

Washington University School of Medicine

Digital Commons@Becker

---

2020-Current year OA Pubs

Open Access Publications

---

5-16-2023

## A single-cell multiomic analysis of kidney organoid differentiation

Yasuhiro Yoshimura

Yoshiharu Muto

Nicolas Ledru

Haojia Wu

Kohei Omachi

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.wustl.edu/oa\\_4](https://digitalcommons.wustl.edu/oa_4)

 Part of the [Medicine and Health Sciences Commons](#)

Please let us know how this document benefits you.

---

---

**Authors**

Yasuhiro Yoshimura, Yoshiharu Muto, Nicolas Ledru, Haojia Wu, Kohei Omachi, Jeffrey H Miner, and Benjamin D Humphreys



# A single-cell multiomic analysis of kidney organoid differentiation

Yasuhiro Yoshimura<sup>a</sup> , Yoshiharu Muto<sup>a</sup>, Nicolas Ledru<sup>a</sup>, Haojia Wu<sup>a</sup>, Kohei Omachi<sup>a</sup> , Jeffrey H. Miner<sup>a</sup> , and Benjamin D. Humphreys<sup>a,b,1</sup>

Edited by Janet Rossant, Gairdner Foundation, Toronto, Canada; received November 22, 2022; accepted April 5, 2023

**Kidney organoids differentiated from pluripotent stem cells are powerful models of kidney development and disease but are characterized by cell immaturity and off-target cell fates. Comparing the cell-specific gene regulatory landscape during organoid differentiation with human adult kidney can serve to benchmark progress in differentiation at the epigenome and transcriptome level for individual organoid cell types. Using single-cell multiome and histone modification analysis, we report more broadly open chromatin in organoid cell types compared to the human adult kidney. We infer enhancer dynamics by cis-coaccessibility analysis and validate an enhancer driving transcription of *HNF1B* by CRISPR interference both in cultured proximal tubule cells and also during organoid differentiation. Our approach provides an experimental framework to judge the cell-specific maturation state of human kidney organoids and shows that kidney organoids can be used to validate individual gene regulatory networks that regulate differentiation.**

kidney organoid | scRNA-seq | scATAC-seq | CUT&RUN | CRISPR interference

The mammalian kidney contains more than 20 specialized cell types acting together to maintain homeostasis by regulating water and electrolyte balance, blood pressure, and waste excretion among other functions. Kidneys are composed of nephrons which consist of a glomerulus and renal tubule which filters blood to ultimately produce urine. Nephron dysfunction leads to chronic kidney disease (CKD) which affects 9.1% of the population worldwide (1). A more detailed understanding of kidney cell types and their development will provide a foundation on which to build replacement cell types for regenerative medicine approaches to treat kidney failure.

Epigenetic regulation plays a critical role in controlling cell differentiation during development. Chromatin accessibility regulates the potential for gene expression in part by enabling access of transcription factors to bind regulatory genomic DNA sequences. The analysis of chromatin accessibility profiles combined with gene transcription at single-cell resolution allows the inference of gene regulatory networks underpinning cellular identity (2–8), including in the developing mouse and human adult kidneys (9–11). Recent technological advances allow the joint profiling of transcription and chromatin accessibility in the same cell providing for enhanced detection of genomic features regulating cell differentiation (12–14).

Kidney organoids differentiated from human pluripotent stem cells (PSCs) have provided unique opportunities to study human kidney development and disease modeling in vitro (15–25). Several established protocols generate organoids containing variable numbers of nephrons, including glomerular podocytes (POD), parietal epithelial cells (PEC), proximal tubules (PT), loop of Henle (LOH), distal nephron epithelial cells (DN) that include early distal tubules (DT) and connecting segment (CS), as well as stromal interstitial cells (ST) through posterior intermediate mesoderm (PIM) induction (26–33). Other protocols have reported methods to generate branching ureteric bud and collecting duct cell types through more rigorous lineage selection (34–39). Studies utilizing single-cell transcriptomics have revealed that kidney organoids contain not only a diverse range of kidney cell types but also off-target cells, such as neurons and melanocytes (40–42). However, how the epigenetic landscape changes during organoid differentiation and how similar the chromatin accessibility patterns are to mature human kidney remain undefined.

Here, we have generated a multimodal atlas of kidney organoid differentiation with combined single-nucleus RNA-sequencing (snRNA-seq) and single-nucleus assay for transposase-accessible chromatin with sequencing (snATAC-seq). We identify cell type-specific chromatin accessibility and predict cis-regulatory links between ATAC peaks and target genes. By direct comparison with human adult kidney (HAK), kidney organoid cell types showed broad chromatin accessibility to their marker gene bodies, but absences of chromatin accessibility in promoter and putative enhancer regions of maturation-related genes, reflecting the epigenetic immaturity of organoid cell types. Changes in transcription

## Significance

During development, changes in chromatin structure regulate gene expression dynamics. A detailed understanding of gene regulatory networks will improve our understanding of organogenesis and disease mechanisms. Using single-nucleus RNA-sequencing (snRNA-seq) and assay for transposase-accessible chromatin with sequencing (snATAC-seq), we comprehensively investigate the single-cell gene regulatory landscape during kidney organoid differentiation.

Author affiliations: <sup>a</sup>Division of Nephrology, Department of Medicine, Washington University in St. Louis School of Medicine, St. Louis, MO 63110; and <sup>b</sup>Department of Developmental Biology, Washington University in St. Louis School of Medicine, St. Louis, MO 63110

Author contributions: Y.Y. and B.D.H. designed research; Y.Y., Y.M., N.L., H.W., K.O., J.H.M., and B.D.H. performed research; Y.Y., Y.M., N.L., H.W., K.O., J.H.M., and B.D.H. analyzed data; and Y.Y. and B.D.H. wrote the paper.

Competing interest statement: B.D.H. is a consultant for Janssen Research & Development, LLC, Pfizer, and Chinook Therapeutics and holds equity in Chinook Therapeutics.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: [humphreysbd@wustl.edu](mailto:humphreysbd@wustl.edu).

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2219699120/-/DCSupplemental>.

Published May 8, 2023.

factor expression during differentiation were associated with enrichment of their corresponding motifs in ATAC peaks and target gene expression. We used this multiomic atlas to identify a putative *HNF1B* gene regulatory network that we validated during organoid differentiation by targeting the promoter and enhancer regions of the *HNF1B* gene with CRISPR interference (CRISPRi), resulting in impaired PT and LOH lineage specification.

## Results

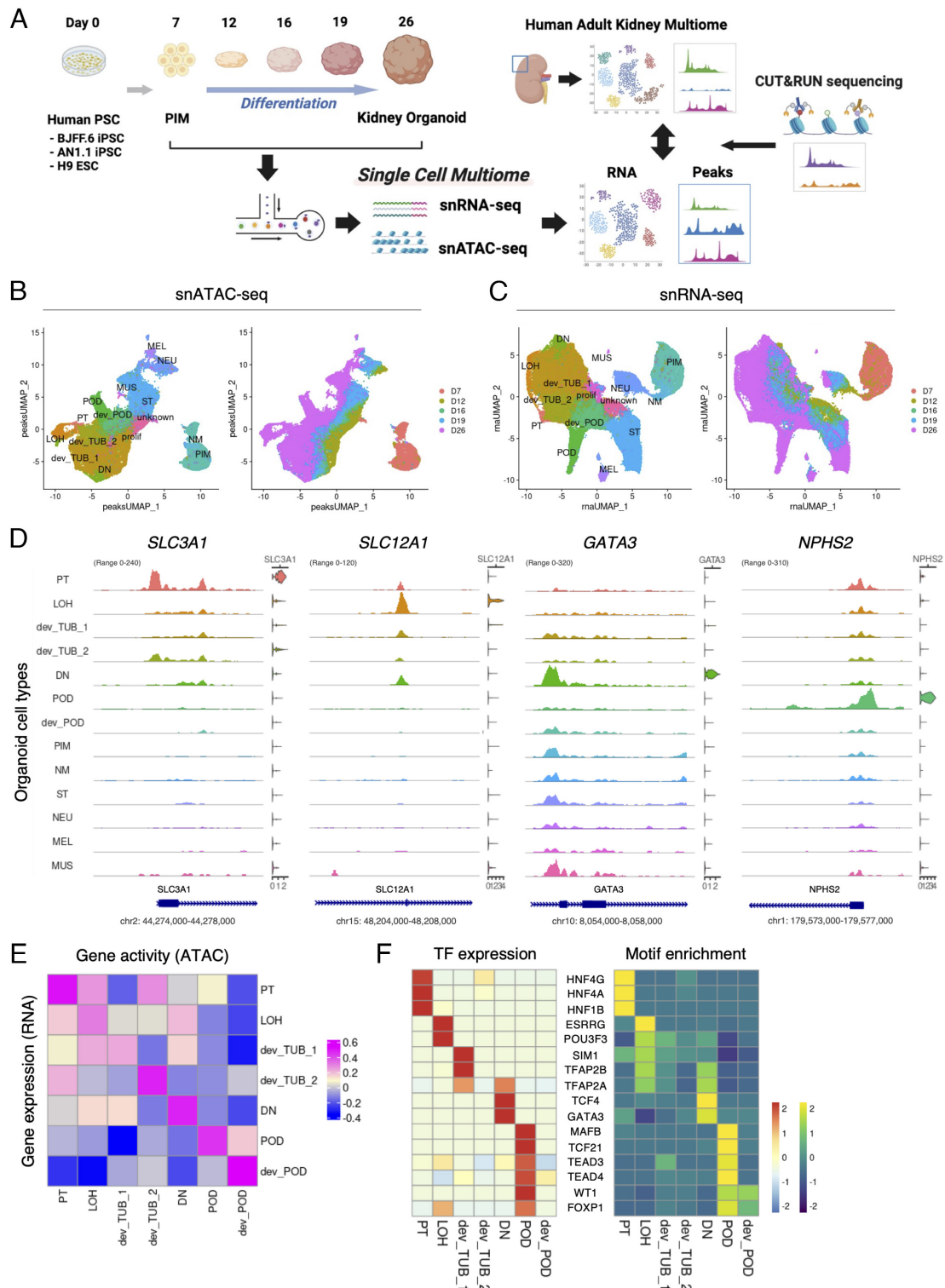
**Single-Cell Multiomic Reconstruction of Kidney Organoid Differentiation.** We induced kidney organoids from three different human PSC lines: BJFF.6, AN1.1 human-induced pluripotent stem cells (hiPSCs), and H9 human embryonic stem cells (hESCs) with a commonly used protocol (27, 43). The resulting kidney organoids showed robust epithelialization and contained multiple nephron segments, including POD (WT1<sup>+</sup>), PT (LTL<sup>+</sup>), and LOH and DN (CDH1<sup>+</sup>) segments (*SI Appendix, Fig. S1 A and B*). We next collected samples at multiple time points from day 7 to day 26 during differentiation and generated snRNA-seq and snATAC-seq libraries from each timepoint using 10X Genomics Multiome kits to simultaneously measure RNA and chromatin accessibility from the same cell (Fig. 1A). After sequencing and quality control, we performed batch correction between samples and merged all time points (*Materials and Methods, SI Appendix, Table S1, and Dataset S1*). The integrated multiome dataset included developing nephron epithelia from earlier time points and at later timepoints contained multiple differentiated epithelial cell types (Fig. 1 B and C and *SI Appendix, Figs. S1 D and E and S2 and Table S2*). As expected, we also observed off-target cells, such as neurons, muscle cells, and melanocytes (Fig. 1 B and C and *SI Appendix, Fig. S1 C–E*), consistent with our previous findings (40). When analyzing cell clusters over time, there was clear progression from a primary mesoderm cluster to intermediary developing clusters at the middle time points and finally to more differentiated cell types at day 26. Each cell type expressed RNA encoding representative marker genes and chromatin accessibility around the transcription start site (TSS) of that marker gene (Fig. 1D and *SI Appendix, Figs. S3 and S4*). We also confirmed that PIM cells at day 7 expressed multiple nephron progenitor cell markers (*SI Appendix, Fig. S5A*). Interestingly, chromatin accessibilities to the distal regulatory peaks of *MEOX1* and *CITED1* genes with high cis-regulatory link scores decreased along differentiation, suggesting that these peaks represent putative enhancers regulating gene expression (*SI Appendix, Fig. S5 B and C*). We calculated gene activity scores from chromatin accessibility and compared this to cognate RNA expression. This revealed good correlations for differentiated cell types, with lesser correlation for developing tubule and LOH (Fig. 1E). Gene activity in the PT also showed moderate correlation to LOH gene expression, and LOH showed moderate correlation to DN.

To identify critical regulators of cell fate during organoid differentiation, we calculated RNA expression and motif enrichment in open chromatin to compile a list of transcription factors (TFs) regulating each cell state (Fig. 1F and *SI Appendix, Table S3*). Many TFs enriched in organoids are also expressed in developing mouse kidneys where they are known to regulate nephrogenesis. For example, *HNF4A* and *HNF1B* are required for PT differentiation and maturation (44–48); *TFAP2B* is essential for DN differentiation (49); *WT1* and *MAFB* regulate POD differentiation and slit diaphragm formation, respectively (50–53). Furthermore, *TCF4* was up-regulated and its motif was enriched in the DN cluster, suggesting high WNT signaling activity in

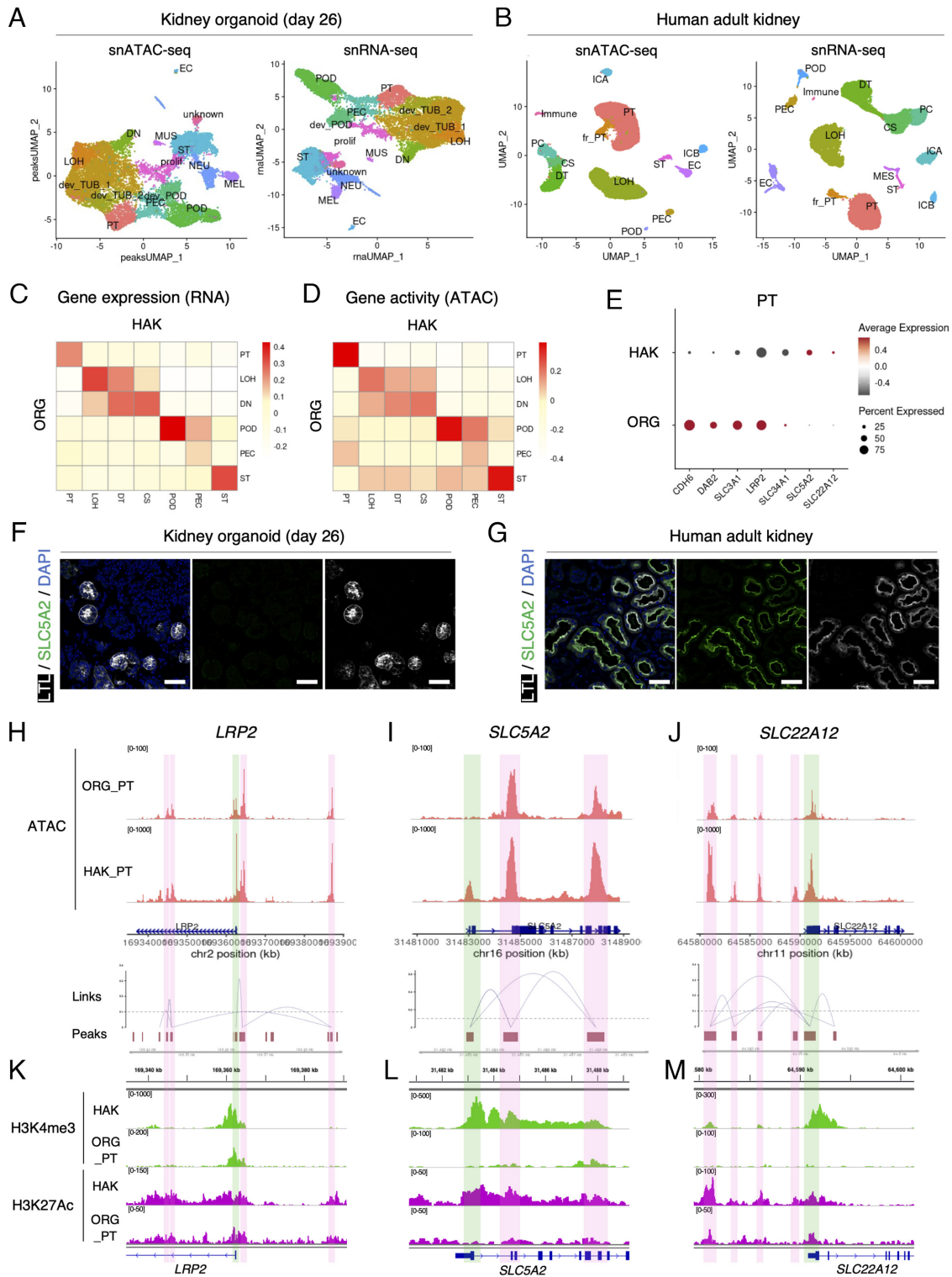
distal nephron, which is consistent with the previous reported WNT/  $\beta$ -catenin gradient in the developing mouse and human kidneys (54). Yap/Taz has been reported to be important for the maintenance of glomerular filtration barrier integrity in mouse podocytes (55, 56). We detected strong gene expression and motif enrichment of *TEAD3* and *TEAD4* in podocytes of human kidney organoids, suggesting the possibility that Hippo pathway signaling plays an important function in human podocyte differentiation. Thus, our multiomic profiling revealed the cell type-specific landscape of gene expression and chromatin accessibility, allowing the prioritization of top-regulatory TFs during human kidney organoid differentiation.

**Comparing Kidney Organoid Epigenetic Features with Human Adult Kidneys.** We next compared day 26 kidney organoids with human adult kidney multiomic datasets (10) to evaluate differences in chromatin accessibility in each cell type (Fig. 1A). The HAK dataset included POD, PT, and LOH cell types but also distal tubules (DT) (*SLC12A3*<sup>+</sup>/*CALB1*<sup>+</sup>), which were not present in the organoids (Fig. 2 A and B and *SI Appendix, Fig. S6A*). HAK also had principal cells (PCs) and intercalated cell type A (ICA) and type B (ICB), which are derived from anterior intermediate mesoderm but not PIM. Each cell type in the HAK expressed appropriate marker genes with corresponding gene activity calculated from the snATAC-seq datasets (*SI Appendix, Fig. S6*). Gene activities in organoids were much less cell specific compared to those in HAK (*SI Appendix, Figs. S6 and S7*). A closer examination of chromatin accessibility peaks for marker genes revealed that the organoid cell types had relatively cell-specific peaks near to promoter regions but low cell-type specificity across the rest of the gene body (*SI Appendix, Fig. S8*). Since gene activity scores reflect the summation of peaks across the whole gene body, this may explain why gene activity scores had lower cell specificity compared to RNA expression in organoid cell types (*SI Appendix, Fig. S6*). When we comprehensively compared gene expression and gene activity between organoids and HAK parenchymal cell types, POD, PT, and ST showed a high correlation in both modalities (Fig. 2 C and D). In contrast, gene expression and gene activity of LOH and DN in organoids showed broad similarity to LOH, DT, and CS in HAK without strong cell type-specific correlation. These results suggest that organoid distal nephron cell types are less differentiated than more proximal cell types.

Next, we compared representative marker gene expression and chromatin accessibility patterns between organoid and HAK PT. Organoid PT expressed the developmental PT marker *CDH6*, whereas this gene was down-regulated in PT from HAK. Typical PT markers such as *LRP2* and *SLC3A1* were also detectable in organoid PT at higher levels than those in HAK PT. Other transport-related genes responsible for reabsorption of glucose (*SLC5A2*, encoding SGLT2) and uric acid (*SLC22A12*, encoding URAT1) were not detected in organoid PT compared to HAK PT (Fig. 2 E–G and *SI Appendix, Fig. S9*). Analysis of chromatin accessibility for *LRP2* revealed that both organoid and HAK PT showed similar peak patterns near the TSS, intronic, and intergenic regions (Fig. 2H). Some peaks with high cis-regulatory link scores calculated by the LinkPeak function (57) are candidate cis-regulatory elements. By contrast, *SLC5A2* entirely lacked promoter accessibility in the organoid PT, while other peaks in the gene body were concordant with HAK (Fig. 2I). A third pattern emerged for the *SLC22A12* gene, where the TSS peaks were similar between organoid and HAK, but the organoid PT lacked most of the upstream putative enhancer peaks (Fig. 2J), suggesting that multiple different epigenetic mechanisms underlie the low expression of PT genes in organoids.



**Fig. 1.** Characterization of gene expression and open chromatin accessibility of human kidney organoid differentiation. (A) Schematic of this study. Samples were collected at multiple time points from day 7 to day 26 during organoid differentiation. The collected samples were dissociated into single cells and nuclei were extracted. snRNA-seq and snATAC-seq libraries were generated from the extracted nuclei using the 10X Genomics single-cell ATAC and gene expression kit. After sequencing, each organoid multiome dataset was merged and processed for analysis. Human adult kidney multiome dataset [Muto et al. (10)] was used to relatively evaluate day 26 kidney organoids. CUT&RUN sequencing was employed to evaluate chemical annotation of ATAC peaks. (B) UMAP plots of chromatin accessible peaks of the merged kidney organoid differentiation multiome dataset. Left plot shows each organoid cell type cluster. Right plot shows origin of time points during differentiation. PT, proximal tubule; LOH, Loop of Henle; dev\_TUB, developing tubule; DN, distal nephron; POD, podocyte; dev\_POD, developing podocyte; PIM, posterior intermediate mesoderm; NM, nascent mesoderm; ST, stromal cell; NEU, neural cell; MEL, melanocyte; MUS, muscle cell; prolifer, proliferating cell; unknown, undefined type of cell. (C) UMAP plots of gene expression of the merged kidney organoid differentiation multiome dataset. Left plot shows each organoid cell type cluster. Right plot shows origin of time point during differentiation. (D) Coverage plots showing ATAC peaks around transcription start site and gene expression levels of *SLC3A1* (PT), *SLC12A1* (LOH), *GATA3* (DN), and *NPHS2* (POD) genes in each cluster of the merged kidney organoid differentiation multiome dataset. (E) Heatmap showing Pearson's correlation coefficient between gene expression and gene activity of each cluster in nephron-lineage cells of the merged kidney organoid differentiation multiome dataset. (F) Heatmap showing gene expression levels (Left) and motif enrichment in ATAC peaks (Right) of top-listed transcription factors in nephron-lineage cells of the merged kidney organoid differentiation multiome dataset.



**Fig. 2.** Comparison of kidney organoid and human adult kidney chromatin accessibility landscapes. (A) UMAP plots of chromatin-accessible peak (*Left*) and gene expression (*Right*) of the day 26 kidney organoid multiome dataset. PT, proximal tubule; LOH, loop of Henle; dev\_TUB, developing tubule; DN, distal nephron; POD, podocyte; dev\_POD, developing podocyte; PEC, parietal epithelial cells in the glomerulus; EC, endothelial cell; ST, stromal cell; NEU, neural cell; MEL, melanocyte; MUS, muscle cell; prolif, proliferating cell; unknown, undefined type of cell. (B) UMAP plots of chromatin-accessible peak (*Left*) and gene expression (*Right*) of the merged human adult kidney multiome dataset. fr-PT, failed repaired proximal tubule; DT, distal tubule; CS, connecting segment; PC, principal cell; ICA, intercalated cell type A; ICB, intercalated cell type B; Immune, immune cell. (C and D) Heatmaps showing correlation between nephron-lineage cells of human adult kidney and kidney organoid (day 26) gene expression (C) and gene activity (D) modalities. (E) Dot plot showing expression levels of representative developing and mature proximal tubule marker genes in the proximal tubule cluster of human adult kidney and kidney organoid (day 26). (F and G) Immunofluorescence images of SLC5A2 (green, SGLT2), LTL (white), and nuclear DAPI (blue) staining in the proximal tubule of kidney organoids (day 26) (F) and human adult kidney (G). (Scale bars indicate 50  $\mu$ m). (H–J) Coverage plots showing ATAC peaks and gene expression level in proximal tubule of human adult kidney and kidney organoid (day 26) (*Upper*) and peaks and peak-to-gene links (*Lower*) of *LRP2* (H), *SLC5A2* (I), and *SLC22A12* (J) genes. Green boxes indicate promoter regions; pink boxes indicate putative enhancer regions. (K–M) Corresponding integrative genomics viewer (IGV) images showing target sites of H3K4me3 (promoter, upper) and H3K27Ac (enhancer and promoter, lower) antibodies in CUT&RUN sequencing with human adult kidney (HAK) and organoid-derived proximal tubule (ORG\_PT) (lower) of *LRP2* (K), *SLC5A2* (L), and *SLC22A12* (M) genes.

To further annotate the gene regulatory landscape of organoids compared to HAK, we next performed cleavage under targets and release using nuclease (CUT&RUN) sequencing (58, 59) using histone modification antibodies with organoid-derived, cell-sorted PT and bulk HAK. This revealed a similar pattern: For PT-associated genes that were highly expressed in organoids, H3K4me3 peaks, which mark active promoters (60), were appropriately detected at the TSS both in bulk HAK and organoid PT (Fig. 2*K*). But for genes such as *SLC5A2* and *SLC22A12* that were poorly expressed in organoid PT, there was a corresponding lack of H3K4me3 at their respective TSSs (Fig. 2*L* and *M*). Similarly, H3K27Ac peaks, which mark active enhancers and promoters (61, 62), were detected at the putative enhancer regions of *LRP2* in both HAK and organoid PT, but absent in the organoid PT at *SLC22A12* putative enhancer regions. These histone biochemical annotations validated the functional features of chromatin-accessible peaks detected in our multiome analysis.

#### Gene Regulatory Dynamics during Kidney Organoid Differentiation.

Next, we examined cell-specific motif enrichment changes during nephron epithelial lineage differentiation. We subset nephron-lineage cells from the organoid differentiation dataset and created a pseudotime trajectory using Monocle 2. The nephron-lineage trajectory showed two major branches: PIM to POD and PIM to renal tubule (TUB) lineages (Fig. 3*A*). A subanalysis of the TUB lineage cells revealed a further divergence into PT and LOH/DN lineages (Fig. 3*B*). These findings are reminiscent of previously reported human fetal nephron trajectories (63). We overlaid sample timepoint with corresponding pseudotime trajectories (*SI Appendix, Fig. S10 A and B*). We next asked how global chromatin accessibility differed between developing and differentiated organoid podocytes and tubule cell types, in comparison to their HAK counterparts. This revealed a clear pattern whereby the developing organoid POD, PT, and LOH had the most unique peaks per cell, with reduced peaks per cell in the more differentiated organoid cell types. However, in comparison to the HAK cell types, all the three organoid cell types had substantially more chromatin accessibility (Fig. 3*C*). This is consistent with broad chromatin accessibility in organoid cell types (*SI Appendix, Figs. S6–S8*) compared to HAK and likely reflects organoid cell immaturity in comparison to their adult kidney counterparts.

Across all organoid cell lineages, motif enrichment changes across pseudotime showed two contrasting patterns: low to high and high to low (Fig. 3*D*). The top-regulating TF motifs (Fig. 1*F*) showed low-to-high patterns in each cell lineage along pseudotime, e.g., *MAFB* and *TCF21* in the POD lineage; *HNF4A* and *HNF4G* in the PT lineage; and *ESRRG* and *TFAP2A* in the LOH/DN lineage (Fig. 3*D*). Interestingly, upregulation of Notch pathway-related motifs, including *HEY1* and *HEYL*, was also observed during later time points of the PT lineage, supporting previous findings implicating the Notch pathway in PT specification (64–66). By contrast, high-to-low pattern motifs included nephron progenitor-related genes such as *SIX1*, *SIX2*, and *MEOX1*. Motifs related to distal nephron-related genes such as *TFAP2A* and *ESRRG* were reduced in the POD lineage, and those of *MAFB* and *HNF4A* were reduced in the LOD/DN lineage. These results indicate that organoid cell types gradually lose accessibility for progenitor-related TFs while acquiring accessibility for cell type-specific TFs during differentiation.

We next focused on the top-regulatory TFs in PT and POD cell types, *HNF1B* and *MAFB*, respectively (Fig. 1*F*). Comprehensive analysis of TF enrichment in the PT lineage cells revealed *HNF1B* as the top-enriched TF during PT differentiation after day 16 (Fig. 3*E*). Along PT differentiation pseudotime, *HNF1B* expression changes varied with enrichment of its corresponding motifs (Fig. 3*F*).

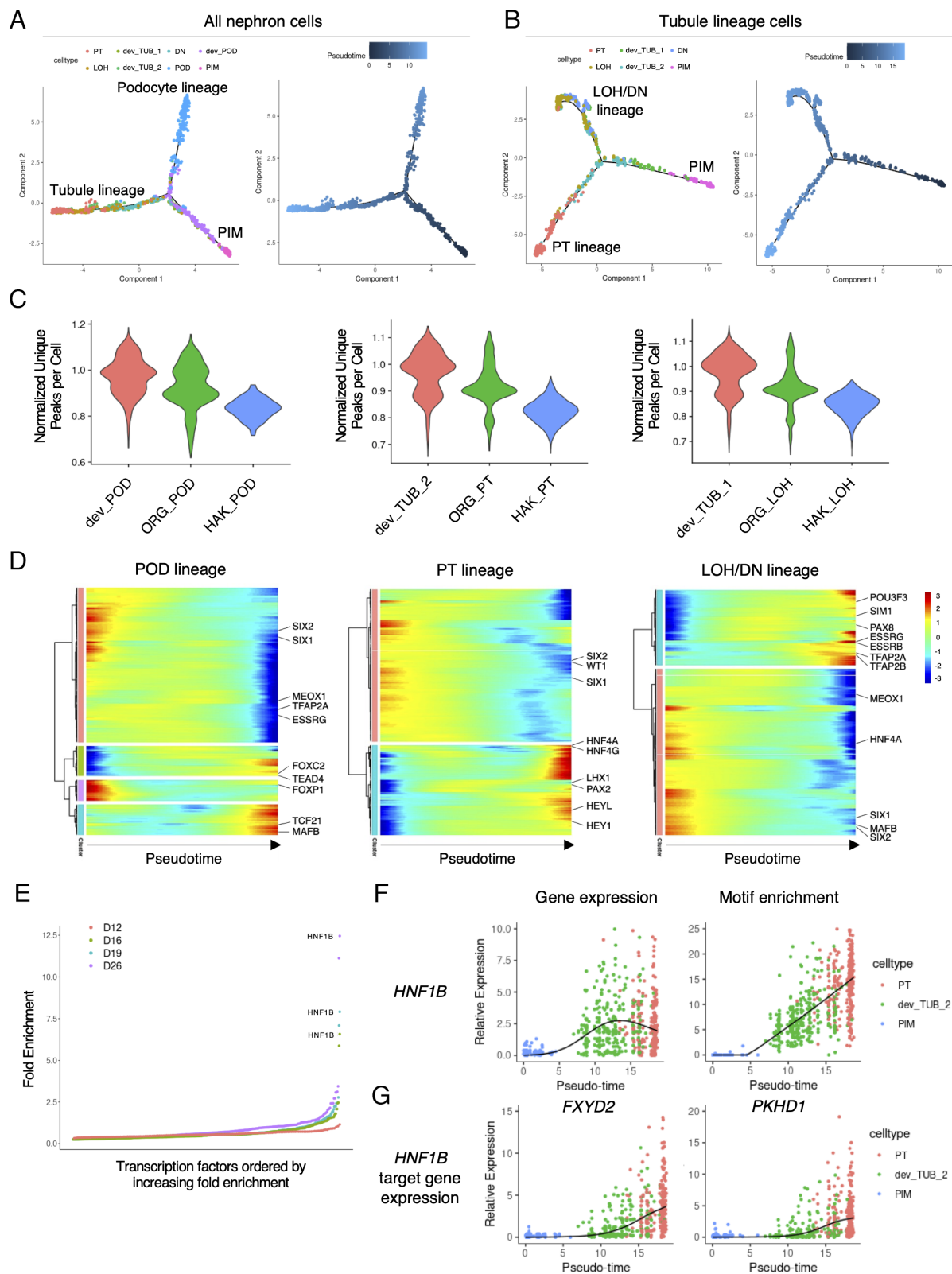
Furthermore, changes in the transcription of *FXYD2* and *PKHD1*, target genes of *HNF1B* (67–69), were concordant with *HNF1B* motif enrichment changes (Fig. 3*G*). Similarly, *MAFB* gene expression and motif enrichment changes varied with *MAGI2* and *NPHS1* expression changes (*SI Appendix, Fig. S10 C and D*), target genes of *MAFB* (70). These results demonstrate the utility of single-cell multiomic analysis of kidney organoid differentiation to reveal cell type differentiation-associated gene regulatory dynamics.

**Interference of the Gene Regulatory Network of *HNF1B* Inhibited PT Differentiation.** We next asked whether we could predict the genomic regulatory features for *HNF1B*, a transcription factor known to be important for PT and LOH differentiation (29, 46–48, 71). We identified three accessible chromatin peaks with high cis-regulatory link scores to *HNF1B* as putative enhancers (Fig. 4*A*). CUT&RUN with anti-H3K27Ac antibody using kidney organoid-derived PT and HAK corroborated that these peaks are potential enhancers regulating *HNF1B* expression (Fig. 4*B*). These three peaks also correlated with elevated *HNF1B* gene expression during organoid differentiation (*SI Appendix, Fig. S11*). To validate the functional contribution of these peaks for gene expression, we epigenetically repressed each promoter and enhancer of *HNF1B* using CRISPR interference (CRISPRi) employing inactive CRISPR enzyme (dCAS9) fused to the KRAB repressor domain (Fig. 4*C*) (72). In cultured human primary renal proximal tubule epithelial cells (RPTEC), CRISPRi with guide RNAs (gRNAs) targeting the promoter region of *HNF1B* resulted in a 70 to 80% reduction of *HNF1B* expression (Fig. 4*D*). The gRNAs targeting enhancer\_2 reduced expression of *HNF1B* by 30 to 50%, whereas targeting of enhancer\_1 or enhancer\_3 reduced *HNF1B* expression by 20 to 40% (Fig. 4*E*). Bulk ATAC-seq confirmed that the chromatin accessibilities of the gRNA target regions were specifically closed after CRISPRi (Fig. S12). These results indicate that all the three potential enhancers regulate *HNF1B* expression in RPTEC.

Next, we examined the functional consequences of modifying *HNF1B* epigenetic regulation during organoid differentiation. We generated hiPSC clones by stably transducing a lentiviral CRISPRi plasmid (dCAS9-KRAB) with gRNAs targeting either promoter\_1 or enhancer\_2-2 of *HNF1B*, and then induced kidney organoids. From day 12 to day 19, control organoids transduced with dCAS9-KRAB plasmid without gRNA substantially up-regulated *HNF1B* expression, whereas CRISPRi organoids targeting both promoter\_1 and enhancer\_2-2 suppressed *HNF1B* expression (Fig. 4*F*). At day 26, CRISPRi organoids targeting the promoter region caused >80% reduction and the enhancer\_2 region caused a 65% reduction in *HNF1B* expression (Fig. 4*F*). Notably, the resultant CRISPRi organoids contained fewer *LTL*<sup>+</sup> PT and *SLC12A1*<sup>+</sup> LOH together with reduced nuclear *HNF1B* expression and presence of CAS9 protein (Fig. 4*G* and *H*, and *SI Appendix, Fig. S13*). These results indicate that the potential enhancers of *HNF1B* identified by our multiome analysis do regulate *HNF1B* expression and functionally contribute to PT and LOH differentiation during human kidney organoid differentiation. These results validate kidney organoids as a valuable tool for investigating gene regulatory mechanisms during human nephrogenesis.

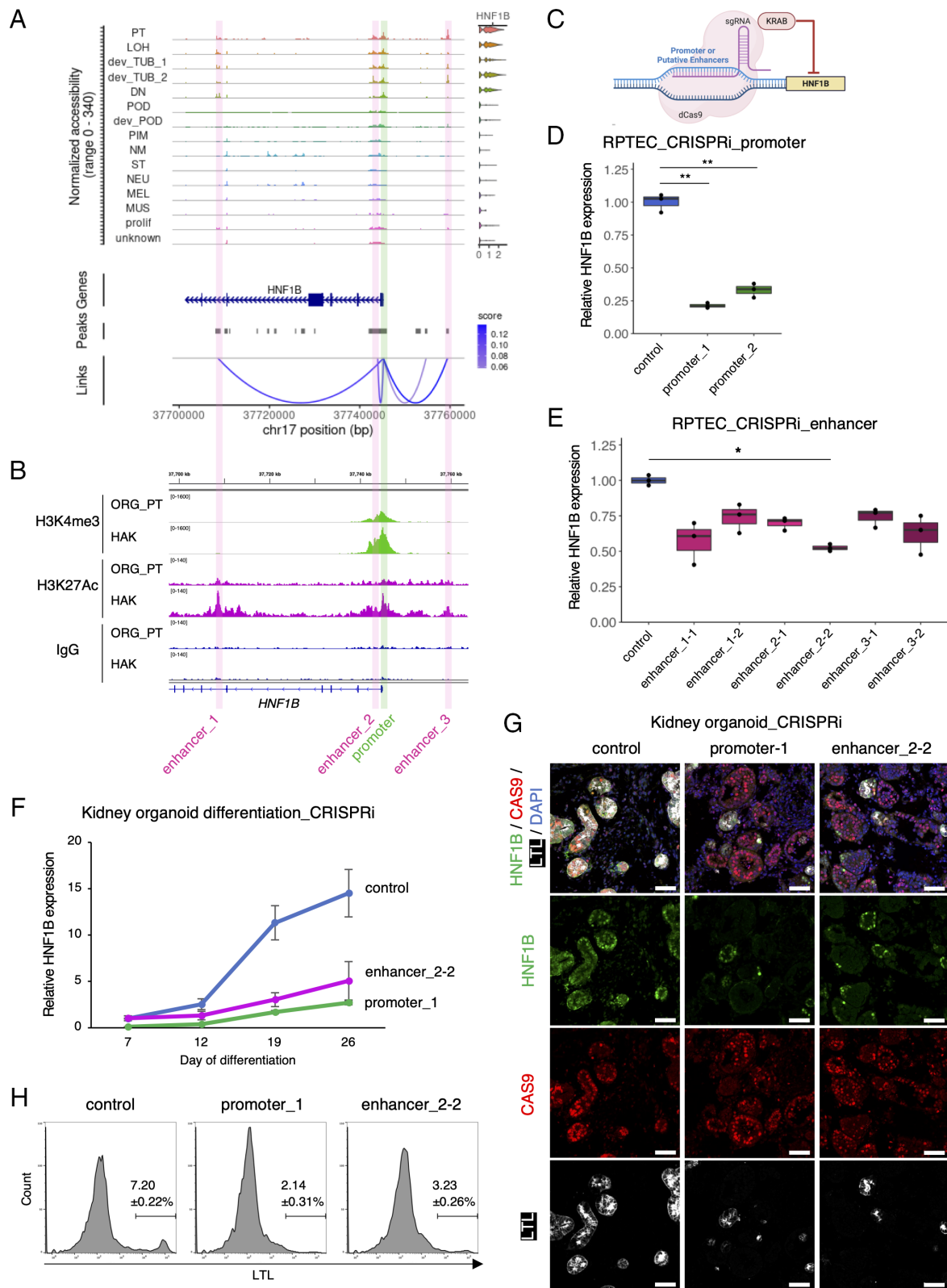
## Discussion

A major challenge to studying the mechanisms that regulate human kidney development is the difficulty of obtaining tissue across developmental stages and the lack of methods to systematically interrogate transcriptional and epigenetic dynamics over time.



**Fig. 3.** Gene regulatory dynamics during kidney organoid differentiation. (A) Pseudotime ordering plots of nephron-lineage cells using snRNA-seq modality showing podocyte and tubule-lineage branches. Colors are based on cell type (Left) and pseudotime (Right). (B) Subordering plots of tubule-lineage cells along pseudotime showing PT and LOH/DN lineage branches. Colors are based on cell type (Left) and pseudotime (Right). (C) Violin plots of number of unique peaks per cell, normalized for read depth, in differentiating (red) and differentiated (green) podocytes (POD, Left), proximal tubules (PT, Middle), and loop of Henle (LOH, Right) of day 26 kidney organoids and those of human adult kidneys (blue). (D) Heatmaps showing motif enrichment changes along pseudotime of podocyte (POD, Left), proximal tubule (PT, Middle), and loop of Henle and distal nephron (LOH/DN, Right) lineages. (E) Transcription factor enrichment in the PT lineage cells at each time point. (F and G) Dynamics of gene expression (F, Left) and motif enrichment (F, Right) of the *HNF1B* gene and their target gene expression changes (G) along pseudotime.





**Fig. 4.** Interference of the gene regulatory network of *HNF1B* inhibited PT differentiation. (A) Coverage plots showing ATAC peaks and gene expression level of *HNF1B* gene of the merged kidney organoid differentiation multiome dataset. Green box indicates promoter region; pink boxes indicate putative enhancer regions. (B) Corresponding integrative genomics viewer (IGV) images showing target sites of H3K4me3 (promoter, Upper), H3K27Ac (enhancer and promoter, Lower), and IgG (negative control) antibodies in CUT&RUN sequencing with day 26 organoid-derived proximal tubule (ORG\_PT) and human adult kidney (HAK). (C) Schematic of *HNF1B* CRISPRi experiment. (D and E) qRT-PCR results of human primary RPTEC with CRISPRi for targeting promoter (D) and putative enhancer (E) regions of *HNF1B*.  $n = 3$  biological replicates. The data are presented as mean  $\pm$  SEM. \*  $P < 0.05$ , \*\*  $P < 0.01$ . (F) qRT-PCR results of kidney organoid differentiation time course with CRISPRi targeting promoter and enhancer\_2 regions of *HNF1B*. Control samples were treated with a dCAS9-KRAB vector without sgRNA.  $n = 4$  biological replicates. The data are presented as mean  $\pm$  SEM. (G) Immunofluorescence images of *HNF1B* (green), LTL (white), CAS9 (red), and nuclear DAPI (blue) staining in kidney organoids (day 26) with CRISPRi for *HNF1B* using control vector (dCAS9-KRAB without sgRNA, Left), sgRNA targeting promoter (Middle), and sgRNA targeting enhancer\_2 (Right). (Scale bars indicate 50  $\mu$ m). (H) Flow cytometry analyses showing LTL-positive cell proportions of kidney organoid (day 26) with CRISPRi using control vector (dCAS9-KRAB without sgRNA, Left), sgRNA targeting promoter (Middle), and sgRNA targeting enhancer\_2 (Right).  $n = 4$  biological replicates. The data are presented as mean  $\pm$  SEM.

Here, we integrated single-cell transcriptome and chromatin accessibility during kidney organoid development in order to provide insight into the mechanisms driving kidney cell differentiation. We find that among the kidney cells represented in organoids, there is broad conservation of cellular differentiation programs in kidney organoids when compared to HAK. Organoid cell types were immature compared to adult, as expected, with distal nephron cell types being more immature than proximal in this differentiation protocol.

We observed increased global chromatin accessibility in tubular progenitors, and with maturation, reduced chromatin accessibility. However, HAK had even less global chromatin accessibility, most likely reflecting a lack of full differentiation in organoids compared to HAK. When we focused on organoid PT, we could define several different patterns of transcription, chromatin accessibility, and histone modification. For highly expressed organoid PT genes such as *LRP2*, chromatin accessibility and H3K4me3 and H3K27Ac marks were all concordant with HAK. But for other PT genes that were not expressed in organoid, some genes such as *SLC5A2* had accessible enhancer regions but inaccessible promoter regions, with variable histone mark concordance. By contrast, *SLC22A12* transcription was also absent in organoid PT, but had accessible promoter chromatin, but not accessible chromatin at putative enhancers when compared to HAK. Taken together, these results paint a complex picture regarding the gene regulatory landscape in organoid cell types. For highly expressed genes, at least, the fact that transcription, chromatin accessibility, and histone marks closely mirror that of HAK indicates that organoids can accurately model complex gene regulatory networks present in mature kidney. Our data validate the use of human kidney organoids to investigate both normal developmental pathways and how pathogenic mutations leading to CAKUT or adult kidney disorders alter cell differentiation and/or cell function.

As validation of our gene regulatory network predictions, we targeted a putative enhancer of *HNF1B* using CRISPRi during organoid differentiation. This resulted in decreased *HNF1B* transcription and impaired PT and LOH differentiation. This result provides a proof of principle for the utility of kidney organoid models to study and decipher the gene regulatory mechanisms controlling kidney development.

There are limitations to our study. First, we compared kidney organoid and human adult kidney single-cell multiomes. Multiomic analysis of human fetal kidneys, although difficult to obtain, will allow for a more precise characterization of the epigenetic status of kidney organoids since kidney organoids most closely resemble fetal kidneys (20, 27, 41, 73). Such an analysis could resolve whether the differences in epigenetic status we observe reflect either discrete developmental states that normally occur during nephrogenesis vs. chromatin accessibility profiles unique to organoid differentiation in vitro and not seen in vivo. Second, we chose a differentiation protocol that in our hands is robust and reproducible, but it does not generate collecting duct cell types and had few endothelial cells. Application of the kind of multiomic analysis performed here on recent protocol improvements will be promising (32). Finally, we validated a single predicted enhancer for the *HNF1B* gene by CRISPRi. Recent advances allow for organoid perturbation screens ranging from dozens of targets to genome wide and these also hold promise (5, 74).

In summary, our study provides a comprehensive investigation of the single-cell gene regulatory landscape during kidney organoid differentiation. We expect that joint single-cell transcriptome and chromatin accessibility datasets will be useful to infer and perturb cell regulatory networks governing kidney cell differentiation.

## Materials and Methods

**Cell Culture.** BJFF.6 and AN1.1 human iPSC lines were obtained from the Genome Engineering & Stem Cell Center at McDonnell Genome Institute, Washington University School of Medicine in St. Louis. The BJFF.6 and AN1.1 lines were reprogrammed from a neonatal foreskin-derived fibroblast sample and a urine sample, respectively, with informed consent from all donors. The H9 human ESC line was obtained from WiCell. The use of all these cell lines has been approved by the Washington University Embryonic Stem Cell Research Oversight (ESCRO) committee. All experiments were approved by an established Institutional Review Board (IRB) protocol approved by Washington University in St. Louis. All PSCs were maintained and expanded in a humidified 5% CO<sub>2</sub> atmosphere at 37 °C in StemFlex medium (Thermo Fisher Scientific, A3349401) on matrigel (Corning, 354277)-coated plate with daily media changes. Cells were passaged every 3 to 4 d by treatment with ReLeSR (STEMCELL Technologies, 05872). The cells were confirmed to be karyotypically normal and *Mycoplasma* free. Experiments were performed below passage 50. 293T cells (ATCC, CRL-3246) were maintained in a humidified 5% CO<sub>2</sub> atmosphere at 37 °C in Dulbecco's modified Eagle's medium (Thermo Fisher Scientific, 11965092) supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, 10437028), 1× GlutaMAX supplement (Thermo Fisher Scientific, 35050061), and 1× penicillin-streptomycin (Thermo Fisher Scientific, 15140122). Human renal proximal tubule epithelial cells (RPEECs; Lonza, CC-2553) were cultured with renal epithelial cell growth medium (Lonza, CC-3190) in a humidified 5% CO<sub>2</sub> atmosphere at 37 °C. 293T cells and RPEECs were passaged by treatment with 0.05% trypsin-EDTA (Thermo Fisher Scientific, 25300054). Experiments were performed on early passages.

**Human Samples.** Nontumor human kidney cortex sample was obtained from a 71-y-old female donor (serum Cr 0.60 mg/dl) from the MTS with informed consent. The participant provided written informed consent in accordance with the Declaration of Helsinki. Samples were frozen or retained in paraffin blocks for future studies. Histologic sections were reviewed by a renal pathologist and laboratory data were abstracted from the medical record. All experiments were approved by an established IRB protocol approved by Washington University in St. Louis.

**Kidney Organoid Induction.** Kidney organoids were generated by Takasato protocol (43) with slight modifications. Briefly, on the next day of passage, PSCs were treated with 8 μM CHIR99021 (TOCRIS, 4423) in basal medium that comprised APEL2 (STEMCELL Technologies, 05275) supplemented with 1.5% Protein-Free Hybridoma Medium II (Thermo Fisher Scientific, 12040077) and 1× antibiotic-antimycotic solution (Corning, 30-004-CI) for 4 d. On day 4, the medium was switched to 200 ng/mL FGF9 (R&D Systems, 273-F9-025) and 1 μg/mL heparin (Sigma-Aldrich, H4784) in the basal medium. On day 7, cells were dissociated into single cells using Accutase (Sigma-Aldrich, SF006) and diluted five-fold with the basal medium with 10 μM Y27632 (TOCRIS, 1254). 2 × 10<sup>5</sup> cells were spun down at 300 g for 2 min (twice with 180° flip) to make cell aggregates in a low cell-binding tube (VWR, 490003-230) and transferred onto a 0.4-μm pore polyester membrane transwell (Corning, 3460). Pellets were incubated with 5 μM CHIR99021 in the basal medium for 1 h, and then cultured with 200 ng/mL FGF9 and 1 μg/mL heparin in the basal medium for 5 d. From day 12 to days 26 to 28, organoids were grown in the basal medium. The medium were changed every other day.

**Nuclei Isolation from Differentiating Kidney Organoids and Library Preparation for Multiome Analysis.** To dissociate organoids into single cells, days 12, 16, 19, and 26 organoids were treated with 0.25% trypsin-EDTA (Thermo Fisher Scientific, 25200056) at 37 °C for 8-10 min. Day 7 posterior intermediate cells were treated with Accutase at 37 °C for 6 min to obtain single cells. Dissociated cells were washed with FACS wash buffer and filtered through a 40-μm strainer (pluriSelect, 43-50040-51). 1 × 10<sup>6</sup> cells were treated with chilled multiome lysis buffer on ice for 5 min to obtain nuclei. The isolated nuclei were washed with multiome wash buffer for a total of three times, then resuspended with chilled multiome nuclei buffer. After nuclei preparation, single-cell libraries were prepared using Chromium Next GEM Single-Cell Multiome ATAC + Gene Expression kit (10X Genomics) following the manufacturer's protocol.

**Kidney Organoid Multiome Data Processing.** Libraries were sequenced on an Illumina NovaSeq instrument. Raw sequencing data were converted into FASTQ files using cellranger-arc mkfastq (v2.0). The FASTQ files were then aligned to the GRCh38 (hg38) reference genome and quantified using cellranger-arc count (v2.0.0). Following sequencing, quality control (QC) was performed on each RNA and ATAC modalities (*SI Appendix, Table S1*). The overall quality of control matrices was high, including molecular counts of each assay, nucleosome signal score, and TSS enrichment scores. We noticed a high percentage of mitochondria in all multiome datasets, which was also observed in the 10× Genomics PBMC multiome dataset using the same nuclei isolation protocol. We reasoned that this was probably due to gentle nuclei isolation treatment, which allowed mitochondria to be attached to the nuclei, and the transposition treatment at 37 °C for 60 min prior to GEM generation. We filtered out low-quality cells and multiplets that showed abnormal QC scores in each dataset (*SI Appendix, Table S2*). The DoubletFinder R package was further used to eliminate cells estimated to be doublets or multiplets. These preprocessing created multiome datasets consisting of quality-assured singlets (*SI Appendix, Table S2*). Singlet datasets of each time point from different PSC clones were merged and subjected to batch effect correction for RNA and ATAC assays using the RunHarmony function in Seurat R package (v.4.1.0). Subsequently, these merged datasets for each time point were then aggregated together to generate an organoid differentiation multiome atlas. This resulted in 56,865 cells after filtering. For a comparison to human adult kidney, we used a merged and batch-corrected day 26 organoid dataset.

**Gene Expression Data Preprocessing and Cell Annotation.** We normalized gene expression UMI count data using SCTransform with regressing out percentage of mitochondrial reads (percent.mt) and read count (nCount\_RNA) in each cell and performed dimensionality reduction by PCA and UMAP embedding ("dim = 1:50") with RunPCA and RunUMAP functions, respectively. Individual clusters were annotated based on the expression of lineage-specific markers (*SI Appendix, Table S2*). Differentially expressed genes among cell types were assessed with the FindMarkers function (Seurat) for transcripts detected in at least 10% of cells using a log-fold-change threshold of 0.25. Bonferroni-adjusted p-values were used to determine significance at an FDR < 0.05.

**DNA Accessibility Data Processing.** ATAC-seq peaks in the dataset were identified on all cells together using MACS2 (75) with the default parameters of Signac's (v1.5.0) CallPeaks function: --gsize 2.7e9 --nomodel False --shift -100 --extsize 200 --d-min 20 --qvalue 0.05 --broad False (57). After removing peaks in nonstandard chromosomes and genomic blacklist regions for the hg38 genome, we quantified counts for the resulting peak set for each cell using the FeatureMatrix function in Signac. A gene activity matrix was constructed by counting ATAC peaks within the gene body and promoter region 2 kb upstream of the transcriptional start site using protein-coding genes annotated in the Ensembl database. The DNA accessibility dataset was then normalized with term-frequency inverse-document-frequency (TFIDF) and was subjected to dimensional reduction by singular value decomposition (SVD) of the TFIDF matrix and UMAP embedding ("dim = 2:50") with RunSVD and RunUMAP functions (Signac). We normalized the number of unique peaks in each cell for read depth by dividing by each cell's cellranger-computed atac\_peak\_region\_fragments metric so that we could compare the number of unique peaks along PT, LOH, and POD differentiation trajectories in day 26 organoid and human adult kidney datasets.

**Correlation of Gene Expression and Gene Activity in Each Nephron Cell Type in the Differentiating Kidney Organoid.** We subset nephron epithelial cells from the kidney organoid differentiation multiome dataset and scaled in each gene expression and gene activity assay. The average gene expression and gene activity values were obtained in each cluster. Pearson's correlation coefficients were calculated among 18,353 genes commonly detected in gene expression and gene activity assays.

**Motif Enrichment and Identification of Key Transcription Factors in Each Cell Type.** Cell type-specific motif enrichment was determined by calculating a per-cell accessibility score for known motifs in the accessible

regions of each cell using the chromVAR package with cisBP motif collections ("human\_pwm\_v2") and the RunChromVAR function in Signac. To identify key transcription factors (TFs) regulating each cell state, we used the presto package to calculate *P*-values based on the Wilcoxon rank-sum test for TFs that yielded significant values for both gene expression data and ChromVAR motif accessibility tests. We then calculated an "AUC" statistic and ranked the TFs for each cell type by taking the average of the AUC values from the two tests. We also used the motif activity scores calculated with chromVAR in order to identify active transcription factors in the PT lineage at each time point. To do this, we computed the average fold change difference in motif activity z-scores for each time point relative to the full dataset with Seurat's FindMarkers function. We then ranked each TF passing FindMarkers' default filters at each time point by fold enrichment for plotting.

**Peak-to-Gene Linkage.** Linkage scores for each peak-gene pair were calculated using the LinkPeaks function with default settings in Signac (57). Briefly, the set of peaks that may regulate each gene was detected by calculating Pearson's correlation coefficients between gene expression and accessibility at peaks within 500 kb of each gene TSS and correcting for bias due to GC content, overall accessibility, and peak size. Significant links were retained above a *P*-value cutoff of 0.05 and a score cutoff of 0.05 and if the gene and open peak were present in at least 10 cells.

**Construction of Pseudotemporal Trajectories.** We used Monocle2 (v2.20.0) (76) to generate pseudotemporal trajectories during nephron epithelial differentiation as described in our previous paper (40). Briefly, we randomly picked equal number of cells from eight types of nephron-lineage clusters. The ordering genes were selected using a semi-supervised approach as follows: They are putatively reported as markers for the kidney developmental state or terminally kidney cell fate; they are specific marker genes differentially expressed in the clusters. We used them to select the ordering genes that covary with these markers using the Monocle function markerDiffTable. We then reduced the data space to two dimensions with DDRTree method and ordered the cells based on their gene expression profiles using the orderCells function. Individual cells were color-coded based on the cell type identity, the collected time points, and pseudotime when the starting state is set as day 7. The trajectory only for tubule lineage was generated in a similar manner using the six types of tubule-lineage clusters.

## Quantification and Statistical Analysis

Details of sample size for each experiment were provided in the figure legends. Data were presented as mean ± SEM. Student's *t* test was used for two-group comparison. One-way ANOVA with the Tukey-Kramer test was used for multiple group comparison. No statistical methods were used to predetermine sample size. Experiments were not randomized, and investigators were not blinded to allocation during library preparation, experiments, or analysis.

**Data, Materials, and Software Availability.** The accession number for sn-multiome and CUT&RUN sequencing datasets in this paper is GEO: [GSE213152](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE213152) (77), and that for bulk ATAC-seq dataset is GEO: [GSE227061](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE227061) (78). Previously published sn-multiome data for human adult kidneys are available in GEO: [GSE151302](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE151302) (79). All data needed to evaluate the conclusions in the paper are present in the paper and/or *SI Appendix*.

**ACKNOWLEDGMENTS.** We thank all members of the Humphreys lab for helpful discussions, and the Genome Technology Access Center and the Genome Engineering & Stem Cell Center at the McDonnell Genome Institute at Washington University School of Medicine for helping in genomic analysis and providing pluripotent stem cell lines, respectively. This work was funded by seed network grant CZF2019-002430 from the Chan Zuckerberg Initiative (to B.D.H.). Additional support was from the Japan Society for the Promotion of Science Overseas Research Fellowships and the Mochida Memorial Foundation for Medical and Pharmaceutical Research (to Y.Y.). B.D.H. holds grant funding from Pfizer.

