

# **Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney**

Yoshiharu Muto<sup>1†</sup>, Parker C. Wilson<sup>2†</sup>, Haojia Wu<sup>1</sup>, Sushrut S. Waikar<sup>3</sup>, Benjamin D. Humphreys<sup>1,4\*</sup>

<sup>1</sup>Division of Nephrology, Department of Medicine, Washington University in St. Louis, St. Louis, MO, USA

<sup>2</sup>Department of Pathology and Immunology, Washington University in St. Louis, St. Louis, MO, USA

<sup>3</sup>Division of Nephrology, Department of Medicine, Boston University Medical School, Boston, MA, USA

<sup>4</sup>Department of Developmental Biology, Washington University in St. Louis, St. Louis, MO, USA

<sup>5</sup>These authors contributed equally.

†These authors contributed equally to this work.

\*Correspondence should be addressed to B.D.H (humphreysbd@wustl.edu)

## Abstract

The integration of single cell transcriptome and chromatin accessibility datasets enables a deeper understanding of cell heterogeneity. We performed single nucleus ATAC (snATAC-seq) and RNA (snRNA-seq) sequencing to generate paired, cell-type-specific chromatin accessibility and transcriptional profiles of the adult human kidney. We demonstrate that snATAC-seq is comparable to snRNA-seq in the assignment of cell identity and can further refine our understanding of functional heterogeneity in the nephron. The majority of differentially accessible chromatin regions are localized to promoters and a significant proportion are closely-associated with differentially expressed genes. Cell-type-specific enrichment of transcription factor binding motifs implicates the activation of NFκB that promotes *VCAM1* expression and drives transition between a subpopulation of proximal tubule epithelial cells. These datasets can be visualized at this resource: <http://humphreyslab.com/SingleCell/>. Our multi-omics approach improves the ability to detect unique cell states within the kidney and redefines cellular heterogeneity in the proximal tubule and thick ascending limb.

**Keywords** (3-5): single cell sequencing, transcriptomics, epigenomics, chromatin, kidney

# Introduction

Single cell RNA sequencing (scRNA-seq) has fostered a greater understanding of the genes and pathways that define cell identity in the kidney<sup>1</sup>. Multiple scRNA-seq atlases of mature human<sup>2–5</sup> and mouse kidney<sup>6,7</sup> have established how transcription contributes to cell type specificity. Recent methods have expanded this approach to single cell profiling of chromatin accessibility<sup>8–10</sup>. Single cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) is an extension of bulk ATAC-seq<sup>11</sup> that employs hyperactive Tn5 transposase to measure chromatin accessibility in thousands of individual cells<sup>9</sup>. Joint profiling by scRNA-seq and scATAC-seq in the adult mouse kidney has provided a framework for understanding how chromatin accessibility regulates transcription<sup>8</sup>, however, the single-cell epigenomic landscape of the human kidney has not been described.

Integration and analysis of multimodal single cell datasets is an emerging field with enormous potential for accelerating our understanding of kidney disease and development<sup>12,13</sup>. Bioinformatics tools can extract unique information from scATAC-seq datasets that is otherwise unavailable by scRNA-seq. Prediction of cell-type-specific cis-regulatory DNA interactions<sup>14</sup> and transcription factor activity<sup>15</sup> are two methods that complement the transcriptional information obtained by scRNA-seq. Long-range chromatin-chromatin interactions play an important role in transcriptional regulation and are influenced by transcription factor binding<sup>16,17</sup>. Chromatin accessibility profiling will help to identify distant regulatory regions that influence transcription via long-range interactions.

The kidney is composed of diverse cell types with distinct subpopulations and single cell sequencing can dissect cellular heterogeneity at high resolution<sup>1</sup>. For example, a small subset of cells within the proximal tubule and Bowman's capsule express vimentin, CD24, and CD133<sup>18,19</sup>. These cells have a distinct morphology and expression profile and undergo expansion after acute kidney injury<sup>19</sup>. Furthermore, they have reported potential for tubular and podocyte regeneration<sup>20–23</sup> and have been implicated in the development of renal cell carcinoma<sup>24</sup>. However, the sparsity of these cells has hampered further characterization. Another example of cellular heterogeneity is seen in the thick ascending limb (TAL). Studies in mouse and human suggest that there are structural and functional differences between medullary and cortical TAL, however, the signaling pathways that drive these differences are not well defined<sup>25</sup>.

Chromatin accessibility is a dynamic process that drives nephron development<sup>26</sup>. Nephron progenitors have distinct chromatin accessibility profiles that change as they differentiate<sup>10,26</sup>. The role of chromatin accessibility in promotion or inhibition of kidney repair and regeneration has important implications for designing therapies for acute and chronic kidney disease<sup>27</sup> and may help to improve directed differentiation of kidney organoids<sup>28</sup>.

We have performed single nucleus ATAC (snATAC-seq) and RNA (snRNA-seq) sequencing to examine how chromatin accessibility can refine our understanding of cell state and function in the mature human kidney. We generated an interactive multimodal atlas encompassing both transcriptomic and epigenomic data (<http://humphreyslab.com/SingleCell/>). Combined snRNA-seq and snATAC-seq analysis improved our ability to detect unique cell states within the proximal tubule and thick ascending limb and redefines cellular heterogeneity that may contribute to kidney regeneration and segment-specific cation permeability.

## Results

### Single cell transcriptional and chromatin accessibility profiling in the adult human kidney

snRNA-seq and snATAC-seq was performed on 5 healthy adult kidney samples (Fig. 1a). Selected patients ranged in age from 50 to 62 years and included men (N=3) and women (N=2). All patients had preserved kidney function (mean sCr = 1.07 mg/dl, eGFR = 64.4 +/- 4.7 ml/min/1.73m<sup>2</sup>). Histologic review showed no significant glomerulosclerosis or interstitial fibrosis and tubular atrophy (Supplementary Table 1). We performed snRNA-seq to determine the cellular composition of our samples, annotate cells based on their transcriptional profiles, and inform our snATAC-seq analysis (Fig. 1a). snRNA-seq identified all major cell types within the kidney cortex (Fig. 1b, Supplementary Fig. 1a) based on expression of lineage-specific markers (Fig. 1c, Supplementary Fig. 1b)<sup>29</sup>. We detected proximal tubule (PT), parietal epithelial cells (PEC), loop of Henle (TAL), distal tubule (DCT1, DCT2), connecting tubule (CNT), collecting duct (PC, ICA, ICB), endothelial cells (ENDO), glomerular cell types (MES, PODO), fibroblasts (FIB), and a small population of leukocytes (LEUK) (Supplementary Table 2, Supplementary Data 1). Notably, there was a subpopulation of proximal tubule that had increased expression of *VCAM1* (PT\_VCAM1). This subpopulation also expressed *HAVCR1* (kidney injury molecule-1), which is a biomarker for acute kidney injury and predictor of long-term renal outcomes<sup>30,31</sup>.

### Integration of single nucleus RNA and ATAC datasets for prediction and validation of ATAC cell type assignments

snATAC-seq captures the chromatin accessibility profile of individual cells<sup>9</sup>. Relatively less is known about cell-type-specific chromatin accessibility profiles; so we leveraged our annotated snRNA-seq dataset to predict snATAC-seq cell types with label transfer<sup>12</sup>. Label transfer was performed by creating a gene activity matrix from the snATAC-seq data, which is a measure of chromatin accessibility within the gene body and promoter of protein-coding genes. Transfer anchors were identified between the 'reference' snRNA-seq dataset and 'query' gene activity matrix followed by assignment of predicted cell types. The distribution of snATAC-seq prediction scores showed that the vast majority of cells had a high prediction score and were confidently assigned to a single cell type (Supplementary Fig. 2). The snATAC-seq dataset was filtered using a 97% confidence threshold for cell type assignment to remove heterotypic doublets. Comparison between snATAC-seq cell type predictions obtained by label transfer (Fig. 1d) and curated annotations of unsupervised clusters (Fig. 1e, f, Supplementary Fig. 1c, d and Supplementary Table 3) indicates that all major cell types were present in both datasets and that snATAC-seq is comparable to snRNA-seq in the detection and assignment of cell identities (Supplementary Fig. 3). We performed downstream analyses with gene-activity-based cell type assignments, which were obtained by unsupervised clustering of the snATAC-seq dataset. Interestingly, snATAC-seq was able to detect two subpopulations within the proximal tubule cluster, which likely represent the proximal convoluted tubule (Fig. 1e, PCT) and the proximal straight tubule (Fig. 1e, PST). PCT showed greater chromatin accessibility in *SLC5A2*, which encodes sodium glucose cotransporter 2 (SGLT2); whereas PST showed greater accessibility in *SLC5A1* (Fig. 1f, Supplementary Fig. 4). SGLT2 reabsorbs glucose in the S1 and S2 segments of the proximal tubule and SGLT1 (*SLC5A1*) is located in S3<sup>32</sup>. The delineation between S1/S2 and S3 was less clear in the snRNA-seq dataset (Supplementary Fig. 4), which suggests that snATAC-seq provides complementary information that may refine cell type assignment; particularly for genes transcribed at low levels or genes that are not detected by snRNA-seq. The chromatin accessibility profile of S1/S2 may be of clinical interest in determining the factors that drive glucose reabsorption, which is the therapeutic target of SGLT2 inhibitors<sup>32</sup>. Together, our multimodal snATAC-seq and snRNA-seq analysis improved our ability to dissect cellular heterogeneity.

## Chromatin accessibility defines cell type

We detected 214,890 accessible chromatin regions among 27,034 cells in the snATAC-seq library. We compared these regions to a previously-published dataset of DNase I-hypersensitive sites (DHS) in bulk glomeruli and tubulointerstitium<sup>33</sup>. DNase hypersensitivity is an alternative measure of chromatin accessibility and approximately 50% of all regions identified by our pipeline were overlapping with a DHS in the glomerulus or tubulointerstitium. The proportion of overlapping regions increased to ~85% when our dataset was filtered for regions contained in at least 10% of nuclei (Supplementary Fig. 5). These data suggest that snATAC-seq is a robust method for the detection of accessible chromatin in the adult kidney.

We used the R package Signac<sup>12</sup> to investigate differences in chromatin accessibility between cell types. Cell types can be distinguished based on whether differentially accessible chromatin regions (DAR) are ‘open’ or ‘closed’ (Fig. 2a, Supplementary Data 2). Approximately 20% (mean proportion=0.203 +/- 0.04) of DAR were closely-associated with differentially expressed genes in their respective cell types (Supplementary Table 4). For example, *LRP2* is a lineage-specific gene expressed in the proximal tubule and a coverage plot in this region shows an increase in number and amplitude of ATAC peaks within its promoter and gene body (Fig. 2a). In fact, the majority of DAR were located in a promoter region within 3kb of the nearest transcriptional start site (Fig. 2b, Supplementary Fig. 6). The second most common location was intronic and the distribution of DAR was relatively conserved across cell types (Fig. 2c). A minority of cell-type-specific differentially expressed genes were closely-associated with a DAR (mean proportion=0.358 +/- 0.07), which raises the question of assigning function to DAR that are not located near differentially expressed genes. Regulatory regions can associate via long-range interactions and a DAR does not necessarily regulate the closest gene<sup>34</sup>. Bioinformatics tools can infer regulatory chromatin interactions and may be useful for assigning function to DAR<sup>14</sup>. Long-range interactions mediate the association between enhancers and promoters via chromatin looping and are regulated in part by transcription factors<sup>16</sup>.

## Chromatin accessibility is associated with cell-type-specific transcription factor activity and chromatin interaction networks

Transcription factors are key determinants of cell fate that drive cellular differentiation in kidney ageing and development<sup>35–37</sup>. Transcription factor ‘activity’ can be predicted for individual cell types based on the presence of binding motifs within differentially accessible chromatin regions (DAR). We used chromVAR<sup>15</sup> to infer transcription-factor-associated chromatin accessibility in our snATAC-seq dataset. We observed that individual cell types can be defined by transcription factor ‘activity’ (Fig. 3a, Supplementary Data 3), suggesting that cell-type-specific transcription factors likely regulate chromatin accessibility. For example, *HNF4A* encodes a key transcription factor that drives proximal tubule differentiation<sup>38</sup>. chromVAR detected an enrichment of HNF4A binding motifs within DAR in the proximal tubule (Fig. 3b, motif activity) that was supported by increased chromatin accessibility in *HNF4A* (Fig. 3b, gene activity) and increased *HNF4A* transcription in the snRNA-seq dataset (Fig. 3b, gene expression). A similar pattern was seen for *TFAP2B*, which regulates development in the distal nephron<sup>39</sup>. There was increased TFAP2B transcription factor ‘activity’ in the thick ascending limb and distal convoluted tubule (Fig. 3b, motif activity), in addition to increased chromatin accessibility in *TFAP2B* (Fig. 3b, gene activity) and increased *TFAP2B* transcription (Fig. 3b, gene expression).

We used the R package Cicero<sup>14</sup> to predict cis-regulatory chromatin interactions for individual cell types. Cis-coaccessibility networks (CCAN) are families of chromatin-chromatin interactions that regulate gene expression by approximating enhancers and promoters. Within *HNF4A*, we observed a robust CCAN in the proximal convoluted tubule with multiple connections (red or blue arcs) between differentially accessible regions (Fig. 3c, red boxes) in the promoter, gene body, and distal 3' region (Fig. 3c). We compared these interactions with the GeneHancer database<sup>40</sup> to determine which connections had been previously-reported in the literature. GeneHancer is a collection of human enhancers and their inferred target genes created using 4 methods: promoter capture Hi-C, enhancer-targeted transcription factors, expression quantitative trait loci, and tissue co-expression correlation between genes and enhancer RNA. The subset of GeneHancer interactions with 'double elite' status is the most stringent set of interactions in the database<sup>40</sup> and the majority of predicted Cicero interactions within 50kb of a cell-type-specific differentially accessible region (DAR) were overlapping with 'double elite' GeneHancer interactions (Fig. 3c, blue arcs). The proportion of Cicero connections present in the 'double elite' GeneHancer database was dependent on the Cicero coaccessibility score, which is a measure of increased confidence of the predicted interaction (Supplementary Fig. 7). The Cicero connections with a lower coaccessibility score (threshold=0.1, mean proportion in GeneHancer=0.68) were less likely ( $p < 0.0001$ , Paired t-test) to be in the GeneHancer 'double elite' database compared to Cicero connections with a higher coaccessibility score (threshold=0.5, mean proportion in GeneHancer=0.75). Within the proximal convoluted tubule, the majority of Cicero connections were either within a promoter region or between a promoter and another location (Fig. 3d) and this distribution was similar in other cell types (Supplementary Fig. 8). In summary, Cicero is a robust method for predicting chromatin-chromatin interactions that may play a role in the regulation of cell-type-specific chromatin accessibility.

### Multi-modal analysis highlights cellular heterogeneity in the thick ascending limb

The thick ascending limb regulates extracellular fluid volume, urinary concentration, and calcium and magnesium homeostasis in the outer medulla and cortex<sup>41</sup>. The majority of bivalent cations are reabsorbed in the cortical segment and are regulated by expression of claudins<sup>42</sup>. Claudins are a family of tight junction proteins that confer segment-specific cation permeability and regulate the reabsorption of Na<sup>+</sup>, K<sup>+</sup>, Cl<sup>-</sup>, Mg<sup>++</sup> and Ca<sup>++</sup><sup>43</sup>. Claudin-10 expression is enriched in the medullary TAL, whereas claudin-16 is expressed predominantly in the cortical TAL<sup>44</sup>. Pathogenic germline variants in *CLDN16* are causative of familial hypomagnesemia, hypercalciuria and nephrocalcinosis, which can lead to end-stage renal disease<sup>45,46</sup>. Interestingly, *Cldn10* deletion causes hypermagnesemia in mice<sup>47</sup>, and even rescues *Cldn16*-deficient mice from hypomagnesemia and hypercalciuria<sup>48</sup>. These data indicate that CLDN10 and CLDN16 may confer opposing segment-specific tight junction cation selectivity. To determine if we could detect subpopulations of cells with variable claudin expression patterns, we performed unsupervised clustering on the thick ascending limb in our snRNA-seq dataset (Supplementary Fig. 9a) to identify 3 groups of cells. There was a group of cells (*SLC12A1+UMOD+*) that expressed cortical TAL markers (*CLDN16*, *KCNJ10* and *PTH1R*) and a second group that expressed the medullary TAL marker, *CLDN10* (Supplementary Fig. 9b). The third group of cells was identified as ascending thin limb (ATL) based on expression of previously-published markers<sup>2</sup>.

We analyzed the thick ascending limb cluster in the snATAC-seq dataset and identified 3 groups of cells that echoed our findings in the snRNA-seq dataset (Supplementary Fig. 9c,d). Our findings suggest that TAL subpopulations can be defined by either transcription or chromatin accessibility profiles. We identified differential transcription factor 'activity' for HNF1B and ESRRB that defined cortical and medullary subpopulations (Supplementary Fig. 9e). Furthermore, HNF1B and ESRRB transcription factor



motifs were enriched in the differentially accessible chromatin regions (DAR) that distinguish between cortical and medullary thick ascending limb populations (Supplementary Fig. 9f). ESRRB is an orphan nuclear receptor with a critical role in early development and pluripotency<sup>49,50</sup> and HNF1B is a homeodomain-containing transcription factor that regulates nephrogenesis. Pathogenic germline *HNF1B* variants are a known cause of autosomal dominant tubulointerstitial kidney disease with hypomagnesemia and hypercalciuria<sup>51</sup>. Interestingly, genetic deletion of *Hnf1b* in mouse kidney was found to increase *Cldn10* expression<sup>52</sup>, suggesting that HNF1B may be involved in the regulation of *CLDN10*. Collectively, our multimodal analysis demonstrates heterogeneity within the TAL at a transcriptomic and chromatin accessibility level and highlights transcription factors that likely contribute to these differences.

### **NF- $\kappa$ B regulates the molecular signature of a subpopulation of proximal tubule that expresses *VCAM1***

We detected a subset of proximal tubule cells that had increased expression and chromatin accessibility of *VCAM1*, which we designated PT\_VCAM1 (Fig. 1). Immunofluorescence studies demonstrated *VCAM1* expression in a scattered distribution amongst proximal tubule epithelium (Fig. 4a). Our single cell studies estimate that PT\_VCAM1 represents ~2% of total cells and 6% of proximal tubular epithelium. Despite the fact that kidney samples originated from patients without kidney injury, the PT\_VCAM1 population showed increased expression of kidney injury molecule-1 (*HAVCR1*, *KIM1*), which is a biomarker that is increased in acute kidney injury<sup>30</sup> and chronic kidney disease<sup>53</sup>. Interestingly, PT\_VCAM1 also expressed *VIM* (vimentin), *CD24*, and *CD133* (*PROM1*) (Supplementary Fig. 10), which is consistent with a previously-described population of cells with progenitor-like features in the proximal tubule<sup>18–20</sup>. These findings suggest that the PT\_VCAM1 cluster may represent an injured or regenerative subpopulation.

We compared the transcriptional profile of PT\_VCAM1 to the remaining proximal tubule to identify differentially expressed genes (Supplementary Data 4). Gene ontology enrichment analysis of the differentially expressed genes showed an enrichment for pathways involved in metabolism, cell migration, angiogenesis, proliferation, and apoptosis. In particular, there was enrichment for genes that control branching morphogenesis of epithelial tubes and the MAPK and Wnt signaling pathways (Supplementary Data 5). These results suggest that the signaling pathways in this subpopulation are distinct from the remaining proximal tubule.

We performed pseudotemporal ordering with Monocle<sup>54</sup> to determine which genes drive the transition from healthy proximal tubule to the PT\_VCAM1 state (Fig. 4b). We identified *VCAM1* and *TPM1* as genes that show increased expression in PT\_VCAM1 cells and *SLC5A12* and *SLC5A4* as genes that show decreased expression (Fig. 4b). *VCAM1* is a key mediator of angiogenesis<sup>55</sup> and *TPM1* encodes tropomyosin 1, which is an actin-binding protein involved in the cytoskeletal contraction<sup>56</sup>. In contrast, *SLC5A12* and *SLC4A4* encode a lactate and bicarbonate transporter. *SLC5A12* and *SLC5A4* are abundantly-expressed in the proximal tubule and *VCAM1* and *TPM1* can be detected in a subset of cells. We constructed a complementary pseudotemporal trajectory with Cicero (13) to examine changes in chromatin accessibility during the transition from PT to PT\_VCAM1. Increased transcription of *VCAM1* and *TPM1* (Fig. 4b) was associated with increased chromatin accessibility within the *VCAM1* gene body and promoter region (Fig. 4c,d). Similarly, decreased transcription of *SLC5A12* and *SLC4A4* (Fig. 4b) was associated with decreased chromatin accessibility (Fig. 4d, Supplementary Fig. 11). We identified transcription factors that likely regulate the transition between proximal tubule and PT\_VCAM1 by assessing chromVAR transcription factor activities. Interestingly, the proximal tubule showed robust

activity of HNF4A, which was decreased in the PT\_VCAM1 cluster and coincided with increased activity of REL and RELA (Fig. 4e). NF- $\kappa$ B is a family of inducible transcription factors that share homology in the Rel domain<sup>57</sup> that has been implicated in the inflammatory response in renal disease<sup>58</sup>. In particular, ischemia-reperfusion injury-induced acute kidney injury activates NF- $\kappa$ B and NF- $\kappa$ B inhibition improves renal function<sup>59</sup>. Consistent with this finding, gene set enrichment analysis<sup>60,61</sup> of the differentially expressed genes in PT\_VCAM1 compared to PT implicated NF- $\kappa$ B signaling (Supplementary Fig. 12). The TNF family of cytokines are well-established activators of NF- $\kappa$ B<sup>62</sup>. To further validate these insights, we used TNF $\alpha$  to induce NF- $\kappa$ B in an *in vitro* model of the proximal tubule (RPTEC). TNF $\alpha$  stimulation increased expression of *VCAM1* and *TPM1* and decreased expression of *SLC5A12* and *SLC4A4* (Fig. 4f). These observations suggest that NF- $\kappa$ B induction may play a role in the transition from proximal tubule to the PT\_VCAM1 cell state.

### **The proportion of PT\_VCAM1 is elevated in acute kidney injury and chronic kidney disease**

We performed deconvolution of bulk RNA-seq obtained from mouse IRI to determine if the proportion of PT\_VCAM1 is related to acute kidney injury. BisqueRNA estimates cell type abundance from bulk RNA-seq using a scRNA-seq reference-based deconvolution<sup>63</sup>. The estimated proportion of PT\_VCAM1 significantly increased 24 hours post-IRI and persisted for at least 7 days; corresponding with a decrease in the proportion of normal PT (Fig. 5a). Interestingly, the estimated proportion of PT\_VCAM1 in the no surgery control mouse kidneys increased in older mice (Fig. 5b) and was accompanied by an increase in leukocytes. These results suggest a role for aging-related chronic inflammation and acute kidney injury in the appearance of PT\_VCAM1. To further characterize the role of PT\_VCAM1 in acute kidney injury, we used a snRNA-seq mouse IRI<sup>64</sup> dataset to predict the corresponding cell type for PT\_VCAM1 in the injured mouse kidney. Label transfer of cell type annotations from mouse IRI to human indicates that the majority of PT\_VCAM1 are related to the failed repair population in the mouse (Failed-repair PT) (Fig. 5c).

We retrieved bulk RNA-seq datasets of healthy and injured human kidneys to estimate the proportion of PT\_VCAM1<sup>63</sup>. We identified 72 non-tumor kidney samples in The Cancer Genome Atlas (TCGA) using the GDC data portal. The TCGA patients had a mean age of 62.5 years (SD=11.9y) and had undergone nephrectomy for renal cell carcinoma. Deconvolution of the non-tumor kidney samples with BisqueRNA<sup>63</sup> estimated that the proportion of PT\_VCAM1 cells was 2.6% (Fig. 5d), which is consistent with our snRNA-seq and snATAC-seq estimates. Next, we analyzed bulk RNA-seq from kidney biopsies of patients with type 2 diabetes<sup>65</sup>. The patients with advanced diabetic nephropathy had a significantly higher proportion of PT\_VCAM1 compared to control or early diabetic nephropathy patients (Fig. 5e), suggesting that PT\_VCAM1 may be related to disease progression in diabetic nephropathy.

### **Discussion**

We performed snRNA-seq and snATAC-seq sequencing in parallel to describe the transcriptional and chromatin accessibility landscape of the adult human kidney. Our analysis demonstrates that snRNA-seq and snATAC-seq are comparable methods for determining cell identity and cell-type-specific chromatin accessibility provides additional information that further elucidates cellular heterogeneity. Multimodal single cell profiling (“multi-omics”)<sup>13</sup> has greatly improved our ability to detect unique cell types and states while introducing a host of bioinformatics challenges and opportunities<sup>12</sup>. In this study, we outline



our integration approach to analyzing paired snRNA-seq and snATAC-seq datasets to highlight functional heterogeneity in the proximal tubule and thick ascending limb.

Studies in mouse and human suggest that there are structural and functional differences between medullary and cortical TAL driven by regional expression patterns of claudins<sup>47,48</sup>. Claudin-10 and claudin-16 regulate paracellular reabsorption of calcium and magnesium in the thick ascending limb. Mice lacking *Cldn16* develop hypercalciuria and hypomagnesemia, which is similar to the phenotype of patients with familial hypomagnesemia with hypercalciuria and nephrocalcinosis (FHHNC) that carry pathogenic variants in *CLDN16*<sup>44</sup>. In contrast, targeted deletion of *Cldn10* in the thick ascending limb results in impaired paracellular sodium permeability and hypermagnesemia<sup>47</sup>. Interestingly, *Cldn10* deletion can partially rescue the phenotype of *Cldn16*-deficient mice (double knockout)<sup>48</sup>. These data suggest that *Cldn10* and *Cldn16* have opposing effects on cation reabsorption in the thick ascending limb and are supported by the observation that *Cldn10* and *Cldn16* are expressed in a mosaic pattern in mice<sup>66</sup>. We observed two distinct subpopulations of *UMOD*+ cells in the thick ascending limb (*CLDN10*+*CLDN16*- and *CLDN10*-*CLDN16*+). In motif enrichment analysis, the *CLDN10*-*CLDN16*+ population had increased transcription factor activity of HNF1B (Supplementary Fig. 9e,f). Pathogenic germline variants in *HNF1B* are causative of autosomal dominant tubulointerstitial kidney disease with hypomagnesemia and hypercalciuria<sup>51</sup>. Deletion of *Hnf1b* the mouse kidney increases *Cldn10* expression, suggesting that *Cldn10* is regulated by HNF1B transcription factor activity. Future studies using human cell lines derived from TAL or kidney organoids may help to validate this hypothesis.

The proximal tubule is the most abundant cell type in the kidney cortex and is divided into segments (S1, S2, S3) with unique functions driven by segment-specific expression of various transporters, including SGLT1 and SGLT2<sup>67</sup>. SGLT2 is a therapeutic target in diabetic nephropathy and the genes and signaling pathways that regulate SGLT2 expression may be of clinical interest<sup>68</sup>. snRNA-seq detected *SLC5A1* (SGLT1) and *SLC5A2* (SGLT2) in the proximal tubule, but lacked the power to clearly distinguish between the S1/S2 segments that express SGLT2 and the S3 segment that expresses SGLT1. In contrast, snATAC-seq was able to separate the S1/S2 and S3 segments based on chromatin accessibility within the gene body and promoter of *SLC5A1* and *SLC5A2*. These data suggest that snATAC-seq may help to further refine segment-specific cell types; particularly those that are defined by genes transcribed at low levels or genes that are not detected by snRNA-seq. Furthermore, snATAC-seq can predict the transcription factors that drive cell-type-specificity, which may improve our understanding of kidney development and directed differentiation of kidney organoids. We used this approach to implicate NFκB signaling in a subpopulation of proximal tubule epithelial cells.

We used snRNA-seq and snATAC-seq to identify a subpopulation of proximal tubule (PT\_VCAM1) that expressed *VCAM1*, *HAVCR1* (KIM-1), vimentin (*VIM*), *PROM1* (CD133), and *CD24*. The PT\_VCAM1 population was also identified in bulk RNA-seq datasets from non-tumor TCGA kidney and human diabetic nephropathy<sup>65</sup>. The proportion of PT\_VCAM1 increased in response to acute and chronic kidney injury in both mouse and human<sup>69</sup>. CD133+CD24+ progenitor-like cells have been previously-described in the human kidney in a scattered distribution<sup>18,19</sup> and *VCAM1* (CD106) expression is present in CD133+CD24+ renal progenitors localized to Bowman's capsule<sup>20</sup>. A separate population of CD133+CD24+CD106- cells are localized to the proximal tubule and both CD133+CD24+CD106+ and CD133+CD24+CD106- cells can engraft in SCID mice to repopulate the tubular epithelium following acute tubular injury<sup>20</sup>. *VCAM1* (*CD106*), *VIM*, *PROM1*, and *CD24* expression was enriched in the PT\_VCAM1 cluster in our snRNA-seq dataset (Supplementary Fig. 10), which differs from the previously-described CD133+CD24+CD106- renal progenitor population localized to the proximal tubule. We used immunofluorescence studies to demonstrate that *VCAM1*+ cells are present in a

scattered distribution within the proximal tubule of human kidneys (Fig. 4a). Comparison of our human data to a mouse snRNA-seq acute kidney injury dataset<sup>64</sup> suggests that PT\_VCAM1 is closely-related to a population of ‘failed-repair’ PT, which has a proinflammatory gene expression signature (Fig. 5c). *HAVCR1* expression in PT\_VCAM1 suggests that PT\_VCAM1 likely represents a subpopulation of PT that is undergoing injury *in situ*, and undergoes expansion in aging and chronic kidney disease (Fig. 5b,d). Pseudotemporal ordering (Fig. 4) indicated that PT\_VCAM1 exist along a continuum with PT (Fig. 4), further supporting the hypothesis that they represent an injured cell state. Motif enrichment analysis showed that PT\_VCAM1 had increased RELA transcription factor activity and NFκB induction by TNFα increased *VCAM1* expression in an *in vitro* model of proximal tubule cells. Our findings suggest that NFκB plays a role in the maintenance of PT\_VCAM1, which may be of clinical interest in designing therapies for acute kidney injury. However, whether proximal tubule repair involves proliferation of a progenitor population or dedifferentiation of mature epithelium still remains controversial<sup>27</sup> and our own results do not support the existence of a fixed intratubular progenitor population<sup>70–72</sup>.

An advantage of snATAC-seq is the ability to measure covariance between accessible chromatin sites to predict cis-regulatory interactions<sup>14</sup>. This approach can link putative regulatory regions with their target genes and has been applied to human pancreatic islets<sup>73</sup>, acute leukemia<sup>74</sup>, and multiple mouse tissues, including: hippocampus<sup>75</sup>, mammary gland<sup>76</sup>, T-cells<sup>77</sup>, and kidney among others<sup>8,9</sup>. In particular, genome wide association study (GWAS) risk loci can be linked to their target genes, which would complement the progress made using chromosome conformation capture (Hi-C)<sup>33</sup>. We generated cell-type-specific cis-coaccessibility networks (CCAN) that had significant overlap with a published database<sup>40</sup>. The remaining interactions may represent the unique chromatin interaction landscape of the kidney. We have made all of our data publicly-available and invite readers to explore cell-type-specific differentially accessible chromatin regions (Supplementary Fig. 13).

The small sample size of this study does not adequately capture the expected heterogeneity of the general population. Furthermore, our study focused on kidney cortex and did not include samples from the medulla. Future studies would benefit from studying diseased kidneys to determine how chromatin accessibility changes with progression. Also, improvements in peak calling algorithms for snATAC-seq data will help to narrow the differentially accessible chromatin regions and identify additional peaks in less common cell types. Despite these limitations, our single-cell multimodal atlas of human kidney redefines cellular heterogeneity of the kidney driven by cell-type-specific transcription factors. Our data enhances the understanding of human kidney biology and provides a foundation for future studies.

## Methods

### Tissue procurement

Non-tumor kidney cortex samples were obtained from patients undergoing partial or radical nephrectomy for renal mass at Brigham and Women's Hospital (Boston, MA) under an established Institutional Review Board protocol. Samples were frozen or retained in OCT for future studies. Histologic sections were reviewed by a renal pathologist and laboratory data was abstracted from the medical record.

### Nuclear dissociation and library preparation

For snATAC-seq, nuclei were isolated with Nuclei EZ Lysis buffer (NUC-101; Sigma-Aldrich) supplemented with protease inhibitor (5892791001; Roche). Samples were cut into < 2 mm pieces, homogenized using a Dounce homogenizer (885302-0002; Kimble Chase) in 2 ml of ice-cold Nuclei EZ Lysis buffer, and incubated on ice for 5 minutes with an additional 2 ml of lysis buffer. The homogenate was filtered through a 40- $\mu$ m cell strainer (43-50040-51; pluriSelect) and centrifuged at 500g for 5 minutes at 4°C. The pellet was resuspended, washed with 4 ml of buffer, and incubated on ice for 5 minutes. Following centrifugation, the pellet was resuspended in Nuclei Buffer (10 $\times$  Genomics, PN-2000153), filtered through a 5- $\mu$ m cell strainer (43-50020-50; pluriSelect), and counted. For snRNA-seq preparation, the RNase inhibitors (Promega, N2615 and Life Technologies, AM2696) were added to the lysis buffer, and the pellet was ultimately resuspended in nuclei suspension buffer (1x PBS, 1% BSA, 0.1% RNase inhibitor)<sup>78</sup>. 10X Chromium libraries were prepared according to manufacturer protocol.

### Single nucleus RNA sequencing bioinformatics workflow

Five snRNA-seq libraries were obtained using 10X Genomics Chromium Single Cell 5' v2 chemistry following nuclear dissociation<sup>78</sup>. Three snRNA-seq libraries (patients 1-3) were prepared for a prior study GSE131882<sup>5</sup>. Libraries were sequenced on an Illumina Novaseq instrument and counted with cellranger v3.1.0 using a custom pre-mRNA GTF built on GRCh38 to include intronic reads. Datasets were aggregated with cellranger v3.1.0 without depth normalization and preprocessed with Seurat v3.0.2<sup>12</sup> to remove low-quality nuclei (Features > 500, Features < 4000, RNA count < 16000, %Mitochondrial genes < 0.8, %Ribosomal protein large or small subunits < 0.4) and DoubletFinder v2.0.2<sup>79</sup> to remove heterotypic doublets (assuming 5% of barcodes represent doublets). The filtered library was normalized with SCTransform<sup>80</sup>, and corrected for batch effects with Harmony v1.0<sup>81</sup>. After filtering, there was a mean of 3997  $\pm$  930 cells per snRNA-seq library and a mean of 1674  $\pm$  913 genes detected per nucleus. Clustering was performed by constructing a KNN graph and applying the Louvain algorithm. Dimensional reduction was performed with UMAP<sup>82</sup> and individual clusters were annotated based on expression of lineage-specific markers. The final snRNA-seq library contained 19,985 cells and represented all major cell types within the kidney cortex (Supplementary Table 1). Differential expression between cell types was assessed with the Seurat FindMarkers function for transcripts detected in at least 20% of cells.

### Single nucleus ATAC sequencing bioinformatics workflow

Five snATAC-seq libraries were obtained using 10X Genomics Chromium Single Cell ATAC v1 chemistry following nuclear dissociation. Libraries were sequenced on an Illumina Novaseq instrument

and counted with cellranger-atac v1.2 (10X Genomics) using GRCh38. Libraries were aggregated with cellranger-atac without depth normalization and processed with Seurat v3.0.2 and its companion package Signac v0.2.1 (<https://github.com/timoast/signac>)<sup>12</sup>. Low-quality cells were removed from the aggregated snATAC-seq library (peak region fragments > 2500, peak region fragments < 25000, %reads in peaks > 15, blacklist ratio < 0.001, nucleosome signal < 4 & mitochondrial gene ratio < 0.25) before normalization with term-frequency inverse-document-frequency (TFIDF). Dimensional reduction was performed via singular value decomposition (SVD) of the TFIDF matrix and UMAP. A KNN graph was constructed to cluster cells with the Louvain algorithm. Batch effect was corrected with Harmony<sup>81</sup>. A gene activity matrix was constructed by counting ATAC peaks within the gene body and 2kb upstream of the transcriptional start site using protein-coding genes annotated in the Ensembl database<sup>83</sup>. The gene activity matrix was log-normalized prior to label transfer with the aggregated snRNA-seq Seurat object using canonical correlation analysis. The aggregated snATAC-seq object was filtered using a 97% confidence threshold for cell type assignment following label transfer to remove heterotypic doublets. The filtered snATAC-seq object was reprocessed with TFIDF, SVD, and batch effect correction followed by clustering and annotation based on lineage-specific gene activity. After filtering, there was a mean of 5408 +/- 1393 nuclei per snATAC-seq library with a mean of 7538 +/- 2938 peaks detected per nucleus. The final snATAC-seq library contained a total of 214,890 unique peak regions among 27,034 nuclei and represented all major cell types within the kidney cortex (Supplementary Table 2). Differential chromatin accessibility between cell types was assessed with the Signac FindMarkers function for peaks detected in at least 20% of cells using a likelihood ratio test. Genomic regions containing snATAC-seq peaks were annotated with ChIPSeeker<sup>84</sup> and clusterProfiler<sup>85</sup> using the UCSC database on hg38<sup>86</sup>.

### **Comparison to previously-published database of DNase hypersensitive sites**

Glomerulus and tubulointerstitial DNase hypersensitive sites (DHS) were downloaded in bed format from Sieber et al<sup>33</sup>. Glomerulus and tubulointerstitial DHS master lists were composed by merging the tissue-specific bed files, converting to a GRanges object with the GenomicRanges package<sup>87</sup>, and collapsing the intervals with the reduce function. cellranger-atac peaks were filtered by the proportion of nuclei containing the snATAC-seq peak and subsequently overlapped with DHS sites.

### **Estimation of transcription factor activity from snATAC-seq data**

Transcription factor activity was estimated using the final snATAC-seq library and chromVAR v1.6.0<sup>15</sup>. The positional weight matrix was obtained from the JASPAR2018 database<sup>88</sup>. Cell-type-specific chromVAR activities were calculated using the RunChromVAR wrapper in Signac v0.2.1 and differential activity was computed with the FindMarkers function. Motif enrichment analysis was also performed on the differential accessible regions with the FindMotif function.

### **Generation of cis-coaccessibility networks with Cicero**

Cis-coaccessibility networks were predicted using the final snATAC-seq library and Cicero v1.2<sup>14</sup>. The snATAC-seq library was partitioned into individual cell types and converted to cell dataset (CDS) objects using the make\_atac\_cds function. The CDS objects were individually processed using the detect\_genes() and estimate\_size\_factors() functions with default parameters prior to dimensional reduction and

conversion to a Cicero CDS object. Cell-type-specific Cicero connections were obtained using the `run_cicero` function with default parameters.

### **Construction of pseudotemporal trajectories with Monocle or Cicero**

Monocle3<sup>54</sup> was used to convert the snRNA-seq dataset into a cell dataset object (CDS), preprocess, correct for batch effects<sup>89</sup>, embed with dimensional reduction and perform pseudotemporal ordering. Cicero<sup>14</sup> was used to generate pseudotemporal trajectories for the snATAC-seq dataset. The CDS was constructed from the snATAC-seq dataset, preprocessed, aligned and embedded. The proximal tubular cells identified with Seurat/Signac were designated as the root cells.

### **Comparison of Cicero co-accessibility Connections to GeneHancer Database**

Cell-type-specific differentially accessible chromatin regions (DAR) were identified with the Signac<sup>12</sup> FindMarkers function using a log-fold-change threshold of 0.25 for peaks present in at least 20% of cells. Cell-type-specific DAR were extended 50kb up- and downstream to create bed files to query the UCSC table browser<sup>86</sup> using the GeneHancer interactions tracks<sup>40</sup>. GeneHancer interactions were compared to cell-type-specific Cicero connections to determine the mean proportion of overlap with increasing Cicero coaccess threshold.

### **Gene ontology enrichment analysis**

Differentially expressed genes in the PT\_VCAM1 cluster (compared to PT) were identified with the FindMarkers function using a log-fold-change threshold of 0.25 for the genes expressed in at least 20% of cells. Genes ontology enrichment was performed with PANTHER (<http://geneontology.org/>)<sup>90,91</sup>.

### **Gene set enrichment analysis**

Differential expressed genes in PT\_VCAM1 cluster (compared to PT) were identified with the FindMarkers function using a log-fold-change threshold of 0.05 for peaks present in at least 5% of cells. The pre-ranked gene list was analyzed with GSEA v4.0.3 (Broad Institute)<sup>60,61</sup>.

### **Deconvolution of bulk RNA-seq data**

For the TCGA (The Cancer Genome Atlas) dataset, HTseq counts and metadata were downloaded from the GDC data portal (<portal.gdc.cancer.gov>) by selecting “kidney”, “TCGA”, “RNA-seq”, and “solid tissue normal”. Bulk RNA-seq counts were normalized with DESeq2<sup>92</sup> and count matrices were deconvoluted with BisqueRNA<sup>63</sup> using snRNA-seq annotations. For the mouse ischemia reperfusion dataset from Liu et al.<sup>69</sup>, a normalized count matrix was downloaded from GSE98622 and converted to human annotations using biomaRt and ensembl prior to deconvolution with BisqueRNA with default parameters. For the diabetic nephropathy dataset<sup>65</sup>, fastq files were downloaded from GSE128736, transcript abundance was quantified with Salmon using GRCh38, count matrices were imported to DESeq2 with tximport, and data was normalized prior to deconvolution with BisqueRNA.



## Inter-species snRNA-seq data comparison

The snRNA-seq dataset for human adult kidneys was converted to mouse annotations using biomaRt and ensembl, and integrated with a mouse IRI snRNA-seq dataset<sup>64</sup> using the FindTransferAnchors function in Seurat. Mouse cell type annotations were transferred to the human dataset.

## Renal proximal tubule epithelial cell culture (RPTEC) and TNF $\alpha$ stimulation

RPTEC (Lonza) were cultured with Renal Epithelium Cell Growth Medium 2 (PromoCell). Cells were maintained in a humidified 5% CO<sub>2</sub> atmosphere at 37°C. Experiments were performed on early passages (passage 2-3). Cells were plated at a density of 1x10<sup>5</sup> cells per well in a 6-well plate, incubated overnight, and subsequently treated with TNF $\alpha$  (R&D systems, 100 ng/ml). Cells were harvested at 24 h or 48 h after treatment.

## RT and real-time PCR analysis

Total RNA was extracted from RPTEC or kidney organoids with the Direct-zol MicroPrep Kit (Zymo) following manufacturer's instructions. The extracted RNA (2 ug) was reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit (Life Technologies). Quantitative PCR (RT-PCR) was performed using iTaq Universal SYBR Green Supermix (Bio-Rad). Data were normalized by the abundance of *GAPDH* mRNA. Primer sequences (sense and antisense, respectively) are as follows:

5'- GACAGTCAGCCGCATCTTCT-3' and 5'- GCGCCCAATACGACCAAATC-3' for *GAPDH*,  
 5'- GGGAAGATGGTCGTGATCCTT-3' and 5'- TCTGGGGTGGTCTCGATTTTA-3' for *VCAM1*,  
 5'- GCCGACGTAGCTTCTCTGAAC-3' and 5'- TTTGGGCTCGACTCTCAATGA-3' for *TPM1*,  
 5'- AGGCAACTTCCCGAGAGTTC-3' and 5'- CCCCAAAGCGGTAGACTTCAG-3' for *SLC5A12*,  
 5'- TGATCGGGAGGCTTCTTCTCT-3' and 5'- GGACCGAAGGTTGGATTCTTG-3' for *SLC4A4*.

## Immunofluorescence studies

Formalin-fixed paraffin embedded tissue sections were deparaffinized and underwent antigen retrieval. Sections were blocked with 1% bovine serum albumin, permeabilized with 0.1% Triton-X100 in PBS and incubated overnight with primary antibodies for VCAM1 (abcam, ab134047) and Biotinylated Lotus Tetragonolobus Lectin (Vector Laboratories, B-1325) followed by staining with secondary antibodies (FITC-, Cy3, or Cy5-conjugated, Jackson ImmunoResearch). Sections were stained with DAPI (4',6'-diamidino-2-phenylindole) and mounted in Prolong Gold (Life Technologies). Images were obtained by confocal microscopy (Nikon C2+ Eclipse; Nikon, Melville, NY).

## Statistical analysis

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. Quantitative data (Fig. 4f) are presented as mean $\pm$ s.d. and were compared between groups with a two-tailed Student's *t*-test unless otherwise indicated. Estimated proportion by deconvolution of RNA-seq data



(Fig. 5a,b,e) were analysed with one-way ANOVA with post hoc Dunnett's multiple comparisons test. A *P* value of <0.05 was considered statistically significant.

## **Acknowledgments**

These experiments were funded by a Chan Zuckerberg Initiative Seed Network Grant and an International Research Fund for Subsidy of Kyushu University School of Medicine Alumni. Y.M. was supported by the Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowships for Research Abroad, a Postdoctoral Fellowship of Uehara Memorial Foundation and a grant from the Nakatomi Foundation.

## **Author contributions**

Y.M. and P.C.W. contributed equally to this study. Y.M., P.C.W. and B.D.H. planned the study. S.S.W. collected the kidney samples. Y.M. prepared the snRNA-seq and snATAC-seq libraries. Y.M., P.C.W. and B.D.H. analysed and interpreted snRNA-seq and snATAC-seq data. Y.M. performed other experiments. H.W. generated the public database for data. Y.M., P.C.W. and B.D.H. wrote the manuscript. B.D.H. coordinated and oversaw the study. All authors discussed the results and commented on the manuscript.

## **Competing financial interests**

The authors declare no competing financial interests.

# References

1. Wilson, P. C. & Humphreys, B. D. Kidney and organoid single-cell transcriptomics: the end of the beginning. *Pediatr Nephrol* **35**, 191–197 (2020).
2. Lake, B. B. *et al.* A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. *Nature Communications* **10**, 1–15 (2019).
3. Young, M. D. *et al.* Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594–599 (2018).
4. Liao, J. *et al.* Single-cell RNA sequencing of human kidney. *Scientific Data* **7**, 1–9 (2020).
5. Wilson, P. C. *et al.* The single-cell transcriptomic landscape of early human diabetic nephropathy. *Proc Natl Acad Sci U S A* **116**, 19619–19625 (2019).
6. Park, J. *et al.* Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* **360**, 758–763 (2018).
7. Karaïskos, N. *et al.* A Single-Cell Transcriptome Atlas of the Mouse Glomerulus. *J. Am. Soc. Nephrol.* **29**, 2060–2068 (2018).
8. Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* (2018) doi:10.1126/science.aau0730.
9. Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324.e18 (2018).
10. Wilson, P. C., Ledru, N. & Humphreys, B. D. Epigenomics and the kidney. *Current Opinion in Nephrology and Hypertension* **29**, 280–285 (2020).
11. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21.29.1–9 (2015).

12. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
13. Method of the Year 2019: Single-cell multimodal omics. *Nature Methods* **17**, 1–1 (2020).
14. Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol. Cell* **71**, 858-871.e8 (2018).
15. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975–978 (2017).
16. Nolis, I. K. *et al.* Transcription factors mediate long-range enhancer–promoter interactions. *PNAS* **106**, 20222–20227 (2009).
17. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the  $\beta$ -globin locus. *Genes Dev.* **20**, 2349–2354 (2006).
18. Lindgren, D. *et al.* Isolation and characterization of progenitor-like cells from human renal proximal tubules. *Am. J. Pathol.* **178**, 828–837 (2011).
19. Smeets, B. *et al.* Proximal tubular cells contain a phenotypically distinct, scattered cell population involved in tubular regeneration. *J. Pathol.* **229**, 645–659 (2013).
20. Angelotti, M. L. *et al.* Characterization of Renal Progenitors Committed Toward Tubular Lineage and Their Regenerative Potential in Renal Tubular Injury. *STEM CELLS* **30**, 1714–1725 (2012).
21. Sagrinati, C. *et al.* Isolation and Characterization of Multipotent Progenitor Cells from the Bowman’s Capsule of Adult Human Kidneys. *JASN* **17**, 2443–2456 (2006).
22. Appel, D. *et al.* Recruitment of Podocytes from Glomerular Parietal Epithelial Cells. *J Am Soc Nephrol* **20**, 333–343 (2009).

23. Ronconi, E. *et al.* Regeneration of Glomerular Podocytes by Human Renal Progenitors. *JASN* **20**, 322–332 (2009).
24. Peired, A. J. *et al.* Acute kidney injury promotes development of papillary renal cell adenoma and carcinoma from renal progenitor cells. *Science Translational Medicine* **12**, (2020).
25. Bankir, L. *et al.* Medullary and cortical thick ascending limb: similarities and differences. *Am. J. Physiol. Renal Physiol.* **318**, F422–F442 (2020).
26. Hilliard, S. *et al.* Defining the dynamic chromatin landscape of mouse nephron progenitors. *Biology Open* **8**, (2019).
27. Chang-Panesso, M. & Humphreys, B. D. Cellular plasticity in kidney injury and repair. *Nat Rev Nephrol* **13**, 39–46 (2017).
28. Wu, H. & Humphreys, B. D. Single Cell Sequencing and Kidney Organoids Generated from Pluripotent Stem Cells. *Clin J Am Soc Nephrol* (2020) doi:10.2215/CJN.07470619.
29. Chen, L. *et al.* Renal-Tubule Epithelial Cell Nomenclature for Single-Cell RNA-Sequencing Studies. *J. Am. Soc. Nephrol.* **30**, 1358–1364 (2019).
30. Yin, C. & Wang, N. Kidney injury molecule-1 in kidney disease. *Ren Fail* **38**, 1567–1573 (2016).
31. Schulz, C.-A. *et al.* Plasma kidney injury molecule-1 (p-KIM-1) levels and deterioration of kidney function over 16 years. *Nephrol. Dial. Transplant.* **35**, 265–273 (2020).
32. Ghezzi, C., Loo, D. D. F. & Wright, E. M. Physiology of renal glucose handling via SGLT1, SGLT2 and GLUT2. *Diabetologia* **61**, 2087–2097 (2018).
33. Sieber, K. B. *et al.* Integrated Functional Genomic Analysis Enables Annotation of Kidney Genome-Wide Association Study Loci. *JASN* **30**, 421–441 (2019).

34. Zhang, J. *et al.* ChIA-PET analysis of transcriptional chromatin interactions. *Methods* **58**, 289–299 (2012).
35. O’Brown, Z. K., Nostrand, E. L. V., Higgins, J. P. & Kim, S. K. The Inflammatory Transcription Factors NFκB, STAT1 and STAT3 Drive Age-Associated Transcriptional Changes in the Human Kidney. *PLOS Genetics* **11**, e1005734 (2015).
36. Schedl, A. Renal abnormalities and their developmental origin. *Nature Reviews Genetics* **8**, 791–802 (2007).
37. Carney, E. Regulation of nephrogenesis. *Nature Reviews Nephrology* **14**, 536–536 (2018).
38. Sasaki, S., Hara, A., Sakaguchi, M., Nangaku, M. & Inoue, Y. Hepatocyte nuclear factor 4α regulates megalin expression in proximal tubular cells. *Biochem Biophys Rep* **17**, 87–92 (2018).
39. Chambers, B. E. *et al.* Tfap2a is a novel gatekeeper of nephron differentiation during kidney development. *Development* **146**, (2019).
40. Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**, (2017).
41. Mount, D. B. Thick Ascending Limb of the Loop of Henle. *CJASN* **9**, 1974–1986 (2014).
42. Blaine, J., Chonchol, M. & Levi, M. Renal Control of Calcium, Phosphate, and Magnesium Homeostasis. *Clin J Am Soc Nephrol* **10**, 1257–1272 (2015).
43. Mineta, K. *et al.* Predicted expansion of the claudin multigene family. *FEBS Letters* **585**, 606–612 (2011).
44. Will, C. *et al.* Targeted deletion of murine Cldn16 identifies extra- and intrarenal compensatory mechanisms of Ca<sup>2+</sup> and Mg<sup>2+</sup> wasting. *American Journal of Physiology-Renal Physiology* **298**, F1152–F1161 (2010).

45. Simon, D. B. *et al.* Paracellin-1, a Renal Tight Junction Protein Required for Paracellular Mg<sup>2+</sup> Resorption. *Science* **285**, 103–106 (1999).
46. Konrad, M. *et al.* Mutations in the Tight-Junction Gene Claudin 19 (CLDN19) Are Associated with Renal Magnesium Wasting, Renal Failure, and Severe Ocular Involvement. *Am J Hum Genet* **79**, 949–957 (2006).
47. Breiderhoff, T. *et al.* Deletion of claudin-10 (Cldn10) in the thick ascending limb impairs paracellular sodium permeability and leads to hypermagnesemia and nephrocalcinosis. *PNAS* **109**, 14241–14246 (2012).
48. Breiderhoff, T. *et al.* Deletion of claudin-10 rescues claudin-16-deficient mice from hypomagnesemia and hypercalciuria. *Kidney Int.* **93**, 580–588 (2018).
49. Martello, G. *et al.* Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* **11**, 491–504 (2012).
50. Festuccia, N., Owens, N. & Navarro, P. Esrrb, an estrogen-related receptor involved in early development, pluripotency, and reprogramming. *FEBS Lett.* **592**, 852–877 (2018).
51. Verhave, J. C., Bech, A. P., Wetzels, J. F. M. & Nijenhuis, T. Hepatocyte Nuclear Factor 1 $\beta$ –Associated Kidney Disease: More than Renal Cysts and Diabetes. *JASN* **27**, 345–353 (2016).
52. Kompatscher, A. *et al.* Transcription factor HNF1 $\beta$  regulates expression of the calcium-sensing receptor in the thick ascending limb of the kidney. *Am. J. Physiol. Renal Physiol.* **315**, F27–F35 (2018).
53. Nakagawa, S. *et al.* Molecular Markers of Tubulointerstitial Fibrosis and Tubular Cell Damage in Patients with Chronic Kidney Disease. *PLoS One* **10**, (2015).
54. Trapnell, C. *et al.* Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions. *Nat Biotechnol* **32**, 381–386 (2014).



55. Reinders, M. E. J., Rabelink, T. J. & Briscoe, D. M. Angiogenesis and Endothelial Cell Repair in Renal Disease and Allograft Rejection. *JASN* **17**, 932–942 (2006).
56. Rajan Sudarsan *et al.* Molecular and Functional Characterization of a Novel Cardiac-Specific Human Tropomyosin Isoform. *Circulation* **121**, 410–418 (2010).
57. Zhang, H. & Sun, S.-C. NF- $\kappa$ B in inflammation and renal diseases. *Cell Biosci* **5**, (2015).
58. Sanz, A. B. *et al.* NF- $\kappa$ B in Renal Inflammation. *JASN* **21**, 1254–1262 (2010).
59. Markó, L. *et al.* Tubular Epithelial NF- $\kappa$ B Activity Regulates Ischemic AKI. *JASN* **27**, 2658–2669 (2016).
60. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545–15550 (2005).
61. Mootha, V. K. *et al.* PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics* **34**, 267–273 (2003).
62. Hayden, M. S. & Ghosh, S. Regulation of NF- $\kappa$ B by TNF Family Cytokines. *Semin Immunol* **26**, 253–266 (2014).
63. Jew, B. *et al.* Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nature Communications* **11**, 1971 (2020).
64. Kirita, Y., Wu, H., Uchimura, K., Wilson, P. C. & Humphreys, B. D. *Cell profiling of mouse acute kidney injury reveals conserved cellular responses to injury.*  
<http://biorxiv.org/lookup/doi/10.1101/2020.03.22.002261> (2020)  
doi:10.1101/2020.03.22.002261.

65. Fan, Y. *et al.* Comparison of Kidney Transcriptomic Profiles of Early and Advanced Diabetic Nephropathy Reveals Potential New Mechanisms for Disease Progression. *Diabetes* **68**, 2301–2314 (2019).
66. Milatz, S. *et al.* Mosaic expression of claudins in thick ascending limbs of Henle results in spatial separation of paracellular Na<sup>+</sup> and Mg<sup>2+</sup> transport. *PNAS* **114**, E219–E227 (2017).
67. Curthoys, N. P. & Moe, O. W. Proximal Tubule Function and Response to Acidosis. *CJASN* **9**, 1627–1638 (2014).
68. Maki, T. *et al.* Amelioration of diabetic nephropathy by SGLT2 inhibitors independent of its glucose-lowering effect: A possible role of SGLT2 in mesangial cells. *Sci Rep* **9**, 4703 (2019).
69. Liu, J. *et al.* Cell-specific translational profiling in acute kidney injury. *J. Clin. Invest.* **124**, 1242–1254 (2014).
70. Humphreys, B. D. *et al.* Repair of injured proximal tubule does not involve specialized progenitors. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9226–9231 (2011).
71. Kusaba, T., Lalli, M., Kramann, R., Kobayashi, A. & Humphreys, B. D. Differentiated kidney epithelial cells repair injured proximal tubule. *Proc Natl Acad Sci U S A* **111**, 1527–1532 (2014).
72. Chang-Panesso, M. *et al.* FOXM1 drives proximal tubule proliferation during repair from acute ischemic kidney injury. *J. Clin. Invest.* **129**, 5501–5517 (2019).
73. Rai, V. *et al.* Single-cell ATAC-Seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell-specific type 2 diabetes regulatory signatures. *Mol Metab* **32**, 109–121 (2020).

74. Granja, J. M. *et al.* Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nature Biotechnology* **37**, 1458–1465 (2019).
75. Sinnamon, J. R. *et al.* The accessible chromatin landscape of the murine hippocampus at single-cell resolution. *Genome Res.* **29**, 857–869 (2019).
76. Chung, C.-Y. *et al.* Single-Cell Chromatin Analysis of Mammary Gland Development Reveals Cell-State Transcriptional Regulators and Lineage Relationships. *Cell Reports* **29**, 495-510.e6 (2019).
77. ElTanbouly, M. A. *et al.* VISTA is a checkpoint regulator for naïve T cell quiescence and peripheral tolerance. *Science* **367**, (2020).
78. Wu, H., Kirita, Y., Donnelly, E. L. & Humphreys, B. D. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *J. Am. Soc. Nephrol.* **30**, 23–32 (2019).
79. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Systems* **8**, 329-337.e4 (2019).
80. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *bioRxiv* 576827 (2019) doi:10.1101/576827.
81. Korsunsky, I. *et al.* Fast, sensitive, and accurate integration of single cell data with Harmony. *bioRxiv* 461954 (2018) doi:10.1101/461954.
82. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4314.
83. Cunningham, F. *et al.* Ensembl 2019. *Nucleic Acids Res* **47**, D745–D751 (2019).

84. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
85. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* **16**, 284–287 (2012).
86. Haeussler, M. *et al.* The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* **47**, D853–D858 (2019).
87. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology* **9**, e1003118 (2013).
88. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).
89. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
90. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* **25**, 25–29 (2000).
91. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
92. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, (2014).

## Figure Legends

### Figure 1. Single Cell Transcriptional and Chromatin Accessibility Profiling on the human adult kidneys

(a) Graphical abstract of experimental methodology. n=5 human adult kidneys were analyzed with snRNA-seq and snATAC-seq. (b) Umap plots of snRNA-seq dataset. PT, proximal tubule; PT\_VCAM1, subpopulation of proximal tubule with VCAM1 expression; PEC, parietal epithelial cells; TAL, thick ascending limb of loop of Henle; DCT, distal convoluted tubule; CNT, connecting tubule; PC, principle cells, ICA, Type A intercalated cells; ICB, Type B intercalated cells; PODO, podocyte; ENDO, endothelial cells; MES, mesangial cells, FIB, fibroblasts; LEUK, leukocytes. (c) Dotplots of snRNA-seq dataset showing gene expression patterns of cluster-enriched markers. (d) Multi-omics integration strategy for processing the snATAC-seq dataset. Following integration and label transfer, the snATAC-seq dataset was filtered using a 97% prediction score threshold for cell type assignment. (e) Umap plots of snATAC-seq dataset with gene activities-based cell type assignments. PCT, proximal convoluted tubule; PST, proximal straight tubule. (f) Dotplots of snATAC-seq dataset showing gene activity patterns of cell type markers.

### Figure 2. Distribution of cell type-specific chromatin accessible regions

(a) Heatmap of average number of Tn5 cut sites within a differentially accessible region (DAR) for each cell type (left). Fragment coverage (frequency of Tn5 insertion) around the DAR (DAR +/-50 Kb) on the LRP2 gene promoter is shown (right). (b) Pie plot of genomic annotations for all DAR in the dataset. (c) Bar plot of annotated DAR locations for each cell type.

### Figure 3. Cell-type-specific transcription factor activity and chromatin interaction networks

(a) Heatmap of average chromVAR motif activity for each cell type. (b) Umap plot displaying chromVAR motif activity (left), gene activity (middle) and gene expression (right) of *HNF4A* or *TFAP2B*. (c) Cis-coaccessibility networks (CCAN, red or blue arcs) near the *HNF4A* locus in the proximal convoluted tubule with multiple connections between differentially accessible regions (red boxes). DAR overlapping with high-confidence GeneHancer interactions are shown as blue arcs. Fragment coverage (frequency of Tn5 insertion) and called ATAC peaks are shown in the lower half. *HNF4A* gene track is shown along the bottom of the image. (d) Genomic features of Cis-coaccessibility networks (CCAN) in the PCT.

### Figure 4. Identification and characterization of previously unrecognized PT subpopulation

(a) Umap plot displaying *VCAM1* gene expression in the snRNA-seq dataset (left), and representative immunohistochemical images of VCAM1 (red) or LTL (Lotus tetragonolobus lectin, green) in the adult kidney (n = 3 patients). Arrowheads indicate the VCAM1+ PT. *VCAM1* was expressed in PEC and a subpopulation of LTL+ PT. Scale bar indicates 100  $\mu$ m (upper right) or 20  $\mu$ m (lower right). (b) Pseudotemporal trajectory from PT to PT\_VCAM1 using snRNA-seq was generated with Monocle3 (left), and gene expression dynamics along a pseudotime trajectory from PT to PT\_VCAM1 are shown (right); *VCAM1* (upper left), *TPM1* (upper right), *SLC5A12* (lower left) and *SLC4A4* (lower right). (c) Fragment coverage (frequency of Tn5 insertion) around the representative DAR (DAR +/-5000 bp) in *VCAM1* locus. (d) Pseudotime trajectory from PT to PT\_VCAM1 using snATAC-seq was generated with

Cicero (right). Chromatin accessibility dynamics along the pseudotemporal trajectory from PT to PT\_VCAM1 are shown (left). chr1:100719411-100719996 (*VCAM1* promoter, upper left); chr15:63040511-63045764 (*TPM1* promoter, upper right), chr11:26714753-26720418 (*SLC5A12* gene body, lower left) and chr4:71338336-71340367 (*SLC4A1* gene body, lower right). (e) Feature plot of single cell chromVAR motif activity of RELA and HNF4A in the entire dataset or PT/PT\_VCAM1 subset. (f) RT and real-time PCR analysis of mRNAs for *VCAM1*, *TPM1*, *SLC5A12* and *SLC4A4* in RPTEC treated with TNF $\alpha$  (100 ng/ml) for 24 h or 48 h. \* $P < 0.05$  (Student's  $t$  test). Bar graphs represent the mean and error bars are the s.d.

### Figure 5. The estimated proportion of VCAM1+ PT increases in acute and chronic kidney disease

(a, b) Deconvolution analysis of bulk RNA-seq mouse kidney IRI dataset (GSE98622) with BisqueRNA. Sham control and IRI (a), or no surgery control (b). (c) Inter-species data integration was performed between mouse IRI snRNA-seq (GSE139107) and human snRNA-seq with Seurat (left). PT and PT\_VCAM1 from human snRNA-seq (middle) are label-transferred from mouse IRI snRNA-seq, and the frequencies of predicted cell types are shown on the heatmap (right). (d) Deconvolution analysis of bulk RNA-seq TCGA non-tumor kidney data (e) Deconvolution analysis of bulk RNA-seq human diabetic nephropathy (DN) data (GSE142025) with BisqueRNA. Box-and-whisker plots depict the median, quartiles and range. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.005$ , one-way ANOVA with post hoc Dunnett's multiple comparisons test.



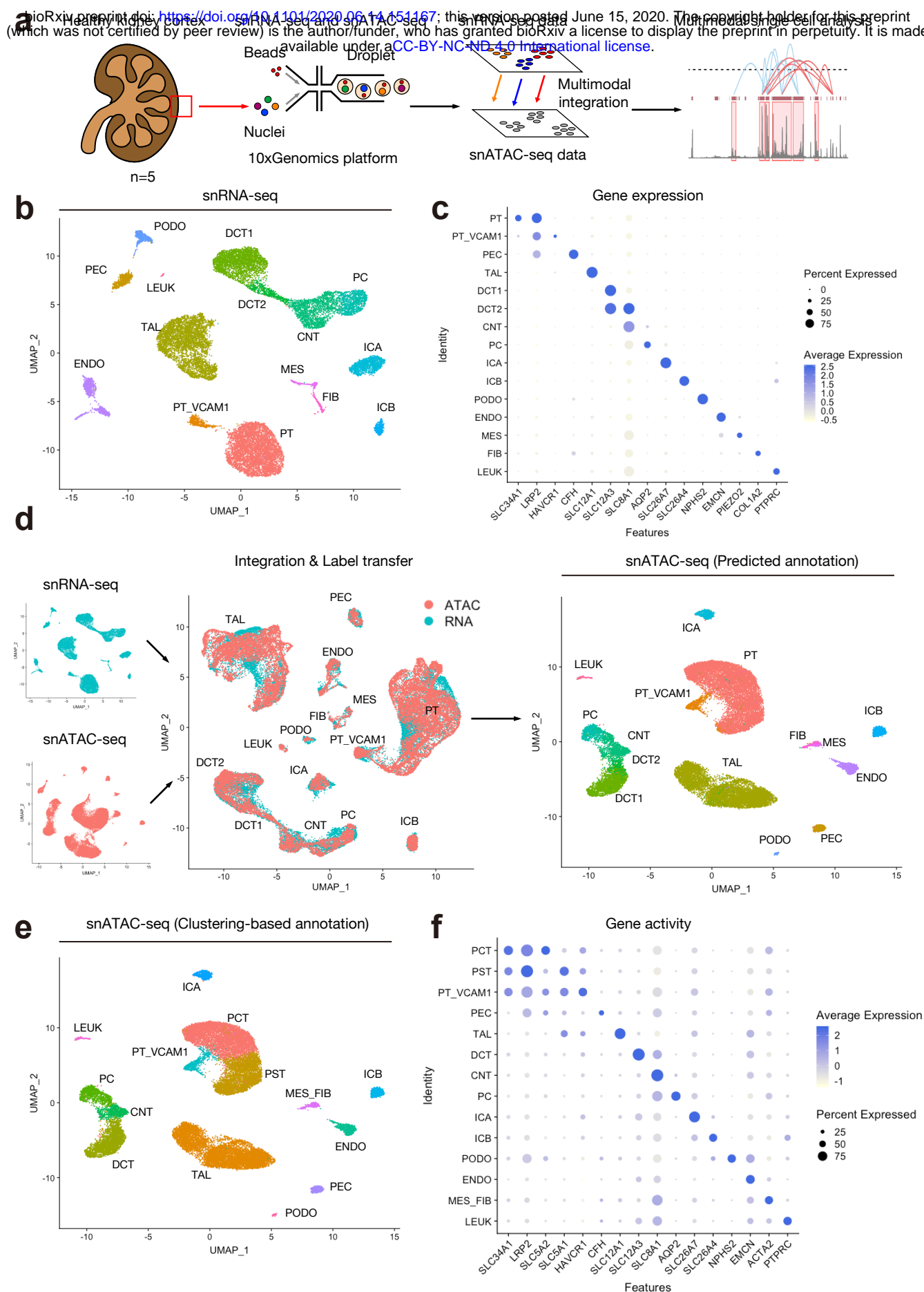
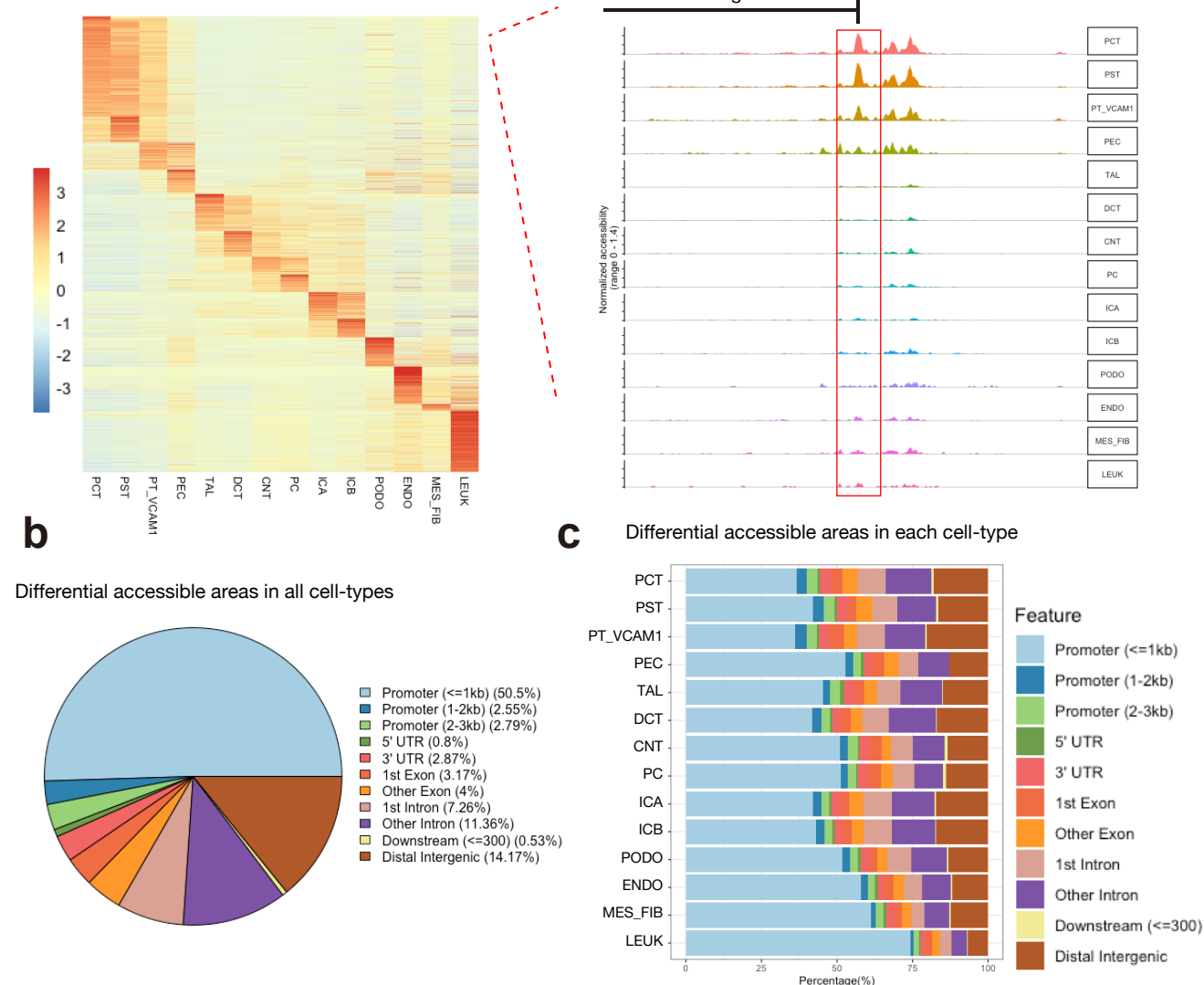


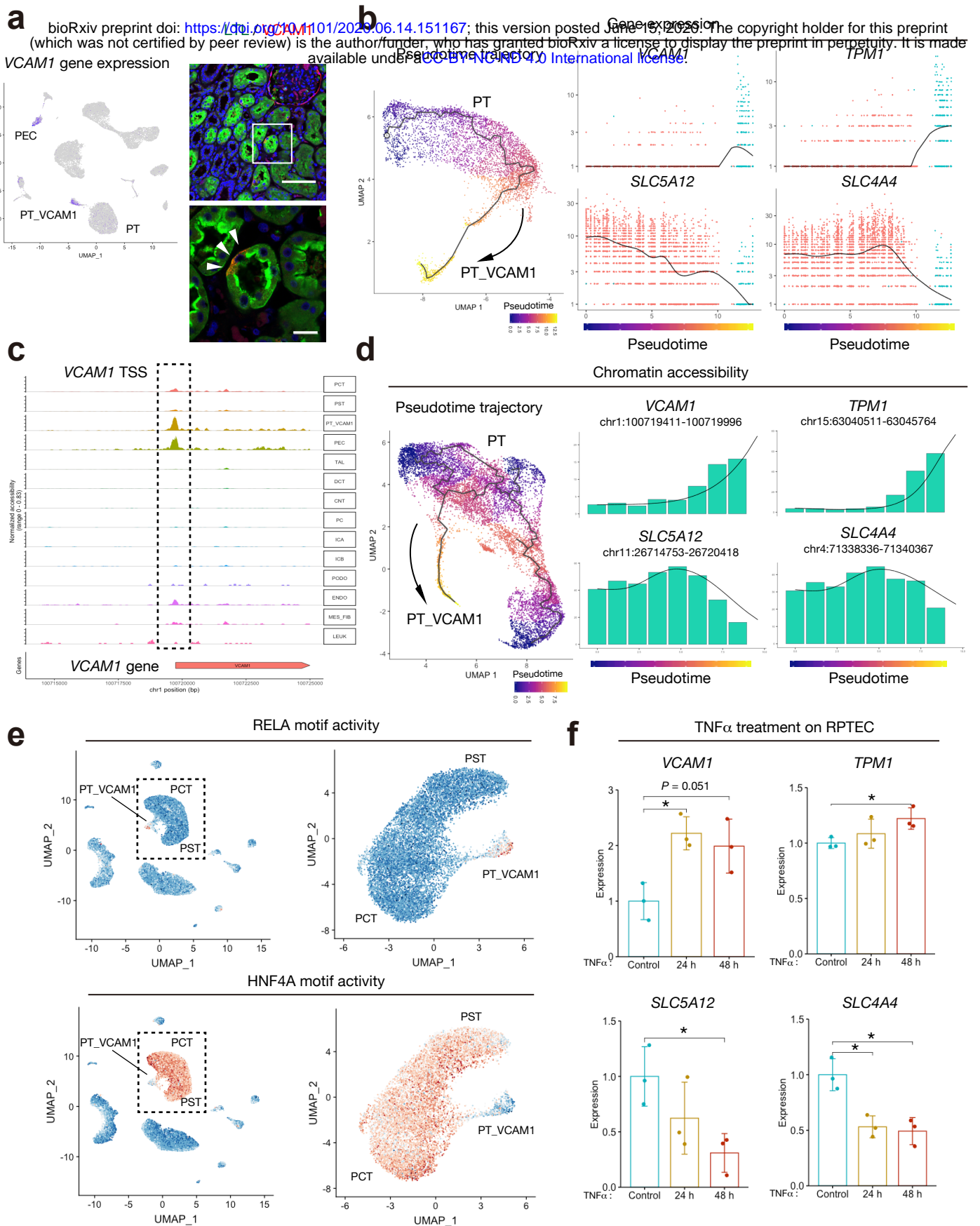
Figure 1



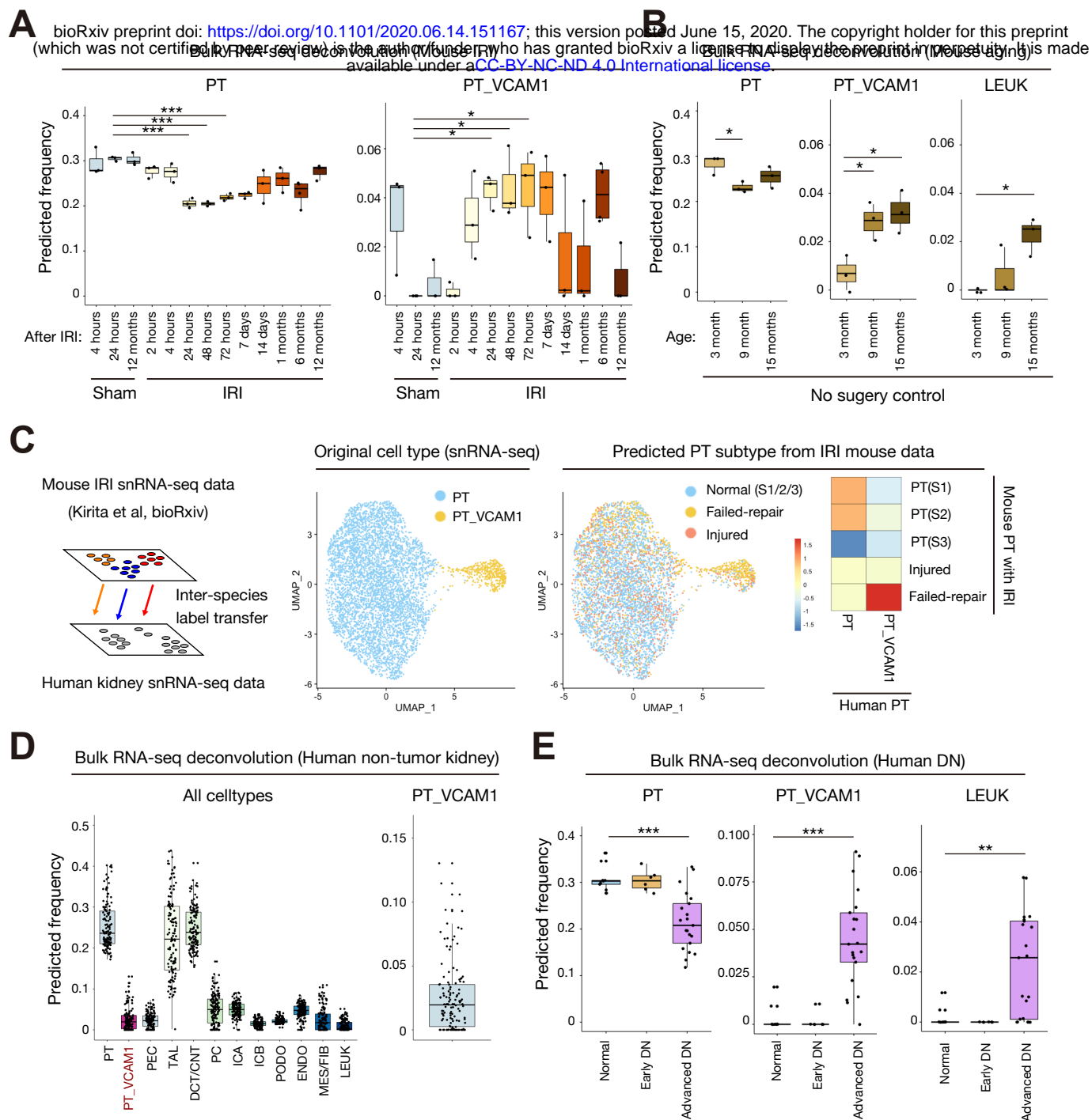
**Figure 2**



### Figure 3



**Figure 4**



**Figure 5**