2$^+$ compartment, and a late Six2^{low}/Wnt4$^+$ compartment called the pretubular aggregate, the precursor of the renal vesicle, and the earliest epithelial precursor of the nephron (1, 2). Single-cell transcriptomics of human and mouse kidneys refined the signatures of NPC states and delineated the developmental trajectories and the potential transcriptional regulatory networks associated with cell fate decisions (3-7). The NPC's choice to self-renew or differentiate and gain a new identity is determined not only by its transcription factor (TF) repertoire but also by the accessibility of the cognate cis-regulatory elements (CREs). Accordingly, to fully characterize the NPC states and fates, it is necessary to have an integrated view of chromatin and gene expression states across cell types, developmental trajectory, and life span.

Chromatin accessibility is a dynamic process that drives tissue development by permitting or restricting access of TFs to cognate CREs. ATAC-seq (Assay for Transposase Accessible Chromatin) has become the assay of choice to decipher open (accessible) chromatin domains composed of promoters and enhancers (8). We previously reported using bulk ATAC-seq in native embryonic and neonatal mouse Six2^{GFP} NPCs that NPCs undergo age-dependent changes in chromatin accessibility at active enhancers of renewal and differentiation genes (9). These differential dynamic

chromatin states were also observed in expanding NPCs ex vivo, suggesting that young and old NPCs are intrinsically different at the epigenetic level (9). Changes in chromatin accessibility to TCF/LEF factors at pro-differentiation enhancers were also shown to mediate gene expression in response to ß-catenin activation (10), underscoring the importance of concerted actions of chromatin and transcription factors in gene regulation. However, given the dynamic changes in NPC states that occur during differentiation, it is difficult to accurately assign a chromatin state to a specific cell population using bulk analysis. In a recent study to address this issue (11) the investigators performed snRNA/snATAC-seq in newborn and adult mouse kidneys revealing key developmental regulators of NPC commitment decisions to the podocyte, proximal and distal fates and linking single nucleotide variants associated with human kidney disease to regulatory elements in key developmental genes. In the present study, we integrated single-cell chromatin accessibility and gene expression profiles obtained from embryonic and neonatal mouse NPCs and subsequently validated our singleomes using single-cell profiles derived from the same cell (multiome). These paired maps recovered a class of genes, frequently lineage-determining transcription factors, enriched in putative enhancers whose accessibility was strongly linked to gene expression.

MATERIALS AND METHODS

Animal protocols utilized in this study were approved by and in strict adherence to guidelines established by the Tulane University Institutional Animal Care and Use Committee.

Mice

Six2GFP NPCs were isolated from kidneys of *embryonic day* E16.5 (n=3) and *postnatal day* P2 (n=4) *Six2GFPCre* (Six2TGC) mice bred on a mixed genetic background by fluorescent-activated cell sorting (FACS), as described (9). For multiomics, cells were isolated from kidneys of E16.5 and P2 C57Bl/6 mice.

Single-cell ATAC seq data generation from E16.5 and P2 Six2GFP kidneys

Nuclei were prepared from freshly explanted kidneys in accordance to the protocol outlined in CG000169 | RevB (10X Genomics). Approximately 100,000 dissociated cells were washed in ice-cold 1X PBS supplemented with 0.04% BSA. Cells were spun down at 500g for 5min at 4°C. The cell pellet was re-suspended by gentle pipetting in 45ul Lysis buffer 10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl2 0.1% NP-40, 0.1% Tween-20, 0.01% Digitonin and 1% BSA. Lysis was allowed to proceed on ice for 5 minutes after which 50ul of ice-cold wash buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% Tween-20 and 1% BSA) was added before spinning down again and re-suspending in 45 μL 1X Nuclei Dilution Buffer (10x Genomics). After the last centrifugation step, 7ul of 1X diluted Nuclei Buffer was added to the nuclei pellet. Two microliters of the nuclei preparation plus 8ul of diluted Nuclei buffer was stained with 10ul of AO/PI staining

6

solution (Nexcelom, MA). Both concentration and viability of the sample was read using the Cellometer Auto 2000 Cell Viability Counter (Nexcelom, MA). The remaining 5ul of nuclei was adjusted for a targeted nuclei recovery of 5000 before proceeding to the Chromium Single Cell ATAC Reagents and Gel Bead Kit protocol CG000168 (PN1000111,10x Genomics).

Paired-end dual index libraries were sequenced on the Illumina NextSeq 550 sequencer. Raw data was extracted and processed using the Cell Ranger pipeline (v1.2.0, 10X Genomics) and the mm10 Mus musculus reference assembly.

Multiome data generation

Cells were isolated from mouse kidneys explanted at E16.5 and P2. The kidneys were dissociated by incubation in Accutase (AT104, Innovative Cell Technologies, Inc.) with slow rotation at 37°C for up to an hour with intermittent gentle pipetting. The reaction was stopped with FBS and the cells were spun down at 500rpm for 5min. The cell pellet was resuspended in 1XPBS supplemented with 10mM EDTA and 2%FBS. Following resuspension, the cells were passed through a 40um cell strainer to obtain a single cell suspension (BD Falcon, catalog# 352235).

Nuclei isolation was achieved with 800,000 cells per age group using the protocol CG000366 RevB (10XGenomics) established for mouse tissues. To achieve cell lysis, we incubated the cells in 0.1X Lysis Buffer for 8min on ice. After the wash and centrifugation steps, the nuclei pellet was resuspended in cold 1X Nuclei Buffer. Approximately 6000 nuclei per sample were used to prepare the libraries in accordance with the steps detailed in the Chromium Next GEM Single Cell Multiome ATAC + Gene

Expression protocol (CG000338, RevD) and the 10X Reagent Kit PN-1000284. Briefly, nuclei suspensions were incubated in a Transposition Mix to transpose and barcode accessible DNA fragments. The transposed nuclei were partitioned into nanoliter-scale Gel Bead-In EMulsion (GEMs). The barcoded transposed DNA and barcoded full-length cDNA from poly-adenylated mRNA were generated and amplified by PCR to obtain sufficient complexity for both ATAC and gene expression library constructions. P7 and a sample index were added to transposed DNA during ATAC library construction via PCR. Barcoded cDNA enzymatic fragmentation, end-repair, A-tailing, and adaptor ligation were followed by PCR amplification. Both the ATAC and 3′- Gene Expression libraries generated contained standard Illumina P5 and P7 paired-end constructs. Library quality controls were performed with an Agilent High Sensitivity DNA kit on the Agilent 2100 Bioanalyzer together with quantitation on a Qubit 2.0 fluorometer. Libraries were sequenced separately with individual parameter settings. Pooled libraries at a final concentration of 650pM were sequenced with paired end dual index configuration on an Illumina NextSeq 550 using Illumina High-Output 150 cycle kits (Cat No.20024907). Cell Ranger-arc version 1.0.1 software (10X Genomics) was used to process raw sequencing data for further downstream analyses.

Analysis using ArchR package

Fragment files generated with cell-ranger -arc version 1.0.1 pipeline (10X Genomics) were used as input for ArchR (Analysis of Regulatory Chromatin in R) (archproject.com)(12). The input fragment file was processed as chunks per chromosome and stored in HDF5 format (hierarchical data format version 5) allowing

8

rapid access and efficient read and write functions during analysis. The QC steps involved the removal of all low-quality cells and predicted heterotypic doublets from our analysis. All cells with a TSS enrichment score greater than or equal to 4 and with ATAC fragment counts greater than 1000 but less than 100,000 per cell were retained in our analysis.

A chromatin accessibility matrix was generated using a genome-wide, non-overlapping, tile matrix of 500bp bins. The gene score algorithm allowed the prediction of gene activity/expression based on ATAC-seq data. It measures accessibility at promoter plus gene body and distal regulatory elements at weighted distances from the gene TSS and TTS while accounting for gene size differences and neighboring gene boundaries. We used a default of 100kb window flanking the gene.

To obtain a low dimensional representation of single-cell ATAC datasets in terms of principal components and UMAP coordinates, we applied an iterative latent semantic indexing approach (13, 14). In brief, Dimensionality reduction was achieved by TF-IDF normalization and Singular Value Decomposition (SVD)/Latent Semantic Indexing (LSI) through user-defined iterations. The first iteration uses top accessible features while the second uses top variable features. Clustering is achieved by Shared Nearest Neighbor algorithm implemented by Seurat followed by 2D embedding in the UMAP (Uniform Manifold Approximation and Projection) space for visualization. Peak analysis was preceded by calling pseudo-bulk replicates of each cluster using default parameters. MACS2 was used for peak calling and the generation of a non-overlapping union peak matrix. ChromVAR (15) (v.1.6) was used to obtain TF accessibility profiles using position weight matrices from the CisBP mouse_pwms_v2 database (16).

Co-accessibility detection in ArchR follows an approach introduced by Cicero (17) by creating low-overlapping aggregates of single cells. ArchR uses optimized methods for calculating significantly correlated peak-peak as well as cis-regulatory peak-to-gene associations (correlation cutoff = 0.45, resolution =1000). The primary difference between peak-to-gene links and peak-peak co-accessibility is that co-accessibility correlates accessibility between two peaks (ATAC-seq only analysis), while peak-to-gene linkage leverages integrates scRNA-seq data to correlate peak accessibility and gene expression which may be a more pertinent approach to identifying gene regulatory interactions.

We applied the GetAssayData function (Seuart 4.0 Package) to recover the log-normalized information stored at the data slot of the RNA assay from the E16.5 Seurat object. This information was used to construct the expression matrix used for regulatory network analysis. Next, to identify TFs and characterize cell states, we employed cis-regulatory analysis using the R package SCENIC v1.1.1 (Aibar et al., 2017), which infers the gene regulatory network based on co-expression and DNA motif analysis. The network activity is then analyzed in each cell to identify recurrent cellular states. We used GENIE3 to identify TFs were identified and compiled into modules (regulons), which were subsequently subjected to cis-regulatory motif analysis using RcisTarget mm9 files with two gene-motif rankings: 10 kb around the TSS and 500 bp upstream. Regulon activity in every cell was then scored using AUCell. Finally, binarized regulon activity was projected onto Seurat-create UMAP clustering.

RNAscope in situ hybridization. RNAscope was applied to formalin-fixed, paraffin-embedded tissue sections from E16.5 kidneys using the Multiplex Fluorescent Assay

(Advanced Cell Diagnostics, Cat# 323100) following the manufacturer's recommendations. The list of probes used are as follows: Foxl1 (Cat# 407401), Foxp1 ((Cat#485221), Foxn3 (Cat# 586011), Nabp1 (Cat#1049221), Mcmdc2 (Cat#477971).

RESULTS AND DISCUSSION

CRE accessibility reveals NPC precursor states and novel cellular markers

To elucidate the cellular heterogeneity of nephron progenitors, we created gene-regulatory datasets using the Chromium platform (10x Genomics) to generate single-cell ATAC-seq (scATAC) and single-cell RNA-seq (scRNA) libraries from fluorescence-activated sorted Six2^{GFP} cells freshly harvested from E16.5 and P2 kidneys of transgenic Six2-GFP-Cre (Six2^{TGC}) mice. We reasoned that while embryonic NPCs are engaged in self-renewal, neonatal NPCs are primed for differentiation nearing the time of cessation of nephrogenesis (P3-P4 in mice). Overall, we obtained 13,506 (E16.5) and 7,232 (P2) single-cell epigenomes and 10,544 (E16.5) and 7,250 (P2) single-cell transcriptomes after quality control and filtering (**Figure S1, A,B and Figure S2 A,B**). After data processing using the Cell Ranger pipeline (10x Genomics), we employed ArchR (12) for downstream pipeline analysis, as described in the Methods section.

To reveal global similarities and differences between individual cells, we performed dimension reduction using uniform manifold approximation and projection (UMAP) and clustering. We employed the latent semantic indexing (LSI) approach for scATAC to obtain a low-dimensional embedding, cell clustering and consensus sets of 753614 (E16.5) and 309584 accessible peaks representing potential CREs. We next called clusters in both scATAC and scRNA (GSE124804) datasets and annotated these clusters using gene activity scores (a metric defined by the aggregate local chromatin accessibility of genes) [**Methods**] and gene expression (**Figure 1 A, B, E, F**). Integrating the derived gene activity scores with gene expression levels using canonical correlation analyses

12

(CCA) to match cells from one modality to their nearest neighbors in the other (**Figure 1 C, G**) showed that the major NPC states were recovered at both E16.5 and P2. Thus, chromatin accessibility is comparable to gene expression in predicting the major cell states of NPCs.

We examined the gene scores matrix (derived from chromatin accessibility) to determine whether we can identify precursor cell states in the embryonic stage NPC. The results revealed that chromatin accessibility can identify precursor states of podocyte, proximal and distal fates (**Figure 1 D**). For example, Cluster 1, which contains markers of cell cycle genes, also features the distal tubule fate marker Hoxd8, whereas clusters 3 and 5 feature markers of proximal progenitors Osr2 and Cdh6. The Clusters 2 and 4, representing early nephron progenitors, feature markers of podocytes. Similarly, progenitor cell states were also identified at P2 based on chromatin accessibility gene scores (**Figure 1 H**), although by this advanced stage of maturation scATAC-seq and scRNA-seq were equivalent in identifying precursor states. Collectively, these findings demonstrate that single-cell chromatin accessibility offers a sensitive means to recover precursor cell states in fetal NPCs.

scATAC-seq also allowed the detection of novel markers, Mcmdc2 (Minichromosone Maintenance Domain Containing-2) and Nabp1(Nucleic acid-binding protein-1), showing selective chromatin accessibility in proximal or distal cell precursors, respectively (**Figure S3 A,C**). Using RNAScope in situ hybridization, we confirmed the RNA expression of *Mcmdc2* in proximal (LTL-positive) tubules, whereas *Nabp1* was expressed in distal (LTL-negative) segments (**Figure S3 B, D**). Interestingly, both genes are involved in DNA repair pathways.

We used the R package Cicero (17) to predict cis-regulatory chromatin interactions for individual cell types. Cis co-accessibility networks (CCAN) are families of chromatin regions that co-vary in their accessibility and can be used to predict chromatin interactions. Within NPCs at E16.5 and P2, most Cicero connections were either within a promoter region or between a promoter and another location like introns and intergenic regions (**Figure S4 A-C**). Overall, there were twice as many CCANs in P2 than E16.5 NPCs reflecting increasing complexity of transcriptional regulation with maturation.

CRE-Gene associations identify potential regulatory chromatin

Because co-variations in chromatin co-accessibility generally have only modest correlations with gene expression, we linked distal peaks to genes in *cis*, based on co-variation in chromatin accessibility and gene expression across all cells to identify CRE-gene pairs during NPC differentiation. Ma and colleagues (18) called chromatin domains enriched in more than 10 CRE-gene associations Domains of Regulatory Chromatin (DORCs). DORCs identified lineage-determining genes in the hair follicle that are accessible prior to onset of gene expression. Trevino and colleagues (BioxRv 2021, pre-print) referred to genes enriched in more than 10 peak-gene links as Genes with Predictive Chromatin (GPCs); GPCs identified lineage-determining factors during neural corticogenesis. In the present study, we applied similar principles by generating pseudobulk aggregates of matched ATAC/RNA annotations across NPCs linking gene-distal CRE accessibility to gene expression. We identified 7039 and 21788 CRE-gene pairs that represent potential regulatory interactions in E16.5 and P2 Six2^{GFP} datasets, respectively (**Figure 2 A, B**). CRE-gene links correlated strongly with cell type-specific

expression of TFs (**Figure S5**). Annotation using the Genomic Regions Enrichment Annotations Tool (GREAT-R) (19) recovered biological processes relating to basic aspects of nephrogenesis ranging from chromatin organization to renal vesicle formation and cell-cell interactions. (**Figure 2 A, B**).

To determine whether CRE-gene associations have a predictive value in identifying key developmental regulators, we ranked genes across all NPCs based on the number of peak-gene associations and exceeding an inflection point ("elbow") (18). This analysis yielded 425 (E16.5) and 412 (P2) genes having more than 4 (E16.5) or 10 CRE-gene associations (P2) (**Figure 2 C, D**). These enriched domains of interactive chromatin identified lineage- and fate-determining genes such as *Wnt4*, *Hox*, *Hnf1, Six2, Emx2, Foxc1, Pou3f3* and *Uncx*, among others (**Figure 2 C, D**). We next wished to determine whether genes enriched in interactive domains are concordantly rich in the active enhancer mark H3K27Ac. For this purpose, we leveraged our ChIP-seq data performed on E13.5 and P0 NPCs (9). Notably, 15-25% of genes have interactive chromatin domains overlapping with the top-ranking H3K27Ac-marked enhancers (**Fig. S6**). Of interest, a similar analysis performed in the hair follicle revealed that 34% of genes harboring domains of regulatory chromatin overlapping with H3K27Ac-superenhancers (18), underscoring that active enhancer formation encompasses additional mechanisms.

We next wished to identify "positive" transcriptional regulators whose gene expression scores are highly correlated to their ChromVAR TF motif deviation z-scores in the top quartile (r>0.5, p<0.01). This analysis yielded key lineage-determining genes concerned with NPC identity and fate decisions (**Figure 2 E,F**). Notably, we observed that in differentiation genes, e.g., *Hnf1b,* chromatin accessibility at CREs preceded gene

15

expression (**Figure S7 A**). Indeed, chromatin profiles show accessibility of the *Hnf1b* locus in precursor cells (**Figure S7 B**) indicating the necessity for additional gain of accessible chromatin at the gene's regulatory elements for cell-type specific expression.

Defining trajectory of transcriptional regulators in NPCs

To identify TFs that may control the gene expression programs across pseudotime, we linked TF motifs enriched in the different cell states with TF gene expression starting from naïve NPCs and ending in the proximo-distal cluster. At E16.5, Six2, Hoxa9, Hoxc8, and Forkhead box (Foxn3, Foxp1, Foxo1, Foxc1, Foxd1) are the initial TFs to engage chromatin (**Figure 3 A**). Additional factors that drive early differentiation (Tcf4/7, Klf10, Rbpj, Pax2/8 and Uncx) and those that specify the proximo-distal lineage fates (e.g., Hoxb2/b4, Lhx1, Hnf1b, Pou3f3, and Klf6 and Elf1/5) are subsequently engaged. Nearing the end of their lifespan, P2-NPCs are actively involved in differentiation, and this is reflected by a TF signature mostly composed of key fate-determining factors of proximal and distal fates, such as Hnf1b, Hnf4a, Hoxd8 and Elf family members (**Figure 3 B**). Interestingly, Nfkb, recently identified by single cell analysis of human kidneys as a potential regulator of proximal tubular integrity (20), is represented (along with its partners Rel-a and Rel-b) in the trajectory (**Figure 3 B**).

The early engagement of Foxn3 and Foxp1 factors in the NPC lineage was intriguing and reminiscent of Foxl1, recently identified in a similar analysis done at P0 (11). So, we explored where these factors might be expressed in the developing kidney. High-sensitivity RNA in-situ hybridization (RNAScope) revealed that Foxn3 and Foxp1 are expressed in the podocytes (albeit not exclusively) (**Figure 3 C**), whereas Foxo1 was

diffusely expressed in the nephrogenic zone (not shown). We also confirmed that Foxl1 is enriched in the podocyte, as recently reported (11) (**Figure 3 C**). These findings imply a potential role for this subset of Fox factors in the specification of the podocyte fate.

Multiome Analysis of the NPC lineage

To validate the above inferences based on integrated singleomes, we generated joint scATAC and scRNA datasets from E16.5 and P2 mouse kidneys using the Chromium Single-Cell Multiome ATAC+GEX platform (10x Genomics). Filtering across both data modalities yielded 7774 and 10,765 cells with high quality transcriptome and epigenome profiles, respectively (**Figure S8**). We obtained 168,566 and 171,147 accessible peaks representing potential CREs at E16.5 and P2, respectively. We clustered the scATAC-seq and scRNA-seq datasets and annotated these clusters using gene expression and gene scores, then selected the cell clusters representing the nephrogenic (NPC) lineage (npc-to-distal) for further analysis. We next integrated the derived gene scores (chromatin accessibility) with gene expression scores (GEX) (**Figure 4 A-D**). Key nephrogenesis factors such as *Six2* and *Pax2* (NPC), *Foxc2* (podocyte), *Hnf4a* (proximal), *Pou3f3* (LOH), and *Gata3* (distal tubule) showed strong cluster specific enrichment using gene and GEX scores (**Figure 4 E**). Marker-based cell clustering of chromatin accessibility defined the cell types of the nephrogenic lineage (**Figure 4 F, Figure S9**). Visualization of accessible peaks across all clusters (C1 to C17) in the *Six2* and *Hnf1b* gene loci (+/-70 kb from the TSS) revealed enriched co-accessible CREs in NPCs and proximo-distal clusters, respectively (**Figure 4 G**). By comparing gene scores and GEX scores, we observed two prevalent patterns: type I genes were characterized

by regulatory chromatin opening preceding gene activity. Examples in this group of genes are transcriptional regulators, enzymes, and G-protein coupled receptors (**Figure S10)**. In group II genes, regulatory chromatin opening coincides with gene activation. The latter group is enriched in channels and transporters. The biological significance of these differential chromatin behaviors is not clear but may be important in lineage priming and terminal differentiation.

Motif enrichment analysis reveals maturational differences in TF activity

We next performed ChromVAR motif enrichment analysis on marker peaks and differentially accessible peaks to determine if these groups of peaks are enriched for binding sites of specific TFs (Cutoff = FDR<=0.1 and Log2FC>0.5) (**Figure 5 A, B**). We found that bHLH motifs are enriched in podocyte and NPCs. In comparison, homeobox motifs are most enriched in NPCs, while Forkhead box motifs are enriched in podocytes. Distal clusters were enriched with accessible Pou motifs, whereas proximal clusters were enriched in Hnf4/Hnf1/Rxr motifs. AP-1 motifs were enriched in intermediate clusters 2, 5 and 8 representing differentiating proximal and distal tubules. Ranking the top differentially accessible motifs based on ChromVAR variability scores across all cells revealed Hnf1 motifs are the most represented in the accessible peaks (**Figure 5 C**). Other TF motifs that are also overrepresented include Six, Hnf4 and Hox motifs. Our previous bulk ATAC-seq analysis revealed that older (P0-P2) NPCs have greater chromatin accessibility to AP-1 factors. Here, we confirm through differential analysis of scATAC seq profiles that NPC maturation is associated with a remarkable shift in TF accessibility from the "progenitor" type (Six2, Hox, Wt1, Fox) to AP-1 factors (**Figure 5**

18

D). Other TFs that gain increased accessibility in P2 are terminal differentiation genes including Hnf1, Tead, Nfkb and Hnf4.

Trajectory of TF motif activity

Integration of TF gene expression with CRE accessibility along the NPC differentiation trajectory illustrated the dynamic accessibility to fate-determining TFs in the progenitor state (Six2, Meis2/3, Tcf21, Hoxc8/a9/a11), committed (E2f6, Stat3, Sox12, Pax2, Pax8), podocytes (Wt1, Fox factors), proximo-distal committed cells (Nfkb1, Essrb, Elf3, Klf6, Hoxb2/b4, Lhx1), distal (Pou3f3), and proximal fate (Hnf4a/g, Hnf1b, Ppara, Nr1h4) (**Figure 6**). To gain further insights into cell type-specific regulons (groups of co-regulated genes), we examined TF regulon activity using Single-Cell rEgulatory Network Inference and Clustering (SCENIC). SCENIC incorporates TF cell type specific expression with motif-based filtration to infer potential direct targets of each TF (regulons). Regulon activity was quantified and was binarized into "on" or "off" based on activity distribution across cells. As shown in **Figure 7 A and B**, SCENIC clustered the cells based on the regulon states of each cell showing strong enrichment of *Meis1, Zeb1, Hoxa10, Hoxd11, Ezh2, Foxc1*, and *Tcf7l1* regulon activity in nephron progenitors, *Mafb* and *Lef1* in podocytes, *Emx2, Hnf1b, Pax8, Lef1, Sox9, Nfkb1* in differentiating cells, *Tfdp1, Brca1, E2f* factors in cycling NPCs, *Hnf1a, and Hnf4g* in proximal tubules, *Atf3 and Paprgc1a* in loop of Henle, and *Hoxd8, Gata3* and *Pou3f3* in distal nephron, respectively. Examples of regulon activity and corresponding TF gene expression are shown in **Figure 7 C**.

CRE-gene associations across the NPC lineage uncover specialized cellular functions

Like our analysis in Six2^{GFP} cells (Figure 2), we applied a correlation-based approach that links CREs to gene expression, identifying 24270 dynamic interactions across all E16.5 and P2 NPC lineage cells and grouped these interactions into 10 k-means clusters (**Figure 8 A**). GREAT analysis of linked genes clustered well into various cell functional types such as metabolic processes, segmental nephron morphogenesis, glomerular/podocyte differentiation and tubular transport (linked genes are listed in Supplemental tables 1-10). Of significant interest, CRE-gene linkages identified genes involved in highly specialized cell functions such as macula densa, mesangial and podocyte morphogenesis.

Using a minimum of 10 CRE-gene links per gene, we defined a set of key developmental TFs and genes that define NPC, podocyte, committed, proximal and distal fates such as *Wt1, Hox, Six2, Foxc2, Hnf1b, Lef1*, *Cdh6, Mecom, Tfap2b,* and *Pou3f3* (**Figure 8 B**), illustrating the value of CRE-gene associations in identifying lineage- and fate-specific genes, and validating the results obtained in Six2^{GFP} cells. Interestingly, the top 2 ranked genes with the highest number of CRE-gene links are *Neat1* and *Malat1*; both genes encode long non-coding RNAs postulated to function in renal epithelial cell protection and repair pathways (21).

Analysis of early NPCs recovers Fox gene activity

To further understand the chromatin landscape of NPCs, we sub-clustered the Cited1-enriched cell population in E16.5 NPCs and obtained 4 cell clusters (**Figure S11 A**). UMAP plots of gene scores and gene expression revealed that C2 is enriched in *Cited1* and *Six2* (**Figure S11 B**). Also, Six2 motif activity is high in C2. In comparison, C4 is enriched in *Fox* family members (*Foxc2, Foxl1, Foxp1 and Foxn3*) when assessed

by gene scores, gene expression and motif activity (**Figure S11 B**, arrows). Interestingly, Wt1 motif activity is also high in C4. Examination of H3K27ac ChIP-seq and ATAC-seq tracks (GSE124804) revealed that Six2 and WT1 bind the *Foxl1* gene at accessible chromatin regions (peaks) upstream of the TSS (**Figure S11 C)**, suggesting that Six2 and Wt1 may directly regulate the cell type-specific expression of Fox factors. Analysis of TF motif enrichment in accessible peaks showed that C2 is enriched with accessible motifs for Six2, and the AP-1 family, while C4 is enriched in Fox motifs (**Figure S11 B, D**). Furthermore, matching TF expression with motif activity across pseudotime revealed high AP-1 and p53 activity (cell cycle/survival) in the beginning of the trajectory as compared to high Fox, Wt1 and Klf15 (podocyte) activity at the end of the trajectory (**Figure S11 E**).

In summary, elucidation of the accessible chromatin landscape in nephron progenitors informs the mechanisms of NPC choices to alternate fates and provides a foundation for future studies in disease states characterized by abnormal nephrogenesis.

Author Contributions

S.H. analyzed the scATAC and multiome sets and participated in manuscript writing and figure preparation; **G.T.** analyzed the scRNA data and assisted in scATAC analysis and figure preparation; **H.L.** participated in kidney cell processing and FACS analysis; **C-H.C.** participated in mouse breeding, genotyping, tissue preparation, and conducted the in situ hybridization studies; **S.E-D.** designed the experiments and participated in writing the manuscript and preparation of figures.

Acknowledgments

Disclosures

None.

Funding

Data Sharing

All data related to scATAC-seq, scRNA and multiome datasets reported in this paper have been deposited to the Gene Expression Omnibus (GEO) under accession number GSE180902 and GSE124804.

References

1. Mugford JW, Yu J, Kobayashi A, and McMahon AP. High-resolution gene expression analysis of the developing mouse kidney defines novel cellular compartments within the nephron progenitor population. *Dev Biol.* 2009;333(2):312-23.
2. Brown AC, Muthukrishnan SD, Guay JA, Adams DC, Schafer DA, Fetting JL, et al. Role for compartmentalization in nephron progenitor differentiation. *Proc Natl Acad Sci U S A.* 2013;110(12):4640-5.
3. Menon R, Otto EA, Kokoruda A, Zhou J, Zhang Z, Yoon E, et al. Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney. *Development.* 2018;145(16).
4. Lindstrom NO, De Sena Brandine G, Tran T, Ransick A, Suh G, Guo J, et al. Progressive Recruitment of Mesenchymal Progenitors Reveals a Time-Dependent Process of Cell Fate Acquisition in Mouse and Human Nephrogenesis. *Dev Cell.* 2018;45(5):651-60 e4.
5. Combes AN, Phipson B, Lawlor KT, Dorison A, Patrick R, Zappia L, et al. Single cell analysis of the developing mouse kidney provides deeper insight into marker gene expression and ligand-receptor crosstalk. *Development.* 2019;146(12).
6. Combes AN, Zappia L, Er PX, Oshlack A, and Little MH. Single-cell analysis reveals congruence between kidney organoids and human fetal kidney. *Genome Med.* 2019;11(1):3.
7. Hochane M, van den Berg PR, Fan X, Berenger-Currias N, Adegeest E, Bialecka M, et al. Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development. *PLoS Biol.* 2019;17(2):e3000152.
8. Buenrostro JD, Wu B, Chang HY, and Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol.* 2015;109:21 9 1-9.
9. Hilliard S, Song R, Liu H, Chen CH, Li Y, Baddoo M, et al. Defining the dynamic chromatin landscape of mouse nephron progenitors. *Biol Open.* 2019;8(5).
10. Guo Q, Kim A, Li B, Ransick A, Bugacov H, Chen X, et al. A beta-catenin-driven switch in TCF/LEF transcription factor binding to DNA target sites promotes commitment of mammalian nephron progenitor cells. *Elife.* 2021;10.
11. Miao Z, Balzer MS, Ma Z, Liu H, Wu J, Shrestha R, et al. Single cell regulatory landscape of the mouse kidney highlights cellular differentiation programs and disease targets. *Nat Commun.* 2021;12(1):2277.
12. Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet.* 2021;53(3):403-11.
13. Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB, et al. A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell.* 2018;174(5):1309-24 e18.
14. Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-Ferreres R, et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature.* 2018;555(7697):538-42.

15. Schep AN, Wu B, Buenrostro JD, and Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods.* 2017;14(10):975-8.

16. Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* 2018;46(D1):D1284.

17. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, et al. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol Cell.* 2018;71(5):858-71 e8.

18. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, et al. Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell.* 2020;183(4):1103-16 e20.

19. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 2010;28(5):495-501.

20. Muto Y, Wilson PC, Ledru N, Wu H, Dimke H, Waikar SS, et al. Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. *Nat Commun.* 2021;12(1):2190.

21. Lu HY, Wang GY, Zhao JW, and Jiang HT. Knockdown of lncRNA MALAT1 ameliorates acute kidney injury by mediating the miR-204/APOL1 pathway. *J Clin Lab Anal.* 2021;35(8):e23881.

Figure Legends

Figure 1. Integrated single-cell ATAC/RNA-seq reveals diversity of cell states in E16.5 and P2 Six2^GFP NPCs.

(A,E) UMAP plots of scATAC datasets with chromatin accessibility (gene scores)-based cell-type assignments in E16.5 (A) and P2 (E) NPCs.

(B,F) UMAP plots of scRNA-based cell-type assignments.

(C,G) UMAP plots of integrated scATAC/scRNA-based cell clustering.

(D,H) Heatmaps of scATAC gene scores for the top 13684 (E16.5) and 13916 (P2) accessible peaks. Representative genes of each cluster are shown.

NP: nephron progenitor; CC cell cycling; Pod: podocyte

Figure 2. Inferring regulatory chromatin from peak-gene links in Six2^GFP NPCs.

(A,B) Correlations of 7039 and 21788 peak-gene associations ATAC and RNA Z-scores across all cell states (k means=10). rGREAT annotation of cluster-related biological processes.

(C,D) Ranking of genes harboring chromatin domains with a minimum of 4 (E16.5) or 10 (P2) peak-gene links.

(E,F) Identification of top positive TF regulators whose correlation between motif and integrated gene expression is greater than 0.5 and 0.2, respectively with an adjusted p-value less than 0.01 and a maximum inter-cluster difference in deviation z-score delta value that is in the top 0.75 quartile.

Figure 3. Six2^{GFP} cellular trajectory based on matched transcription factor expression and motif activity.

(A,B) Side-by-side heatmaps of TF gene expression and ChromVar motif scores across pseudotime and cell states.

(C) Confocal microscopy images (x60 magnification) of section RNA in situ hybridization (RNAscope) showing expression of Foxc2, *Foxl1*, *Foxn3* and *Foxp1* in podocytes in E16.5 kidney (white arrows).

Figure 4. Multiome single-cell ATAC-seq/Gene expression (GEX) of E16.5/P2 kidneys.

(A) scATAC UMAP-based clustering.

(B) scRNA UMAP-based clustering.

(C,D) Matched ATAC/RNA UMAP-based clustering.

(E) Heatmap of gene scores and GEX of representative cell-type specific genes.

(F) UMAP plot (top) and heatmap (bottom) representing E16.5/P2 chromatin accessibility-based assignment of cell clusters.

(G) Representative chromatin accessibility tracks of all clusters for the *Six2* and *Hnf1b* genes. Boxed areas denote gene regulatory regions.

Figure 5. Multiome analysis of differential chromatin accessibility in E16.5/P2 NPC lineage.

(A) scATAC-based UMAP plot with cell-type specific annotations.

(B) ChromVAR TF motif enrichment per cell cluster.

(C) ChromVAR deviation scores for the most represented TF motifs in NPCs.

(D) Age-based motif enrichment analysis in E16.5 and P2 NPCs.

Figure 6. Multiome NPC trajectory based on matched transcription factor expression and motif activity.

Side-by-side heatmaps of TF gene expression and motif scores across pseudotime.

Figure 7. Regulon activity of the NPC lineage.

(A) SCENIC-generated heatmap of cell type-specific regulons. Regulons are binarized to "on" (black) or "off" (white). The genes are colored by the cell cluster.

(B) tSNE plot of regulon density representing regulon states as inferred by SCENIC algorithm.

(C) tSNE plots showing regulon activity of selected regulons for nephron progenitors (*Zeb1, Tcf7l1*), differentiating (*Hnf1b*), podocyte (*Mafb*), proximal (*Hnf4a*), and distal (*Gata3, Hoxd8, Pou3f3*) tubules.

Figure 8. Multiome analysis: inferring regulatory chromatin and TF trajectories.

(A) Correlations of 24270 CRE-gene associations ATAC and RNA Z-scores across all cell states (k means=10). rGREAT annotation of cluster-related biological processes.

(B) Inset, bar graph depicting ranking of genes based on the number of correlated peak-gene associations (+/-250 kb from TSS). Line graph depicts ranking of genes harboring chromatin domains with a minimum of 10 peak-gene links.

Supplemental Figure 1. Quality control for the scATAC-seq data.

(A) Insert size distribution of the scATAC-seq samples showing characteristic nucleosomal periodic patterns.

(B) Transcription start sites (TSS) signal enrichment of the scATAC-seq samples.

Supplemental Figure 2. TSS and unique nuclear fragments enrichment in scATAC-seq data.

(A) Violin plots showing TSS enrichment scores per sample at E16.5 and P2.

(B) Violin plot showing the log(base10) of unique nuclear fragments per sample.

Supplemental Figure 3. Chromatin accessibility gene scores identify new markers of proximal and distal tubules.

(A,C) UMAP plots of P2 Six2^{GFP} scATAC depicting selective enrichment of chromatin activity for *Mcmdc2* and *Nabp1* in proximal and distal progenitors, respectively.

(B) Confocal microscopy images (x40) of section in situ hybridization for *Mcmdc2* RNA in E16.5 and P2 kidneys showing predominant expression in proximal (LTL^{+}) tubules.

(D) Confocal microscopy images (x40) of section in situ hybridization for *Nabp1* RNA showing predominant expression in LTA-negative (distal) segments.

Supplemental Figure 4. Cicero cis co-accessibility networks.

(A,B) E16.5 and P2 Six2^{GFP} NPCs. Promoter-promoter, Intron-intron, Intergenic-intergenic connections are predominant in both age groups.

(C) Age-related comparison of Cicero cis co-accessibility networks (CCANs) showing enhanced Cicero connections with age.

Supplemental Figure 5. Peak-gene links representing co-variation in regulatory elements accessibility with gene expression and cell type-specific expression. Boxed areas depict differential peaks with significant correlation to gene expression. The strength of correlation is depicted by the color of the line.

Supplemental Figure 6. Ranking of ChIP-seq H3K27Ac-enriched peaks per gene in E13.5 and P0 NPCs. Datasets derived from (9)

Supplemental Figure 7. Chromatin activity and gene expression at the *Hnf1b* locus in P2 Six2^GFP NPCs.

(A) The peak counts of all *Hnf1b* correlated peaks (left) and *Hnf1b* gene expression (right) colored in UMAP. The bracketed areas point to regions with differential signals.

(B) Browser view of cell state specific peak-gene links at the *Hnf1b* locus. Putative enhancer peak-gene links are active in differentiating progenitors and those fated to become proximal/distal progenitors (box). Proximal-specific activity is associated with a gain of two additional peak-gene links (arrowheads).

Supplemental Figure 8. Quality control of Multiome scATAC/GEX (10x Genomics).

Supplemental Figure 9. Differential gene-specific chromatin accessibility across cell clusters. Representative genes of progenitors (*Six2*), early (*Wnt4, Lef1*) and late (*Lhx1, Hnf1b*) differentiating, proximal (*Hnf4a*), LOH/distal (*Pou3f3 and Tfap2b*) segments are shown.

Supplemental Figure 10. Multiome analysis of chromatin accessibility and gene expression defines two types of genes. (A) Type I: regulatory chromatin opening precedes gene activation. (B) Type II: regulatory chromatin opening coincides with gene activation.

Supplemental Figure 11. Chromatin landscape of the early progenitor cluster.

(A) UMAP sub-clustering of E16.5 multiome Cited1-high cluster. Numbers in parentheses represent the cell number per sub-cluster.

(B) UMAP plots highlighting chromatin accessibility (gene scores), gene expression and TF motif activity. *Fox* factors and *Six2/Cited1* have reciprocal cellular activity in C4 and C2, respectively. *WT1* motif activity overlaps with that of *Fox* factors.

(C) Browser tracks of ChIP-seq (*Six2, Osr1*, and *WT1*) and bulk ATAC-seq in NPC-P0 (GSE804) highlighting the *Foxl1* locus. Six2 and WT1 co-occupy accessible chromatin regions upstream of the TSS (boxed area).

(D) Enrichment of accessible AP-1 motifs in cluster 2 and Fox factors motifs in cluster 4. Note that the TF identity represents consensus elements and thus do not precisely reflect the exact member of the TF family.

(E) Trajectory heatmap of TF expression and motif activity. Start of trajectory is cluster 2 and end is cluster 4.

Figure 1

Figure 2

A full-page scientific figure (Figure 2) with panels A–F showing heatmaps of ATAC and RNA Z-scores, GO biological process enrichment bar charts, ranked gene list plots for peak-gene associations, and TF motif scatter plots.

E16.5-Six2^GFP NPCs

Regulatory element-Gene expression linkages

GO Biological Process (-log10 pvalue)
Hypergeometric Expected

epithelial cell differentiation	18.45
morphogenesis of an epithelium	16.92
reproductive system development	16.00
reproductive structure development	15.83
gland development	15.70

GO Biological Process
-log10(Binomial p value)

paramesonephric duct development	41.48
regulation of epithelial cell differentiation	23.00
renal vesicle morphogenesis	22.44
metanephric nephron morphogenesis	21.86

GO Biological Process
-log10(Binomial p value)

DNA replication-dependent nucleosome assembly	49.71
chromatin silencing at rDNA	39.68
DNA packaging	38.05
chromatin assembly	34.09
chromatin assembly or disassembly	32.95

GO Biological Process
-log10(Binomial p value)

regulation of intrinsic apoptotic signaling pathway	25.29

GO Biological Process
-log10(Binomial p value)

positive regulation of mitochondrion organization	6.33
cell-cell junction organization	6.12
cell-substrate adhesion	5.89
cell junction assembly	5.33
negative regulation of translational elongation	4.79
oocyte development	4.63
Peyer's patch morphogenesis	4.51
oocyte differentiation	4.28
bone remodeling	4.27
female gamete generation	

Regulatory element-Gene expression linkages

P2-Six2^GFP NPCs

E

Positive TFs

Hoxb2,4
Foxc1,d1,n3
Lef1, Pax8, Pou3f1
Tcf7,21

TFRegulator   NO   YES

F

Positive TFs

Hoxa11,b5,c11,d8,d11
Foxc2,l1, Lhx1, Jun,
Fos, Fosb, Uncx,
Hnf1b,Klf5,11,12

TFRegulator   NO   YES

Peak-gene associations N=7139

DORCs (n=422)

425 genes

Wnt4
Hnf1b
Six2
Lhx1
Foxc2
Pou3f3
Lef1
Wt1
Mafb

Peak-gene associations n=21788

412 genes

Hoxb3
Hoxb5
Hoxb4
Hoxb2
Hoxa9
Foxc1
Six2
Hnf1b
Meox1, Emx2
Uncx

Number of Correlated Peaks

Ranked Gene List (x1000)

ATAC Z-Scores
21788 P2GLinks
-2 ... 2

RNA Z-Scores
21788 P2GLinks
-2 ... 2

groupBy
Diff NPC
Early Distal
Early Prox.
NPC1
NPC2
Podocyte
Prox. Prog

Figure 3

E16.5-Six2^GFP NPCs

P2- Six2^GFP NPCs

C

Figure 4

Figure 5

Multiome
E16.5+P2 kidneys

A

scATAC

C1-C4: proximal
C5: proximo-distal
C6: early proximal
C7: early LOH/distal
C8-C10: LOH/distal
C11: podocytes
C12,C13: early podocyte
C14-C17: nephron progenitors

UMAP_Combined of LSI_Combined colored by
colData : Clusters

D

if enrichment per age group
(HOMER)

UMAP_Combined of LSI_Combined colored by
colData : Sample

ChromVAR_deviation z-score

color ■ 1-E165dual ■ 2-P2dual

color
- ■ C1 ■ C5 ■ C9
- ■ C2 ■ C6 ■ C10
- ■ C3 ■ C7 ■ C11
- ■ C4 ■ C8 ■ C12

Six2.Homeobox_255 (137)
Hoxc9.Homeobox_132 (132)
HOXA9.Homeobox_129 (112)
Six1.Homeobox_254 (104)
CDX4.Homeobox_31 (70)
Tcf21.bHLH_287 (63)
Pitx1.Ebox.Homeobox.bHLH_224 (60)
Cdx2.Homeobox_30 (57)
FOXA1.AR.Forkhead.NR_5 (53)
HOXA2.Homeobox_128 (53)
FOXA1.Forkhead_87 (52)
ZBTB18.Zf_307 (51)
Ap4.bHLH_4 (49)
FOXA1.Forkhead_88 (47)
HOXD13.Homeobox_133 (47)
Ptf1a.bHLH_232 (47)
FoxL2.Forkhead_93 (42)
Unknown.Homeobox_153 (42)
Myf5.bHLH_169 (42)
NeuroD1.bHLH_173 (41)
Hoxb4.Homeobox_131 (39)
Olig2.bHLH_202 (39)
MyoG.bHLH_171 (38)
FOXM1.Forkhead_94 (36)
Pdx1.Homeobox_219 (36)
HOXB13.Homeobox_130 (34)
Tcf12.bHLH_286 (32)
Fox.Ebox.Forkhead.bHLH_90 (31)
E2A.bHLH_51 (31)
SCL.bHLH_252 (30)
HEB.bHLH_121 (30)
Oct4.Sox17.POU.Homeobox.HMG_197 (28)
NF1.FOXA1.CTF.Forkhead_174 (27)
Foxa2.Forkhead_89 (25)
MyoD.bHLH_170 (23)
Lhx3.Homeobox_152 (22)
Phox2a.Homeobox_221 (21)
OCT4.SOX2.TCF.NANOG.POU.Homeobox.HMG_198 (21)
Tcf4.HMG_289 (20)
Fosl2.bZIP_86 (298)
BATF.bZIP_20 (292)
Atf3.bZIP_12 (289)
Jun.AP1.bZIP_141 (281)
Fra1.bZIP_98 (280)
Fra2.bZIP_99 (278)
AP.1.bZIP_1 (266)
Bach2.bZIP_17 (182)
Bach1.bZIP_16 (77)
NF.E2.bZIP_179 (76)
Nrf2.bZIP_192 (61)
TEAD.TEA_294 (48)
TEAD2.TEA_292 (40)
MafK.bZIP_158 (36)
Hnf1.Homeobox_125 (36)
TEAD4.TEA_293 (28)
NF1.CTF_176 (27)
HNF4a.NR..DR1_126 (22)
NFkB.p65.Rel.RHD_180 (21)
c.Jun.CRE.bZIP_142 (20)

E16.5   **P2**

Norm. Enrichment -log10(P-adj) [0-Max]

0   100

10(P-adj) [0-Max]

Essrb
Nr5a2
Tcfcp2l
Hnf1

Lhx

Ap-1

Hnf4a
Tr4
Ppar
Rxr
Erra
Nf1

Tcfcp2l1 Tead Lhx Hox Pax Pou Fox
Tcf21

Rank Sorted Annotations

combinedVars
20 40 60 80

Figure 6

Left column labels:

hr13:Foxd1
hr3:Foxo1
hr8:Maf
hr2:Mafb
hr13:Sox4
hr17:Six2
hr13:Foxc1
hr17:Six3
hr9:Rora
hr8:Nr2f6
hr6:Hoxa9
hr15:Hoxc8
hr1:Pbx1
hr12:E2f6
hr6:Hoxa5
hr3:E2f5
hr12:Mycn
hr17:Pbx2
hr11:Patz1
hr5:Klf3
hr3:Lef1
hr15:Klf10
hr5:Rbpj
hr2:Sp5
hr11:Tcf7
hr4:Creb3
hr16:Gabpa
hr2:Xbp1
hr10:Hmg20b
hr2:Pbx3
hr1:Klf7
hr2:Phf21a
hr19:Pax2
hr6:Hoxa7
hr2:Hoxd4
hr4:Klf4
hr7:E2f8
hr15:Atf1
hr5:Cux1
hr2:Pax8
hr1:Creb1
hr5:Kdm2b
hr8:E2f4
hr11:Hoxb2
hr2:Mga
hr11:Lhx1
hr12:Id2
hr15:Sp1
hr5:Fosl2
hr4:Pou3f1
hr10:Smarcc2
hr19:Emx2
hr11:Hoxb4

z: labels (column 2):

z:Foxd1_330
z:Foxo1_314
z:Maf_793
z:Mafb_136
z:Sox4_751
z:Six2_445
z:Foxc1_320
z:Six3_516
z:Rora_682
z:Nr2f6_657
z:Hoxa9_508
z:Hoxc8_405
z:Pbx1_554
z:E2f6_264
z:Hoxa5_510
z:E2f5_262
z:Mycn_65
z:Pbx2_496
z:Patz1_153
z:Klf3_804
z:Lef1_734
z:Klf10_810
z:Rbpj_824
z:Sp5_238
z:Tcf7_726
z:Creb3_870
z:Gabpa_273
z:Xbp1_103
z:Hmg20b_729
z:Pbx3_513
z:Klf7_171
z:Phf21a_15
z:Pax2_612
z:Hoxa7_509
z:Hoxd4_531
z:Klf4_143
z:E2f8_263
z:Atf1_106
z:Cux1_474
z:Pax8_854
z:Creb1_109
z:Kdm2b_249
z:E2f4_257
z:Hoxb2_582
z:Mga_96
z:Lhx1_422
z:Id2_33
z:Sp1_142
z:Fosl2_113
z:Pou3f1_620
z:Smarcc2_646
z:Emx2_535
z:Hoxb4_511

z:Emx2_535
z:Hoxb4_511

Middle labels:

Tcf
Klf
Rbpj
Pax2,8
Uncx

Hoxb2,b4
Lhx1
Hnf1b
Emx2
Pou3f

Prrx1
Klf6
Elf1,5

Middle-right z: labels:

z:Hoxa5_510
z:E2f5_262
z:Mycn_65
z:Pbx2_496
z:Patz1_153
z:Klf3_804
z:Lef1_734
z:Klf10_810
z:Rbpj_824
z:Sp5_238
z:Tcf7_726
z:Creb3_870
z:Gabpa_273
z:Xbp1_103
z:Hmg20b_729
z:Pbx3_513
z:Klf7_171
z:Phf21a_15
z:Pax2_612
z:Hoxa7_509
z:Hoxd4_531
z:Klf4_143
z:E2f8_263
z:Atf1_106
z:Cux1_474
z:Pax8_854
z:Creb1_109
z:Kdm2b_249
z:E2f4_257
z:Hoxb2_582
z:Mga_96
z:Lhx1_422
z:Id2_33
z:Sp1_142
z:Fosl2_113
z:Pou3f1_620
z:Smarcc2_646
z:Emx2_535
z:Hoxb4_511

Legends:

GeneIntegrationMatrix
-1.5   1.5
MotifMatrix
-1.5   1.5

motif activity

GeneIntegrationMatrix
-1.5   1.5
MotifMatrix
-1.5   1.5

Bottom-right labels:

Esrrg
Pou3f3
Esrra
Elf3
Klf6
Emx2
Hoxb4
Hoxb2
Lhx1
Sox11
Relb
Hnf1b
Hnf4g
Hnf4a
Rela
Nr3c1
Srebf2
Mafg
Ppara
Junb
Nr1h4

Figure 7

Figure 8

Multiome (E16.5+P2)

CRE-gene linkages

■ E16.5
■ P2

A

Job ID: 20210421-public-4.0.4-va3bBn
Display name: E16.5_Kmeans2.txt

GO Biological Process
Hypergeometric Expected

0 2 4 6 8 10 12 14 16 18

epithelial cell differentiation — 18.45
morphogenesis of an epithelium — 16.92
reproductive system development — 16.00
reproductive structure development — 15.83
gland development — 15.70

Monocarboxylic acid metabolism (49.4)
Fatty acid metabolism (33.7)
Anion transport (31.1)
Microvillus length (30.5)

20210421-public-4.0.4-1maEiM
name: E16.5_Kmeans5.txt

GO Biological Process
-log10(Binomial p value)

0 5 10 15 20 25 30 35 40

paramesonephric duct development — 41.48
regulation of epithelial cell differentiation — 23.00
renal vesicle morphogenesis — 22.44
metanephric nephron morphogenesis — 21.96

Renal tubule development (61.2)
Regionalization (40.2)
Macula densa development (34.9)
DCT/LOH cell differentiation (34.9)
Mesoderm cell fate (69.2)
Renal vesicle morphogenesis (53.5)

Job ID: 20210421-public-4.0.4-VGI0A0
Display name: E16.5_Kmeans6.txt

GO Biological Process
-log10(Binomial p value)

0 5 10 15 20 25 30 35 40 45

DNA replication-dependent nucleosome assembly — 49.71
chromatin silencing at rDNA — 38.05
DNA packaging — 34.09
chromatin assembly — 32.95
chromatin assembly or disassembly —

A/P patterning (55.8)
Mes-epi transition (45.8)
Mesoderm cell fate (42.8)

Job ID: 20210421-public-4.0.4-hj8pWC
Display name: E16.5_Kmeans8.txt

GO Biological Process
-log10(Binomial p value)

0 2 4 6 8 10 12 14 16 18 20 22 24

regulation of intrinsic apoptotic signaling pathway — 25.29
regulation of organelle organization —

Cell junction organization (37.5)
Mesonephros proliferation (28.0)
Visceral epithelial development (25.3)
Actin cytoskeleton organization (24.3)

Job ID: 20210421-public-4.0.4-CLMsCB
Display name: E16.5_Kmeans10.txt

GO Biological Process
-log10(Binomial p value)

0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5

positive regulation of mitochondrion organization — 6.33
cell-cell junction organization — 6.12
cell-substrate adhesion — 5.89
cell junction organization — 5.33
negative regulation of translational elongation — 4.79
oocyte development — 4.70
Peyer's patch morphogenesis — 4.63
oocyte differentiation — 4.31
bone remodeling — 4.28
gamete generation — 4.27

Sodium transport (120.0)
Potassium transport (104.5)
ATP hydrolysis coupled transport

CREClusters

metanephr

meta

AA

npc1
npc2
diff

ATAC Z-Scores
24270 P2GLinks
-2 — 2

RNA Z-Scores
24270 P2GLinks
-2 — 2

Regulat...   ...pression linkages

Job ID: 2021042...
Display name:

GO Biological Process
Hypergeometric Expected
-log10 (Binomial p value)

0 2 4 6 8 10 12 14 16 18

...fferentiation — Epithelial differentiation — 18.45
...n epithelium — Morphogenesis of epithelium — 16.92
reproductive system development — 16.00
reproductive structure development — 15.83
gland development — 15.70

B

Peak-gene associations
N= 7139

Number of Genes

Correlated Associations per Gene
1 2 3 4 5 6 7-10 >10

Ranked Gene List

0 5000 10000 15000 20000 25000 30000

Ranked Gene List

DORCs
(n=422)

Neat1
Malat1
Foxc2
Hoxc8, Uncx
Hoxc9
Pou3f3
Cdh1
Wt1
Pax2, Slc34a1
Hnf1b, Hoxd11
Six2, Mecom
Pbx1, Frz7
Essrg, Cldn10
Foxc1, Hoxa11, Slc12a1
Bmper, Cdh6
Trap2b

Wnt4
Emx2, Hnf1b, Kctd1, Pax2
Lhx1, Sp5
Bdnf, Foxc2, 2xWt1, Pou3f3, Six2
Hoxa3, Hoxc8, Igf2, Klf16, Lef1, Ncam1
Ehmt2, Fxyd6, Gas1, Mafb, Notch1, Smarcc2, Uncx, Wt1

Peak-gene associations
n = 21788

Genes

2500
2000
1500
1000
500
0

Correlated associations per gene
1 2 3 4 5 6 7-10 >10

Peaks

100
90
80
70

Hoxb3
Hoxb1
Hoxb5

Supplemental Figures

Single-cell chromatin and gene-regulatory dynamics of mouse nephron progenitors

Sylvia Hilliard,[1] **Giovane Tortelote,**[1] **Hongbing Liu,**[1] **Chao-Hui Chen,**[1] **and Samir S. El-Dahr**[1*]

[1]Section of Pediatric Nephrology, Department of Pediatrics, Tulane University School of Medicine, New Orleans, LA 70112, USA

Table of contents:

A

B

E16.5_sample 1

Insert Size Distribution
Sample SH2_E165 -

Enrichment around TSS (normalized)
Sample SH2_E165 -

E16.5_sample 2

Insert Size Distribution
Sample SH9_E165 -

Enrichment around TSS (normalized)
Sample SH9_E165 -

P2_sample 1

Insert Size Distribution
Sample SH3_P2 -

Enrichment around TSS (normalized)
Sample SH3_P2 -

P2_sample 2

Insert Size Distribution
Sample SH7_P2 -

Enrichment around TSS (normalized)
Sample SH7_P2 -

scATAC-seq: TSS and unique nuclear fragments enrichment

A

B

Mcmdc2 LTL Dapi

A *Mcmdc2*

UMAP Dimension 2 / UMAP Dimension 1

NormCounts

0.1 0.2 0.3 0.4 0.5

scATAC

B

E16.5

P2

Nabp1 LTL Dapi

C *Nabp1*

UMAP Dimension 2 / UMAP Dimension 1

NormCounts

0.3 0.6 0.9

scATAC

D

E16.5

P2

Cis-coaccessibility Networks (CCANS)

A

E16.5 CCANS

B

P2 CCANS

C
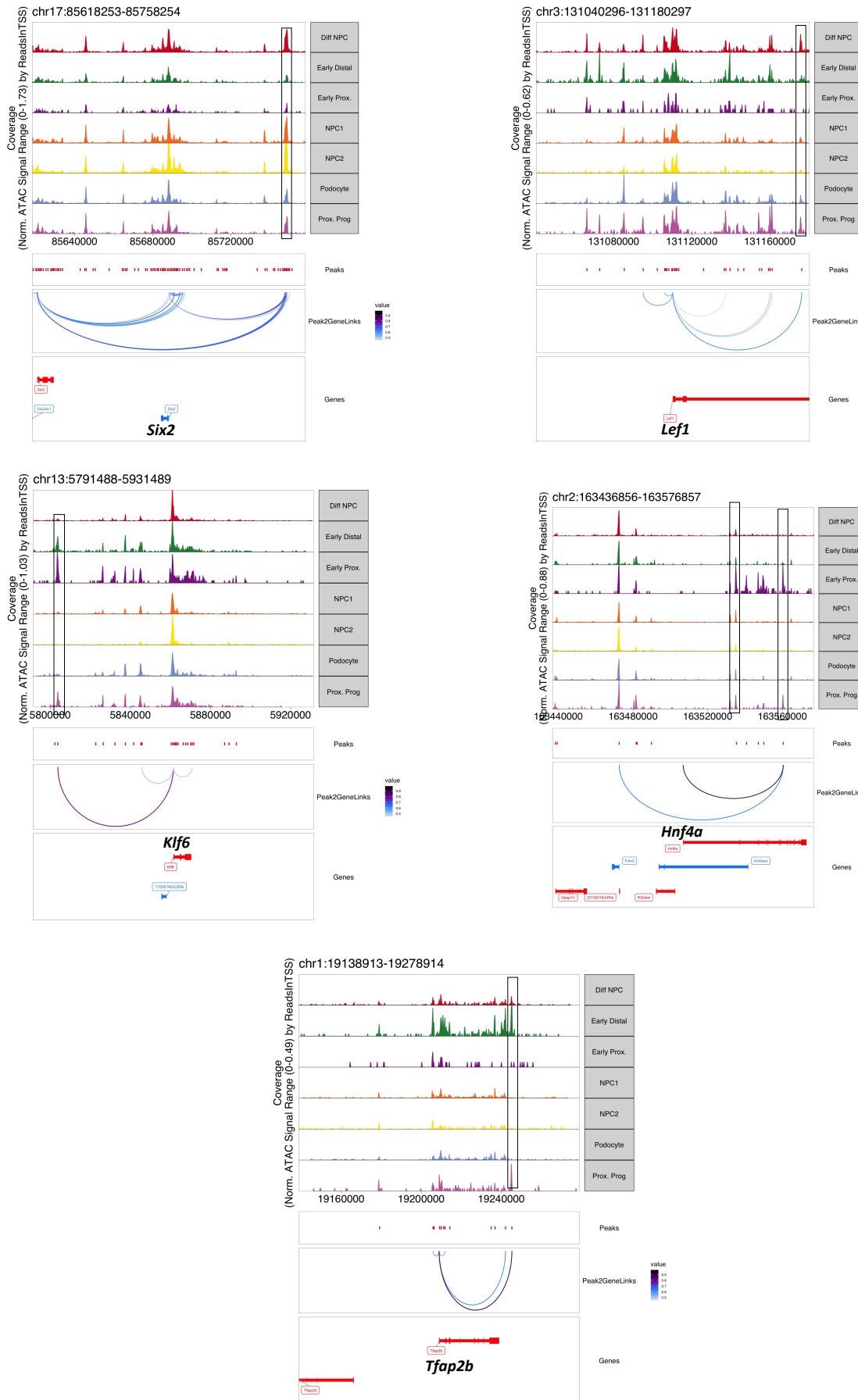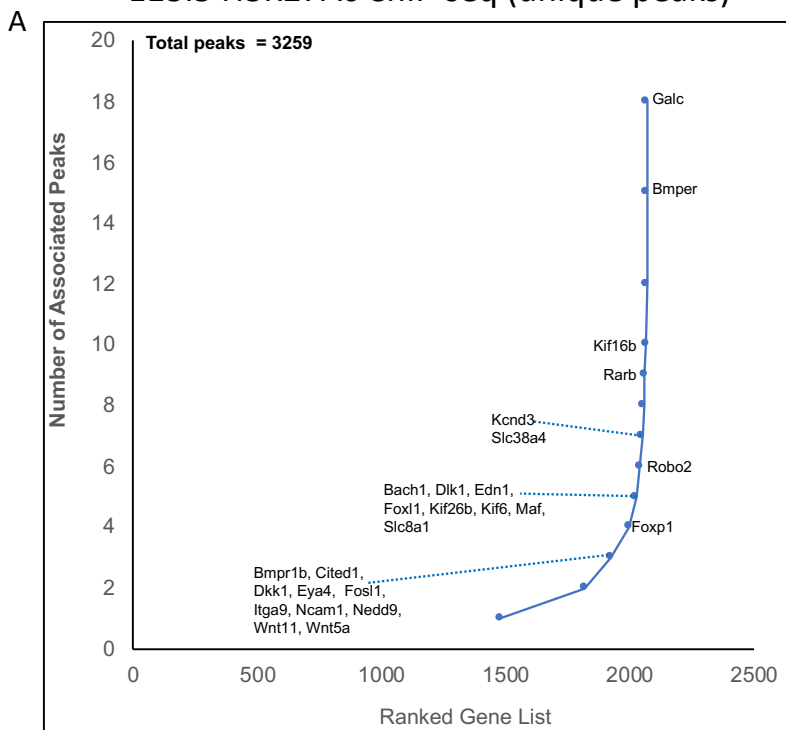

Cis-Coaccessibility Networks

CRE-gene associations correlate with cell type-specific gene expression

Fig. S5

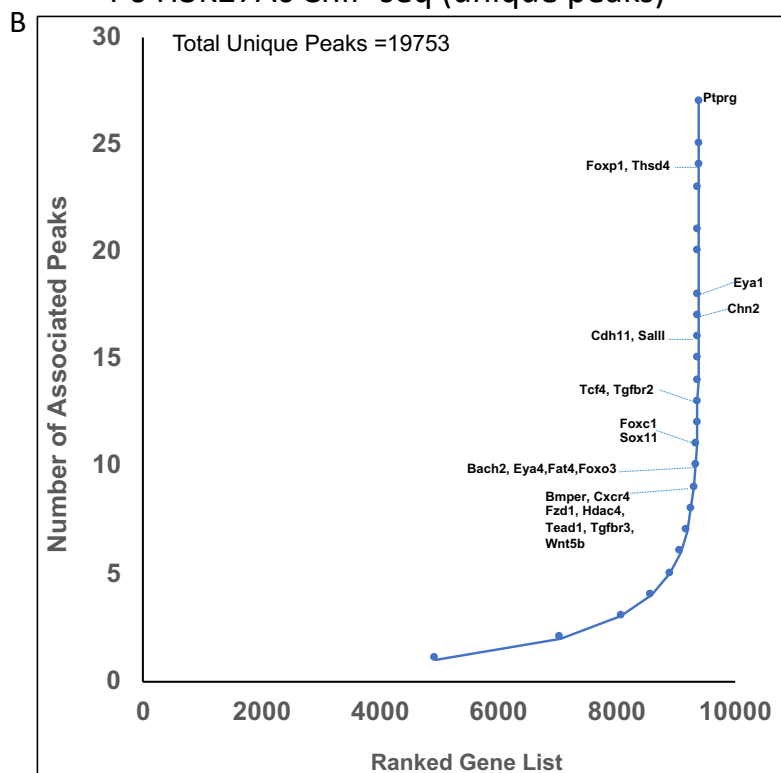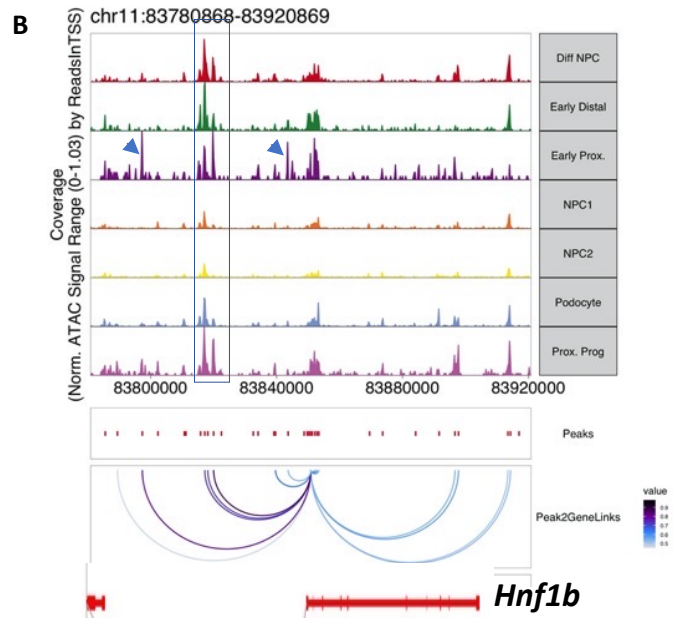## E13.5 H3K27Ac ChIP-seq (unique peaks)

A

Total peaks = 3259

Labels on curve (from top): Galc (18), Bmper (15), Kif16b (10), Rarb (9), Kcnd3, Slc38a4 (7), Robo2 (6), Bach1, Dlk1, Edn1, Foxl1, Kif26b, Kif6, Maf, Slc8a1 (5), Foxp1 (4), Bmpr1b, Cited1, Dkk1, Eya4, Fosl1, Itga9, Ncam1, Nedd9, Wnt11, Wnt5a

X-axis: Ranked Gene List (0–2500)
Y-axis: Number of Associated Peaks (0–20)

P0 H3K27Ac ChIP-seq (unique peaks)

B

Total Unique Peaks =19753

Labels on curve (from top): Ptprg (27), Foxp1, Thsd4 (24), Eya1 (18), Chn2 (17), Cdh11, Sall1 (16), Tcf4, Tgfbr2 (14), Foxc1, Sox11 (11), Bach2, Eya4, Fat4, Foxo3 (10), Bmper, Cxcr4, Fzd1, Hdac4, Tead1, Tgfbr3, Wnt5b

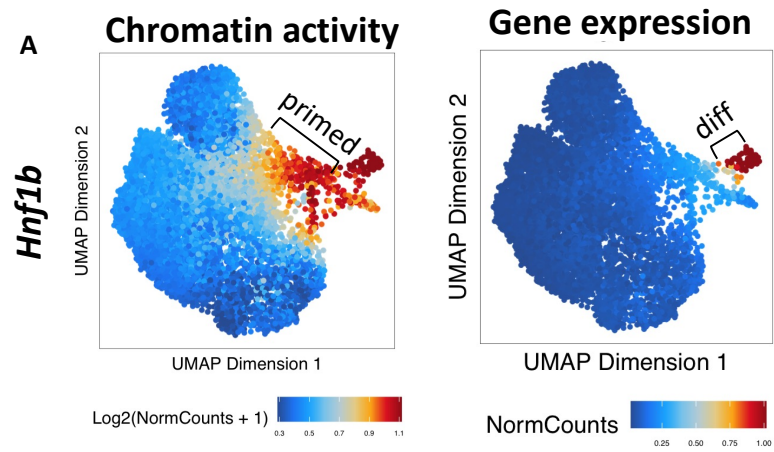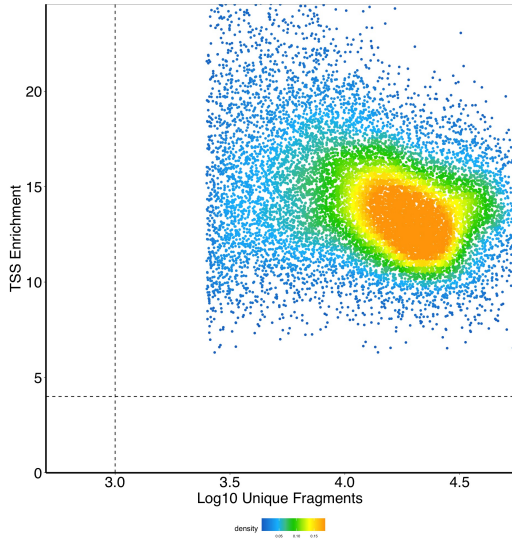X-axis: Ranked Gene List (0–10000)
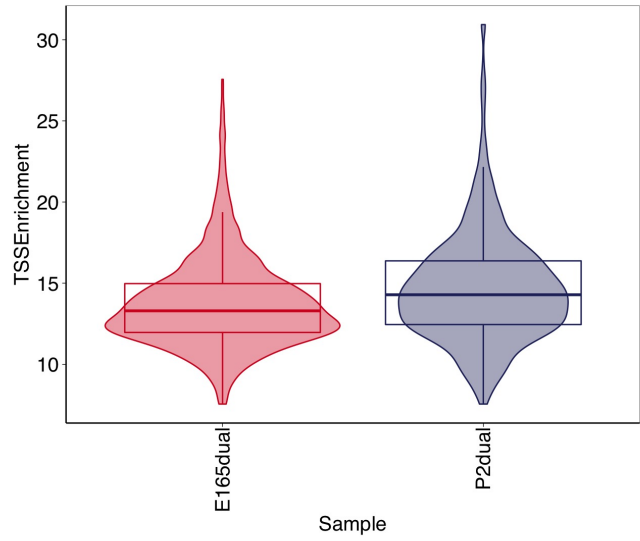Y-axis: Number of Associated Peaks (0–30)

Multiome: scATAC/GEX (E16.5 + P2)
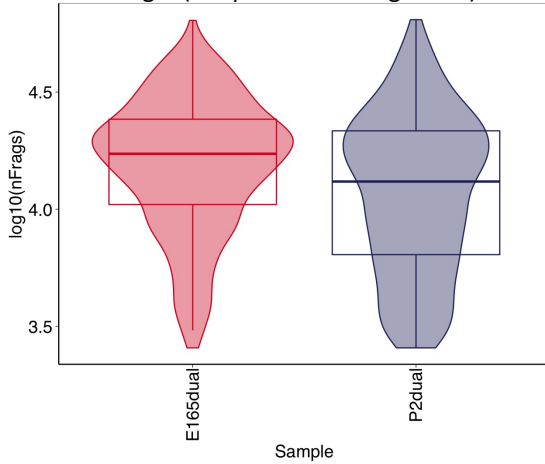## QC metrics
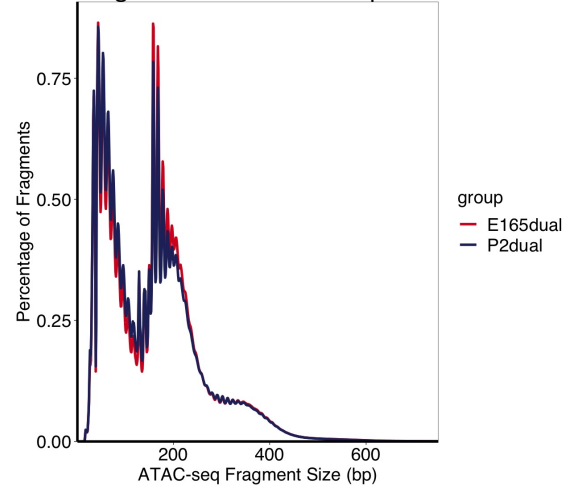

log10(Unique Fragments) vs TSS enrichment score

TSS enrichment score

Log10(unique nuclear fragments)

Fragment size distribution profiles

TSS enrichment profiles per sample

Differential gene-specific chromatin accessibility   Fig. S9

Multiome
E16.5+P2 kidneys

A

scATAC

C1-C4: proximal
C5: proximo-distal
C6: early proximal
C7: early LOH/distal
C8-C10: LOH/distal
C11: podocytes
C12,C13: early podocyte
C14-C17: nephron progenitors

UMAP_Combined of LSI_Combined colored by colData : Clusters

UMAP_Combined of LSI_Combined colored by colData : Sample

color

C1	C5	C9
C2	C6	C10
C3	C7	C11
C4	C8	C12

color   ■ 1-E165dual   ■ 2-P2dual
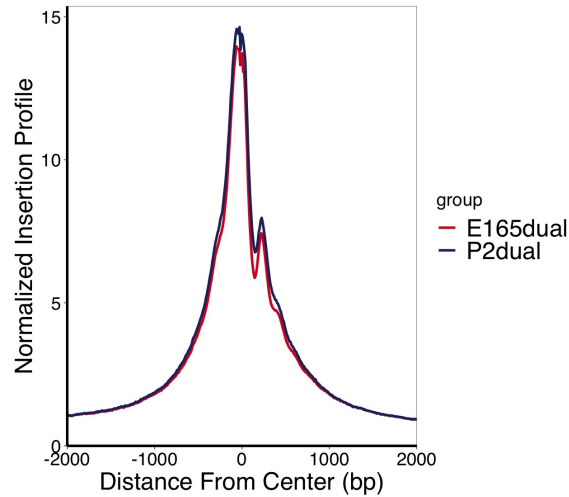
Type I genes
Regulatory chromatin access <u>precedes</u> gene activation

A
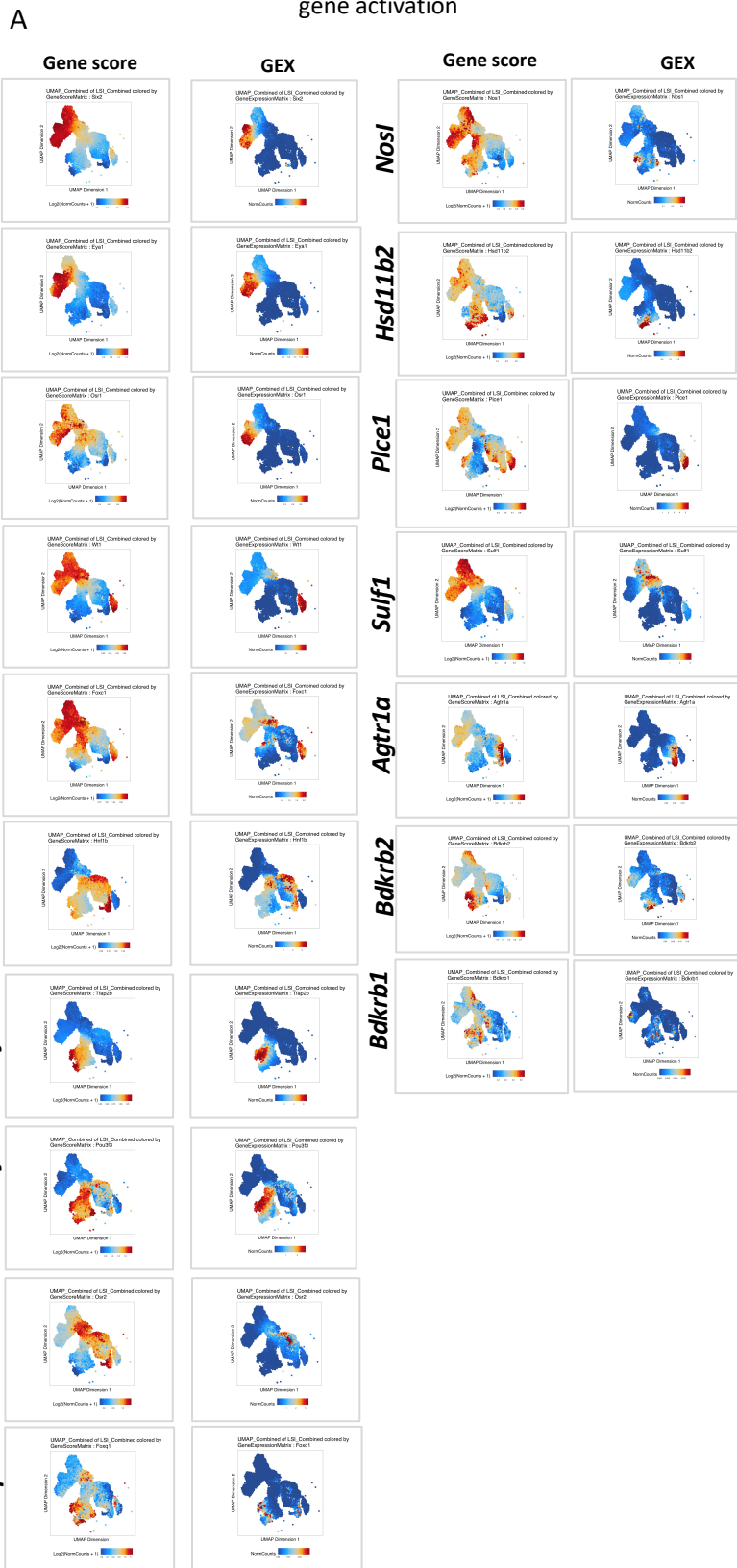
Type II genes
Regulatory chromatin access <u>coincides</u> with gene activation

B

A

C1(329)   C3(162)
C2(246)   C4(248)

B

Gene scores | Gene expression | TF motif activity

Cited1
Six2
Wt1
Foxc2
1

C

ChIP — Six2, Osr1, Wt1
ATAC
Foxl1

D

TF motif enrichment per cluster

— Smarcc1_843 (130)
— Fos_104 (119)
— Bach1_108 (97)
— Bach2_119 (72)
— Fosb_98 (72)
— Jund_135 (59)
— Junb_127 (55)
— Fosl1_107 (42)
— Nfe2_132 (37)
— Batf3_789 (34)
— Foxl1_381 (43)
— Foxd3_327 (42)

E

trajectory

NPC Trajectory — TF expression
NPC Trajectory — TF motif activity

Jun
Fosb
Jund
Rfx2
Fgr3
Trp53
Fgr2
Fos
Hivep3
Nfatc2
Junb
Atf1
Stat3
Meis2
E2f5
Uncx
Tcf3
Tcf4
— Hoxb8
— Hoxc6
— Zfp281
— Ybx1
— Klf7
— E2f1
— Foxl1
— Wt1
— Klf6
— Foxc2
— Klf15

ExpressionMatrix   1.5
Matrix   1.5

UMAP_Combined4 of LSI_Combined4 colored by
GeneScoreMatrix : Cited1

low   high

UMAP_Combined4 of LSI_Combined4 colored by
GeneExpressionMatrix : Cited1

NormCounts   2 3 4 5