2020

Multidimensional Zonation in the Renal Vasculature

Hannah Black
Pomona College

Multidimensional Zonation in the Renal Vasculature

Hannah Black

Katy Muzikar, Daniel J. O'Leary

Department of Chemistry
Pomona College
Claremont, CA
April 2020

Thesis submitted in partial fulfillment of the requirements for the Pomona College Bachelor of Arts in Chemistry

Table of Contents

List of Abbreviations

scRNA-seq: Single Cell RNA Sequencing

nUMI: Number of unique molecular identifiers

nGene: Number of genes

Percent.mito: Percent mitochondrial content

DEG: Differentially expressed gene

TSNE: T-stochastic neighbor embedding

SWNE: Similarity weighted non-negative embedding

DVR: Descending vasa recta

AVR: Ascending vasa recta

List of Tables and Figures

Abstract

The kidneys are two bean-shaped organs located in the posterior abdominal wall. They perform the vital task of filtering the blood and concentrating its waste products into urine. As a result, the kidneys play a role in regulating blood pressure, blood pH, and electrolyte balance.[4] Kidney disease is a broad heading for a variety of illnesses pertaining to the kidney, but the end result is that the kidneys have a greatly reduced ability to clean the blood.[11] The overall burden of kidney disease is massive because the treatments available are extremely expensive, in short supply, and the condition itself greatly increases the chances of developing other serious conditions such as cardiovascular disease.[11] The burden of kidney disease is so immense that it is recognized a major public health issue alongside conditions such as diabetes and obesity, and there is a tremendous need for treatment options that address kidney disease in its earliest stages to prevent serious adverse health consequences down the line.[11] Single cell RNA sequencing (scRNA-seq) is a new tool that enables researchers to studies tissues and organs at the resolution of individual cells.[12] The technology is unique in that it enables researchers to take extremely high resolution data for thousands of cells at once, data which is then analyzed using sophisticated bioinformatics techniques.[12] In the interest in adding to the body of knowledge about the kidney and vascular systems throughout the body, I matched scRNA-seq data to an anatomical map of the renal vasculature and identified transcriptional differences between the distinct endothelial compartments, as well as identified "zonation" in the transcriptional profiles of endothelial cells along the arterio-venous axis and the cortico-medullary axis. Lastly, I compared these findings to a dataset of P0 endothelial cells, which showed an arterio-venous distribution but appeared to reveal an underdeveloped cortico-medullary axis.

Introduction

Overview of the Kidney

The kidneys are two bean-shaped organs located in the posterior abdominal wall.[1] They are reddish brown in color and surrounded by an adipose (fat) capsule that serves to keep the kidneys in place and protect them from injury.[1] In adults, each kidney is roughly the size of a closed fist, and they are organized into three layers: the cortex, outer medulla, and inner medulla.[1] The functional unit of the kidney is the nephron, and human kidneys contain on average 1 million nephrons each, all of which are present at birth.[2,3]

The kidneys perform the vital task of filtering the blood and concentrating its waste products into urine.[4] As a result, the kidneys play a role in regulating blood pressure, blood pH, and electrolyte balance.[4] The kidneys also produce select hormones such as erythropoietin and degrade others such as parathyroid hormone, glucagon, and insulin.[4]

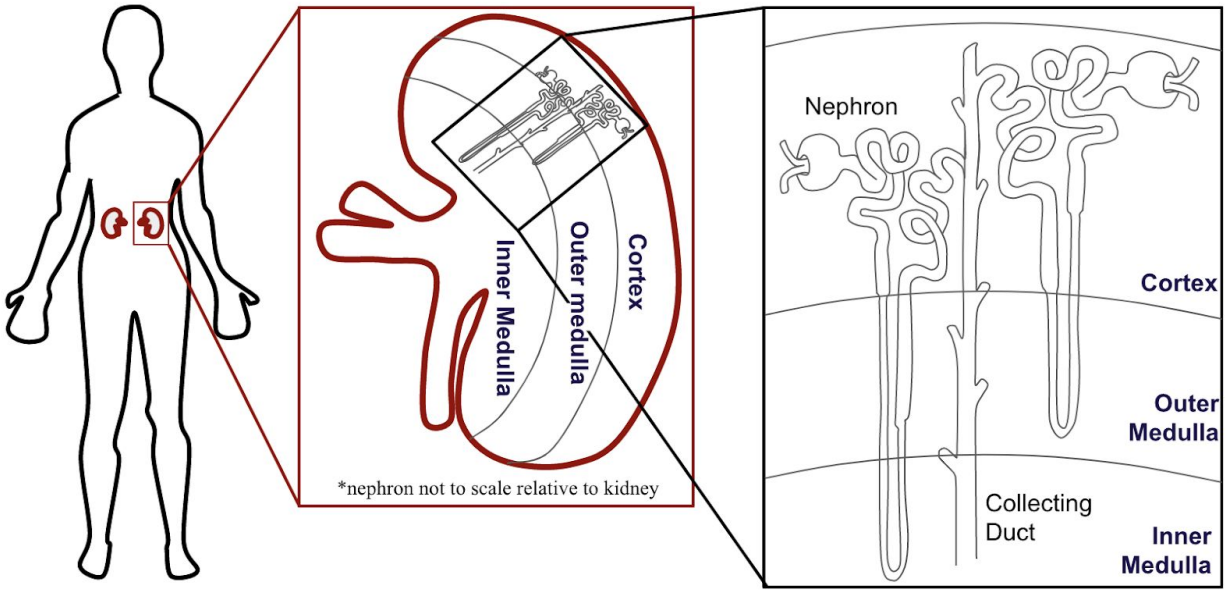

Figure 1. Schematic showing kidney location in the body and overview of nephron structure

Overview of Kidney Disease Causes and Treatment

Kidney disease is a broad heading for a variety of illnesses pertaining to the kidney, but regardless of cause, kidney disease is typically diagnosed by glomerular filtration rate.[5] Glomerular filtration rate is a number calculated for each individual based on the amounts of different filtration markers such as urine albumin and creatinine found in a urine sample.[5] In general, the lower the glomerular filtration rate, the poorer the kidneys are filtering blood.[5]

Major causes of kidney disease include glomerular diseases, diabetic nephropathy, hypertensive nephrosclerosis, and polycystic kidney disease, with glomerular disease being the most common cause.[6] However, regardless of the cause, the issue remains the same: the kidney has a reduced ability to clean the blood of toxins.[7] When kidney disease becomes severe and requires renal replacement therapy (RRT) such as dialysis or transplantation, the prognosis is comparable to a diagnosis of metastatic cancer.[7]

Kidney disease has been recognized as a public health priority since 1995 as a result of the prevalence of the disease and the need for immediate treatments.[8] According to the Global Burden of Disease study, in 1990 kidney disease was identified as the 27th leading cause of death globally. By 2010, kidney disease had climbed to the 18th.[5] It is also estimated that the true burden of kidney disease is even higher as a substantive portion of diabetic patients who pass away were additionally diagnosed with renal failure, but without it being recorded as the official cause of death.[5]

The prevalence of chronic kidney disease is extremely high, with 10-13% of the overall populations of the United States, China, Canada, and India showing signs of the disease.[9] In the 64 and above age cohort, this prevalence jumps to 35.8%.[9] These figures are also increasing over

time.[9] In the United States, 10% of the population had kidney disease between 1988-1994. This figure jumped to 13.1% from 1999-2004. In Taiwan, from 1996 to 2003 the prevalence jumped from 2% to 9.3%.[9] Overall, it is estimated that 5-10 million people die each year due to kidney disease or kidney injury.[10]

With such massive global presence, kidney disease sits alongside diabetes, hypertension, and obesity a part of the major noncommunicable disease burden and can have confounding effects with these diseases.[5] Mortality from cardiovascular disease is 30 times higher in patients with end stage renal failure.[5]

As discussed above, the two most commonly used treatment options for end stage renal failure are dialysis and renal replacement therapies. With the vast scope of kidney disease, the number of people in need of a transplant vastly exceeds the number of kidneys available.[10] And while renal replacement therapies are more widely available, they are extremely expensive.[10] Furthermore, while those in lower socioeconomic status have higher rates of kidney disease, the cost of treatment is so expensive that it poses a high financial burden on families.[5] In addition to cost to families, it is estimated that 2-3% of all health care dollars in developed countries go towards the treatment of patients with end stage renal failure.[5]

Due to the epic proportions of the disease, high cost of treatment, and increased risk of other serious conditions such as cardiovascular disease, CKD is considered a major public health issue and there is a need for treatment options that address kidney disease as early as possible to help with long-term treatment of the disease and prevent more costly treatments down the line.[11]

Single Cell RNA Sequencing

Single cell RNA sequencing (scRNA-seq) is a new tool that enables researchers to studies tissues and organs at the individual cell resolution.[12] Like other RNA analysis techniques, scRNA-seq aims to study cells based on the type and abundance of RNAs present in the cells, in other words the cell's RNA profile.[12] However, scRNA-seq differs from other techniques such as the northern blot or RNA microarrays in two important ways.[12] First, scRNA-seq collects RNAv data at the individual cell level. In other words, the output of an scRNA-seq analysis is a dataset where the complete RNA profile of each individual cell is tagged to that specific cell.[12] This is important in studying highly heterogeneous populations or populations with rare cell types because unique or rare signatures are not "drowned out" by other cell RNA profiles in the same sample.[12] Second, scRNA-seq has a high throughput, where researchers can take the profiles of thousands of cells in a sample simultaneously.[12] In other words, scRNA-seq enables researchers to create high resolution data for thousands of cells at once.[12]

While there are multiple techniques and platforms that enable scRNA-seq, the two most popular are the Smart-seq2 and 10X Chromium 3' platforms. The Smart-seq2 approach uses a SMART (Switching Mechanism at 5' End of RNA Template) workflow and the 10X Chromium 3' uses a microfluidics, droplet based approach.[13] Both techniques have pros and cons, but because microfluidics techniques enable the scRNA-seq of much larger samples and was the technique[13] used to collect data for this thesis, I will be specifically referring to microfluidics scRNA-seq from this point on. More specifically, the details included about procedure will be specifically referring to the 10X Chromium 3' platform.[13]

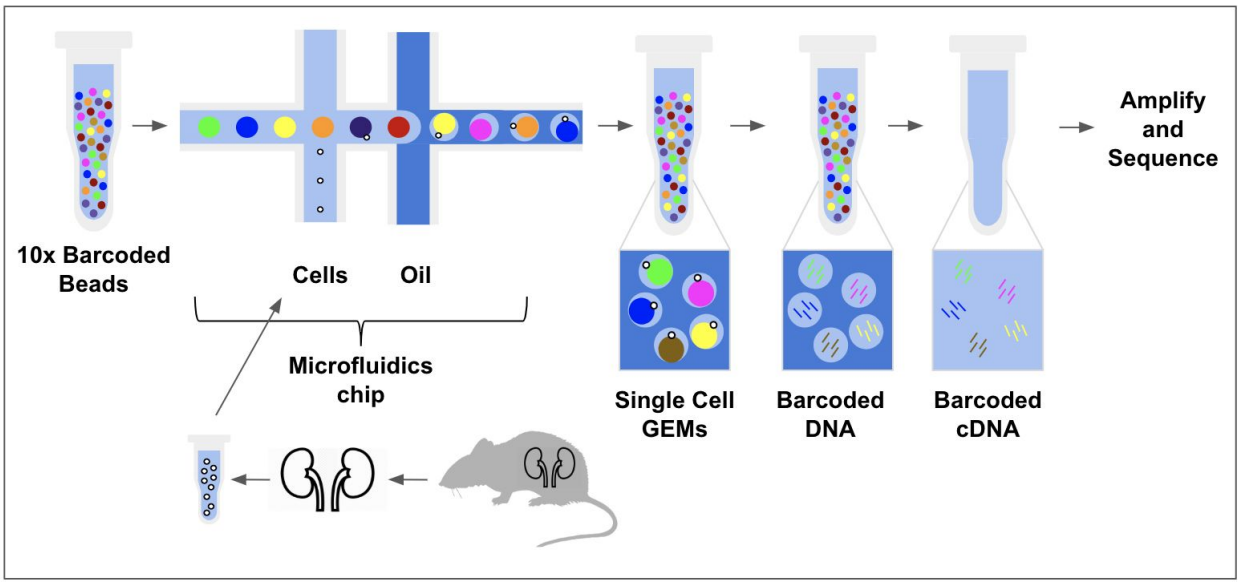

Figure 2. 10X Genomics Chromium 3' scRNAseq workflow

The 10x Chromium 3' workflow, at a high overview, requires the following steps. First, the researcher must isolate the tissue of interest, then dissociate it into single cells.[14] Then the cell suspension is added to a microfluidics chip, which the cells pass through one by one.[14] At a certain junction on the microfluidics chip, the cells are combined with the 10X functionalized gel beads (which contain the barcode sequences for tagging the RNA in the cell) and each gel/bead combination is separated into its own reagent droplet in an oil solution.[14] The gel and bead in emulsion (GEM) allows each individual cell to be lysed and tagged in its own micro-reaction bubble.[14] After the cell is lysed in the droplet, the RNAs for each cell are tagged with the specific barcode's from their bead, the RNAs are separated from the rest of the mixture, undergo reverse transcription to generate tagged cDNA, then the cDNA is amplified and sequenced.[14] It is possible to amplify and sequence the cDNA together because each cDNA is tagged via barcode to which bead was in the micro-reaction droplet with its cell.[14] Overall, this process enables a

one-day single cell sequencing of tens of thousands of cells at once to produce the RNA profile of each individual cell in the sample.[14]

In order to generate high quality data, is is important to ensure that the cells in suspension are minimally disrupted from their natural state.[12,15] The RNA profile of a cell is in constant flux, and the dissolution techniques used to create a cell suspension for a scRNA-seq study can impact the transcriptional profile of the cell.[15] This is because breaking down an organ or tissue into a suspension of single cells typically requires a complex protocol and the use of one or more enzymes, such as collagenase, trypsin, and pronase, all of which are active at 37°C.[15] However, cells are still metabolically active at 37°C, and because they have been displaced from their organism and are in a foreign environment their RNA profile can rapidly shift.[15] This can create a large stress signature in the RNA profile that overwhelms the cell's natural signature.[15] However, if the tissue could be kept at low temperatures during dissolution, enabling the process to take place 'on ice', it is possible to effectively freeze the RNA profile of the cells in their native tissue form during dissolution.[15]

Cold active proteases make it possible to perform the single cell suspension at cold temperatures.[16] In many different niches around the world, there are organisms called psychrophiles that live in extremely cold conditions. In order to survive in these environments, their cellular machinery has evolutionarily adapted to function in the extreme cold.[16] *Bacillus licheniformis* is one such microorganism[16] *Bacillus licheniformis* was originally isolated from fermented Thai fish sauce and is a proteolytic enzyme with a molecular mass of 31 kDa, an isoelectric point >9.3, and an optimal pH of 10.[17] A protocol was developed wherein the *Bacillus licheniformis* cold active protease could be used to create a dissociation protocol that can occur

as “on ice” temperatures of 6°C.[15] The procedure was originally developed for use in newborn kidneys, but the McMahon Lab adapted this protocol to dissociate complete adult mouse kidneys with greatly reduced stress signature.[15]

Single Cell RNA Sequencing and Kidney Research

In the broader kidney research field, single cell RNA sequencing has been used to study several facets of the kidney. The Suszták lab at Perelman Genomics and Computational Biology compiled a healthy adult mouse single cell kidney dataset made up of 59,979 cells which were used to identify a potential transitional cell type between collecting duct cell types.[18] They also used transitional fate mapping to find that intercalated and principal cells in the collecting duct undergo transitions mediated by the Notch signalling pathway.[18] The Epithelial Systems biology laboratory, led by Mark Knepper, used scRNA-seq data to specifically examine the differences in three cell types that makeup the collecting duct.[19] The Tabula Muris Consortium undertook the massive task of single cell sequencing cells from 20 different organs in *Mus musculus* and creating an online repository of genes identifying major organ systems and their high level organization.[20]

The McMahon lab undertook a highly detailed, anatomical based, approach to studying the cell types along each segment of the nephron and collecting duct in the kidney, elucidating the differences between male and female kidneys as well as the different transcriptional signatures of the two types of nephrons in the kidney.[10] This highly detailed analysis was also made into an online searchable tool with the information from the study, eliminating the barrier of coding needed to study the raw single cell data.[10] This single-cell atlas enables clinicians and

researchers to easily access high-resolution RNA-seq data for each segment of the highly complex nephron and aid in the effort to develop new treatment options for CKD. [10]

The Renal Vasculature and Motivations for this Thesis

While much research has been done focusing on the epithelium of the kidney, the nephron and collecting duct, interfacing with the nephron along its entire length is the less-studied vasculature, which actually carries blood through the kidneys.[21] The vasculature has its own highly unique organization, with both a vertical and lateral heterogeneity and distinct capillary beds at different depths in the kidney.[21]

The blood enters the kidney via the renal artery from the inner medulla and then branches into interlobar arteries which travel towards the corticomedullary boundary where they travel along as arcuate arteries, then branch and travel upwards towards the cortex as cortical radiate arteries. From these cortical radiate arteries, afferent arterioles branch off and enter the glomerular capsules to form the tightly interwoven glomerular capillaries.[21] The glomerular capillaries exit the glomerulus as efferent arterioles which take one of two tracks. The efferent arterioles from cortical nephrons form the cortical capillary beds, which flow into veins which exit the kidney.[21] The efferent arterioles from juxtamedullary nephrons flow towards the medulla, turning into the descending vasa recta which makes a loop in the inner medulla to become the ascending vasa recta, which then feeds into a large vein which carries the blood out of the kidney ultimately via the renal vein. Some of the descending vasa recta also branch off in the medulla to form the medullary capillary bed, which also drains into the venous system.[21]

This structure with three paths of blood flow, through the cortical capillary bed, medullary capillary bed, or vasa recta, is repeated throughout the kidney with each nephron.[21]

While this repeated vasculature structure is a vital part of the kidney's filtration mechanism, there is comparatively minimal research into the vasculature's role in the filtration process and the cell types in the vasculature as compared to the wealth of research on the nephron.[21]

Several recent papers used single cell data to describe the vasculature in other organs, such as the brain and heart.[22,23] However, these studies focus on the broader vasculature headings of "arterial" and "venous", and leave questions about how these broad categories of cell types vary in an organ or tissue.

In the interest in learning more about the heterogeneity of vasculature throughout the body, the kidney provides an ideal organ system to do that because of the highly repetitive and organized vasculature. Therefore, in this study, I have matched scRNA-seq data to an anatomical map of the renal vasculature with the goal of furthering our understanding of a highly complex organ with high clinical significance.

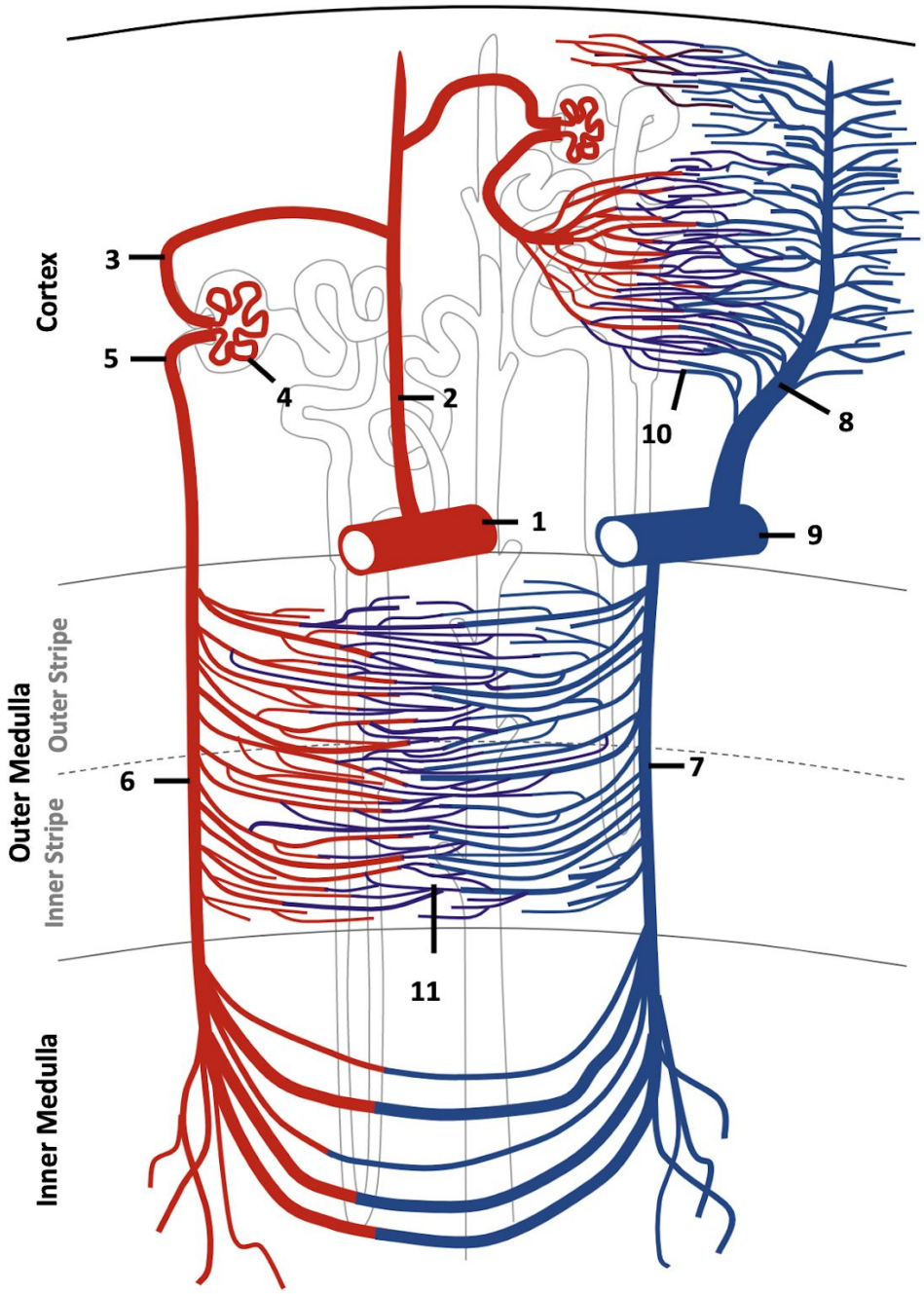

Figure 3. Schematic of the vasculature organization, (1) Arcuate artery (2) Cortical radiate artery (3) Afferent arteriole (4) Glomerular capillaries (5) Efferent arteriole (6) Descending Vasa Recta (7) Ascending Vasa recta (8) Cortical radiate vein (9) Arcuate vein (10) Cortical Capillary Bed (11) Medullary Capillary Bed

Brief Overview of Methodology

To study the renal vasculature using scRNA-seq data, I used three bioinformatic packages in the "R" integrated development environment. To perform the initial quality control

filtering steps and initial clustering, I employed the widely used "Seurat" package.[24] This included performing principal component analysis (PCA) and t-stochastic neighbor embedding (TSNE) to organize the dataset according to similarities between cells. The resulting TSNE plot enabled me to identify what populations in the single cell data represent which portions of the renal vasculature. [24] In order to study anatomical relationships between different cell types in the renal vasculature, a package called SWNE (similarity weighted non-negative embedding) was employed.[25] This software uses embedded factors, in this case expression of certain genes, to add a biological factor to the structure of the single cell embedding.[25] Lastly, in order to study in greater detail the idea of "zonation" in the renal vasculature, a package called Cellrouter was used to study the transition from artery to vein in the cortical capillary bed and the vasa recta loop.[26] Cellrouter does this by computationally recreating these vascular trajectories in order to study how gene expression varies along the trajectory.[26]

Materials and Methods

The primary endothelial cells studied in this paper are from the same dataset discussed in Ransick et. al.[10] The full methodology for the single cell dataset collection can be found there but is briefly summarized here. Additionally, the imaging performed for this project was done by a research technician in the lab. For clarity, the researcher responsible for each section of the methodology is listed beside the section heading.

Mouse Adult Kidney Single Cell Data Collection[10] ***(Performed by Andrew Ransick, Senior Research Associate, McMahon Lab)***

Adult mice were euthanized by CO_2 inhalation and then transfused with saline in order to remove blood. The kidneys were then harvested, the outer capsule was removed and ureter trimmed away. Each kidney was dissected into three zones, the cortex (zone 1), outer medulla (zone 2), and inner medulla (zone 3). In order to dissociate the adult kidney cells, a new 'cold dissociation' protocol was developed around Bacillus licheniformis cold active protease. According to a quantitative PCR assessment, this procedure improved cell dissociation times compared to ice-cold dissociation and did not induce a significant transcriptional response.

Samples for single cell RNA sequencing were prepared such that suspensions contained roughly 7000-9000 cells and were then combined with 10X Chromium reagent mix. Samples were divided by replicate and by zone and loaded into the 10X microfluidics chip. The data was then collected by following the manufacturer's instructions for the Chromium v2 Single Cell 3' protocol. Briefly, the steps include cell capture, lysis, and mRNA reverse transcription in microdroplets within the fluidics chip. The resulting cDNA was collected, cleaned of waste products, amplified, then processed into sequencing libraries. Sequencing using the HiSeq 4000

platform was performed at The Translational Genomics Center at the Children's Hospital Los Angeles Center for Personalized Medicine. The resulting sequencing reads were matched to a mouse reference genome and compiled using the Cell Ranger Single-Cell Software Suite 2.0 from 10X Genomics.

The resulting matrices representing each of the 12 samples (four mice with three kidney zones each) were loaded into R using the Read10X() function. These 12 replicates were then converted into Seurat-format objects, labeled by sample, and combined into a single object using CreateSeuratObject(), AddMetaData(), and MergeSeurat(), respectively. Stringent filtration metrics were then applied to the entire dataset such that only cells with 1,000-4,000 genes/cell, 1,000-16,000 mRNA transcripts/cells (UMI's), and a mitochondrial percentage of genes/cell of 35% or fewer were kept in the dataset. All cells outside of these parameters were filtered out using FilterCells(). This analysis created a dataset with 19,125 distinct genes expressed over a sample of 31,265 cells which was then clustered using the Seurat package.[24] Clusters were subsequently assigned to one of five categories: Interstitial, Immune, Vascular, Ureteric Epithelium, and Nephron. Ransick et al. proceeds to analyze the ureteric epithelium and nephron cells.[10]

Identification and Separation of the Endothelial Cells ***(Performed by Hannah Black, Thesis Author)***

The work discussed in this paper begins with the identification and separation of the endothelial cells from the rest of the adult dataset using packages hosted in R. All bioinformatics were performed in the RStudio integrated development environment using R version 3.5.1. First, the filtered adult dataset from Ransick et al.[10] was loaded. Because a new version of Seurat had

been released after the publication of that paper, the UpdateSeuratObject() command was used to update the object's formatting such that the vascular analysis could be performed in the newest version of Seurat. The full adult dataset was then visualized using Dimplot(), and, as identified by Ransick et al.,[10] the six clusters representing the endothelial cell populations were visualized by confirmation of expression of the canonical endothelial genes Kdr and Cdh5. Upon confirming that the expected endothelial populations showed these endothelial markers as well as others including Pecam1 and Plvap, a standard workflow was followed to analyze the endothelial cells.

Seurat Workflow ***(Performed by Hannah Black, Thesis Author)***

The Seurat analysis was performed using the Seurat version 3.0 package developed by the Satija Lab.[24] The endothelial cells were first subsetted from the full adult kidney dataset using the SubsetData() command. The data was then normalized using NormalizeData() and variable features in the dataset were identified using FindVariableFeatures(). The genes identified as variable features were then scaled using ScaleData(), and the first 50 principal components were calculated based on the variable features using RunPCA(). The principal components were then visualized using ElbowPlot(), DimHeatmap(), and JackStrawPlot() in order to identify which principal components represent true variation and dimensionality in the dataset. Upon visual inspection, the first 25 principal components were selected and t-stochastic neighbor embedding was performed using RunTSNE(). The results of the embedding were visualized using TSNEPlot() and the R object containing the now-processed endothelial data was saved.

Gene markers for each of the clusters in the preliminary object were identified using FindAllMarkers() and exported into a CSV that could be visualized in Microsoft Office Excel. ViolinPlots for the quality control metrics nGene, nUMI, and percent.mito were generated for all clusters using VlnPlot(). Percent.mito refers to percent mitochondrial content and is a marker of cell distress as highly distressed cells tend to have a disproportionately high percentage of mitochondrial genes. The number of genes (nGene) detected in a cell is used as a quality control metric because an extremely low number might indicate a low-quality cell or empty droplet. Similarly an unusually high number of genes could indicate a cell doublet or multiplet, where two or more cells are captured in a single reaction droplet rather than one. UMI's are unique molecular identifiers (nUMI) and are used as a quality control metric similar to number of genes. Using these three quality control metrics enabled the identification of any clusters of disproportionately poor quality relative to the other clusters in the dataset.

From the preliminary clustering, one cluster showed a higher-than-average mitochondrial percentage, which was reflected in the high number of mitochondrial genes in its top DEG list. As per standard lab protocol, this cluster was removed from a subset from the data following the same standard Seurat workflow. The data was then re-analyzed from the NormalizeData() step in order to ensure that the remaining cells in the dataset can be clustered without the influence of the cells that were removed.

Upon performing the second clustering, the top differentially expressed genes (DEGs) enabled the identification of a small cluster with a distinct red blood cell profile and a small cluster with a distinct nephrogenic profile. These clusters were subsetted from the data and a fresh analysis was again performed using the Seurat workflow detailed above.

Upon performing the third clustering, the quality control metrics represented by nUMI, nGene, and percent.mito appeared consistent across the clusters, and the DEG lists showed no clusters with distinct non-endothelial signatures and this pruned dataset was carried forward for further analysis.

The first step in analyzing this object was to adjust the resolution of the TSNE such that the number of clusters reasonably matched our understanding of the vasculature in the kidney. There is no one "correct" way to choose the resolution for a single cell TSNE object,[27] but it is valuable to visualize the data with the approximate amount of detail you seek to identify in the biological understanding of the system you are studying. Changing the resolution does not impact the placement of cells in the embedding, but changes the number of clusters these cells are grouped into (visualized in this paper by different colors) where fewer clusters split the data according to the most prominent differences in the dataset and greater numbers of clusters splits the data according to finer differences. Thus, objects with resolutions ranging from 0.6-1.1 were generated at 0.1 intervals, with a resolution of 1 being the default set by the Satija Lab.[24] After analyzing the objects at each resolution, as well as their respective DEG lists for each clustering, the object generated with a resolution of 0.7 showed slightly fewer distinct populations than we would expect given our understanding of the renal vasculature, and 0.8 showed slightly more. Thus, a final object was generated at a resolution of 0.75 and this object was selected to continue the analysis.

To identify how the anatomy of the renal vasculature aligned with the single cell data, genes represented in different populations of the renal vasculature were identified from the scientific literature, and these genes were used in conjunction with the cortical depth data

inherent to the dataset (zones 1, 2 and 3) to identify which clusters represented which portion of the endothelium.

SWNE Analysis ***(Performed by Hannah Black, Thesis Author)***

Similarity Weighted Nonnegative Embedding (SWNE) analysis was performed using the SWNE package developed by Yan Wu in the Kun Zhang Lab.[25] In order to perform this analysis, SWNE embedding was performed using the RunSWNE() command and visualized using PlotSWNE(). SWNE embedding was performed for a range of k values, from 11-19, where k represents the number of nearest neighbors. A k value of 14 was ultimately selected as it appeared to minimize the distortion of the 2D representation of the data and the finalized object was saved.

Cellrouter Analysis ***(Performed by Hannah Black, Thesis Author)***

CellRouter analysis was performed using the Cellrouter package developed by Edroaldo Lummertz da Rocha of the George Daley Lab.[26] In order to perform cellrouter analysis, the clusters for analysis (those representing either the medullary capillary bed or the vasa recta loop) were subsetted from the main seurat object using the subset() function. The raw gene reads for these clusters were pulled out as a list using GetAssayData() and converted into a matrix using as.data.frame(as.matrix()). A cellrouter object was created from these raw counts using CellRouter() and the cluster identities, sex, zone, and replicate information for these cells were added back to this object with addInfo(). The cell router object was then processed using Normalize(), scaleData(), and computePCA(). The principal components were then visualized using plot(). A TSNE was then computed using computeTSNE(), added to the object using customSpace(), and plotted using plotReducedDimension(). A nearest neighbors network was

generated using the jaccard simulation type and the buildKNN() function. This network was then visualized using plotKNN(). The source and target clusters for the algorithm to walk between were selected and put into a list, and paths between the source and target were identified using findPaths(). The identified trajectories were processed using processTrajectories(), unique paths were pulled out and then trajectories were formed with correlationPseudotime(), and topGenes() were pulled out for the genes identified in these trajectories. The dynamics of the genes changing along the identified trajectories were processed using smoothDynamics() and clusterGenesPseudotime(). The genes moving along these trajectories were then subsequently visualized using plottrajectories() and plotPathHeatmap().

RNAScope and Antibody Imaging ***(Performed by Kari Koppitch, Research Lab Technician, McMahon Lab)***

Kidneys were harvested from adult C57BL/6 and transgenic p0 Hoxb7-Venus Six2-TGC-tdT mice. Adult mice were perfused with HPBS followed by 4% PFA in PBS. All kidneys were fixed for 1 hour in 4% PFA-PBS at 4°C, washed and equilibrated overnight in 30% sucrose-PBS at 4°C. Kidneys were embedded in OCT and cryosectioned at 10-12 micron thickness and stored at -80°C. Before RNAscope was performed according to manufacturer instructions, slides were post-fixed in 4%PFA-PBS for 1hour at 4°C. RNascope followed the RNAScope® Multiplex Fluorescent V2 Assay (323100-USM) manufacturer's protocol for Fixed Frozen tissue starting from slide dehydration. Antibody stains followed manufacturer's protocol (323100-TN) with the following exceptions. Sections were first blocked in 1.5% Sea-block -TBS-0.1%BSA for 1 hour at room temperature followed by primary antibody staining overnight at 4°C and secondary development with previously tested secondary AlexaFluor dyes for 1hour

at room temperature instead of consecutive TSA developments. Images were taken as 40x tilescans on a Leica Sp8 confocal microscope and further processed with Leica LAS X software.

ACD RNascope probes used were Gata5 (549061-C1), Pi16 (451311-C2), Ifit1 (500071-C2, Apela (416811-C3), AplnR (436171-C1), Apln (415371-C2) and Akr1b3 (custom, C3 probe). Antibodies used in immunostaining were Plvap (rat, BioRad- MCA2539GA, 1:100), CD-31 (rat, BG Pharmingen-550274, 1:400), Aqp1 (rabbit, abcam-ab168387, 1:750), Aqp2 (rabbit, Kerafast-ENH129, 1:500), Tamm-Horsfall (rabbit, Alfa Aesar-J645429, 1:250), Igf1 (mouse IGF1, abcam-ab176523, 1:500), Lef1 (rabbit, Cell Signaling-2230, 1:100), Jag1 (goat, R&D-AF599, 1:250), Fabp5 (goat, R&D-AF1476, 1:300), and Gap43 (rabbit, Sigma-AB5220, 1:1000) .

P0 Kidney Cell Collection ***(Performed by Andrew Ransick, Senior Research Associate, McMahon Lab)***

P0 kidneys (those from the first day of postnatal life) were dissected, the outer capsule removed, ureters trimmed away, and the tissue digested as described in the adult kidney dissociation. Digestion times and enzyme concentrations were altered to adjust for the decreased kidney size and unique composition of interstitial cell types at this earlier developmental stage. Data were collected on the same 10X platform using the same chemistry as employed for adult data sets.

P0 Single Cell Bioinformatics ***(Preformed by Hannah Black, thesis author)***

The P0 single cell data was analyzed according to the same procedures described above for the adult. In brief, this entailed identifying the P0 endothelial cells using canonical endothelial cell markers, performing a preliminary clustering of the endothelial subset in Seurat,

then finally embedding the data using a SWNE analysis. All bioinformatics were performed using the same sequence steps as described for the adult.

Results and Discussion

The goal of this project is to use single cell RNA sequencing data to describe, in detail, the vasculature in the kidney. This project is motivated by the need for a rigorous understanding of the kidney, an organ with high clinical significance, and aims to contribute to that understanding by analyzing the specialized endothelial cells that exist within it.

Analysis of Endothelial Cells in Seurat

The first step in the analysis was to select out the endothelial cells from the broader dataset, which included all cell types in the adult mouse kidney. The genes used to identify the endothelial cells from the larger dataset were canonical endothelial markers (Figure 4). Kdr (aliases: Flk1, Vegfr) is one of two endothelial cell specific receptors for vascular endothelial growth factor (Vegf).[28,29] Pecam1 (alias: Cd31) is a cell adhesion molecule expressed at endothelial cell junctions.[30] Cd34 is a membrane-spanning protein expressed widely in vascular endothelium as well as most hematopoietic progenitor cell types.[31] Plvap is a vesicle associated protein critical to the formation of fenestration in the vascular endothelium.[32] Emcn is a widely expressed regulator of angiogenesis[33] and Tek (alias: Tie2) is a receptor tyrosine kinase specific to endothelial cells.[34]

Based on the specificity of these endothelial cell markers within the full kidney dataset, all clusters predominantly expressing these endothelial clusters were chosen for further analysis (Figure 4). It is worth noting at this stage that while we occasionally refer to this dataset as the "vascular" dataset, we are looking solely at the endothelial cells of the blood vessel, which are the cells that form the main vessel wall and are directly in contact with the blood that flows through the vessels. There are other cell types that contribute to the vascular structure, such as

smooth muscle and pericyte cells.[35] These mural cells can be identified in the full kidney dataset based on the canonical marker Cnn1[10,36] but will not be studied as part of the endothelial dataset here.

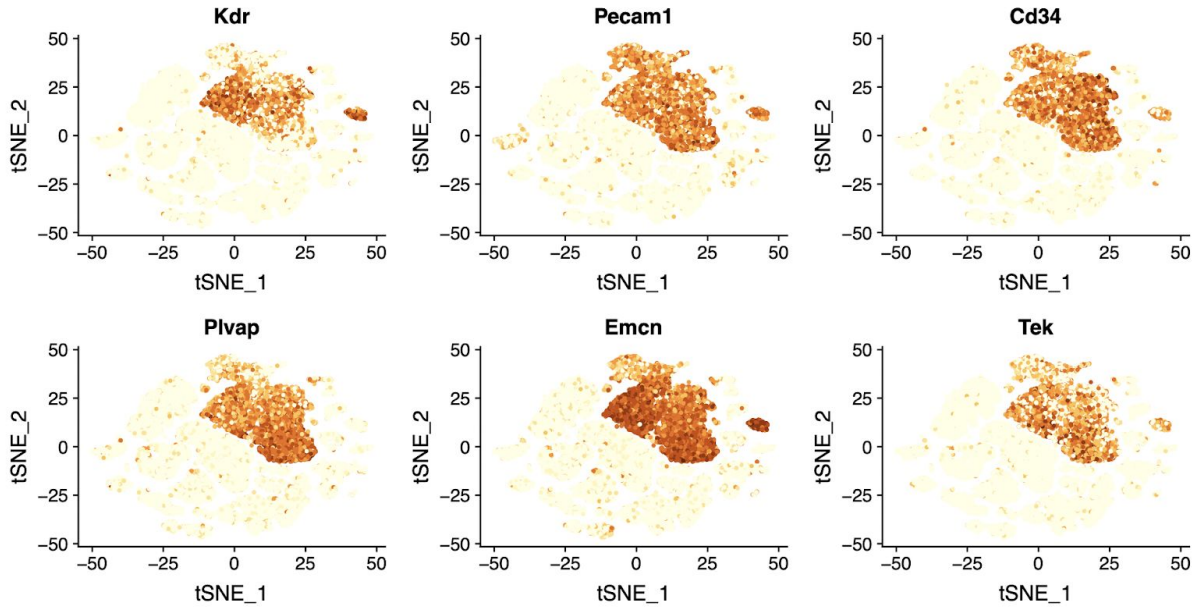

Figure 4. FeaturePlots showing the expression of canonical endothelial-cell markers (Kdr, Pecam1, Cd34, Plvap, Emcn, and Tek) in the full adult kidney dataset. Expression of each of these markers is primarily isolated to a specific population, identifying those cells as the endothelial cells, which were subsequently subsetted for the endothelial cell analysis (this full adult kidney dataset is also described in greater detail in Ransick et al.[10]).

Upon subsetting the endothelial cells from the full kidney dataset, the endothelial cells were clustered using Seurat[24] and three small trace populations of contaminating non-endothelial cell types were identified and removed. Upon generating a finalized dataset, the zonal, replicate, and sex distribution of the cells were visualized and a level of resolution was selected that approximately mirrored the number of distinct regions in the vasculature identified in the literature (Figure 5). Visuals representing quality control metrics nUMI, nGene, and percent.mito

were also generated (Figure 6) alongside a table summarizing the number of cells in each cluster according to Replicate, Sex, and Zone (Table 1).

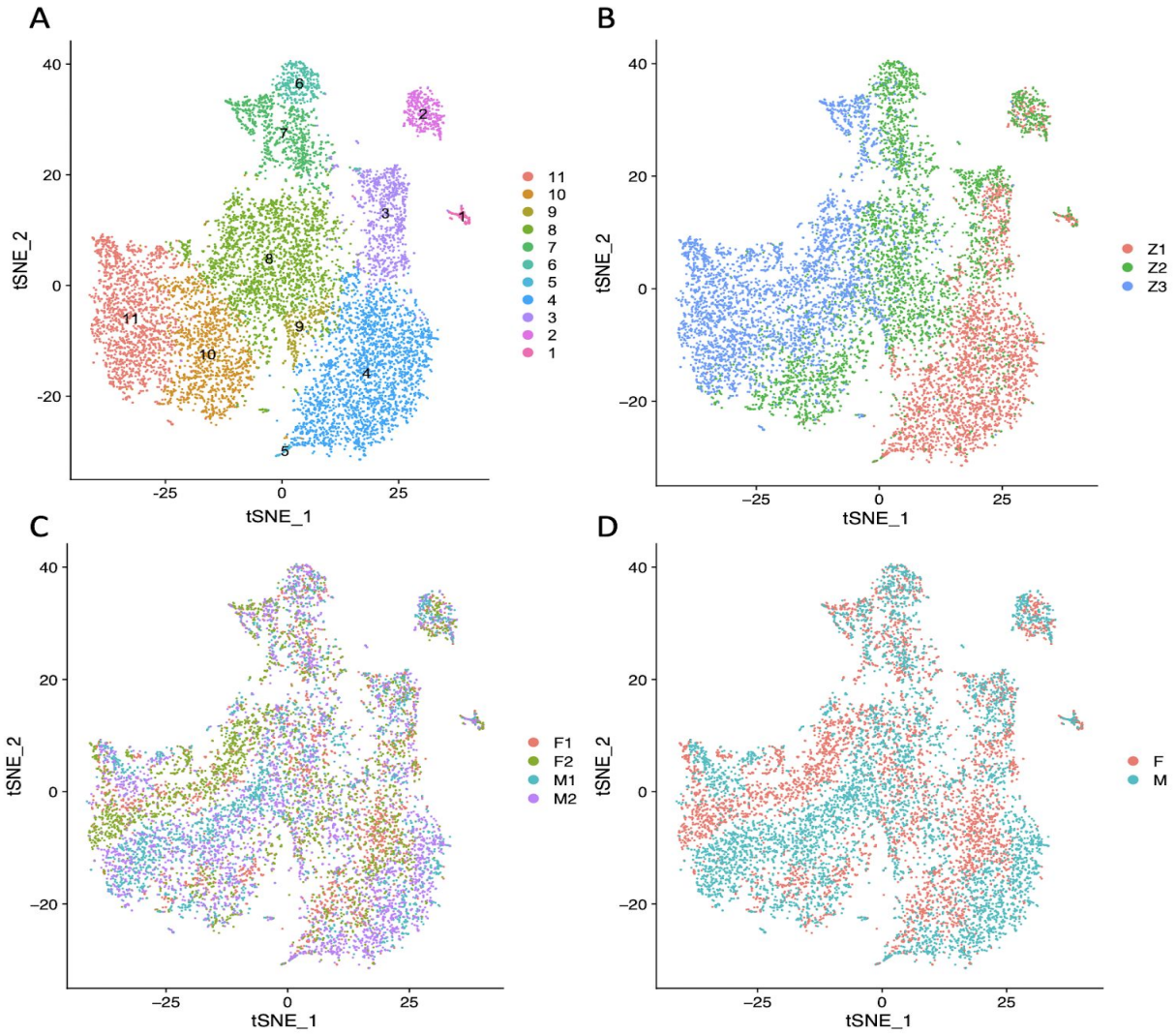

Figure 5. Output from the TSNE analysis performed in Seurat, where each dot on the plot represents a single cell in the endothelial dataset and is spatially located near those cells most similar to it according to the TSNE embedding. A) TSNE plot colored by cluster B) TSNE plot colored by zone C) TSNE plot colored by replicate D) TSNE plot colored by sex

Table 1. Number of cells in each cluster belonging to each replicate, zone, and sex.

		Cluster Number										
		11	10	9	8	7	6	5	4	3	2	1
Replicate	F1	219	221	39	370	158	69	5	566	163	52	24
	F2	451	240	97	700	276	50	7	516	224	98	23
	M1	500	351	28	561	140	84	6	426	181	108	18
	M2	479	377	82	553	208	111	24	924	232	98	23
Zone	Z1	2	11	46	10	0	1	30	2144	360	150	67
	Z2	98	644	176	1139	400	295	12	279	428	202	21
	Z3	1549	534	24	1035	382	18	0	9	12	4	0
Sex	F	670	461	136	1070	434	119	12	1082	387	150	47
	M	979	728	110	1114	348	195	30	1350	413	206	41

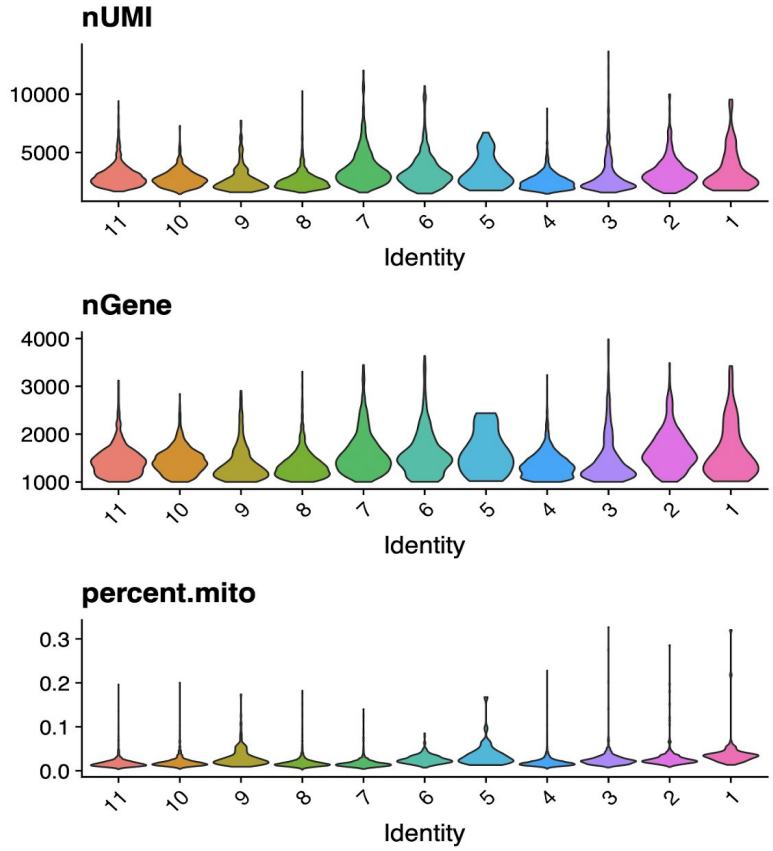

Figure 6. ViolinPlots showing nUMI, nGene, and percent.mito organized by cluster identity on the x-axis.

At first glance, the TSNE embedding appears to recreate the cortical-medullary organization that exists in the kidney. This can be seen by the organization of the zones in the TSNE, where zones 1, 2, and 3 represent cells from the cortex, outer medulla, and inner medulla, respectively, and are ordered as they are found in the kidney (Figure 5B). Additionally, our replicates appeared to mix evenly across the clusters, indicating that our samples are mixing well (Figure 5C; Table 1). While there is some visual striping of the sexes within the clusters, none of the clusters are dominated by cells from one sex, indicating that sex differences are not driving the clustering (Figure 5D; Table 1) Additionally, the quality control metrics nGene, nUMI, and percent.mito looked strong and consistent across the clusters. Dotplots were generated to show the differentially expressed genes that define each cluster (DEGs) and begin to draw correlations between different clusters and anatomical compartments (Figure 7).

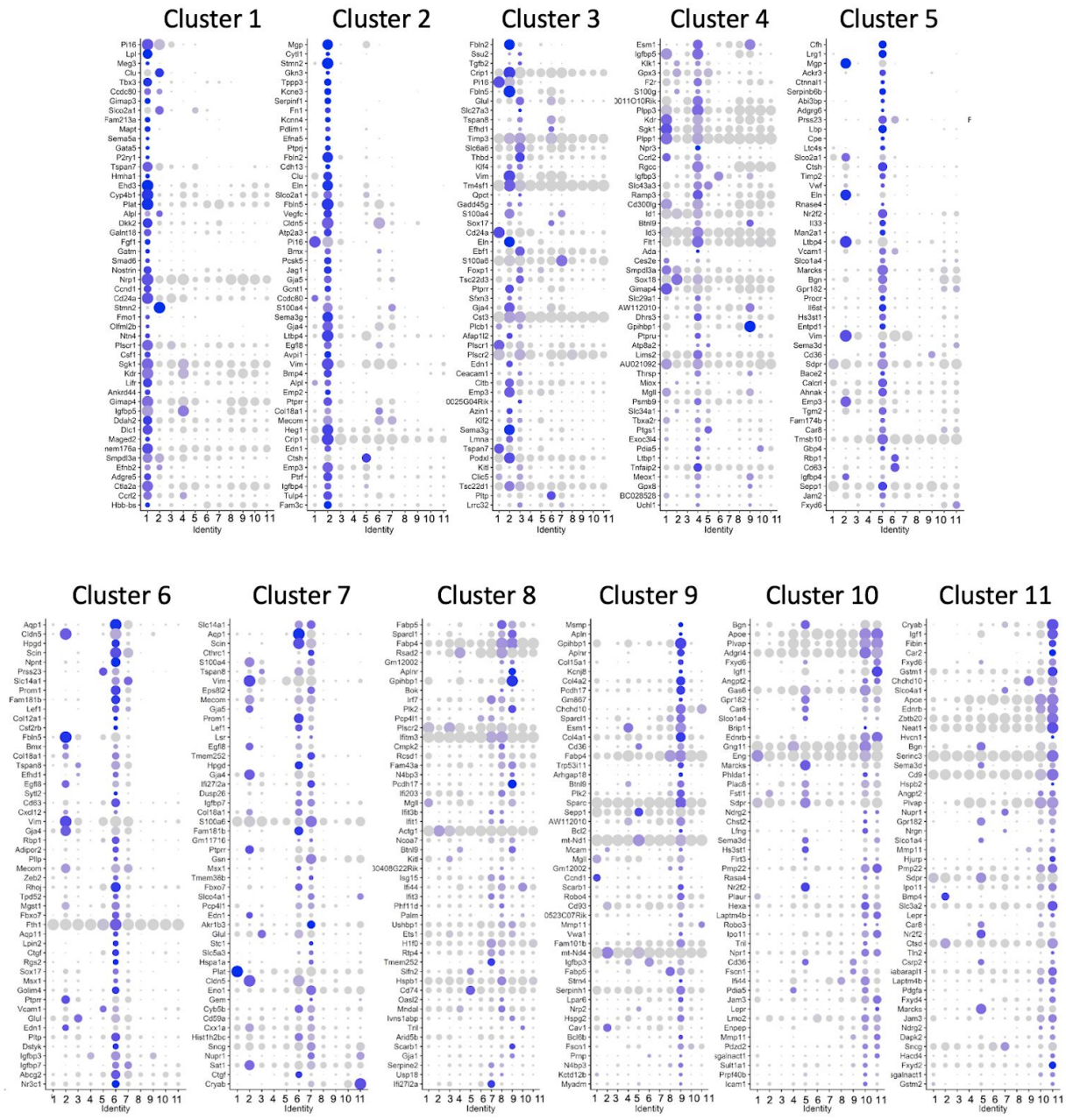

Figure 7. Dotplots showing the top 50 differentially expressed genes (DEGs) for each cluster. Cluster identity is named on the x axis and the top 50 DEG's are ordered from most differentially expressed to least from top to bottom. Together, the DEGs show distinct endothelial transcriptional profiles for each of the eleven clusters in the sample.

Connecting the single cell data to the renal biology

With confidence in the replicability of the data and the quality of cells amongst the clusters, the differentially expressed gene lists were used alongside canonical genes identified in the literature to assign each cluster an anatomical identity. For example, Ehd3 and Lpl are top DEGs for cluster 2 and are documented as highly enriched in the glomerular endothelium.[37, 38] It is also known from previous studies that while the glomerular endothelium is fenestrated (has gaps in the endothelial membrane), it does not express the diaphragmatic endothelial fenestration gene, Plvap.[39] Thus we are able to assign cluster 2 as the glomerular endothelium (Table 2).

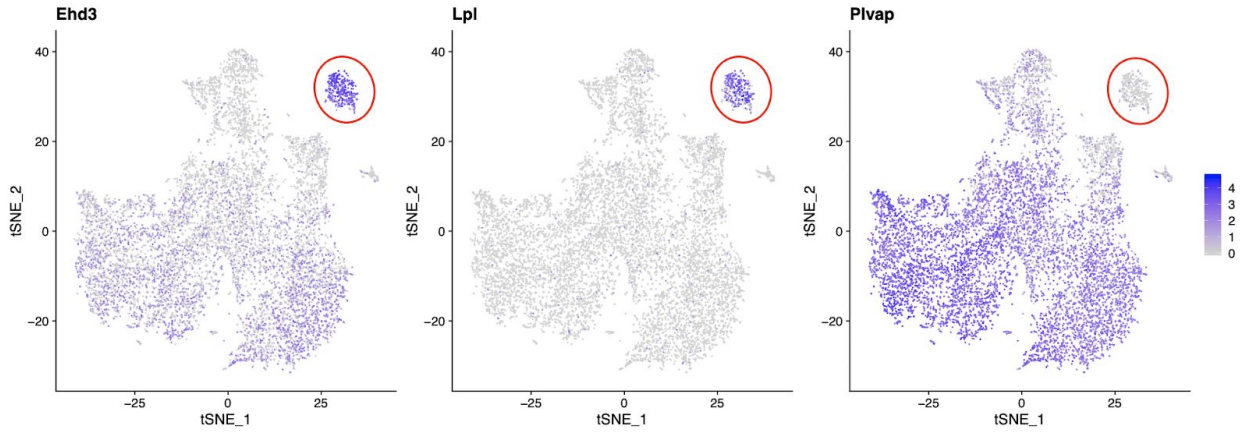

Figure 8. Feature plots showing expression of genes identifying the glomerular endothelium. Circled is cluster 2, which represents the glomerular endothelium.

While certain endothelial compartments such as the glomerular endothelium have distinct transcriptional profiles that are well characterized in the literature, this is not true for all compartments of the endothelium. For instance, there is no well-established marker distinguishing medullary versus cortical capillaries in the kidney, thus we must use other information such as the zonal identities of the cell (whether cells came from the cortex, outer

medulla, or inner medulla) to help us infer the identities of certain populations. We can also use canonical arterial/venous markers to help us generally identify whether a cluster is more arterial or venous in nature, giving another clue as to its anatomical correlation.

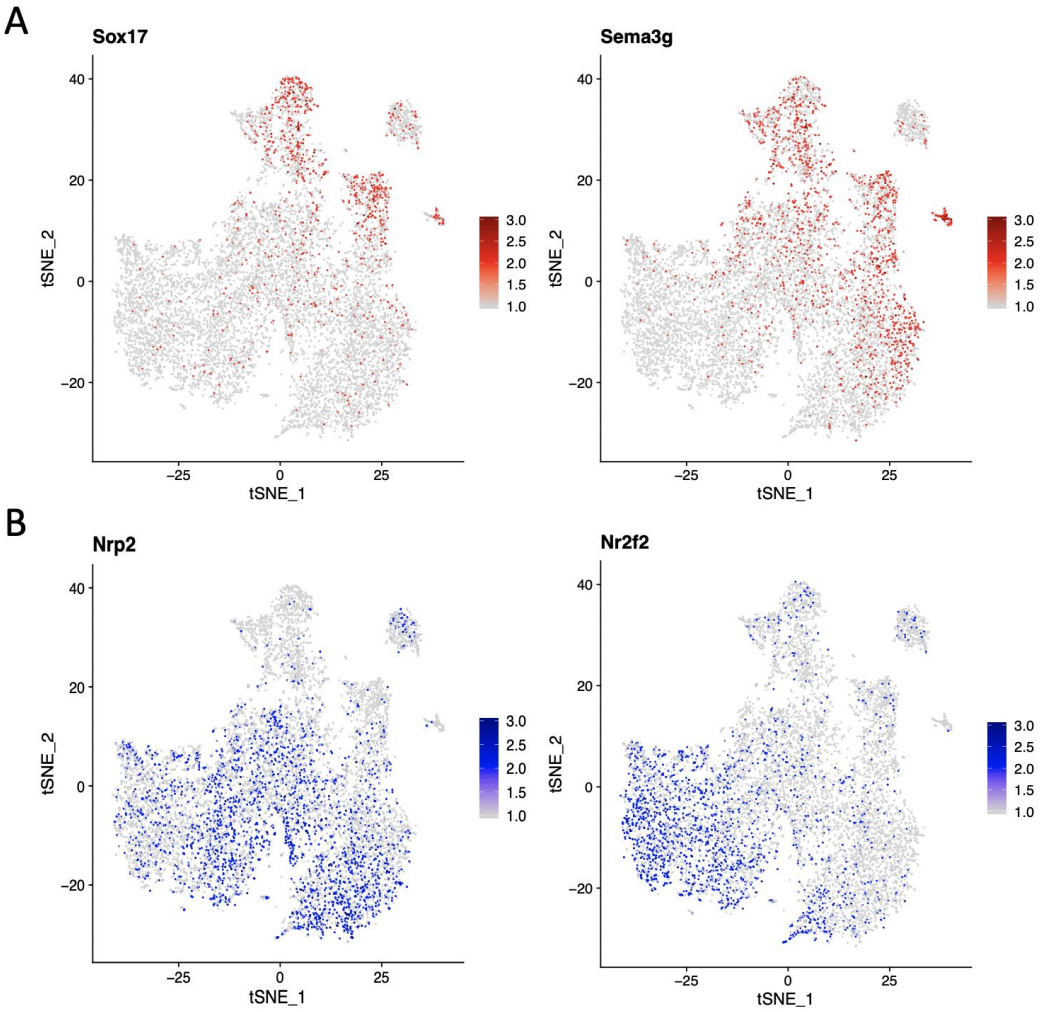

Figure 9. Feature Plots showing the expression of genes related to arterial and venous identity. A) Expression pattern of genes enriched in arterial endothelium. Red color chosen to symbolize more oxygenated blood. B) Expression pattern of genes enriched in venous endothelium. Blue color chosen to symbolize more deoxygenated blood. Together these show distinct arterial and venous regions within the TSNE embedding.

Between genes identified in the literature, expression of canonical arterial/venous markers, and the cortical depth information embedded in the dataset via the zones, each cluster in

the TSNE object was assigned an identity and subsequently renumbered to roughly reflect the flow of blood through the kidney. The default numbers of clusters given in Seurat are assigned solely on the number of cells in the cluster, with largest cluster assigned cluster zero and the second largest assigned cluster 1, and so forth.[24] Thus this default numbering holds no anatomical information other than the number of cells collected for a certain cell type, and all TSNE plots in this paper reflect the updated numbering system rather than the default convention. Rationale for cell identities are briefly summarized in Table 2 and can be visualized on the TSNE with reference to the anatomy in Figure 10.

Table 2. Cluster Identities and Identifying Factors

Anatomical Identity	Cluster #	Identifying Factors
Large Cortical Arteries	1	Sox17 (arterial) Mgp and Eln (large vessels)
Glomerular Endothelium	2	Ehd3, Plvap(-), Nostrin, Gata5, Kdr Enriched
Arterioles/Capillaries of Cortical Capillary Plexus	3	Inferred due to arterial identity, large number of cells, and zone 1 identity
Venules/Capillaries of Cortical Capillary Plexus	4	Inferred due to venous identity, large number of cells, and zone 1 identity
Large Veins	5	Eln, Mgp, Nr2f2 & Nrp2 (venous)
Descending Vasa Recta	6 & 7	Slc14a1, Aqp1, Sox17 (arterial), zone 2 and 3
Arterioles/Capillaries medullary microcirculation	8	inferred due to arterial identity, location on zone 2/3 identity
Apln/Apela enriched cells	9	Apela and Apln rich cells, mixed zonal identity
Venules/Capillaries Medullary Microcirculation	10	inferred due to venous identity, and zone 2 identity
Ascending Vasa Recta	11	Prox1, Nrp2 & Nr2f2 (venous), zones 2 and 3

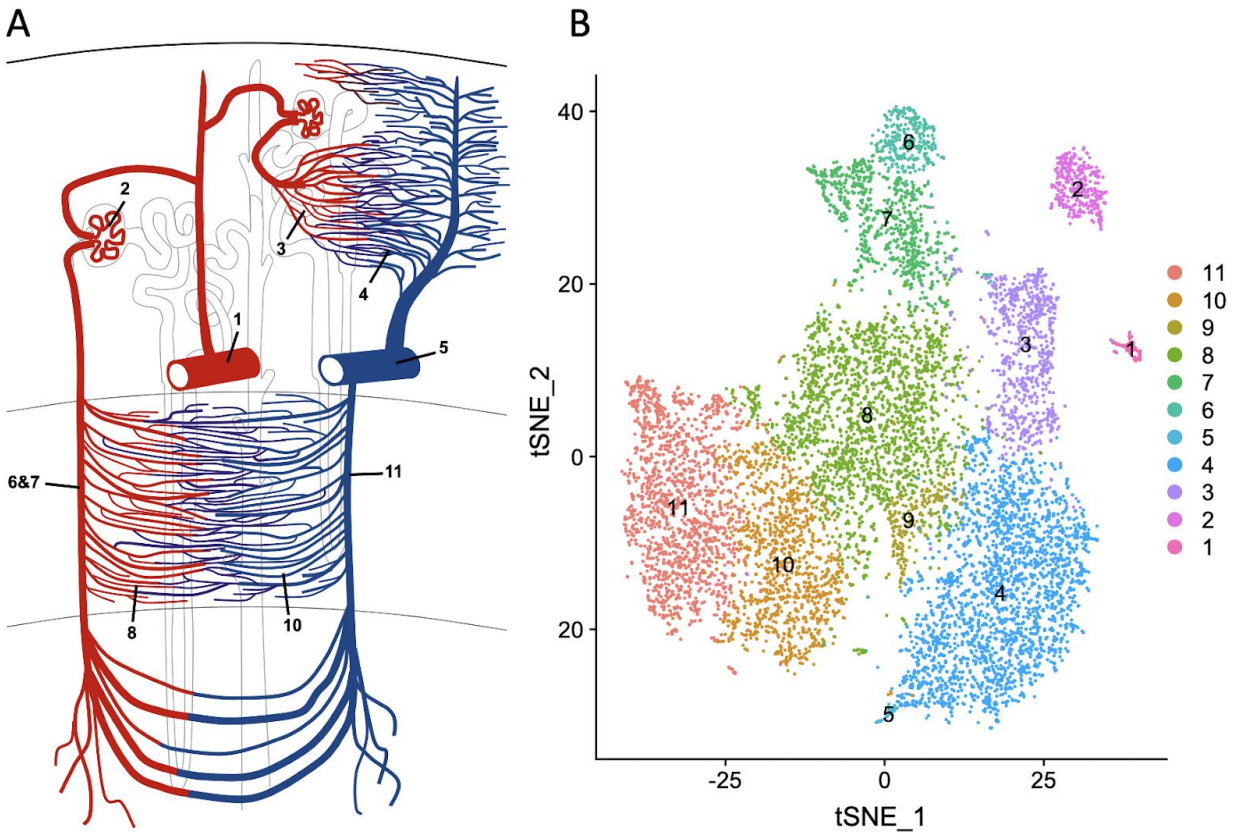

Figure 10. A) Map of the renal vasculature labeled by corresponding cluster. B) TSNE embedding of the endothelial cells. (1) Large Cortical Arteries (2) Glomerular Endothelium (3) Arterioles/Capillaries of Cortical Capillary Plexus (4) Venules/Capillaries of Cortical Capillary Plexus (5) Large Veins (6&7) Descending Vasa Recta (8) Arterioles/Capillaries medullary microcirculation (9) Apln/Apela enriched cells (10) Venules/Capillaries Medullary Microcirculation (11) Ascending Vasa Recta

From the TSNE analysis, all clusters in the dataset were assigned an anatomical compartment with the exception of cluster 9, which had a highly distinct transcriptional profile without a clear anatomical correlate. Notably, these cells are the only cells in the entire kidney that express Apln, the gene coding for the protein apelin.[40] These cells are also enriched in the apelin receptor, Aplnr,[40] and Msmp, which is a gene that has been studied in reference to

angiogenic activity, particularly in certain tumors.[41,42] Research on this population in the kidney is limited because when slices of kidney are stained for Apln, there is no clear localization of Apln$^+$ cells, only a very rare and scattered expression.[40] However, this data suggests that while the Apln$^+$ population is rare, at just 2.44% of the 10,082 endothelial cells in our sample, it has a highly distinct profile. A deeper investigation into this small population is beyond the scope of this study but is a future research interest as the apelinergenic system regulates many homeostatic processes and Apln$^+$ and Msmp$^+$ cells are of substantial interest due to their roles in angiogenic regulation.[40,41,42]

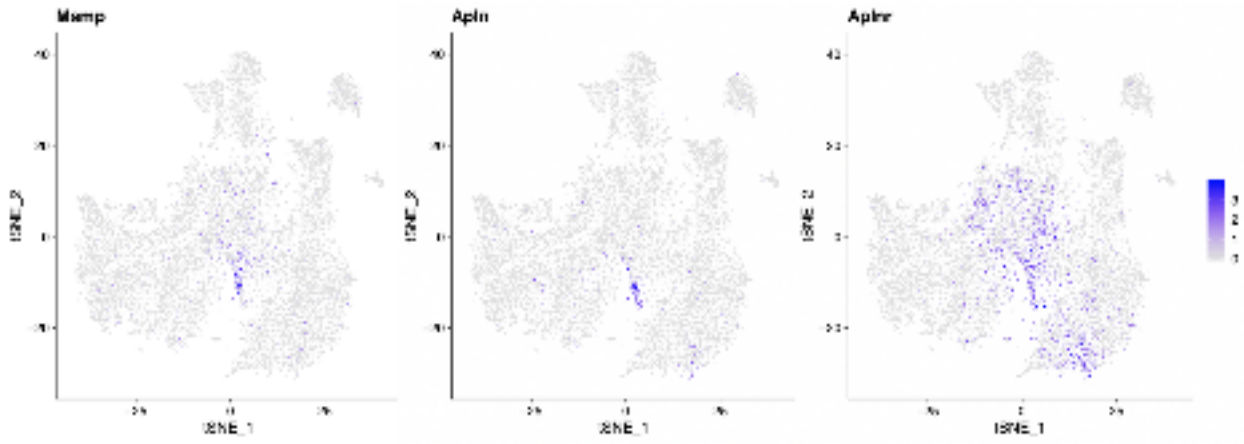

Figure 11. Expression pattern of genes enriched in cluster 9, a unique cluster without a clear anatomical correlate but a highly distinct transcriptional profile

Validation of Single Cell Data in vivo

Upon labeling the clusters based on genes based on genes identified in the literature and biological information embedded in our data (cortical depth, arterial/venous identity), we looked to see if our single cell model, now separated into different vascular compartments, could predict gene expression in the kidney. The goal was to identify distinct gene expression patterns in the single cell model and see if those patterns are reflected *in vivo*. However, when looking to verify

genes *in vivo*, it was important to consider expression in the endothelium as well as expression in other cell types in the kidney. When staining a slice of kidney tissue, the probe will come in contact with tubular and interstitial cells as well as the endothelial cells. If a gene is specific to the vasa recta in the endothelial dataset but also has broad expression in the nephron tubules, then the stain would show both of these signals, and because the tubular components are tightly interwoven with the vasculature throughout the kidney, it can be challenging to visually separate a tubular signal from a vascular signal. Despite these challenges, high quality stains were generated for many genes, eight of which are shown here in conjunction with the single cell data. These genes are Gata5, Pi16, Akr1b3, Igf1, Jag1, Ifit1, Fabp5, and Lef1.

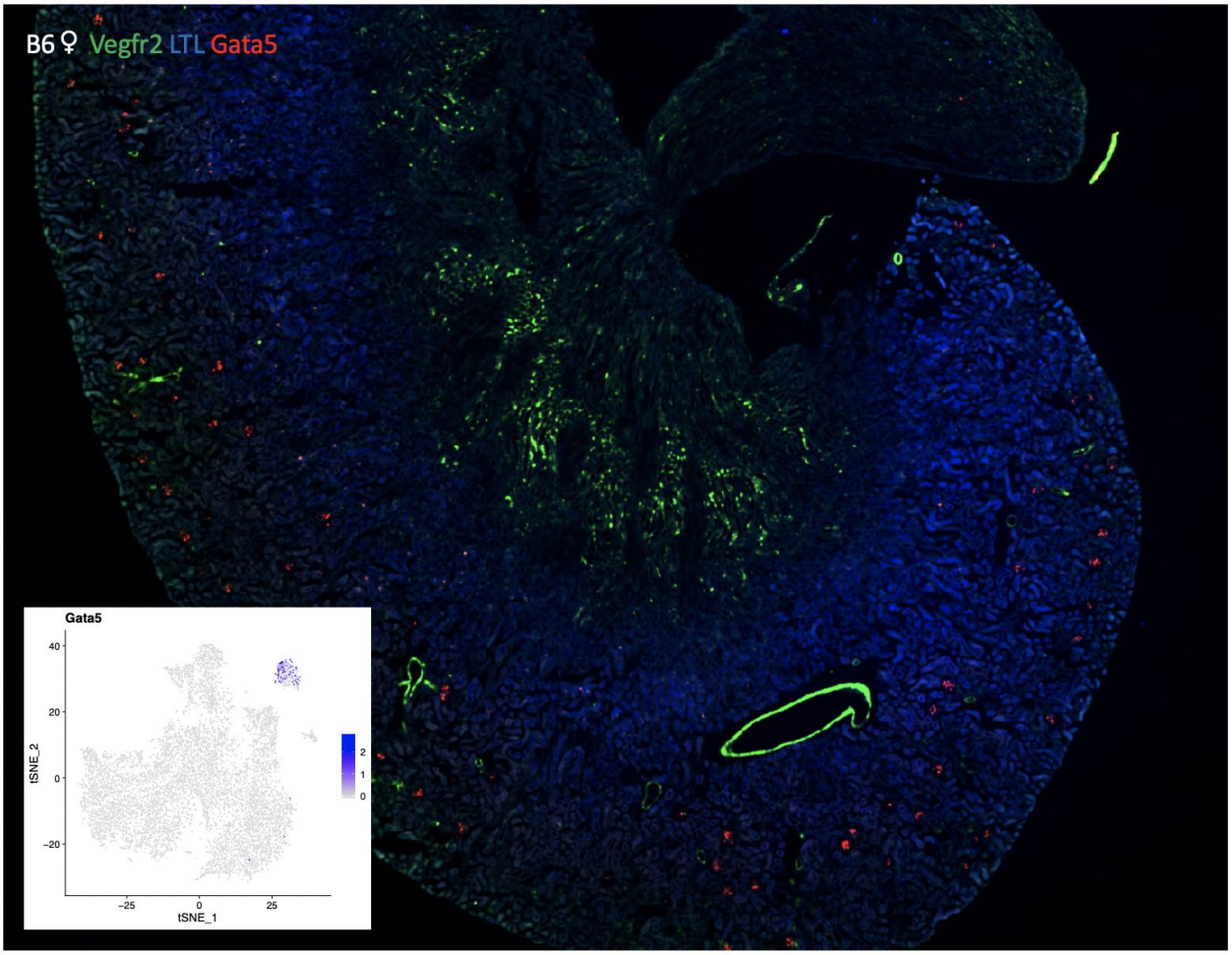

Figure 12. Kidney tissue showing Gata5 expression and predicted expression from TSNE. The TSNE plot shows expression of Gata 5 (in blue) restricted to cluster 2, predicting expression only in the glomerular endothelium. The stain shows Gata5 (in red) restricted to the glomerular endothelium as predicted. Immunohistochemistry and RNAscope *in situ* hybridization by Kari Koppitch.

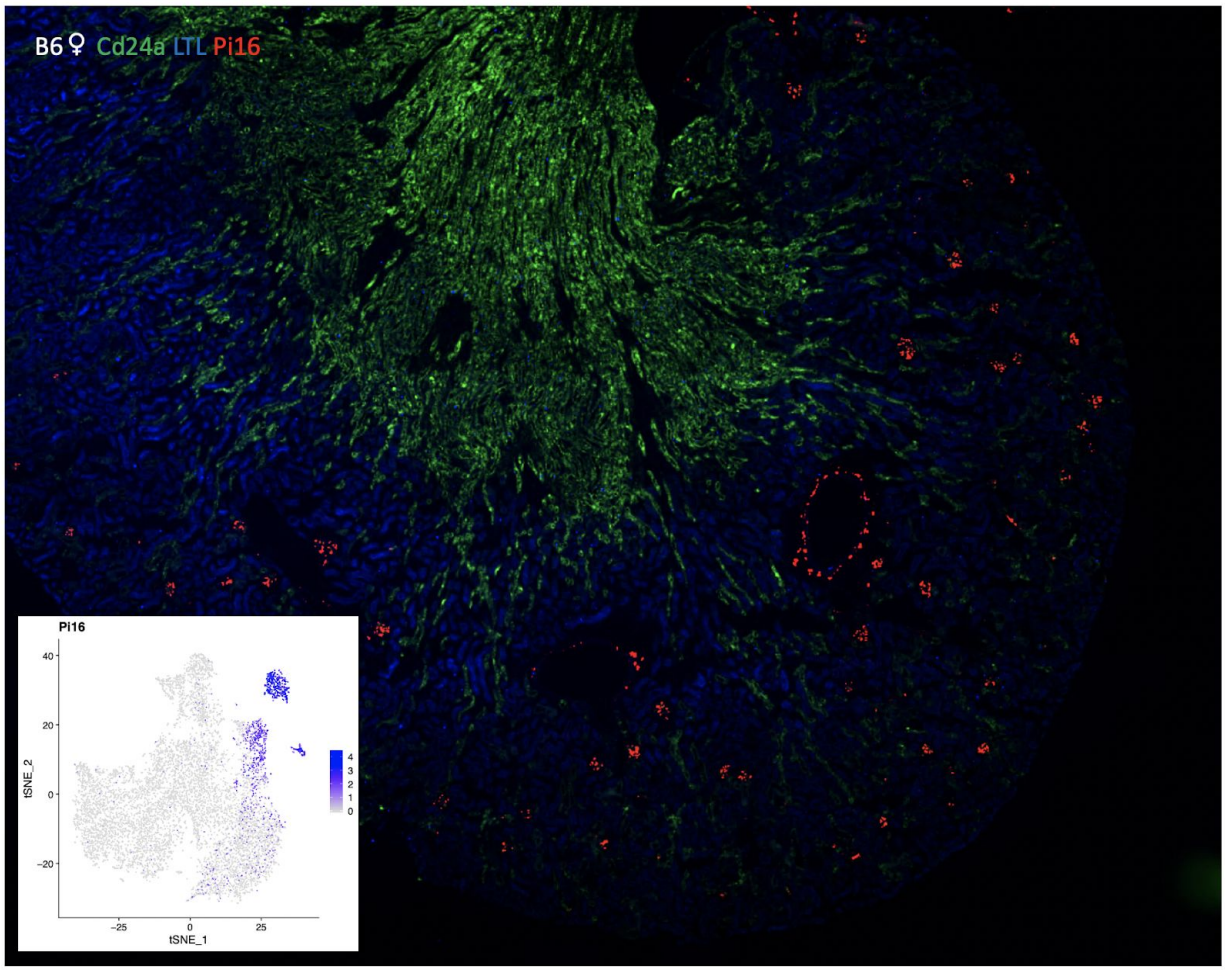

Figure 13. Kidney tissue showing Pi16 expression and predicted expression from TSNE. The TSNE plot shows expression of Pi16 (in blue) enriched in clusters 1, 2, and 3, predicting expression in the large arteries, glomerular endothelium, and arteries of the cortical capillary bed. The stain shows Pi16 (in red) enriched in the large arteries and glomerular endothelium as predicted (close up of the cortical arteries not available from this scan). Immunohistochemistry and RNAscope *in situ* hybridization by Kari Koppitch.

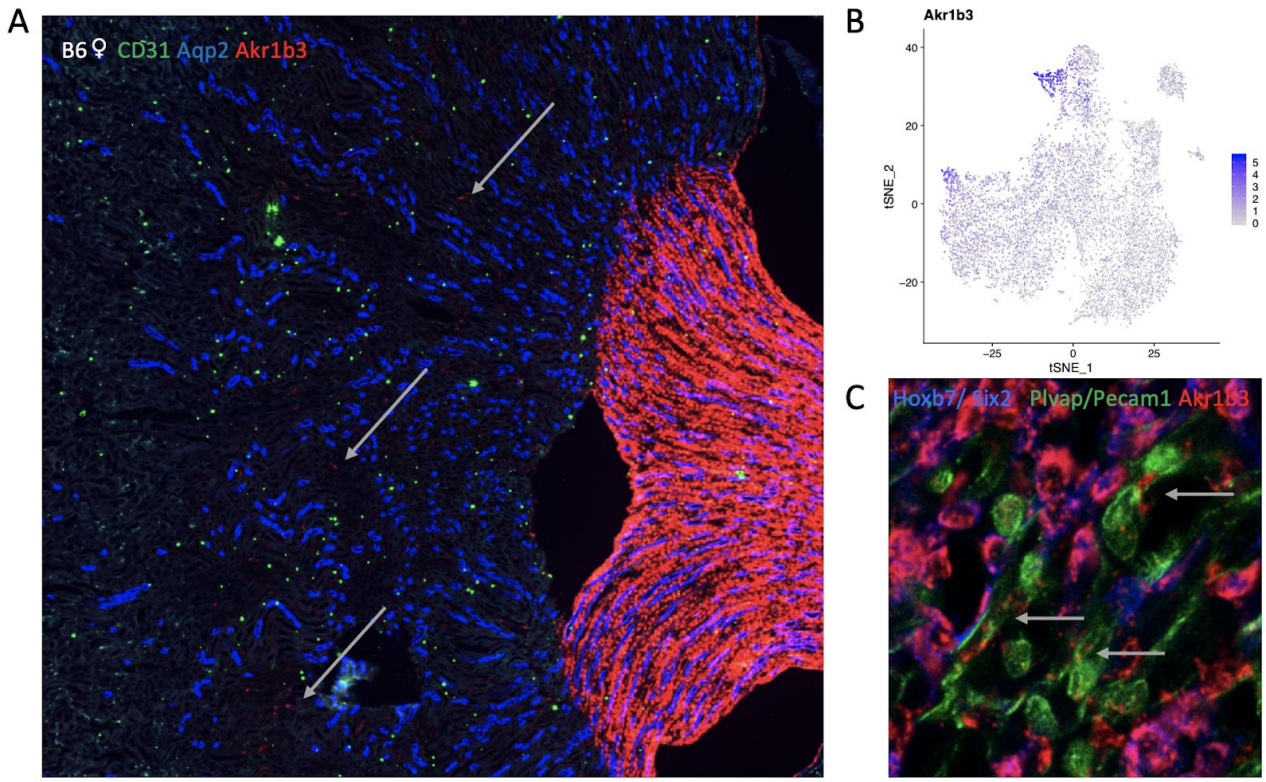

Figure 14. A) Kidney tissue showing Akr1b3 expression. Arrows point to expression of Akr1b3 outside the papilla, which is easy to miss due to the bright signal from the papilla. B) Akr1b3 predicted expression from TSNE, where expression is upregulated in the clusters representing the ascending and descending vasa recta (7 and 11), which exist in the inner medulla. C) up close image showing Akr1b3 expression within endothelial cells via coexpression with endothelial cell markers Plvap/Pecam1. Immunohistochemistry and RNAscope *in situ* hybridization by Kari Koppitch.

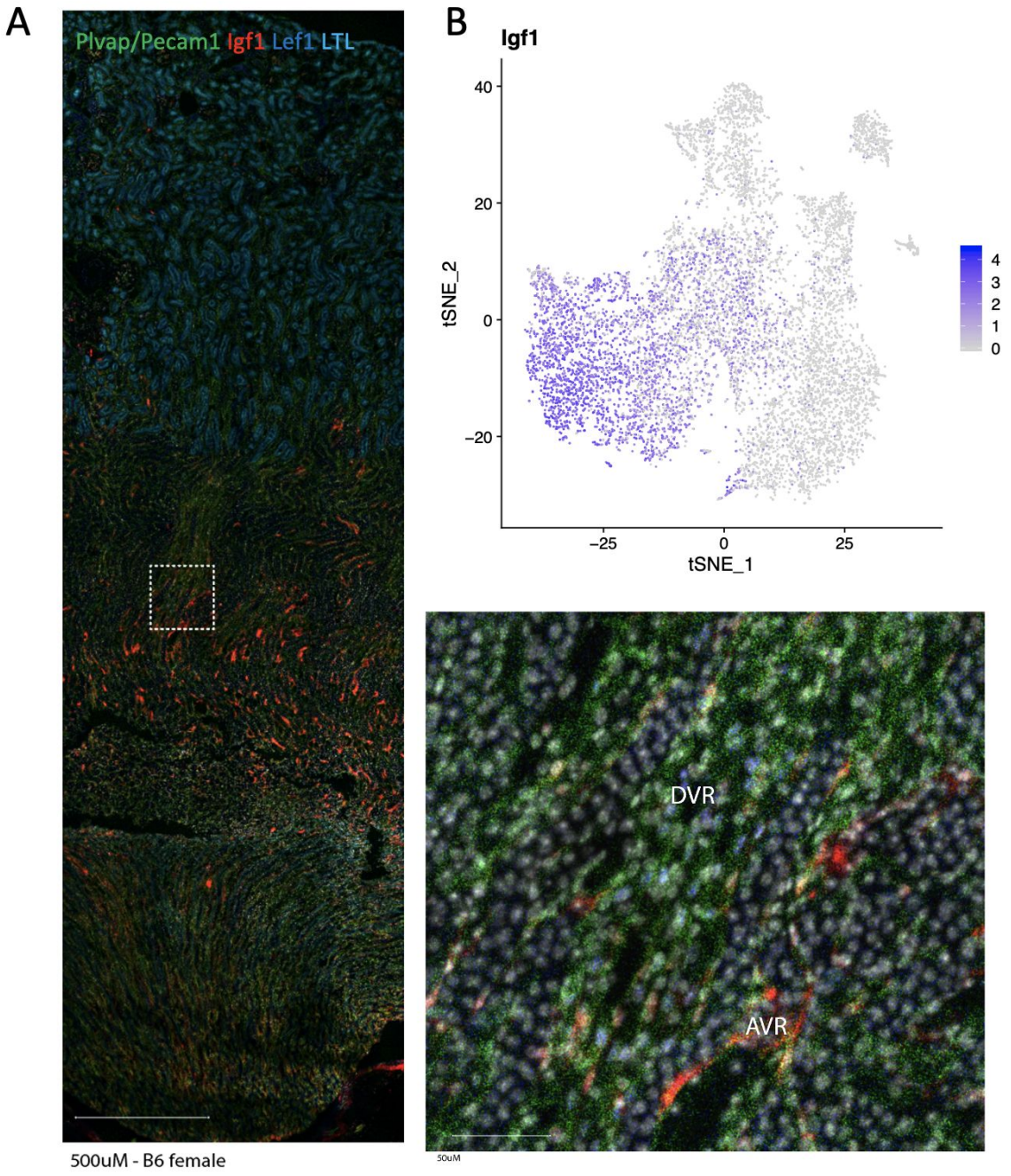

Figure 15. A) Igf1 stain shows signal in the inner and outer medulla. B) Igf1 expression in the TSNE predicts expression restricted to the inner and outer medulla and enriched in the endothelial cells of the ascending vasa recta and not in the endothelium of the descending vasa recta. C) Up-close image showing Igf1 expression within certain endothelial cells (ascending vasa recta, Igf1 positive) but not others (descending vasa recta, Lef1 positive). Immunohistochemistry and RNAscope *in situ* hybridization by Kari Koppitch.

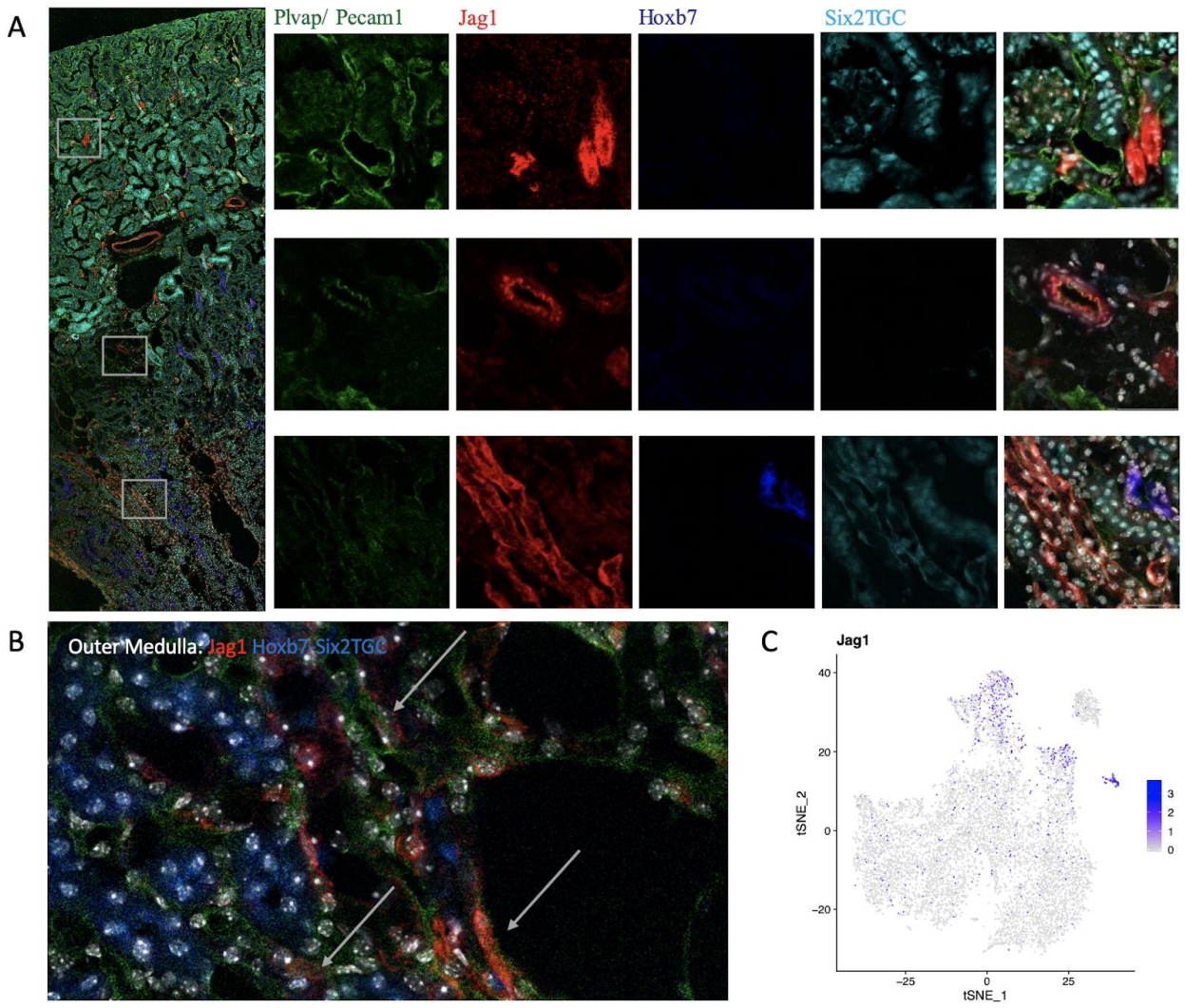

Figure 16. A) Kidney tissue showing Jag1 expression in the cortex, outer medulla, and inner medulla, in arterioles, large arteries, and vasa recta endotelium respectively. B) Jag1 expression in endothelial cells in the outer medulla demonstrated by coexpression with Plvap/Pecam1 (green) C) TSNE predicts Jag1 expression in the large arteries (cluster1), arterioles of the cortical capillary bed (cluster 3), and descending vasa recta cells (clusters 6&7). Immunohistochemistry and RNAscope *in situ* hybridization by Kari Koppitch.

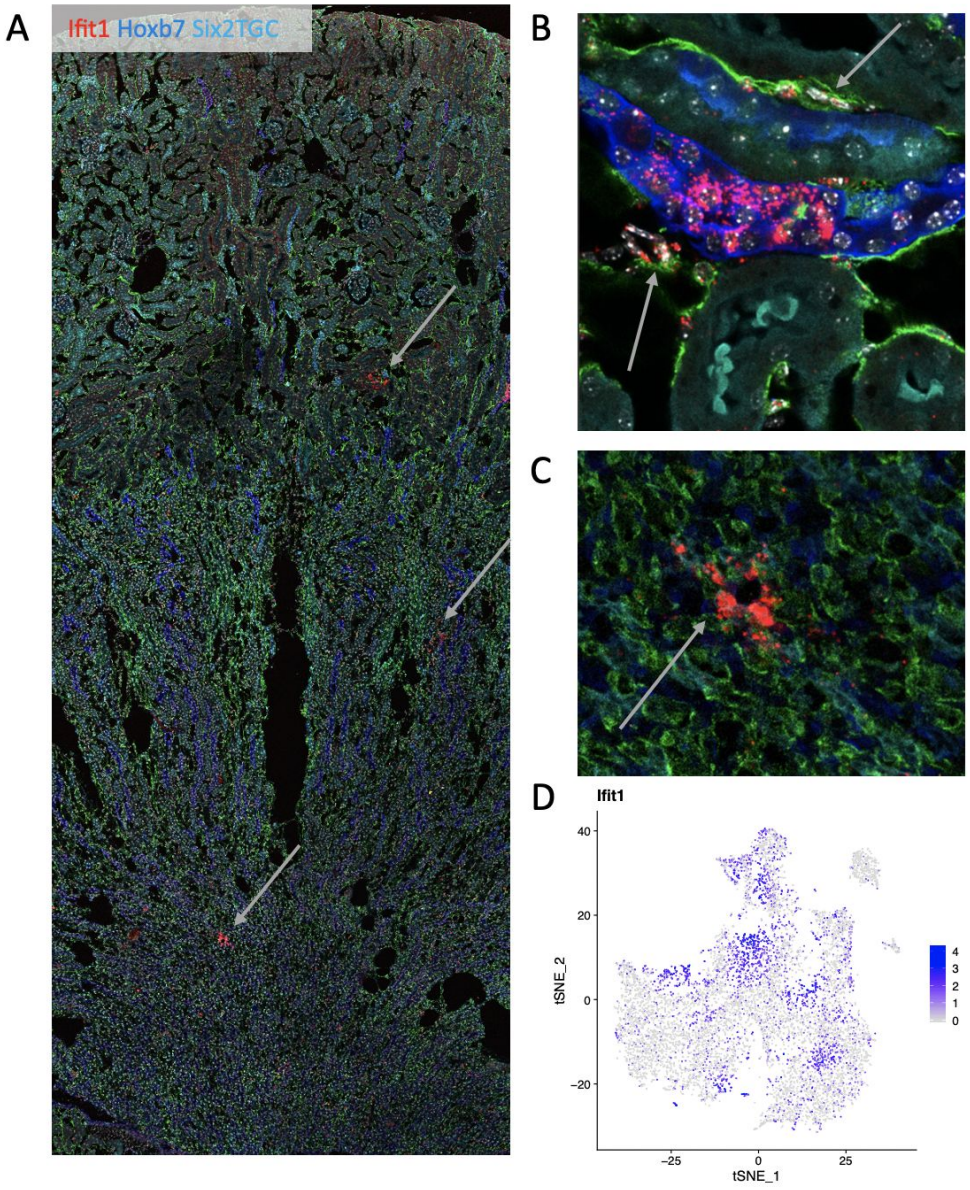

Figure 17. A) Ifit1 expression in kidney slice highlighted by arrows B) Up-close image showing Ifit1 expression in tubules as well as adjacent endothelial cells as demonstrated by coexpression with Plvap/Pecam1 (green) C) Ifit1 predicted expression from TSNE in scattered endothelial populations. Immunohistochemistry and RNAscope *in situ* hybridization by Kari Koppitch.

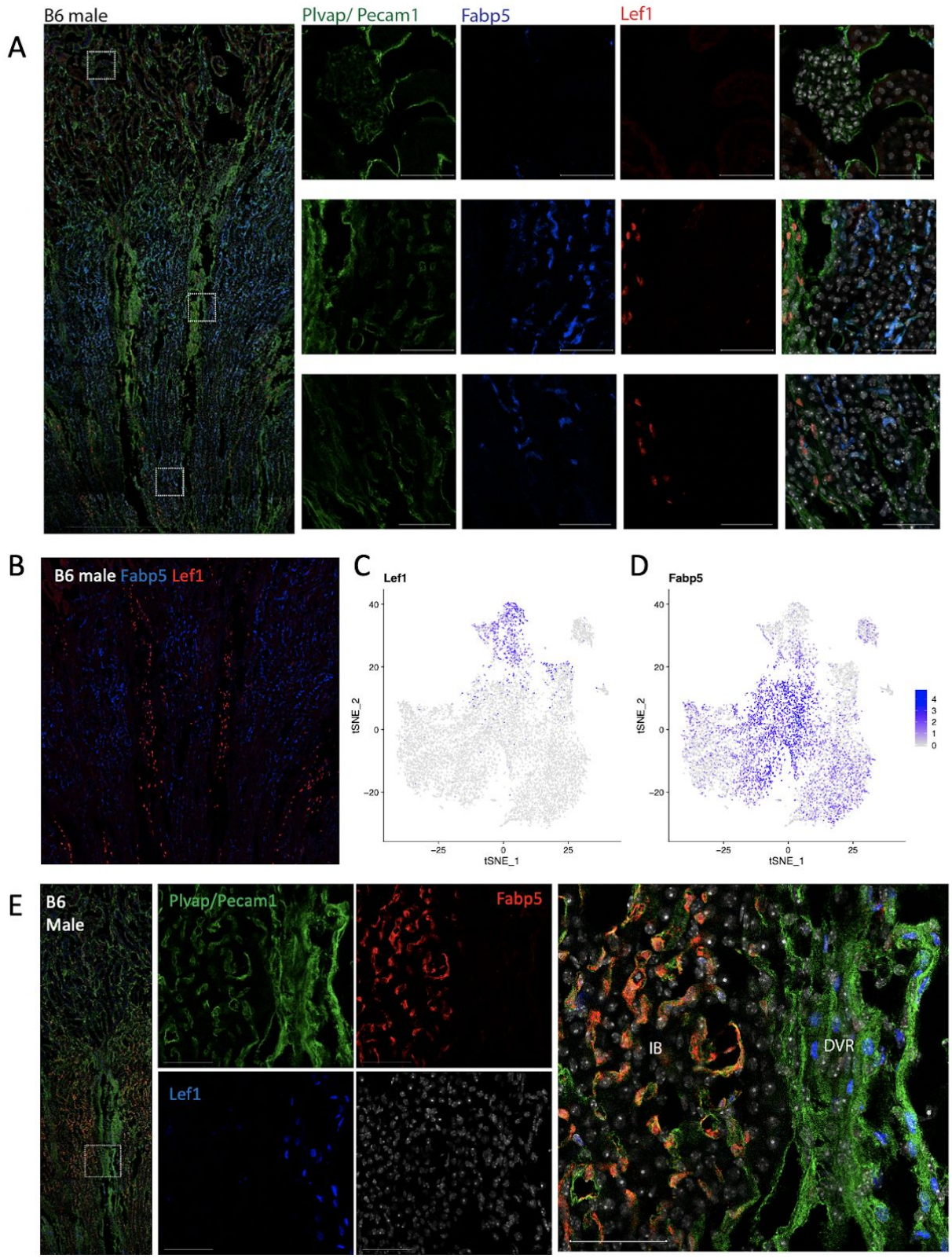

Figure 18. A) Lef1 and Fabp5 expression in the cortex, outer medulla, and inner medulla B) Lef1 and Fabp5 expression highlighted in the interbundle region without Pecam1/Plvap signal C) Lef1 predicted expression from TSNE D) Fabp5 predicted expression from TSNE E) Up close view of the interbundle plexus and vascular bundle junction showing both Fabp5 and Lef1 in the endothelium. Immunohistochemistry and RNAscope *in situ* hybridization by Kari Koppitch.

Collectively, these stains serve as a proof of concept that information learned through our single cell endothelium dataset can be validated *in vivo*. This lends credibility to our assignments of cell identity within the single cell data, and demonstrates how powerful single cell data can be in mining large sums of information from complicated organs and tissues. Our endothelial dataset had 19,125 distinct RNA's expressed across 10,082 endothelial cells that could successfully be organized in such a way that genes mined from the single cell data could be demonstrated to be expressed as predicted *in vivo.* This type of study enables researchers to examine these countless genes across many cell types simultaneously and prevents the need for more tedious, challenging and time consuming studies that might be needed to study these gene expression patterns or cell types individually. By collecting a comprehensive dataset using single cell RNAseq, labeling it through established markers and anatomical landmarks embedded in the data (corticomedullary depth, as represented in the zones), researchers can rapidly uncover new genetic targets for further research.

SWNE Analysis

With *in vivo* stains confirming the assigned anatomical identities in the single cell data, we wanted to return to the bioinformatics and investigate the idea of zonation in the renal vasculature. Zonation is the idea of a gradual phenotypic range in cell types, and it was described in the context of the vasculature in the brain in 2018.[22] This paper isolated brain endothelium, collected single cell data, and then computationally recreated the blood vessel transition from arteries through capillaries to veins.[22] In doing so, they found that these endothelial cells demonstrated a gradual change in genotype rather than having highly distinct genotypes for

arterial, capillary, and venous endothelial cells.[22] The kidney has a highly organized vasculature, so we sought to see if we could identify the same trend in our single cell data.

It's worthwhile noting that while the brain certainly has a well organized vasculature within the organ, it does not contain a fundamental repeating unit of vasculature like the endothelium in the kidney, which is paired with the repeating unit of the nephron. It is also worthwhile to note that the software they used to identify trajectories through their single cell data was housed in a different language than the bioinformatics performed here.[22] Thus, we opted to use the package Cellrouter,[26] hosted in R, which was released in 2018 and built to identify trajectories through single cell data. This would serve to simplify the analysis process and, if zonation could be identified in our endothelial dataset, it would corroborate the idea that zonation is a feature of the endothelium and could be identified using a variety of computational techniques.

In the renal endothelium, there are two distinct gradients we wanted to study, the transition from artery to vein, and the loop through the vasa recta, which passes into and then back out of the highly hypoxic inner medulla. Thus, we will be investigating the arteriovenous transition as well as the cortico-medullary transition.

Preliminary analysis performed on an older iteration of the data aligning cells along the first principal component suggested that we might find zonation in our data (data not shown). However, aligning cells along the first principal component is not the most rigorous way to study zonation because while the results may suggest zonation, by visualizing the cells along a principle component (which are defined here by gene expression profiles) you may be inadvertently ordering the cells by increasing or decreasing expression of the genes of interest

rather than by the order they would exist in in the biology. In other words, depending on what factors play into the principle component, you might be artificially creating something that looks like zonation due to your analysis technique.

However, before we began to rigorously investigate the idea of zonation in our data, we wanted to re-visualize the data using a new tool called SWNE which allows single cell data to be plotted with a level of biological information embedded in the visualization itself, creating a visualization technique that can more accurately reflect the biology.[25] While the data in our TSNE does appear to be ordered with a gradual change from artery to vein, and is indeed organized via depth (as we know the biology is), the TSNE computational technique does not intrinsically use any biological factors when plotting the data, and having this secondary visualization technique would bolster our confidence that the patterns (rather than simply the cluster identities) that appear in the TSNE are indeed reflecting the biological organization.

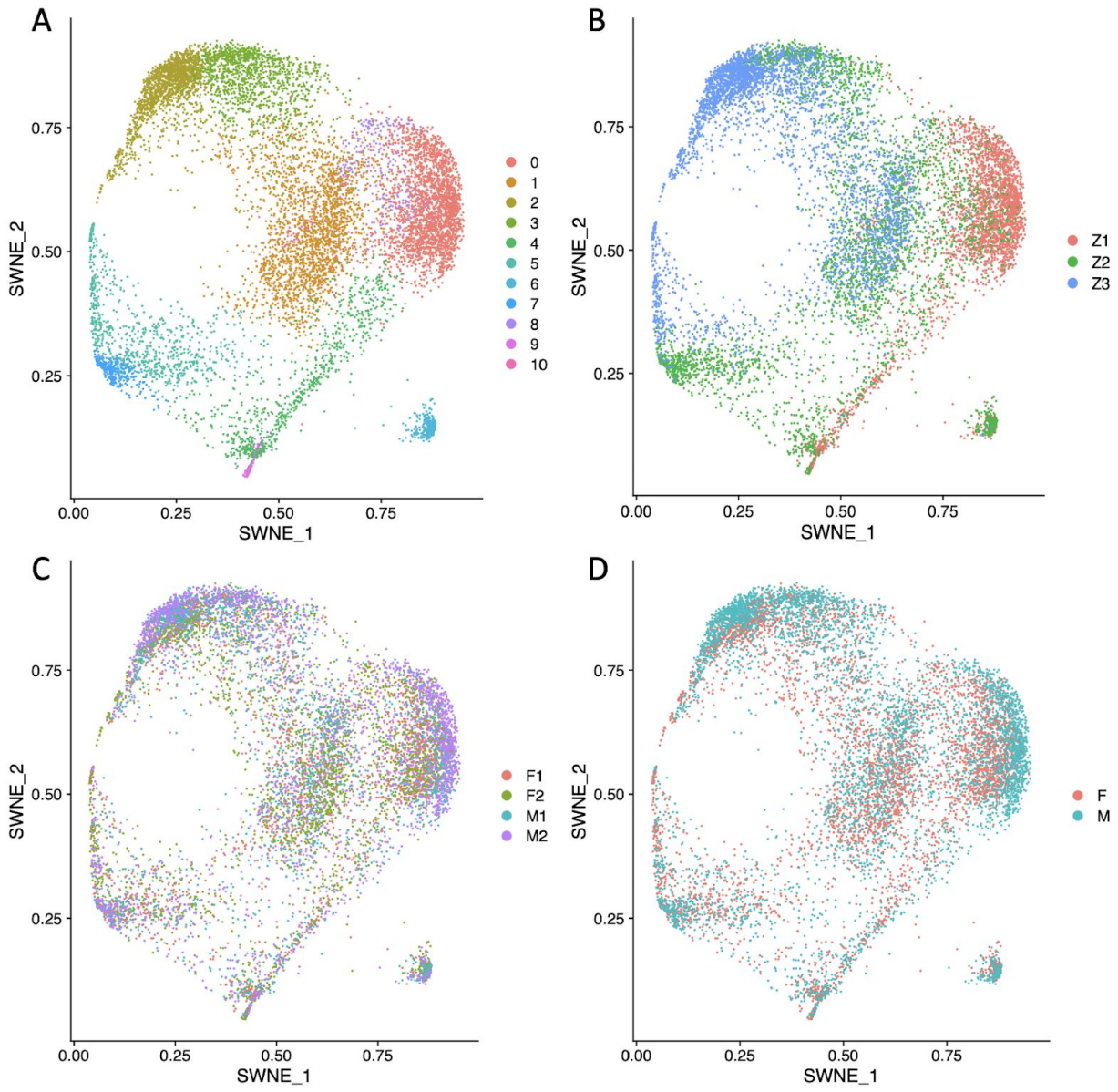

Figure 19. SWNE plots of the adult renal endothelium, where (as in a TSNE plot) each dot represents one cell in the endothelial dataset. A) SWNE plot colored by cluster B) SWNE plot colored by zonal identity C) SWNE plot colored by replicate D) SWNE plot colored by sex

Upon analyzing the data in SWNE, we found that the overall structure of the data showed that which we would expect to see based on the anatomy. As with the TSNE, the SWNE plot showed an arteriovenous gradient (Figure 23B) and a cortical-medullary axis (Figure 19B), as

well as a good distribution between the replicates and sexes in our dataset (Figure 19C&D). The SWNE also visually revealed two new relationships in our data.

Notably, the populations we predicted represented the descending and ascending vasa recta formed a loop in the SWNE diagram, which we would expect based on the anatomy because the descending vasa recta flows into the ascending vasa recta, forming a loop in the inner medulla. In the TSNE, there was no direct connection between these populations (Figure 10). Additionally, the SWNE plot appeared to recreate the three paths that blood through the vasculature, through the cortical capillary bed, medullary capillary bed, and the vasa recta loop. Again, this was an independent visualization technique that includes a layer of biological information in the embedding itself, and by superimposing the cluster identities from our initial Seurat analysis onto this embedding we can see that the SWNE plot placed the cells exactly where we would expect them based on the anatomy, even recreating the three paths the blood can take through the kidney.

It is important to note that an exception to this organization is that the glomerular endothelial cells form a small "island cluster" off on their own, rather than continuous with the rest of the endothelial cells in the sample (Figure 19A). However, this is not unexpected and the reason is quite intuitive. The glomerular endothelial cells are a very special subset of endothelial cells that exist within the unique environment of Bowman's capsule. They are surrounded by unique cells called podocytes and an extracellular matrix distinct to this environment.[43] With this unique role and environment, it is understandable, perhaps even expected, that the genetic profile of these endothelial cells differs markedly from that of endothelial cells in the rest of the kidney, and thus causes them to computationally organized as distinct cluster that is discontinuous with

the rest of the endothelial cells, which share much more in common with each other. This distinction can be seen in both the TSNE and SWNE plot where the glomerular endothelial cells exist as a distinct cluster separate from the main body of endothelial cells (Figure 10, Figure 19).

Zonation Analysis using Cellrouter

With confidence in our analysis and the assignments of cluster identities we proceeded to investigate the idea of zonation in the renal endothelium using Cellrouter. First, we went to investigate the idea of zonation along the artery to vein transition. Unlike in the brain zonation analysis,[22] where all endothelial cells were all considered together, there are three distinct tracts identified in our single cell dataset where vessels transition from artery to vein. In order to most clearly be able to compare the results from our artery-vein transition to that of the brain zonation analysis, we chose to first study the arteriovenous transition through the cortical capillary bed as opposed to the medullary capillary bed or the vasa recta loop. The reason for this is that the cortical capillary beds is not as directly involved with the countercurrent exchange system as the medullary capillary bed and vasa recta loop, which have a unique extracellular environment subject to wide fluctuation as a result of this specialized role.[44]

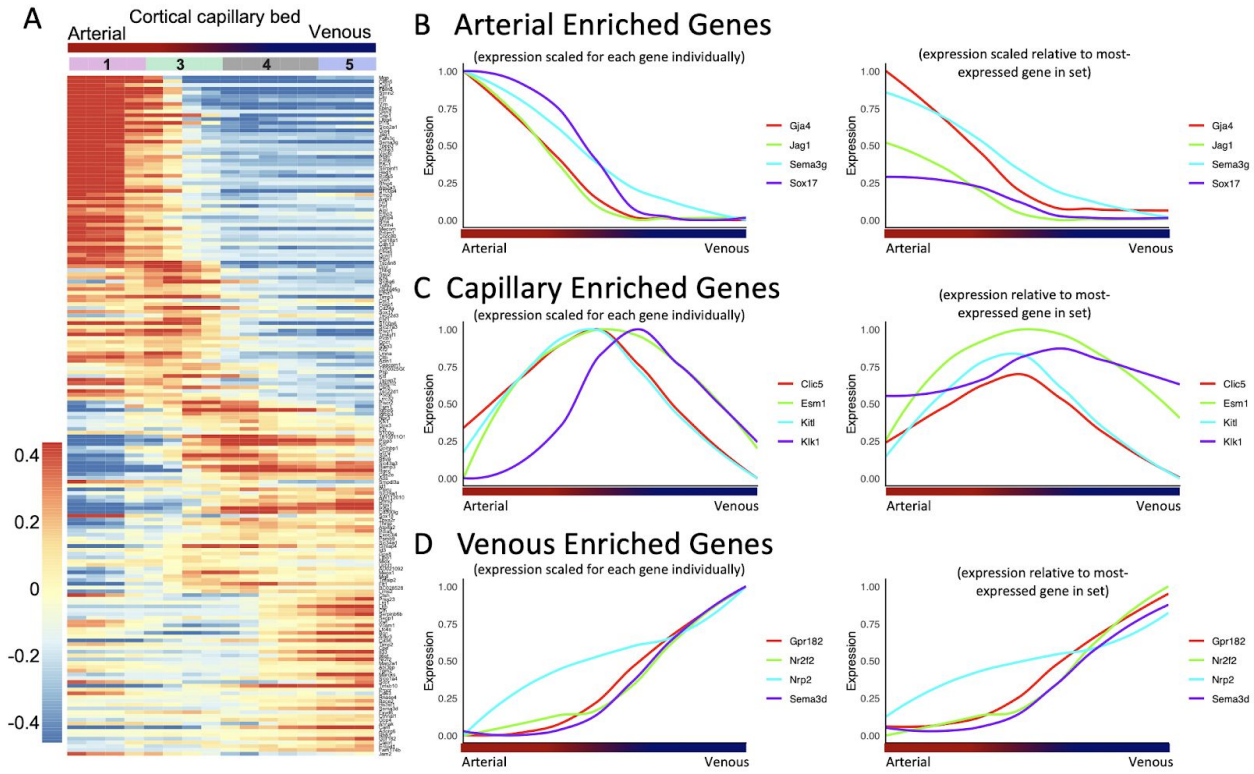

Figure 20. A) Heatmap showing genes differentially expressed along the cellrouter trajectory from artery to vein in the cortical capillary bed. This heatmap shows a gradual transition of transcriptional profile from artery to capillary to vein rather than 3 highly transcriptionally distinct and homogenous populations, suggesting zonation. B, C, &D) Variation in individual gene expression patterns shown along the cellrouter trajectory from artery to vein in the cortical capillary bed for arterial, capillary, and venous enriched genes, respectively. In each set of genes, expression is shown scaled for each gene individually (to show how each gene's trajectory varies, left) and scaled relative to the gene most expressed in the set of genes (to show how absolute levels of expression compare between the genes, right).

The Cellrouter analysis of the cortical capillary bed cells revealed evidence of zonation. Markers of arterial expression change such as Gja4 and Sox17 were downregulated across the trajectory from artery to vein, and markers of venous expression change such as Nr2f2 and Plvap upregulated across the trajectory. These transitions occur in a gradual fashion where there is overlap in expression rather than sharp cut offs associated with the transition between arteries and veins. We also found several genes upregulated in the capillaries such as Esm1 and Kitl,

which were upregulated and then downregulated across the trajectory, similar to those genes identified as being enriched in the capillaries in the brain endothelium.[22] Ultimately, we pulled out the genes differentially expressed across the trajectory and plotted a heatmap to show how the DEGs changed along the trajectory (Figure 20).

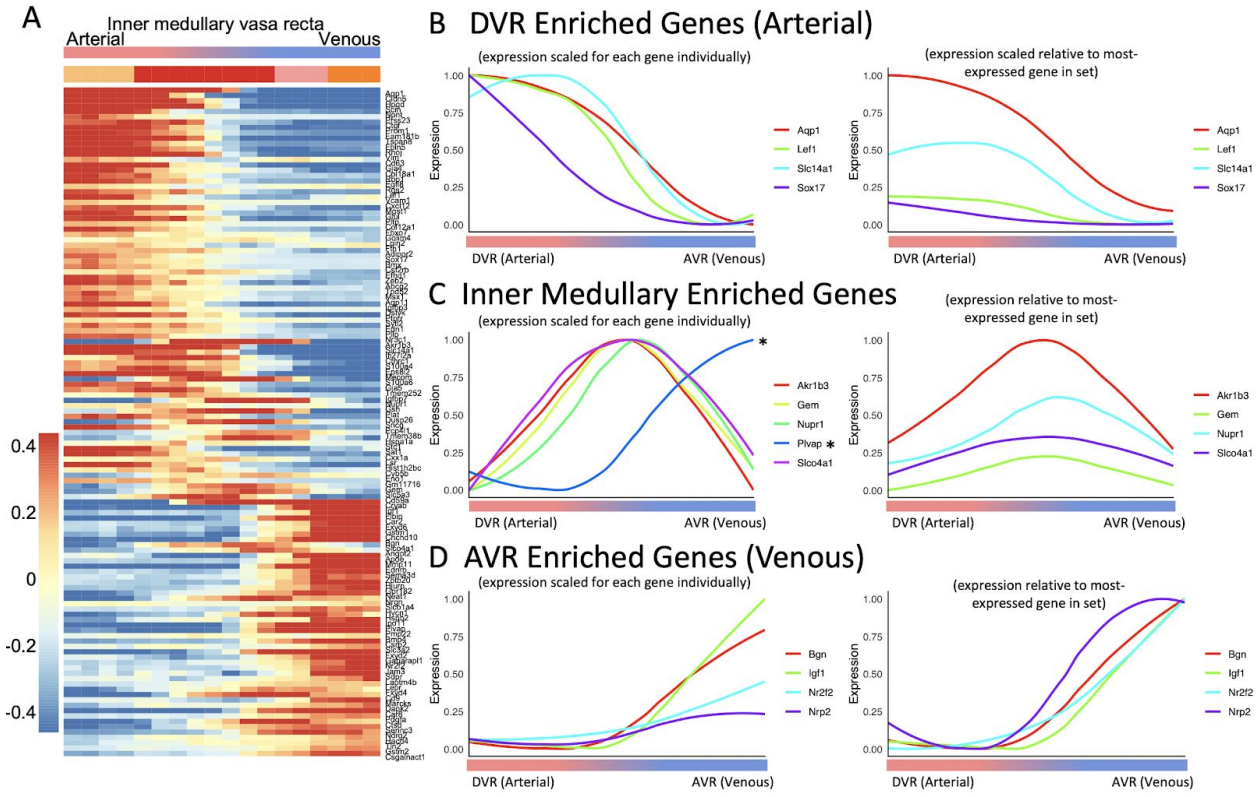

Figure 21. A) Heatmap showing genes differentially expressed along the cellrouter trajectory from artery to vein via the vasa recta loop from DVR to the AVR through the inner medulla. The heatmap shows a gradual upregulation and downregulation of transcription of genes associated with each celltype, suggesting zonation. B, C, &D) Variation in individual gene expression patterns shown along the cellrouter trajectory from artery to vein in the cortical capillary bed for DVR, Inner Medullary, and AVR enriched genes, respectively. In each set of genes, expression is shown scaled for each gene individually (to show how each gene's trajectory varies, left) and scaled relative to the gene most expressed in the set of genes (to show how absolute levels of expression compare between the genes, right).

We then investigated the idea of zonation in the vasa recta loop, which passes through the inner medulla of the kidney. This loop is critical to the concentration of the urine and the reabsorption of critical nutrients as part of the countercurrent exchange system,[45] and the endothelium in this vasculature transitions from artery to vein as well as into and then out the inner medulla. As expected genes enriched in the descending vasa recta are gradually down-regulated, genes enriched in the inner medulla are up and then down regulated, and genes enriched in the ascending vasa recta are gradually upregulated. Similarly, genes expressed in the inner medulla are gradually upregulated and then downregulated across the trajectory. Together, these patterns corroborate the idea of zonation that the endothelial cells experience a graduate change in transcriptional profile along the vessel rather than existing as highly distinct cell types.

An additional interesting facet of this transition is the upregulation of Plvap relative to the expression of genes enriched in the inner medulla. As previously discussed, Plvap is a gene associated with the fenestration of venous endothelium, and as seen in Figure 21C, it is upregulated before the upregulation of genes associated with medullary depth. This would suggest that the vasa recta loop becomes fenestrated before reaching the deepest point of the vasa recta loop in the inner medulla. This point has been previously identified in the literature where fenestrated endothelium has been identified in descending portions of the vasa recta.[46] Although a subtle point, this corroborates the idea that the transition from non-fenestrated to fenestrated endothelium occurs before the vasa recta loop passes through its deepest point in the inner medulla.

Comparison of P0 Endothelium to Adult Endothelium

At birth, human kidneys are largely developed and have their full number of nephrons. At birth, rat kidneys are not fully developed and are still developing.[47] Here, we sought to identify if some of the cell types and patterns we identified in the adult endothelium showed up in the P0 endothelium, and if so how similar their transcriptional profiles are to those found in the adult dataset.

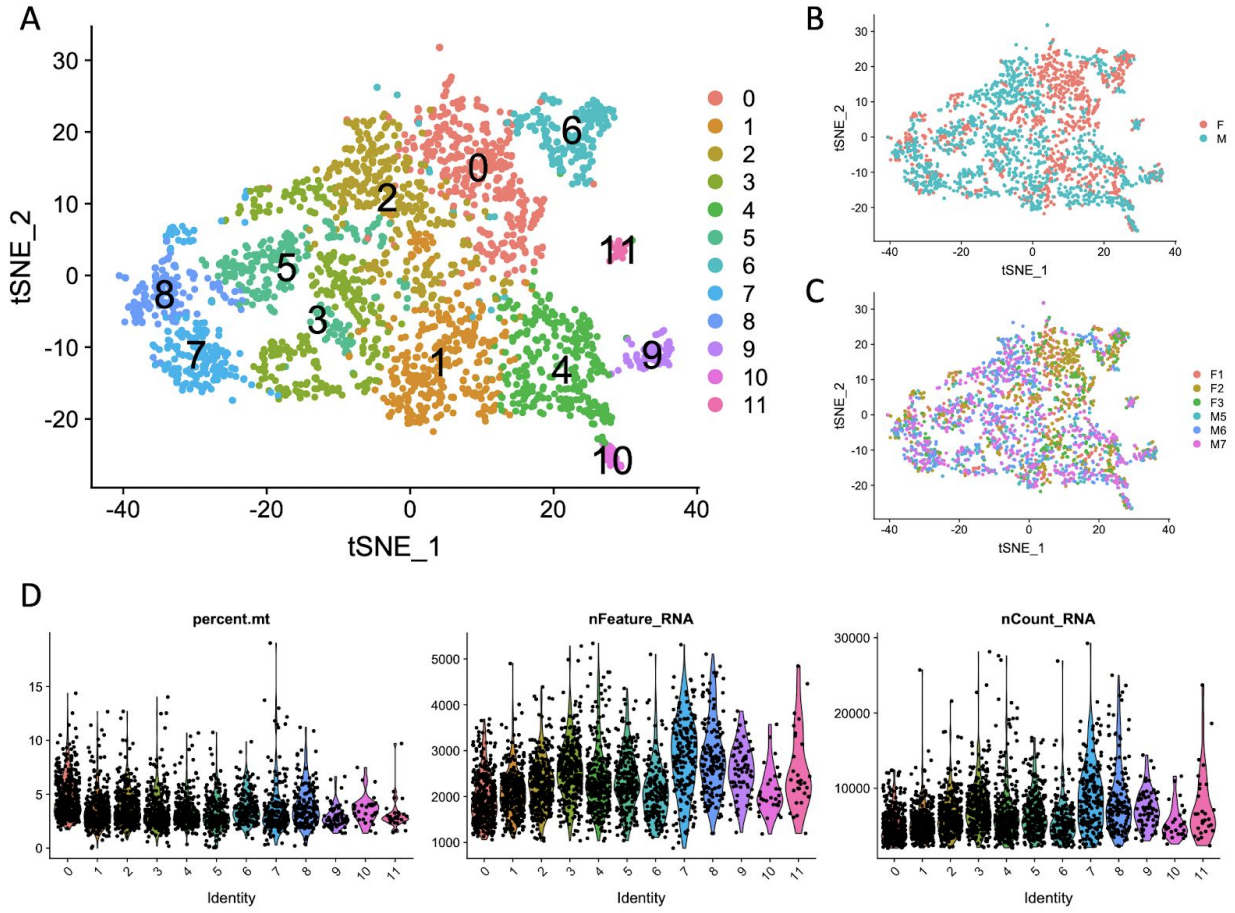

Figure 22. Summary of P0 endothelial single cell data. A) TSNE representing the P0 endothelium B) TSNE colored by sex C) TSNE colored by replicate D) Violin plots showing showing quality control metrics percent.mt (percent mitochondrial RNA), nFeature_RNA (number of distinct RNA's), and nCount_RNA (total number RNA) for the cells in each cluster.

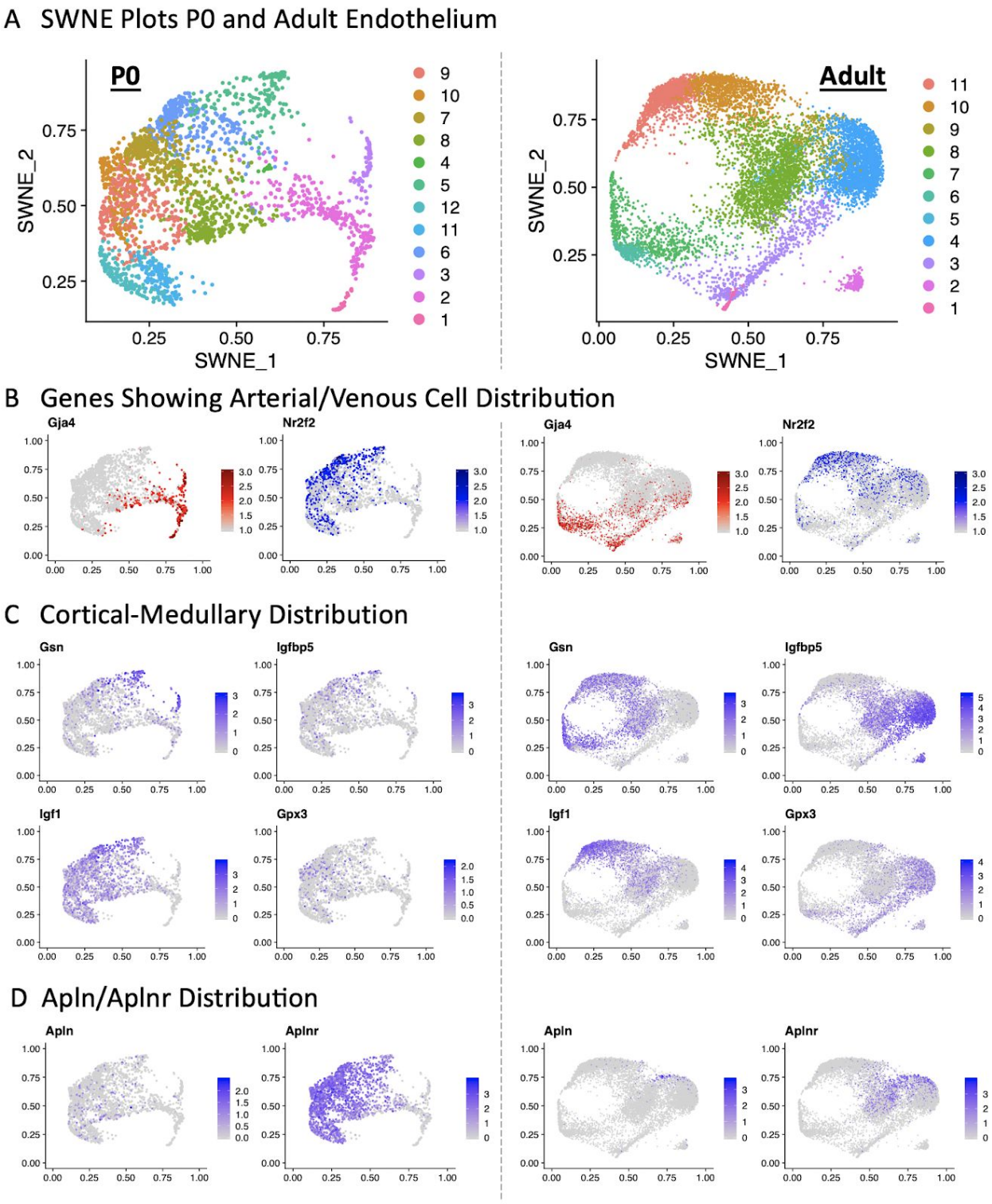

Figure 23. Comparison of endothelial single cell organization between adult and P0 kidneys A) SWNE plots for each population B) Genes representing arterial (Gja4) and venous (Nr2f2) distribution C) Cortical-medullary distribution represented by Gsn, Igfbp5, Igf1, and Gpx3 D) Apln and Aplnr distribution

Upon a preliminary clustering in seurat (Figure 22), we analyzed the data with SWNE as SWNE reflected more anatomical organization in the adult dataset. Within the SWNE, we could see that there seemed to be an arterio/venous gradient as seen in the adult (Figure 23). In contrast to the adult dataset though, the distinction between the cortical and medullary capillary bed endothelium does not appear in the single cell data. Interestingly, there also seems to be a different apln/aplnr organization in this dataset. The single cell analysis of the P0 dataset does not reveal a distinct population of $Apln^+$ cells. Instead, these cells are scattered throughout the SWNE embedding suggesting that, relative to other transcriptional differences in the dataset, the $Apln^+$ cells do not share as much in common with each other as with other cells in the dataset. Additionally, the P0 endothelium showed a much broader expression of Aplnr.

Conclusions and Future Work

In this study, we collected renal samples from adult mice to create a full kidney dataset of 10,083 cells. From this full dataset, we isolated the vasculature compartment based on canonical endothelial markers and proceeded to match clusters with their anatomical compartments to create a single cell atlas of the renal vasculature. The identities assigned to the relative clusters were then validated within *in vivo* scans. We then investigated the idea of zonation in the renal vasculature. In order to do this we first re-analyzed our data in SWNE such that our identities and their orders had a level of biological information embedded in the overall structure. We then used Cellrouter to study two distinct gradients in the kidney: the transition from artery to vein and the transition through the inner medulla. In studying the transition from artery to vein we looked at the cortical capillary bed and saw gradual transcriptional change consistent with the idea of zonation. In studying the transition through the medulla we studied the vasa recta and again found evidence of zonation as well as several genes specifically upregulated in the inner medulla. Lastly, we collected renal samples from p0 in order to investigate key gradients identified in the adult and found that the single cell data appeared to reflect an arterial/venous organization and unique Apln/Aplnr distribution.

In building upon this research, my first aim would be to convert the endothelial single cell dataset with its anatomical correlations into a form that could be easily accessed by another researcher. There is more information available in the dataset than could be fully mined by a single researcher and effectively communicated via traditional publishing avenues. It is for this reason that Ransick et al.[10] built a publicly accessible searchable website to help spread the information they learned about the nephron via single cell data to a broader audience

(https://cello.shinyapps.io/kidneycellexplorer/). The need for tools like this is clear because of the amount of coding knowledge needed to engage with single cell data in the first place. Until the tools available to study single cell data are accessible with click-based softwares, easily searchable websites can be used to communicate the data mined from a single cell paper to a broader audience. In alignment with this goal, a personal future aim is to identify more genes of possible clinical significance. Furthermore, I hope to mine the data further to study the apela/apln/aplnr distributions in the kidney with hopes of discovering more about the role of these genes in the kidney and throughout the body.

Acknowledgements

I would like to thank the McMahon Lab for mentorship and support throughout the two years I was fortunate enough to work there. I would also like to thank the Pomona College Chemistry Department for all of their support and guidance throughout the thesis process and my studies more broadly.

Bibliography

(1)

Vanlandewijck, M.; He, L.; Mäe, M. A.; Andrae, J.; Ando, K.; Del Gaudio, F.; Nahar, K.; Lebouvier, T.; Laviña, B.; Gouveia, L.; et al. A Molecular Atlas of Cell Types and Zonation in the Brain Vasculature. *Nature* **2018**, *554* (7693), 475–480. https://doi.org/10.1038/nature25739.

(2)

Schwartz, G. J.; Furth, S. L. Glomerular Filtration Rate Measurement and Estimation in Chronic Kidney Disease. *Pediatr Nephrol* **2007**, *22* (11), 1839–1848. https://doi.org/10.1007/s00467-006-0358-1.

(3)

Bertram, J. F.; Douglas-Denton, R. N.; Diouf, B.; Hughson, M. D.; Hoy, W. E. Human Nephron Number: Implications for Health and Disease. *Pediatr Nephrol* **2011**, *26* (9), 1529–1533. https://doi.org/10.1007/s00467-011-1843-8.

(4)

Tanner, G. A. (2009). Kidney function. Medical Physiology-Principles for Clinical Medicine. 3rd ed. Philadelphia: Lippincott Willams and Wilkins, 391-418.

(5)

Jha, V.; Garcia-Garcia, G.; Iseki, K.; Li, Z.; Naicker, S.; Plattner, B.; Saran, R.; Wang, A. Y.-M.; Yang, C.-W. Chronic Kidney Disease: Global Dimension and Perspectives. *The Lancet* **2013**, *382* (9888), 260–272. https://doi.org/10.1016/S0140-6736(13)60687-X.

(6)

Wang, H.; Zhang, L.; Lv, J. Prevention of the Progression of Chronic Kidney Disease: Practice in China. *Kidney International* **2005**, *67*, S63–S67. https://doi.org/10.1111/j.1523-1755.2005.09416.x.

(7)

Eknoyan, G.; Lameire, N.; Barsoum, R.; Eckardt, K.-U.; Levin, A.; Levin, N.; Locatelli, F.; Macleod, A.; Vanholder, R.; Walker, R.; et al. The Burden of Kidney Disease: Improving Global Outcomes. *Kidney International* **2004**, *66* (4), 1310–1314. https://doi.org/10.1111/j.1523-1755.2004.00894.x.

(8)

Perneger, T. V.; Brancati, F. L.; Whelton, P. K.; Klag, M. J. Studying the Causes of Kidney Disease in Humans: A Review of Methodologic Obstacles and Possible Solutions. *American Journal of Kidney Diseases* **1995**, *25* (5), 722–731. https://doi.org/10.1016/0272-6386(95)90548-0.

(9)

Stenvinkel, P. Chronic Kidney Disease: A Public Health Priority and Harbinger of Premature Cardiovascular Disease. *Journal of Internal Medicine* **2010**, *268* (5), 456–467. https://doi.org/10.1111/j.1365-2796.2010.02269.x.

(10)

Ransick, A.; Lindström, N. O.; Liu, J.; Zhu, Q.; Guo, J.-J.; Alvarado, G. F.; Kim, A. D.; Black, H. G.; Kim, J.; McMahon, A. P. Single-Cell Profiling Reveals Sex, Lineage, and Regional Diversity in the Mouse Kidney. *Developmental Cell* **2019**, *51* (3), 399-413.e7. https://doi.org/10.1016/j.devcel.2019.10.005.

(11)

Levey, A. S.; Coresh, J. Chronic Kidney Disease. *The Lancet* **2012**, *379* (9811), 165–180. https://doi.org/10.1016/S0140-6736(11)60178-5.

(12)

Hrdlickova, R.; Toloue, M.; Tian, B. RNA-Seq Methods for Transcriptome Analysis: RNA-Seq. *WIREs RNA* **2017**, *8* (1), e1364. https://doi.org/10.1002/wrna.1364.

(13)

Baran-Gale, J.; Chandra, T.; Kirschner, K. Experimental Design for Single-Cell RNA Sequencing. *Briefings in Functional Genomics* **2018**, *17* (4), 233–239. https://doi.org/10.1093/bfgp/elx035.

(14)

Anna Malkina- By-Anna Malkina- Last full review/revision Jul 2018| Content last modified Aug 2018 - https://www.merckmanuals.com/home/kidney-and-urinary-tract-disorders/kidney-failure/acute-kidney-injury-aki

(15)

Adam, M.; Potter, A. S.; Potter, S. S. Psychrophilic Proteases Dramatically Reduce Single-Cell RNA-Seq Artifacts: A Molecular Atlas of Kidney Development. **2017**, 8.

(16)

Joshi, S.; Satyanarayana, T. Biotechnology of Cold-Active Proteases. *Biology* **2013**, *2* (2), 755–783. https://doi.org/10.3390/biology2020755.

(17)

Toyokawa, Y.; Takahara, H.; Reungsang, A.; Fukuta, M.; Hachimine, Y.; Tachibana, S.; Yasuda, M. Purification and Characterization of a Halotolerant Serine Proteinase from Thermotolerant Bacillus Licheniformis RKK-04 Isolated from Thai Fish Sauce. *Appl Microbiol Biotechnol* **2010**, *86* (6), 1867–1875. https://doi.org/10.1007/s00253-009-2434-5.

(18)

Park, J.; Shrestha, R.; Qiu, C.; Kondo, A.; Huang, S.; Werth, M.; Li, M.; Barasch, J.; Suszták, K. Single-Cell Transcriptomics of the Mouse Kidney Reveals Potential Cellular Targets of Kidney Disease. *Science* **2018**, *360* (6390), 758–763. https://doi.org/10.1126/science.aar2131.

(19)

Chen, L.; Lee, J. W.; Chou, C.-L.; Nair, A. V.; Battistone, M. A.; Păunescu, T. G.; Merkulova, M.; Breton, S.; Verlander, J. W.; Wall, S. M.; et al. Transcriptomes of Major Renal Collecting Duct Cell Types in Mouse Identified by Single-Cell RNA-Seq. *Proc Natl Acad Sci USA* **2017**, *114* (46), E9989–E9998. https://doi.org/10.1073/pnas.1710964114.

(20)

The Tabula Muris Consortium; Overall coordination; Logistical coordination; Organ collection and processing; Library preparation and sequencing; Computational data analysis; Cell type annotation; Writing group; Supplemental text writing group; Principal investigators. Single-Cell Transcriptomics of 20 Mouse Organs Creates a Tabula Muris. *Nature* **2018**, *562* (7727), 367–372. https://doi.org/10.1038/s41586-018-0590-4.

(21)

K.A. Munger, C.K. Kost, B.M. Brenner. The renal circulations and glomerular ultrafiltration. Brenner & rector's the kidney (9th ed.), Elsevier/Saunders (2012), pp. 94-137

(22)

Vanlandewijck, M., He, L., Mäe, M. A., Andrae, J., Ando, K., Del Gaudio, F., ... & Sun, Y. (2018). A molecular atlas of cell types and zonation in the brain vasculature. *Nature*, *554*(7693), 475.

(23)

Su, T., Stanley, G., Sinha, R., D'Amato, G., Das, S., Rhee, S., ... & Roper, W. A. (2018). Single-cell analysis of early progenitor cells that build coronary arteries. *Nature*, *559*(7714), 356-362.

(24)

Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W. M.; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177* (7), 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

(25)

Wu, Y.; Tamayo, P.; Zhang, K. Visualizing and Interpreting Single-Cell Gene Expression Datasets with Similarity Weighted Nonnegative Embedding. *Cell Systems* **2018**, *7* (6), 656-666.e4. https://doi.org/10.1016/j.cels.2018.10.015.

(26)

Lummertz da Rocha, E.; Rowe, R. G.; Lundin, V.; Malleshaiah, M.; Jha, D. K.; Rambo, C. R.; Li, H.; North, T. E.; Collins, J. J.; Daley, G. Q. Reconstruction of Complex Single-Cell Trajectories Using CellRouter. *Nat Commun* **2018**, *9* (1), 892. https://doi.org/10.1038/s41467-018-03214-y.

(27)

Kiselev, V. Y.; Andrews, T. S.; Hemberg, M. Challenges in Unsupervised Clustering of Single-Cell RNA-Seq Data. *Nat Rev Genet* **2019**, *20* (5), 273–282. https://doi.org/10.1038/s41576-018-0088-9.

(28)

Katoh, O.; Tauchi, H.; Kawaishi, K.; Kimura, A.; Satow, Y. Expression of the Vascular Endothelial Growth Factor (VEGF) Receptor Gene, KDR, in Hematopoietic Cells and Inhibitory Effect of VEGF on Apoptotic Cell Death Caused by Ionizing Radiation. *Cancer Research* **1995**, *55* (23), 5687–5692.

(29)

Gerber, H.-P.; Condorelli, F.; Park, J.; Ferrara, N. Differential Transcriptional Regulation of the Two Vascular Endothelial Growth Factor Receptor Genes Flt-1, but Not Flk-1/KDR, Is up-Regulated by Hypoxia. *Journal of Biological Chemistry* **1997**, *272* (38), 23659–23667.

(30)

Duncan, G. S.; Andrew, D. P.; Takimoto, H.; Kaufman, S. A.; Yoshida, H.; Spellberg, J.; De La Pompa, J. L.; Elia, A.; Wakeham, A.; Karan-Tamir, B. Genetic Evidence for Functional Redundancy of Platelet/Endothelial Cell Adhesion Molecule-1 (PECAM-1): CD31-Deficient Mice Reveal PECAM-1-Dependent and PECAM-1-Independent Functions. *The Journal of Immunology* **1999**, *162* (5), 3022–3030.

(31)

Natkunam, Y.; Rouse, R. V.; Zhu, S.; Fisher, C.; van de Rijn, M. Immunoblot Analysis of CD34 Expression in Histologically Diverse Neoplasms. *The American journal of pathology* **2000**, *156* (1), 21–27.

(32)

Strickland, L. A.; Jubb, A. M.; Hongo, J.; Zhong, F.; Burwick, J.; Fu, L.; Frantz, G. D.; Koeppen, H. Plasmalemmal Vesicle‐associated Protein (PLVAP) Is Expressed by Tumour Endothelium and Is Upregulated by Vascular Endothelial Growth Factor‐A (VEGF). *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* **2005**, *206* (4), 466–475.

(33)

Park-Windhol, C.; Ng, Y. S.; Yang, J.; Primo, V.; Saint-Geniez, M.; D'Amore, P. A. Endomucin Inhibits VEGF-Induced Endothelial Cell Migration, Growth, and Morphogenesis by Modulating VEGFR2 Signaling. *Scientific reports* **2017**, *7* (1), 1–13.

(34)

Yano, M.; Iwama, A.; Nishio, H.; Suda, J.; Takada, G.; Suda, T. Expression and Function of Murine Receptor Tyrosine Kinases, TIE and TEK, in Hematopoietic Stem Cells. *Blood, The Journal of the American Society of Hematology* **1997**, *89* (12), 4317–4326.

(35)

Chen, Q.; Zhang, H.; Liu, Y.; Adams, S.; Eilken, H.; Stehling, M.; Corada, M.; Dejana, E.; Zhou, B.; Adams, R. H. Endothelial Cells Are Progenitors of Cardiac Pericytes and Vascular Smooth Muscle Cells. *Nature communications* **2016**, *7* (1), 1–13.

(36)

Miano, J. M.; Thomas, S.; Disteche, C. M. Expression and Chromosomal Mapping of the Mouse Smooth Muscle Calponin Gene. *Mammalian Genome* **2001**, *12* (3), 187–191.

(37)

Brunskill, E. W.; Potter, S. S. Gene Expression Programs of Mouse Endothelial Cells in Kidney Development and Disease. *PloS one* **2010**, *5* (8).

(38)

Camps, L.; Reina, M.; Llobera, M.; Vilaro, S.; Olivecrona, T. Lipoprotein Lipase: Cellular Origin and Functional Distribution. *American Journal of Physiology-Cell Physiology* **1990**, *258* (4), C673–C681.

(39)

Strickland, L. A.; Jubb, A. M.; Hongo, J.; Zhong, F.; Burwick, J.; Fu, L.; Frantz, G. D.; Koeppen, H. Plasmalemmal Vesicle‐associated Protein (PLVAP) Is Expressed by Tumour Endothelium and Is Upregulated by Vascular Endothelial Growth Factor‐A (VEGF). *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* **2005**, *206* (4), 466–475.

(40)

O'Carroll, A.-M.; Salih, S.; Griffiths, P. R.; Bijabhai, A.; Knepper, M. A.; Lolait, S. J. Expression and Functional Implications of the Renal Apelinergic System in Rodents. *PloS one* **2017**, *12* (8).

(41)

Mitamura, T.; Pradeep, S.; McGuire, M.; Wu, S.; Ma, S.; Hatakeyama, H.; Lyons, Y. A.; Hisamatsu, T.; Noh, K.; Villar-Prados, A. Induction of Anti-VEGF Therapy Resistance by Upregulated Expression of Microseminoprotein (MSMP). *Oncogene* **2018**, *37* (6), 722–731.

(42)

Apte, R. S.; Chen, D. S.; Ferrara, N. VEGF in Signaling and Disease: Beyond Discovery and Development. *Cell* **2019**, *176* (6), 1248–1264.

(43)

Satchell, S. The Role of the Glomerular Endothelium in Albumin Handling. *Nature Reviews Nephrology* **2013**, *9* (12), 717–725.

(44)

Guder, W.; Beck, F.; Schmolke, M. Regulation and Localization of Organic Osmolytes in Mammalian Kidney. *Klinische Wochenschrift* **1990**, *68* (22), 1091–1095.

(45)

Wilde, W. S.; Vorburger, C. Albumin Multiplier in Kidney Vasa Recta Analyzed by Microspectrophotometry of T-1824. *American Journal of Physiology-Legacy Content* **1967**, *213* (5), 1233–1243.

(46)

Yuan, J.; Pannabecker, T. L. Architecture of Inner Medullary Descending and Ascending Vasa Recta: Pathways for Countercurrent Exchange. *American Journal of Physiology-Renal Physiology* **2010**, *299* (1), F265–F272.

(47)

Carey, A. V.; Carey, R. M.; Gomez, R. A. Expression of Alpha-Smooth Muscle Actin in the Developing Kidney Vasculature. *Hypertension* **1992**, *19* (2_supplement), II168.