Uned Ymchwil Arennol Cymru
Wales Kidney Research Unit

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Single-nucleus RNA sequencing of proximal tubular cells in progressive renal fibrosis

Yueh-An Lu, MD

Thesis presented for the degree of Philosophiae Doctor

May 2022

Wales Kidney Research Unit, Division of Infection & Immunity,

School of Medicine, Cardiff University

Declaration

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed..... (candidate) Date .....

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD

Signed..... (candidate) Date .....

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed..... (candidate) Date .....

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed..... (candidate) Date .....

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans after expiry of a bar on access previously approved by the Graduate Development Committee.

Signed..... (candidate) Date .....

To my family,

how lucky I am to have a family like you.

Thank you for always supporting me and believing in me.

Acknowledgement

I would like to express my gratefulness to my supervisor Prof Donald Fraser along with my co-supervisors Prof Philip Taylor, Dr Robert Andrews, Dr Timothy Bowen and Dr Chia-Te Liao for offering me the opportunity to study a PhD in their laboratory and their continuous guidance. I would particularly like to thank Prof Donald Fraser for his close supervision, extraordinary patience, helpfulness and understanding in times of stress and difficulties with my experiments.

I am thankful to Dr Chia-Te Liao, who developed the mouse model of aristolochic acid nephropathy (AAN) and helped me in the study design and the animal handling; to Dr Bnar Talabani who always helped me with my experiment and became my first friend in UK; to Dr Katherine Simpson, Dr Emma Woods and Dr Alexa Wonnacott who helped me with the molecular biology technique and tissue culture; to Dr Irina Grigorieva who helped me with histology staining and microscope; to Dr Robert Steadman and Dr Soma Meran who always gave me advise from their point of view.

I am thankful to Dr Rachel Raybould who did the library preparation for the single-cell experiment. Without her, I could not have such a high quality of data, to Dr Angela Marchbank for her great help in next generation sequencing. I would also like to thank Dr Magdalena Czubala and Mark Gurney who helped my animal work.

I would also like to thank all the people in the WKRU and Tenovus lab for their help and patient when we working together in the lab. Thanks to all the people that shared the laboratory and study office with me during my PhD; to all those who shared advise, guidance, scientific and non-scientific.

I am thankful to my hospital in Taiwan for supporting this PhD studentship, especially to Prof Ya-Chung Tien who pushed me to start a PhD as one of the most critical carrier development plans, helped me find this great lab and came to visit me in my first year of PhD.

I am most grateful to my parents. This thesis would have never been possible without their sacrifice, education, support and encouragement. Thank for all my friends in Taiwan who spent time with me on phone when I was alone.

Thesis summary

Proximal tubular cells (PTCs) are the most abundant cell type in the kidney. PTCs are central to normal kidney function, and to kidney regeneration versus organ fibrosis following kidney injury. This study determined PTC phenotype in healthy and fibrotic kidneys by single-nucleus RNA sequencing (snRNA-Seq), aiming to improve understanding of the character of PTCs in chronic kidney disease and fibrosis. The nuclear isolation protocol was optimised to achieve the best yield and nuclear RNA quality. SnRNA-Seq using healthy kidneys and fibrotic kidneys induced by aristolochic acid injection from adult male mice was performed. PTCs mapped to five abundant clusters, corresponding to tubular segments S1, S1-2, S2-cortical S3, and medullary S3. Novel clusters that were present at low abundance in normal kidneys and in increased number in kidneys undergoing regeneration and fibrosis following injury were identified. These clusters exhibited clear molecular phenotypes and were categorised as, proliferating, dedifferentiated-intermediate, dedifferentiated-regenerating, and a dedifferentiated-senescent category that was present only after injury. Using trajectory and RNA velocity analysis, two major processes of PTC transition and differentiation were described, the path toward cellular senescence and the path toward tubular regeneration. Comprehensive pathway analyses revealed metabolic reprogramming and various immune activations in new PTC clusters. In ligand-receptor analysis, new PTC clusters promoted fibrotic signalling to fibroblasts and inflammatory activation to macrophages. The new identified PTC clusters were validated using confocal microscope. SnRNA-Seq using growing mouse kidneys from 2 and 4-week-old mice were than carried out to investigate PTC proliferation and differentiation. The 2-week-old mouse kidneys had a larger proportion of proliferative cells and the male to female difference of PTCs became significant in the 4-week-old mouse kidneys. These data identified unappreciated heterogeneity in PTC phenotypes, revealed novel PTCs associated with fibrosis and regeneration and inferred the possible PTC differentiation pathways after kidney injury.

Glossary of Abbreviations

AA	aristolochic acid
AAN	aristolochic acid nephropathy
AKI	acute kidney injury
ATAC-Seq	assay for transposase-accessible chromatin using sequencing
BMPs	bone morphogenetic proteins
bulk RNA-Seq	bulk-RNA sequencing
CKD	chronic kidney disease
CNT	connecting tubule
DCT	distal convoluted tubule
DEG	Differentially expressed gene
dNTP	deoxynucleotide
ECM	extracellular matrix
EMT	endothelial-mesenchymal transition
FCS	foetal calf serum
FDR	false Discovery Rate
FFPE	formalin-fixed paraffin-embedded
GEMs	gel beads in emulsions
GESA	gene set enrichment analysis
GFR	glomerular filtration rate
GSH	intracellular glutathione

GWASs	genome-wide association studies
IF	immunofluorescence
IHC	immunohistochemistry
IRI	ischemia-reperfusion injury
LOH	loop of Henle
MMPs	matrix metalloproteinases
mRNA	messenger RNA
NGS	next-generation sequencing
OAT	organic anion transport
OCT	organic cation transporter
PC	principal component
PCA	principal component analysis
PT	proximal tubule
PTC	proximal tubular cell
QC	quality control
RIN	RNA integrity number
RNA-Seq	RNA sequencing
SASP	senescence-associated secretory phenotype
scRNA-Seq	single-cell RNA sequencing
snRNA-Seq	single-nucleus RNA sequencing
t-SNE	t-Distributed Stochastic Neighbor Embedding
UMAP	Uniform Manifold Approximation and Projection

UMI	unique molecular identifier
αSMA	alpha-smooth muscle actin

Publications and Presentations arising from this thesis

Publications

Lu YA, Liao CT, Raybould R, Talabani B, Grigorieva I, Szomolay B, Bowen T, Andrews R, Taylor PR, Fraser D*. Single-Nucleus RNA Sequencing Identifies New Classes of Proximal Tubular Epithelial Cells in Kidney Fibrosis. J Am Soc Nephrol. 2021 Oct;32(10):2501-2516.

Woods EL, Grigorieva IV, Midgley AC, Brown CVM, **Lu YA**, Phillips AO, Bowen T, Meran S, Steadman R*. CD147 mediates the CD44s-dependent differentiation of myofibroblasts driven by transforming growth factor-β 1. J Biol Chem. 2021 Sep;297(3):100987.

Newbury LJ, Simpson K, Khalid U, John I, de Rivera LB, **Lu YA**, Lopez-Anton M, Watkins WJ, Jenkins RH, Fraser DJ, Bowen T*. miR-141 mediates recovery from acute kidney injury. Sci Rep. 2021 Aug 13;11(1):16499.

Presentations

Yueh-An Lu, Chia-Te Liao, Rachel Raybould, Bnar Talabani, Barbara Szomolay, Timothy Bowen, Robert Andrews, Philip R. Taylor, Donald Fraser. Single nucleus RNA sequencing identifies new classes of renal proximal tubular epithelial cell in a chronic renal fibrosis model. UK Kidney Week 2020, Oct. 2020, United Kingdom, Oral presentation.

Yueh-An Lu, Chia-Te Liao, Rachel Raybould, Bnar Talabani, Barbara Szomolay, Timothy Bowen, Robert Andrews, Philip R. Taylor, Donald Fraser. Single-nucleus RNA sequencing identifies new classes of renal proximal tubular epithelial cell in kidney fibrosis. American Society of Nephrology Kidney Week 2020, Oct. 2020, Online, Poster.

Contents

Chapter 1

Introduction

1.1 Renal proximal tubular cells (PTCs)

1.1.1 Overview of kidney

The kidneys are two bean-shaped retroperitoneal organs that locate between the transverse processes of T12-L3 vertebrae. A normal adult human kidney is about 10-15 cm in length and weighs 160 grams. The kidneys are the major organs to maintain homeostasis by regulating fluid balance and excreting metabolic wastes into the urine. They also have endocrine functions that produce hormones to maintain blood pressure, bone health and control red blood cell production.

The nephron, the kidney's basic structural and functional unit, is a highly complex structure comprising various components and cell types (Figure 1.1). A healthy adult kidney has approximately 1 million nephrons per kidney, while numbers for individual kidneys range from 200,000 to > 2.5 million (Bertram et al. 2011). The number of effective nephrons declines with age (Charlton et al. 2021). An average loss of about 6500 nephrons per year has been reported (Hoy et al. 2003; Denic et al. 2017). Nephron loss can be accelerated by many environmental and disease factors, including air pollution, preterm birth and hypertension (Charlton et al. 2021; Chen et al. 2021).

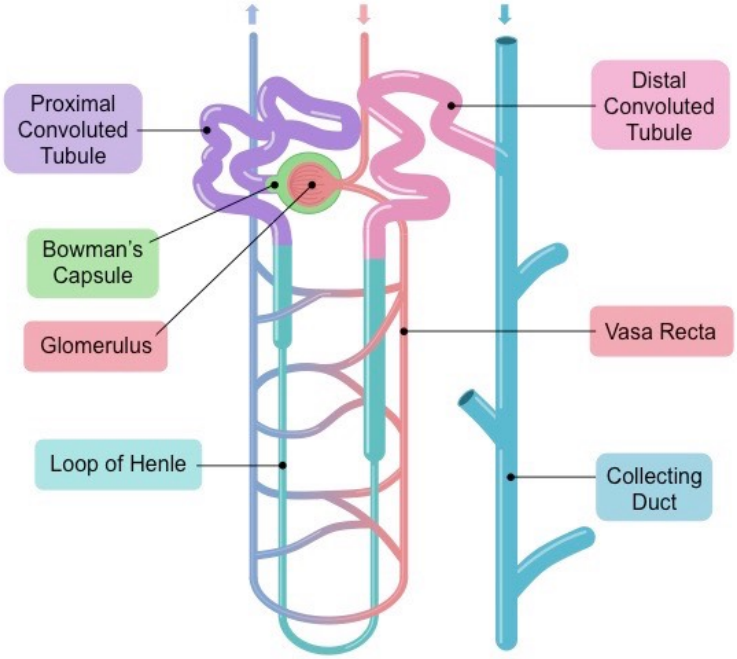


Figure 1.1 Structure of a nephron. Nephron highly complex structure comprising various components and cell types (https://ib.bioninja.com.au/).

The glomerulus and the renal tubule are the two major components of the nephron. The glomerulus is the blood filtration component that consists of fenestrated glomerular capillaries, mesangial cells, parietal cells, and podocytes. Blood from the renal vessels is filtered through the fenestrated capillaries, producing glomerular filtrate. The aggregate glomerular filtration rate (GFR) across the nephrons in both kidneys is commonly used as a main indicator for renal function. The tubular component connects the glomerulus to the renal pelvis. To achieve homeostasis, the glomerular filtrate is processed into urine via selective transcellular and paracellular reabsorption and secretion in the tubular components. Highly specialised, sequential urine processing occurs in discrete nephron segments, reflecting underlying cellular specialisation (Lee et al. 2015).

The tubular component comprises five anatomically and functionally unique parts of polarised monolayer epithelium that surround a central lumen: the proximal tubule (PT), the loop of Henle (LOH), the distal convoluted tubule (DCT), the connecting tubule (CNT) and the collecting duct.

1.1.2 Structure and physiology of PT

The PT is the beginning part of the tubular component, and connects the renal pole of the glomerulus to the hairpin-shaped LOH. The PT is divided into, first, the convoluted section (*pars convoluta*) that is located exclusively in the renal cortex and, second, the straight section (*pars recta)* that starts from the renal cortex and terminates at the outer stripe of the outer medulla (Zhuo and Li 2013). The PT is further divided into segments S1, S2 and S3, based on microscopic appearances. The S1 segment includes the beginning and middle portion of the convoluted section, and the S2 segment runs from the late convoluted section to the beginning of the straight section, followed by the S3 segment that includes the remaining portion of the straight section. Clerk *et al.* reported that the S1, S2 and S3 segments contribute approximately 20%, 20% and 5% of total renal cells, respectively (Clark et al. 2019).

The S1 segment has extensive apical microvilli, basolateral infoldings, cytoplasmic complexities, numerous long mitochondria, well-developed Golgi apparatus, and a prominent endocytic compartment that provides the highest capacity for solute reabsorption among all renal tubular segments (Zhuo and Li 2013). These

microscopic features are less evident in the S2 segment, which demonstrates a gradual transition from the S1 segments and additionally displays more numerous peroxisomes and larger secondary lysozymes. The S3 segment is more distinct, comprising simple cuboidal cells without the above features.

Functional heterogeneity of transport activities among S1-S3 segments has also been described. PT is the primary site of fluid, electrolyte, glucose and amino acid reabsorption, comprising active transport of ~150 litres per day of solute-rich fluid in a healthy human (Meinild et al. 2000). Segment-specific localisation of transporters/channels contributes to the solute transport and the maintenance of the local environment. Polarised localisation of transporters/channels on the membrane determines the physiological characteristics of reabsorption and secretion, e.g., glucose and amino acid reabsorption. Glucose is freely filtered from plasma into the glomerular filtrate and is typically completely reabsorbed in PT in healthy adults, resulting in glucose-free urine. Sodium/glucose cotransporter 2 (SGLT2), the low-affinity high-capacity glucose transporter, is responsible for glucose reabsorption on the apical membrane of S1 (mainly) and S2. In contrast, sodium/glucose cotransporter 1 (SGLT1), the high-affinity low-capacity glucose transporter, is responsible for glucose reabsorption on the apical membrane of S3 (Ghezzi et al. 2018). The intracellular glucose is then transported toward plasma through the basolateral glucose transporter 2 (GLUT2). Similarly, the b(0,+)-type amino acid transporter 1 (b(0,+)AT, (encoded by *SLC7A9*) on the S1 segment works together with the neutral and basic amino acid transport (rBAT,

encoded by *SLC3A1*) and the AGT1 (encoded by *SLC7A13*) in the S3 segment to achieve cysteine reabsorption (Broer and Palacin 2011; Nagamori et al. 2016). The unique pattern of segment-specific transcriptomes reflects the distribution of these transporters along the nephron, thus can be used for phenotypic identification of cells, for example in a single-cell experiment (https://hpcwebapps.cit.nih.gov/ESBL/Database/NephronRNAseq/) (Lee et al. 2015; Clark et al. 2019).

1.1.3 Non-tubular cells of kidney

Apart from the tubular cells, there are a lot of different cell type in the kidney. Glomeruli is the beginning part of a nephron that does selective filtration of blood into an ultrafiltrate (Zanetti 2020). The microscopic appearances and compound deposition of the glomerulus is used to define immune mediated disease of kidney. Renal vasculature comprises glomerulus and peritubular capillary network. Pathological changes of renal vasculature are associated with atheromatous disease and hypertension. The fibroblast is the major cell type of the interstitium. Fibroblasts can be stimulated by various cytokines, growth factors and hormones from nearby cells after kidney injury. TGF-β is the main mediator of renal fibrosis that triggering the Smad2 and EGF/EGFR pathway, promoting cellular differentiation and the subsequent myofibroblast transformation (Meran and Steadman 2011; Duffield 2014). The renal leucocytes will be recruited in kidney injury and modulate the inflammatory response and

fibrosis (Eleftheriadis et al. 2021; Linke et al. 2022). Injured PTCs secret inflammatory cytokines such as MCP-1 which recruit macrophages. TGF-β from recruited macrophages and PTCs can stimulate nearby fibroblasts to produce matrix components and undergo myofibroblast transformation (Yu et al. 2003; Gewin and Zent 2012). Cross-talk has been demonstrated for these cells with PTCs in kidney diseases and in renal fibrosis. For example, tubuloglomerular feedback influences the interaction between PTCs and glomerulus (Chen et al. 2020). Both PTCs and fibroblasts can secret TGF-β under pathological conditions and this cytokine acts on both types of cells (Tan et al. 2016).

1.1.4 Role of PTCs in acute kidney injury (AKI)

The PT has high energy requirements, reflecting the large amount of solute reabsorption occurring in this part of the nephron. The scale of proximal tubular resorptive activity makes the kidney a highly metabolically active organ and makes the PTCs vulnerable to injury (Berg JM 2002; Chevalier 2016). AKI, which refers to the prompt decline of renal function, is defined as any of the following: (1) an increase in serum creatinine by 0.3 mg/dl within 48 hours; or (2) an increase in SCr to x 1.5 times of baseline, which is known or presumed to have occurred within the last seven days; or (3) urine volume < 0.5 ml/kg/h for 6 hours (KDIGOAcuteKidneyInjuryWorkGroup 2012). PTCs are the primary target of AKI in many cases, especially in kidney injury induced by renal blood supply deprivation. Loss of proximal tubular epithelium decreases sodium and fluid reabsorption from PT, increases urinary loss of NaCl and fluid, then reduces GFR via tubuloglomerular feedback (Fattah and Vallon 2018). The regeneration of damaged PTCs with the restoration of normal solute reabsorption is critical for AKI recovery. The severity and frequency of proximal tubular injury are independent risk factors of renal outcome post AKI (Takaori et al. 2016).

With the heterogeneous causes of AKI, current treatments of AKI focus on managing the background diseases and renal supportive therapy. Current recommendations include removing renal toxic drugs or environmental toxins, treating underlying diseases, and maintaining renal perfusion by the fluid supplement. Effective therapy that specifically targets the common renal pathway

of AKI is still under laboratory investigation.

Inhibition of tubular solute transport has long been proposed as an approach to limiting injury or promoting recovery from AKI. Loop diuretics, for example furosemide, are commonly used in patients with AKI, principally for their effects on salt and water balance, but are proposed to have wider potential benefits including metabolic. Inhibition of PT solute transport is also proposed to have potential benefits, relating to the reservation of energy supplies for epithelial regeneration (Fattah and Vallon 2018; Hegde 2020). Pharmacological candidates are emerging for AKI prophylaxis, based on metabolic PT effects. SGLT2 inhibitors are a relatively new therapy for diabetes mellitus that act by inhibiting glucose reabsorption from early PT and increasing urinary glucose excretion. Long-term SGLT2 inhibitor use protects against CKD progression and, interestingly, may lower risk of AKI (Wanner et al. 2016; Neal et al. 2017). Genetic knock out of *Sglt1,* the aforementioned proximal tubular glucose transporter found in the S3 segment, was a benefit for AKI recovery in a mouse model of ischemia-reperfusion injury (IRI) by accelerating GFR recovery with a lesser rise in renal mRNA expression of injury markers, including kidney injury molecule-1 (Kim-1), chemokine (C-C motif) ligand 2 (CCL2), fibronectin-1 and collagen type I-α_1. However, genetic deletion of *Sglt2* failed to ameliorate AKI in the IRI mouse model (Nespoux et al. 2020). Pannexin1 (Panx1) is a ubiquitously expressed nonselective membrane transport channel in epithelia. Activation of Panx1 causes ATP efflux and contributes to inflammation (Poudel and Okusa 2019). Both PT and vascular

endothelial tissue-specific *Panx1* knockout mice were protected from AKI (Jankowski et al. 2018). Pharmacological trials regarding the protective effect of Panx1 inhibition on AKI are undergoing.

1.1.5 Role of PTCs in chronic kidney disease (CKD)

CKD is defined as abnormalities of kidney structure or function for more than three months with health implications, on the basis of clinical, laboratory and image diagnosis (Stevens et al. 2013). The prevalence of CKD increases with age, where CKD accounts for 18.2% of the population aged > 60 years in the UK, resulting in considerable expenditure in the NHS (Hirst et al. 2020). The outcome of patients with CKD correlates to the cause and the residual function.

Common causes of CKD in the adult include diabetes mellitus, glomerulonephritis, malignant hypertension, nephrolithiasis, malignancy, polycystic kidney disease, environmental toxins, and nephrotoxic drugs (Figure 1.2). Glomerular sclerosis and tubulointerstitial renal fibrosis are common pathological features of CKD across these diagnostic entities. In genome-wide association studies (GWASs), hundreds of loci associated with lower GFR and CKD have been identified, which provides one explanation for the various phenotypes of CKD with the same cause and CKD with an unknown underlying disease. PT shows the greatest enrichment for target quantitative trait locus (QTL) of CKD in GWAS (Qiu et al. 2018). Tubule-specific eQTL effect of CKD on the DAB2 gene was identified in this GWAS study. DAB2 is the endocytic adaptor protein of megalin and cubilin. Altered DAB2

function results in a filtration defect, TGFβ-induced profibrotic gene expression and CKD development (Qiu et al. 2018; Schutte-Nutgen et al. 2019; Long et al. 2022).

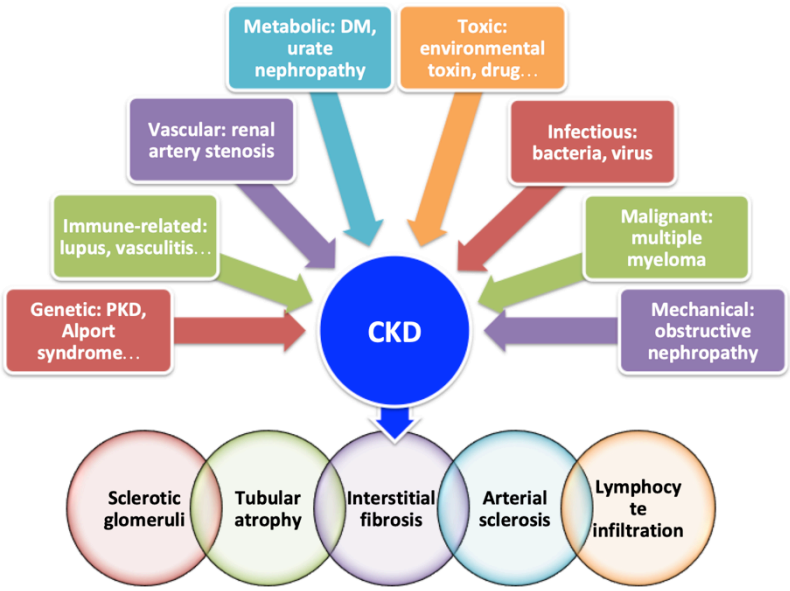


Figure 1.2 Common etiologies and pathological findings of CKD. The figure shows common causes of CKD. Across these diagnostic entities, kidneys with CKD have common pathological features: glomerular sclerosis, tubular atrophy, interstitial fibrosis, arterial sclerosis and lymphocyte infiltration.

Alterations in PTC gene expression and responses occur following both glomerular and tubular injury, and such changes may play key roles in CKD pathophysiology. Diabetes-induced proximal tubulopathy is an early disease event that could predict and contribute to CKD development in diabetes (Gilbert 2017). Similarly, in the unilateral ureter obstruction (UUO) mouse model of

progressive renal fibrosis, persistent activation of autophagy in PTCs resulted in atubular glomeruli and promoted fibrogenesis and CKD (Forbes et al. 2011; Livingston et al. 2016). PTC activation in the injured kidney leads to metabolic changes, inflammatory response and extracellular matrix synthesis. These changes may be required for tissue repair following injury but in progressive CKD, maladaptive repair processes may result in fibrosis progression (Schnaper 2017). Alterations of apoptosis, proliferation, differentiation, cellular adhesion, cellular transport, signal transduction, immune and metabolic pathways of PTCs are implicated in both profibrotic and protective mechanisms in proteinuric nephropathies (Rudnicki et al. 2007). The upregulated extracellular matrix (ECM)-receptor interaction pathway of PTCs during hypoxia drives structural change and interstitial collagen accumulation, which is a critical process of EMT (Yu et al. 2016).

1.1.6 Proximal tubulopathy as a determinant of AKI to CKD transition

AKI was considered an independent event of acute and reversible renal function decline. Conversely, CKD refers to a long-term structural or functional abnormality of the kidneys. However, recent evidence suggests that AKI and CKD are interconnected, where AKI from various etiologies can lead to residual CKD, and CKD is an independent risk factor for developing AKI (Figure 1.3) (Heung et al. 2016). AKI and CKD may share common mechanisms of recovery and fibrosis, regardless of aetiology (Heung et al. 2016).

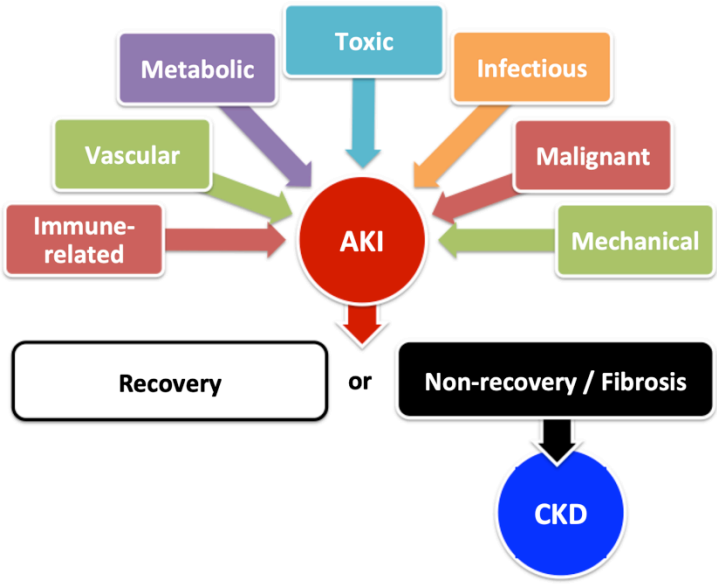


Figure 1.3 Etiologies and outcomes of AKI. There are various causes of AKI, including immune dysfunction, vascular disease, metabolic disease, toxin, infection, malignancy and mechanical obstruction. AKI was considered a reversible independent event of kidney injury, but current evidence shows that AKI is interconnected with CKD.

Maladaptive repair of proximal tubulopathy is considered a common mechanism in AKI and CKD and plays a critical role in AKI to CKD progression (Ferenbach and Bonventre 2015). In both AKI and CKD of diverse aetiology, responses of PTCs to injury were identified as key nodes in recovery versus progression (Chevalier 2016).

The potential pathophysiology underlying AKI to CKD transformation involves the TGF-β signalling pathway, the P53 mediated tubular cell injury and death, HIF activation, inflammation, mitochondrial dysfunction and oxidative stress (He et al. 2017). Higashi *et al.* disclosed the genetic profiling of the profibrotic versus the anti-fibrotic pathway of AKI and CKD, mainly focusing on PGE2 and TGF-β signal

transductions (Higashi et al. 2019). Persistent activation of EGFR in PTCs triggers profibrotic and pro-inflammatory factors via the ADAM17-AREG-EGFR pathway, resulting in progressive renal fibrosis after AKI (Kefaloyianni et al. 2016; Kefaloyianni et al. 2019). These evidence suggested that proximal tubulopathy plays an important role in AKI recovery, and PTCs are a potential therapeutic target for AKI.

1.2 Manipulation of PTCs in kidney injury

1.2.1 Heterogeneous response of PTCs to kidney injury

With the highly specialised nature of the structure, each nephron segment reacts uniquely to fibrotic stimuli. It is not surprising that the heterogeneity in response to injury also occurs in mature PT, which shows significant structural and functional differences along the different proximal tubular segments. Transporter-dependent cytotoxicity results in segment-specific damage of PTCs in various drug-induced kidney injuries and heavy metal injuries. For example, organic cation transporter 2 (OCT2) is mainly located on the basolateral membrane of the S2 and S3 of PT and is responsible for eliminating the anti-neoplastic drug cisplatin. Transportation of cisplatin into PTCs by OCT2 is the central mechanism of cisplatin-induced nephrotoxicity (Karbach et al. 2000). Organic anion transport 1 (OAT1) and 3 (OAT3) are responsible for elimination of multiple drugs and heavy metals, including ochratoxin A, aristolochic acid, mercury, methotrexate and tenofovir (Nigam et al. 2015). The S2 specific

expression of OAT1 and OAT3 determines the pathophysiology and phenotype of kidney injury induced by these agents (Hwang et al. 2010).

Variability in peritubular vascular perfusion will determine the outcome of ischemic kidney injury of each individual nephron. In this case, significant epithelial necrosis may be noted in some nephrons while adjacent ones are spared. The heterogeneous reaction of PTCs to injury also happens within individual nephrons, where some PTCs may be severely injured while others remain intact (Bonventre 2014). Cellular senescence and cell cycle status affect the response to fibrotic stimuli and cellular damage of each cell; thus, the cellular outcome can be recovery, maladaptation or death after injury (Yang et al. 2010b). The heterogeneous response of PTCs to injury makes it more complicated to investigate the responses of PTCs in progressive renal fibrosis. The average genetic profiles and signal transductions of PTCs may not comprehensively represent the phenotype and function of each PTC. Therefore, there is a need to investigate PTCs response to injury at the single-cell level.

1.2.2 Dedifferentiated PTCs in PT regeneration

1.2.2.1 The Source of PTC renewal following injury

The PT retains a strong capacity to repair following acute injury. Proliferation of a distinct subpopulation of PTCs, the dedifferentiated PTC phenotype, may be responsible for the regeneration process after kidney injury. The dedifferentiated PTCs enter the cell cycle, proliferate and re-differentiate to normal mature PTCs.

These cells contain less cytoplasm, fewer mitochondria, and absent brush border when compared to normal mature PTCs, and retain expression of progenitor markers CD24, CD133 and mesenchymal marker vimentin (Smeets et al. 2013). Occasional such dedifferentiated PTCs can be detected scattered in normal human kidneys, suggesting existing of baseline turnover of proximal tubular epithelium in normal circumstances. The number of dedifferentiated PTCs significantly increases in injured tubule or epithelial necrosis, possibly reflecting the undergoing PTCs regeneration.

The source of the dedifferentiated PTCs has been widely discussed. Evidence suggests a tubular origin for dedifferentiated PTCs, rather than from bone marrow progenitor cells in the lineage study (Lin et al. 2005). Progenitor cells have been identified in renal papilla with the ability to migrate to the upper papilla and transform into proliferating cells (Oliver et al. 2009). However, there is a long distance from the papilla to the renal cortex. In the rat model of segment-specific injury of PT, localised necrosis of PTCs followed by dedifferentiation of pre-existing regional PTCs for epithelial regeneration was noted (Fujigaki et al. 2006). Kusaba *et al.* proved that fully differentiated PTCs transiently expressed injury and regeneration markers, CD133, CD24, Ki67, KIM1 and Vimentin, after injury, with no evidence of intratubular progenitors during PT repair using genetic lineage tracing methods (Kusaba et al. 2014). Further investigation performed by the same group showed that the proliferation of the dedifferentiated PTCs is regulated by the EGFR-FOXM1 signalling pathway, which has been identified as a

promoter for EMT in various cancers (Chang-Panesso et al. 2019). Although these landmark studies suggested the residual mature tubular epithelium transforms to dedifferentiate PTCs after injury, it is still unclear whether every mature PTC has equal regeneration ability. To maintain the daily epithelial turnover in healthy kidneys, it is still unclear whether the dedifferentiated PTCs identified in healthy human kidneys derive from mature PTCs or migrate from papillary progenitor cells. It is conceivable that the two mechanisms may both exist, variably contributing to tubular regeneration.

1.2.2.2 Crosstalk between dedifferentiated PTCs and interstitial cells

The interstitium of the kidney lies outside of the nephrons and provides mechanical support for the tubular structures. Crosstalk between PTCs and adjacent interstitial cells also participates in governing PT responses during regeneration and fibrosis. The renal interstitium is composed of various cell types, including fibroblasts, pericytes and immune cells. Over 90% of the interstitial cells express PDGFR-β, including fibroblasts and pericytes. Following renal injury, PDGFR-β positive interstitial cells migrate toward the injured site before the onset of dedifferentiation of PTCs, and may support epithelial regeneration (Schiessl et al. 2018). Injured PTCs release the kidney injury response and profibrotic cytokine TGF-β, which activates Smad2, Smad3 and other pathways in interstitial fibroblasts, leading to extracellular matrix accumulation and fibroblast to myofibroblast differentiation (Castrop 2019). The Wnt/β-catenin interstitial-

epithelial signal transduction increases cell survival and cell cycle progression of PTCs, but persistent activation leads to endothelial - mesenchymal transition (EMT) and extracellular matrix accumulation (Tan et al. 2016).

1.2.2.3 Cell cycle arrest of injured PTCs

Accumulation of G2M phase PTCs in the recovered PT is a common phenomenon in various kidney injury models, e.g. mouse model of IRI, aristolochic acid nephropathy (AAN) and UUO (Yang et al. 2010b). The G2M arrest phenomenon is restricted to PTCs and not detected in the interstitial cells. The persistent presence of G2M arrest PTCs in the kidney after AKI is considered a mechanism of AKI to CKD transition (Ferenbach and Bonventre 2015).

DNA damage of PTCs activates the DNA damage checkpoint pathway, ATM – Chk2 – P53 – P21, which leads to phosphorylation and disassociation of the CDK1 / cyclin-B1 maturation complex, and results in cell cycle arrest (Jenkins et al. 2014; Romanov et al. 2015). The G2M arrest protects cells from regeneration without DNA repair at the beginning of AKI. However, persistent G2M arrest activates fibrotic signals, PFGF and TGFβ-1, and promotes fibrosis (Lombardi and Lasagni 2016). Inhibition of G2M arrest could ameliorate fibrosis, whereas stimulation of G2M arrest could increase fibrosis (Cianciolo Cosentino et al. 2013; Wu et al. 2013). This makes G2M arrest alteration a potential treatment for AKI.

1.2.3 Cellular senescence of PTCs after injury

The cellular senescent programme is an irreversible biological process of cessation of mitosis. Renal senescence was first described in normal ageing. Accumulation of senescent cells in the kidney was linked to age-related renal function decline (Sturmlechner et al. 2017). Senescence-associated secretory phenotype (SASP) refers to a distinct metabolically active cellular phenotype with cell cycle arrest and senescence-associated proteins secretion. Features of SASP include (1) the presence of senescence-associated β-galactosidase (SA-β-gal), which is a β-galactosidase isoform from overexpression or accumulation of the lysosomal β-galactosidase, reflecting the increase in lysosomal content in senescent cells (2) presence of cell-cycle arrest markers, including cyclin-dependent kinase inhibitors P21, P16, P27, P15 and cell cycle regulator P53 (3) absence of proliferation markers Ki67 and DNA replication licensing factor MCM2 (4) presence of DNA damage response markers such as H2AX and P53 (5) senescent cytokines secretion, including IL-1a, IL-1b, IL-6, IL-8, TGFβ1, WNT16B, plasminogen activator inhibitor 1 (PAI-1), CCL2 (also known as MCP-1), CXCL1, TNF-α, and CCN2 (Kurz et al. 2000; Sturmlechner et al. 2017; Docherty et al. 2019).

Maladaptive repair of PT after AKI shares common features with renal ageing, including the cell-cycle arrest of tubular epithelium, altered cytokine secretion, chronic *Kim-1* expression and chronic inflammatory cell infiltration (Ferenbach and Bonventre 2015). The G2M arrested PTCs act as the senescent cells that form

a cellular component named target of rapamycin - autophagy spatial coupling compartments and promote profibrotic factors secretion (Narita et al. 2011; Canaud et al. 2019).

Accumulation of senescent PTCs was described not only in the maladaptive recovery of AKI but also in CKD, autosomal dominant polycystic kidney disease and following renal transplantation, and may be an important contributor to progressive fibrosis (Docherty et al. 2019). Senescent cells were identified in patients with diabetic nephropathy, where hyperglycemia caused tubular senescence via SGLT2 and P21-dependent pathways (Kitada et al. 2014). The reactive oxygen species - NF-κB - P53 pathway is also involved in the cellular senescence of PT, leading to fibrotic gene expression and proliferation inhibition (Shimizu et al. 2011).

Persistence of PTCs with cell cycle arrest at G2M recruits immune cells, promotes fibrosis, and increases the susceptibility to further injury. Cytokines and growth factors released from senescent PTCs alter cellular response to stimulation through paracrine and autocrine mechanisms, and affects the fate of the kidney (Ferenbach and Bonventre 2015). Also, the senescent PTCs alter the microtubule and actin cytoskeletal structure, which may contribute to the failure of senescent kidney cells to organise centrosomes, proliferate and migrate (Moujaber et al. 2019). Cellular senescence of PT may thus contribute to AKI to CKD transition and progression of fibrosis.

1.2.4 Partial EMT of PTCs after injury

EMT is a biological process by which epithelial cells lose polarity and transform into a mesenchymal cell phenotype. Markers of EMT include loss of E-cadherin expression, and expression of N-cadherin and vimentin. Loss of intercellular connecting structures is an early hallmark of EMT, with repression of proteins that form tight junctions, adhesion junctions, desmosomes and gap junctions, eventually resulting in loss of epithelial apical-basal polarity (Lamouille et al. 2014). Actin in cells undergoing EMT is assembled into contractile stress fibers, increasing cellular motility. Cells undergoing EMT also express matrix metalloproteinases (MMPs) that can degrade extracellular matrix (ECM) proteins, promoting cell migration and invasion. EMT has been classified in three distinct biological conditions: EMT during implantation, embryogenesis, and organ development (type 1); EMT associated with tissue regeneration and organ fibrosis (type 2); and EMT associated with cancer progression and metastasis (type 3) (Kalluri and Weinberg 2009).

The TGF-β family proteins, compromising TGF-β 1-3 and bone morphogenetic proteins (BMPs), are inducers and regulars of EMT. The TGF-β family proteins also drive fibrosis in the kidney and other organs. Proteins that associated with EMT include SNAIL factors (SNAIL1 and SNAIL2), ZEB factors (ZEB1 and ZEB2) and basic helix–loop–helix factors (TWIST1 and TWIST2). These proteins down-regulate epithelial proteins and upregulate proteins associated with the mesenchymal phenotype. Other transcription factors of EMT have also been reported, including

the FOX families, HUMA2, GATA4 and GATA6 (Lamouille et al. 2014). Upregulation of other mesenchymal markers, including fibronectin, collagen, alpha-smooth muscle actin (α -SMA), and MMPs can be detected in cells post-EMT in addition to N-cadherin and vimentin.

Potential sources of myofibroblast in the fibrotic process include local fibroblast proliferation, pericyte, blood derived fibrocytes and the tubular epithelial cells. Pro-inflammatory and profibrotic cytokines secreted by injured PTCs or senescent cells may induce a partial EMT phenotype in adjacent renal tubular epithelial cells. The PTCs undergoing partial EMT express mesenchymal markers but do not convert into interstitial fibroblasts and remain inside the tubule (Lovisa et al. 2016). The PTCs with partial EMT regulates renal fibrosis progression (Zeisberg et al. 2007). Inhibition of EMT could restore repair and regeneration of kidney (Grande et al. 2015; Lovisa et al. 2015).

A recent landmark finding in nephrology is the beneficial effect of the drug class SGLT2 inhibitors in retarding rate of CKD progression. Work in mice suggests that some potential mechanisms may involve the EMT program of PTCs. Renal fibrosis could be suppressed by SGLT2 inhibitors, empagliflozin and canagliflozin, in the association of proximal tubular EMT inhibition in diabetic mice (Li et al. 2020b). The restoration of renal fibrosis was associated with inhibiting SGLT2 mediated aberrant glycolysis through *Sirt3* dependent pathway, where the pathological lesions were preserved in diabetic mice treated with insulin. Das *et al.* proved that Empagliflozin restored hyperglycemia-induced E-cadherin suppression and

suppressed fibronectin, α-SMA and vimentin expression (Das et al. 2020). In this study, the authors showed that empagliflozin inhibited hyperglycemia-induced oxidative stress, NF-κB activation, p38 MAPK activation, prevented miR-21-dependent RECK (Reversion Inducing Cysteine Rich Protein with Kazal Motifs) suppression and eventually suppressed EMT and migration of PTCs.

1.3 Aristolochic Acid Nephropathy

CKD has many aetiologies including direct tubular damage by toxins (Figure 1.2). One such class of toxin is the aristolochic acids, which are both causative of historical forms of CKD and used experimentally to model CKD *in vivo*.

1.3.1 History of aristolochic acid nephropathy

Aristolochic acids (AAs, represented mainly by AA I and II) are an active component of herbs of *Aristolochia* and some *Asarum* species. Herbal remedies with AAs have been used as anti-inflammatory drugs, diuretics and in weight loss regimens. AAs exhibit nephrotoxicity and are carcinogenic. AA was identified as a causative factor in epidemic nephropathy that was first reported between 1955 and 1957 (Ceovic et al. 1992). Increased frequency of renal failure and papillary transitional cell tumour of the urinary tract was reported in Balkan countries (Petronic et al. 1991). An outbreak of rapid progressive interstitial nephritis was also reported in more than 100 patients who used AA-containing Chinese herbs in 1992 (Mei et al. 2016). Long-term renal composite outcomes included renal

insufficiency, need for long-term dialysis and urinary tract cancer. AA also causes Fanconi syndrome, a proximal tubulopathy of polyuria, solute loss and hypokalemia (Lee et al. 2004). Clinical characteristics of AAN are mild proteinuria, anaemia and elevated serum creatinine. The use of AA-containing herbs has been banned in many countries, including the United Kingdom.

1.3.2 Pathological and molecular characteristics of AAN

Administration of AAs induces both acute and chronic kidney injury. In the acute stage, the kidney shows organelle swelling, tubular necrosis and lymphocyte infiltration under the microscope (Pozdzik et al. 2008). In the chronic condition, tubular atrophy, progressive interstitial fibrosis along the medullary rays up to the subscapular area, and persistent lymphocytes infiltration are the typical pathological findings. Glomeruli are preserved in both acute and chronic conditions.

AA has selective proximal tubular epithelial toxicity in the kidney (Jadot et al. 2017). The S3 segment is the preferential target of AA. The protein expression level of the proliferation markers, Ki-67 and PCNA, is increased in AAN kidneys. The injured PTCs secrete cytokines, including MCP-1, that activate neighbouring fibroblasts, and induce extracellular matrix synthesis and fibrosis. The expression level of the fibrosis-related genes, *Collagen I* and *III* and *TGF-β*, was increased after AA administration (Pozdzik et al. 2008). Cellular senescence was also identified in kidneys with AAN. Increased mRNA expression of senescent genes,

P53, P21 and P16, were reported together with positive SA-β-gal staining in AAN kidneys (Urate et al. 2021).

1.3.3 Mechanism of AA-induced nephropathy

AA is an albumin-binding compound in the bloodstream and displays anionic properties. AA is transported from the renal peritubular capillaries into the proximal tubular cells through the OAT family. The metabolites of AA bind to DNA, leading to AA-DNA adduct formation and DNA damage. The AA-DNA adducts are crucial for AA-induced carcinogenesis, cytotoxicity and apoptosis. AA-DNA adducts are a biomarker of AA exposure and AA related uroepithelial carcinoma. A:T to T:A transversions induced by the AA-DNA compound in *the P53* gene contribute to cancer formation (Chen et al. 2012).

The specific mechanism of AA-induced nephrotoxicity is yet not clear. AA induces oxidative stress by activating the MEK/ERK1/2 signalling pathways, followed by depletion of intracellular glutathione (GSH) and cell cycle arrest in the G2/M phase (Romanov et al. 2015). Subsequently, oxidative stress triggers DNA damage and activation of the MAPK pathway, resulting in direct apoptosis (Jadot et al. 2017). AA-induced fibrosis is linked to activation of TGF-β signalling. Activation of C3a/C3aR complement in PTCs was also reported in AA-induced fibrosis (Ye et al. 2019).

1.3.4 Mouse model of aristolochic nephropathy

AA-induced kidney injury presents a biphasic evolution in both experimental animal models and clinical patients (Jadot et al. 2017). Acute proximal tubular necrosis with interstitial inflammatory cell infiltration develops within the first three days of short-term AA injection in a rat model of AAN (Lebeau et al. 2005). Resolution and regeneration of PTCs starts a few days after AA injection. With consistent stimulation of AA, the kidneys display tubular atrophy and interstitial fibrosis, indicating CKD development. The AAN is a good model to investigate renal tubular damage in recovered AKI, AKI transition to CKD and chronic fibrosis in CKD.

Different regimens were used to induce AAN in the literature. AA is usually administrated orally or via intraperitoneal injection. Huang et al. reviewed and optimised the mouse model of AA-induced CKD (Huang et al. 2013). A high single dose of intraperitoneal AA injection > 10.0 mg/kg in C57BL/6 adult male mice caused severe AKI and death. Intraperitoneal injection of 2.5 mg/kg AA once weekly was suboptimal to induce chronic fibrosis. The intraperitoneal injection regimen of 3.0 mg/kg every three days for six weeks followed by six weeks of recovery / fibrotic stage was recommended. After 12 weeks, mice with AAN developed renal fibrosis with increased serum creatinine. Male mice had higher serum creatinine than female mice, while the baseline creatinine was similar between both sexes. More severe extent of fibrosis and collagen deposition were also noted in male mice, indicating that the males were more susceptible to AA-

induced kidney injury.

1.3.5 Other models of CKD and AKI to CKD transition

AAN is a simple and reproducible model for studying AKI to CKD transition and renal fibrosis with relevance to human disease, especially for investigating PT pathophysiology. The IRI model is also suitable for studying AKI and AKI to CKD transition of PTCs. However, the need for surgery makes it more complicated. In the UUO model without nephrectomy, which is another commonly used model for fibrosis, mice have normal creatinine without proteinuria (Yang et al. 2010a). Moreover, UUO is a rare cause of CKD in humans. Therefore, a mouse model of chronic AAN was selected as my primary experimental model of CKD.

There are two established mouse models of AAN at Cardiff University: the model of acute renal tubular injury through a single injection of AA, and the model of chronic renal fibrosis caused by repetitive injections of AA. They are stable and straightforward mouse models of kidney injury induced by intraperitoneal injection of AA. These models are modified versions of protocols of AAN in published studies (Yang et al. 2010b; Huang et al. 2013). Both models were set up and optimised by my co-supervisor, Dr Chia-Te Liao.

The chronic AAN model typically employs C57BL/6 wild-type mouse. In this model, intraperitoneal injection of AA (2.5 mg/kg) is performed twice weekly for two weeks, followed by two weeks of recovery/fibrosis (Figure 1.4). The experiment ends on the 28th day after the first AA injection. The model has been characterised

using C57BL/6 male mouse by testing the response to various AA protocols in detail (CT Liao, unpublished).

Figure 1.4 Mouse model of chronic AAN. In this model, mice receive an intraperitoneal injection of AA twice weekly for two weeks. Mice will be euthanised on the 28th day after the first AA injection.

1.4 Single-cell RNA Sequencing

1.4.1 RNA sequencing overview

Messenger RNA (mRNA) acts as a template for protein synthesis. RNA sequencing (RNA-Seq) is a tool for measuring mRNA expression profiles in biological samples. mRNA processing in eukaryotes starts from DNA transcription to synthesise pre-mRNA, which is 5' capped and 3' polyadenylated, followed by splicing to remove introns, then the product is finally transported out of the nucleus in the form of mature mRNA. The workflow of RNA-Seq begins with RNA extraction, followed by RNA purification and fragmentation, cDNA synthesis, cDNA amplification and adaptor ligation to make a cDNA sequencing library (Figure 1.5). After library quality control, the library is then sequenced using next-generation sequencing (NGS) to a read depth of 10–30 million reads per sample on a high-throughput platform (Stark et al. 2019). More than 90% of NGS relies on Illumina sequencing

technology that images a fluorescently labelled reversible terminator for each deoxynucleotide (dNTP). The sequencing result (presented as the fastq file format) is processed through high-throughput computing to align the sequencing reads to a transcriptome or a genome. Bioinformatics analysis of read counts of each gene from different samples are then performed to filter, normalise and compare the datasets to obtain the differential gene expression (DGE) profiles.

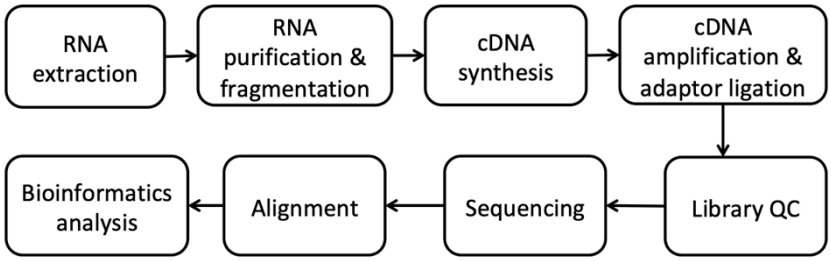


Figure 1.5 Workflow of RNA-Seq. The figure shows the general workflow of RNA-Seq. In the bioinformatics analysis, biological samples from different conditions are compared to show the difference.

1.4.2 Identify expression profiles at the single-cell level

The conventional technique of RNA-Seq, termed bulk-RNA sequencing (bulk RNA-Seq) is widely employed as a method for investigating gene expression profiling. Bulk RNA-seq captures the average signal of all cells within a sample with high sequencing depth. Signals from rare but unique cell types are not easily studied because of the nature of the bulk RNA-Seq, but can now be uncovered by single-cell RNA sequencing (scRNA-Seq). scRNA-Seq is an innovation that allows investigators to obtain RNA profiles from individual cells using NGS technology. ScRNA-Seq provides gene expression profiles from each cell rather than average signals from the whole sample. Table 1.1 shows the comparison of bulk RNA-seq and scRNA-seq.

Table 1.1 Difference between bulk RNA-seq and scRNA-seq. Comparison of bulk RNA-Seq and scRNA-Seq.

	Bulk RNA-Seq	**ScRNA-Seq**
Cost	Less cost	Expensive
Protocol	RNA extraction and purification, cDNA synthesis and amplification. No need for barcoding and unique molecular identifier.	Similar to bulk RNA-Seq but includes a barcoding step in which RNA is labelled with a unique molecular identifier.
Depth	10–30 million reads per sample.	Minimum 20,000 reads per cell on the 10X platform.
Number of genes acquired	More than 20,000 genes per sample in human and mouse.	About 2,000 genes per cell on the 10X platform.
Output	Average gene expression across all cells.	Expression profiles of an individual cell.
Analyses	- Gene expression - Differential gene expression between different conditions - Alternative splicing	- Dimensional reduction, doublet removal and cell clustering - Cell type identification and differential gene expression
Application	Compare RNA expression profiles between different conditions with high sequencing depth.	Build up cell atlas, identify new cell types/subtypes, lineage and differentiation study.
Visualisation of the main result	Heat map and volcano plot	UMAP plot or t-SNE plot

Single-cell sequencing helps to understand cellular differences, signal transduction and cellular responses, especially in an organ with complex structure or high heterogeneity such as the brain or kidney. Furthermore, current innovations in technique have the potential to acquire integrated data of DNA, RNA and proteomic expression from a single cell.

ScRNA-Seq is increasingly used for (1) building a cell atlas from a highly complex organ (e.g., brain or kidney), (2) identifying cellular lineages in organogenesis and developing tissue, (3) identifying new cell types, (4) identifying rare but important signals/cells that determine the fate of an organ or disease outcome, so far with a focus on cancer research and immunology.

1.4.3 Methodology of single-cell sequencing

Since the first single-cell sequencing method of mRNA was developed by Tang et al. in 2009 (Tang et al. 2009), more than 100 single-cell sequencing methods have been published. The cell barcoding system is the cornerstone of the single-cell sequencing technique. Cell-specific nucleotide barcodes are tagged to sequencing reads of each cell, allowing investigators to sort transcriptomic or genomic information from individual cells. cDNA amplification is a common step of all sequencing library preparation protocols after barcoding, owing to the small amount of starting material (Svensson et al. 2017).

Three different approaches have been employed to capture single-cells for sequencing: microwell plates, microfluidics chips and nanoliter droplets. In the

microwell system, cells and reagents are dispensed into a multi-well plate, where the cell barcodes are tagged to the cells, followed by amplification. In the microfluidics system, the cellular suspension is added to the integrated fluidic circuit, and the cells are shuttled into separated microfluidic reaction chambers, where the cell barcoding and amplification happen. In the nanoliter droplet system, cells are attached to gel beads then embedded into oil droplets with reaction enzymes. Cells lysis and barcoding take place in the oil droplets.

Platform selection for single-cell sequencing should be based on the experimental design and aim of the research. Wang et al. compared the scRNA-Seq results across microwell plates (Fluidigm C1 96 and HT), microfluidics chips (Clontech iCell8) and nanoliter droplets (10xGenomics Chromium) methods with four commonly used commercial platforms (Wang et al. 2019). Fluidigm C1 96 could achieve 6,000 genes per cell with up to 96 cells per run and the most expensive cost per good cell. Conversely, 10xGenomics Chromium provided 2000-3000 genes per cell with high throughput per run (5000 cells in the study but up to 30,000 cells in 2022) and low cost per good cell. Figure 1.6 shows a summary of the study that compares the four platforms.

Imaging capability	Yes	Yes	Yes	No
Throughput per run	Up to 96 cells	Up to 800 cells	~ 1000 cells	~ 5000 cells
Transcriptomic data	Full-length	3’	3’	3’ or 5’
Selection of individual cell	Yes	Per row	Yes	No
# genes per cell	~ 6100	~ 4000	~ 4200	~ 2000-3000
Cost per good cell	$83	$17	$18	$1.2

Figure 1.6 Comparison of 4 commonly used commercial scRNA-Seq platforms. Fluidigm C1 96 and HT use the microwell plates method, Clontech iCell8 uses the microfluidics chips technique, and 10xGenomics Chromium used the nanoliter droplets technique. The table is modified from the result table of the study performed by Wang et al. (Wang et al. 2019).

1.4.4 Workflow of the scRNA-Seq

The general workflow of scRNA-Seq starts from single-cell or single-nucleus suspension preparation, followed by sample loading and barcoding, reverse transcription and cDNA amplification, then sequencing and data analysis. Figure 1.7 describes the workflow of the scRNA-Seq using the 10x platform.

Biological samples are dissociated to make single-cell suspension or processed with a nuclear isolation protocol to make a single-nucleus suspension. Well-preserved cellular or nuclear RNA are critical for getting good data without bias or noise signals. When managing fresh samples, tissue harvest time to library preparation is a predominant determinant of the RNA quality. With frozen samples, the method of sample preservation may have a greater effect. Mechanical destruction of tissue should be avoided in this step if possible.

The barcoding system is the core step of scRNA-Seq. In the 10x system, cell suspension and barcode gel beads are loaded into a 10x chip. A cell attaches to a gel bead, and the gel bead is then encapsulated in an oil droplet with the enzyme. The enzyme lyses the cell/nuclear membrane to release the RNA, and barcodes attach to the poly-A tail of the RNA in the oil droplet. All RNA in the cell or nucleus shares the same 10x barcode, which is a 16 base pair nucleotide. Barcodes help to identify the cellular origin of the RNA in data analysis. After barcoding, the RNA is pulled together for reverse transcription, cDNA amplification, and sequencing. These steps are similar to bulk RNA-Seq.

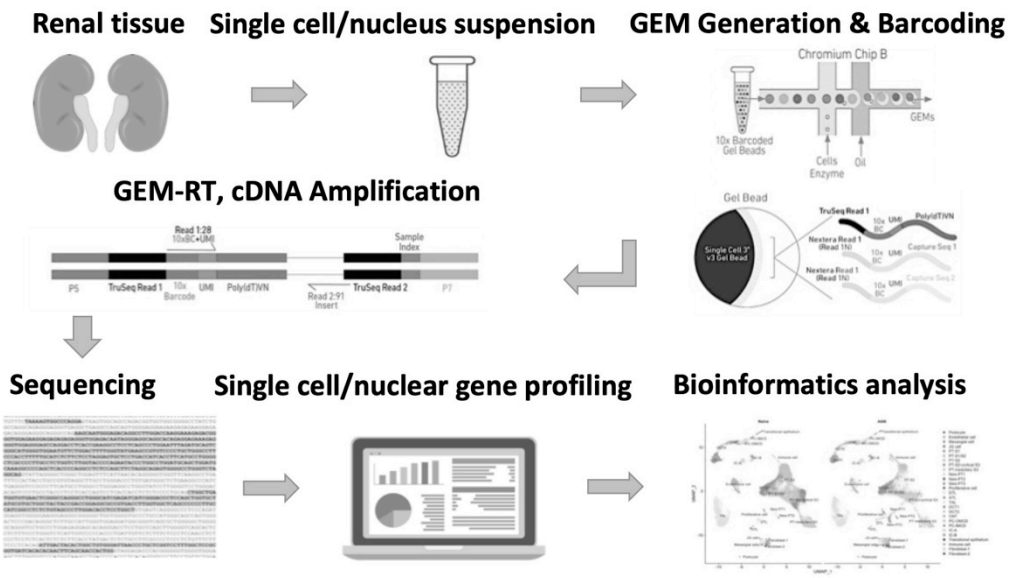


Figure 1.7 Workflow of the scRNA-Seq using the 10x platform. The workflow includes sample loading and barcoding, reverse transcription and cDNA amplification, sequencing, and data analysis. 10x Genomics Chromium uses the nanoliter droplet method for attaching a barcode to each RNA in the cell.

The best approach to bioinformatics analysis for scRNA-Seq results has not been determined. With increases in file sizes as more cells are included, and the maturation of single-cell multi-omics sequencing methods, new analytic tools keep developing. These tools are mostly R- or python-based packages. Figure 1.8 describes the basic approach of scRNA-Seq data analysis. The raw sequencing data is typically generated in the format of fast files. Genome or transcriptome mapping and gene counting is done on high-performance cluster computing, and generates a barcode-gene matrix. A barcode will be identified as a cell when it passes quality control (QC). Low-quality nuclei and dubious doublets are then removed. The remaining barcodes are then integrated for further processing and cell clustering. The cell type of each cluster is identified using canonical, or anchor, genes (e.g., *Nphs1* for podocyte). The downstream analyses include trajectory and pseudotime analysis, pathway analysis and ligand-receptor analysis.

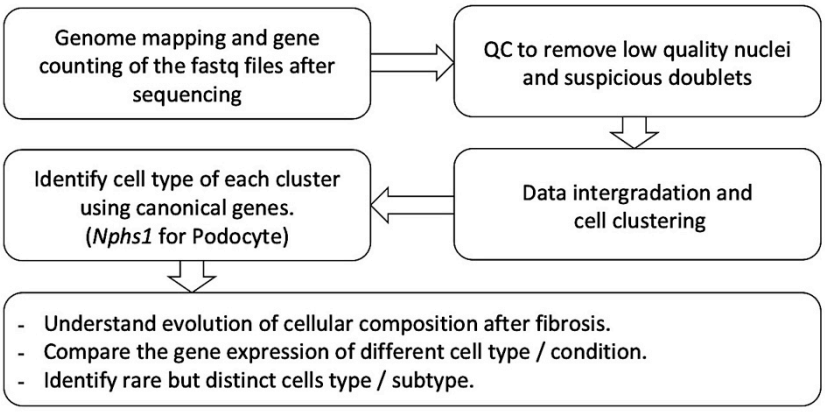


Figure 1.8 Workflow of data processing and analysis of scRNA-Seq. The figure summarises the basic workflow of scRNA-Seq data analysis. Method of mapping, QC and doublet removal determines the data present in all downstream analyses. QC = quality control.

1.4.5 ScRNA-Seq in kidney research

ScRNA-Seq has been used to recognise new cell types, identify rare but important events, disclose gene regulation in a particular disease, and establish a more precise biological model in various studies (Lindstrom et al. 2019). The kidney is a highly complex organ comprising more than ten cell types. Heterogeneity exists between different cell types and within the same type of cells. For example, the aforementioned heterogeneous response of PTCs to injury and the persistent presence of senescent PTCs after injury indicates the existence of unique PTC phenotypes. ScRNA-seq is, therefore, an appropriate tool to investigate PTCs heterogeneity.

1.4.5.1 ScRNA-Seq of healthy kidney

The first landmark study of scRNA-Seq was published in 2018. Park *et al*. performed the scRNA-Seq experiment using seven healthy male mice kidneys on the droplet-based single-cell RNA sequencing platform (Park et al. 2018). The study sequenced 57,979 cells and eventually characterised 43,745 cells after quality control. The study identified 16 distinct cell types. A previously unrecognised cell type of collecting duct, CD-Trans, which behaves as a transitional cell between principal cells and intercalated cells, was identified and validated.

The nature of the healthy human kidney at the single-cell level was firstly described in 2018. Aiming to disclose the cellular identity of renal tumor, normal

tissue biopsies from surgically resected kidneys were taken and sequenced (Young et al. 2018). The study identified the major cell types in normal kidneys, including epithelial cells, fibroblasts, myofibroblasts, vascular endothelial cells. Among 42,809 normal tissue biopsy cells, 37,951 mature kidney cells were epithelial cells, with a large proportion of PTCs.

1.4.5.2 Single-cell versus single-nucleus in kidney research

ScRNA-Seq requires live cells with intact intracellular RNA to build a high-quality library. Because the kidney is a solid organ with a well organised interstitial component which has high matrix content, enzymatic digestion by collagenase is commonly used when preparing a single-cell suspension. In the aforementioned scRNA-Seq studies of healthy kidneys, the resolution of PTCs was not high, evidenced by a failure to separate different segments of PT.

Wu et al. raised a recommendation to use single-nucleus rather than single-cell for investigating PTC heterogeneity at the single-cell level, especially when studying injured PTCs. Drawbacks of using scRNA-Seq while investigating injured PTCs include (1) the fragile injured PTCs may have a propensity to die or be destroyed during the dissociation process, leading to selection bias during sample preparation; (2) PTCs are sensitive to nutritional supply due to their mitochondria-rich nature. Stress-induced transcriptional artefacts of PTCs can be quickly introduced during kidney digestion; (3) scRNA-seq is incompatible with frozen archival material such as renal biopsy tissue, limiting its clinical application

(O'Sullivan et al. 2019; Wu et al. 2019).

The concern of using the nucleus is that the nucleus contains only 10% RNA of a cell, which gave rise to doubt about whether the RNAseq profile of a nucleus is representative of the cell's profile and active biological processes. However, Wu et al. found that single-nucleus RNA sequencing (snRNA-Seq) provided equivalent gene detection compared to scRNA-Seq, with reduced dissociation bias, on the droplet-based single-cell sequencing platform (Wu et al. 2019). Employing snRNA-Seq in the UUO mouse model of chronic renal fibrosis, two distinct minor categories of PTCs, the dedifferentiated PTCs and proliferative PTCs, were identified in chronic renal fibrosis. In contrast, ScRNA-Seq profiles were found to be enriched in leukocytes and glomerular cells (O'Sullivan et al. 2019). These data suggest that compared to scRNA-Seq, snRNA-Seq is a superior tool for studying PTC heterogeneity.

1.5 Aims

Focusing on PTCs heterogeneity, this project aimed to improve understanding of the character of PTCs in CKD and fibrosis. The overall aim was to study the features of PTCs in fibrosis and regeneration at the single-nucleus level.

Hypothesis: Unique PTC phenotypes occur after fibrotic stimuli, which are key to nephron recovery versus fibrosis following kidney injury and determine the fate of kidney in progressive CKD.

Four over-arching experimental aims were proposed.

1. *Optimise the protocol of nucleus isolation from mouse kidneys.*

   My project started in Jan. 2019, when single-cell sequencing techniques and applications were rapidly developing. The nuclear isolation protocol was optimised to get the best nuclear RNA yield and quality. In order to get the best library quality, fresh samples were tested and used in snRNA-Seq experiments. I also tested frozen preserved samples to understand the best tissue preservation method for a single-cell experiment.

2. *Carry out gene expression profiling of renal PTCs in normal kidney and progressive CKD using an established mouse model of AAN.*

   The experiments investigated the function of PTCs during progressive renal fibrosis by using snRNA-Seq. A mouse model of chronic renal fibrosis due to

AA was used. The aim was to characterise the PTCs in normal kidneys and to disclose the gene profiling and regulation of PTCs in CKD.

3. *Verify the new PTC phenotypes identified in the AAN snRNA-Seq experiment.*

   To validate the new PTC phenotypes identified in the AAN snRNA-Seq experiment, immunohistochemistry (IHC) and immunofluorescence (IF) staining for markers of discrete PTC phenotypes markers were performed.

4. *Investigate PTC phenotypes in healthy growing mouse kidneys, focusing on PTC proliferation and female-to- male differences.*

   Tubular elongation during kidney growth was studied, as a physiological scenario in which PTC proliferation was predicted to be high. The intent was to uncover PTC proliferation and differentiation pathways, and to relate these to similar process occurring during PTC renewal in recovery following kidney injury.

Chapter 2

Methods

2.1 Nuclear isolation and quality evaluation

2.1.1 Cell culture

2.1.1.1 HK-2 cell

HK-2 cell line (ATCC® CRL-2190™) is an adult human immortalised cortical proximal tubular epithelial cell line that was first isolated and cultured in 1984 (Detrisac et al. 1984). Cells from the primary culture were transduced with human papillomavirus 16 E6/E7 genes, resulting in an immortalised, well-differentiated proximal tubule cell line (Ryan et al. 1994). The HK-2 cell culture used DMEM:F12 medium (Sigma Dulbecco's Modified Eagle's Medium and Gibco™ Ham's F-12 Nutrient Mix (Gibco™ 11765054) 1:1) supplemented with 10% (v/v) foetal calf serum (FCS), L-Glutamine 2 mM, penicillin 100 U/mL, streptomycin 100 µg/mL, hydrocortisone 0.4 µg/mL, transferrin 5 µg/mL, and sodium selenite 5 ng/mL.

For *in vitro* growth of HK-2 cells, frozen HK-2 cells (passage 6) in a cryovial stored in liquid nitrogen were quickly thawed by gently swirling in a 37 ^{o}C water bath. The cryovial was decontaminated by spraying 70% (v/v) ethanol in a laminar flow hood. The defrosted cell suspension was transferred into a 15 mL conical tube containing 10 mL of culture medium. The cell suspension was centrifuged at 1500 rpm 20 ^{o}C for 5 minutes. The supernatant was removed, and the pellet was resuspended with 10 mL culture medium. The cell suspension was transferred into a 75 cm^2 tissue culture flask with a vented cap. The flask was incubated at 38 ^{o}C in a humidified incubator with 5% CO_2 in air atmosphere. Cells were monitored every day or every other day. The culture medium was changed every 2-3 days.

When the cells were over 90% confluence monolayer, cells would be passed into new culture flasks by trypsinisation. When passaging cells, the culture medium was removed from the flask. The cells were washed with 10 mL DPBS to rinse off any remaining medium. After removing PBS, 5 mL of trypsin (trypsin-EDTA 0.05% (w/w), 0.53 mM) mixed 1:1 with PBS) was added to the flask to detach the cells from the bottom of the flask. The trypsin containing flask was intubated at 38 ^{o}C for 3-5 minutes. Gentle shaking or tapping of the flask was done to help cells detach. A microscope was used to confirm that the cells were fully detached. Then 5-10 mL of culture medium (containing 10% (v/v) FCS) was added into the cell suspension to inactive trypsin and wash down cells through gentle pipetting. The cell suspension was transferred into a sterile 50 mL conical tube and centrifuged at 1500 rpm 20 ^{o}C for 5 minutes. The supernatant was removed, and the pellet was resuspended with the desired dilution of cells in a total of 20-50 mL of culture medium. Every 10 mL of the cell suspension was transferred into a new 75 cm^2 flask with a vent cap. The flask was incubated at 38 ^{o}C in a humidified incubator with 5% CO_2 in air atmosphere. Unused cells with passage 25 would be discarded without further passaging.

Nuclear isolation was performed using HK-2 cells with proper viability during cell growth and good morphology under the microscope from a 90-100% confluent flask.

2.1.1.2 GEnC cell

GEnC cell line was a conditionally immortalised glomerular endothelial cell line invented at the University of Bristol. GEnC cells were cultured in EGM-2 MV Microvascular Endothelial Cell Growth Medium-2 BulletKit (catalogue number 3202, containing CC-3156 and CC-4147). All components in the BulletKit except VEGF were added into the EBM-2 medium.

Dr Melissa Thomas kindly gave the GEnC cells. For *in vitro* growth of GEnC cells, frozen GEnC cells (passage 20) in a cryovial stored in liquid nitrogen were quickly thawed by gently swirling in a 37 ^{o}C water bath. The cryovial was decontaminated by spraying 70% (v/v) ethanol in a laminar flow hood. The defrosted cell suspension was transferred into a 15 mL conical tube containing 5 mL of culture medium. The cell suspension was centrifuged at 1500 rpm 20 ^{o}C for 5 minutes. The supernatant was removed, and the pellet was resuspended with 5 mL culture medium. The cell suspension was transferred into a 25 cm^2 tissue culture flask with a vented cap. The flask was incubated at 33 ^{o}C in a humidified incubator with 5% CO_2 in air atmosphere. Cells were monitored every day or every other day. The culture medium was changed every 2-3 days. Cells would be split to the next passage by trypsinisation when the cells were over 90% confluence.

When the cells were over 90% confluence monolayer, cells would be passed into new culture flasks by trypsinisation. When passaging cells, the culture medium was removed from the flask. The cells were washed with 5 mL DPBS to rinse off any remaining medium. After removing PBS, 2.5 mL of trypsin (trypsin-EDTA

0.05% (w/w), 0.53 mM) mixed 1:1 with PBS) was added to the flask to detach the cells from the bottom of the flask. The trypsin containing flask was intubated at 33 ^{o}C for 2-3 minutes. Gentle shaking or tapping of the flask was done to help cells detach. Microscope was used to confirm that the cells were fully detached. Then 5 mL of culture medium (containing 10% (v/v) FCS) was added to the cell suspension to inactive trypsin and wash down cells through gentle pipetting. The cell suspension was transferred into a sterile 50 mL conical tube and centrifuged at 1500 rpm 20 ^{o}C for 5 minutes. The supernatant was removed, and the pellet was resuspended with the desired dilution of cells in 10-15 mL of culture medium. Every 5 mL of the cell suspension was transferred into a new 25 cm^2 flask with a vent cap. The flask was incubated at 33 ^{o}C in a humidified incubator with 5% CO_2 in air atmosphere. Unused cells with passage 25 would be discarded without further passaging.

Cells for nuclear isolation would be grown to 80% confluence at 33 ^{o}C followed by thermo-switching to a 38 ^{o}C humidified incubator with a 5% CO_2 in air atmosphere. The thermo-switching process allowed GEnC cells to differentiate and mature in 5-7 days. Nuclear isolation was performed using GEnC cells with proper viability during cell growth and proper morphology under the microscope from a 90-100% confluent flask.

2.1.2 Nuclear isolation

2.1.2.1 Nuclear isolation protocol for in vitro cell line

This protocol was modified from the recommended nuclear isolation protocol of *in vitro* cell line from a prepared single-cell suspension for about $2.5x10^6$ cells (10x Genomics 2018).

Trypsin was used to detach and suspend cells from a 90-100% confluent flask. The culture medium was removed from the flask. The cells were washed with 5-10 mL DPBS to rinse off any remaining medium. After removing PBS, 2-5 mL of trypsin (trypsin-EDTA 0.05% (w/w), 0.53 mM) mixed 1:1 with PBS) was added to the flask to detach the cells from the bottom of the flask. The trypsin containing flask was intubated for 3-5 minutes. Gentle shaking or tapping of the flask was done to help cells detach. Microscope was used to confirm that the cells were fully detached. Then 5-10 mL of culture medium (containing 10% (v/v) FCS) was added into the cell suspension to inactive trypsin and wash down cells through gentle pipetting. The cell suspension was centrifuged at 300 g for 5 minutes at room temperature. The supernatant was removed, and the cell pellet was resuspended with 1 mL of NP40 lysis buffer, which contained NP40 0.1% (v/v), Triz-HCl 10 mM, NaCl 10 mM, and MgSO4 3 mM, and incubated on ice for 5 minutes. The sample was then centrifuged at 500 g 4 ^{o}C for 5 minutes. The supernatant was removed, and the nuclear pellet was resuspended by 1 mL of wash and resuspension buffer (1xDPBS supplemented with 1.0% bovine serum albumin (BSA) and 0.2 U/µL Protector RNase Inhibitor (Sigma 3335399001)). Lysis efficacy and cell viability would be

assessed by trypan blue nuclear stain (see section 2.1.4.1) when working with a new cell or tissue type in this step.

After confirming the sample had adequate lysis efficacy, the cell suspension was centrifuged again at 500 g for 5 minutes at 4 ^{o}C. The supernatant was removed, and the pellet was resuspended with 1 mL of wash and resuspension buffer. To wash and clean up the debris, the sample was centrifuged again at 500 g for 5 minutes at 4 ^{o}C. After removing the supernatant, the nuclear pellet was resuspended in 1 mL of wash and resuspension buffer and filtered through a 40 µm cell strainer. The nuclear yield was evaluated by hemocytometer and Muse cell counter. The nuclear RNA quality was assessed by Bioanalyzer (Agilent 2100).

2.1.2.2 Nuclear isolation protocol for mouse kidney

This protocol was modified from a published study on the *in vivo* renal nuclear isolation protocol of mouse kidney (Wu et al. 2019). A quarter of an adult mouse kidney was used for each experiment.

Lysis buffer was prepared freshly using Nuclei EZ Lysis buffer (Sigma NUC101) supplemented with protease inhibitor (Sigma 5892970001, 1 tablet per 10 mL of lysis buffer), RNasin Plus RNase inhibitors (Promega N2615) and SUPERaseIN RNase inhibitors (Life Technologies AM2696). For “lysis buffer-1”, a higher concentration of RNase inhibitors, 5 µl/mL RNasin Plus and 5 µl/mL SUPERaseIN, were used. For “lysis buffer-2”, a lower concentration of RNase inhibitors, 1 µl/mL RNasin Plus and 1 µl/mL SUPERaseIN, were used. Resuspension buffer was

prepared freshly using 1xDPBS, 1.0% BSA and 0.2 U/µL Protector RNase Inhibitor (Sigma 3335399001).

After schedule 1 termination, mouse kidneys were harvested after left ventricle perfusion with chilled PBS. The renal capsule was removed, and the kidney was preserved in PBS on ice. The harvested kidney was minced into < 2 mm pieces and transferred into a Dounce tissue grinder containing 2 mL of lysis buffer-1. The kidney was homogenised using a Dounce tissue grinder. The homogenisate was transferred into a 50 mL centrifuge tube, and then another 2 mL of lysis buffer-1 was added. The sample was incubated for 5-7 minutes on ice, filtered through a 40 µm cell strainer, and centrifuged at 500 g for 5 minutes at 4 ^{o}C. The supernatant was removed, and the pellet was resuspended using 4 mL of lysis buffer-2. Samples were incubated for another 5-7 minutes on ice and centrifuged at 500 g for 5 minutes at 4 ^{o}C. The supernatant was removed, and the pellet was resuspended with 4 mL of wash & resuspension buffer. The resuspended nuclei suspension was filtered through a 20 µm cell strainer. Nuclear abundance, lysis efficacy and concentration of the nuclei suspension were evaluated using the Muse cell counter.

The recommended concentration of nuclei was 700-1200 nuclei/µL. In most conditions, I adjusted the nuclei concentration to 1000 nuclei/µL. The nuclei suspension was used immediately for the 10x Genomics® Single Cell Protocol, to minimise the time between preparation of nuclei and chip loading.

2.1.3 Mouse kidney dissociation protocol

The number of kidney cells yielded from kidney dissociation was used to reference the total cell number of a mouse kidney. A comparison of cell and nuclear numbers yielded from kidney dissociation, and nuclei isolated from the contralateral mouse kidney were used to evaluate the lysis efficacy of the nuclear isolation protocol.

After schedule 1 termination, mouse kidneys were harvested after left ventricle perfusion with chilled PBS. The renal capsule was removed, and the kidney was preserved in PBS on ice. Pre-warmed DMEM-F12 medium (Gibco™, ThermoFisher Scientific, 11320033) was aliquoted into a 24-well flat-bottom cell culture plate (1 mL per mouse kidney per well) supplied with 0.2 mg/mL Liberase TL and 100 U/mL DNase I. Kidney was then cut into small pieces (~1 mm^3) and transferred into the culture plate carefully. The kidney was incubated at 37 ^{O}C for 30 minutes. After incubation, renal tissue fragments were transferred to a 70 µm mesh cell strainer top on a 50 mL centrifuge tube containing 1 mL FCS. The sample was gently mashed with a 5 mL syringe plunger and washed with cold DMEM-F12 up to 5 mL. The sample was then centrifuged at 4 ^{o}C, 500 g for 5 minutes. The supernatant was removed, and the pellet was resuspended with 10 mL cold PBS. Cell number was counted by Muse cell analyser. The cell counting result was recorded as the reference for the total cell number of a kidney.

2.1.4 Cell viability and lysis efficacy

2.1.4.1 Trypan blue stain for lysis efficacy evaluation

The Trypan blue dye exclusion test helps evaluate cell viability in a cell suspension. The test is based on the principle that a viable cell has an intact cell membrane that excludes trypan blue dye from the cell, whereas dead cell uptakes and retains trypan blue dye, therefore being stained with a blue colour (Strober 2015). Trypan blue dye exclusion test was used to evaluate lysis efficacy in the nuclear isolation process. Isolated nuclei would be stained with trypan blue dye due to the lack of a cell membrane. Conversely, un-lysed viable cells would show clear cytoplasm under a light microscope. The percentage of viable cells over total cells was counted by using a hemocytometer. If a high fraction of viable cells was presented, lysis time should be extended to achieve a better lysis efficacy. Samples with adequate lysis efficacy might reduce selection bias in nuclear isolation and prevent cytoplasmic RNA contamination in subsequent snRNA- Seq.

Trypan blue stain was used to evaluate the lysis efficacy of nuclear isolation from the cell line. Total cell count and the percentage of viable cells over total cells were calculated using a hemocytometer. The cell suspension was diluted to an appropriate concentration for cell counting by hemocytometer to perform a trypan blue dye exclusion test. After pipette mixing, 20 µL of cell suspension was taken into a new Eppendorf tube. Then 20 µL of trypan blue dye (0.4% (w/w)) was added to the cell suspension to achieve a 1 to 2 dilution. The cell/trypan blue dye

mixture was incubated for 2-3 minutes at room temperature. Intubation longer than 5 minutes might result in cell death.

The coverslip was placed on the hemocytometer over the chambers. 10 µL of the trypan blue-stained cell suspension was loaded into each chamber. The hemocytometer was then placed on the stage of a light microscope with 10X magnification. The stained nuclei and unstained intact cells were counted in each of the four outside squares of the hemocytometer using a hand tally counter, including cells on the bottom and left-hand perimeters but not on the top and right-hand perimeters. Since each major square represents a total volume of 0.1 mm^3, the cells' concentration equals the average cell count per square x dilution factor x 10^4.

2.1.4.2 Muse cell analyser

Muse Cell Analyzer uses miniaturised fluorescent detection and micro-capillary technology to analyse cells 2-60 µm in diameter.

For cell staining, 20 µL of well pipette-mixed cell or nucleus suspension was added into 1.5 mL Muse microcentrifuge tubes with screw caps containing 180 µL of Muse count & viability reagent (usually with a 10X dilution to achieve a concentration close to 100 particles/µL while 10-500 particles/µL was acceptable by the system). The sample was stained for 10 minutes. Complete system clean was performed before and after cell counting by using Guava Instrument Cleaning Fluid and distilled water. After vortex mix, the stained cell suspension was loaded

to the Muse cell analyser. Cell counting was performed using Count & Viability function with fine-tuning the settings for the VIABILITY vs CELL SIZE INDEX plot and the VIABILITY vs NUCLEATED CELLS plot.

Muse cell analyser provides information on total cell concentration, viable cell concentration and viability. Since nuclei accounted for most dead cells, the nucleus number was recorded by calculating the difference between the number of total cells and the number of viable cells. Muse cell analyser was used for counting and lysis efficacy evaluation for nucleus suspension from both cell line and mouse kidney. All single-cell experiments in my work used the Muse cell analyser nucleus counting result as the reference for chip loading.

2.1.4.3 Luna automated fluorescence cell counter

Luna automated fluorescence cell counter measures cell viability, counts cells and shows images from each channel (bright field, green and red fluorescence channel). For cell or nucleus counting, 18 µL of nuclei suspension and 2 µL of Acridine Orange were mixed. 10-12 µL mixture was loaded into the chamber of a PhotonSlide. The slide was inserted entirely into the slide port of the counter. Fluorescence Cell Counting Mode was selected, and cell size gating was adjusted by cell type. Focus and exposure were adjusted based on the image on the monitor. The Luna cell counter evaluated the lysis efficacy of nuclei suspension isolated from snap-frozen mouse kidney samples.

2.1.4.4 ImageStream

ImageStream (Amnis® ImageStream®X Mk II Imaging Flow Cytometer) is an imaging flow cytometry that allows multiparameter cell analysis. It acquires both fluoresce signals and high-resolution images from a large number of cells. ImageStream was used to directly visualise DPAI stained renal cell suspension acquired from kidney dissociation and nuclei isolated from mouse kidneys.

Cell and nucleus suspensions from a whole mouse kidney were centrifuged at 500 g 4 ^{o}C for 5 minutes. The supernatant was removed, and the pellet was resuspended in 100 µL PBS. The sample was stained with 1 µL of DAPI solution (10 µg/mL). The DAPI stained sample was then loaded to ImageStream.

2.1.5 RNA quality assessment

2.1.5.1 Nuclear RNA extraction

Nuclear RNA extraction and purification were carried out using the column method. MiRNeasy Mini Kit (Qiagen Cat No./ID: 217004) could purify total RNA from 18 nucleotides (nt) upwards for 50–100 mg of tissue. Since 1x10^{6} typical mammalian cells may contains 10–30 µg total RNA (https://www.qiagen.com/nl/resources/faq? id=06a192c2-e72d-42e8-9b40-3171e1eb4cb8&lang=en) and nuclear RNA accounts for 10-20% of total RNA (Piwnicka et al. 1983), 1-6 µg RNA would be expected from every 1x10^{6} nuclei.

For RNA extraction, nuclei suspension was centrifuged at 500 g for 5 minutes at 4°C, and the supernatant was removed. 700 µL of QIAzol lysis reagent was added

to the sample and homogenised using pipet mix. The homogenate was incubated at room temperature for 5 minutes. 140 µL of chloroform was added to the homogenate and shook vigorously for 15 seconds. The sample was incubated at room temperature for 2-3 minutes, then centrifuged for 15 minutes at 12,000 g at 4°C. After centrifuge, the upper aqueous phase was transferred to a new collection tube, avoiding transferring any interphase. 525 µL of 100% ethanol was added and mixed thoroughly by pipetting. Up to 700 µL sample was transferred into a RNeasy® Mini column in a 2 mL collection tube each time and centrifuged at ≥8000 x g for 15 seconds at room temperature. The flow-through was discarded. This step was repeated using the remainder of the sample. 700 µL of buffer RWT was pipetted onto the RNeasy Mini column and centrifuged at ≥ 8000 g for 15 seconds at room temperature. The flow-through was discarded. 500 µL of buffer RPE was pipetted onto the RNeasy Mini column and centrifuged at ≥ 8000 g for 15 seconds at room temperature. The flow-through was discarded. This step was repeated, and the RNeasy Mini column was centrifuged at ≥ 8000 g for 2 minutes. Then the RNeasy Mini column was transferred to a new 1.5 mL collection tube. 30 µL RNase-free water was pipetted onto the RNeasy Mini column membrane and centrifuged at ≥8000 g for 1 minute at room temperature to elute. The nuclear RNA product was stored in a -80 ^{o}C fridge.

2.1.5.2 Nanodrop analysis

The extracted nuclear RNA would be analysed using Nanodrop. Nanodrop provided results of RNA concentration and spectrophotometer measurement of absorbance of the extracted RNA. For Nanodrop analysis, 1 µL RNase-free water was loaded into the machine for blanking. Then 1 µL sample was loaded for measurement. The absorbance ratio at 260 nm, 280 nm and 230 nm was used to evaluate the purity of DNA and RNA. A 260/280 ratio of ~1.8 was generally accepted as pure DNA, and a ratio of ~2.0 was generally accepted as pure RNA. The 260/230 ratio will be expected in 2.0-2.2. A ratio beyond the normal range might indicate protein, EDTA, carbohydrates or phenol contamination.

2.1.5.3 Measurement of RNA integrity number (RIN)

Nuclear RNA quality was evaluated by nuclear RNA integrity number (RIN). RIN measurement was carried out using Bioanalyzer (Agilent 2100), which runs electrophoretic separation of RNA samples through a chip approach instead of traditional agarose gel. Bioanalyzer provides information on the integrity, quality, and concentration of the RNA. It can be utilised for RNA quality control by assessing RNA integrity. RIN is determined from an algorithm using visual inspection of the electropherogram generated by Bioanalyzer to identify the ribosomal peaks and calculate the ratio of 18S/28S ribosomal peaks. RIN is presented as a value of 1 to 10. A RIN of 10 is the highest and indicates intact RNA,

whereas 5 indicates partially degraded RNA and 3 indicates strongly degraded RNA.

RIN has been used in quality control for snRNA-seq of total RNA from frozen human prefrontal cortex tissue (Krishnaswami et al. 2016). Slane et al. proved that Bioanalyzer could provide similar electropherogram, gel picture and RIN for nuclear and total RNA in *Arabidopsis thaliana* (Slane et al. 2015). I, therefore, used the RIN of the extracted nuclear RNA as an indicator for sample quality when optimising the nuclear isolation protocol.

RIN of nuclear RNA was tested in the following conditions step by step to achieve a comprehensive evaluation of the nuclear isolation protocols.

A. To evaluate the feasibility of RIN measurement of nuclear RNA, the RIN of nuclear RNA extracted from HK-2 and GEnC cell lines were tested.
B. To evaluate the impact of RNase inhibitor on RNA quality, a comparison of RIN of nuclei extracted from HK-2 and GEnC cells with/without RNase inhibitor in wash & resuspension buffer was performed.
C. RIN of nuclear RNA of mouse kidney was measured. RNase inhibitors from different suppliers with various dosages were tested to acquire the best RIN of nuclear RNA in mouse kidneys.
D. To address the nuclear RNA degradation of *in vivo* experiment, RINs of cellular/nuclear RNA from each critical step of the mouse nuclear isolation protocol were measured.

E. Finally, two kinds of lysis buffer, NP40 and EZ lysis buffer, were compared directly in the nuclear isolation of mouse kidneys.

I performed the sample preparation and nuclear isolation in all cases, while Dr Amanda Redfern, the facility leads of Microarray and NGS of Central Biotechnology Services of Cardiff University, kindly helped with the Bioanalyzer and RIN measurement.

2.2 Animal model of chronic AAN

A clinically relevant mouse model of AAN induced chronic fibrosis was used to compare healthy and fibrotic kidneys. The experiment used C57BL/6 male mice aged 8-10 weeks bred by Charles River Laboratories. Male mice were used because they were more fibrogenic than female mice. Chronic AAN was induced by intraperitoneal injection of AA (Sigma A5512) twice weekly for two weeks (inject on D0, D4, D7, D11) with a dosage of 2.5 mg per kg body weight each time (Figure 2.1). The inflammation and tissue fibrosis developed after repetitive injuries, followed by tissue remodelling and fibrosis in the following two weeks after the last injection. Mice were housed with free access to chow and tap water on a 12-hour day/night cycle in a specific pathogen-free environment. Appearance, activity and bodyweight of the experimental mice were monitored daily during the experiment. Mice were terminated using Schedule 1 procedure (usually CO2 asphyxiation followed by cervical dislocation) on the defined time point (D28). After termination, blood sampling was obtained through cardiac

puncture. Chilled PBS (1x) was perfused via the left ventricle to the circulation system before kidney harvest. Both kidneys were harvested and processed. A quarter of the kidney was processed for snRNA-Seq. The rest of the renal tissue would be processed for histopathology. Hematoxylin & eosin stain and Masson's trichrome stain for the formalin-fixed paraffin-embedded sections were performed for each mice to confirm the fibrotic status.

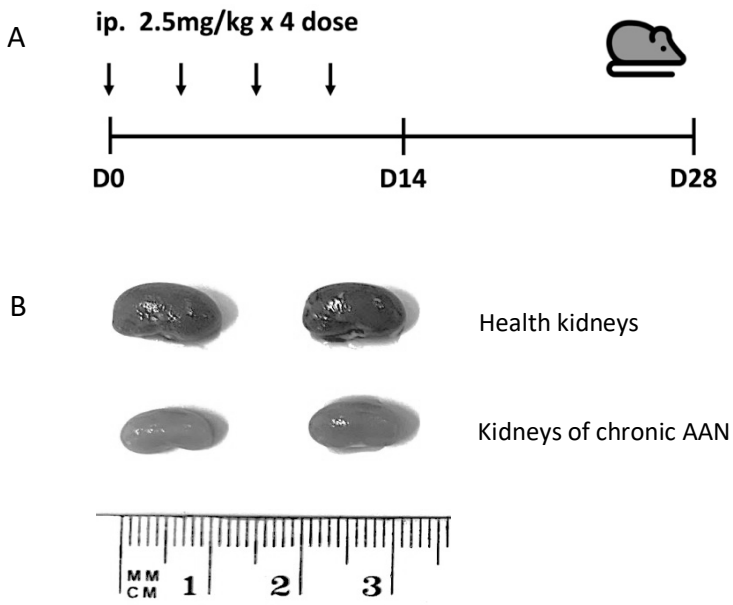


Figure 2.1 Mouse model of AAN. (A) Chronic AAN was induced by intraperitoneal injection of aristolochic acid (AA, Sigma A5512) twice weekly for two weeks. (B) Kidneys from chronic AAN mice were grossly different from healthy kidneys.

2.3 SnRNA-Seq

2.3.1 Gene expression library preparation

The 10x Genomics platform is a microfluidic droplet-based platform for single-cell and single-nucleus RNA sequencing, which enable up to 10,000 cells/nuclei per sample. Libraries preparation was performed using the 10x Chromium Single Cell Gene Expression Solution (single-cell 3' and reagent kits v3/v3.1, 10x Genomics). Nuclei suspension, gel beads and oil were loaded into the 10x chip carefully immediately after nuclear isolation. A gel bead contained primers with TruSeqRead (22 nt partial Illumina TruSeqRead 1 sequence), 10x barcode (16 nt), unique molecular identifier (UMI, 12 nt) and poly(dT)VN (30 nt poly(dT) sequence) (Figure 2.2A) . All primers in a gel bead had the same 10x barcode, and each primer had a unique UMI. UMIs tagged RNA molecules that helped to reduce errors and quantitative bias introduced by PCR amplification during sequencing. After sample loading, the chip was put into the 10x Chromium controller to create gel beads in emulsions (GEMs). In the 10x Chromium controller, the gel beads ran through the channel and captured the nuclei loaded using the microfluidic method (Figure 2.2B). The gel bead-nucleus complexes were then embedded in oil to form oil droplets. Lysis buffer within oil droplets would lyse the nuclei membrane. The gel beads captured poly-adenylated RNA molecules of the released RNAs that labelled RNAs from each nucleus with the same 10x barcode. Then the labelled RNAs were pulled together for cleanup, cDNA amplification, fragmentation, adaptor ligation, QC and quantification. cDNA quality was

evaluated by Fragment Analyzer (5200 Fragment Analyzer System, Agilent). The library preparation was kindly helped by Dr Rachel Raybould, MRC Centre for Neuropsychiatric Genetics and Genomics of Cardiff University and Dr Katherine Simpson, Wales Kidney Research Unit of Cardiff University. Figure 2.3 shows the structure of the final single-cell 3' gene expression library.

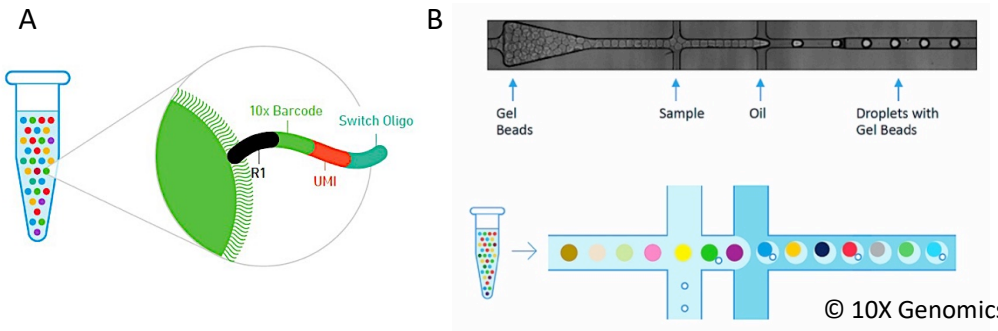


Figure 2.2 Sample loading and barcoding. (A) A gel bead contained primers with TruSeqRead, 10x barcode, unique molecular identifier and poly(dT)VN. (B) Gel beads ran through the channel and captured the nuclei loaded using the microfluidic method. The gel bead-nucleus complexes were then embedded in oil to form oil droplets.

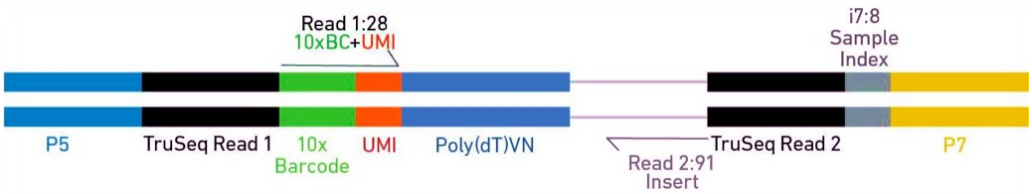


Figure 2.3 The structure of the final single-cell 3' gene expression library. The 10x barcode is a 16 bp nucleotide, and the UMI is a 12 bp nucleotide.

2.3.2 Next-generation RNA sequencing

A minimum of 20,000 reads per cell/nucleus was required for a convincing result of the 10x Chromium Single Cell Gene Expression Solution kits v3 and v3.1, 10x. The sequencing platform was chosen by calculating the multiple of targeted recovery nuclei number and appropriate sequencing depth. For the snRNA-Seq of mouse kidney with chronic AAN, sequencing was carried out using Illumina NextSeq 550 System, which enabled up to 400 million reads in one run, for 2 runs, targeting 3000-4000 nuclei recovered from each sample for 4 samples. For the snRNA-Seq of healthy growing mouse kidneys, sequencing was carried out using Illumina NovaSeq, which enabled up to 1.5 billion reads in one run, targeting 6000 nuclei recovered from each sample for 8 samples. The sequencing was performed by the Genome Research Hub of the School of Biosciences, Cardiff University and Wales Gene Park.

2.4 Bioinformatics analysis

2.4.1 Genome mapping and gene counting

Contemporary snRNA-Seq requires analysis of hundreds to thousands of transcripts originating from thousands of nuclei. Transcripts from nuclei were assigned a UMI, and data were tabulated in a matrix by UMI and barcode (each unique barcode identifying a nucleus of origin). The sequencing data were processed using the zUMIs pipeline (version 2.3.0) (Parekh et al. 2018). The pipeline discarded reads with low-quality barcodes and UMIs and then map reads

to the mouse reference assembly (Mus_musculus.GRCm38.95). zUMIs assigned the mapped reads to genes by an R package called RsubreadfeatureCounts and generated barcode-gene matrix utilising UMI counts of exon, intron, and exon+intron overlapping reads. zUMIs outputted the data matrix as a rds file for downstream analysis in R. Genome mapping and gene counting were performed on Supercomputing Wales Cardiff high performance computing cluster, Hawk.

2.4.2 Data integration, quality control and biological insight identification

The barcode-gene matrix generated by zUMIs was analysed using the R package, Seurat (version 3.1.3)(Butler et al. 2018; Stuart et al. 2019). The first step of quality control and data filtering is based on the data distribution, aiming to remove significant outliers and poor-quality cells. In the AAN experiment, cells for individual samples were retained if they contained ≧ 400 genes and genes identified in ≧ 3 nuclei. After merging the 4 naïve kidneys and 4 AAN kidneys with *merge* function, cells were filtered again to remove those with nuclei expressed ≦ 400 genes or ≧ 7500 genes or had mitochondrial gene expression ≧ 10%. In the growing mouse kidney experiment, cells for individual samples were retained if they contained ≧ 200 genes and genes identified in ≧ 3 nuclei. After merging samples with *merge* function, cells were filtered again to remove those with nuclei expressed ≦ 200 genes or ≧ 6000 genes or had mitochondrial gene expression ≧ 2.5%.

The feature counts were normalized by using *NormalizeData* function with LogNormalize (scale.factor = 10000). Highly variable genes were identified using the *FindVariableFeatures* function to identify the top 2000 variable genes. Gene expression levels of the top 2000 variable genes were scaled by *ScaleData* function. The principal component analysis (PCA) result of the scaled data was obtained by *the RunPCA* function. The number of principal components (PCs) included in the downstream analysis was determined by identifying the knee point of the elbowplot generated after running *the JackStraw* procedure. *FindNeighbors* and *FindClusters* function was applied based on previously identified PCs to cluster the nuclei. To visualise the dataset in low-dimensional space, the *RunTSNE* function was used to generate an a t-Distributed Stochastic Neighbor Embedding (t-SNE) plot, and the *RunUMAP* function was used to generate a Uniform Manifold Approximation and Projection (UMAP) plot. An R package, DoubletFinder (version 2.0.2), was used to predict and exclude dubious doublets in the dataset(McGinnis et al. 2019).

After doublet removal, the naïve and AAN datasets were integrated (dims = 1:50) using *FindIntegrationAnchors* and *IntegrateData* function. Then the integrated data was processed with *ScaleData, RunPCA, RunUMAP, FindNeighbors and FindClusters*, which have already been described above. Final clustering results were visualised using a UMAP. *DotPlot* function was used to show the expression level and percentage of genes expressed among different clusters.

Differentially expressed gene (DEG) analysis was performed using *the*

FindMarkers command (Wilcox method as the default) to identify canonical cell type marker genes of each cluster. Significance was defined as a gene with an adjusted p-value < 0.05, a ≧ 0.25 average log fold difference between the two groups of cells, and presence detected in at least 10% of cells in either of the two populations. P-value adjustment was performed using Bonferroni correction based on the total number of genes in the dataset.

The cell cycle analysis used *CellCycleScoring* to identify the cell in the G2/M and S status. Cells with G2M.Score > 0.15 and G2M.Score > S.Score were assigned as G2M status. Cells with S.Score > 0.15 and S.Score > G2M. Scores were assigned as G2M status. Cell with G2M Score <0.15 and S.Score < 0.15 were assigned as G1/G0 phase (Wu et al. 2019). An R package, EnhancedVolcano (version 1.2.0), was used to generate a volcano plot (Blighe 2018).

2.4.3 Trajectory, RNA velocity and pseudotime analysis

To predict the biological progression of PTC differentiation in normal and fibrotic status, trajectory, RNA velocity and pseudotime analysis were performed using all phenotypes of PTCs.

2.4.3.1 Trajectory analysis

Trajectory analysis could show how the gene expression profile changes continuously along an axis spanning the phenotypes. An R package, Monocle 3 (version 0.2.2.0), was used for trajectory and pseudotime analysis (Cao et al.

2019).

To run trajectory analysis in the snRNA-Seq dataset of mouse kidneys of chronic AAN, the metadata from the integrated Seurat object and the top 2000 variable genes from the integrated assay were loaded to the Monocle. The analysis involved all PTC clusters and the proliferative PTCs identified in proliferative cells using default parameters and removing clusters where low gene numbers were detected. The PTCs were re-clustered by Monocle, and then the trajectory analysis was performed using the *learn_graph* function.

2.4.3.2 RNA velocity analysis

RNA velocity analysis could predict the future state of each cell from the direction and rate of change of its gene expression profile within the following few hours using the ratio of spliced and unspliced RNA counts (La Manno et al. 2018). In combination with trajectory analysis, RNA velocity analysis provided insights into the speed and the direction of transcriptional dynamics. RNA velocity of each cell was calculated by velocyto.py and velocyto.R (version 0.6). The spliced and unspliced RNA counts provided by the 10X cellranger package were introduced and combined into the result of trajectory analysis. *SCTransform* function was used here as an alternative to the *NormalizeData*, *FindVariableFeatures, ScaleData* workflow for the spliced and unspliced RNA counts. *RunVelocity* function was used to calculate the RNA velocity. The result was shown by arrows with various sizes and directions on the UMAP generated by the trajectory

analysis.

2.4.3.3 Pseudotime analysis

Pseudotime analysis using *the Order_cells* function of Monocle was performed based on the RNA velocity result. The result of pseudotime analysis inferred the biological progression of PTC differentiation along the trajectory.

2.4.4 Pathway analysis

I conducted gene set enrichment analysis (GESA) to understand pathways of the new-PT clusters by using an R package, WebGestaltR (version 0.4.3)(Liao et al. 2019). Differentially expressed genes and average log fold change of the target clusters (new-PT1, new-PT2 and new-PT3) obtained from DEG analysis, comparing gene expression of the target PTC cluster to all other PTCs, were used for GESA. We evaluated the pathway enrichment in four major functional databases: KEGG, Panther, Reactome and WikiPathways. The recommended False Discovery Rate (FDR) cutoff of 0.25 was used (https://www.gsea-msigdb.org/gsea/index.jsp).

2.4.5 Ligand-Receptor analysis

The analysis was based on a well-reviewed dataset with 2557 ligand-receptor pairs from a published study (Ramilowski et al. 2015). PT S1-S3 were combined to normal PT and fibroblast-1, 2 were combined to fibroblast in ligand-receptor analysis. Ligands and receptors with an average fold-change ≥ 0.25 in the DGE analysis were paired. Pairing results of ligand genes of New-PT1, New-PT2, New-PT3 and receptor genes of fibroblast, immune cells and normal PT were shown using the cross products of ligand/receptor gene expression on a heatmap. I also showed individual ligand-receptor pairs with the ligands from the New-PT clusters and receptors from fibroblast, immune cell and normal PTCs using a circular visualisation tool, circlize (version 0.4.8), in R (Gu et al. 2014). A workflow of bioinformatics analysis is shown in Figure 2.4.

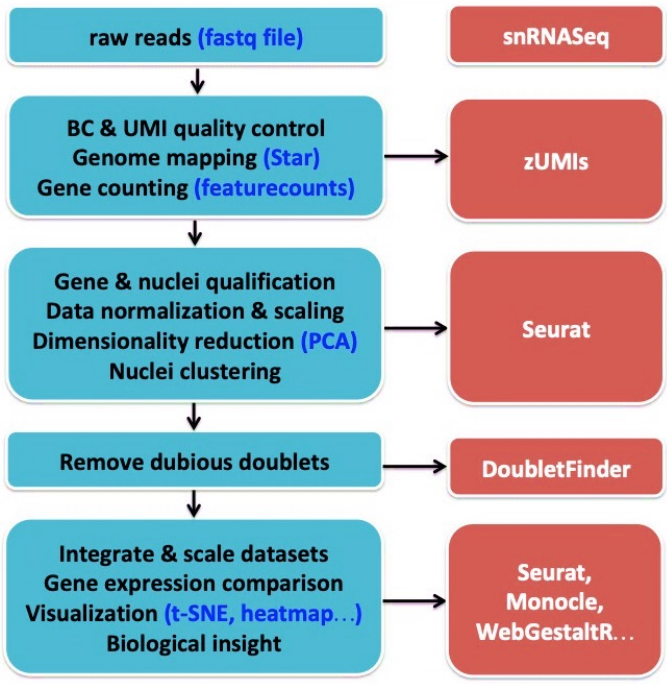


Figure 2.4 Workflow of snRNA-Seq bioinformatics analysis. The fastq file was processed by zUMIs for genome mapping and gene counting. The data qualification, normalisation, scaling, dimensional reduction, clustering and visualisation was processed by Seurat. The dubious doublets were identified using Doublet Finder. Final data integration, clustering, and gene expression comparison were processed by Seurat, Monocle, WebGestaltR and more.

2.4.6 Combine analysis with a published dataset

For validation, a combined analysis of snRNA-Seq results using my AAN induced CKD dataset and the murine ischemic reperfusion injury induced AKI dataset published by Kirita et al. was done (Kirita et al. 2020). The two datasets were merged, and the top 2000 variable genes from the combined dataset were obtained for PCA. Harmony (version 1.0), another batch integration method, was used to combine the datasets (Korsunsky et al. 2019; Tran et al. 2020). Harmony helped to combine datasets from different sequencing platforms. The clustering

result of all cells from both datasets was shown by UMAP. I analysed PTCs from both datasets using Seurat integrate data function for PTC analysis.

2.5 Microscopic validation

Microscopic validation was used to validate the newly identified cell type in an snRNA-Seq experiment. Formalin-fixed paraffin-embedded kidney sections from the AAN and the control group were used for immunohistochemistry (IHC) and immunofluorescence (IF) stain. All steps were performed at room temperature except for primary antibody staining.

For deparaffinisation, slide sections were intubated at 60 ^{o}C for 20-30 minutes. The sections were rehydrated by 3 changes of Xylene for 7 minutes each, 3 changes of 100% ethanol for 7 minutes each, 96% ethanol for 7 minutes, 70% ethanol for 7 minutes, 50% ethanol for 7 minutes, and 2 changes of deionised H_2O (dH2O) for 5 minutes. The dH_2O was replaced by citrate buffer (10 mM citric acid, 0.05% (v/v) Tween 20, pH 6.0). Antigen retrieval was performed for 20 minutes at 120 ^{o}C in the autoclave. The slide staining jar was cool down to room temperature after antigen retrieval, followed by 2 changes of PBS for 5 minutes each. Wax was drawn around the tissue with a wax pen.

2.5.1 IHC stain (except for goat primary antibodies)

For IHC stain, the rehydrated sections were blocked with 3% (v/v) hydrogen peroxide for 10 minutes, followed by 2 changes of PBS for 5 minutes each. If using

primary mouse antibody, mouse on mouse block (Vector, MKB-2213-1) was applied for 1 hour followed by 2 changes of PBS for 5 minutes each. UltraVision LP HRP Polymer enhancer system (Thermo Scientific, 12624007) was used to enhance the staining strength. Ultra V block was applied for 5 minutes, followed by 2 changes of PBS for 5 minutes each. Sections were stained with primary antibodies for Ki67 (Abcam, ab15580, 1:800 in 1% (v/v) goat serum) and alpha-smooth muscle actin (Invitrogen, MA5-11547, 1:800 in 1% (v/v) goat serum) at 4 ^{o}C overnight. 4 changes of PBS wash for 5 minutes each was performed in the following day. Sections were stained with primary antibody enhancer for 10 minutes, followed by 4 changes of PBS wash for 5 minutes each. Then HRP polymer was applied for 15 minutes in the dark, followed by 4 changes of PBS wash for 5 minutes each. DAB (Vector, SK-4100) stain was used for 5-10 minutes, followed by tab water wash. The counterstain used Hematoxylin (Vector, H-3404-100) for 2-5 minutes, followed by tab water wash.

The sections were dehydrated by 2 changes of dH2O for 30 seconds, 50% ethanol for 1 minute, 70% ethanol for 1 minute, 96% ethanol for 3 minutes, 3 changes of 100% ethanol for 2 minutes each, 3 changes of Xylene for 1 minute each. Cytoseal (Thermo scientific, 8310-4) was applied to the dehydrated section, and the tissue was covered by a coverslip.

2.5.2 IHC stain for goat primary antibodies

For IHC using primary antibody made from goat, sections were blocked with 3% (v/v) hydrogen peroxide for 10 minutes, followed by 2 changes of PBS for 5 minutes each. 10% (v/v) of donkey serum block was applied for 30 minutes, followed by 2 changes of PBS for 5 minutes each. The Avidin/Biotin block (Vector, SP-2001) was applied using Avidin block for 15 minutes followed by 2 changes of PBS for 5 minutes each and Biotin block for 15 minutes followed by 2 changes of PBS for 5 minutes each. If using primary mouse antibody, mouse on mouse block (Vector, MKB-2213-1) was applied for 1 hour followed by 2 changes of PBS for 5 minutes each. Sections were stained with goat anti-HAVCR1 primary antibody (R&D, AF3689, 1:800 in 1% (v/v) donkey serum) at 4 ^{o}C overnight. 4 changes of PBS wash for 5 minutes each were performed on the following day. Sections were stained with biotinylated donkey anti-goat antibody (Abcam, ab6884) for 30 minutes, followed by 4 changes of PBS wash for 5 minutes each. Then VECTASTAIN ABC-HRP enhancer kit (Vector, PK-4000) was applied for 30 minutes, followed by 4 changes of PBS wash for 5 minutes each. DAB (Vector, SK-4100) stain was used for 5-10 minutes, followed by tab water wash. The counterstain used Hematoxylin (Vector, H-3404-100) for 2-5 minutes, followed by tab water wash. The dehydration and slide sealing steps were the same as IHC stain with non-goat primary antibodies in section 2.5.1

2.5.3 IF stain

For IF stains, the rehydrated sections were incubated with the mouse-on-mouse block (Vector, MKB-2213-1) for 1 hour, followed by 2 changes of PBS for 5 minutes each if using the primary mouse antibody. 10% (v/v) of goat serum block was applied for 30 minutes, followed by 2 changes of PBS for 5 minutes each. Sections were stained with primary antibodies at 4 ^{o}C overnight. Primary antibodies included anti-SLC4A4 (Invitrogen, PA5-57344), anti-VCAM1(Invitrogen, MA5-11447), anti FODX1(LifeSpan BioSciences, LS-B9155-LSP), anti-AKAP12 (Abcam, ab49849), anti-WT1 (Sigma, MAB4234), anti-NCAM1 (Abcam, ab220360), anti-Tenascin C (Abcam, ab108930) anti-P21 (Novus bio, NBP2-29463) and anti-HAVCR1. Primary antibodies were diluted with 1% (v/v) goat serum with a dilution ratio of 1:200-1:800. 4 changes of PBS wash for 5 minutes each was performed on the following day. Sections were stained with secondary goat anti-rabbit Alexa Fluor 488 or 594 conjugated antibodies (Invitrogen, 1:500 in 1% (v/v) goat serum) for 1 hour in the dark, followed by 4 changes of PBS wash for 5 minutes each. Hoechst 33342 stain (10 µg/mL, diluted with PBS from 1mg/dL stock) was applied for 5 minutes in the dark, followed by 4 changes of PBS wash for 5 minutes each. Fluorescent mounting medium was applied to the section, and the tissue was covered by a coverslip. The IF-stained sections were preserved in the dark at 4 ^{o}C, and the images would be taken within 3 months.

2.5.4 Microscope

IHC stained tissue slides were visualised and digitised using an Olympus DP27 5MP colour camera attached to a Leica DMLA microscope. IF stained tissue slides were visualised and digitised using a confocal laser scanning microscope (LSM800, Carl Zeiss).

2.5.5 Quantitative image analysis

Immunostained tissue slides were visualised and digitised using an Olympus DP27 5MP colour camera attached to a Leica DMLA microscope or a confocal laser scanning microscope (LSM800, Carl Zeiss). Images were analysed with the ZEN2012 software (Zeiss), and quantification was performed with Qupath software(Bankhead et al. 2017). Pixel classifier was used to detect the collagen (cyan) stain, and thresholds were determined to express the positive stain as a percentage of total area (in >10 fields of view per animal, n=3 per group). Quantification of the alpha-smooth muscle actin DAB signal was used to confirm fibrosis in the AAN model. Nuclear Ki67 DAB signal was used to quantify proliferating cells as a percentage of all hematoxylin stained cells using QuPath's cell detection algorithm. Dr Irina Grigorieva, Wales Kidney Research Unit, Cardiff University, did the quantitative image analysis.

Chapter 3

Optimisation of Nuclear Isolation Protocol

3.1 Introduction

The single-cell sequencing technique was first published in 2009 and commercialised in 2018. The 10x platform emerged as the most commonly employed technique, as it provided comparatively high throughput (up to 10,000 cells per sample in 2019) with a lower cost per cell. The pipeline of approaching renal PTCs at the single-cell level had not been well-established when I started this project in early 2019. As mentioned in chapter 1, snRNA-Seq had apparent advantages over scRNA-Seq when investigating epithelial cells in the renal injury model (Wu et al. 2019). High quality RNA is crucial for recovering accurate biological information from the sequencing result. In snRNA-Seq, kidney samples are quickly processed, resulting in minimal inflammatory noise in PTCs, as scRNA-Seq has a longer kidney dissociation time that may induce inflammatory signals. Also, fresh kidney samples were used to achieve the best nuclear RNA quality as possible. Taken together, snRNA-Seq using fresh kidney samples on the 10x platform was selected as the research method.

First, nuclear isolation from human renal cell lines was tested to set up an optimal workflow. PTC was the most abundant cell type in the kidney. Therefore HK-2 cell line, which was a human proximal tubule cell line (ATCC® CRL-2190™), was used in the *in vitro* nuclear isolation development. A non-epithelial cell line, the GEnC cell line, a conditionally immortalised glomerular endothelial cell line developed by Professor Satchell at the University of Bristol, was also selected to evaluate the nuclear isolation efficacy in a different cell type. Two nuclear isolation protocols

were used for reference when establishing my protocol: the prepared single-nucleus suspension from cell line by 10x Genomics (10x Genomics 2018) and the *in vivo* renal nuclear isolation protocol of mouse kidney from a published study (Wu et al. 2019). Nuclear isolation protocols were tested and modified to maximise yield and RNA quality while minimising processing time and transcriptomic profile alterations.

After the nuclear isolation workflow was set up using cell lines, the *in vivo* nuclear isolation experiments were carried out. Nuclear isolation from fresh kidney samples of healthy adult mice was then tested to optimise the *in vivo* nuclear isolation protocol. Yielded nuclei number and nuclear RNA quality were carefully evaluated. The protocol would be reviewed and validated when working with a new species or mouse of different ages.

The workflow of nuclear isolation from frozen renal tissue was also tested for samples that need to be stored from weeks to months before processing. This helped to expand the application of snRNA-Seq to complicated experiments and samples that are difficult to collect, like human renal biopsy samples. Two tissue preservation methods, RNA later and snap-frozen, were evaluated by comparing the isolated nuclei yielded and the nuclear RNA quality.

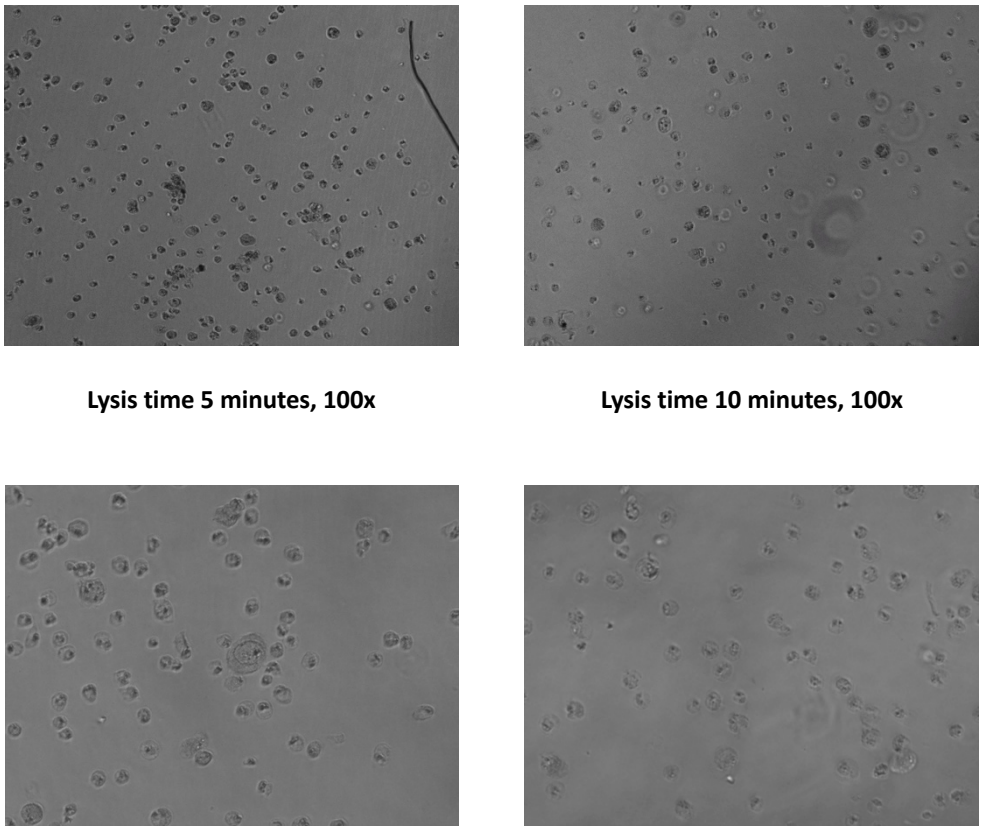
The experimental aim of the chapter was to optimise nuclear isolation protocol from various kidney cells.

3.2 Results

3.2.1 Nuclear isolation from cell line

3.2.1.1 Determine the optimal lysis time of HK-2 cells

HK-2 cells from a 90-100% confluent T75 flask were detached by trypsinisation. The HK-2 cell suspension was processed following the nuclear isolation protocol for prepared single-cell suspensions. Nuclear isolation experiments with a lysis time of 5 and 10 minutes were carried out to determine the optimal cell lysis time for HK-2 cells. The nuclei number yielded from the sample with a lysis time of 5 minutes was more than the sample with a lysis time of 10 minutes under the microscope (Figure 3.1). Samples from both conditions had good lysis efficacy, defined as less than 2% of viable unlysed cells. The result suggested that the optimal lysis time of HK-2 cells should be between 5-and 10 minutes.



Lysis time 5 minutes, 100x **Lysis time 10 minutes, 100x**

Lysis time 5 minutes, 200x **Lysis time 10 minutes, 200x**

Figure 3.1 Nuclear isolation with a lysis time of 5 and 10 minutes. Nuclei yield with a lysis time of 5 minutes was more than samples with a lysis time of 10 minutes. Both samples had a low proportion of unlysed cells, indicating adequate lysis efficacy.

3.2.1.2 Quantification of isolated nuclei from HK-2 cells

The nuclear isolation experiment was repeated for quantification using two fully confluent T75 flasks of HK-2 cells with a lysis time of 5 minutes. Cell counting was done using a hemocytometer before and after nuclear isolation and a Muse cell analyser after nuclear isolation. Trypan blue stain was performed before hemocytometer counting. By the hemocytometer, the cell counts before nuclear isolation were 1.19×10^6 and 2.21×10^6, and the count of the isolated nuclei were

$1.06x10^6$ and $7.82x10^5$, respectively. Samples showed good lysis efficacy under the microscope. The nuclei numbers counted by the Muse cell analyser were $1.12x10^6$ with a viability of 0.9% and $9.47x10^5$ with a viability of 1.2%. Results of cell counting from hemocytometer and Muse cell analyser were compatible. The results also supported that adequate lysis efficacy could be achieved with a lysis time of 5 minutes.

3.2.1.3 Nuclear isolation from HK-2 and GEnC cell lines

Nuclear isolation from HK-2 and GEnC cell lines was performed to evaluate the lysis efficacy across different renal cell types. Nuclear RNA quality was evaluated by measuring RNA integrity number (RIN). To confirm the effectiveness of RNase inhibitor for preserving nuclear RNA quality, a control experiment without using RNase inhibitor in wash & resuspension buffer was done.

Nuclei yield and RINs are shown in Table 3.1 and Figure 3.2. A difference was noted between the nuclei yield counted by hemocytometer and Muse cell analyser in sample 1. In sample 1, the hemocytometer yielded $1.58x10^6$ and $3.26x10^5$ before and after nuclear isolation, whereas the nuclei count was $3.26x10^6$ by the Muse cell analyser. Underestimation of the cell/nucleus number by hemocytometer could happen when the concentration of cell/nuclei suspension was high. Also, cell/nucleus number counted using hemocytometer sometimes varied when the cell counting was repeated, whereas repeating the cell/nuclei counting using Muse cell analyser usually had consistent results. In the

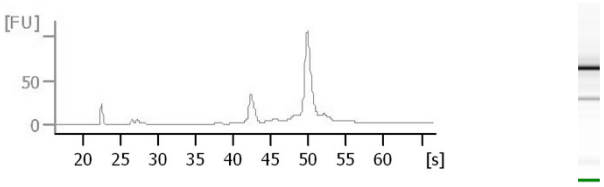
GEnC cells nuclear isolation experiments (sample 3 and sample 4), the isolated nuclei/cells ratio was about 50% with relative high viability (3.8%). Extension of lysis time should be considered when working with GEnC.

Nanodrop analysed the isolated nuclei to measure the yielded RNA amount. The total amounts of nuclear RNA in sample 1 and sample 3 were 9507 ng and 1818 ng, which correlated to the cell counting results of the Muse cell analyser. As snRNA-Seq required accurate cell count when loading samples to the chip, the Muse cell analyser was applied as the standard method of cells/nuclei counting in all single-cell experiments in this project. The ratios of 260/280 and 260/230 of nuclei samples were beyond the expected scope. This will be further discussed in the discussion section.

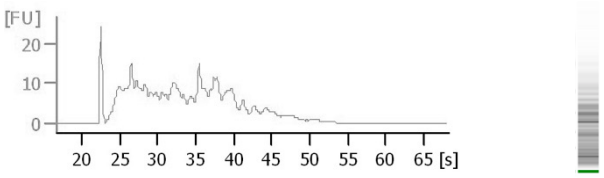
Using the standard nuclear isolation protocol, the RINs of the isolated HK-2 nuclei and the isolated GEnC nuclei were 9.2 and 9.0. For the samples without RNase inhibitor, the RINs were only 3.2 and 1.7, indicating that RNase inhibitor was essential for protecting nuclear RNA from degradation.

Table 3.1 Nuclear isolation from HK-2 and GEnC cell lines. Nuclear isolation works for both HK-2 and GEnC cell lines. RNase inhibitor is essential for protecting nuclear RNA from degradation.

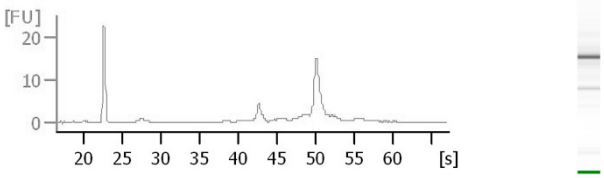
Sample No.	**1**	**2**	**3**	**4**
Cell type	HK-2	HK-2	GEnC	GEnC
Source	100% confluent T75	100% confluent T75	100% confluent T25	100% confluent T25
RNase inhibitor	Yes	No	Yes	No
Hemocytometer cell/nucleus count before and after nuclear isolation				
Initial cell number	1.58×10^6	2.63×10^6	1.06×10^6	4.99×10^5
Nuclei number	3.26×10^5	1.73×10^5	5.02×10^5	2.44×10^5
Muse cell/nucleus count after nuclear isolation				
Total cell	3.31×10^6		5.22×10^5	
Viability	1.4%		3.8%	
Isolated nuclei	3.26×10^6		5.02×10^5	
Nanodrop analysis				
RNA conc. (ng/ul)	316.9	228.1	60.6	63.4
Volume (ul)	30	30	30	30
Total nuclear RNA amount (ng)	9507	6843	1818	1902
260/280	2.04	2.17	1.69	1.85
260/230	1.77	1.67	0.63	0.56
Bioanalyzer analysis				
RIN	9.2	3.2	9.0	1.7



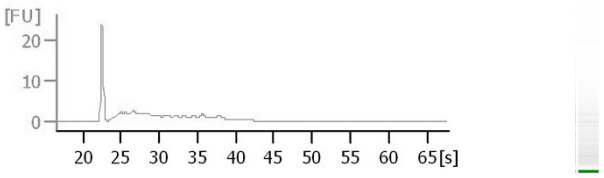
Electropherogram - sample 1 (HK-2, with RNase inhibitor)



Electropherogram - sample 2 (HK-2, without RNase inhibitor)



Electropherogram - sample 3 (GEnC, with RNase inhibitor)



Electropherogram - sample 4 (GEnC, without RNase inhibitor)

Figure 3.2 Electropherograms of the nuclear RNA from Bioanalyzer analysis. The RINs of the isolated nuclei from HK-2 and GEnC cells were 9.2 and 9.0. RINs were only 3.2 and 1.7 in the samples without RNase inhibitor. The result supported the efficacy of RNase inhibitor for preserving RNA quality.

3.2.2 Nuclear isolation from fresh mouse kidney

3.2.2.1 Lysis efficacy of mouse kidney nuclear isolation

The kidney dissociation procedure utilised Liberase to isolate renal cells and make a single-cell suspension. The number of cells acquired from a whole kidney was used as a reference for the total cell number of a mouse kidney. Lysis efficacy of the nuclear isolation protocol was evaluated by comparing cells/nuclei yield of the kidney dissociation and the nuclear isolation protocols using the contralateral kidneys from the same mice.

Kidney dissociation and nuclear isolation were done using the bilateral kidneys of an adult mouse. Cells/nuclei yielded from kidney dissociation, and nuclear isolation were 5.48×10^7 and 3.27×10^7, respectively (Table 3.2). The viability of the nuclei suspension was 1.0%. Renal cells/nuclei were directly evaluated under a light microscope and Imagestream (Figure 3.3). Renal cells had various sizes and morphology, while nuclei had a relative equivalent size and morphology. The result showed that the nuclear isolation protocol achieved adequate lysis efficacy.

Table 3.2 Cells/nuclei yield from kidney dissociation and nuclear isolation procedures. The lysis efficacy of the nuclear isolation protocol was good.

Process bilateral kidneys from an adult mouse		
Procedure	Kidney dissociation	Nuclear isolation
Muse cell/nucleus count		
Total cells	5.48×10^7	3.30×10^7
Viability	57.30%	1.00%
Viable cells / nuclei (nuclei show “dead” in Muse counter)	3.14×10^7	3.27×10^7

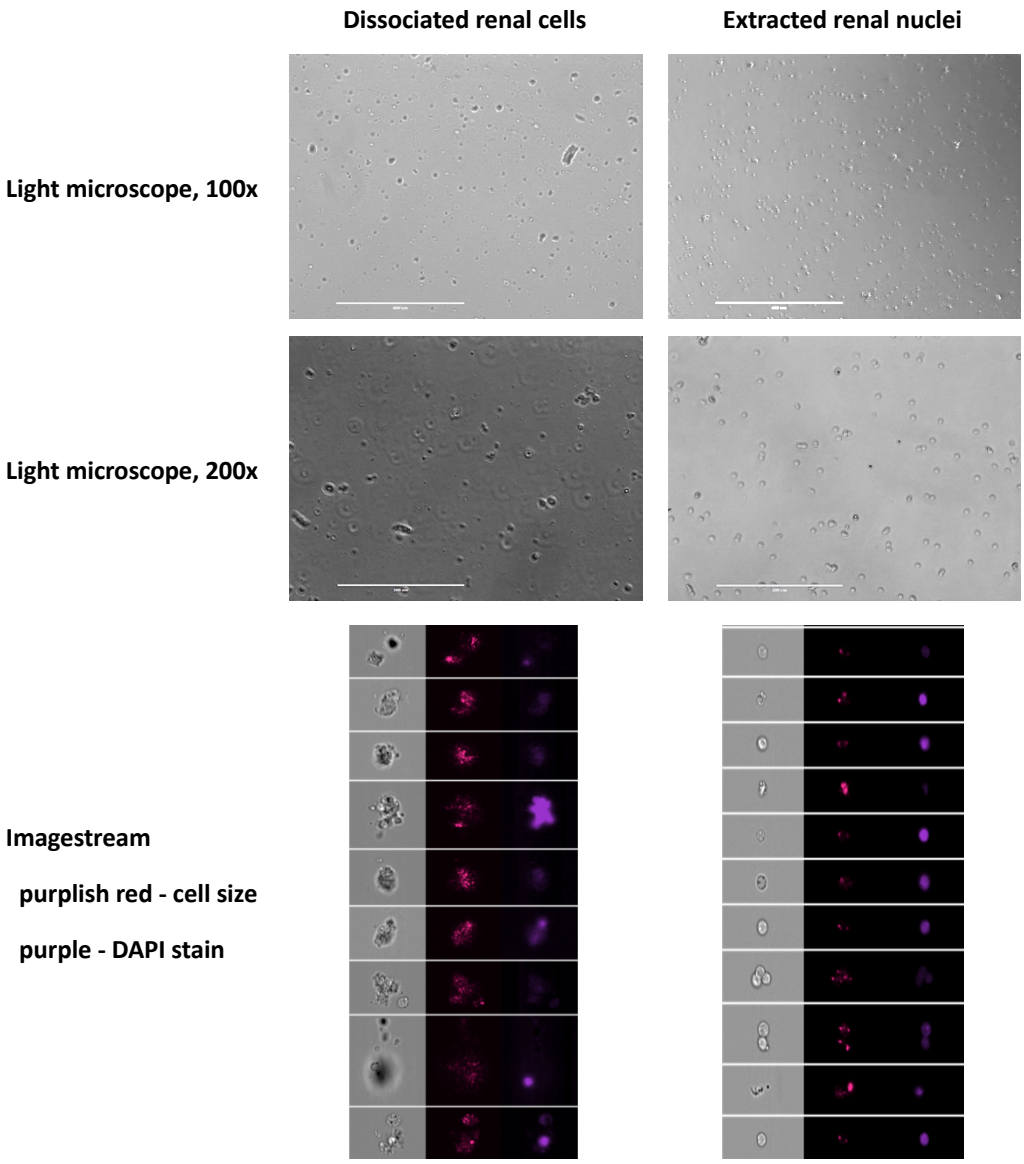
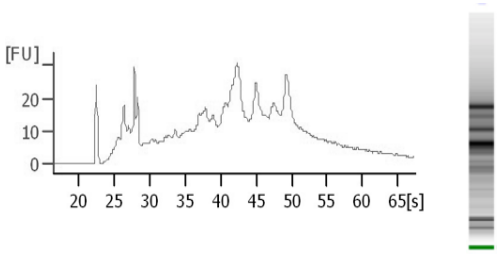


Figure 3.3 Dissociated renal cells (left) and extracted renal nuclei (right) under light microscope and Imagestream. Renal cells had various sizes and morphology under Imagestream. Nuclei had relative equivalent size and morphology. Multiplets was easily found in the single-nucleus suspension. (scale bar = 200 µm)

3.2.2.2 Isolated nuclear RNA quality of mouse kidney

Nuclear isolation from mouse kidneys was repeated, and the nuclear RNA quality was measured using Bioanalyzer. Nuclear isolation of an adult mouse kidney yielded 5.29×10^7 nuclei. The nuclei suspension was centrifuged at 500 g for 5 minutes at 4 °C. The pellet was lysed by QIAzol, and the nuclear RNA was extracted using the column method. The RNA extraction yielded 16.63 µg nuclear RNA with a 260/280 ratio of 2.10 and a 260/230 ratio of 2.10 on the Nanodrop. Bioanalyzer analysis of the nuclear RNA showed a RIN score of 4. Partially to severely degraded nuclear RNA was noted (Table 3.3).

Table 3.3 Cells/nuclei yield from kidney dissociation and nuclear isolation procedures. Partially to severely degraded nuclear RNA was noted.

Sample	**A mouse kidney**		
RNase inhibitor	Yes		
Nuclei count (Muse)	5.29×10^7		
Nanodrop analysis		**Bioanalyzer analysis**	
RNA conc. (ng/µL)	554.3	RIN score	4
Total volume (µL)	30	Electropherogram 	
Total nuclear RNA amount (ng)	16629		
260/280	2.10		
260/230	2.10		

3.2.2.3 Evaluation of the impact of various RNase inhibitors

RNA degradation of the isolated mouse renal nuclei was noted in the previous experiment. In the *in vitro* experiments using HK-2 and GEnC cell lines, the RINs from the extracted nuclear RNA of cell suspension were above 9. Also, RNase inhibitor was essential for preserving nuclear RNA from degradation.

In this experiment, various RNase inhibitor recipes were tested for evaluating the impact of different RNase inhibitors on the RIN of the nuclear RNA. Nuclei extraction was performed using adult mouse kidneys in the following conditions: (1) with/without proteinase inhibitor in the lysis buffer, (2) with RNase inhibitors from different companies, (3) with different concentrations of RNase inhibitors (Table 3.4). The RNase inhibitors recipes were modified from the protocol of prepared single-nucleus suspension for cell line from 10x Genomics (10x Genomics 2018), and the *in vivo* nuclear isolation protocol of mouse kidney from a published study (Wu et al. 2019).

Bioanalyzer analysis showed that the nuclear RNA was partially severely degraded based on lower RINs. The concentration of RNase inhibitors or using RNase inhibitors from different companies did not significantly improve nuclear RNA quality.

Table 3.4 Impact of proteinase inhibitor and RNase inhibitor on the nuclear RNA quality. Nuclear isolation was performed with/without proteinase inhibitor and with different concentrations of RNase inhibitor. RIN scores still showed partially to severely degraded nuclear RNA across all samples.

Sample No.	1	2	3	4	5	6
Proteinase Inhibitor	No	No	Yes	Yes	Yes	Yes
RNase inhibitor of lysis buffer-1 (each 4 mL)	No	20 µL Protector	20 µL RNasin Plus & 20 µL SUPERaseIN	20 µL RNasin Plus & 20 µL SUPERaseIN	20 µL Protector	20 µL Protector
RNase inhibitor of lysis buffer-2 (each 4 mL)	No	10 µL Protector	4 µL Rnasin Plus & 4 µL SUPERaseIN	4 µL Rnasin Plus & 4 µL SUPERaseIN	10 µL Protector	10 µL Protector
RNase inhibitor of resuspension buffer (each 4 mL)	10 µL Protector	10 µL Protector	2 µL Rnasin Plus	10 µL Rnasin Plus	20 µL Protector	No
RIN	4.0	3.7	3.1	3.3	3.1	2.6

3.2.2.4 Identification of the critical step of RNA degradation

Nuclei isolated from cell lines had high-quality nuclear RNA, but partial to severe nuclear RNA degradation was noted in the *in vivo* nuclear isolation experiments. After reviewing the two protocols, mouse kidneys had to be minced and homogenised by Dounce homogeniser before cell lysis since the kidney is a solid organ with a firm texture. Mechanical destruction of the nuclei or heat damage during grinding could damage the nuclear RNA. The protocol was adjusted to minimise mechanical destruction and heat damage, that the times of loose and tight homogenisation steps were reduced from 10-15 to 2-3 times.

The critical step of nuclear RNA degradation was investigated by measuring the RINs of the samples from different steps of the nuclear isolation protocol. Six samples from the following steps were collected:

1. Cut the kidney into < 2 mm pieces with a scalpel. => sample 1
2. Loose homogenisation by a Dounce homogeniser. => sample 2
3. Tight homogenisation by a Dounce homogeniser. => sample 3
4. Incubate the homogenate on ice for 5 minutes, then filter through a 40 µm cell strainer. => sample 4
5. Centrifuge at 500 g, 4 ^{o}C for 5 minutes. Remove the supernatant, resuspend the pellet with lysis buffer and incubate on ice for 5 minutes. => sample 5
6. Centrifuge at 500 g, 4 ^{o}C for 5 minutes. Remove the supernatant, resuspend the pellet with nuclei suspension buffer and filter through a 20

µm cell strainer. => sample 6

Bioanalyzer results are shown in Figure 3.4. The RIN of sample 1 was 9.1, and the electropherogram showed clear peaks at 18s and 28s, indicating the nuclear RNA was intact after kidney harvest and when the kidney was minced into small pieces. Partial RNA degradation was noticed at homogenisation steps after reducing the loose and tight homogenisation to 2-3 times (sample 2 and 3). Samples 2-6 had similar electropherograms, and their RINs were all-around 7. The result showed that homogenisation was the primary cause of nuclear RNA degradation. The nuclear RNA quality was significantly improved after reducing the number of times of homogenisation. Since homogenisation was an essential step of kidney nuclei extraction, a RIN of 6-7 was considered acceptable.

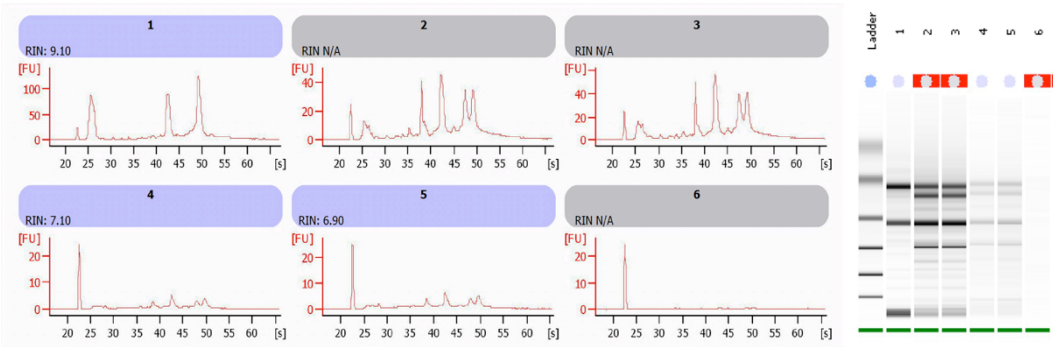


Figure 3.4 Bioanalyzer analysis of the nuclear RNA from different steps of the nuclear isolation protocol of mouse kidney. The RIN of sample 1 was 9.1. Samples 2-6 had RINs around 7 and similar electropherograms. This showed that partial RNA degradation happened mainly in the tissue homogenisation steps.

3.2.2.5 Lysis buffer and lysis time of nuclear isolation from mouse kidney

The 10x snRNA-Seq platform allows 500-10000 cells per run for each sample. Nuclear isolation from an adult male mouse kidney usually yielded 30-50 million nuclei, far exceeding the demand. A large tissue in a tissue grinder might increase mechanical damage during homogenisation. Therefore, the final protocol used a quarter of adult mouse kidneys in each nuclear isolation experiment.

To evaluate the appropriate lysis time of the protocol, nuclear isolation from a quarter of mouse kidney was repeated using the modified protocol with loose and tight homogenisation 2-3 times for 2 samples. Nuclei yields of the two samples were 9.05×10^6 and 9.81×10^6, and both samples had acceptable nuclear RNA quality (Table 3.5 and Figure 3.5, sample 1 and 2). Nevertheless, a high percentage of the unlysed cells, 1.80% and 5.50%, were noted. Therefore, the lysis time was extended from 5 minutes to 7 minutes.

The impact of the lysis buffer was also tested using two different lysis buffers, NP40 and EZ lysis buffer, in mouse kidney nuclear isolation (Table 3.5, sample 3 and 4). Compared to the sample managed using NP40, the sample with EZ lysis buffer yielded more nuclei and a higher RIN (Figure 3.5). A higher risk of selection bias was considered when using NP40 because of the weaker power of nuclei extraction. With a lysis time of 7 minutes, a low percentage of viable cells was noted in both samples, indicating a better lysis efficacy.

Taking all the nuclear isolation experiments of mouse kidneys together, using EZ lysis buffer would yield 10 million nuclei/cells from a quarter of a mouse kidney.

This confirmed that the result of nuclear isolation was reproducible and reliable. With acceptable RNA quality and adequate lysis efficacy, this protocol was used as the final protocol for nuclear isolation of mouse kidneys.

Table 3.5 Appropriate lysis time and lysis buffer assessment. The nuclear isolation experiment was repeated using the modified protocol with a lysis time of 5 minutes (sample 1 and 2). High viability of nuclei suspension was noted in both samples with acceptable nuclear RNA quality. Sample 3 and 4 used an extended lysis time of 7 minutes with two different lysis buffers, NP40 and EZ lysis buffer. Sample 4 had higher nuclei yielded and a higher RIN.

Sample No.	**1**	**2**	**3**	**4**
Source	¼ from the same mouse kidney		¼ from the same mouse kidney	
Lysis buffer	EZ lysis	EZ lysis	NP40	EZ lysis
Lysis time	5 min + 5 min	5 min + 5 min	7 min + 7 min	7 min + 7 min
Muse cell analyser				
Total cells	9.05×10^6	9.81×10^6	5.19×10^6	1.09×10^7
Viability	1.80%	5.50%	0.30%	0.20%
Isolated nuclei	8.89×10^6	9.27×10^6	5.16×10^6	1.09×10^7
Nanodrop analysis				
RNA conc. (ng/µL)	142.2	122.9	73.1	159.3
Total volume (µL)	30	30	30	30
Total nuclear RNA amount (µg)	4.27	3.69	2.20	4.78
260/280	2.06	2.06	1.84	2.08
260/230	1.34	1.73	0.41	0.75
Bioanalyzer analysis				
RIN	6.5	6.1	6.9	7.5

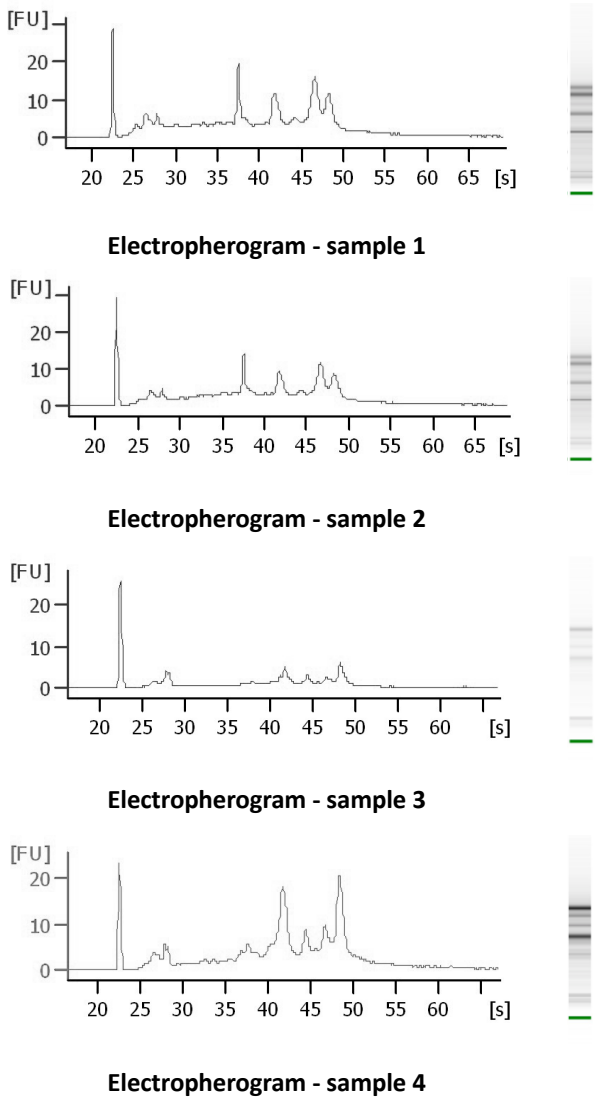


Figure 3.5 Bioanalyzer analysis of the nuclear RNA from samples with different lysis time and lysis buffers. All samples were partially degraded with a RIN of 6.1-7.5.

3.2.3 Nuclear isolation from preserved mouse kidney

In order to have the best nuclear RNA quality, nuclear isolation was performed using freshly harvested mouse kidneys in the previous experiments. In some cases, samples need to be harvested on a specific date and preserved for later use. Thus, nuclear isolation from preserved renal tissue was also tested.

Two different ways of kidney preservation, snap-frozen and RNA later, were tested. Mouse kidney preserved with snap-frozen was frozen by liquid nitrogen immediately after kidney harvest and stored in a -80 ^{o}C freezer. Kidney preserved in RNA later was put into chilled RNA later after kidney harvest and stored in a -80 ^{o}C freezer.

Nuclear isolation was performed 1-4 weeks after kidney harvest. Each sample was prepared using an adult mouse kidney. The two kidneys harvested from the same mouse were preserved in RNA later and by snap-frozen separately. Nuclear isolation from bilateral kidneys of the same mouse was performed on the same day to make the result comparable. The nuclei suspension was measured by Muse cell analyser for the cell count and viability, Luna automated fluorescence cell counter for the cell count and nuclear morphology, Nanodrop for RNA amount and Bioanalyzer for the RNA quality.

Compared to samples preserved in the RNA later, samples preserved by snap-frozen had almost 2 times of nuclei yield and better lysis efficacy (Table 3.6). The cell count was similar in the Luna cell counter, but all samples had higher viability. The difference in viability between Muse and Luna cell counters could be related

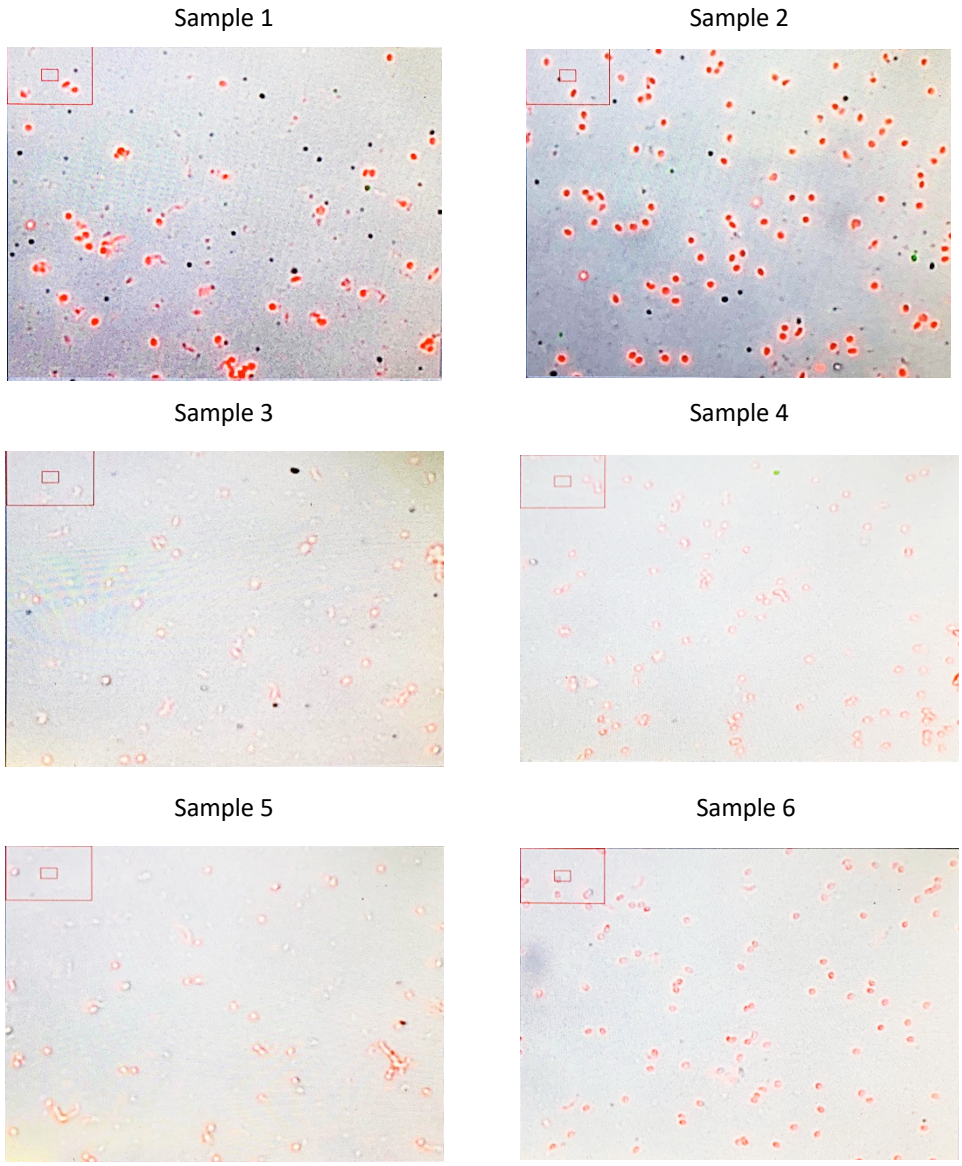
to the way of sample measurement, in which the Muse cell analyser used the flow and gating method and the Luna cell counter used fluorescence staining. Multiplets was easily found in samples preserved in RNA later and lesser in samples preserved by snap-frozen (Figure 3.6). Although the calculated nuclei numbers of samples stored in different ways were different, there was no such difference in the extracted nuclear RNA amounts between the two conditions. All samples had good RINs, indicating both methods were suitable for preserving nuclear RNA for at least a month after tissue harvest. Samples preserved by snap-frozen had slightly better nuclear RNA quality compared to samples stored in RNA later. Snap-frozen was selected as the method for preserving renal tissue for snRNA-Seq because of higher nuclei yield, fewer multiplets and higher RIN compared to samples preserved using RNA later.

Table 3.6 Nuclei isolated from preserved mouse kidneys. The experiment used an adult mouse kidney for each sample. Bilateral kidneys from the same mouse were preserved in RNA later and liquid nitrogen separately and processed on the same day. The table shows the result of the Muse cell analyser, Luna automated fluorescence cell counter, Nanodrop and Bioanalyzer.

Mouse ID								
	1	1	2	2	3	3	4	4
Sample No.								
	1	2	3	4	5	6	7	8
Duration from kidney harvest to nuclear isolation (weeks)								
	1	1	2	2	3	3	4	4
Method of kidney preservation								
	RNA.L	LN	RNA.L	LN	RNA.L	LN	RNA.L	LN
Muse cell analyser								
Count ($x10^7$)	1.61	3.50	1.14	2.41	1.21	2.47	1.05	2.06
Viability (%)	2.4	0.7	2.0	0.7	1.3	0.5	1.4	0.6
Luna automated fluorescence cell counter								
Count ($x10^6$/mL)*	2.36	6.49	2.01	4.12	2.40	5.38		
Viability (%)	2.7	2.3	4.7	3.5	5.8	1.8		
Nanodrop analysis for the extracted nuclear RNA								
RNA conc. (ng/µL)	428.9	426.8	504.7	286.9	479.4	499.9	388.0	372.7
Total volume (µL)	30	30	30	30	30	30	30	30
Total nuclear RNA amount (µg)	12.87	12.80	15.14	8.61	14.38	15.00	11.64	11.18
260/280	2.06	2.11	2.12	2.09	2.12	2.10	2.12	1.94
260/230	0.64	0.76	1.21	0.54	0.93	0.67	1.37	1.13
Bioanalyzer analysis								
RIN	7.0	7.2	7.9	8.2	6.7	7.2	7.4	7.9

RNA.L = RNA later; LN= liquid nitrogen. *Each nuclei suspension was 4ml in volume.

Figure 3.6 Fluorescence-stained nuclei under Luna automated fluorescence cell counter. Nuclei were stained by an orange fluorescence. Samples preserved in RNA later (sample 1, 3, 5) had fewer nuclei than samples preserved by snap-frozen (sample 2, 4, 6). Multiplets were easily found in samples preserved in RNA later.



3.3 Discussion

In this chapter, I optimised the nuclear isolation protocol from human kidney cell lines, freshly harvested mouse kidneys and preserved mouse kidneys. Nuclear isolation experiments of cell lines helped to set up the workflow. It also provided evidence that Bioanalyzer could appropriately evaluate nuclear RNA quality. GEnC cells had relatively high viability (3.8%) with a lysis time of 5 minutes, indicating the best lysis time should be different across cell types. The representativeness of cellular composition from the snRNA-Seq result should not be over-interpreted. Severe nuclear RNA degradation was noted in the nuclear isolation experiments of mouse kidneys using the reference protocol. Mechanical destruction by tissue grinder was proved as the major cause of nuclear degradation. The quality of nuclear RNA was improved after the times of homogenisation were reduced. For renal tissue that nuclear isolation could not be processed immediately, snap-frozen was an optimal method for tissue preservation.

The differences between cytosolic and nuclear RNA were widely-discussed. 10x Chromium Single Cell 3' Solution profiles polyadenylated transcripts. As the nucleus comprises only 10% of the total RNA of a whole cell, understanding the nucleo-to-cytosolic difference of polyadenylated transcripts helps with data interpretation. The nucleus contains both unspliced and spliced pre-mRNA, whereas only spliced mRNA can be transported out of the nucleus. More than 90% of human genes undergo alternative splicing and influence the final product of translation (Wang et al. 2008). Zaghlool et el. compared the result of DGE

analysis between cytosolic and nuclear RNA between fetal and adult frontal cortex (Zaghlool et al. 2021). The number of DGEs was significantly higher in the cytosolic RNA analysis. Unique DEGs from the nuclear RNA analysis were mainly enriched for mitochondrial translation, which mitochondria was pelleted together with the nuclear fraction in the study.

Accurate cell counting is critical for library preparation of snRNA-Seq. In this chapter, the hemocytometer, Muse counter, and Luna automated fluorescence cell counter have been used for cell counting. Cell/nuclei suspension had to be kept in a particular concentration with a careful pipet mix to get a more reliable result from the hemocytometer. A significant difference in cell count between the hemocytometer and Muse cell counter was noted in Table 3.1. Since the total nuclear RNA amount could refer to the actual nuclei number, the result of cell counting at the Muse cell counter was considered more reliable. Both Muse and Luna cell counter provided consistent results. Poor accuracy and reliability of hemocytometer were proved compared to flow cytometry (Viall and LeVine 2020). Muse cell counter was used as the standard method for the snRNA-Seq experiment because of its convenience and accessibility in our laboratory.

Nanodrop measured nuclear RNA for RNA amount and purity (Desjardins and Conklin 2010). Nucleic acids have absorbance maxima at 260 nm. The ratio of absorbance at 260nm to 280nm was used to evaluate the extracted DNA/RNA purity. The 260/280 ratios when the five nucleotides were measured independently were: Guanine: 1.15, Adenine: 4.50, Cytosine: 1.51, Uracil: 4.00,

and Thymine: 1.47 (www.nanodrop.com). Since Uracil is present in RNA and Thymine is present in RNA, pure RNA had a higher 260/280 ratio than pure DNA. Generally, pure DNA had a 260/280 ratio of ~1.8, and pure RNA had a 260/280 ratio of ~2.0. A low 260/280 ratio could be related to chemical contamination during the RNA extraction procedure, e.g., phenol, guanidine, or protein contamination. The ratio of absorbance at 260nm to 230nm was also used as a secondary measurement of nucleic acid purity. The 260/230 ratio was expected to be 2.0-2.2. A low 260/230 ratio could be related to contamination by materials with high 230nm reabsorption, e.g., TRIzol/QIAzol, EDTA, guanidine thiocyanate, phenol, or carbohydrates. Most nuclear RNA samples showed an adequate 260/280 but low 260/230 ratio, while the RNA quality measured by RIN was acceptable. Contamination with guanidine thiocyanate in the lysis buffer QIAzol was the most common cause of a low 260/230 ratio when using the column RNA extraction method. (https://www.qiagen.com/nl/resources /faq?id=c59936fb-4f1e-4191-9c16-ff083cb24574&lang=en). This would not compromise the reliability of downstream applications. It is also worth mentioning that RNA purity does not indicate integrity and quality.

Chapter 4

Single-Nucleus RNA Sequencing of Mouse Kidney with Chronic Aristolochic Acid Nephropathy

4.1 Introduction

CKD is a worldwide health issue that is clinically defined as the presence of renal structural or functional abnormalities for ≥ 3 months with implications for health (Levin 2013). The mechanisms of renal fibrosis have not been clearly described in the literature, thus hindering the development of therapy for CKD and renal fibrosis.

Aristolochic acid (AA), which can be extracted from the Aristolochiaceae family of plants, may induce direct proximal tubular injury, renal interstitial fibrosis and end-stage kidney disease in humans. A clinically relevant mouse model of chronic aristolochic acid nephropathy (AAN) was introduced in this experiment. Chronic renal fibrosis was induced in 8-9 weeks old male mice by intraperitoneal injection of AA (2.5mg/kg for 4 doses) over 2 weeks, followed by 2 weeks of renal recovery-fibrotic phase. SnRNA-Seq of kidney was then carried out to investigate the responses of PTCs during kidney fibrosis.

The hypothesis was that there might be underappreciated heterogeneity in PTCs, especially in fibrotic kidneys. The experimental aim of this chapter was to investigate the heterogeneity of PTCs and the responses of PTCs during kidney fibrosis using a clinically relevant mouse model of chronic AAN.

4.2 Results

4.2.1 Induction of chronic AAN in C57BL/6J mice

Six male C57BL/6J mice aged 8-9 weeks were randomly assigned to the chronic AAN group (n=3) and naïve control group (n=3) (Experiment ID: snRNA-Seq_AAN1). Chronic AAN was induced by intraperitoneal injection of 4 doses of AA within 2 weeks. There was no intervention for the naïve control group. All mice were euthanised 4 weeks after the first injection of the AAN group. Both kidneys were harvested. A quarter of the left kidney from two mice (randomly selected) in each group were processed for the snRNA-Seq experiment (Figure 4.1). Nuclear isolation was performed immediately following tissue harvesting. Rapid tissue processing to nuclei was employed to minimise artefacts arising during processing that may have limited discrimination of PTC sub-clusters in previous reports of single-cell sequencing from the kidney (Park et al. 2018; Wu et al. 2019). The rest of the kidney samples were preserved for further experiments, including histopathology.

The experiment was repeated to confirm the reproducibility of the result, increase the number of samples to four in each group in snRNA-Seq and increase the total number of cells (Experiment ID snRNA-Seq_AAN2). Data from the two experiments were combined in downstream bioinformatics analysis.

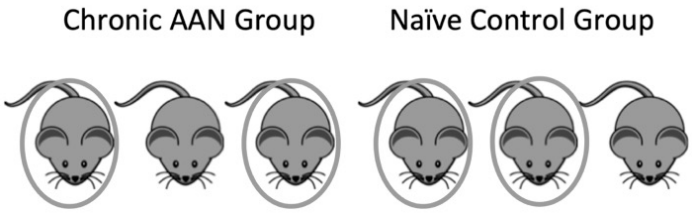


Figure 4.1 Study design. Six male C57BL/6J mice aged 8-9 weeks were randomly assigned to the chronic AAN group (n=3) and the naïve control group (n=3). Chronic AAN was induced in the AAN group by intraperitoneal injection of AA. A quarter of the left kidney from two mice in each group were processed for the snRNA-Seq experiment 4 weeks after the first injection.

4.2.2 Body weight and creatinine changes

A single dose of AA injection induces proximal tubular injury, whereas repeat injection induces tubular injury, followed by chronic inflammation and renal fibrosis. Bodyweight loss is a common adverse effect of AA injection and must be carefully monitored. The chronic AAN and the naïve group had a significant difference in actual body weight change ($p < 0.001$, *t*-test) and percentage of body weight change ($p < 0.001$, *t*-test) during the experiment (Table 4.1, Figure 4.2A, 2B).

Serum creatinine measurement was done. In the first experiment (snRNA-Seq_AAN1), mice in the AAN group had significantly higher serum creatinine than mice in the control group at the end of the experiment ($p = 0.03$, *t*-test, Figure 4.2C). In the second experiment (snRNA-Seq_AAN2), there were two outliers in serum creatinine values in the AAN group, these were a negative creatinine value (mouse ID AAN_4) and an extreme low creatinine value (mouse ID AAN_6).

Tracing back to the animal record, the two mice had had body weight loss after AA injection. Also, renal fibrosis was confirmed using a microscope. Therefore, the creatinine data in the second experiment was considered a measurement error.

Table 4.1 Body weight and serum creatinine of mice.

Experiment ID: SnRNA-Seq_AAN1						
Mouse ID	**Naïve_1**	**Naïve_2**	**Naïve_3**	**AAN_1**	**AAN_2**	**AAN_3**
BW D0 (g)	26.4	25.6	24.3	25.9	26.7	22.2
BW D28 (g)	29.0	28.7	25.8	25.1	26.2	22.2
Serum creatinine (μmol/dL)	14.24051	22.78481	25.63291	39.87342	37.02532	37.02532
Experiment ID: SnRNA-Seq_AAN2						
Mouse ID	**Naïve_4**	**Naïve_5**	**Naïve_6**	**AAN_4**	**AAN_5**	**AAN_6**
BW D0 (g)	25.3	26.4	21.2	23.8	25.3	24.3
BW D28 (g)	28.8	30.3	25.1	23.6	24.6	24.9
Serum creatinine (μmol/dL)	*12.46040*	*5.34017*	*7.71358*	*-3.85679*	*29.37093*	*0.59335*

BW=body weight. AAN=aristolochic acid nephropathy. Squares mark the mice whose kidneys were proceeded for snRNA-Seq.

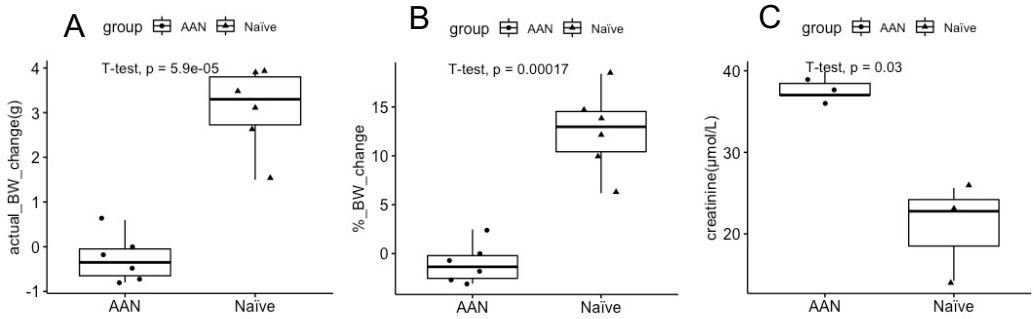


Figure 4.2 Body weight change and serum creatinine at the end of the experiment (day 28). Boxplots show the mean (thick line in the box) +/- standard deviation (upper and lower edge of the box) and the extreme value (end of the line) of the data. A significant difference **in** actual body weight change (A) and percentage of body weight change (B), and serum creatinine (C) between the AAN and naïve group were noted. Creatinine data only included results from experiment 100619-snRNAseq-AAN.

4.2.3 Nuclei number and retrospective RNA quality evaluation

The nuclear isolation step yielded 10-20 million nuclei from each quarter of the kidney, indicating that the estimated total cell number from an adult male mouse kidney was 40-80 million (Table 4.2). A slightly higher viability of the isolated nuclei was noted in the second experiment (snRNA-Seq_AAN2). The library preparation was performed by Dr Rachel Raybould, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University. 6400 nuclei of each sample were loaded into the 10x chip, targeting 4000 cells/sample recovery. The RNA sequencing was performed by the Genome Research Hub of School of Biosciences, Cardiff University. The nuclear RNA quality of the residual samples was measured after library preparation (Table 4.2). The RINs of Naïve_5, AAN_5 and AAN_6 showed partial degradation of nuclear RNA.

Library preparation was immediately processed after nuclear isolation on the experiment day. Nuclear RNA extraction of the residual nuclei suspension was usually done a few hours after kidney harvest and library preparation. Therefore, RNA degradation might happen during that period, and the RINs of the nuclei suspension samples were only for reference.

Table 4.2 Nuclei number and RIN of the isolated nuclei. The table shows nuclei yield, viability and nuclear RNA integrity. Naïve_5, AAN_5 and AAN_6 showed partial degradation of the nuclear RNA.

Experiment ID	**snRNA-Seq_AAN1**			
Mouse ID	**Naïve _1**	**Naïve_2**	**AAN_1**	**AAN _2**
Isolated nuclei (1/4 kidney)	$1.07x10^7$	$2.10x10^7$	$1.17x10^7$	$1.35x10^7$
Viability	0.20%	0.20%	0.90%	0.90%
RIN	6.0	6.7	7.5	7.0
Experiment ID	**snRNA-Seq_AAN2**			
Mouse ID	**Naïve _5**	**Naïve_6**	**AAN_5**	**AAN _6**
Isolated nuclei (1/4 kidney)	$1.37x10^7$	$9.9x10^6$	$1.47x10^7$	$1.39x10^7$
Viability	1.80%	0.90%	2.20%	0.80%
RIN	4.4	7.5	3.5	4.4

* RIN score was measured after library preparation on the experimental date, which was usually few hours after nuclei isolation. Nuclear degradation might happen in this period.

4.2.4 Histology evidence of fibrosis

The presence of renal fibrosis was confirmed histologically in mice exposed to AAN and normal histology in control animals. Figure 4.3A shows Masson's Trichrome stain of kidneys from healthy mice and mice with chronic AAN on day 28. Masson’s Trichrome stains cytoplasm in red, collagen in blue and nuclei in dark brown, and helps to identify renal fibrosis. Significant fibrotic change developed in mouse kidneys of aristolochic AAN. Micrographs shown are representative of 4 naïve mice and 4 AAN mice. Immunohistochemical staining of alpha-SMA showed increased interstitial staining in the AAN kidneys, which is compatible with increased myofibroblasts present in renal fibrosis.

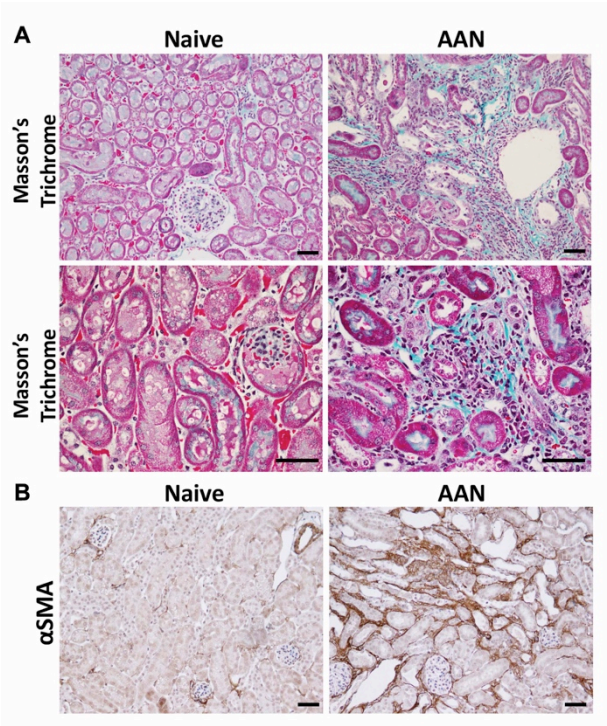


Figure 4.3 Microscopic evaluation of fibrosis in the naïve and the AAN mice. (A) Masson's Trichrome stain of kidneys from healthy mice (left, Naïve) and mice with chronic AAN (right, AAN) was taken on day 28. The cytoplasm is stained red, collagen is blue, and nuclei are dark brown, which helps to identify renal fibrosis. Significant fibrotic changes developed in mouse kidneys due to AAN. (B) Immunohistochemical stain of alpha-SMA and HAVCR1 for Naïve/AAN kidneys. Micrographs shown are representative of 4 naïve mice and 4 AAN mice.

4.2.5 Sequencing and genome mapping

The results of genome mapping and gene counting are shown in Table 4.3. Only nuclei with ≧ 400 genes and genes identified in ≧ 3 nuclei went into the downstream analysis. The Naive samples had 1207 genes/nucleus in median and 1906 UMIs/nucleus in median. The AAN samples had 1555 genes/nucleus in median and 2631 UMIs/nucleus in median. The low RINs in Naïve _5, AAN_5 and AAN_6 did not significantly affect the number of genes detected in this step.

Table 4.3 Number of nuclei and genes detected by using the zUMIs pipeline. The table shows the result of genome mapping of the 4 Naïve and the 4 AAN mice. The median of genes detected per nuclei varies between samples.

Experiment ID: SnRNA-Seq_AAN1				
Mouse ID	**Naïve_1**	**Naïve_2**	**AAN_1**	**AAN_2**
Number of nuclei	4505	2207	3964	3791
Genes detected	21514	21309	19974	20507
Genes detected per nuclei (mean)	1220.38	2057.85	1011.367	1320.546
Genes detected per nuclei (median)	931	2006	895.5	1290
UMIs detected per nuclei (mean)	2175.691	4130.239	1667.377	2332.055
UMIs detected per nuclei (median)	1339	3727	1315	2094
Experiment ID: SnRNA-Seq_AAN2				
Mouse ID	**Naïve_5**	**Naïve_6**	**AAN_5**	**AAN_6**
Number of nuclei	2146	2460	2587	3120
Genes detected	22596	22509	20944	21600
Genes detected per nucleus (mean)	2057.569	2105.51	1263.063	1570.371
Genes detected per nucleus (median)	1950	1995.5	1224	1466.5
UMIs detected per nucleus (mean)	4342.903	4374.669	2256.821	2867.268
UMIs detected per nucleus (median)	3479.5	3692	1946	2419.5

4.2.4 Quality control, data integration and doublet removal

Barcode-gene matrixes generated by zUMIs with the same condition were combined by the *merge* function of the Seurat package. After merging the 4 naïve kidneys, and 4 AAN kidneys, the number of feature counts, RNA counts and percentage of mitochondrial genes of each nucleus are shown in Figure 4.4 A-B. Nuclei with ≦ 400 genes or ≧ 7500 genes, or ≧ 10% mitochondrial genes, were removed, aiming to remove a few significant outliers. Only 94 nuclei were removed in this step. Then, the data was normalised and scaled for PCA. T-SNE plots were built on PCA results (Figure 4.4 C-D).

The estimated multiplet rate of each sample was ~3.1% by the user guide from the 10x company (https://www.10xgenomics.com/resources/user-guides/). Doublets were detected using an R package, Doublet Finder (McGinnis et al. 2019). Clusters with a high proportion of suspicious doublets, the cluster 12 of naïve kidney and the cluster 17 of AAN kidney, were completely removed from the downstream analysis (Figure 4.4E, 4.4F). Those labelled as dubious doublets but not in these 2 clusters were also excluded to minimise the impact of doublets on data integration and clustering. After quality control and doublet removal, a total of 23,885 nuclei were included in the downstream analysis (Figure 4.4G-H).

The nuclei number of each sample that went into the following quality control step and final analysis are listed in Table 4.4.

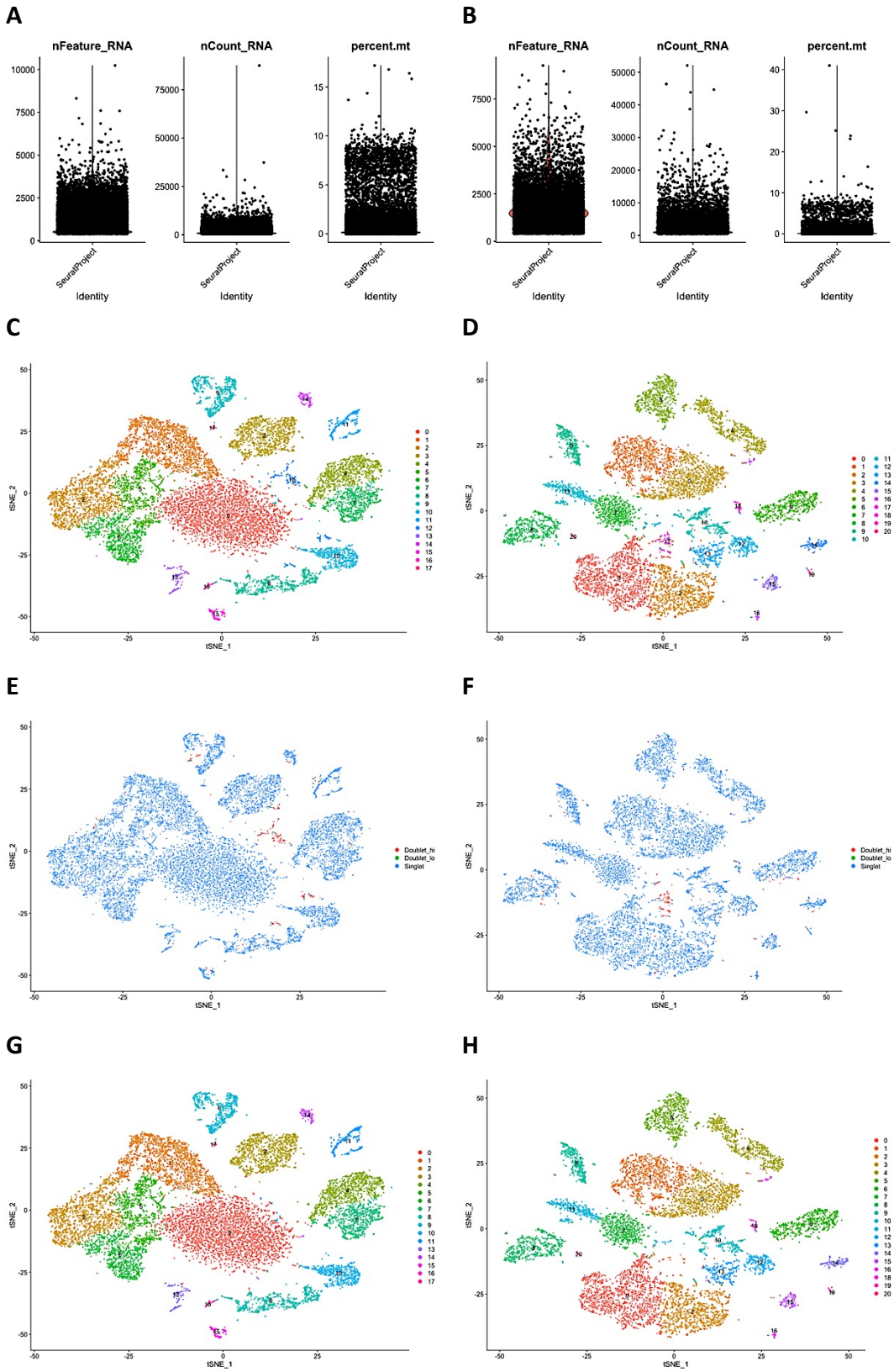


Figure 4.4 General information of the combined data of the naïve (left) and the AAN samples (right). (A)(B) Violin plots show the number of feature counts, RNA counts and percentage of mitochondrial genes of each nucleus. (C)(D) t-SNE plots shows clustering result of naïve and AAN

data. (E)(F) Results of doublet analysis. Nuclei identified as having a high probability of doublet were coloured in red, and a low probability of doublet were coloured in green. Nuclei that were identified as singlet were coloured in blue. (G)(H) t-SNE plots after doublet removal.

Table 4.4 Nuclei number of each sample after quality control steps.

Experiment ID: SnRNA-Seq_AAN1				
Mouse ID	**Naïve_1**	**Naïve_2**	**AAN_1**	**AAN_2**
Nuclei loaded into the chip	6400	6400	6400	6400
Nuclei detected	4505	2207	2146	2460
Quality control: keep nuclei with nFeature_RNA > 400 & nFeature_RNA < 7500 & percent.mt < 10	4495	2206	2139	2455
Nuclei number involved in the final analysis after removing suspicious doublets	4224	2137	2069	2350
Experiment ID: SnRNA-Seq_AAN2				
Mouse ID	**Naïve_5**	**Naïve_6**	**AAN_5**	**AAN_6**
Nuclei loaded into the chip	6400	6400	6400	6400
Nuclei detected	3964	3791	2587	3120
Quality control: keep nuclei with nFeature_RNA > 400 & nFeature_RNA < 7500 & percent.mt < 10	3933	3778	2562	3118
Nuclei number involved in the final analysis after removing suspicious doublets	3885	3689	2532	2999

4.2.5 Data integration and clustering

After doublet removal, data from the two conditions were integrated and clustered. Unbiased hierarchical clustering results were visualised using UMAP. To determine appropriate resolution of clustering, resolution 0.5, 1.0, 2.0, 3.0, 4.0, 8.0 of the *FindClusters* function were tested. A higher resolution would result in a higher number of clusters without affecting the shape of UMAP. With resolution of 0.5, 1.0, 2.0, 3.0, 4.0, 8.0, there were 20, 26, 37, 46, 52, 79 clusters identified in the combined dataset (Figure 4.5 A-F). A resolution of 3.0 was used in the following cell-type identification step, considering the number of cell types in a mouse kidney in the literature without missing any rare cell type, especially in renal fibrosis.

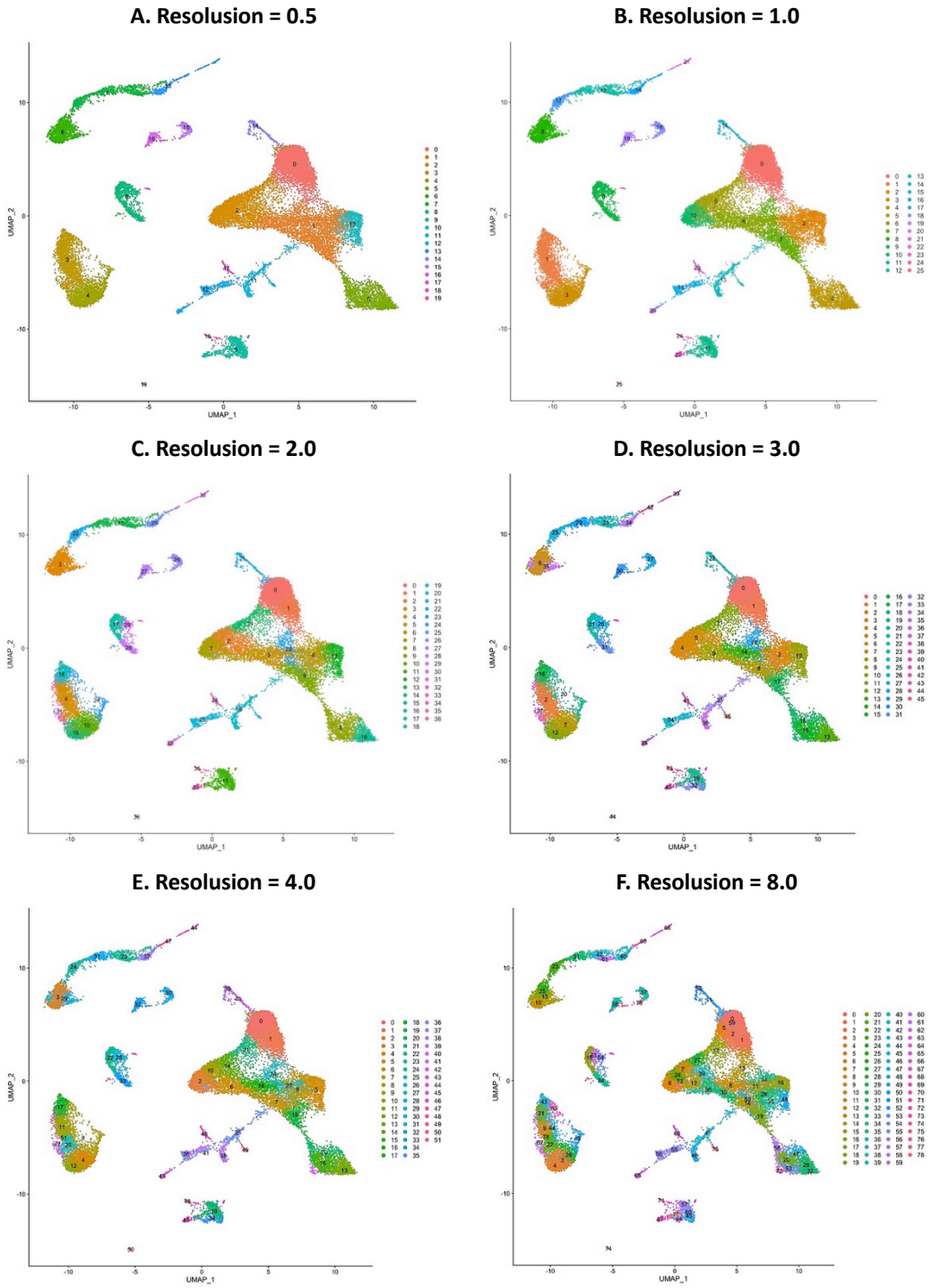


Figure 4.5 UMAP plots of the integrated dataset with different resolutions when performing nuclei clustering. A higher resolution would result in a higher number of clusters without affecting the shape of UMAP. With resolution of 0.5, 1.0, 2.0, 3.0, 4.0, 8.0, there were 20, 26, 37, 46, 52, 79 clusters identified in the combined dataset.

Nuclei on the UMAP were coloured by their sample origin to see if there was a significant batch effect from the biological replicates. Nuclei from different samples were distributed heterogeneously on the UMAP, indicating the data were well integrated (Figure 4.6). Results of data integration and clustering showed that the snRNA-seq result was reproducible with two conditions and two biological replicates of each condition in two identical experiments.

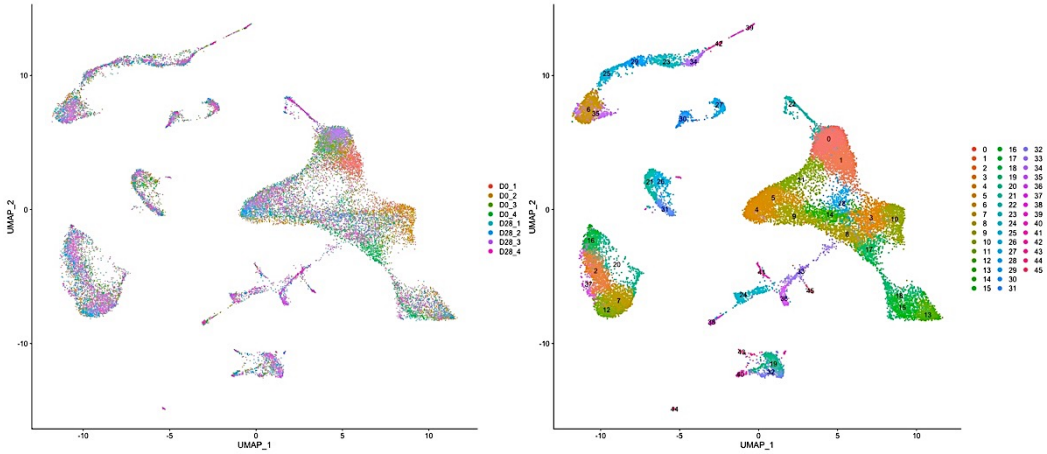


Figure 4.6 UMAP plots of the integrated dataset with a resolution of 3.0. In the right plot, the nuclei were coloured by their sample origin. Nuclei from different samples were distributed in the UMAP heterogeneously, indicating the data was well integrated.

T-distributed stochastic neighbour embedding (t-SNE) plots, another well-known nonlinear dimensionality reduction method, were also used to visualise the clustering result with resolution 3.0 (Figure 4.7). Compared to the t-SNE plot, clusters on the UMAP plot were more condensed and clearer.

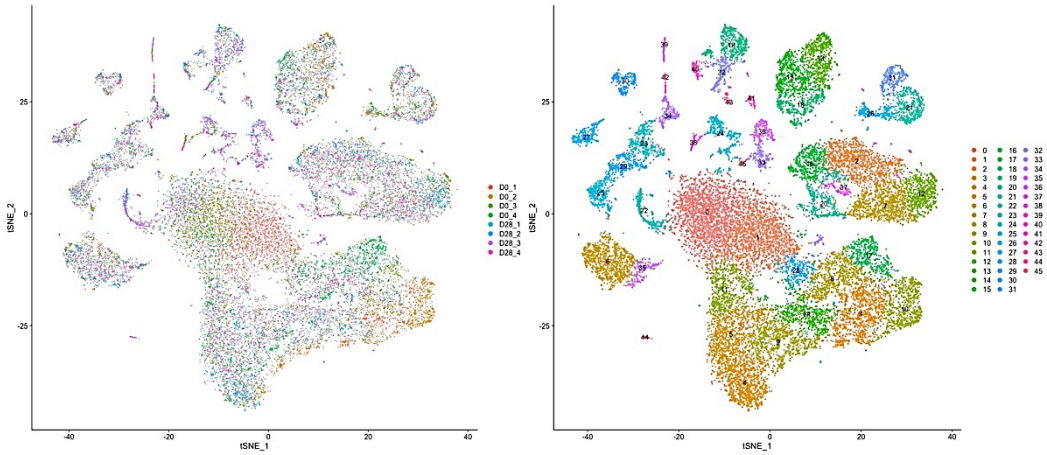


Figure 4.7 t-distributed stochastic neighbour embedding (t-SNE) plots of the integrated dataset with a resolution of 3.0. In the right plot, the nuclei were coloured by their sample origin. In the left plot, the nuclei were coloured by clusters.

4.2.6 Cell type identification

4.2.6.1 Identification of common cell-type of kidney

Canonical markers of kidney cell populations were used to identify major cell types in the kidney: podocyte *(Nphs1),* endothelial cells *(Flt1),* mesangial cells *(Igfbp5),* juxtaglomerular cells (JG cells) *(Ren1),* PTCs *(Slc34a1),* proliferative cells *(Top2a* and *Mki67),* descending thin limb *(Aqp1),* ascending thing limb *(Clcnka),* thick ascending limb (TAL) *(Slc12a1 and Umod),* distal convoluted tubule (DCT) 1 *(Slc12a3)* and 2 *(Slc12a3* and *Slc8a1)*, connecting tubule *(Slc8a1)*, principle cell-outer medullary collecting duct and inner medullary collecting duct *(Aqp2*, *Slc8a1* and *Atp6v1b1*, different expression level and percentage*),* intercalated cells type A *(Atp6v1b1* and *Slc4a1)* and type B *(Atp6v1b1* and *Slc26a4),* transitional epithelium *(Upk1b),* immune cells *(Ptprc)* and fibroblasts *(Pdgfrb and Cfh).*

Nphs1 encodes Nephrin, a podocyte-specific protein at the slit diaphragm with 8 extracellular immunoglobulin-like modules, a fibronectin type III-like domain, and a cytosolic C-terminal tail (Garg 2018). Mutation of *Nphs1* causes a congenital nephrotic syndrome, autosomal recessive congenital nephrotic syndrome of the Finnish type. *Nphs2,* which encodes another podocyte-specific protein, podocin, also showed specific expression with a high expression level in the podocyte cluster.

Flt1 was used as a marker gene for endothelial cells. *Flt1 (Vefgr1)* and *Flk-1/Kdr (Vegfr-2)* encode tyrosine kinases vascular endothelial growth factor receptor (VEGFR) 1 and 2, which acts as cell surface receptors for VEGFs, placenta growth

factor (PlGF) and endocrine gland-derived vascular endothelial growth factor (EG-VEGF) on endothelial cells (Ferrara 2004; Melincovici et al. 2018). *Kdr* also showed endothelial cellular specific expression with a high expression level in the endothelial cell cluster. Apart from endothelial cells, *Flt1* can be detected on inflammatory cells, monocytes/macrophages, bone marrow-derived hematopoietic progenitor cells, trophoblastic cells, renal mesangial cells, tumour cells and vascular smooth muscle cells (VSMCs) (Melincovici et al. 2018). The expression of *Flt1* on the cells mentioned above was not significant in my data. Marker genes used for cell-type identification are listed in Table 4.5.

Table 4.5 Marker genes used for cell-type identification.

Gene name	Protein name	Cell type
Nphs1	Nephrin	Podocyte
Flt1	Vascular endothelial growth factor receptor 1	Endothelium
Igfbp5	Insulin-like growth factor-binding protein 5, IGF-binding protein 5	Mesangial cell
Ren1	Renin-1	JG cell
Slc34a1	Sodium-dependent phosphate transport protein 2A, Sodium-phosphate transport protein 2A	PTC
Slc5a2	Sodium/glucose cotransporter 2, Na(+)/glucose cotransporter 2	PTC-S1
Slc22a6	Solute carrier family 22 member 6, Organic anion transporter 1	PTC-S2
Slc13a3	Solute carrier family 13 member 3, Na(+)/dicarboxylate cotransporter 3	PTC-S2
Slc5a10	Sodium/glucose cotransporter 5, Na(+)/glucose cotransporter 5	PTC-S3
Slc7a13	Solute carrier family 7 member 13, Sodium-independent aspartate/glutamate transporter 1, X-amino acid transporter 2	PTC-S3
Cyp7b1	Cytochrome P450 7B1	PTC-medullary S3
Vcam1	Vascular cell adhesion protein 1, V-CAM 1, VCAM-1, CD antigen CD106	New-PT1
Havcr1	Hepatitis A virus cellular receptor 1 homolog, HAVcr-1, Kidney injury molecule 1, KIM-1, T-cell immunoglobulin mucin receptor 1, TIM-1, CD antigen CD365	New-PT3
Akap12	A-kinase anchor protein 12	New-PT2&3
Top2a	DNA topoisomerase 2-alpha	Proliferative cell
Ki67	Proliferation marker protein Ki-67	Proliferative cell
Aqp1	Aquaporin-1, AQP-1	DTL
Clcnka	Chloride channel protein ClC-Ka, Chloride channel Ka, ClC-K1	ATL

Slc12a1	Solute carrier family 12 member 1, Bumetanide-sensitive sodium-(potassium)- chloride cotransporter 2, Kidney-specific Na-K-Cl symporter	TAL
Umod	Uromodulin, Tamm-Horsfall urinary glycoprotein, THP	TAL
Slc12a3	Solute carrier family 12 member 3, Na-Cl symporter, Thiazide-sensitive sodium-chloride cotransporter	DCT1&2
Slc8a1	Sodium/calcium exchanger 1, Na(+)/Ca(2+)-exchange protein 1. Solute carrier family 8 member 1	DCT2&CNT
Aqp2	Aquaporin-2, ADH water channel	PC
Atp6v1b1	V-type proton ATPase subunit B, kidney isoform, V-ATPase subunit B1	IC-A&B
Slc4a1	Band 3 anion transport protein, Anion exchange protein 1, Anion exchanger 1, MEB3, Solute carrier family 4 member 1, CD antigen CD233	IC-A
Slc26a4	Pendrin (Sodium-independent chloride/iodide transporter) (Solute carrier family 26 member 4)	IC-B
Upk1b	Uroplakin-1b	Transitional epithelium
Ptprc	Receptor-type tyrosine-protein phosphatase C, Leukocyte common antigen, Lymphocyte antigen 5, T200, CD antigen CD45	Immune cell
Pdgfrb	Platelet-derived growth factor receptor beta, Beta platelet-derived growth factor receptor, CD140 antigen-like family member B, Platelet-derived growth factor receptor 1, PDGFR-1, CD antigen CD140b	Mesangial cell & fibroblast
Cfh	Complement factor H, Protein beta-1-H	Mesangial cell & fibroblast

JG cell, Juxtaglomerular cells; PT, proximal tubule; S1/S2/S3, segment 1/2/3 of proximal tubule; DTL, descending thin limb; ATL, ascending thing limb; TAL, thick ascending limb; DCT1/DCT2, distal convoluted tubule1/2; CNT, connecting tubule; PC-OMCD, principal cell-outer medullary collecting duct; PC-IMCD, principal cell-inner medullary collecting duct; IC-A, intercalated cells, type A; IC-B, intercalated cells, type B.

Figure 4.8A shows the results of cell type indication on the UMAP. UMAP of each mouse and the number of nuclei comprising each cluster are shown in Figure 4.9 and Table 4.6. Nuclei from all mice contributed to every cluster, except for that subsequently labelled “new-PT3”, which was seen in 4 of 4 AAN-treated mice and none of 4 naïve controls. Figure 4.8B shows the expression level and the percentage of expression of the canonical genes per cluster on the dotplot.

Regional-specific genes of mouse kidneys identified by Ransick et al. were used to localize clusters in cortical-medullary and outer-inner medullary regions (Ransick et al. 2019). Fig 4.8C shows a regional expression of *Cyp2e1* in cortical PTCs and *Cyp7b1* in medullary PTCs. The expression of canonical genes for PC-OMCD and PC-IMCD indicates that PCs from the two different regions were well-clustered (Fig 4.8D). The top 100 genes with the lowest adjusted P value of each cluster are shown in the appendix.

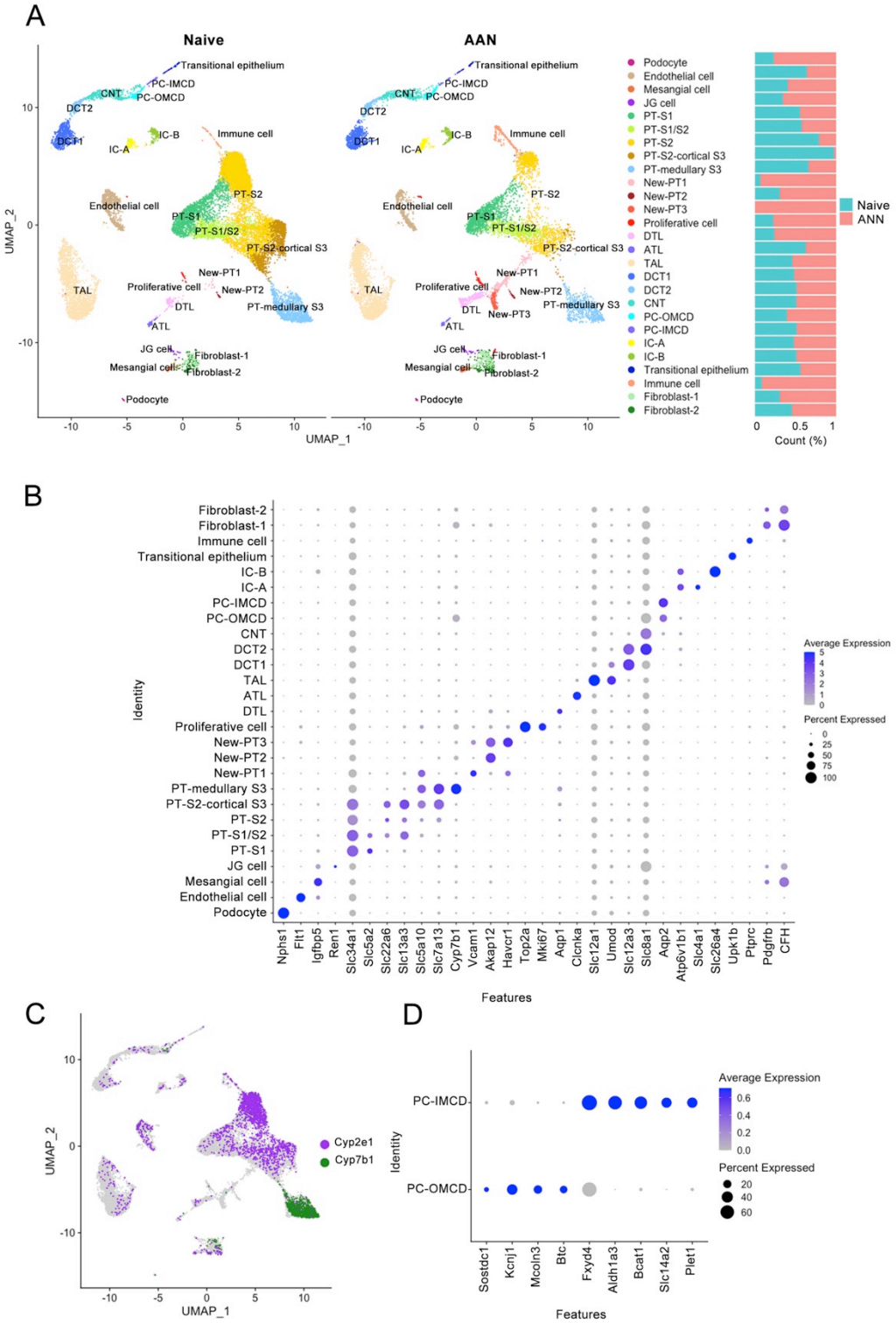


Figure 4.8. Clustering and cell-type identification of 23,885 nuclei using combined datasets from four naïve and four AAN mice. (A) The UMAP plot of the combined dataset is shown by splitting conditions. We identified all major cell types in the kidney and four new classes of cells labelled as proliferative cells and New-PT clusters 1-3. (B) Dotplot shows the expression levels and the percentage of gene expression of the canonical genes in each distinct cell type. (C) Feature plot of

regional-specific genes shows the expression of *Cyp2e1* (purple) in cortical PTCs and *Cyp7b1* (green) in medullary PTCs. (D) The expression of canonical genes of principal cells (PCs) in the outer and inner medullary collecting duct shows that PCs from the two regions were well-clustered.

JG cell, Juxtaglomerular cells; PT, proximal tubule; S1/S2/S3, segment 1/2/3 of proximal tubule; DTL, descending thin limb; ATL, ascending thing limb; TAL, thick ascending limb; DCT1/DCT2, distal convoluted tubule1/2; CNT, connecting tubule; PC-OMCD, principal cell-outer medullary collecting duct; PC-IMCD, principal cell-inner medullary collecting duct; IC-A, intercalated cells, type A; IC-B, intercalated cells, type B.

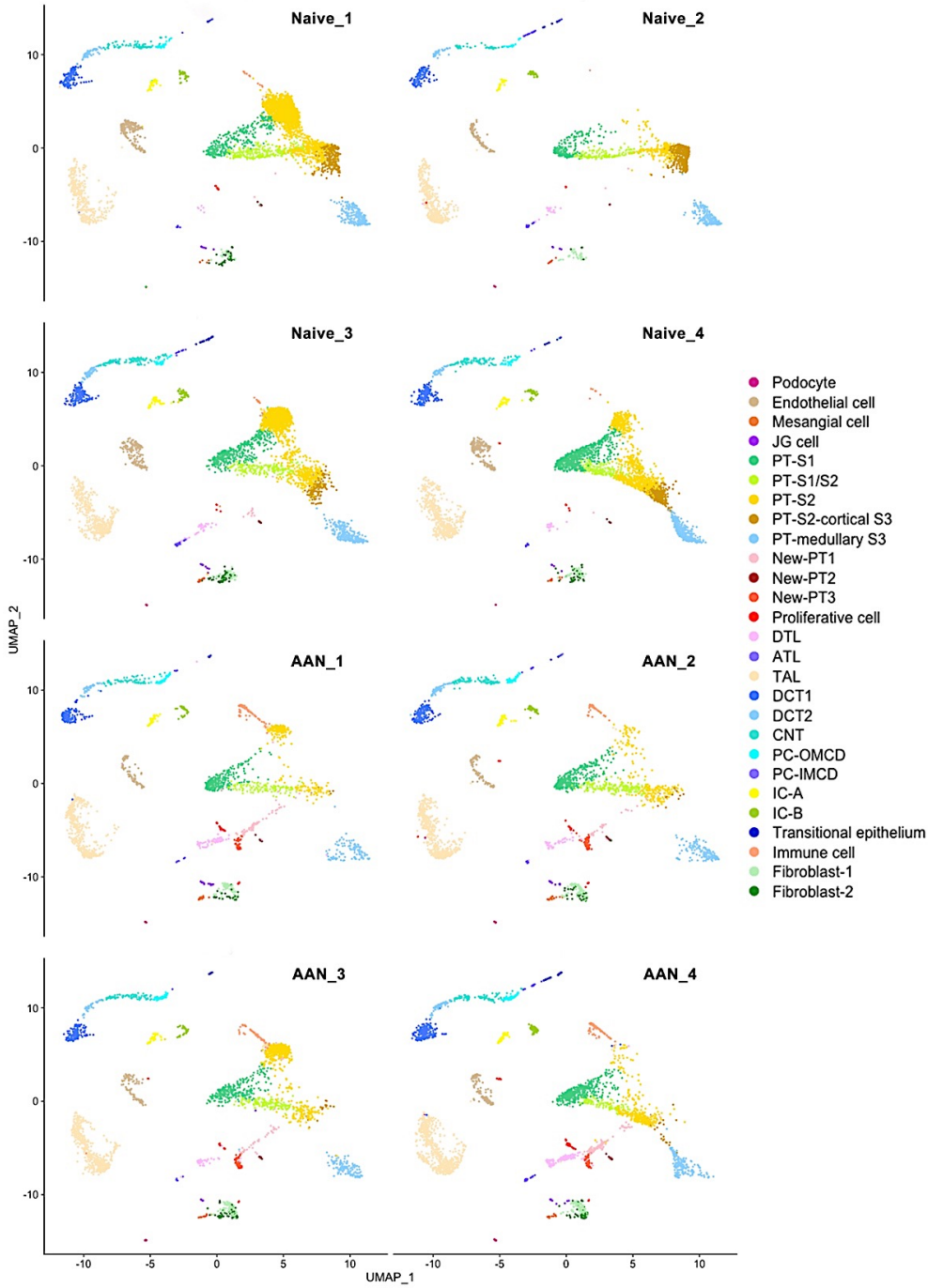


Figure 4.9 Distribution of nuclei from each mouse on the UMAP plots. Nuclei from all mice contributed to every cluster, except for that subsequently labelled "new-PT3", which was seen in 4 of 4 AAN-treated mice and none of 4 naïve controls.

Table 4.6 Number of nuclei comprising each cluster from each mouse.

Cell-type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Naïve_1	4	200	5	6	290	250	2103	282	239	2	5	0	8	8	5	357	144	51	85	29	1	32	37	12	14	15	40
Naïve_2	3	93	8	5	199	145	195	422	229	2	2	0	4	22	20	415	113	42	49	20	29	31	31	13	1	39	5
Naïve_3	4	163	20	10	381	181	1567	127	254	12	6	0	4	43	47	463	156	57	117	32	13	49	50	26	11	38	54
Naïve_4	2	203	11	4	590	205	858	381	394	2	4	0	6	21	8	482	161	43	106	28	4	33	42	11	7	49	34
Naïve total	13	659	44	25	1460	781	4723	1212	1116	18	17	0	22	94	80	1717	574	193	357	109	47	145	160	62	33	141	133
% of all naïve	0.09	4.73	0.32	0.18	10.5	5.61	33.9	8.70	8.01	0.13	0.12	0.00	0.16	0.68	0.57	12.3	4.12	1.39	2.56	0.78	0.34	1.04	1.15	0.45	0.24	1.01	0.96
Cell-type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
AAN_1	11	82	18	16	251	115	219	4	108	76	9	38	22	81	5	429	136	40	68	43	3	36	27	6	130	70	26
AAN_2	11	93	20	16	336	192	203	10	163	28	9	65	19	48	5	508	154	53	96	35	7	45	36	12	66	89	31
AAN_3	6	89	12	5	251	159	601	15	131	38	6	87	11	45	6	512	132	34	82	38	1	30	34	7	72	70	58
AAN_4	16	117	16	11	371	119	292	19	189	123	15	50	26	131	34	619	205	58	105	57	34	50	59	25	109	100	49
AAN total	44	381	66	48	1209	585	1315	48	591	265	39	240	78	305	50	2068	627	185	351	173	45	161	156	50	377	329	164
% of all AAN	0.44	3.83	0.66	0.48	12.2	5.88	13.2	0.48	5.94	2.66	0.39	2.41	0.78	3.07	0.50	20.8	6.30	1.86	3.53	1.74	0.45	1.62	1.57	0.50	3.79	3.31	1.65
Cell-type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
All	57	1040	110	73	2669	1366	6038	1260	1707	283	56	240	100	399	130	3785	1201	378	708	282	92	306	316	112	410	470	297
% of all nuclei	0.24	4.35	0.46	0.31	11.2	5.72	25.3	5.28	7.15	1.19	0.23	1.01	0.42	1.67	0.54	15.8	5.03	1.58	2.96	1.18	0.39	1.28	1.32	0.47	1.72	1.97	1.24

Cell type: 1 = "Podocyte", 2 = "Endothelial cell", 3 = "Mesangial cell", 4 = "Juxtaglomerular cells", 5 = "PT-S1", 6 = "PT-S1/S2", 7 = "PT-S2", 8 = "PT-S2-cortical S3", 9 = "PT-medullary S3", 10 = "New-PT1", 11 = "New-PT2", 12 = "New-PT3", 13 = "Proliferative cell", 14 = "Descending thin limb", 15 = " Ascending thing limb", 16 = " Thick ascending limb", 17 = "Distal convoluted tubule-1", 18 = "Distal convoluted tubule-2", 19 = "connecting tubule", 20 = "Principle cell-outer medullary collecting duct", 21 = "Principle cell-inner medullary collecting duct", 22 = " Intercalated cells type A", 23 = " Intercalated cells type B", 24 = " Transitional epithelium", 25 = " Immune cell", 26 =”Fibroblast-1", 27=”Fibroblast-2”.

4.2.6.2 Clarifying the Pdgfrb+ clusters

Some functionally distinct renal cell types share lineage and have many common features in their gene expression profiles, complicating their recovery in discrete clusters in recent landmark studies (Stewart et al. 2019; Wu et al. 2019). These include mesangial cells, JG cells and fibroblasts, which differentiate from *Foxd1*+ cortical stromal cells and acquire similar genetic features after profibrotic stimulation (Johnson et al. 1992; Kobayashi et al. 2014; Karaiskos et al. 2018). Four clusters closely proximal to one another on UMAP shared expression of the shared markers for mesangial cells, fibroblasts and JG cells: *Cfh, Fhl2* and *Pdgfrb* (Figure 4.10). Mesangial cells were identified from the expression of the mesangial-restricted *Igfbp5* and *Itga8* (http://www.proteinatlas.org) (Uhlen et al. 2015; Lu et al. 2017), while JG cells were the only cluster expressing *Ren1*. Two renal fibroblast clusters were identified, of which the Fibroblast-1 cell number increased >200% in the fibrotic kidney. We provisionally identified Fibroblast-1 as myofibroblast-containing based on *Meis1* expression (Chang-Panesso et al. 2018). Our dataset is limited in this regard by low *Acta2* detection (Figure 4.10B), and further characterisation of renal fibroblast populations may benefit from supplementary technical approaches.

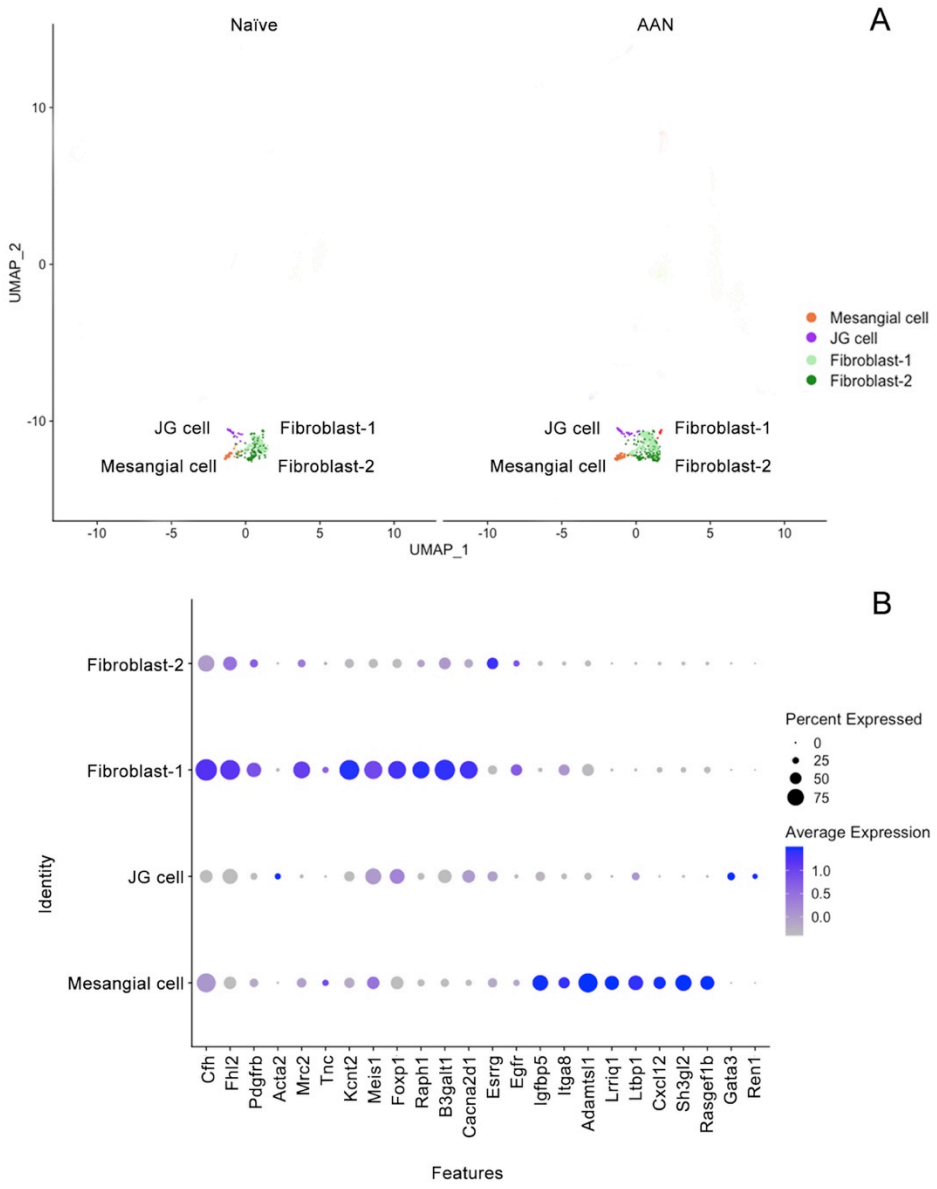


Figure 4.10 Characteristics of the Pfgfbr+ clusters. (A) UMAP of the Pdgfrb+ clusters, split by naïve and AAN conditions. (B) Dot plot of the canonical genes of the Pdgfrb+ clusters.

4.2.6.3 Analysis of proliferative cells and cell cycle

One cluster expressed markers of proliferative cells not seen in other clusters, notably including *Top2a* and *Mki67*. Uniquely, this cluster comprised four distinct and separated subclusters of nuclei. These four subclusters were located in close proximity to the endothelial, PTC, TAL and fibroblast clusters, respectively (Figure 4.11A). Consistent with this cluster comprising proliferating cells from the adjacent clusters, the four subclusters also expressed canonical markers for the cell type (e.g., *Flt1*-endothelial, *Slc34a1*-PTC, *Slc12a1*-TAL, *Pdgfrb*-fibroblast, Figure 4.11B). Cell cycle analysis identified cells in G2/M and S phase, localized mainly to the proliferative and immune cell clusters (Figure 4.11C). G2/M arrest of PTCs contributes to fibrogenesis after kidney injury (Yang et al. 2010b), and cell numbers of G2M and S phase PTCs increased in the AAN kidneys.

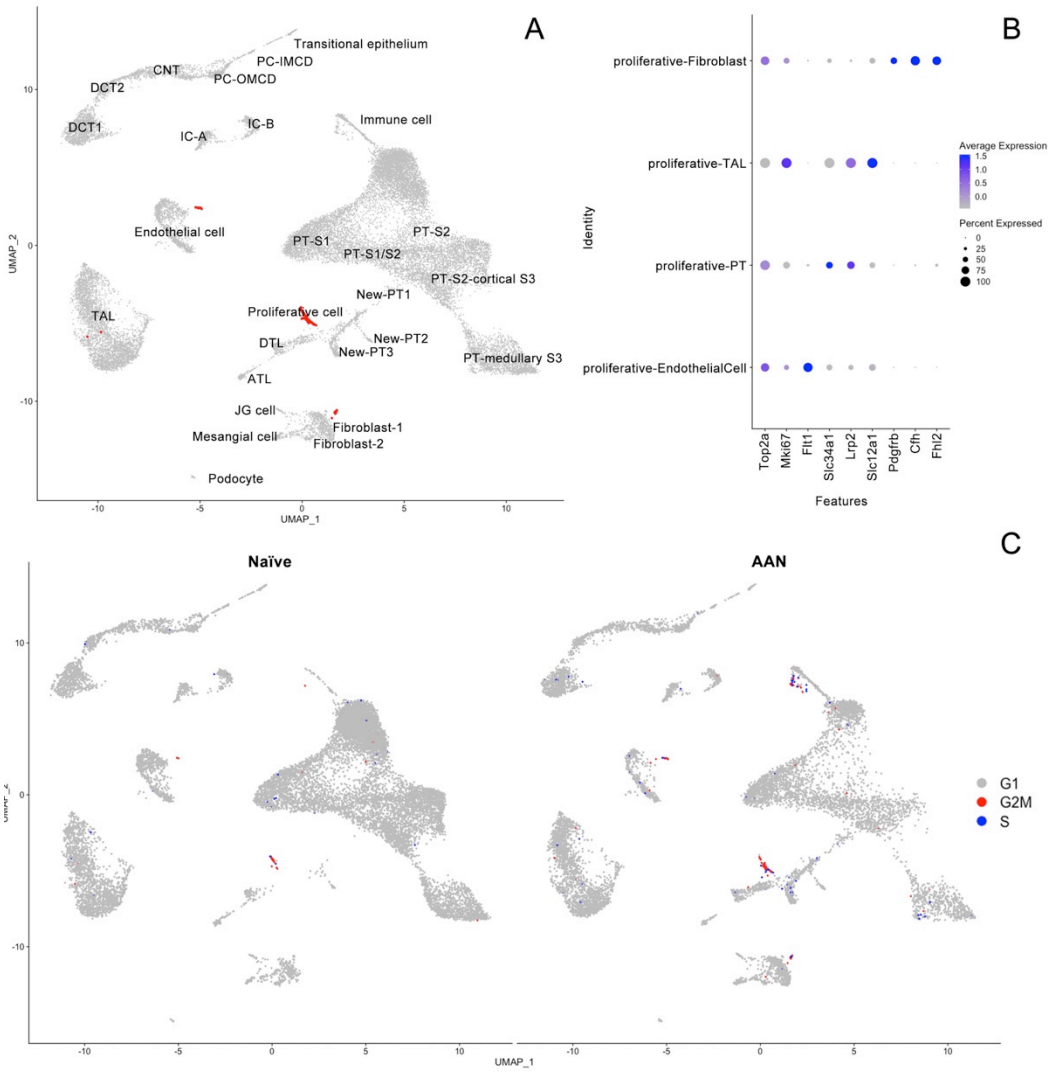


Figure 4.11 Cell cycle analysis. (A) The proliferative cells were segregated into 4 subclusters on the UMAP. The location of the subclusters on the UMAP was close to their original cell types. (B) The dotplot of the subclusters showed these proliferative cells persistent expressed canonical genes of their original cell types. (C) Cell cycle analysis of the combined dataset showed theG2/M and S phase cells were mainly in the immune cell and proliferative cell cluster. An increased number of G2/M and S phase cells was noted in the fibrotic kidney.

4.2.6.4 Analysis of PTC subclusters

The proximal tubule is divided into segments S1, S2 and S3 based on microscopic characteristics. The S1 segment is the longest and comprises PTCs with extensive apical microvilli, basolateral infoldings, cytoplasmic complexities, numerous long mitochondria, and a prominent endocytic region (Jacobson 1981). These features are present but less evident in the S2 segment, which demonstrates a gradual transition from the S1 segments and additionally displays more numerous peroxisomes and larger secondary lysozymes. The S3 segment is more distinct, comprising simple cuboidal cells without the above features.

Five clusters were identified corresponding to proximal tubular segments S1-S3: “S1”, “S1-2”, “S2”, “S2-cortical S3” and “medullary S3”. These PTC clusters were labelled according to their expression of genes localized to PTC segments in previous studies. *Slc34a1 encodes* Na^+*-*P_i *cotransporter 2a (NaPi-2a)*, localized in fully differentiated PTCs along with all segments except for medullary S3(Kusaba et al. 2014; Chen et al. 2019a). *Slc5a2* encodes a low affinity/high capacity Na^+/glucose cotransporter, sodium/glucose cotransporter 2 (Sglt2), responsible for 80-90% of the glucose reabsorption in S1(Kanai et al. 1994). *Slc22a6* encodes organic anion transporter 1 (Oat1) in S2, which plays a critical role in drug and xenobiotic elimination, and has been linked to AAN(Dantzler and Wright 2003; Hwang et al. 2010; Breljak et al. 2016; Chen et al. 2019a; Li et al. 2020a). *Slc13a3* encodes sodium-dependent dicarboxylate transporter (NaDC3) in S2, which is responsible for the transport of succinate and other Krebs cycle intermediates.

Slc7a13, also known as Agt1, is an amino acid transporter localized to the apical membrane of the S3 segment and is considered an S3-canonical gene(Nagamori et al. 2016; Chen et al. 2019a).

The expression levels of segment-specific solute transporter related genes were reviewed to validate established PTC categories. Relevant genes were identified in the first instance from the study from the Knepper lab, in which deep bulk sequencing was applied to microdissected renal tubules to identify nephron segment-specific transcriptomes (Lee et al. 2015). Additional studies are indicated for the validation of key anchor genes below. These included markers for S1 (*Slc5a2, Slc5a12*), S2 (*Slc22a6, Ca4, Slc13a3*), S3 (*Slc5a10*, *Slc7a13, Atp11a*), medullary S3 (*Cyp7b1*, *Slc6a13, Slc34a1(-)*), and pan-PTC markers (*Slc34a1, Lrp2, Slc4a4*) (Figure 4.12A) (Thiagarajan et al. 2011; Kamiyama et al. 2012; Lee et al. 2015; Cao et al. 2018; Ransick et al. 2019). The differences in genes enriched in each PT cluster supported their attribution to specific PT segments (Figure 4.13).

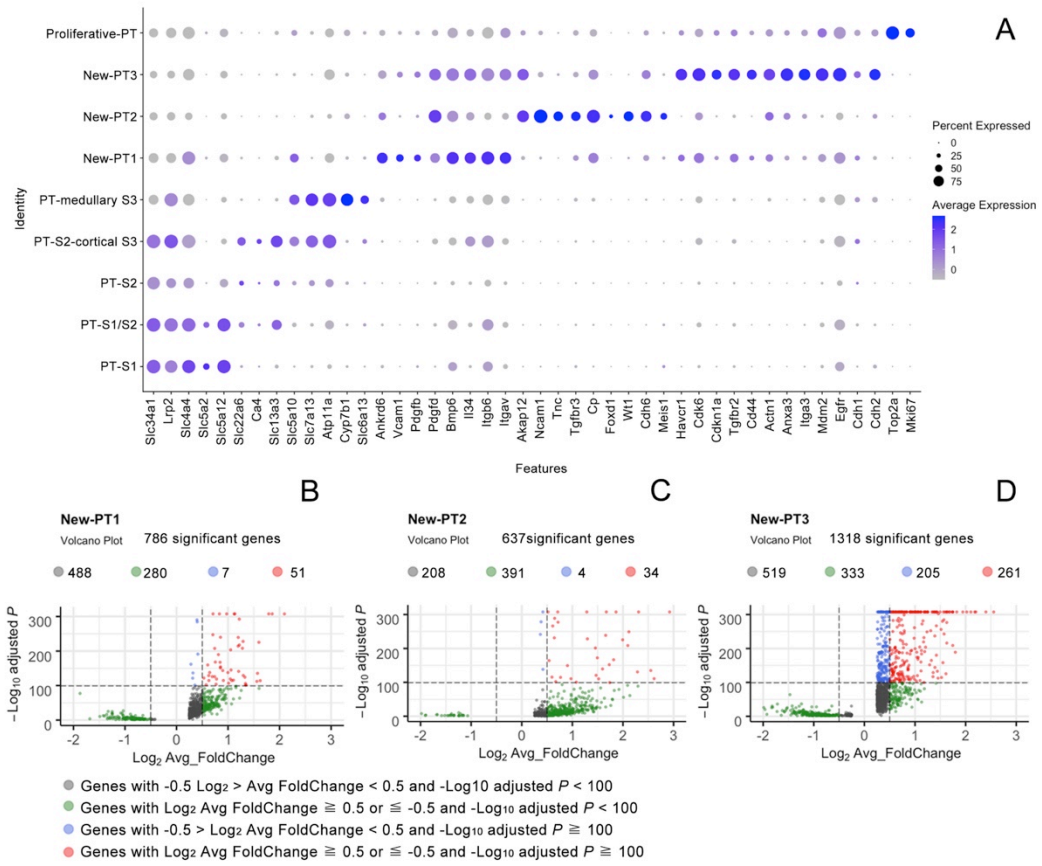


Figure 4.12 Gene expression profiles of Proximal Tubular Cell clusters. (A) Dotplot shows the expression levels and the percentage of gene expression in each cluster of the canonical genes in the normal and new classes of PTCs. (B) (C) (D)Volcano plots show the significant genes in differentially expressed gene analysis of the three New-PT clusters by comparing the RNA profiles of one cluster to all other clusters of the dataset. Significance was defined as a gene with an adjusted p-value < 0.05, a ≧ 0.25 average log fold difference between the two groups of cells, and presence detected in at least 10% of cells in either of the two populations.

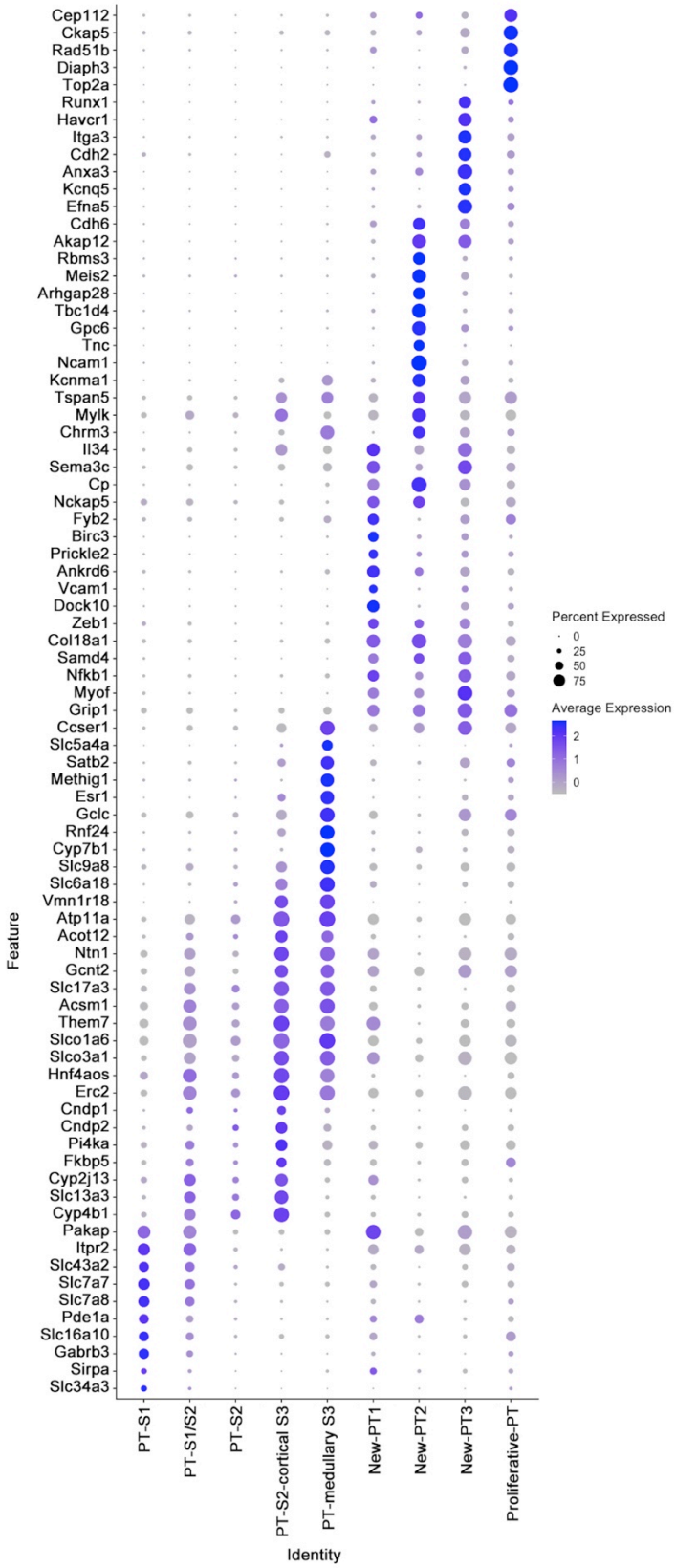


Figure 4.13 Dotplot of the cluster-enriched gene expression of the nine PT clusters. The difference in cluster-enriched gene expression of these PT clusters shows the identity of each PT cluster.

4.2.6.5 New PTC clusters identification

Four additional clusters were located in close proximity to clusters S1, S1-2, S2, S2-cortical S3, and medullary S3 on UMAP. First, the most significant component of the distributed proliferative cluster. This part of the proliferative cluster, labelled proliferative PTCs, expressed proximal tubular markers and proliferative marker genes, including *Ki67* and *Cdca3*, that Park *et al.* highlighted as identifying a novel cell type in normal mouse kidney, and that Wu *et al.* identified as proliferative-PT in the unilateral ureter obstruction model of renal fibrosis(Park et al. 2018; Wu et al. 2019). The other three clusters showed strong signals for genes upregulated in tubules following kidney injury in previous bulk sequencing and other historical approaches (*Vcam1, Havcr1 and Akap12*), demonstrated expression of canonical PTC genes, and were labelled as New-PT clusters. Clusters New-PT1 and 2 were increased in abundance in kidneys undergoing fibrosis following injury, in which circumstances the "New-PT3" were also found. The discrete gene expression signatures of New-PT1-3 clusters included genes associated with renal injury response and fibrosis progression (Figure 4.12A). DGE analysis comparing each cluster to all other clusters of the whole dataset detected 786 differentially expressed genes in New-PT1, 637 in New-PT2 and 1318 in New-PT3. Analysis of adjusted *p*-value and average log-fold change of differentially expressed genes in New-PT 1-3 clusters demonstrated that each had a discrete identifying signature (Figure 4.12 B-D).

4.2.7 Characteristics of the new classes of PTCs

4.2.7.1 Trajectory and velocity analysis

Trajectory analysis was performed to infer paths of cell state transitions within the PTC clusters. Trajectory analysis showed a continuous change in RNA profile in normal PTCs, ordered along the anatomical axis from S1 to S3 tubular segments (Figure 4.14A). RNA velocity analysis was then employed, in which the RNA processing activity evident in the transcriptome of each cell is used to evaluate transcriptional reprogramming and predict the future state of cells. Represented graphically by an arrow, the velocity of each cell indicates its rate and direction of transcriptional change. RNA velocity analysis showed that the New PT 1-3 and the proliferative PTCs exhibited strong directional change towards other states, while the normal PT clusters exhibited stable transcriptional profiles, concordant with mature cell phenotypes (Figure 4.14B). The data further indicated that NewPT1 is an intermediate cell type, from which cells may differentiate in two directions, “NewPT1 - NewPT2 - proliferative PT - normal PTCs”, and "NewPT1 - NewPT3" (Figure 4.14B). Pseudotemporal ordering was also performed with time zero set at NewPT1, based on the results of the RNA velocity analysis (Figure 4.14C). An ordered progression of cell states in pseudotime was seen from New-PT states through anatomically distinct tubular segments (Figure 4.14C).

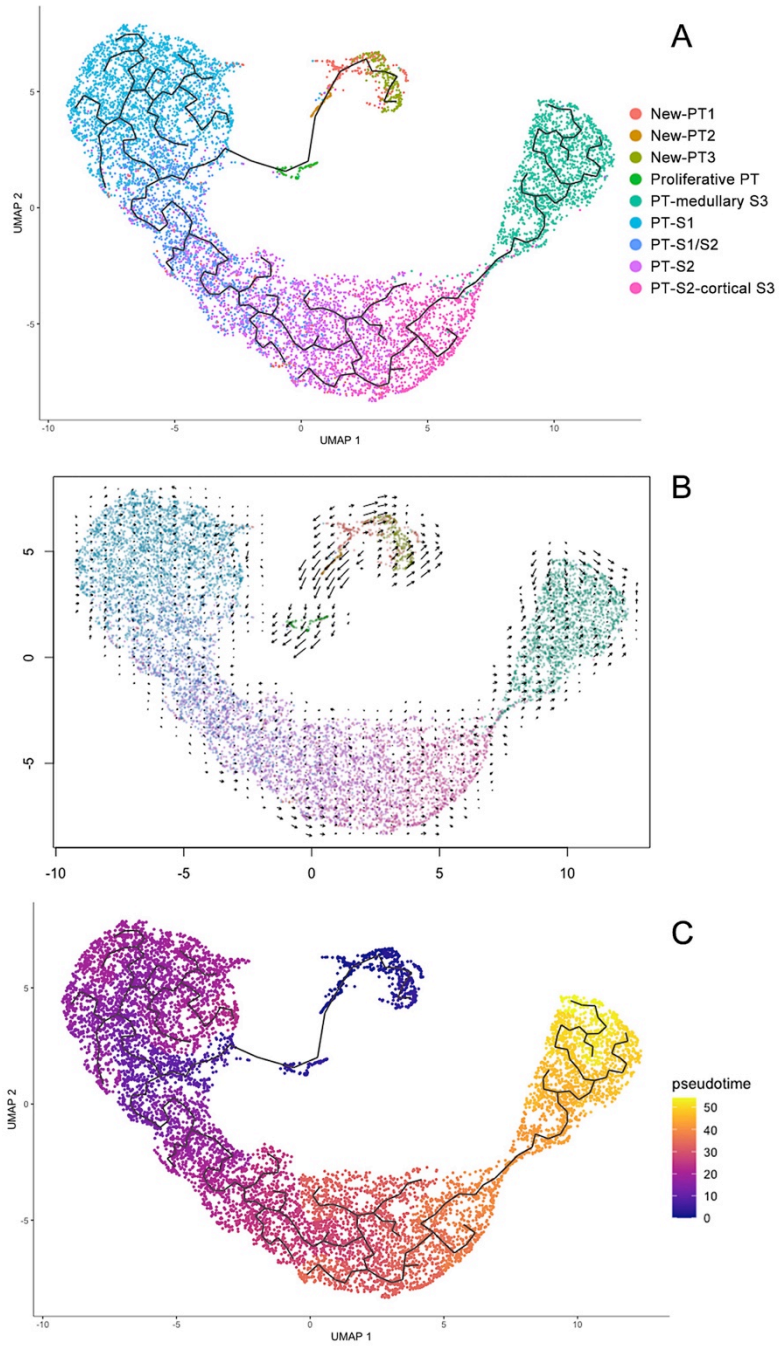


Figure 4.14 Trajectory, RNA velocity and pseudotime analysis of the Proximal Tubular Cells. (A) Trajectory analysis shows the dynamic process of the PTCs. RNA expression profiles change continuously along the anatomic axis in normal PTCs (PT-S1 to PT-S3). (B) RNA velocity analysis shows New-PT1 is an intermediate cell type that may differentiate toward two directions "New-PT1 – New-PT2 - proliferative PT - normal PTCs and "New-PT1 – New-PT3". The New-PT 1-3 and the proliferative PTCs are more likely to differentiation, and the normal PTs are the mature populations (C) PTC pseudotime analysis.

Genetic profiles suggested that the cells comprising these new PTC classes were dedifferentiated. Expression of *Kim-1, Vcam1, CD44, Anax3, Akap12* and *Ncam1* has been reported in dedifferentiated PTCs in the literature (Abbate et al. 1999; Smeets et al. 2013; Wu et al. 2019). Current evidence suggests that, following kidney injury, surviving differentiated PTCs can transform into dedifferentiated PTCs and then undergo proliferation and re-differentiation to restore normal proximal tubular morphology and function (Kusaba et al. 2014). We identified clusters labelled New-PT1 and New-PT2 and proliferative PTCs in normal kidneys, suggesting that this dedifferentiation-proliferation-differentiation process occurs in normal circumstances.

Trajectory and RNA velocity analysis suggested that the New-PT1 represented an intermediate cell type bridging normal PTCs, New-PT2, New-PT3 and proliferative-PT. New-PT1 expressed injury markers, including *Vcam1, Pdgfb, Pdgfd, Bmp6, Il34, Itgb6,* and *Itgav*. New-PT1 also partially preserved typical PTC markers, *Slc4a4* and *Slc5a10*. Some cells in New-PT1 also expressed *Havcr1*. This subset of *Havcr1*-positive New-PT1 cells was only identified in AAN mice (Figure 4.15). The New-PT1 cluster was labelled as “dedifferentiated- intermediate PTCs". New-PT2 cells expressed genes characteristic of the developing kidney and reactivated during tubular regeneration, including *Ncam1*, *Tnc*, *Tgfbr3, Foxd1* and *Wt1*, and was labelled as “dedifferentiated-regenerating PTC”. New-PT3 cells were detected only in AAN mice, indicating a PTC phenotype specific to progressive kidney fibrosis. *Havcr1*, also known as *Kim1*, was a prominent marker

of New-PT3, which also had a prominent expression of *Chd2*. Apart from genes related to kidney injury and fibrosis, this cluster expressed *Ckd6* and *Cdkn1a* (also known as *p21*), which are associated with the cell cycle and also with cell death. The SASP pertains to senescent cells with cell-cycle arrest that remain metabolically active and release senescence-associated proteins(Sturmlechner et al. 2017; Docherty et al. 2019). SASP-related genes, including *Cdkn1a (P21), Cdkn2b (P15), Tp53, Tgfb1, Serpine1 (Pai1), Ccl2 (Mcp1), Cxc1,* and *Ccn2*, were enriched in New-PT3 (Figure 4.16). This cluster was identified as "dedifferentiated-senescent PTC".

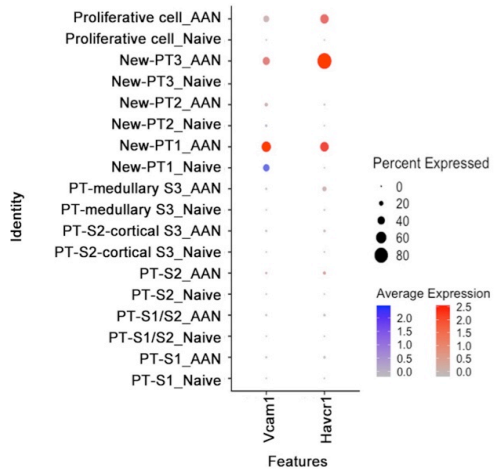


Figure 4.15 Expression of *Vcam1* and *Havcr1* in PT clusters. The dotplot shows that *Havcr1* positive New-PT1 cells were only identified in the AAN kidney.

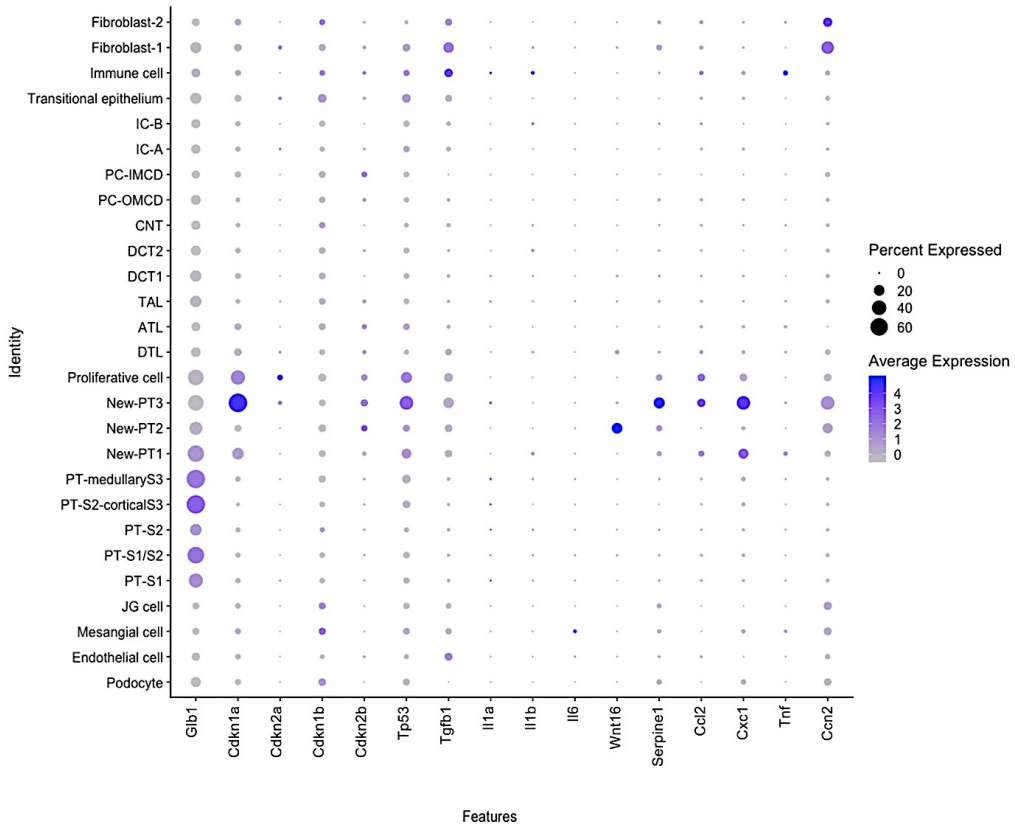


Figure 4.16 The new-PT3 cluster expressed the senescence-associated genes. Senescence-associated genes, including Cdkn1a (P21), Cdkn2b (P15), Tp53, Tgfb1, Serpine1 (Pai1), Ccl2 (Mcp1), Cxc1, and Ccn2, were enriched in New-PT3.

4.2.7.2 Pathway analysis of the new classes of PCTs

Pathway enrichment analysis was carried out for a further understanding of the molecular interaction network of new-PT1-3. Functional and signalling pathway analysis in KEGG showed a general reduction of metabolic pathways among the three new-PT clusters (Figure 4.17) (https://www.kegg.jp/kegg/). Environmental information processing was enriched, including signal transduction (KEGG category 3.2) and signalling molecules and interaction (KEGG category 3.3) pathways, except for the AMPK signal pathway, which acts as a sensor of cellular energy status. Consistent with categorisation as "dedifferentiated-senescence PTCs", new-PT3 showed enrichment of cell growth and death (KEGG category 4.2), including cell cycle, apoptosis, p53 signalling pathway cellular senescence. The cellular community (KEGG category 4.3) and cell motility (KEGG category 4.3) pathways were highly evident in the three new classes of PTCs, corresponding to their dedifferentiated phenotype. Immune system-activation pathways were also enriched in each new-PT cluster. A complete document summarising the pathway analysis results based on KEGG, Panther, Reactome, and WikiPathways databases is provided in Appendix.

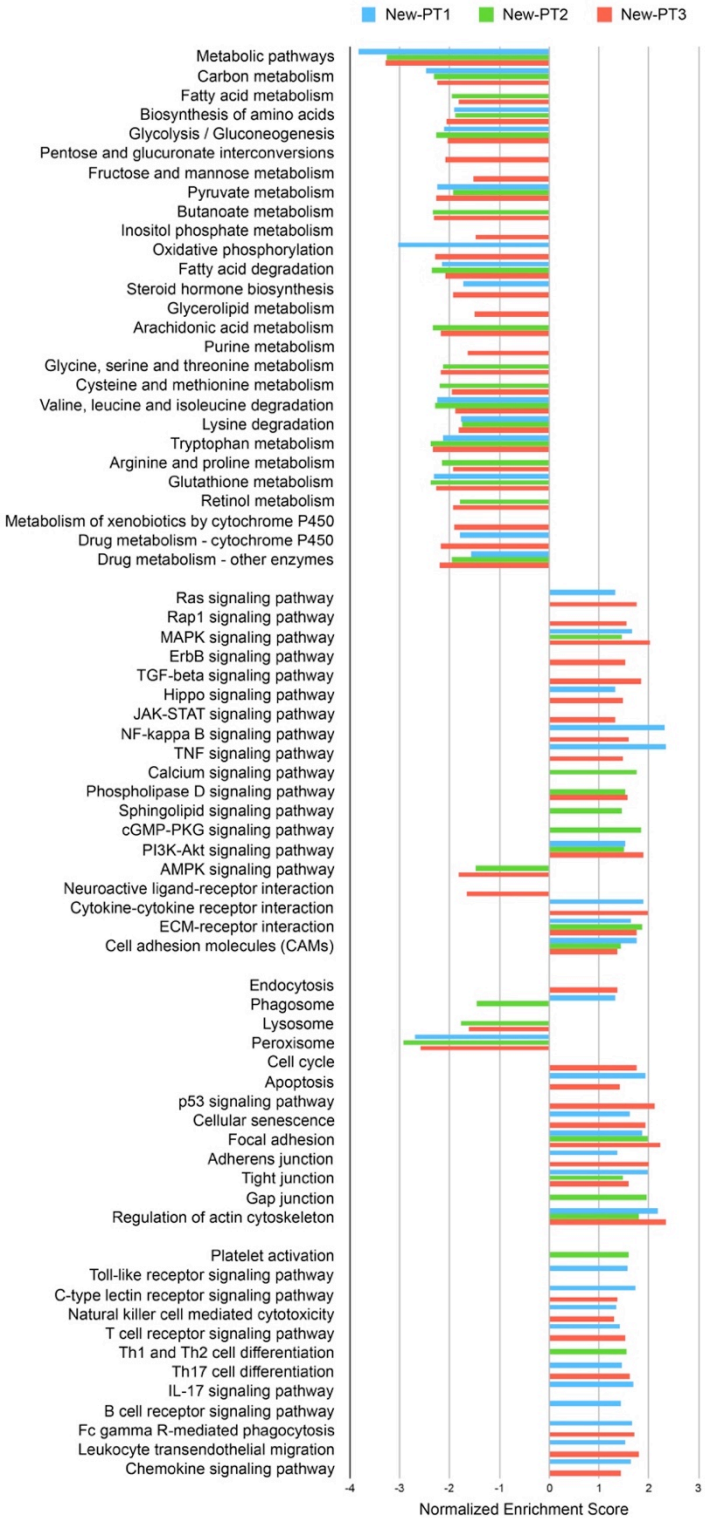


Figure 4.17 Pathway analysis based on KEGG database. Results show metabolic reprogramming, enrichment of cellular communication and cell motility, and various immune activations in New-PT clusters. The enrichment was calculated by comparing the RNA profiles of one New-PT cluster to all other PTC clusters. Pathways with a false discovery rate < 0.25 were listed.

4.2.7.3 Intercellular cross-talk between the new classes of PTCs and adjacent cell types

Ligands and receptors detected in clusters under analysis with a positive average fold-enrichment were paired to elucidate intercellular signal transduction networks. For this analysis, PT S1-3 were combined to "normal PT", and fibroblast-1, -2 were combined to “fibroblast” (Figure 4.18A). We summarised ligand-receptor pairs for ligands from the New-PT clusters and receptors from fibroblast, immune cell and normal PT clusters (Figure 4.18 B-D). The New-PT clusters expressed multiple profibrotic signals to fibroblasts, and analysis of the interaction of the New-PTs with immune cells showed several ligand-receptor pairs associated with macrophage activation. New-PTs also expressed signals to EGF receptors, expressed abundantly by normal PTCs. In comparison, normal PTs demonstrated low expression of ligands when compared to New-PTs. The strongest ligand signal from normal PTs was for netrin, a secreted laminin-related protein linked to suppressing inflammatory and injury signals in the kidney. The expression level of ligands and receptors in each cluster are shown in Figure 4.19.

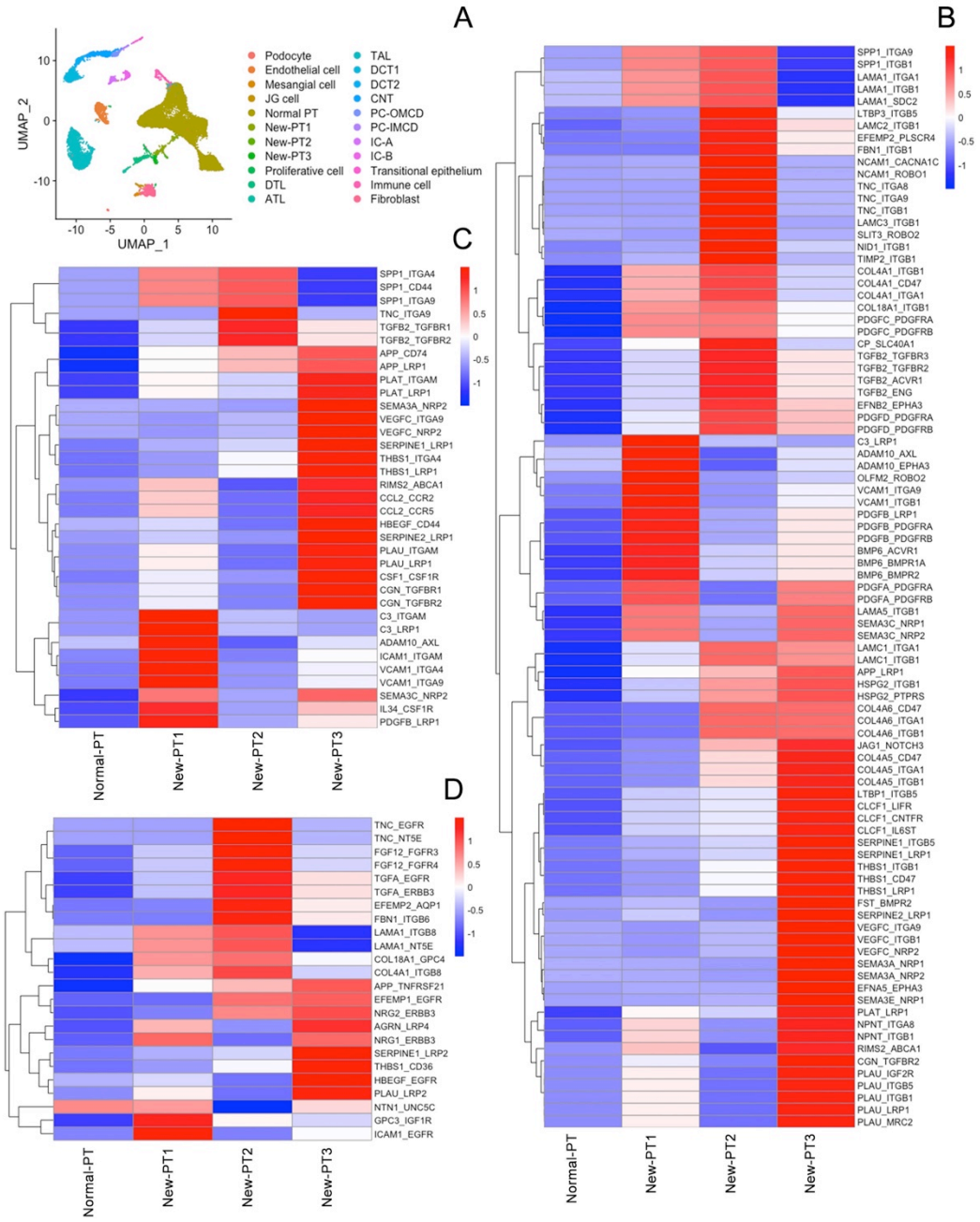
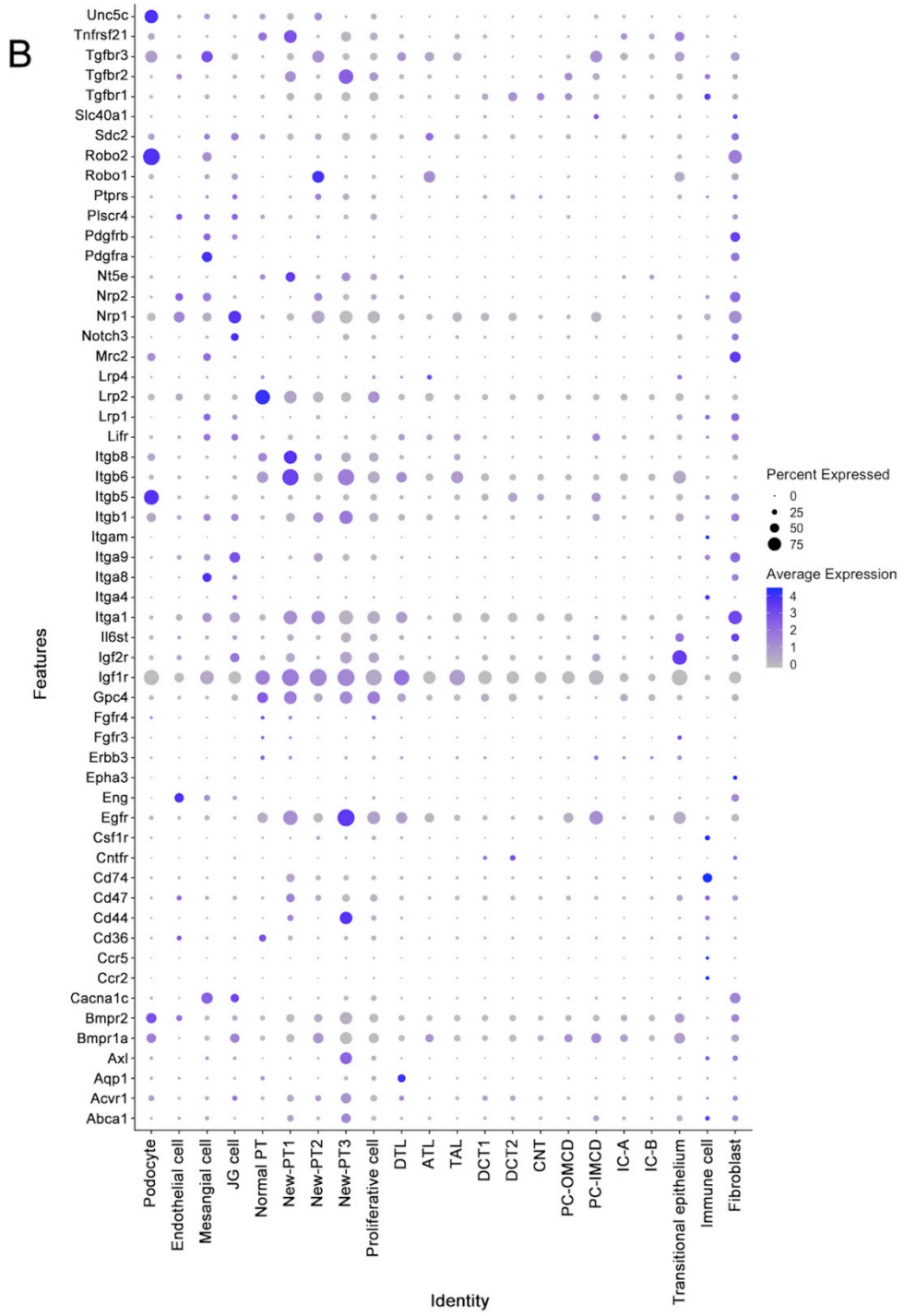


Figure 4.18 Ligand-receptor analysis. (A) PT S1-S3 are combined to normal PT and fibroblast-1, 2 are combined to Fibroblast for ligand-receptor analysis. Heatmap shows the cross product of ligand gene expression from normal PT, New-PT1, New-PT2, New-PT3 and receptor gene expression from (B) fibroblast, (C) immune cells, (D) normal PT.

Figure 4.19 Expression of ligands and receptors in each cell type. Dotplot shows the expression level of all (A) ligands (B) receptors in the ligand-receptor paring result.


B
Features
Unc5c
Tnfrsf21
Tgfbr3
Tgfbr2
Tgfbr1
Slc40a1
Sdc2
Robo2
Robo1
Ptprs
Plscr4
Pdgfrb
Pdgfra
Nt5e
Nrp2
Nrp1
Notch3
Mrc2
Lrp4
Lrp2
Lrp1
Lifr
Itgb8
Itgb6
Itgb5
Itgb1
Itgam
Itga9
Itga8
Itga4
Itga1
Il6st
Igf2r
Igf1r
Gpc4
Fgfr4
Fgfr3
Erbb3
Epha3
Eng
Egfr
Csf1r
Cntfr
Cd74
Cd47
Cd44
Cd36
Ccr5
Ccr2
Cacna1c
Bmpr2
Bmpr1a
Axl
Aqp1
Acvr1
Abca1
Podocyte
Endothelial cell
Mesangial cell
JG cell
Normal PT
New-PT1
New-PT2
New-PT3
Proliferative cell
DTL
ATL
TAL
DCT1
DCT2
CNT
PC-OMCD
PC-IMCD
IC-A
IC-B
Transitional epithelium
Immune cell
Fibroblast
Identity
Percent Expressed
0
25
50
75
Average Expression
4
3
2
1
0

4.2.8. Combined analysis with AKI

For further validation, I performed a combined analysis of snRNA-Seq results using this AAN-induced CKD dataset and the murine ischemia-reperfusion injury (IRI)-induced AKI dataset published by Kirita et al. (Kirita et al. 2020). In the study, kidneys from mice with 4 and 12 hours and 2, 14, and 42 days after IRI were harvested and processed. A total of 99,935 mouse AKI single-cell transcriptomes were analysed. The studies identified a distinct proinflammatory and profibrotic PTC status, named failed-repaired PTCs. The failed-repaired PTCs were found mainly in 14 days and 6 weeks after IRI and ageing rat kidneys and transplant kidneys.

The two datasets were pooled together and integrated by an R package, Harmony (Korsunsky et al. 2019). Results of cell clustering and cell-type identification were comparable. Nuclei identified as the same cell type in the two models were clustered together in the same cluster (Figure 4.20A).

Analysis of PTCs from the two datasets was also carried out. PTCs from the two datasets were subset and combined using integrate function in the R package, Seurat. Significantly fewer injured and severely injured PTCs in the CKD was noted (Figure 4.20B). UMAP of PTCs also addresses the relationship of new PT clusters identified in the two studies, where the proliferative PT in my data could be mapped to the repairing PT in the AKI model, and the new-PT 1-3 could be mapped to the failed repair PT.

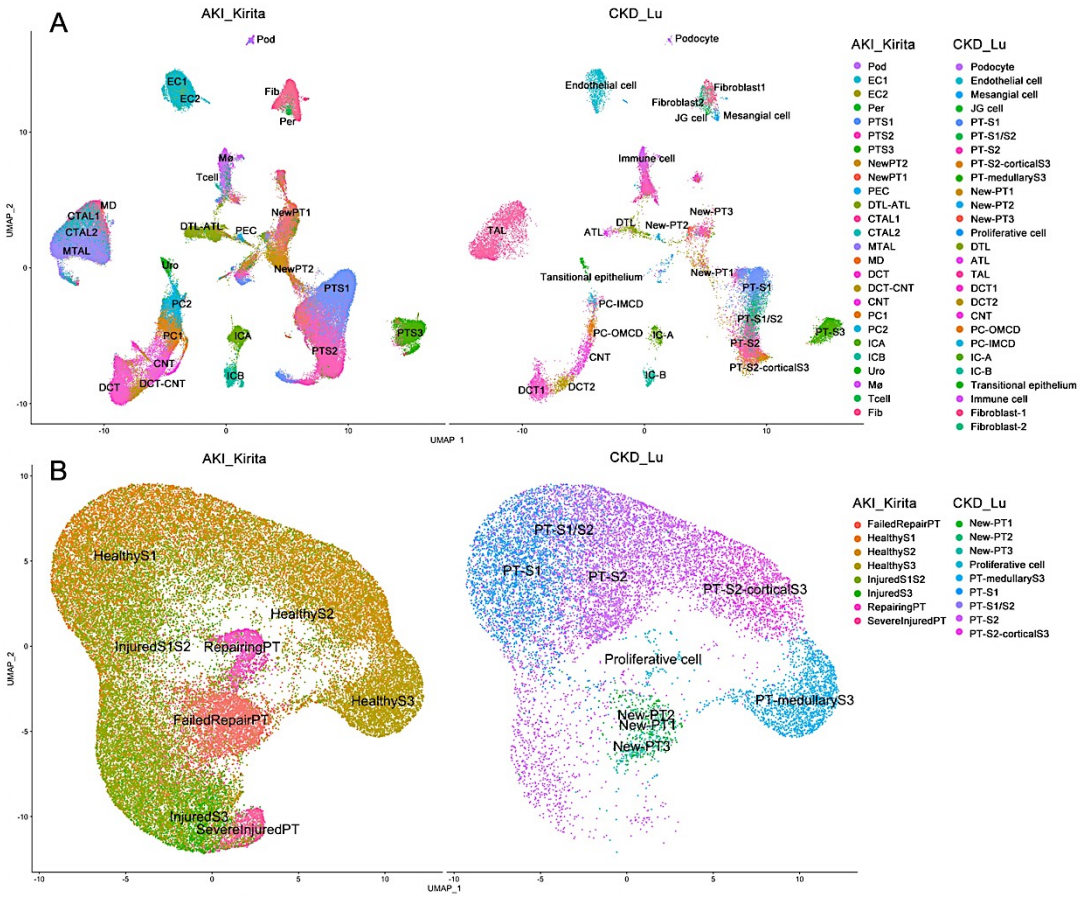


Figure 4.20 Clustering result of combined single-nucleus RNA sequencing datasets of the murine ischemic-reperfusion injury-induced AKI published by Kirita et el. and the AAN induced CKD. (A) UMAP of all cell types split by the nuclear origin shows comparable results of cell clustering and cell-type identification; (B) UMAP of PTCs shows the relationship of new identified PT clusters in two studies.

4.3 Discussion

Here I carried out and repeated the snRNA-seq of chronic AAN using 4 fibrotic and 4 healthy kidneys in 2 experiments. I have delineated PTC subtypes found in the normal mouse kidney, and further describe novel PTC phenotypes associated with kidney fibrosis. The experiments characterised the cellular composition of the adult mouse kidney and compared healthy animals to those with significant fibrosis caused by aristolochic acid-induced toxic proximal tubule injury. There was no obvious batch effect between the biological replicates while cell clustering, which indicated that the results were reproducible. A reliable workflow of nuclear isolation protocol, library preparation and bioinformatics analysis has been set up through the experiments.

The experimental approach of snRNA-seq benefitted from refinements introduced by other investigators, demonstrating the benefits of rapidly processing unsorted whole kidney nuclear preparations prior to transcriptomic profiling. I was thus able to delineate major populations of cells that have proved challenging to resolve in previous studies, including mesangial cells, fibroblasts and juxtaglomerular cells. Also, the proliferative cells were identified as distinct lineages by showing marker genes of the endothelial cell, PTC, TAL, and fibroblast, respectively.

PTCs make a predominant contribution to the wet weight of the normal kidney, are highly metabolically active, and play central roles in kidney recovery vs fibrosis following injury. Nine unique classes of PTCs were identified from more than

10000 PTC nuclei in this experiment. The result improved the resolution of PTCs clustering in single-cell analyses from kidney tissue, where existing reports commonly grouped PTCs in a single large cluster, numerically dwarfing all other clusters presented.

Five clusters of PTCs abundant in normal kidney tissue, each mapping to proximal tubule segments on the basis of canonical marker expression. These comprised clusters with expression profiles consistent only with a single tubular segment (S1, S2, medullary S3) and those enriched in genes found across neighbouring segments (S1-2 and S2-cortical S3). These data uncover significant complexity in PTC phenotype and provide an expression map of abundantly expressed genes within the major cell phenotypes at single-cell resolution.

Kidneys from animals treated with recurrent doses of AA to induce renal injury and subsequent fibrosis contained an increased representation of proliferating cells, immune cells and fibroblasts (Huang et al. 2013). Apart from proliferating PTCs, three new PTC clusters more prominent in kidneys undergoing fibrosis were further identified. These new PTCs suggested the persistent existence of senescent cells after injury and the ongoing dedifferentiation-regeneration process of PTCs in the kidney.

Cluster new-PT1 displayed an expression profile intermediate between canonical PTC clusters and clusters new-PT2 and new-PT3, and trajectory analysis further suggested that this cluster may represent PTCs in transition between canonical and these rarer phenotypes. Intriguingly, clusters new-PT2 and new-PT3

demonstrated enriched expression of a panel of genes expressed in proximal tubules following injury (e.g., *Kim1*). Rather than the diffuse expression in PTC suggested by prior bulk analyses, however, our data reveal restricted expression of specific markers by cluster.

Cluster new-PT2 expressed multiple genes associated with tubular regeneration following injury and was labelled dedifferentiating-regenerating PTCs on this basis. Within these regeneration-associated genes, *Ncam1* is an early nephron progenitor marker that is also seen in proximal tubules after injury, and may contribute to the recovery of PTC function (Abbate et al. 1999; Buzhor et al. 2013). *Tnc* protects against kidney injury and promotes tubular regeneration (Chen et al. 2019b). *Tgfbr3* attenuates TGF-beta signalling through processes including glycosaminoglycan modifications of the type I and type II TGF-beta receptors (Eickelberg et al. 2002). New-PT2 also demonstrated enriched expression of *Foxd1* and *Wt1*, genes reactivated during tubular regeneration processes.

Cluster new-PT3 exhibited unique enrichment for *Havcr1*, a transcript that is nearly undetectable in normal kidneys but occurs promptly after acute kidney injury. *Havcr1* expression may be upregulated chronically after kidney injury, and its persistent expression leads to renal fibrosis (Humphreys et al. 2013; Kirk 2013). New-PT3 cells also expressed other genes linked to fibrotic responses in tubular cells, namely *Cdh2 (N-Cadherin)*, which is associated with fibroblast growth factor signalling and cell invasiveness, and several genes linked to the SASP, including *Cdkn1a (P21), Cdkn2b (P15), Tp53, Tgfb1, Serpine1 (Pai1), Ccl2 (Mcp1), Cxc1,* and

Ccn2. New-PT3 were accordingly labelled dedifferentiated-senescent PTCs. Cellular senescence, which can be identified in ageing kidneys and CKD, was a key determinant of renal regenerative capacity and a mediator of post-injury fibrosis and function (Mylonas et al. 2021).

These data identify principal cellular phenotypes existing in the proximal tubule of the kidney. They further uncover PTC clusters with discrete, fibrosis-associated phenotypes delineated by unique expression profiles of disease-associated markers. Absence of spatial information of cells was a major limitation of the snRNA-Seq experiment. Localisation of the newly identified PTC clusters and their communication with other cell types (e.g., locally recruits inflammatory leukocytes or fibroblasts) needed to be further evaluated.

Chapter 5

Validation of New Identified Proximal Tubular Cells

5.1 Introduction

In chapter 4, I performed snRNA-Seq comparing kidneys from healthy mice to fibrotic kidneys caused by AA-induced toxic proximal tubule injury. Kidneys from animals treated with recurrent doses of AA had renal injury and subsequent fibrosis. The fibrotic kidney had an increased representation of proliferating cells, immune cells and fibroblasts. Cellular composition of healthy adult mouse kidney and kidney with AAN were characterised. The experimental approach benefitted from refinements introduced by other investigators, demonstrating the benefits of rapidly processing unsorted whole kidney nuclear preparations prior to transcriptomic profiling. Thus, the results delineated major populations of cells that have proved challenging to resolve in previous studies, including mesangial cells, fibroblasts and juxtaglomerular cells. Furthermore, I identified proliferating cells of distinct lineages and characterised PTCs in detail.

PTCs make a predominant contribution to the wet weight of healthy kidney, are highly metabolically active, and play central roles in kidney recovery versus fibrosis following injury. In the mouse AAN snRNA-Seq experiment, PTCs were classified into five clusters that were abundant in healthy kidney tissue, mapping to specific proximal tubule segments on the basis of canonical marker expression (Figure 4.8). These comprised clusters with expression profiles consistent only with a single tubular segment (S1, S2, medullary S3), as well as those enriched in genes found across neighboring segments (S1-2 and S2-cortical S3). These results uncover complexity in PTC phenotype, and provide an expression map of

abundantly expressed genes within the major cell phenotypes at single-cell resolution.

Apart from PTC S1-S3, four new PTC clusters were described and named as New-PT1-3 and proliferative PT. These new PT clusters were more prominent in kidney undergoing fibrosis. Cells in these clusters did not express canonical markers of PTC e.g., *Slc34a1 and Lrp2.* These clusters were identified as PTCs in bioinformatics analysis because of the following reasons. First, their location on the UMAP adjacent to canonical PTC S1-S3. As UMAP preserves global structure of the dataset and creates meaningful separation for clusters, relative location and distance on the UMAP of each cluster can be used to infer their phenotypic nature. Second, some of New-PT clusters still preserved marker genes of canonical PT segment, e.g., New-PT1. New-PT1 displayed an expression profile intermediate between canonical PTC clusters and clusters New-PT2 and New-PT3. The trajectory analysis further suggested that this cluster may represent PTCs in transition between canonical and these rarer phenotypes. Third, these New-PT clusters demonstrated enriched expression of a panel of genes expressed in proximal tubules following injury. Many of the marker genes have been described in injured PTCs in the literature that helps to link these cells to AKI and CKD.

10X single-cell RNA sequencing platform, which uses droplet-based method, provides results of a large number of genes in large number of cells. Validation of scRNA-Seq or snRNA-Seq result is an important but sometimes difficult task. Validation of the presence of marker genes at the level of RNA (e.g., RNA flow, *in*

situ hybridization, fluorescent in situ hybridization, real-time reverse transcription polymerase chain reaction…) or protein (e.g., IHC, IF, flow cytometry, western blot…) is important when discovering new cell types.

The key limitation of the 10x single-cell platform and other scRNA-Seq is the lack of spatial information. Although the new-PT clusters had unique gene expression profiles included injury markers seen in PT injury and fibrosis, this does not confirm that they were PTCs. Also, the location and the distribution of the new PT cells would be of great interest that helped to understand tubular injury and interaction of injured PTCs with other cell types.

The experimental aim of this chapter was microscopic validation for the New-PT clusters and to provide spatial information of cell and their phenotype straightforwardly. Selection of marker genes for microscopic validation was based on the result of DEG analysis and reviewed using feature plots.

5.2 Result

5.2.1 Gene selection for microscopic validation

5.2.1.1 Canonical PTCs (PTC segment 1-3)

Co-expression of canonical markers of canonical PTCs and New-PT clusters was evaluated, in order to confirm that clusters labelled "New-PT" were really tubular. Ideally, a gene found in all of the five canonical PTC clusters with high expression level and that was not detected in other cell type would be the best marker for validation. Eight genes were cconsidered as candidates for miccroscopic validation for canonical PTCs, including three pan-PTC marker genes from Figure 4.12 (*Slc34a1, Lrp2, Slcc4a4*) and five genes selected from DEG analysis that compared canonical PTCs to all other cell types (*B3gat2, Grip1, Mylk, Slcc1a1, Slco3a1*). Expression level of the target genes and their distribution on UMAP were evaluated using feature plots.

Slc34a1 was highly expressed in cortical PTCs (segment S1, S1-2, S2, S2-cortical S3) but it was nearly undetectable in medullary S3 (Figure 5.3A). *Lrp2* and *Slc4a4* were strongly expressed in all types of PTCs and showed low expression in other cell-types (Figure 5.3B, 3C). *Slc4a4* was also expressed in the New-PT clusters, which made it a good marker for both canonical PTCs and newly identified PTC phenotypes. *B3gat2* was specific to PTCs but it had lower percentage of expression among PTCs and lower level of expression than *Lrp2* and *Slc4a4* (Figure 5.3D). *Grip1* was a gene of interest because it was widely expressed in all New-PT clusters (Figure 5.3E). Although *Grip1* could be detected in descending thin limb,

principal cells and transitional epithelium, these cell types located at renal medullary region, indicating the cortical expression of *Grip1* might be equivalent to canonical PTC and New-PT clusters. *Mylk* expressed in many types of cells including PTCs, thick ascending tubule, fibroblast and mesangial cell (Figure 5.3F). Like *B3gat2, Slc1a1* was specific to PTCs but with lower percentage of expression among PTCs and lower level of expression than *Lrp2* and *Slc4a4* (Figure 5.3G). *Slco3a1* was not a gene specific to PTCs (Figure 5.3H). According to gene expression on feature plots, antibodies for LRP2, SLC4A4, and GRIP1 were tested in IHC stain. The one with best performance in IHC stain would be used in IF staining.

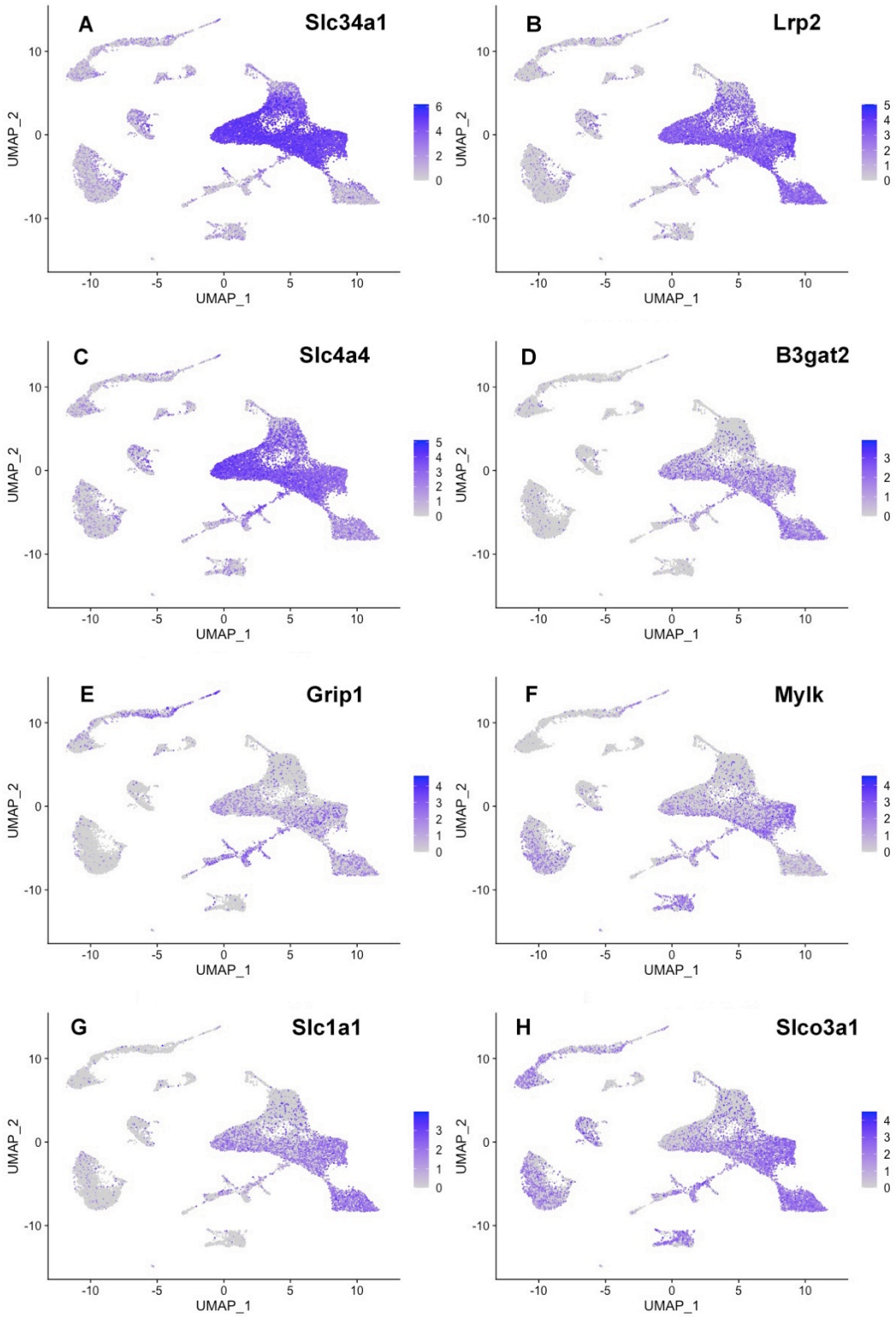


Figure 5.3 Feature plots of candidates for canonical PTC marker genes. Figure shows level of gene expression and distribution of *Slc34a1, Lrp2, Slcc4a4, B3gat2, Grip1, Mylk, Slcc1a1,* and *Slco3a1* on the UMAP plot (A-H).

5.3.1.2 New-PT1

Rather than the diffuse expression in PTC suggested by prior bulk analyses, my data revealed restricted expression of specific markers by cluster, especially the new-PT clusters. New-PT 1-3 were named as “dedifferentiated PTC - intermediate”, “dedifferentiated PTC - regenerating” and “dedifferentiated PTC - senescence”, respectively. New-PT1 was identified as the transitional status between canonical and rarer PTC phenotypes. Many marker genes of New-PT1 overlapped with marker genes of other PTC clusters.

Canonical genes of New-PT1 on Figure 4.12, including *Ankrd6, Vcam1, Pdgfb, Pfgfd, Bmp6 Il34, Itgb6,* and *Itgav*, were selected as candicates for microscopic evaluation. Although these genes were broadly expressed in New-PT1 and had a high expression level, many of them were not specific to New-PT1. *Ankrd6, Pfgfd, Bmp6 Il34, Itgb6,* and *Itgav* also expressed in canonical PTCs, other new-PT clusters and other cell types (Figure 5.4 A, D-H). Since cell number of new-PT1 was much lesser than canonical PTCs, it was important to have a specific marker for New-PT1 for a reliable validation.

Both *Vcam1* and *Pdgfb* were specific to New-PT1 (Figure 5.4B, 4C). *Vcam1* was one of the top ranked marker genes of New-PT1 ($p < 0.001$). In DEG analysis comparing New-PT1 to all other clusters, *Vcam1* had an average log-fold change 1.86. It was detected in 50.9% of New-PT1 cells and only prescence in 2.7% of other cell types. *Vcam1* encodes vascular cell adhesion molecule-1 (VCAM1), which was reported to be present on some (but not all) parietal epithelial cells

lining Bowman's capsule in normal kidney and PTC in human biopsy samples with nephritis (Seron et al. 1991). *Pdgfb,* another specific marker to New-PT1, was detected in 45.2% of New-PT1 and presence in 6.3% of other cell types with an average log fold change 0.90, $p < 0.001$. Therefore, VCAM1 was the prefer marker for microscopic validation for New-PT1.

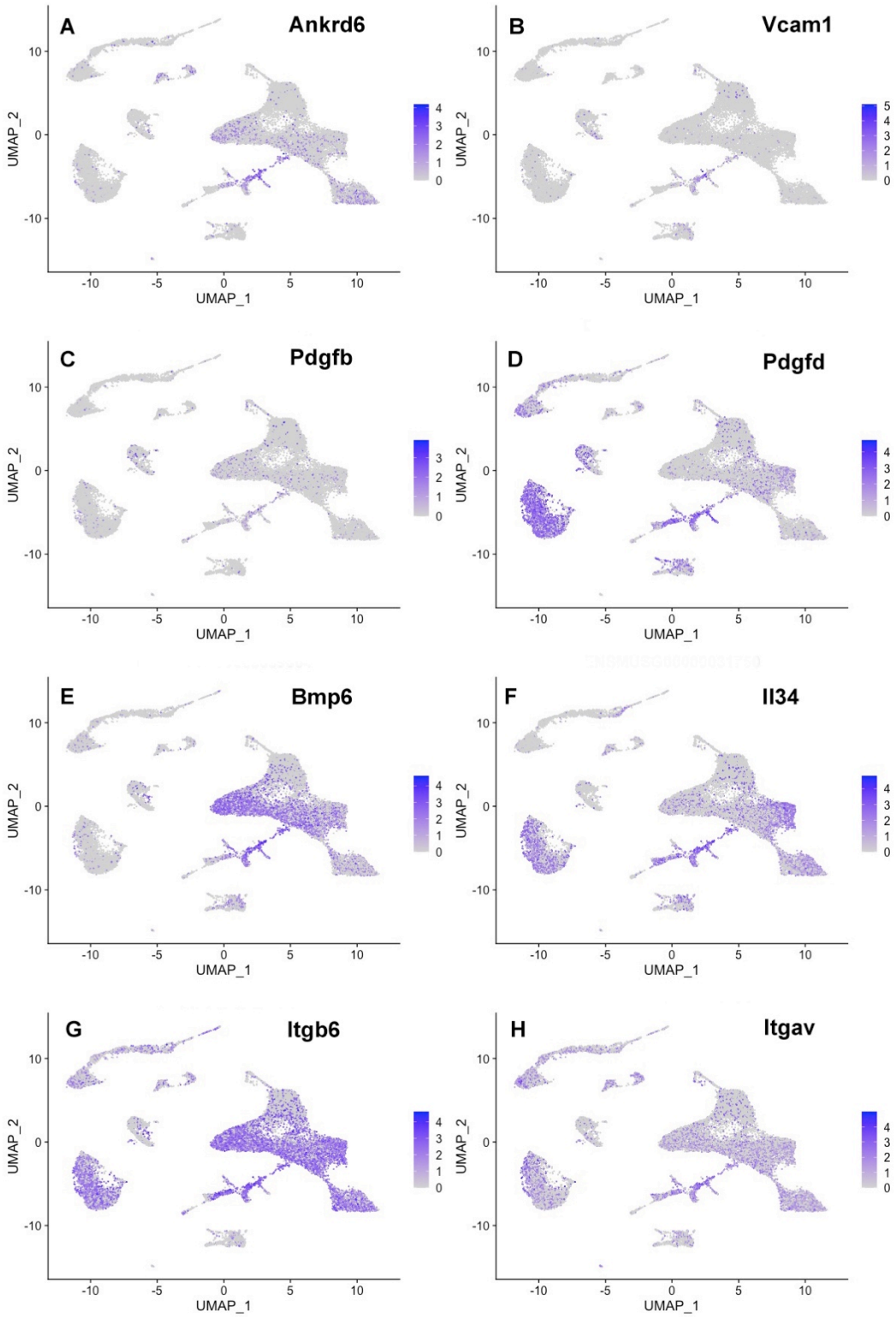


Figure 5.4 Feature plots of candidates for New-PT1 marker genes. Figure shows level of gene expression and distribution of *Ankrd6, Vcam1, Pdgfb, Pdgfd, Bmp6 Il34, Itgb6, and Itgav* on the UMAP plot (A-H).

5.3.1.3 New-PT2

Percentage of New-PT2 cells in kidney increased following kidneuy injury, where New-PT2 comprised 0.12% and 0.23% of healthy kidney and AAN kidney, respectively. Cluster New-PT2 expressed multiple genes associated with tubular regeneration and was labelled dedifferentiating-regenerating PTCs on this basis. Within these regeneration-associated genes, *Ncam1* is an early nephron progenitor marker that is also seen in proximal tubules after injury, and may contribute to recovery of PTC function (Abbate et al. 1999; Buzhor et al. 2013). *Tnc* protects against kidney injury and promotes tubular regeneration (Chen et al. 2019b). *Tgfbr3* attenuates TGF-beta signalling through processes including glycosaminoglycan modifications of the type I and type II TGF-beta receptors (Eickelberg et al. 2002). New-PT2 also demonstrated enriched expression of *Foxd1* and *Wt1*, genes reactivated during tubular regeneration processes. Therefore, protein expression of these markers was evaluated.

Figure 5.5 shows expression of marker genes of New-PT2: *Akap12, Ncam1, Tnc, Tgfbr3, Cp, Foxd1, Wt1 and Chd6,* on feature plots. *Akap12* was highly specific to New-PT2 but it was also a marker gene of New-PT3. Since New-PT3 cells were only identified in AAN kidney, *Akap12* was still a good marker for New-PT2 in healthy kidney, and its co-expression with other New-PT2 marker genes would be helpful to identity New-PT2. *Ncam1, Tnc* and *Wt1* were specific to New-PT2 among tubular cells and selected for microscopic validation although these genes were also detecetd in fibroblast or mesangial cell. *Foxd1* only expressed in a small

amount of cells but it was highly specific to New-PT2 and considered as a marker for validation.

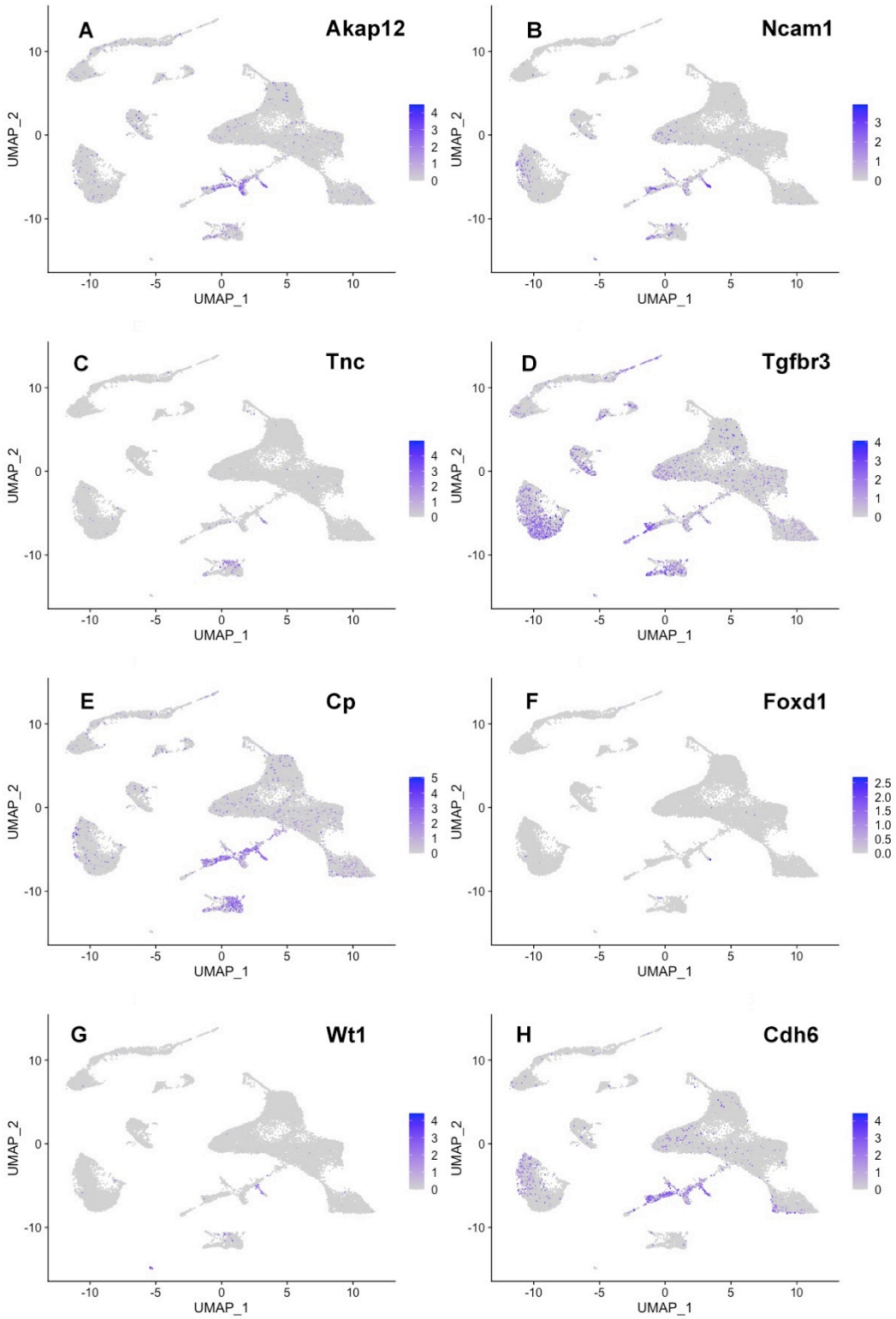
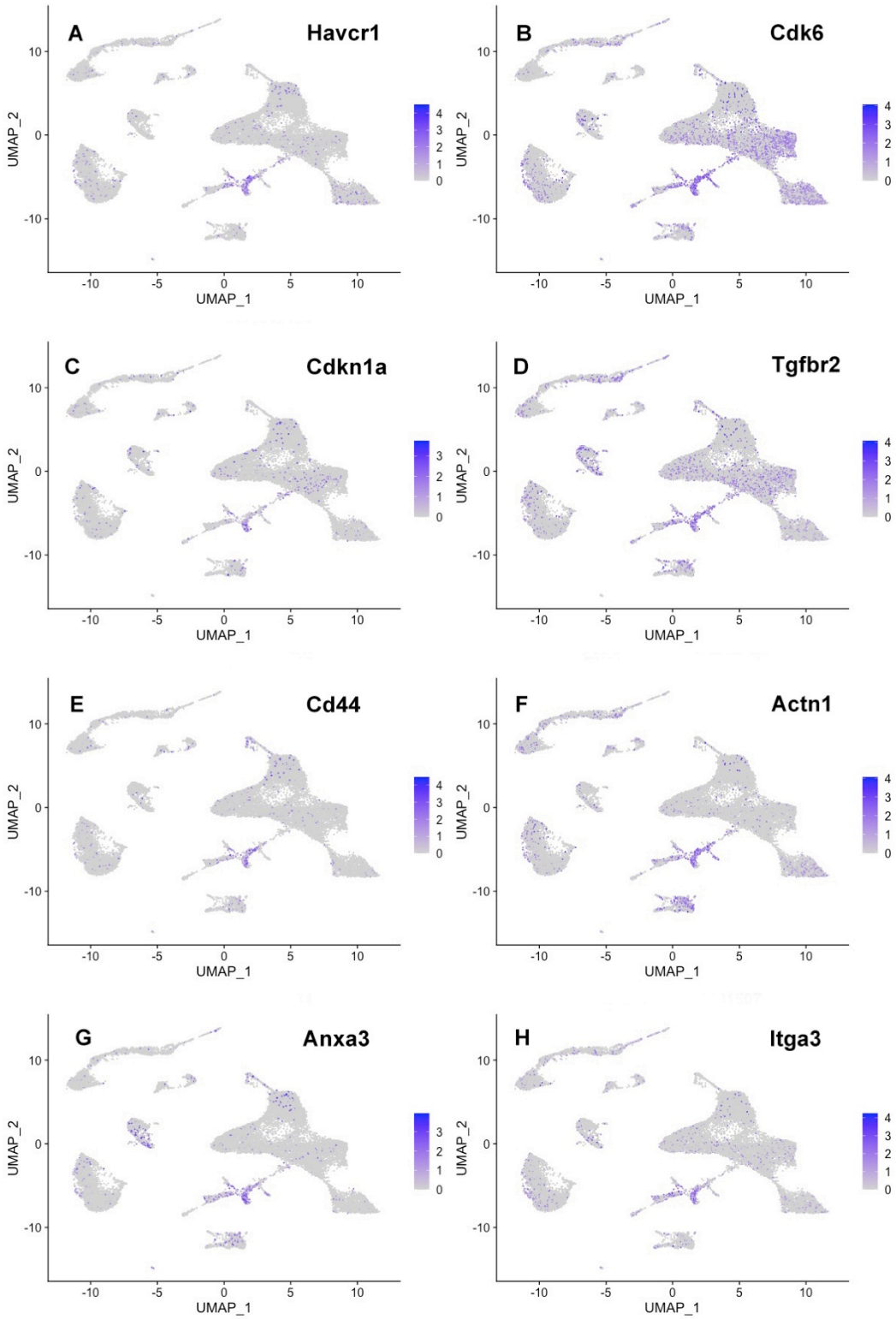


Figure 5.5 Feature plots of candidates for New-PT2 marker genes. Figure shows level of gene expression and distribution of *Akap12, Ncam1, Tnc, Tgfbr3, Cp, Foxd1, Wt1 and Chd6* on the UMAP plot (A-H).

5.3.1.4 New-PT3

New-PT3, named as “dedifferentiated PTC - senescence”, was only present in AAN kidney and expressed genes associated with maladaptation after kidney injury: *Havcr1, Cdk6, Cdkn1a, Tgfbr2, Cd44, Actn1, Anaxa3, Itga3, Mdm2, Egfr, and Cdh2.* Figure 5.6 shows feature plots of the New-PT3 marker genes. Cluster New-PT3 exhibited unique enrichment for *Havcr1*, a transcript that is nearly undetectable in healthy kidney but that is expressed promptly after AKI. *Havcr1* expression may be upregulated chronically after kidney injury, and its persistent expression leads to renal fibrosis (Humphreys et al. 2013; Kirk 2013). New-PT3 cells also expressed genes linked to senescence-associated secretory phenotype, that are found in cells exhibiting cell cycle arrest and inflammatory cytokine secretion. *Cdkn1a* encodes P21 protein, a key node of cellular senescence, which is a downstream mediator of P53 activity and associated with cell cycle arrest. Therefore, HAVCR1 and P21 were prefer markers for microscopic validation. *Cd44, Anxa3, Igta3* and *Cdh2* were also specific to New-PT3, whereas *Cdk6, Tgfbr2, Actn1, Mdm2, Egfr* and *Cdh1* could be detected in other cell types.


A
Havcr1
B
Cdk6
C
Cdkn1a
D
Tgfbr2
E
Cd44
F
Actn1
G
Anxa3
H
Itga3
UMAP_1
UMAP_2

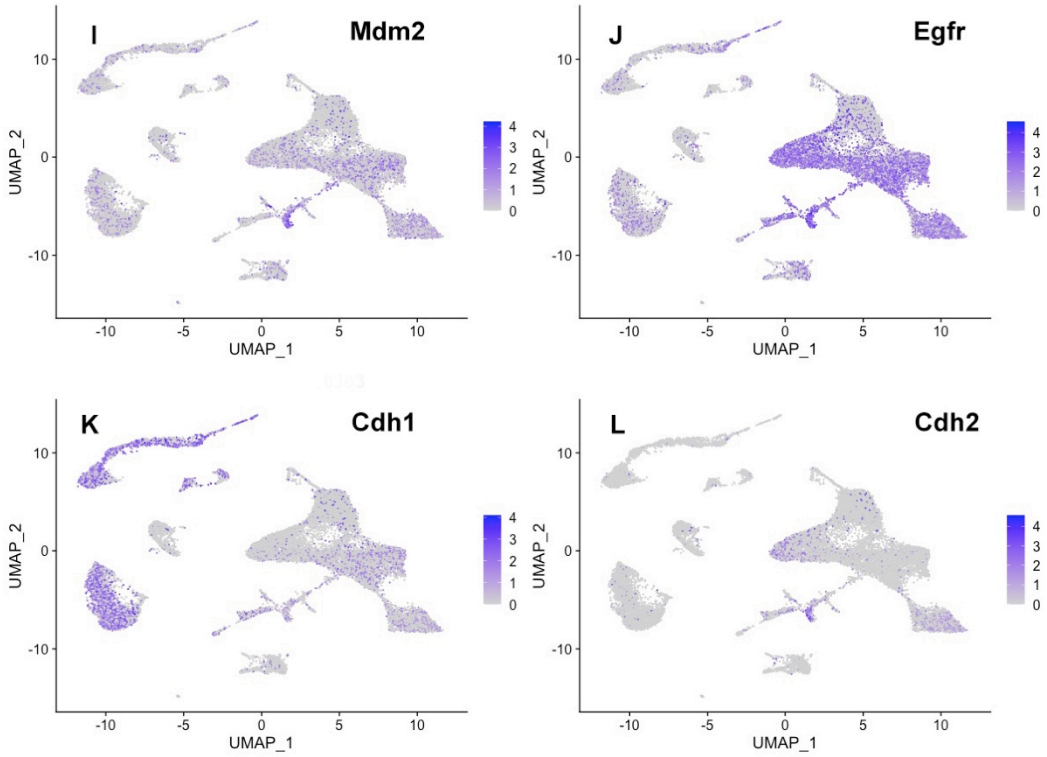


Figure 5.6 Feature plots of candidates for New-PT3 marker genes. Figure shows level of gene expression and distribution of *Havcr1, Cdk6, Cdkn1a, Tgfbr2, Cd44, Actn1, Anaxa3, Itga3, Mdm2, Egfr, Cdh1 and Cdh2* on the UMAP plot (A-L).

5.3.1.5 Proliferative PT

Mki67 encodes nuclear Ki-67 protein, which is a proliferation-associated protein that presents in all active cell cycle phase (Bruno and Darzynkiewicz 1992). *Top2a* encodes DNA topoisomerase 2-alpha, a key node for mitotic phase, that expression level increases in mid-S phase through mitosis and rapidly decreases upon mitotic completion (Lee and Berger 2019). Both *Mki67* and *Top2a* were ideal for microscopic validation.

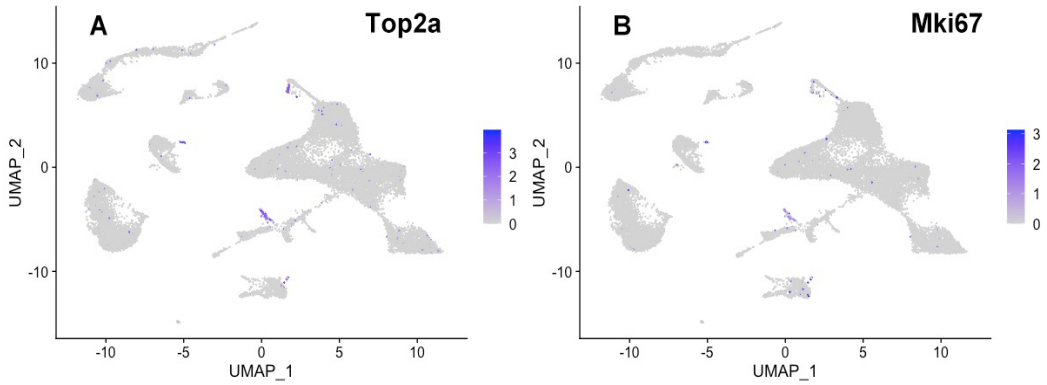


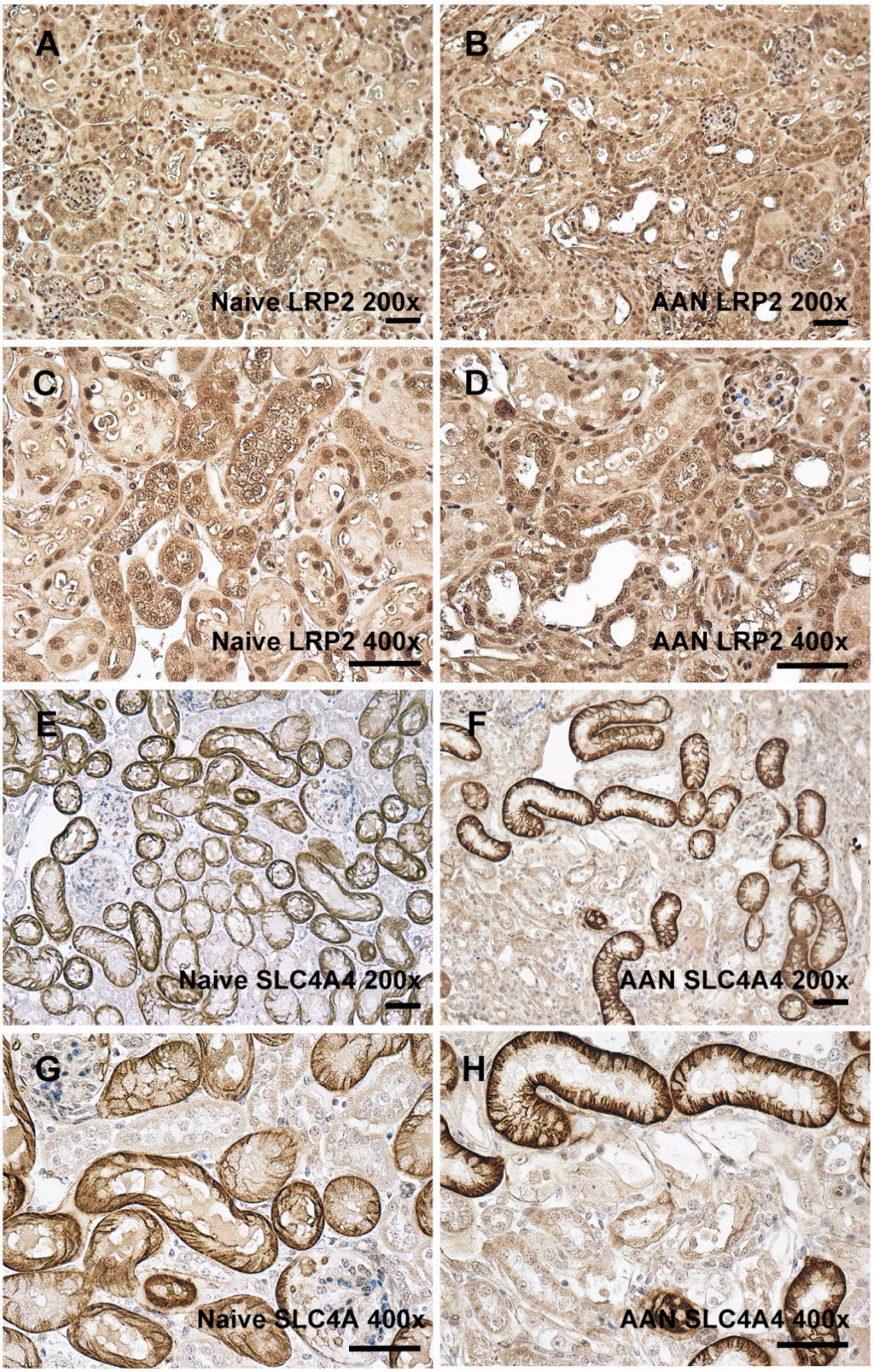
Figure 5.7 Feature plots of candidates for proliferative PT marker genes. Figure shows level of gene expression and distribution of *Top2a* and *MKi67* on the UMAP plot (A and B).

5.3.2 Validation for canonical PTCs

New-PT clusters did not significantly express traditional PTC marker genes such as *Slc34a1*, therefore cell type of New-PT clusters needs to be confirmed by microscopic validation. Primary antibodies for LRP2 (Novusbio NB110-96417, mouse, 1:200 dilution), SLC4A4 (Invitrogen PA5-57344, rabbit, 1:200 dilution) and GRIP1 (Alomone labs APZ-015, rabbit, 1:200 dilution) were tested in IHC stain. Concentrations of the primary antibodies for the test run were decided by their instructions on the manual. Aim of IHC stain was to confirm antibody efficacy on formalin-fixed paraffin-embedded (FFPE) mouse kidney and set a reference for IF stain.

Primary antibody for LRP2 stained PTCs but significant background and nuclear staining were noted in both naïve and AAN kidneys at 1:200 dilution (Figure 5.7A-D). LRP2 encodes megalin, which is an apical membrane endocytic receptor and responsible for protein and middle molecular weight ligand reabsorption. Unspecific binding of anti-LRP2 antibodies might be due to inappropriate antibody concentration, the antigen retrieval process or the mouse to mouse binding due to its mouse host nature. IHC stain of SLC4A4 showed clear staining on basolateral membrane of canonical PTCs in both naïve and AAN kidney with spared injured PTCs in AAN kidney (Figure E-H). SLC4A4 encodes electrogenic sodium bicarbonate cotransporter 1 (NBCe1), which is a basolateral membrane transporter for Na^+ and HCO_3^- and responsible for bicarbonate reabsorption and acid-base maintenance. Moreover, *Slc4a4* broadly expressed on New-PT1 with

high level of gene expression, which could be a good marker of canonical PTCs for IF stain. IHC stain of GRIP1 showed positive staining in normal PTC in health kidney and injured dedifferentiated PTC in AAN kidney (Figure 5.7 I-L). It was interesting that anti-GRIP1 also stained mesangial area and glomerular parietal epithelium. Taking the IHC stains together, SLC4A4 was selected to be the marker of canonical PTCs.


A
Naive LRP2 200x
B
AAN LRP2 200x
C
Naive LRP2 400x
D
AAN LRP2 400x
E
Naive SLC4A4 200x
F
AAN SLC4A4 200x
G
Naive SLC4A 400x
H
AAN SLC4A4 400x

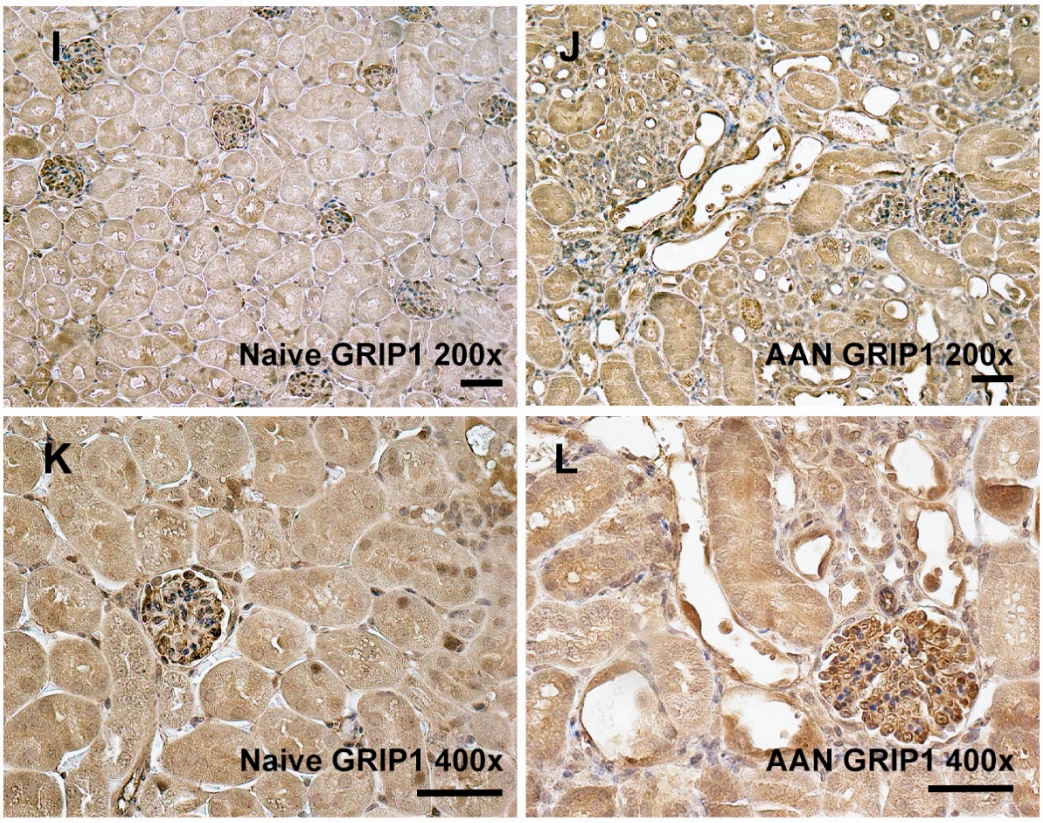


Figure 5.7 IHC stain of canonical PTC markers in the naïve and the AAN kidney. Primary antibody for LRP2 stained PTCs but significant background and nuclear staining (A-D). Anti-SLC4A4 clearly stained basolateral membrane of canonical PTCs in both naïve and AAN kidney with spared injured PTCs in the AAN kidney (E-H). Anti-GRIP1 stained normal PTC in the healthy kidney, injured dedifferentiated PTC in the AAN kidney and glomerular parietal epithelium (I-L). (Scale bar = 50 µm)

5.3.3 Validation for New-PT1

IHC stain of anti-VCAM1 (Invitrogen MA5-11447, mouse, 1:100 dilution) showed occasional PTC cytosomal stain in health kidney and cytosomal stain of injured dedifferentiated PTC in AAN kidney (Figure 5.8). Diffuse nuclear DAB background was also noted.

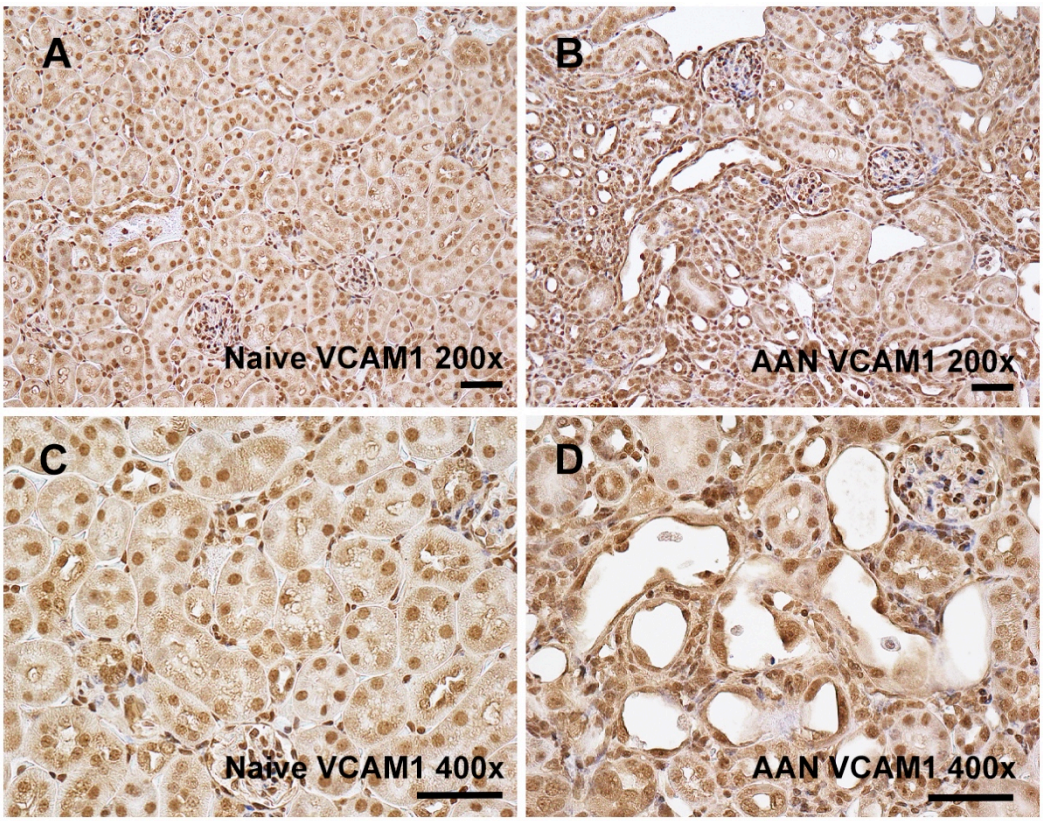


Figure 5.8 IHC stain of New-PT1 marker VCAM1 in the naïve and the AAN kidney. Occasional PTC cytosomal stain in the healthy kidney and cytosomal stain of injured dedifferentiated PTC in the AAN kidney were noted with diffuse nuclear DAB background. (Scale bar = 50 µm)

Phenotype of cells co-expression of SLC4A4 and VCAM1 were confirmed using IF stain and a confocal microscope. Anti-VCAM1 antibody had occasional cytosolic staining (red) of non-injured PTCs with positive staining of mesangial and interstitial cells. Anti-SLC4A4 antibody had basolateral staining of normal proximal tubule (green). The injured PTCs was negative for SLC4A4 staining. Occasional PTC's exhibiting VCAM-1 and SLC4A4 positive staining were identified as New-PT1 and were present in both healthy and AAN kidneys (Figure 5.9).

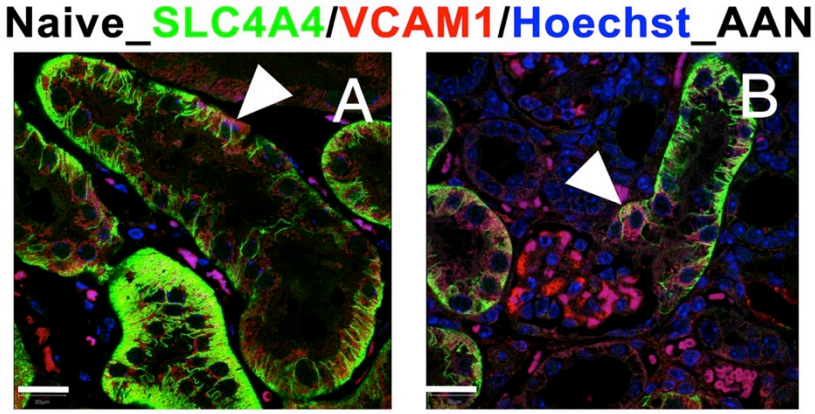


Figure 5.9 IF stain of SLC4A4(green) and VCAM1 (red) in the naïve and the AAN kidney. Occasional PTC's exhibiting VCAM-1 and SLC4A4 positive staining were identified as New-PT1 and were present in the health and the AAN kidney. (Scale bar = 20 µm)

5.3.4 Validation for New-PT2

New-PT2 comprised 0.12% of the healthy kidneys and 0.39% of the AAN kidneys. IHC stain with anti-AKAP12 (Abcam ab49849, mouse, 1:1000) antibody was positive for a small number of PTCs in the healthy kidney and injured dedifferentiated PTCs in the AAN kidney. Positive staining for AKAP12 in parietal epithelial cells was also noted in both AAN and healthy kidneys (Figure 5.10 A-D). Anti-NCAM1 (Abcam ab220360, rabbit, 1:200) antibody stained the basolateral membrane of a few PTCs in the healthy kidney and injured PTCs in the AAN kidney (Figure 5.10 E-H). Positive parietal epithelial staining for NCAM1 was also noted in both healthy and AAN kidneys. IHC stain with anti-WT1 (Sigma MAB4234, mouse, 1:800) showed strong nuclear staining in on podocytes and weaker nuclear staining in occasional PTCs and parietal epithelial cells (Figure 5.11 A-D). WT1 is a canonical marker of podocytes and its expression on PTC or parietal epithelium has been rarely mentioned. IHC stain with anti-TNC (Abcam ab108930, rabbit, 1:800) antibody showed occasional PTC stain in the canonical PTCs and basolateral stain in injured dedifferentiated PTCs (Figure 5.11 E-H). Mesangial and strong interstitial staining in the fibrotic kidney of TNC was also noted, which matched the gene expression profiles on the feature plot. PTCs positive for AKAP12/NCAM1/WT1/TNC were easily to be found in the AAN kidney and were occasionally found in the naïve kidney.

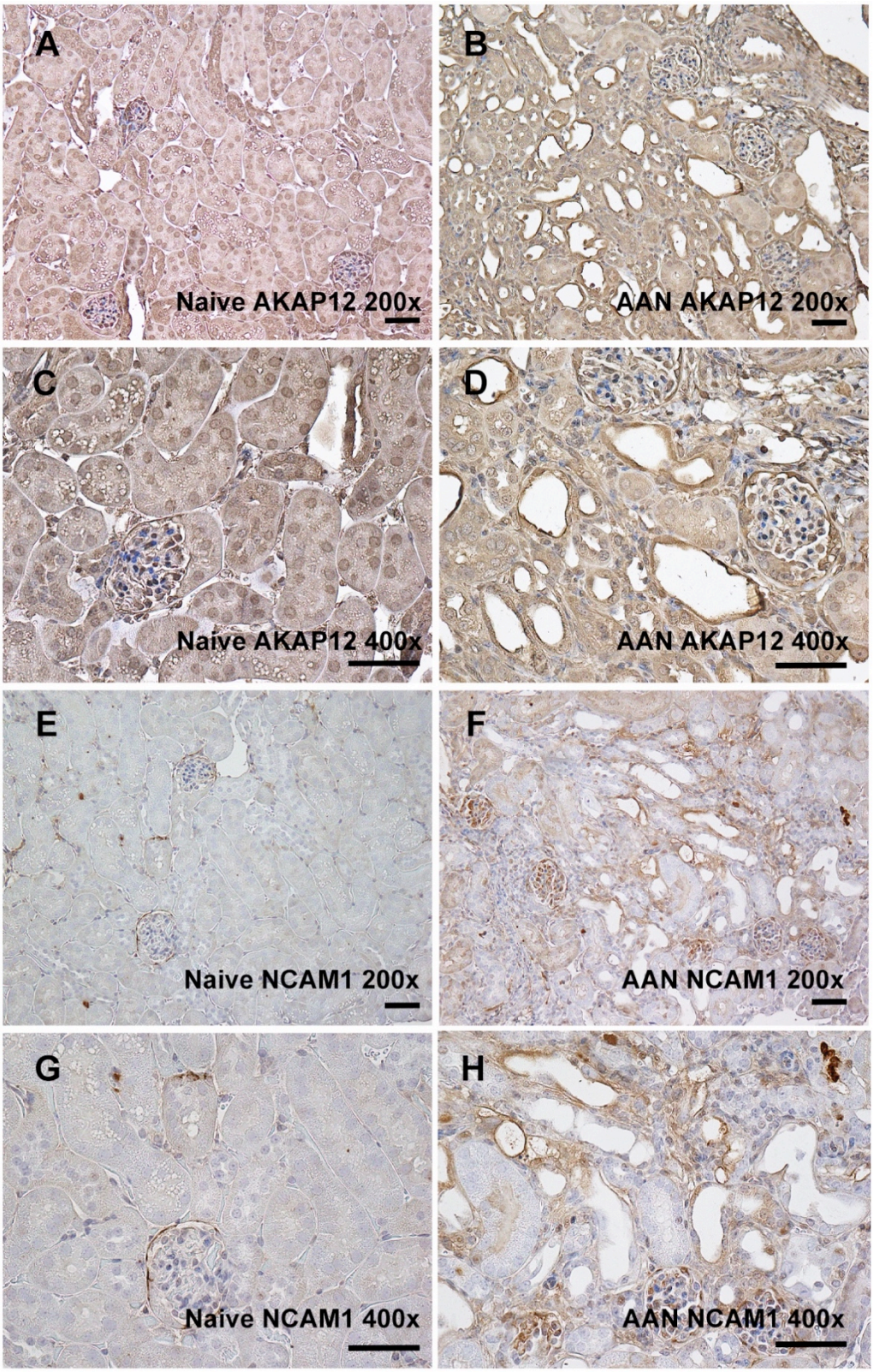


Figure 5.10 IHC stain of New-PT2 markers AKAP12 and NCAM1 in the naïve and the AAN kidney. Positive staining for a small number of PTCs in the healthy kidney, injured dedifferentiated PTCs in the AAN kidney and parietal epithelium were noted. (Scale bar = 50 µm)

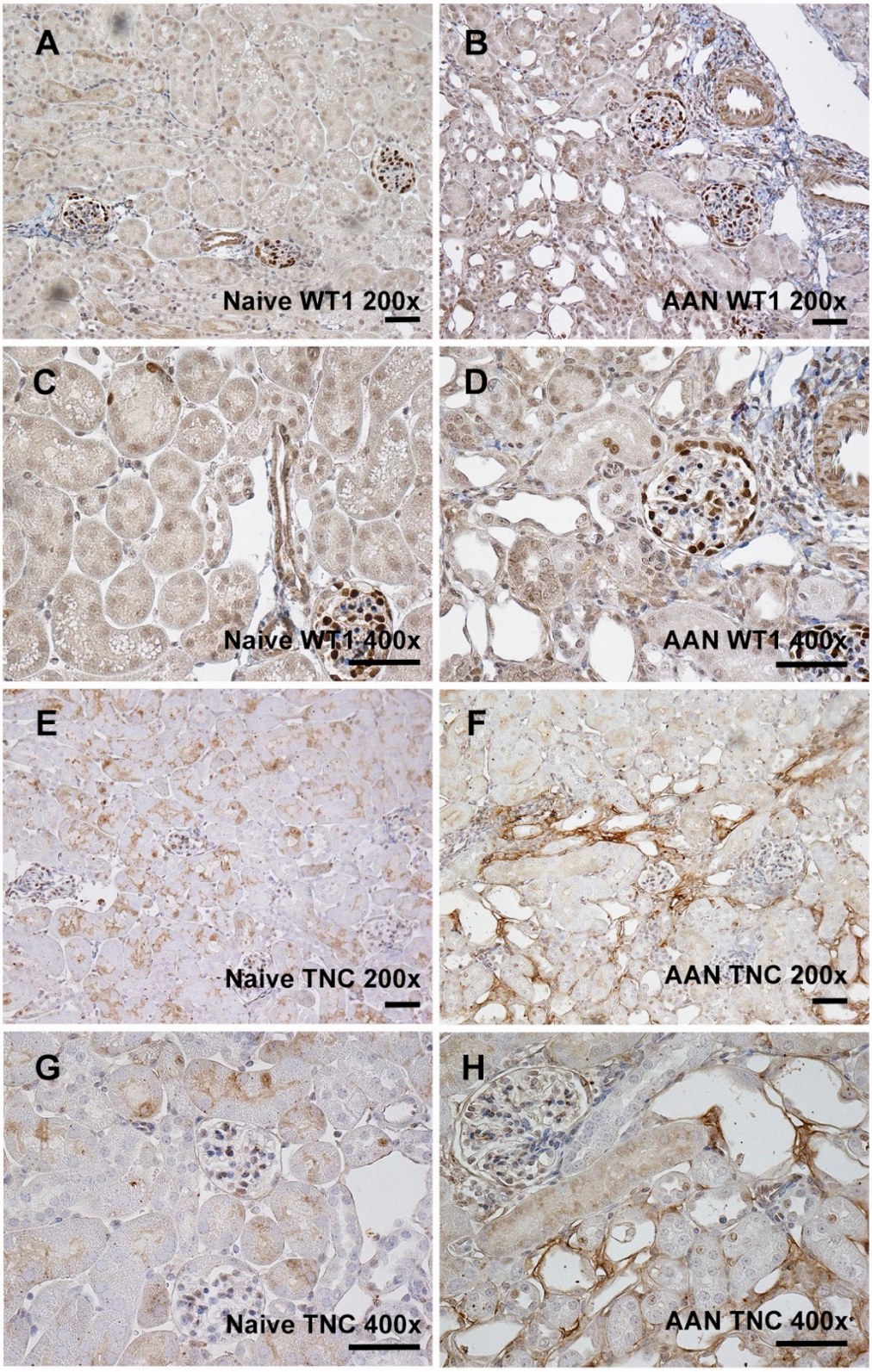


Figure 5.11 IHC stain of New-PT2 markers WT1 and TNC in the naïve and the AAN kidney. IHC stains were positive for a small number of PTCs in the healthy kidney, injured dedifferentiated PTCs in the AAN kidney and parietal epithelium were noted. (Scale bar = 50 µm)

Microscopic evaluation of New-PT2 markers identified not only PTCs but also glomerular parietal epithelial cell staining. Another two markers, FOXD1 and KCNMA1, were also used to validate the New PT-2 cluster. *Kcnma1* was expressed mainly in in medullary S3 and New-PT2, therefore, it was selected as an additional marker for validation (Figure 5.12).

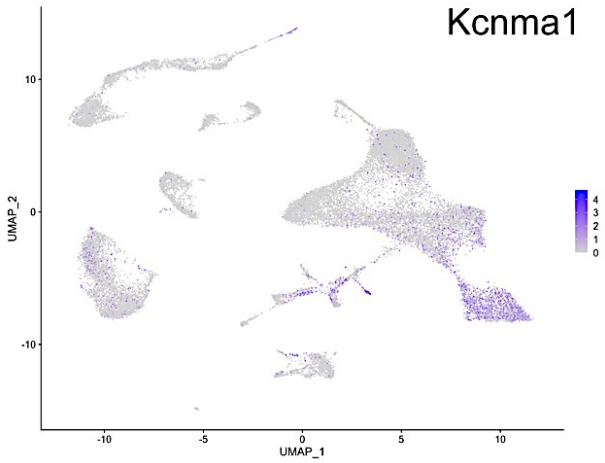


Figure 5.12 Feature plots of *Kcnma1*. *Kcnma1* mainly expressed in in medullary S3 and New-PT2 therefore it was selected as an additional marker for New-PT2 validation.

Results of IHC stain of anti-FOXD1 (LifeSpan BioSciences LS-B9155-LSP, rabbit, 1:50 dilution) are shown in Figure 5.13 A-D. RNA expression of *Foxd1* was only detected in a very small number of cells in the snRNA-Seq experiment. However, IHC staining was positive in normal PTCs, injured PTCs and parietal epithelium, potentially revealing the difference between gene and protein expression. IHC stain with anti-KCNMA1 (Alomone labs APC-107, rabbit, 1:200) showed the same findings as FOXD1 and previous markers, where PTCs of healthy kidney, injured PTCs of AAN kidney and glomerular epithelial cells were stained.

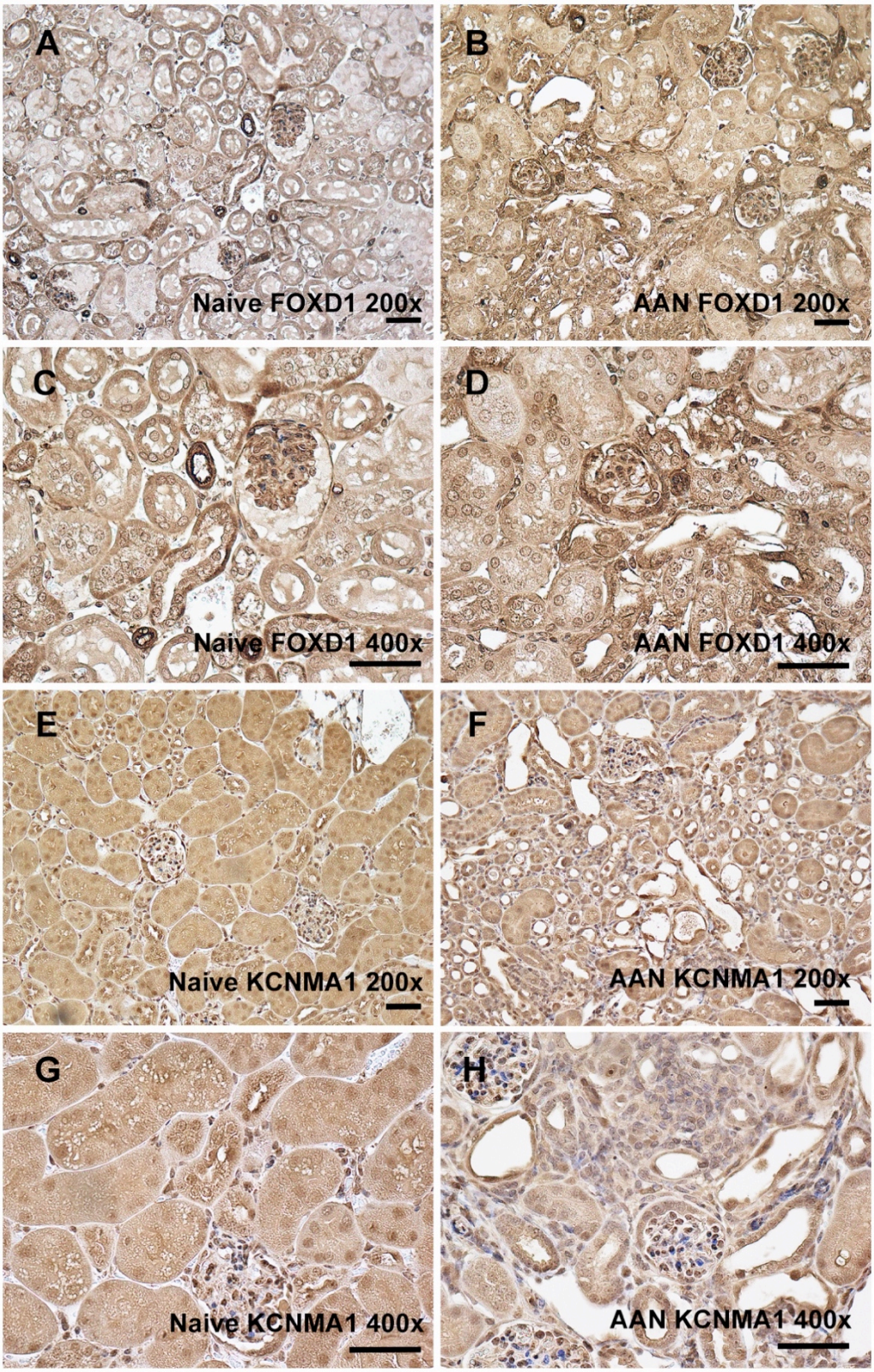
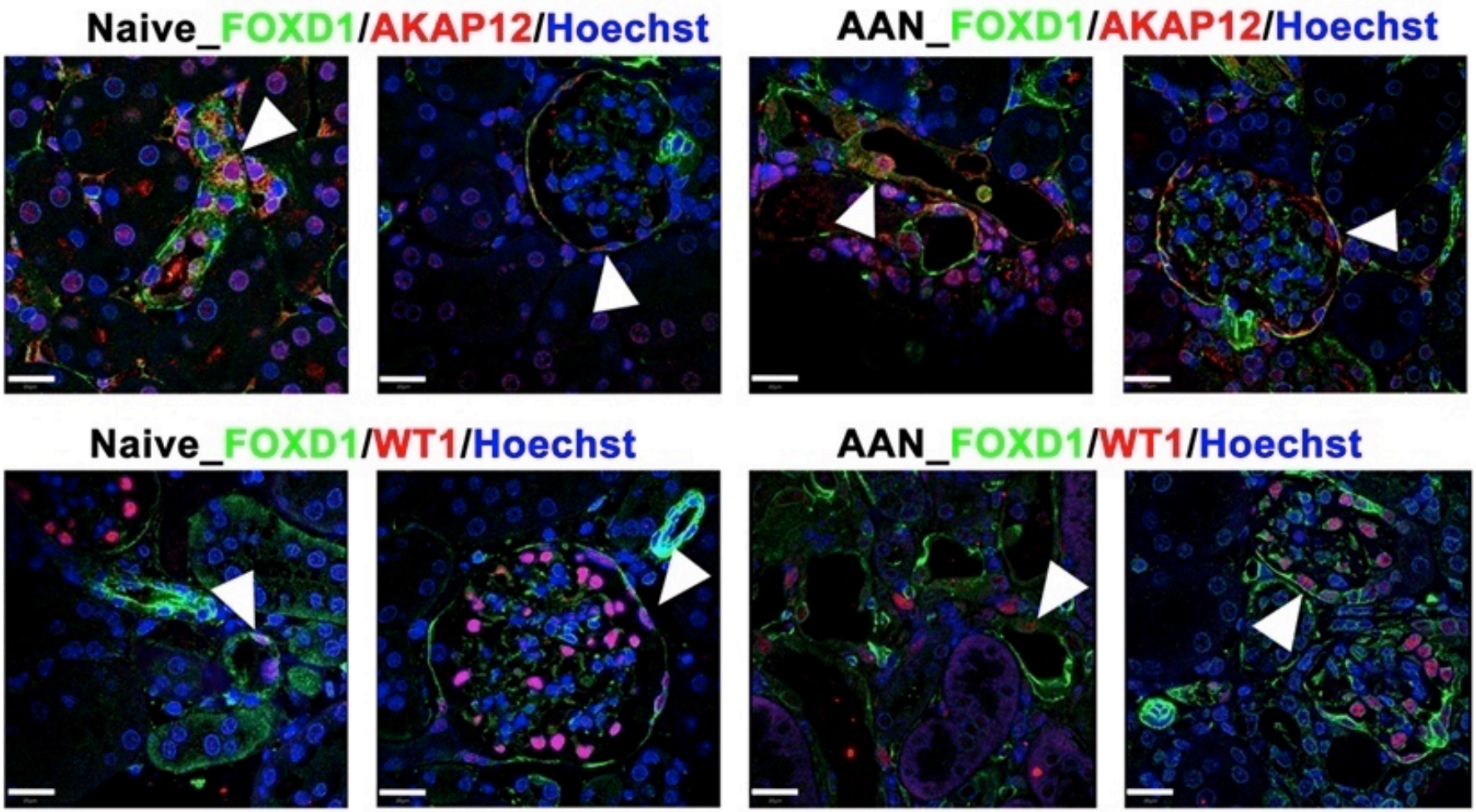


Figure 5.13 IHC stain of New-PT2 markers FOXD1 and KCNMA1 in the naïve and the AAN kidney. Positive staining of a small number of PTCs in the healthy kidney, injured dedifferentiated PTCs in the AAN kidney and parietal epithelium were noted. (Scale bar = 50 µm)

IF stain was used to identify cells with New-PT2 marker co-expression. Merged images from confocal microscopy showed co-expression of FOXD1/AKAP12, FOXD1/WT1, NCAM1/AKAP12, TNC/AKAP12 in the naïve and the AAN kidney. Tubular staining for each marker was evident, with co-expression in rare tubular cells discernible through merged signal in the healthy mouse kidney, and in increased numbers of cells in kidneys from the AAN group (Figure 5.14). Co-expression of markers was also evident through merged signal in parietal epithelial cells, consistent with the presence also of parietal epithelial cells in the New-PT2 cluster. Therefore, the New-PT2 comprised PTC, which significantly increased in number of cells staining positive for New-PT2 markers after kidney injury, together with parietal epithelial cells.


Naive_FOXD1/AKAP12/Hoechst
AAN_FOXD1/AKAP12/Hoechst
Naive_FOXD1/WT1/Hoechst
AAN_FOXD1/WT1/Hoechst

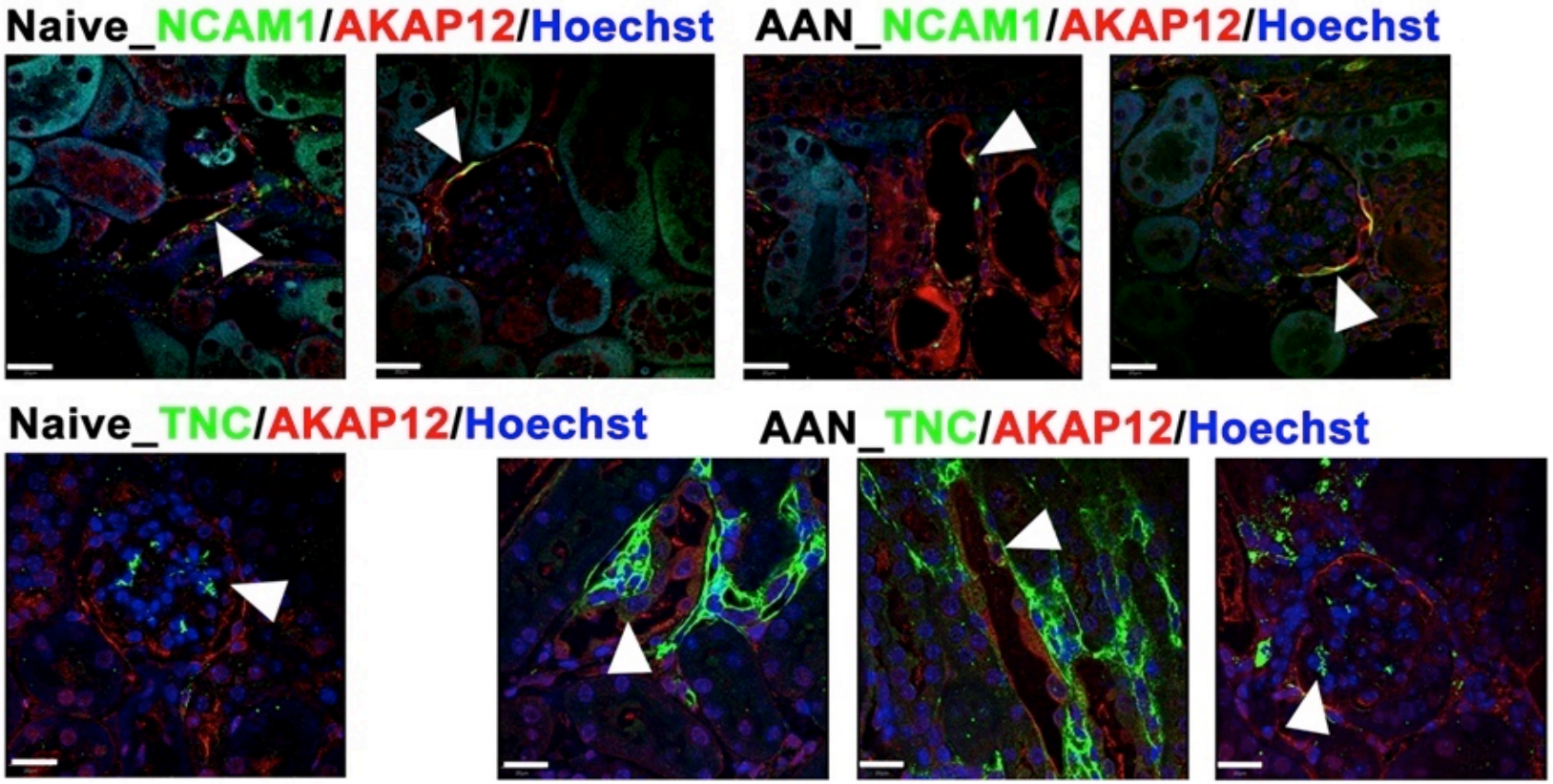


Figure 5.14 IF stain of the New-PT2 markers. Proximal tubular staining of FOXD1/AKAP12, FOXD1/WT1, NCAM1/AKAP12, and TNC/AKAP12 for New-PT2 was identified in both naïve and AAN kidneys. Markers for New-PT2 also stain the parietal epithelial cells. (Scale bar = 20 µm).

5.3.5 Validation for New-PT3

HAVCR1 and P21 were the preferred markers for New-PT3. IHC stain with anti-HAVCR1 (R&D AF3689, goat, 1:800 dilution) showed HAVCR1 was nearly undetectable in healthy kidney but was seen in a proportion of PTCs in the AAN kidneys (Figures 5.15). IF stain showed co-expression of HAVCR1 and P21 (anti-P21, Novusbio NBP2-29463, mouse, 1:200 dilution) in injured dedifferentiated PTCs in the AAN kidney (Figure 5.16)

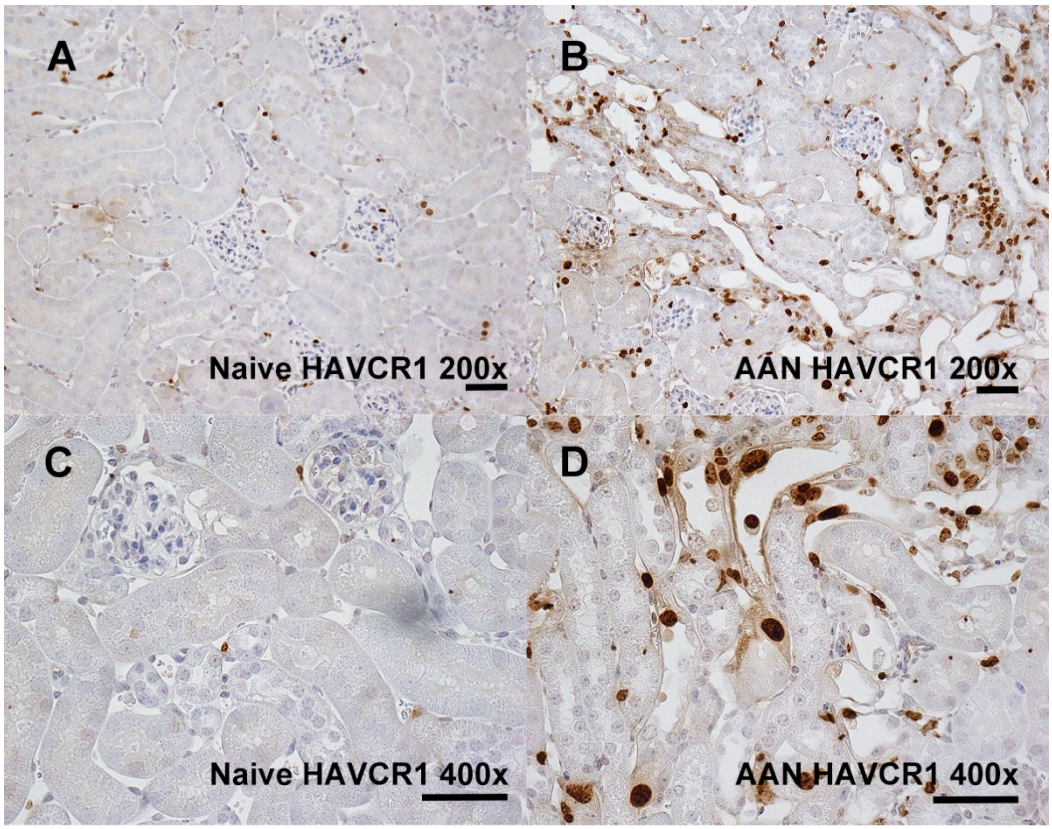


Figure 5.15 IHC stain of New-PT3 markers HAVCR1 in the naïve and the AAN kidney. HAVCR1 was nearly undetectable in health kidney but were seen in a proportion of PTCs in AAN kidneys. (Scale bar = 50 µm)

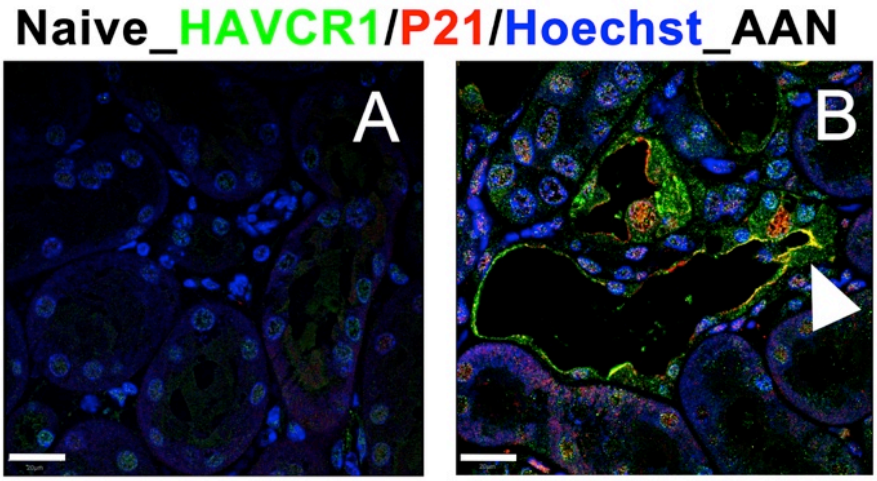


Figure 5.16 IF stain of New-PT3 markers HAVCR1 and P21. Co-expression of HAVCR1/P21 in injured dedifferentiated PTC in the AAN kidney. Notably New-PT3 was only identified in the AAN kidney. (Scale bar = 20 µm).

5.3.6 Validation and quantification for Proliferative PT

IHC stain with anti-Ki67 (Abcam ab15580, rabbit, 1:800 dilution) showed a significant increase in Ki67 positive PTCs (Figure 5.17A). Compared to healthy kidneys, AAN kidneys demonstrated a higher percentage of Ki-67 positive cells (Figure 5.17B).

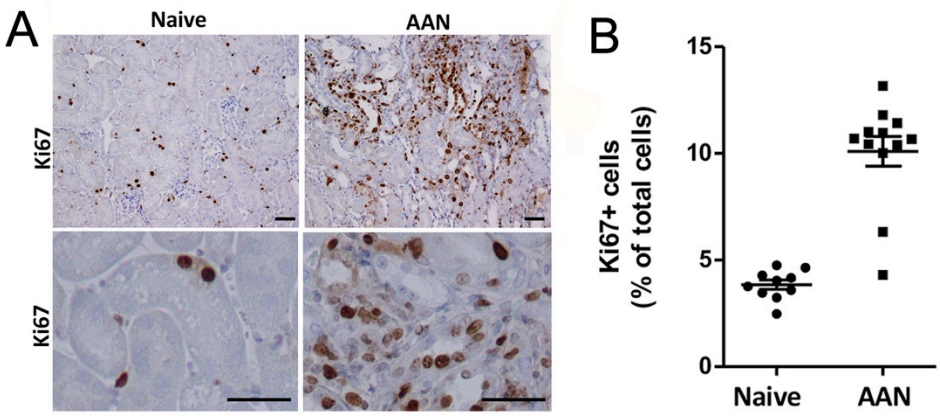


Figure 5.17 IHC stain of proliferative marker, Ki67, in the Naïve and the AAN kidneys. (A) Increased Ki67 positive PTC in the AAN kidney (B) Ki67 DAB signal was used to quantify proliferating cells as a percentage of all hematoxylin stained cells. (Scale bar = 50 µm)

5.4 Discussion

In this chapter, microscopic validation of the new-PT phenotypes identified in the AAN snRNA-Seq experiment was performed. The expression level of the canonical genes of each new-PT cluster were reviewed using feature plots. Co-expression of at least two canonical markers in IF stain was used to confirm the existence of the new-PT 1-3 phenotypes. Antibodies for VCAM1 and SCL4A4 were used for the new-PT1 and HAVCR1 and P21 were used for the new-PT3. For the new-PT2, a combination of antibiotics for FOXD1, AKAP12, WT1, NCAM1 and TNC were used. The new-PT2 markers stained PTCs significantly increased after kidney injury and these markers also stained parietal epithelial cells.

Independent verification for new cell types or cellular differentiation pathways identified in single-cell experiments is important but presents challenges. When working with a solid organ such as kidney or brain, studying the spatial localisation of the cell clusters helps to disclose the relationship between the cellular localisation and its RNA expression, and may also facilitate the interpretation of cellular interactions. The RNA or protein expression on organ sections are usually used to verify new cell types identified in scRNA-Seq/snRNA-Seq, e.g., RNA scope or IHC/IF stain. As an alternative, spatially resolved transcriptomic platforms combine single-cell technology and spatial information in one experiment. A recent publication showed regional-specific gene expression profiles of injured kidneys and the dynamic landscape of cell-cell interactions (Dixon et al. 2022). Resolution is limited in current approaches, but it is likely that

improvements in the cellular resolution and the development of the analytic tools will expand the research application of single-cell spatially resolved technology in the future.

IHC and IF stain both depend on antigen-antibody interactions to label the target protein with chemicals / radioactive elements or fluorescent dyes. IHC and IF may have different characters in validation. Typically, IHC staining preserves more structural information than IF staining, whereas double or triple staining improves the accuracy and the specificity of the IF stain for a unique cell type. Therefore, IF stain was used for new-PT validation.

Isotype control staining of IHC and IF helps to support the true positive of target protein when using a monoclonal primary antibody. Isotype control is a negative control using antibody of the same isotype, clonality, conjugate, host species, and concentration if known as the primary antibody. Performing isotype control for the New-PTs could further validate the presence of these new classes of PTCs in the kidney. Quantification was only performed for proliferative cells. Quantification of the positive dual-stained cells in IF stain could help to validate the percentage of New-PTs in both health and fibrotic kidneys, which should be considered in further single-cell experiments.

The IHC and IF data presented here showed that the New-PT2 markers stained not only PTCs but also the parietal epithelium. The New-PT2 cluster expressed genes that are seen during nephrogenesis, which may imply a PTC phenotype related to tubular regeneration. Activation of the glomerular parietal epithelium

may play a major role in glomerulopathy, especially in rapid rapidly progressing glomerulonephritis and focal and segmental glomerulosclerosis (Moeller and Smeets 2014). Parietal epithelial cells have also been identified as a podocyte progenitor population, that may help podocyte replacement in glomerular diseases (Poulsom and Little 2009; Kaverina et al. 2019). Diversity of parietal epithelial cells has been described in the literature and classified as parietal podocyte, plat parietal epitheliums, intermediate parietal epitheliums and cuboidal parietal epitheliums (D'Agati and Shankland 2019). The cuboidal parietal epithelium connects the Bowmen's capsule to the beginning part of PTC and shares morphological features with PTCs. A subset of parietal epithelial cells expressing de-differentiated PTC markers CD24+CD133+ was identified, and could differentiate to mature PTCs (Sagrinati et al. 2006). It is interesting to speculate that these cuboidal parietal cells might contribute to tubular regeneration or elongation in growing kidneys.

Chapter 6

Single-Nucleus RNA Sequencing of Growing Mouse Kidney

6.1 Introduction

In Chapter 4 and 5, new classes of PTCs were identified using snRNA-Seq and validated by confocal microscopy. The new PTC clusters were found at low abundance in normal kidney and in increased number in kidneys undergoing regeneration/fibrosis following injury. These clusters exhibited clear molecular phenotypes, permitting labeling as, proliferating PTCs, New-PT1 (dedifferentiated - intermediate), New-PT2 (dedifferentiated - regeneration) and (present only following injury) New-PT3 (dedifferentiated -senescence). These rare but distinct PTC clusters exhibited gene expression signatures suggesting roles in renal injury responses and fibrosis progression. They might send fibrotic signalling to fibroblasts and inflammatory signaling to macrophages. It was proposed that these New-PTs had important roles in regulating recovery versus fibrosis after kidney injury.

These previous data led to inferences about PTC regeneration and proliferation pathways. Epithelial regeneration is the key process of tubular recovery after AKI; however, its mechanisms remain unclear. Whether the new epithelial cells are derived from dedifferentiation of the mature tubular epithelium, or from scattered progenitor cells, remains controversial. In a pivotal study focusing on PTC regeneration using lineage-tracing, there was no evidence for PT stem-cells (Kusaba et al. 2014).

Here I further characterised PTC phenotypes during mouse growth, aiming to understand the PTC proliferation and acquisition of a mature PTC phenotype.

Epithelial tubule elongation from differentiated PTCs plays a key role in kidney growth (Kusaba et al. 2014), which may share common mechanisms with tubular regeneration after kidney injury. Mouse kidney development begins at E8.5 with the formation of the nephric duct primordium from the nephrongenic cord (Davidson 2008). The process of ureteric bud branching and nephron formation of embryonic kidney has been studied using scRNA-Seq (Combes et al. 2019). The final wave of nephrogenesis happens between postnatal day 2 (P2) and 4 (P4) in mouse (Little and McMahon 2012). Cessation of nephrogenesis controls the total nephron number.

Although the nephron number is fixed a few days after birth, the kidney size keeps increasing. Along with the increase of body size and weight after birth, cell number and weight of kidney keep increasing, whereas nephron number is nearly fixed (Falconer et al. 1978; Murawski et al. 2010). The mouse shows its highest rate of body and kidney size enlargement from birth till 6 weeks old, and tubular elongation makes a predominant contribution to this kidney growth. Understanding the cellular composition and the RNA expression profiles of kidney in the growth stages could disclose the tubular elongation process. For this purpose, snRNA-Seq using 2-week-old and 4-week-old health mouse kidney was performed.

Only male mice were used in the previous AAN experiment. In a single-cell RNA sequencing experiment using adult healthy mouse kidneys, differences of gene expression profiles between sexes were noted, especially in the proximal tubule

where gender-specific PTC clusters were identified (Ransick et al. 2019). Male and female had different susceptibility to kidney injury, which could be linked to sex-specific gene expression of PTCs in AKI (Vinas et al. 2020). Therefore, I included both male and female mouse in the snRNA-seq experiment of growing mouse kidney.

The approach taken to data integration will significantly affect the result of cell clustering in a single-cell experiment. Since two different variables, age and gender, were included, selecting an appropriate method of data combination and batch effect correction was critical for proper analysis. Here I introduced another batch correction method, Harmony (Korsunsky et al. 2019). Harmony groups cells into multi-dataset clusters based on PCA embedding, then runs an iterative algorithm to adjust for dataset-specific effects (Figure 6.2).

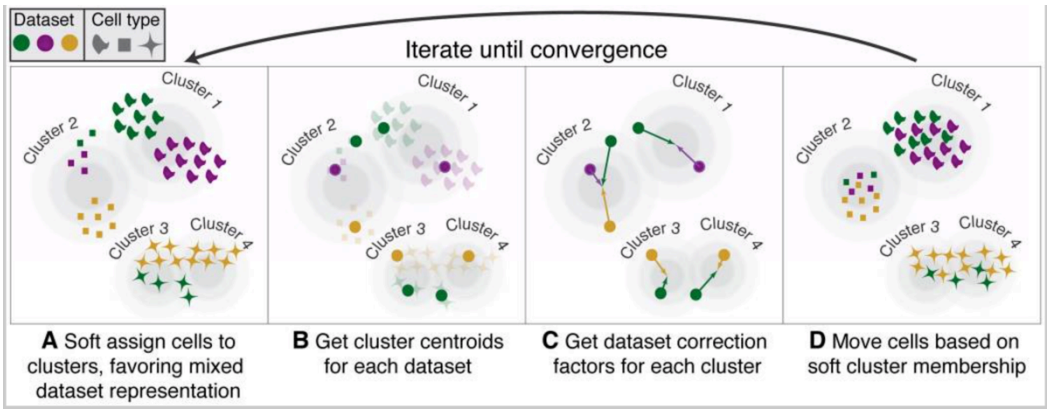


Figure 6.2 Harmony method of dataset integration (Korsunsky et al. 2019). The algorithm shows how Harmony works for dataset correction and to build up the adjusted PCA embedding.

This chapter shows the results of snRNA-seq using 2- and 4-week old healthy mouse kidneys from both sexes. In the chapter 4, I have demonstrated PTC phenotypes in healthy adult kidney and changes in response to injury (AA). These previous data led to inferences about PTC regeneration/proliferation pathways. Here I further characterised PTC phenotypes during mouse growth.

The hypothesis was that distinct PTC phenotypes represent a key node in both kidney growth and tubular regeneration. Investigation of PTC proliferation and epithelial tubular elongation at single-cell level helps to understand the mechanisms of PTC differentiation and renal recovery after kidney injury. The aim of this chapter was to understand the cellular composition and the gene expression profiles of growing mouse kidneys using snRNA-seq, focusing on PTC proliferation and sex differences.

6.2 Result

6.2.1 Optimisation of nuclear isolation protocol for growing mouse kidney

Kidney from young mouse was smaller and fragile. The nuclear isolation protocol was reviewed and optimized, aiming to get the best yield of nuclei and RNA quality from growing mouse kidneys.

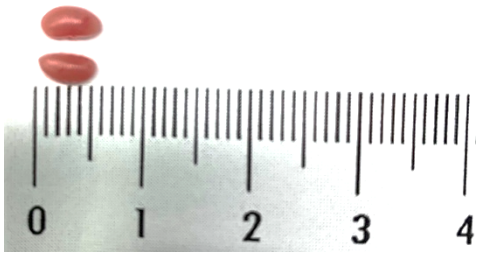
The experiments used C57BL/6J mice bred in the Biological Services facility, Heath Campus of Cardiff University, with standard care of housing and husbandry. Mice were euthanized using Schedule 1. Both kidneys were harvested, preserved in chilled PBS and processed.

6.2.1.1 Body weight and size of kidney of mouse from different ages

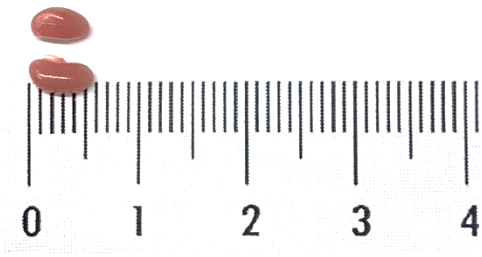
To understand the difference of body weight and kidney size of mouse from different ages, body weight, weight of kidney and length of kidney from 2, 4, 12 weeks old healthy C57BL/6J mice were measured (Figure 6.3). With body weight increasing, weight and length of mouse kidney also increased. The weight of kidney of 2, 4, 12 week old mouse were 0.035 g (2 week male), 0.03 g (2 week female), 0.10 g (4 week male), 0.10 g (4 week female), 0.23 g (12 week male), 0.125 g (12 week female). The length of kidney of 2, 4, 12 week old mouse were 0.5 cm (2 week male), 0.6 cm (2 week female), 0.8 cm (4 week male), 0.7 cm (4 week female), 1.0 cm (12 week male), 0.8 cm (12 week female).

In the snRNA-seq experiments of AAN in chapter 4, each sample used a quarter of male mouse kidney. To have a result that could represent the whole kidney

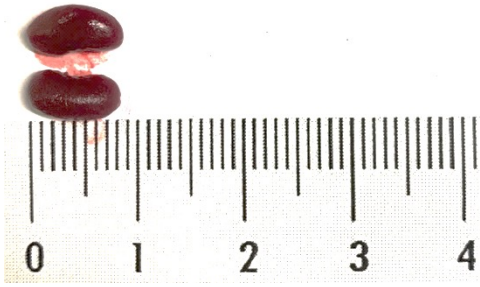
with a nuclei yield similar to the previous experiments, a kidney from 2 weeks old mouse and half kidney from 4 weeks old mouse were used for nuclear isolation in the growing mouse kidney snRNA-seq experiment.



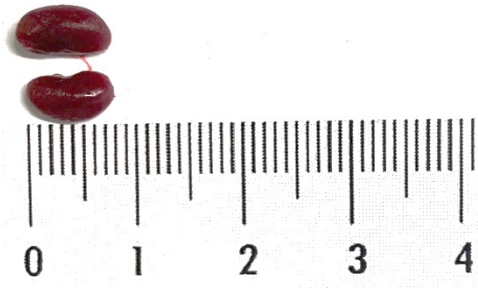
2wk male; BW 5.01 g; KW 0.035 g



2wk female; BW 5.93 g; KW 0.03 g

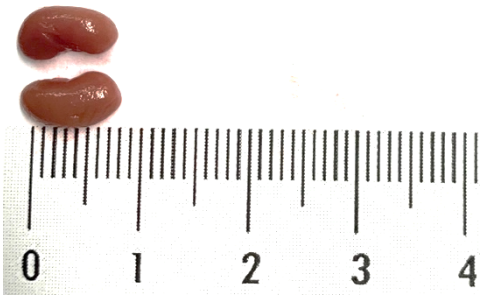


4wk male; BW 17.41 g; KW 0.10 g



4wk female; BW 14.27 g; KW 0.10 g

12wk male; BW 29.50 g; KW 0.23 g



12wk female; BW 19.87 g; KW 0.125 g

Figure 6.3 Body weight, weight and length of kidney of male and female mouse in different ages. The images were taken directly after kidney harvest. The scale ruler is shown in centimeters. Sexual difference of body weight and kidney size became significant in adult mice. This provided a clue for the material of snRNA-seq experiment of growing mouse kidneys.

BW=body weight, KW=kidney weight (g/per kidney).

6.2.1.2 Nuclear isolation protocol optimisation

Preliminary nuclear isolation experiments from 2-week and 4-week old healthy mouse kidney were performed to evaluate lysis time, lysis efficacy and nuclear

RNA quality. Results of nuclear isolation from a whole 2-week-old mouse kidney and half 4-week-old mouse kidney are shown in Table 6.1. The number of nuclei were $1.58x10^7$ and $1.28x10^7$ from 2-week-old kidney and $1.65x10^7$ and $2.56x10^7$ from 4-week-old kidney. The nuclei yields were compatible with previous nuclear isolation experiments of adult mouse kidney. All samples had good lysis efficacy with current protocol.

Kidneys from 2-week-old mice were processed together on the same day and kidneys from 4-weeks-old mice were processed together on another day. A new RNA isolation kit, RNeasy Mini Kit (Qiagen Cat No. 74104), was used instead of the previous used miRNeasy Mini Kit (Qiagen Cat No. 217004). Partial sample loss happened in the samples from 4-week-old kidney during RNA extraction and resulted lower amount of extracted nuclear RNA in sample 3 and 4. Electropherogram of Bioanalyzer showed one big peak between 18S and 28S, thus RINs of these samples were unavailable (Figure 6.4).

The nuclear isolation used the same protocol with the previous AAN experiment and the experiment went smoothly. Therefore, the Bioanalyzer results were considered related to the new RNA extraction reagent rather than poor quality of isolated nuclei. SnRNA-seq of growing kidney was carried out without repeating this experiment. MiRNeasy Mini Kit was used for retrospective nuclear RNA extraction for RIN measurement of the following snRNA-seq experiment.

Table 6.1 Nuclear isolation from 2-week-old and 4-week old health mouse kidney. The nuclei yield of growing mouse kidneys were as expected and the lysis efficacy was good. Partial sample loss happened in the 4-week-old kidney samples during RNA extraction that resulted in lower amount of extracted nuclear RNA in sample 3 and 4.

Sample ID	**1**	**2**	**3**	**4**
Age	2 week	2 week	4 week	4 week
Gender	Male	Male	Male	Female
Kidney processed for nuclear isolation	one	one	half	half
Muse cell analyzer				
Isolated nuclei	$1.58x10^7$	$1.28x10^7$	$1.65x10^7$	$2.56x10^7$
Viability	1.20%	0.40%	0.00%	0.10%
Nanodrop analysis				
RNA conc. (ng/ul)	115.7	129.1	58.8	58.5
Volume (ul)	30	30	50	50
Total nuclear RNA amount (ug)	3.47	3.87	2.94	2.93
260/280	1.87	1.89	1.90	1.92
260/230	0.56	1.08	0.26	0.09

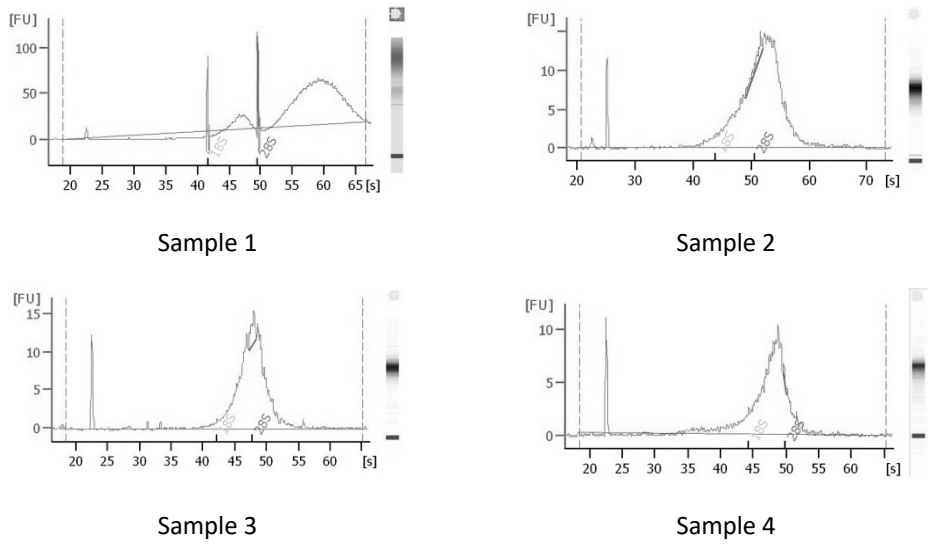


Figure 6.4 Electropherogram of Bioanalyzer of nuclear RNA from growing mouse kidneys. All samples had only one big peak between 18S and 28S. RINs from these samples were unavailable.

6.2.2 Nuclear isolation and retrospective RNA quality evaluation

The experimental design of growing kidney snRNA-seq used kidneys harvested from 2-week-old and 4-week-old mice from both sexes, thus included 4 conditions in the experiment. Each condition had two biological replicates. The experiments used C57BL/6J mice bred in Heath campus of Biological Services department of Cardiff University with standard care of housing and husbandry. Schedule 1 followed by chilled PBS perfusion was performed before kidney harvest. Nuclear isolation and library preparation were carried out immediately after kidney harvest.

Nuclei number, viability and RIN of the nuclear isolation procedure are shown in Table 6.2. Ten millions of nuclei were isolated from each sample. The nuclei yield from female and male kidneys were similar in 2-week-old kidneys ($1.4x10^7$ and $9.48x10^6$ vs. $1.07x10^7$ and $9.81x10^6$). Slightly lower nuclei yield was noted in female kidneys than male kidneys in 4 weeks old mice($1.09x10^7$ and $1.08x10^7$ vs. $1.32x10^7$ and $1.18x10^7$). The lysis efficacy was good.

RNA extraction of the residual samples from nuclei suspension was performed several hours after nuclear isolation. Bioanalyzer assessment of the nuclear RNA RINs suggested partial degradation of the nuclear RNA. RNA degradation might occur when samples stayed on ice.

Table 6.2 Nuclei number and retrospective RINs of growing mouse snRNA-seq experiment. The nuclei yield from female and male mice were similar in 2-week-old kidneys but slightly lower in female in 4-week-old kidneys. All samples had good lysis efficacy. RINs showed degradation of nuclear RNA of the residual samples. The nuclear RNA extraction was performed hours after nuclear isolation therefore RINs were only for reference.

Experiment	**2-week-old mouse kidney snRNA-seq**			
Mouse ID	**2wk_F1**	**2wk_F2**	**2wk_M1**	**2wk_M2**
Isolated nuclei (one kidney)	$1.4x10^7$	$9.48x10^6$	$1.07x10^7$	$9.81x10^6$
Viability	0.50%	0.40%	0.10%	0.50%
RIN	3.2	3.1	4.1	3.2
Experiment	**4-week-old mouse kidney snRNA-seq**			
Mouse ID	**4wk_F1**	**4wk_F2**	**4wk_M1**	**4wk_M2**
Isolated nuclei (1/2 kidney)	$1.09x10^7$	$1.08x10^7$	$1.32x10^7$	$1.18x10^7$
Viability	0.50%	0.50%	0.80%	0.80%
RIN	4.8	4.4	5	4.3

RIN = RNA integrity number

The library preparation was performed by Dr Rachel Raybould, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University. 9600 nuclei of each sample were loaded onto 10x chip, targeting recovery of 6000 cells/sample. The RNA sequencing was performed by Genome Research Hub of School of Biosciences, Cardiff University using NovaSeq.

6.2.3 Sequencing and genome mapping

The sequencing data were processed using the zUMIs pipeline (version 2.3.0)(Parekh et al. 2018). The results of genome mapping and gene counting is shown in Table 6.3. Only nuclei with ≧ 200 genes and genes identified in ≧ 3 nuclei went into the downstream analysis.

A total of 44,455 nuclei were recovered from 8 samples. The nuclei recovery rate was 57.88%. Data from all samples were merged together for downstream analysis. The combined data had median 2,289 genes per nucleus and median 4,463 UMIs per nucleus. Compared to the snRNA-seq experiment of AAN in Chapter 4, which the Naïve samples had median 1,207 genes per nucleus and the AAN samples had median 1,555 genes per nucleus, the number of genes detected per nucleus was substantially increased. This provided more information for cell clustering and further bioinformatics analysis. This increment might represent the benefit from the increased depth of sequencing by NovaSeq.

Table 6.3 Number of nuclei and genes detected using zUMIs pipeline. SnRNA-seq of 2-week-old and 4-week-old mouse kidneys were performed on the 10X platform. Genome mapping and gene counting were carried out using zUMIs pipeline. Only nuclei with ≧ 200 genes and genes identified in ≧ 3 nuclei went into the downstream analysis.

Experiment: 2-week-old mouse kidney snRNA-seq				
Mouse ID	**2wk_F1**	**2wk_F2**	**2wk_M1**	**2wk_M2**
Number of nuclei	5198	5784	4233	5350
Genes detected	24903	25358	25172	24912
Genes detected per nuclei (mean)	2200.165	2180.577	2536.06	2055.729
Genes detected per nuclei (median)	2112	2086.5	2439	1939.5
UMIs detected per nuclei (mean)	4580.867	4402.029	5677.734	4066.497
UMIs detected per nuclei (median)	4098.5	3923.5	5061	3509
Experiment: 4-week-old mouse kidney snRNA-seq				
Mouse ID	**4wk_F1**	**4wk_F2**	**4wk_M1**	**4wk_M2**
Number of nuclei	6896	6150	5661	5183
Genes detected	25848	25527	25560	25554
Genes detected per nuclei (mean)	2324.976	2419.973	2528.285	2750.85
Genes detected per nuclei (median)	2251	2360	2457	2663
UMIs detected per nuclei (mean)	4712.792	4988.909	5466.53	6372.983
UMIs detected per nuclei (median)	4258	4601.5	4909	5714

6.2.4 Quality control, batch-effect correction and doublet removal

6.2.4.1 Quality control of the growing kidney dataset

After merging the 8 samples, the number of feature counts, RNA counts and percentage of mitochondrial genes of each nucleus are shown in Figure 6.5A. Nuclei with ≦ 200 genes or ≧ 6000 genes, or ≧ 2.5% mitochondrial genes were removed. This step removed a total of 162 nuclei significant outliers. The data was then normalized and scaled for principal component analysis (PCA, Figure 6.5B).

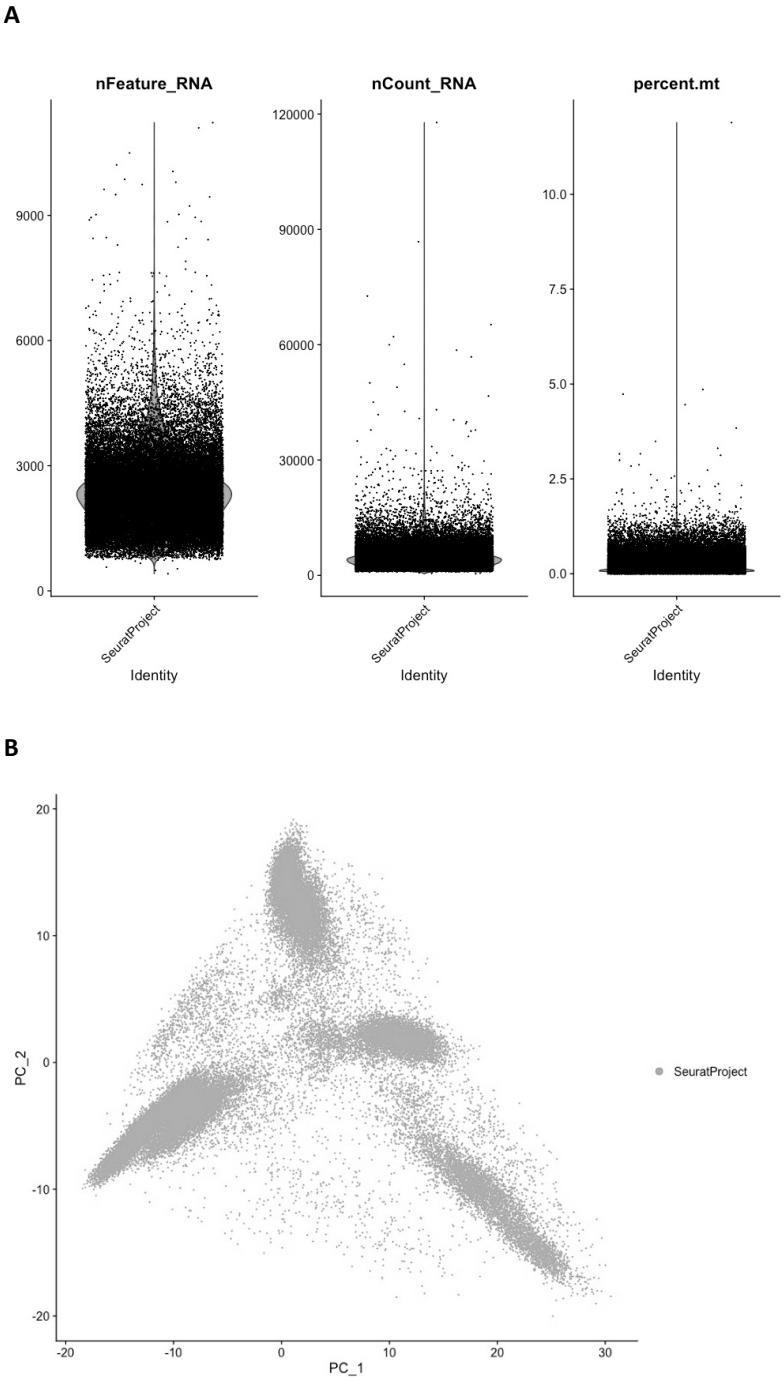


Figure 6.5 General information of the combined dataset of 8 growing mouse kidney. Each dot represents a nucleus. (A) Violin plots show number of feature counts, RNA counts and percentage of mitochondrial genes of each nucleus. Significant outliers on the violin plots were removed. (B) Result of principal component analysis.

6.2.4.2 Batch-effect correction

Batch-effect resulted from sample preparation, condition and sequencing. Data intergradation and batch-effect correction were performed using an R package, harmony (Korsunsky et al. 2019; Tran et al. 2020). After PCA embeds cells into a 2-dimensional space, Harmony algorithm inputs a PCA embedding of cells, along with their cell-specific status, and returns a batch corrected embedding to adjust for dataset specific effects.

Figure 6.6 shows the distribution of nuclei on the PCA plot in different cell-specific status and their embedding values of principal component (PC)_1 in the PCA. In this dataset, PC_1 was largely affected by age. Since libraries preparation of samples with the same age were performed on one chip, both age and sample preparation could cause this batch-effect. The embedding values of PC_2-PC_9 are shown in Figure 6.7. PC_ 2-9 were all affected by age and PC_4 was also affected by gender in 4-weeks-old kidneys. Base on this, age was set as the variable to correct in Harmony.

Results of harmony corrected PCA shows that the dataset specific effect of nuclei distribution was diminished (Figure 6.8A). Harmony corrected embedding values of PC_1, named as harmony_1, of each sample were relative equivalent, as well as PC_2 to PC_9, named as harmony_2 to harmony_9 (Figure 6.8B).

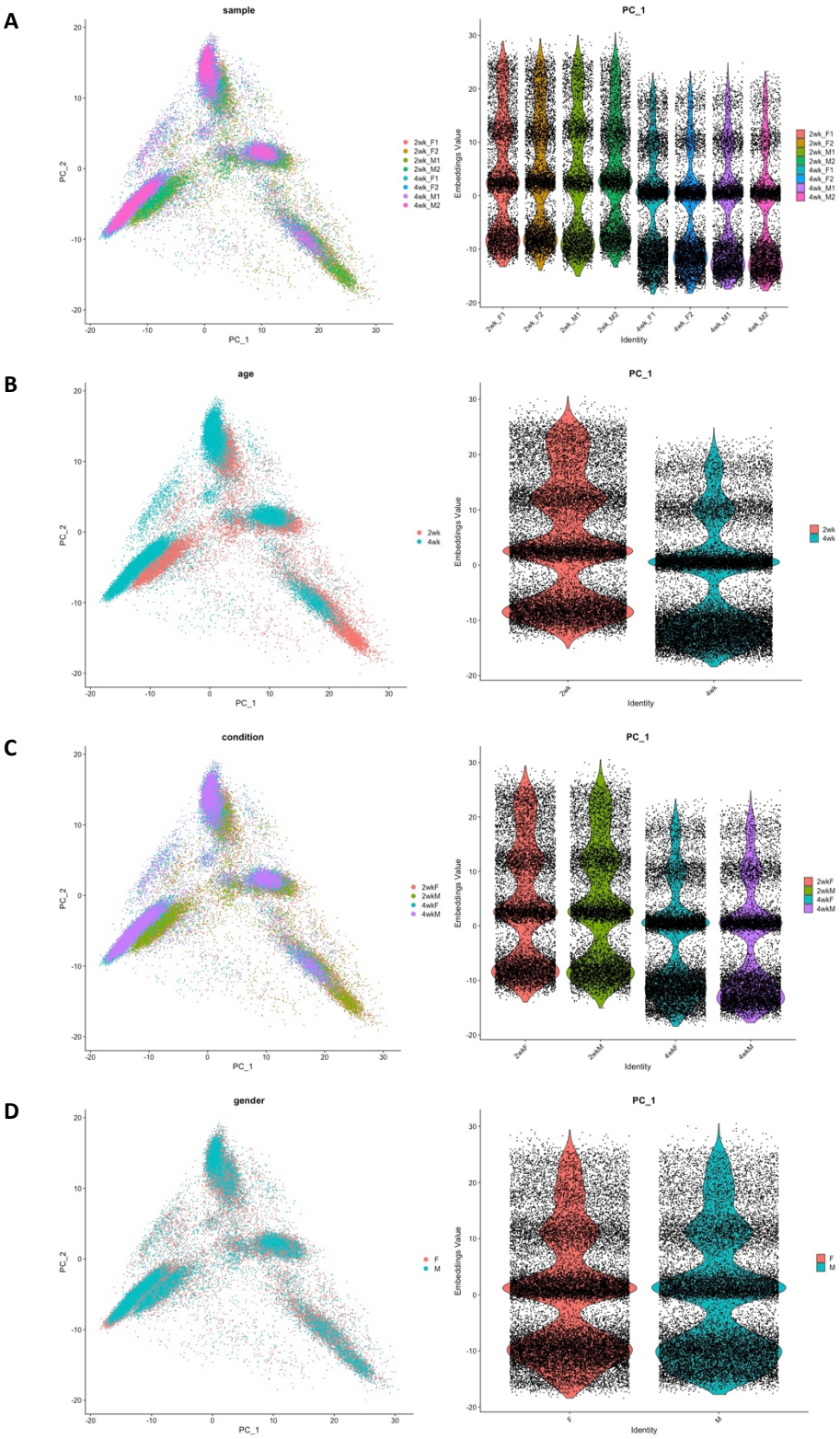


Figure 6.6 PCA in different cell-specific status before batch-effect correction. Left side of the figures shows PCA plots coloured by their cell-specific status: (A) sample; (B) age; (C) condition (gender and age); (D)gender. Right side of the figures shows the embedding value of each nucleus in PC_1 on violin plots. PC_1 was largely affected by age. F=female; M=male; PC= principal component.

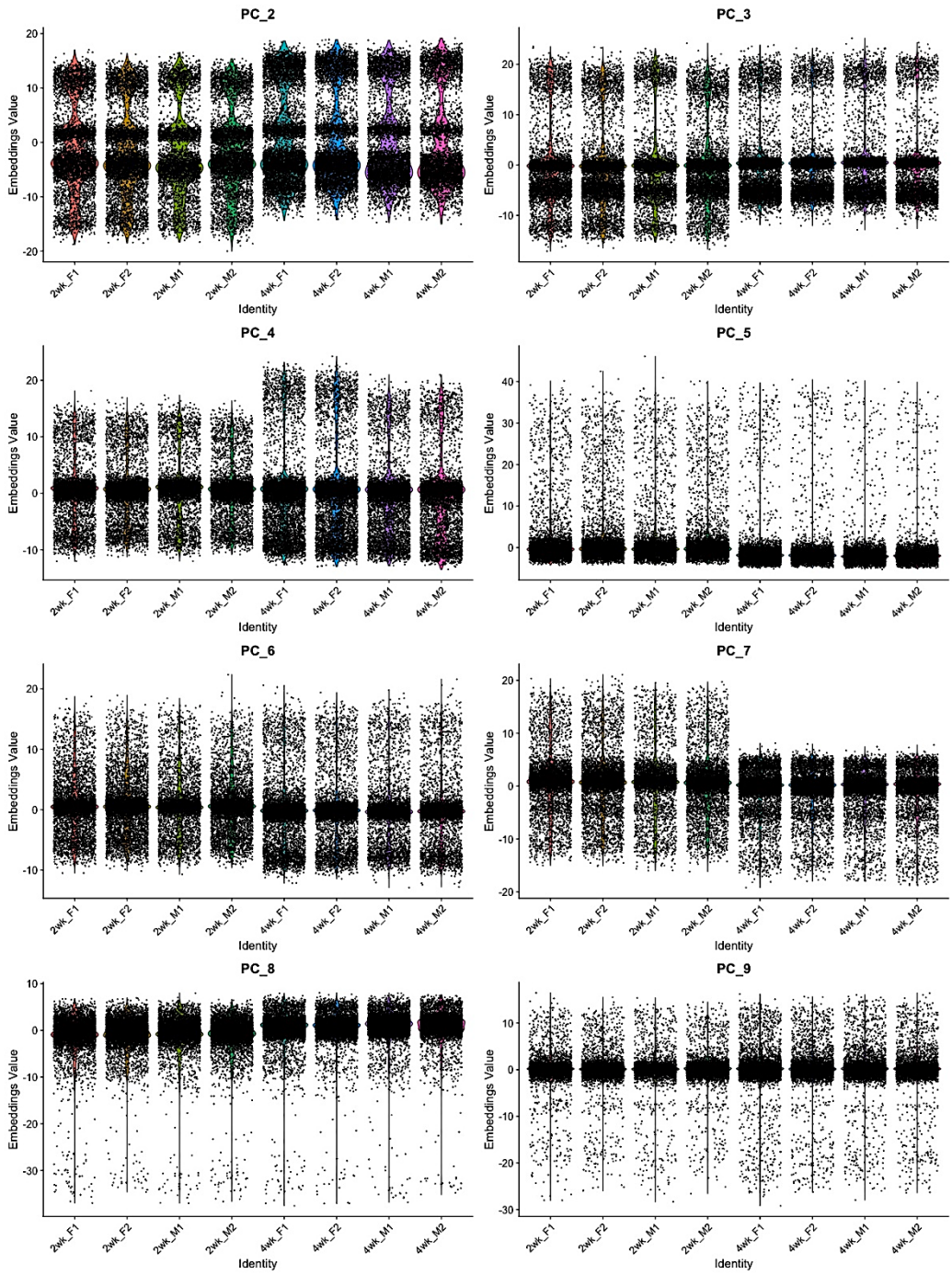


Figure 6.7 Violin plots of embedding values of PC_2 to PC_9 of each sample. Embedding values of each nucleus were significantly affected by age in PC_2 to PC_9. PC_4 also affected by gender. Age was set as the variable to correct in Harmony to correct the batch difference.

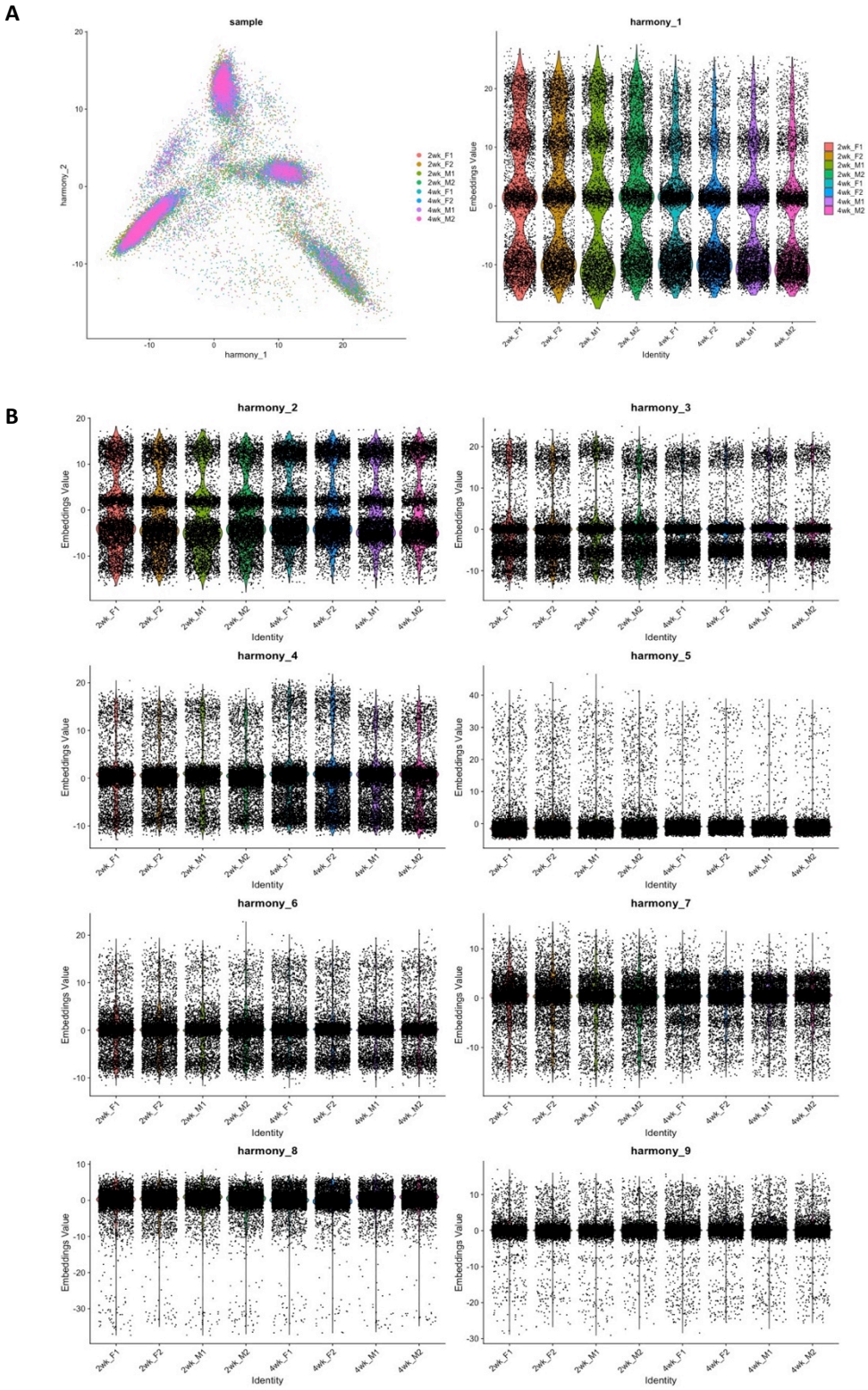


Figure 6.8 Harmony corrected PCA embedding. (A) PCA plot after the dataset was adjusted by Harmony (left) and the harmony corrected embedding values of PC_1, named as harmony_1, of each sample (right); (B) Harmony corrected embedding values of PC_2 to PC_9, named as harmony_2 to harmony_9. Dataset specific effect was diminished after harmony correction.

6.2.4.3 Removal of suspicious doublets

Cell clustering, t-SNE and UMAP were built on harmony corrected PCA results (Figure 6.9 A and B). The cluster 10, 13, 19, 23, 25, 27, 31, 36, 37 had higher nFeatures than other clusters (Figure 6.9C). These clusters might be clusters with multiplets or undergoing proliferation.

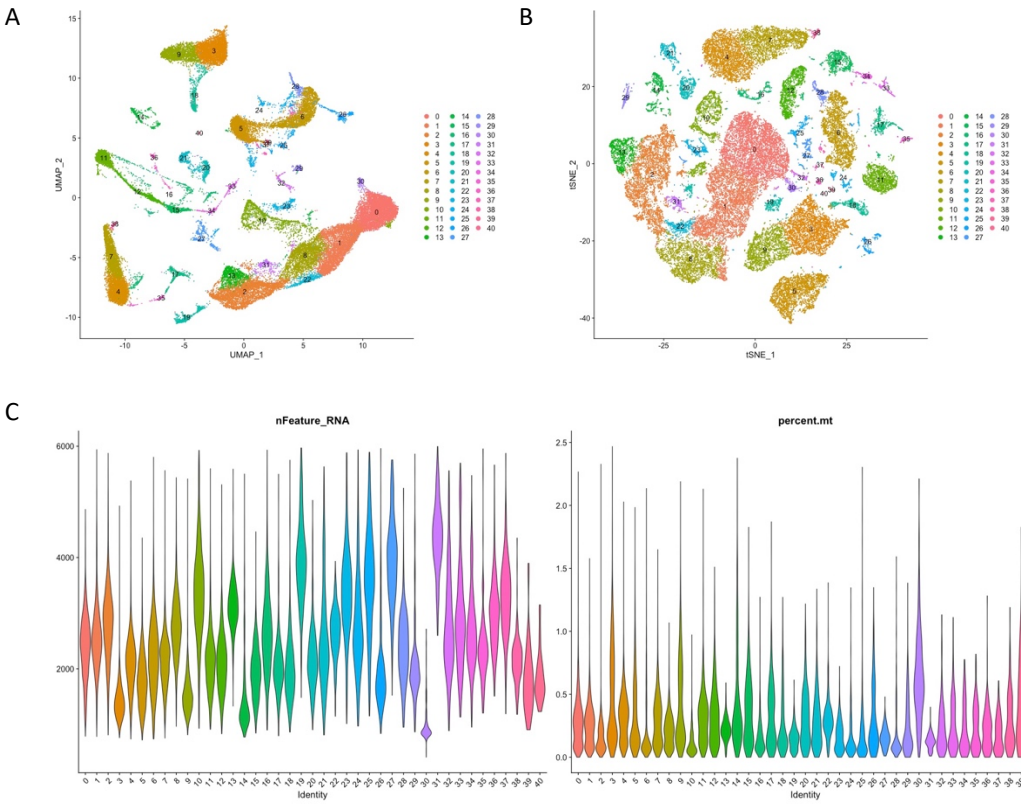


Figure 6.9 Result of cell clustering built on harmony corrected PCA. The clustering result is visualized using (A) UMAP plot and (B) t-SNE plot. (C) Violin plot shows that cluster 10, 19, 23, 25, 27, 31 had higher number of features (nFeature) than other clusters, which implied that these clusters were clusters of multiplets or undergoing proliferation.

Estimated multiplet rate of each sample was ~4.6% by the user guide from 10X company (https://www.10xgenomics.com/resources/user-guides/). Doublets were detected using an R package, Doublet Finder (McGinnis et al. 2019). Result of DoubletFinder analysis is shown on Figure 6.10 The clusters that most cells were identified as suspicious doublets, including the cluster 19, 23, 25, 27, 36, 37, were completely removed from downstream analysis. These clusters were compatible with clusters with higher nFeatures in Figure 6.9C.

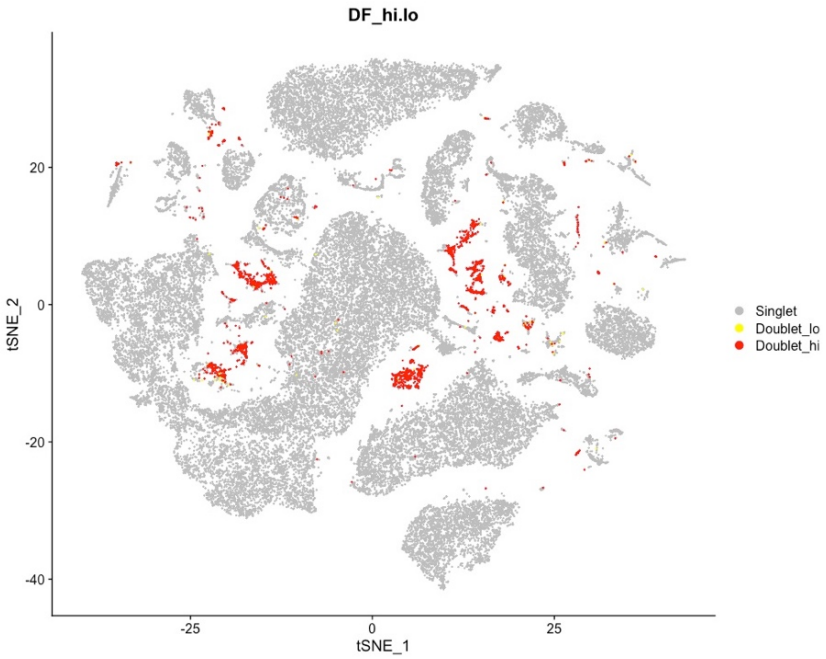


Figure 6.10 DoubletFinder identified suspicious doublets on t-SNE plot. Suspicious doublets were coloured by red or yellow on T-SNE plot. Nuclei with red colour indicated they had high probability of doublet, nuclei with yellow our indicated they had lower probability of doublet, nuclei with grey colour were identified as singlet.

Nuclei labelled as dubious doublets but not in these 2 clusters were also excluded to minimize the impact of doublets on clustering. T-SNE plot before and after doublet removal is shown in Figure 6.11. After quality control and doublet removal, a total of 41789 nuclei were included into downstream analysis. Nuclei number of each sample that went into the next quality control step and final analysis are listed in Table 6.4.

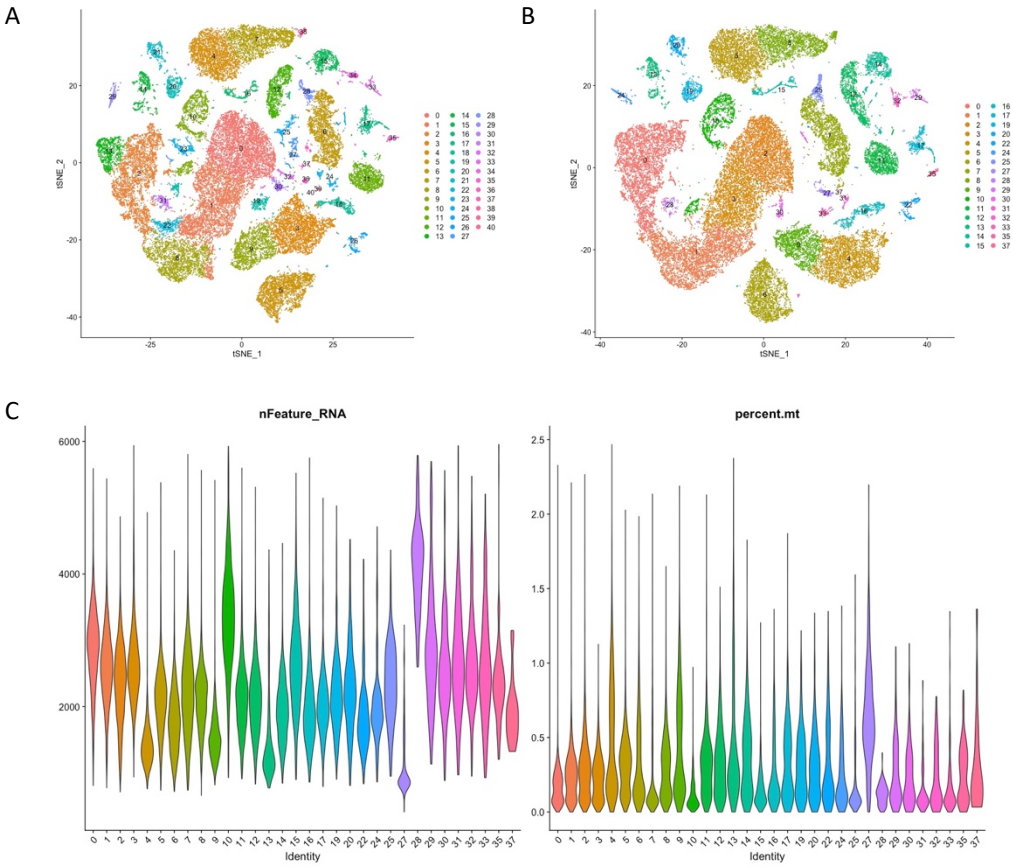


Figure 6.11 Result of doublet removal using DoubletFinder. T-SNE plot before (A) and after (B) doublets removal. (C) Violin plots shows the genes detected per cell (nFeature_RNA) and percentage of mitochondria gene (percent.mt) of each cluster.

Table 6.4 Nuclei number of each sample after quality control and doublet removal. Number of nuclei withdrawn from each step was relatively equivalent in each sample. A total of 41789 nuclei was included into downstream analysis.

Experiment: 2-week-old mouse kidney snRNA-seq				
Mouse ID	**2wk_F1**	**2wk_F2**	**2wk_M1**	**2wk_M2**
Nuclei loaded into chip	9600	9600	9600	9600
Nuclei detected	5198	5784	4233	5350
Quality control: keep nuclei with nFeature_RNA > 200 & nFeature_RNA < 6000 & percent.mt < 2.5	5181	5777	4213	5337
Nuclei number involved in final analysis after remove suspicious doublets	4944	5506	4030	5080
Experiment: 4-week-old mouse kidney snRNA-seq				
Mouse ID	**4wk_F1**	**4wk_F2**	**4wk_M1**	**4wk_M2**
Nuclei loaded into chip	9600	9600	9600	9600
Nuclei detected	6896	6150	5661	5183
Quality control: keep nuclei with nFeature_RNA > 200 & nFeature_RNA < 6000 & percent.mt < 2.5	6876	6140	5629	5140
Nuclei number involved in final analysis after remove suspicious doublets	6430	5783	5240	4776

6.2.5 Clustering and cell type identification

6.2.5.1 Clustering of nuclei

After doublet removal, the dataset was normalized, scaled and clustered again. PCA followed by harmony batch-effect correction was done before clustering. Unbiased hierarchical clustering results were visualized using UMAP plot. To determine appropriate resolution of clustering, resolution 0.5, 1.0, 2.0, 3.0, 4.0, 8.0 of the *FindClusters* function were tested. A higher resolution would result in a higher number of clusters without changing the shape of the UMAP plot. With resolution of 0.5, 1.0, 2.0, 3.0, 4.0, 8.0, there were 32, 36, 48, 58, 65, 100 clusters identified in the combined dataset (Figure 6.12). A resolution of 3.0 was selected, taking the range of expected number of cell types in a mouse kidney and the structure of the UMAP plot into consideration, aiming to have a proper number of cell types without missing any rare but distinct clusters in the growing kidneys.

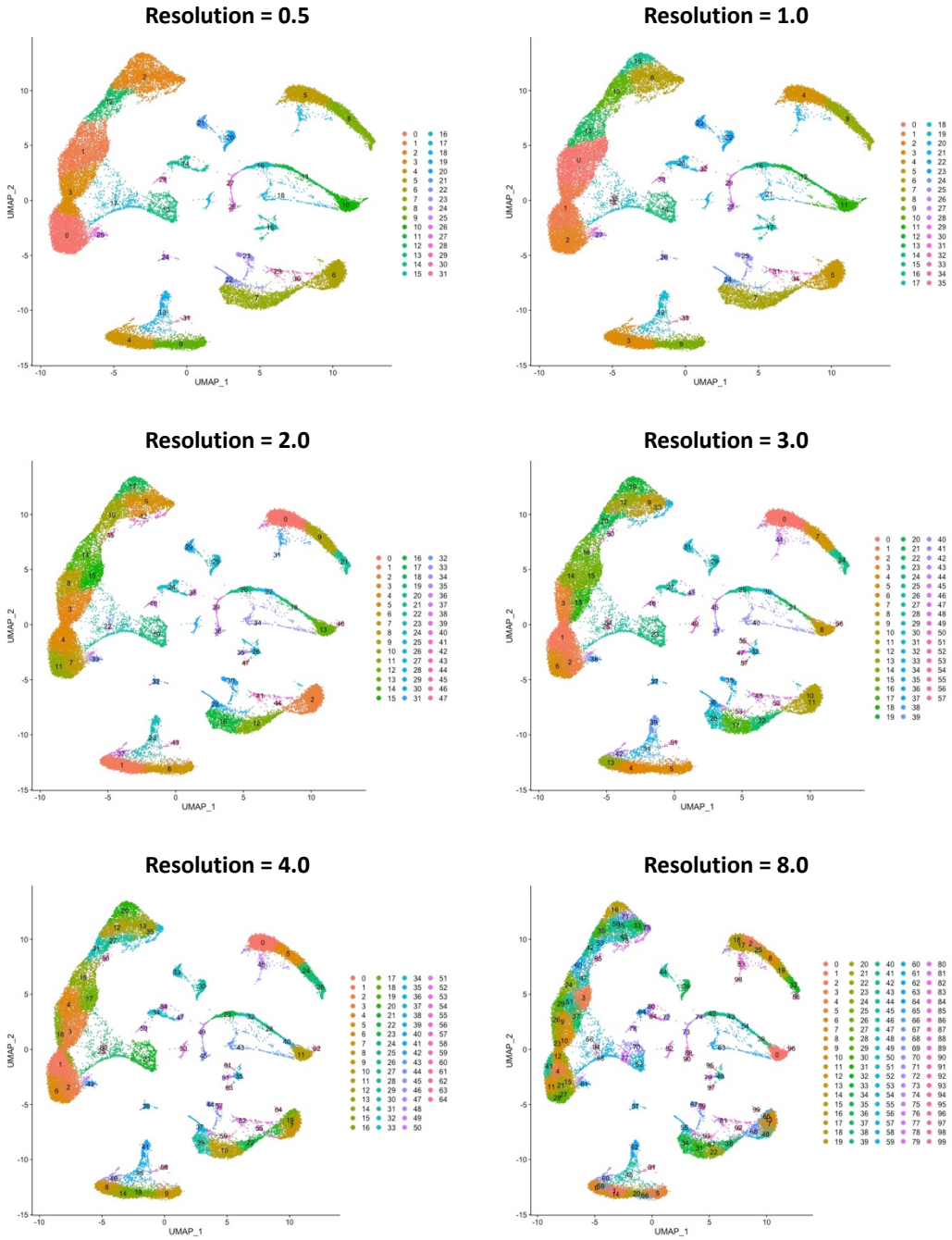


Figure 6.12 UMAP plots of the growing mice kidney dataset with different resolution when performing nuclei clustering. A higher resolution results in a higher number of clusters without affecting the shape of UMAP. With resolution of 0.5, 1.0, 2.0, 3.0, 4.0, 8.0, there were 32, 36, 48, 58, 65, 100 clusters identified in the combined dataset.

Nuclei on the UMAP were coloured by their sample origin to confirm that there was no significant dataset specific effect of clustering (Figure 6.13). Nuclei from different samples distributed heterogeneously on the UMAP, indicating the data was well integrated and clustered. T-SNE plot with resolution 3.0 is shown in Figure 6.14.

Both UMAP and t-SNE show the same clustering result. Distribution of nuclei is more condensed on the UMAP plot and the distance of clusters on UMAP plot could infer the difference of genetic expression profiles between the clusters, therefore UMAP was superior for visualizing single-cell RNA sequencing result. Violin plots shows the genes detected per cell (nFeature_RNA) of each cluster on Figure 6.15. Clusters with higher nFeature_RNA were considered as proliferative clusters.

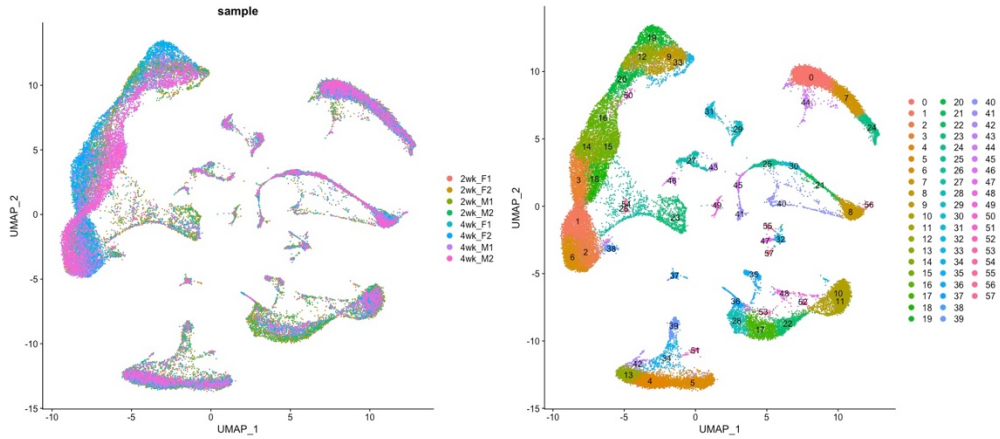


Figure 6.13 UMAP plots of clustering result with a resolution of 3.0. In the left plot, the nuclei were coloured by their sample origin. Nuclei from different samples distributed in the UMAP heterogeneously, indicating the data was well integrated.

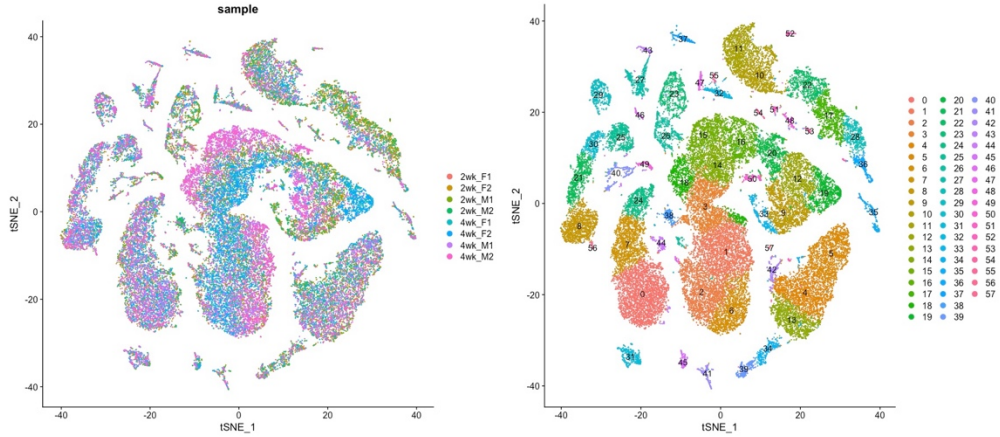


Figure 6.14 T-SNE plots of t clustering result with a resolution of 3.0. In the left plot, the nuclei were coloured by their sample origin. In the right plot, the nuclei were coloured by clusters.

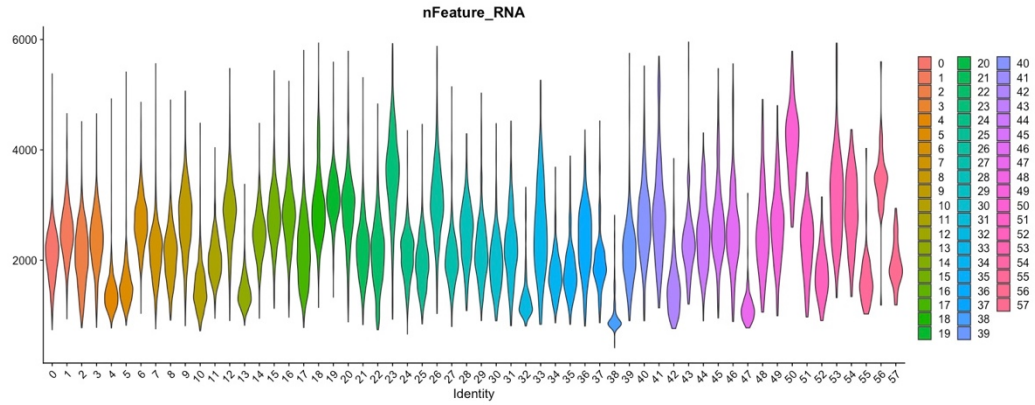


Figure 6.15 Violin plots of the clustering result. Violin plots shows the genes detected per cell (nFeature_RNA) of each cluster.

6.2.5.2 Cell type identification

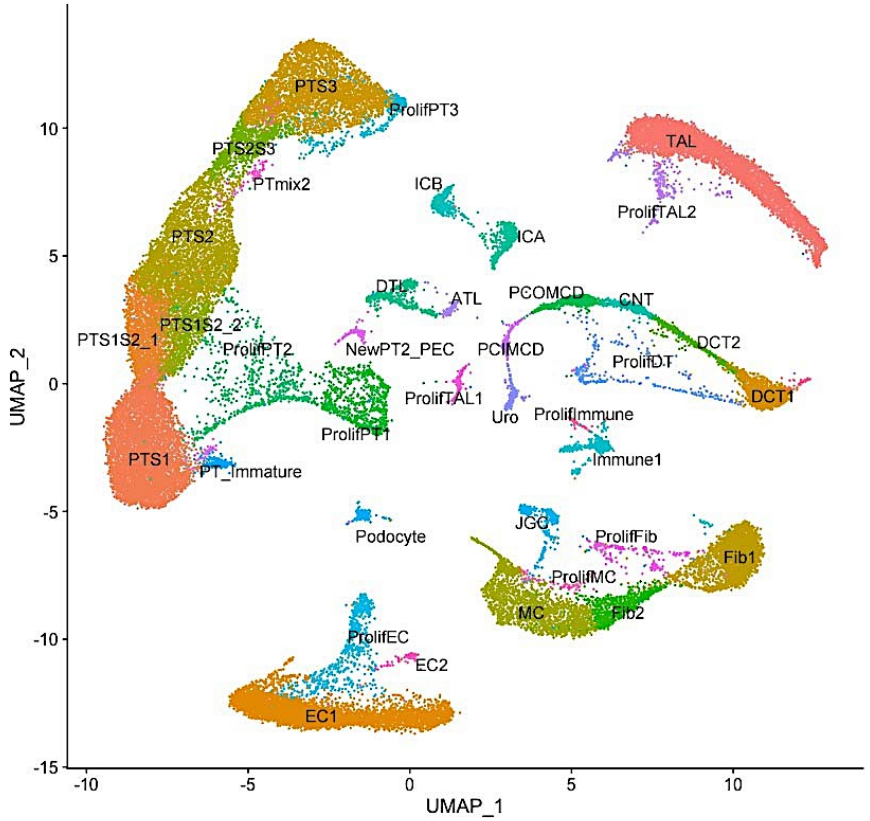
Marker genes used in the AAN experiment (Table 4.5) were used for cell type identification for the 2-week-old and 4-week-old growing kidneys. Using the same marker genes, all major cell types were identified in the growing kidneys (Figure 6.16A). The contribution of each condition to the clusters is shown in Figure 6.16B. Result of cell cycle analysis helped to identify proliferative clusters (Figure 6.17).

In the AAN experiment, adult health kidney segregated PTCs to five abundant clusters, mapping to S1, S1-2, S2, S2-cortical S3, and medullary S3 segments. Four additional new PTC clusters were identified and showed increased abundance in fibrotic kidneys, which permitting labelling as, proliferating, New-PT1, New-PT2 and (present only following injury) New-PT3. Growing kidneys segregated PTCs to 12 clusters, including 6 mature PTC clusters, PTmix, 3 proliferative PTC clusters, NewPT2_pareital epithelial cell (NewPT2_PEC), and PT_Immature. The S1, S1-S2, S2, S2-S3 shared the same marker genes with adult mouse kidney (Figure 6.16C). The marker gene of medullary localisation, *Cyp7b1*, was not detected in the S3 cluster of growing kidney, therefore the cluster was labelled as S3 but not medullary S3. This might relate to the lack of cortical-medullary differentiation in the growing kidneys. The PTmix cluster expressed marker genes for all proximal tubular segments. Cells in ProlifPT1 expressed *Top2a* and *Ki67* and they were labelled as G2M phase in the cell cycle analysis. The ProlifPT2 and ProlifPT3 were labelled as S phase cells, which expressed *Hells* and *Pola1.* ProlilfPT2 expressed marker genes of S1 and S2 and ProlifPT3 expressed marker gene of S3,

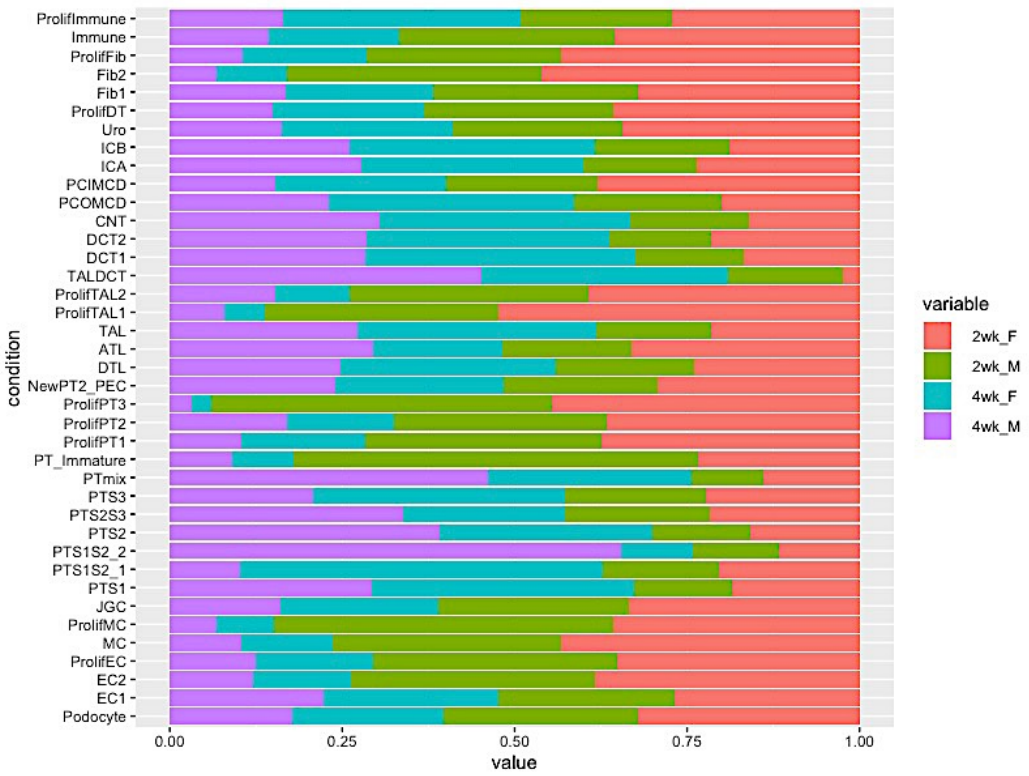
respectively. The NewPT2_PEC expressed the New-PT2 marker gene, *Akap12*, without the marker gene of New-PT3, *Havcr1*, expression. In the AAN experiment, New-PT2 markers identified not only PTC but also glomerular parietal epithelial cell staining in normal mouse kidney, therefore this cluster was labelled as NewPT2_PEC. The PT_Immature cluster expressed *Slc34a1* with weak *CFH* expression, and was considered to contain immature PTCs. Also, PT_Immature cluster was the cluster 38 on Figure 6.15, which had low number of genes detected per nuclei. Characteristics of the PT_Immature cluster should be further investigated and validated.

The results revealed the cellular composition of kidneys at different ages. Compared to the 2-week-old kidneys, the proportion of podocytes, endothelial cells, mesangial cells, JG cells, fibroblasts, and immune cells decreased whereas the proportion of proximal and distal tubular cells increased in 4-week-old kidney (Figure 6.16B).

A


PTS3
ProlifPT3
PTS2S3
PTmix2
ICB
TAL
ProlifTAL2
PTS2
ICA
DTL
ATL
PCOMCD
CNT
PTS1S2_2
PTS1S2_1
ProlifPT2
NewPT2_PEC
PCIMCD
DCT2
ProlifDT
ProlifTAL1
DCT1
Uro
ProlifImmune
ProlifPT1
Immune1
PTS1
PT_Immature
Podocyte
JGC
ProlifFib
Fib1
ProlifMC
MC
Fib2
ProlifEC
EC2
EC1
UMAP_2
UMAP_1
10
5
0
-5
-10
-15
-10
-5
0
5
10

B


ProlifImmune
Immune
ProlifFib
Fib2
Fib1
ProlifDT
Uro
ICB
ICA
PCIMCD
PCOMCD
CNT
DCT2
DCT1
TALDCT
ProlifTAL2
ProlifTAL1
TAL
ATL
DTL
NewPT2_PEC
ProlifPT3
ProlifPT2
ProlifPT1
PT_Immature
PTmix
PTS3
PTS2S3
PTS2
PTS1S2_2
PTS1S2_1
PTS1
JGC
ProlifMC
MC
ProlifEC
EC2
EC1
Podocyte
condition
0.00
0.25
0.50
0.75
1.00
value
variable
2wk_F
2wk_M
4wk_F
4wk_M

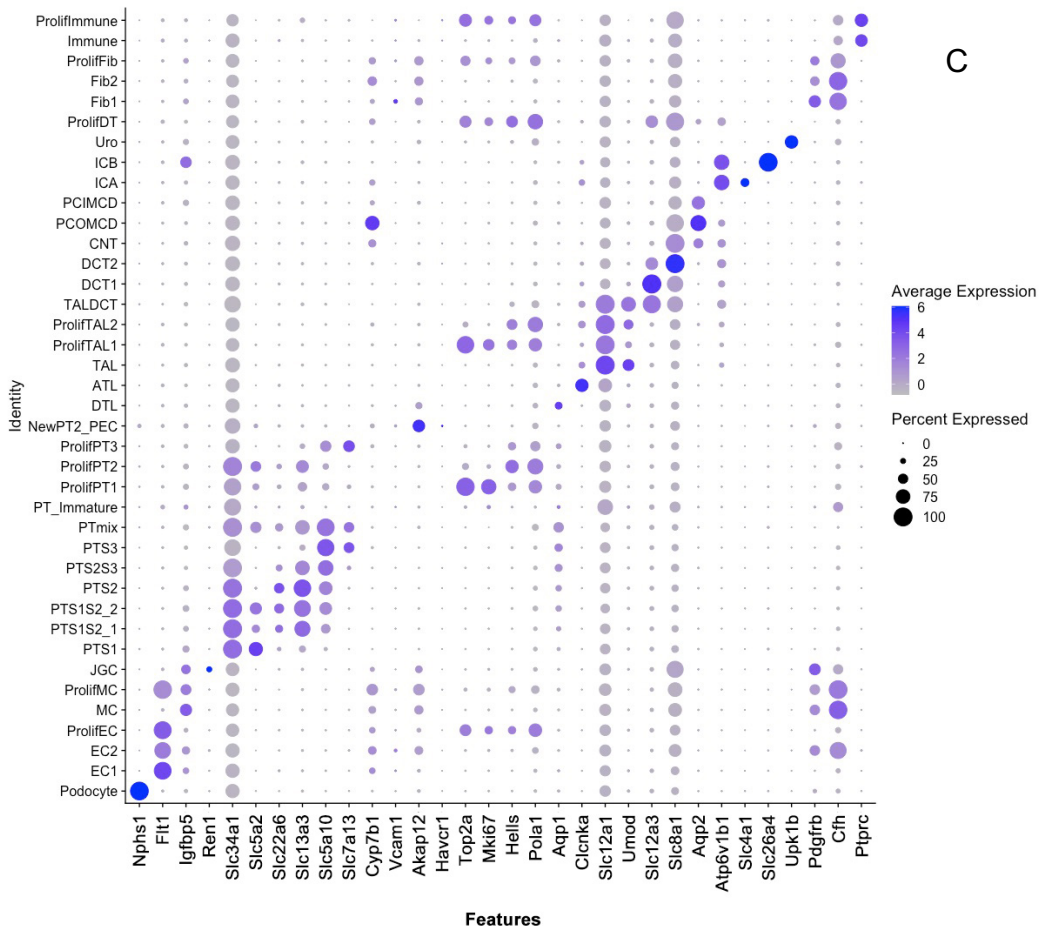


Figure 6.16 Clustering and cell-type identification of 41789 nuclei using combined datasets from eight 2-week and 4-week old mouse kidneys. (A) UMAP plot of the combined dataset. All major cell types of kidneys and their proliferative cells were identified. Growing kidney segregated PTCs to 12 clusters, including 5 mature PTC clusters, 1 PTmix cluster, 3 proliferative PTC clusters, NewPT2_parital epithelial cell (NewPT2_PEC), and PT_Immature. (B) Contribution of different age and gender of kidneys to each cell type. (C) Dotplot shows the expression levels and the percentage of gene expression of the canonical genes in each distinct cell type.

EC, endothelium; MC, mesangial cell; JGC, Juxtaglomerular cells; PT, proximal tubule; S1/S2/S3, segment 1/2/3 of proximal tubule; DTL, descending thin limb; ATL, ascending thing limb; TAL, thick ascending limb; DCT1/DCT2, distal convoluted tubule 1/2; CNT, connecting tubule; PC-OMCD, principal cell-outer medullary collecting duct; PC-IMCD, principal cell-inner medullary collecting duct; IC-A, intercalated cells, type A; IC-B, intercalated cells, type B; ProlifDT, proliferative distal tubule; Fib, fibroblast.

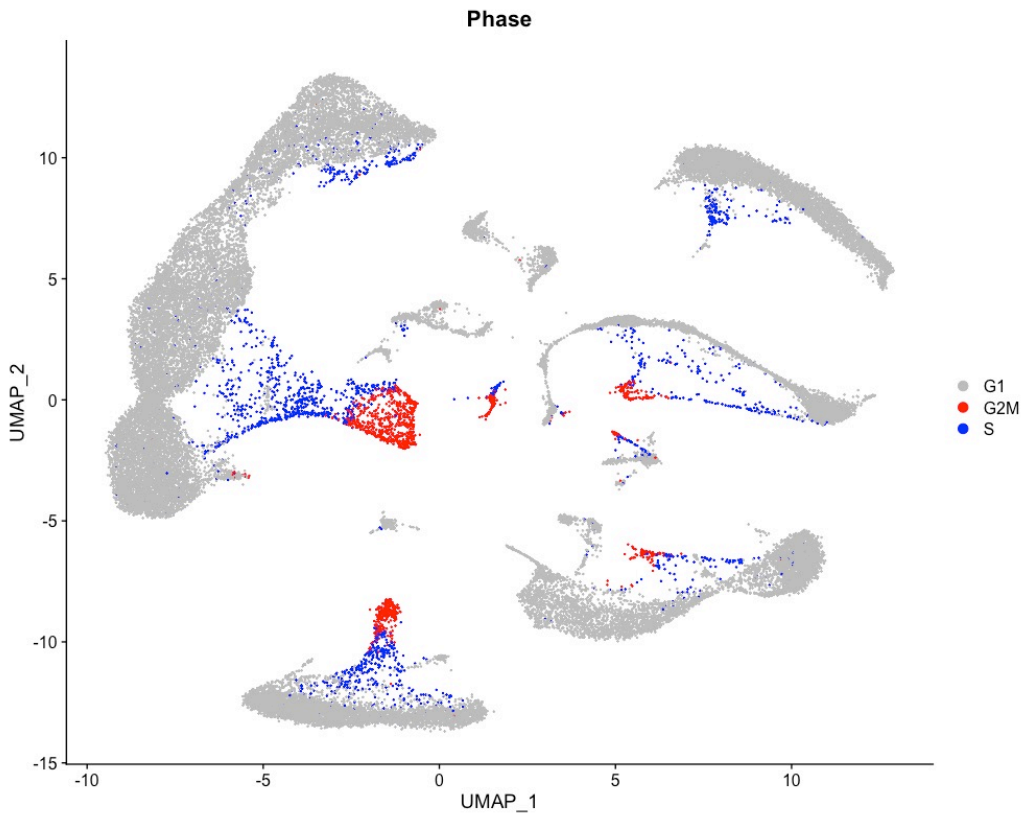


Figure 6.17 Cell cycle analysis. Result of cell cycle analysis helped to identify cells in proliferative status. The location of the proliferative clusters on the UMAP were close to their original cell types.

6.2.5.3 Proliferative cells

Proliferative cells were identified in all major cell types. They expressed both proliferative genes and marker genes of the cell type to which they are expected to differentiate, and localized next to their cognate mature cell type on the UMAP plot. Here I identified ProlifPT1, ProlifPT2, ProlifPT3, ProlifEC, ProlifTAL1, ProlifTAL2, ProlifDT, ProlifFib, ProlifMC, and ProlifImmune (Figure 6.16A). Results of cell cycle analysis showed the cell cycle status of each cell. The proliferative PTCs and the proliferative TALs were divided into more than one cluster. ProlifPT1 and ProlifTAL1 were in G2M phase and ProlifPT2, ProlifPT3 and ProlifTAL2 were in S phase. Compared to the 4-week-old mouse kidneys, the 2-week-old mouse kidneys had a larger proportion of proliferative cells (Figure 6.18A). Cell proliferation is more prominent in young mouse and contributes to kidney growth.

6.2.5.4 Female to male difference

Puberty is the final stage of maturation. The puberty of C57B6/J mouse happens at 4 weeks old. Female to male difference would become significant from genetic profiles to gross appearance after the onset of puberty. Grossly on the UMAP plot of the growing mouse kidneys, the female to male difference was more significant in mature PTCs than other cell types (Fig 6.18B). Further analysis of the gender difference of different cell types is described in the following section.

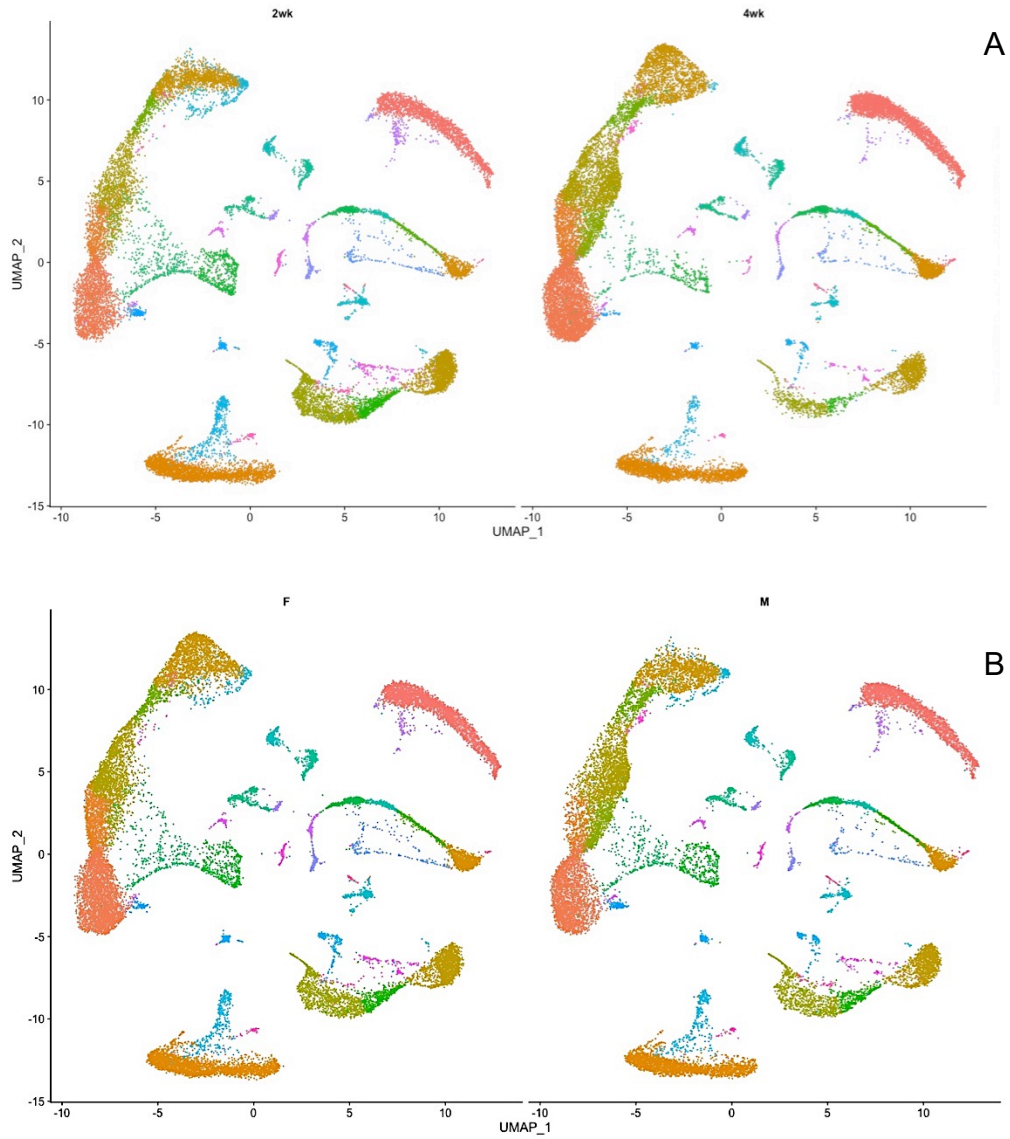
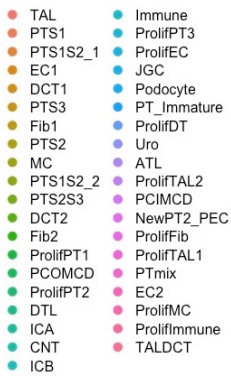


Figure 6.18 UMAP plot of growing kidneys in different conditions. Cells on the UMAP plot was split to (A) 2-week-old versus 4-week-old and (B) female versus male. The 2-week-old kidneys had a larger proportion of proliferative cells. The male to female difference was more significant in PTCs than other cell types.

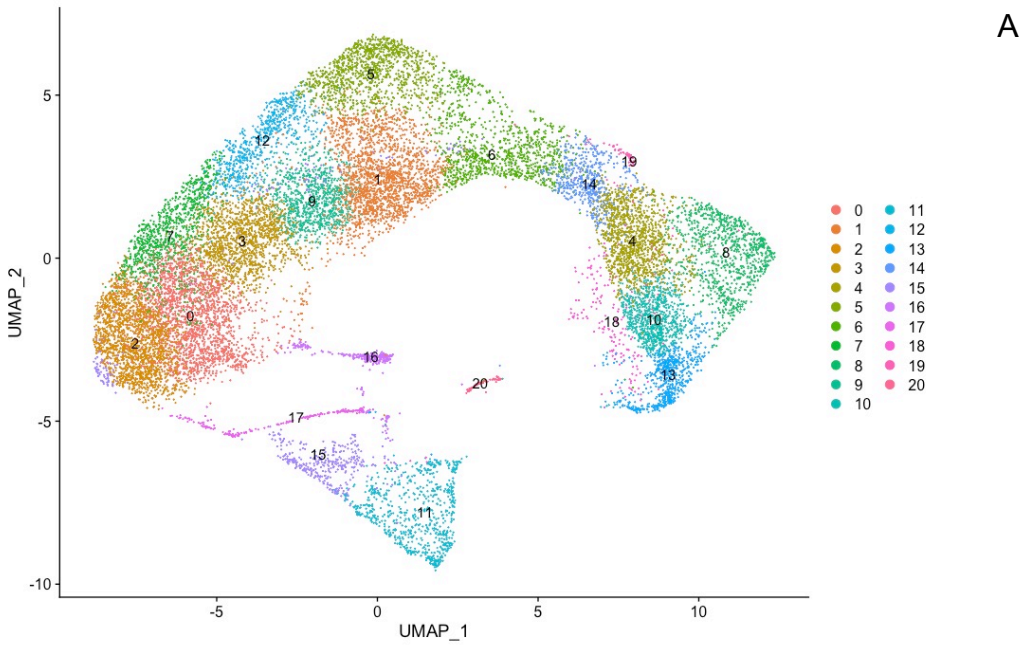


6.2.6 Analysis of PTC clusters

For further analysis of PTCs, the 12 PTC clusters were selected and cells were re-clustered. Data from 2 and 4-week-old PTCs were integrated using Seurat *FindIntegrationAnchors* and *IntegrateData* function. This provided an integrated assay of top 2000 variable genes that helped downstream trajectory and velocity analysis.

There were 21 clusters identified after PTC re-clustering with Seurat (Figure 6.19). Cells classified as the same cell type in the whole dataset analysis were still clustered together when re-clustering (Figure 6.19B). The dotplot shows the expression level and the percentage of expression of the canonical genes of PTC clusters identified in the AAN experiment, including mature PTC, New-PT clusters and proliferative PTCs (Figure 6.20).

The marker genes of New-PT1 and New-PT3, *Vcam1* and *Havcr1*, were rarely detected in these PTC clusters. The marker genes of New-PT2 expressed in NewPT2_PEC, and the cluster 17 in the re-clustering analysis. *Xist* codes for a long noncoding RNA that is part of a core mechanism for X Chromosome inactivation and is described as exclusively expressed in female cells (Ransick et al. 2019). In this dataset, *Xist* was detected in 91.7% of nuclei from female cells but only 0.4% nuclei from male cells.



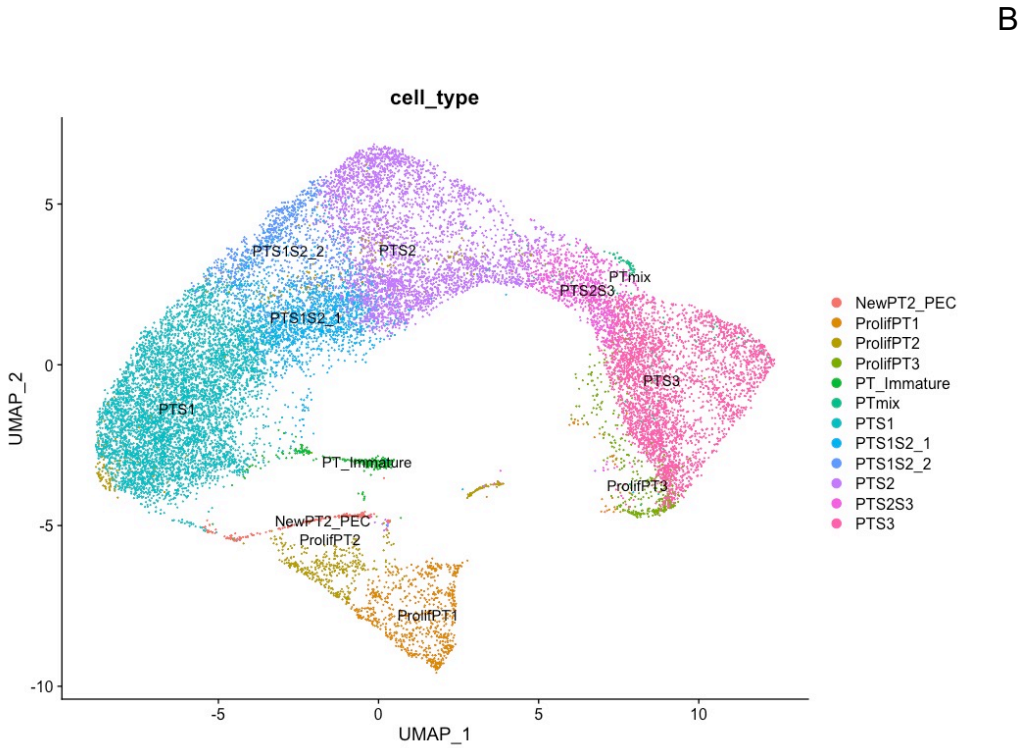


Figure 6.19 Result of PTC re-clustering. The re-clustering result is visualized using the UMAP plot. Cells are colored by their clustering result in re-clustering analysis (A) and their cell type (B). Cells classified as the same cell type in the original analysis were still clustered together in the re-clustering analysis.

PT, proximal tubule; S1/S2/S3, segment 1/2/3 of proximal tubule; ProlifPT, proliferative proximal tubule; NewPT2_PEC, new proximal tubule 2_pariatal epithelial cell.

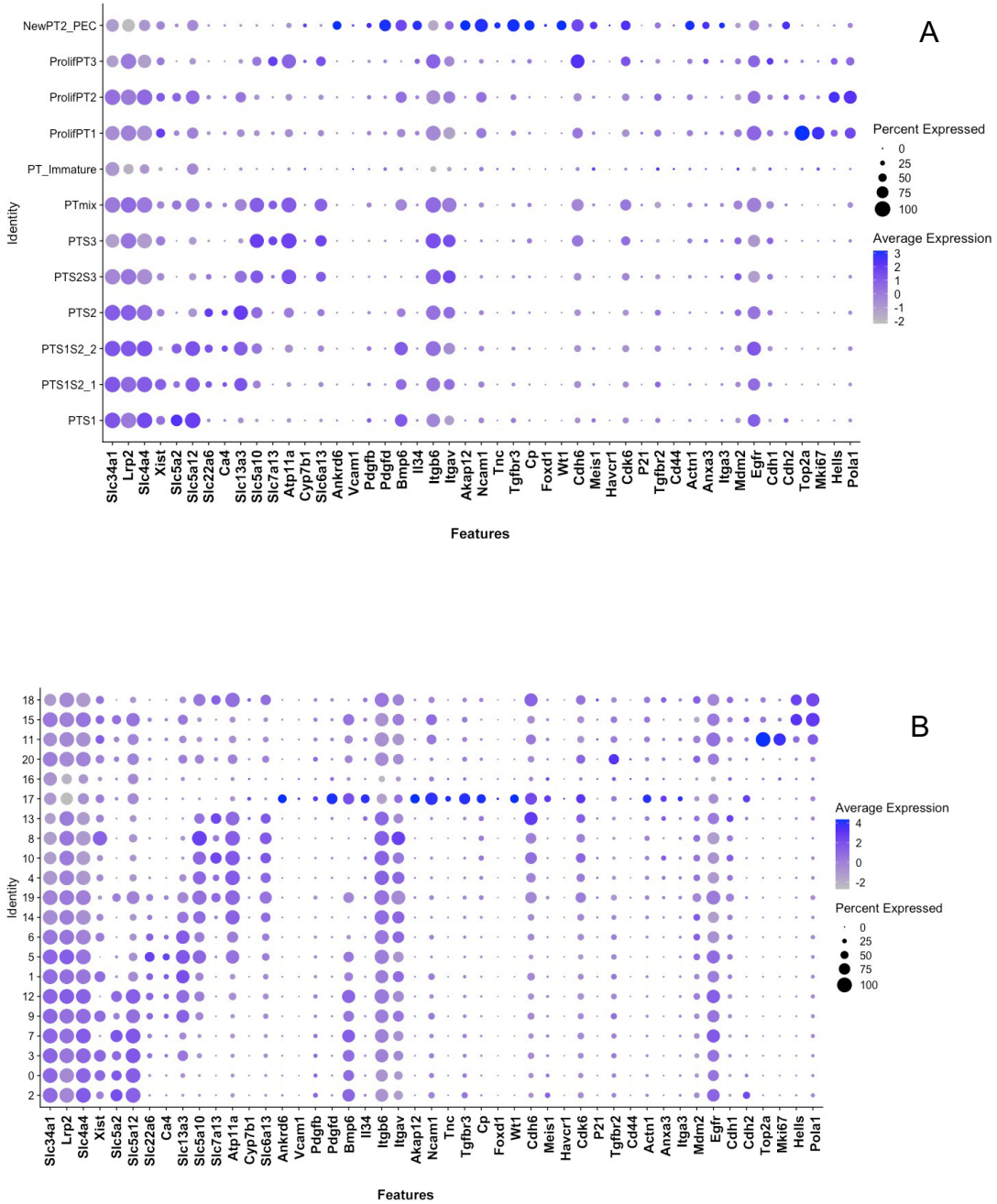
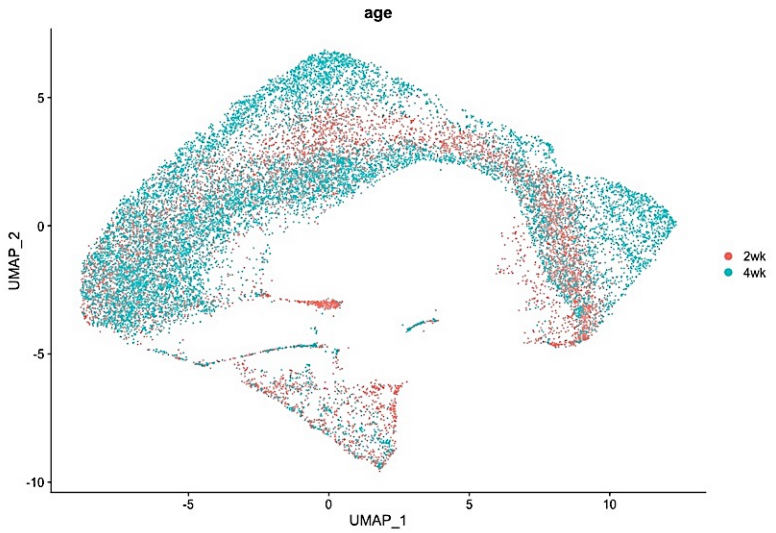


Figure 6.20 Dotplot of PTC clusters. The dotplot shows the expression level and the percentage of expression of a canonical gene of (A) PTC clusters identified in the whole data analysis (B) The clusters identified in re-clustering. The dotplot used the gene list on from the PTC analysis result in Chapter 4, which involved canonical genes of mature PTC, New-PT clusters and proliferative PTCs. *Xist* is a female specific gene

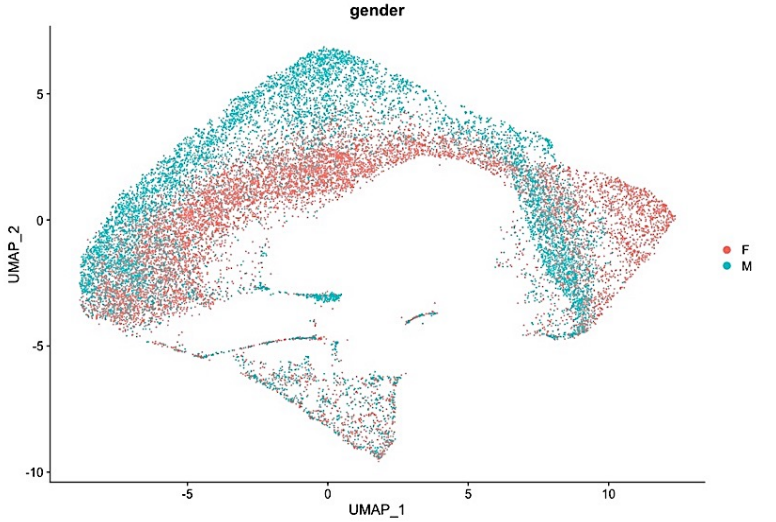
The result of PTC re-clustering provided a clear view of difference between different ages and gender in mature PTCs (Figure 6.21). The mature PTCs of 2-week-old mice from both sexes were in the middle of the UMAP plot whereas cells from the 4-week-old male and 4-week-old female were localised on the two sides of the UMAP plot. From this I inferred that gender difference became more significant at 4-weeks-old. Sex-specific clusters were detected after re-clustering. Cluster 0, 3, 9, 1, 8 were almost specific to female and cluster 7, 12, 5, 19 were almost specific to male. The proliferative clusters comprised cells from both male and female.

The re-clustering was also performed on *Pdgfbr+* clusters, involving fibroblast mesangial cell and JG cells. Important differences between male and female were not evident in these cell types (Figure 6.22).

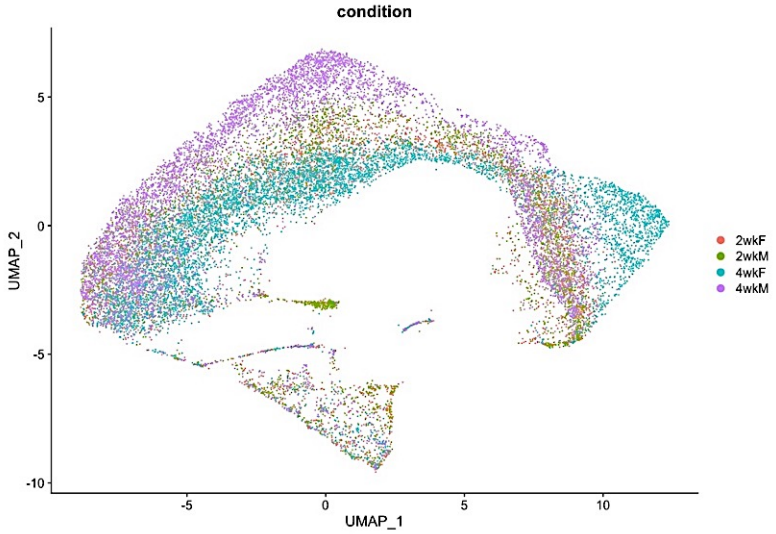
A


age
UMAP_2
UMAP_1
5
0
-5
-10
-5
0
5
10
2wk
4wk

B


gender
UMAP_2
UMAP_1
5
0
-5
-10
-5
0
5
10
F
M

C


condition
UMAP_2
UMAP_1
5
0
-5
-10
-5
0
5
10
2wkF
2wkM
4wkF
4wkM

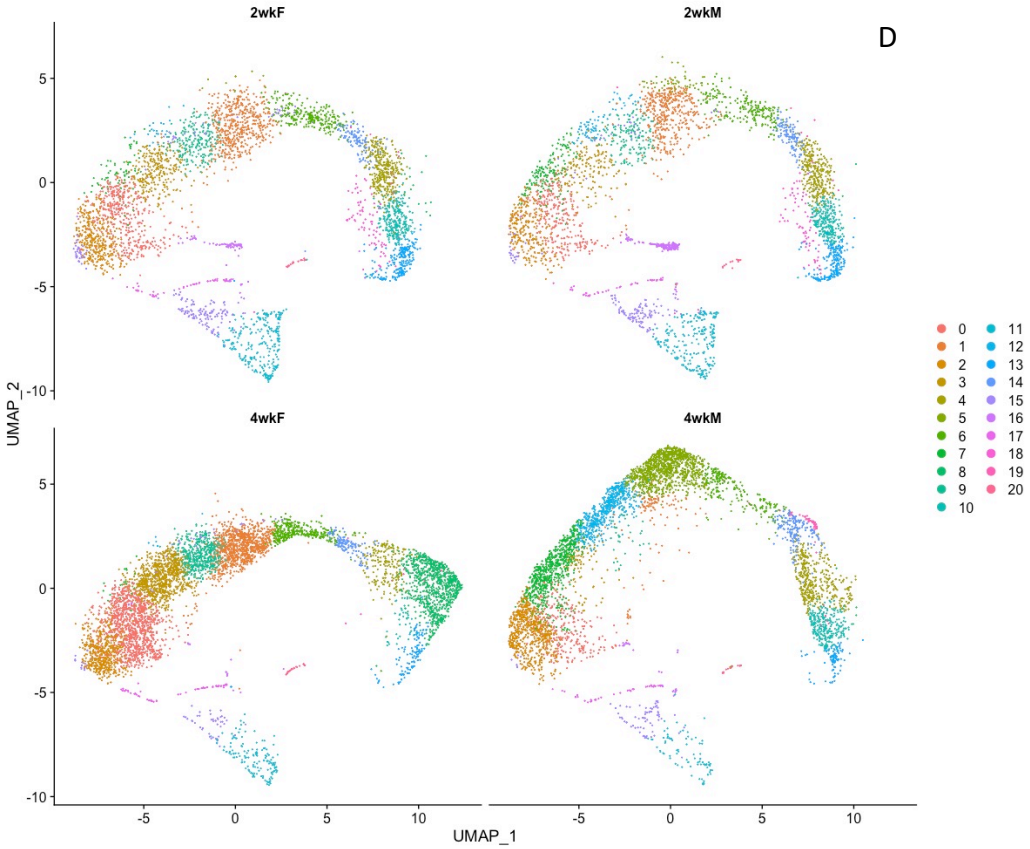


Figure 6.21 UMAP of PTCs in different conditions. Cells on the UMAP plot was coloured by (A) age, which is 2-week-old versus 4-week-old and (B) gender, which is female versus male and (C) condition, which includes both age and gender (D) The UMAP was split by age which shows relative even cell number of each condition.

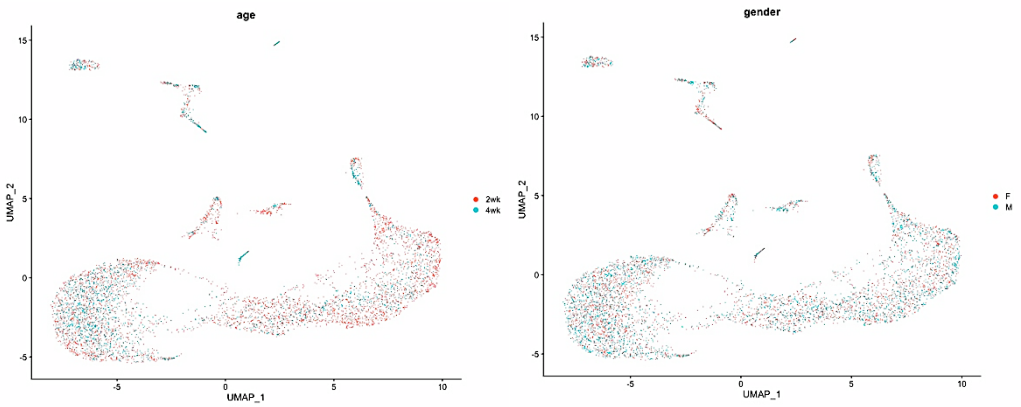


Figure 6.22 Result of re-clustering of the *Pdgfbr+* cells. The mesangail cell, fibroblast and JG cell were selected and re-clustered. The result is visualized using a UMAP plot. Unlike the PTCs, differences in cells according to age or gender were not evident.

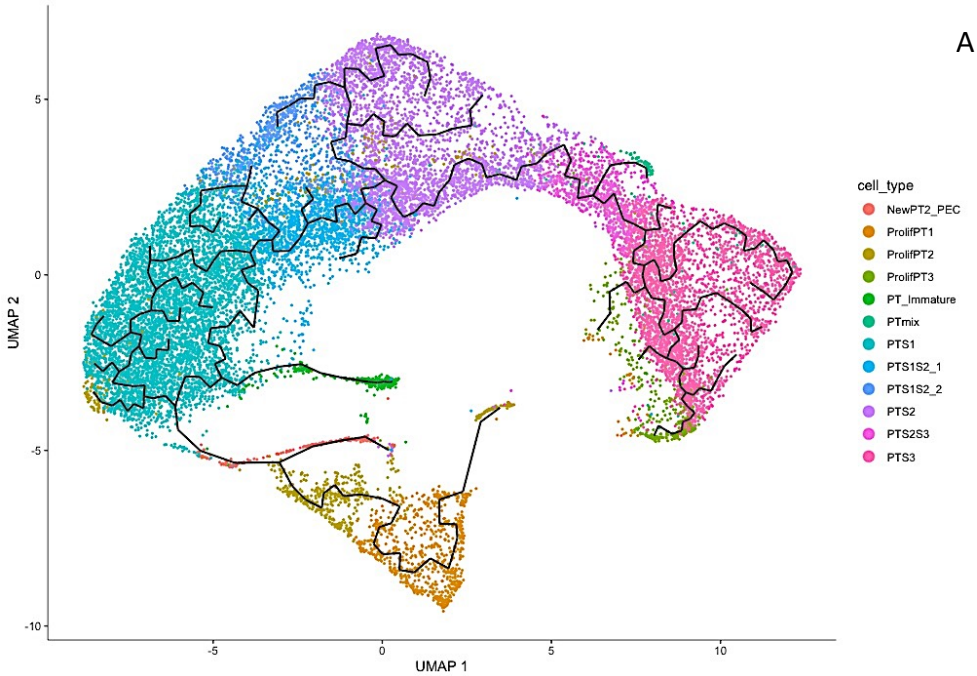
6.2.7 Trajectory and velocity analysis of PTCs

To understand the potential transition paths of cells and estimate the pathways of cell differentiation, trajectory and velocity analysis of PTCs were performed. Trajectory showed a continuous change in RNA profile from proliferative PTCs to normal PTCs, ordered along the anatomical axis from S1 to S3 tubular segments (Figure 6.23A). The result was quite similar to the trajectory analysis of PTCs from adult health/AAN mice in Chapter 4.

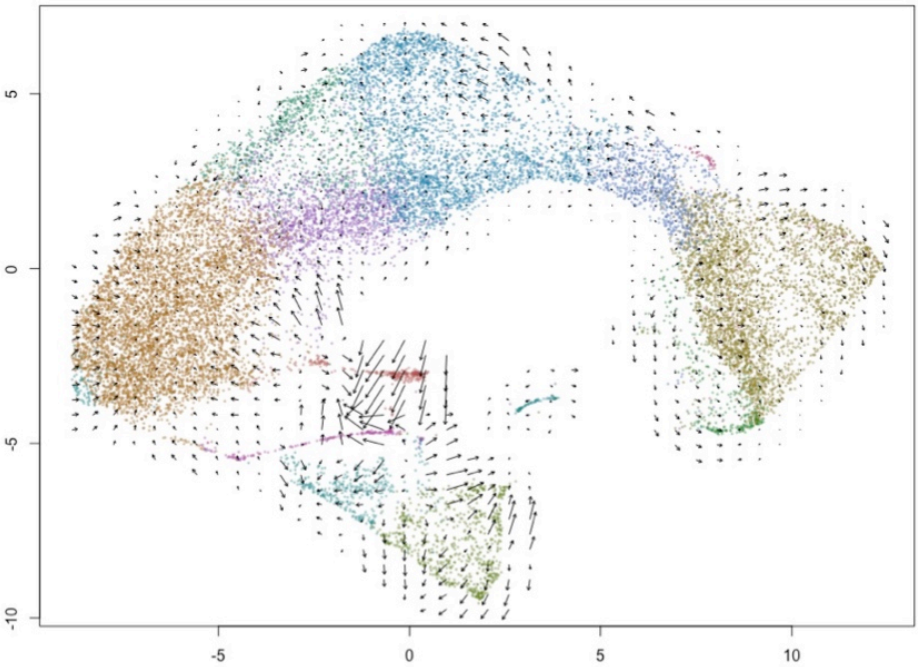
The RNA velocity analysis was then performed, which infers the direction and the rate of transcriptional reprogramming over a time course of hours, using the ratio of unspliced to spliced RNA. PT_Immature showed strong directional change towards NewPT2_PEC and proliferative PTCs, followed by mature PTC clusters (Figure 6.23B). The dynamic change of the ProlifPTs implies rapid proliferation before entering mature differentiated status in growing kidneys. Result of velocity analysis provided a possible cellular transition path “PT_Immature - NewPT2_PEC - ProlifPT - mature PTCs”.

The Pseudotime analysis was perform with time zero set at PT_Immature and NewPT2_PEC based on the results of the RNA velocity analysis. An ordered progression of cell states in pseudotime was seen, from states through anatomically distinct tubular segments (Fig 6.23C).

A


cell_type
NewPT2_PEC
ProlifPT1
ProlifPT2
ProlifPT3
PT_Immature
PTmix
PTS1
PTS1S2_1
PTS1S2_2
PTS2
PTS2S3
PTS3
UMAP 1
UMAP 2
5
0
-5
-10
-5
0
5
10

B


5
0
-5
-10
-5
0
5
10

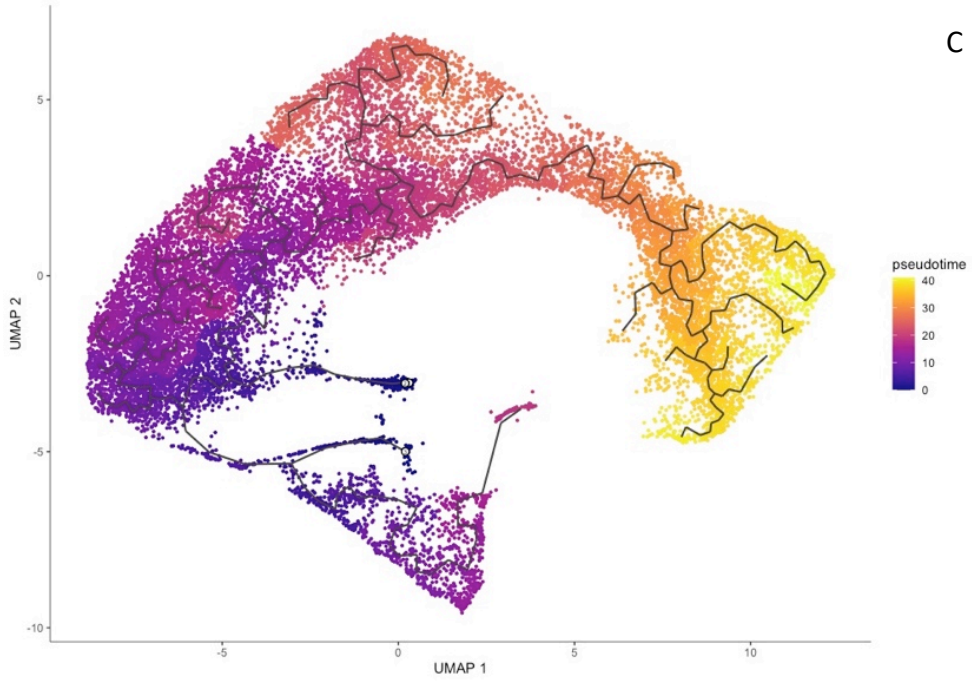


Figure 6.23 Trajectory, RNA velocity and pseudotime analysis of the PTC of growing mouse kidneys (A) RNA expression profiles change continuously from ProlifPT to the anatomic axis in mature PTCs (S1 to S3). (B) RNA velocity analysis shows the cellular transition path "PT unknown - NewPT2_PEC - ProlifPT - mature PTCs". (C) PTC pseudotime analysis.

6.3 Discussion

The cellular composition and RNA expression of growing mouse kidneys in both sexes was investigated at a single-cell level in this chapter. After optimizing the kidney nuclear isolation protocol from young mice, snRNA-Seq of 2 and 4 weeks old mouse kidney was carried out. In this experiment, the 2-week-old mouse kidneys had a larger proportion of proliferative cells and the male to female difference of PTCs became significant in the 4-week-old mouse kidneys.

The AAN experiment included 23,885 nuclei, whereas 41,789 nuclei went into the final analysis in this growing kidney experiment. Also, the number of genes detected per nucleus increased from a median of 1,207 genes per nucleus in the AAN experiment to 2,289. This might represent the benefit from changing the sequencing system from NextSeq 550 to NovaSeq and the improvement of nuclear isolation skill. Taking both parameters together, this snRNA-seq dataset of growing mouse kidneys had nearly 4 times the resolution, compared to the AAN dataset.

The data integration method in this chapter, Harmony, was used to generate the plots of PCA embedding value which were employed to review and decide variables for correction. Two variables, age (with sample preparation) and gender, were included in this experiment, but only age significantly contributed to the difference of PCA embedding. After Harmony adjustment, the dataset was well-integrated, with the difference of proliferative cells between ages and the sex difference of PTCs preserved.

Mature PTCs were segregated to 6 clusters when analyzing the whole dataset, labelling as S1, S1-S2_1, S1S2_2, S2, S2-S3 and S3. After re-clustering of cells assigned to specific cell-type with Seurat, PTCs but not *Pdgfbr+* cells presented sex-specific clusters, suggesting that male to female difference are more prominent in PTCs, which may relate to the protective mechanism of female from AKI episode.. Female and male kidneys exhibit different responses to AKI and differences in AKI to CKD transition (Lima-Posada et al. 2017). Sex hormones, namely the presence of testosterone or absence of estrogen, are linked to the susceptibility to AKI (Park et al. 2004; Kang et al. 2014). Sex differences of renal transporters on PTCs may contribute to the molecular mechanisms of AKI susceptibility of male (Hosszu et al. 2020). In this experiment, sex specific PTC clusters in different segments were identified using snRNA-seq. This provided a dataset for identifying and comparing the expression of sex-specific genes in different segments. It is important to validate the PTC clusters identified here. High-resolution confocal microscopy, RNA scope or spatial transcriptomic single-cell RNA sequencing can be considered as tools for validation.

The result suggested that PTC proliferation makes the main contribution to kidney growth. Tubular elongation derived from flow-induced shear stress and modulated by mesenchymal, was proved to be the major process of kidney enlargement during growth and maturation (Conrad et al. 2021). Three proliferative PTC clusters were identified, labelled as ProlifPT1, ProlifPT2 and ProlifPT3 and identified as G2M, S, S phase in cell cycle analysis. Compared to the

4-week-old kidneys, the 2-week-old kidneys had a larger proportion of all three ProlifPT clusters and the PT_Immature, indicating the decrease of PTC proliferation in 4-week-old kidneys. In velocity analysis, the strong directional change of PT_Immature cells towards NewPT2_PEC and proliferative PTCs was evident. This may also imply that PT_Immature might be the beginning of the PTC differentiation and proliferation pathway. Velocity analysis also showed the cellular transition of mature PTCs toward PT_Immature. This was compatible with the literature that fully differentiated PTCs contribute to the new epithelium in both tubule elongation during kidney growth and tubule regeneration after AKI (Kusaba et al. 2014). The result raised a possible path of PTC proliferation, “mature PTCs - PT_Immature - NewPT2_PEC - ProlifPTs - newly differentiated mature PTCs”. Future work includes validation and localisation of the PCT phenotypes identified in growing mouse kidney and further investigation of the PTC differentiation path.

Apart from validation and localisation, “real time” instead of “pseudotime” analysis of cellular proliferation and sex difference is a valuable approach to consider in future work. A further experiment that includes healthy mouse kidneys from earlier stages of growth and also adult tissue may help to definitively disclose the path of PTC proliferation and further elucidate sex difference of PTCs.

Chapter 7

General Discussion

CKD is a worldwide health issue associated with high mobility and mortality. The mechanisms of CKD development and AKI to CKD transition have been investigated using different approaches, and PTCs were considered the common pathway of AKI and CKD. Single-cell sequencing technique is a newly developed tool that discloses gene expression profiles at the single-cell level. Compared to the traditional bulk RNA-Seq that shows the average signals from whole tissue, scRNA-Seq provides information from individual cells that helps in understanding tissue heterogeneity. There are now global efforts to build a cell atlas using scRNA-Seq, showing the cellular composition of healthy and diseased kidneys and other organs, and the intercellular communication between different cells.

The hypothesis of my research was that there was underappreciated heterogeneity in PTC phenotypes. PTCs play a crucial role in determining the fate of kidneys after kidney injury and in the CKD status, but much of the experimental data pertaining to their biology and responses in kidney disease considers them *en masse*. Therefore, snRNA-Seq was used to investigate PTC heterogeneity, aiming to clarify PTC phenotypes in healthy and fibrotic kidneys, and to study the cellular differentiation pathways of PTC recovery after kidney injury.

I optimised experimental techniques for nuclear isolation from both fresh harvested and frozen preserved mouse kidneys. Minimising the mechanical destruction steps and rapid processing of the nuclei improved the nuclear RNA quality, and is likely to have been an important contributor to the better resolution of PTC phenotypes than in previous studies. PTCs have high oxygen

demand; therefore, they are highly susceptible to ischemic injury in AKI. Laborious preparative approaches for making cell / nuclei suspension are likely to induce hypoxic and inflammatory signals and disturbance the metabolic pathways, disturbing data interpretation.

Mature PTs are described as having three continuous segments based on morphology and localisation. Segment-specific transporters and susceptibility to injury have been described in the literature. In my experiment, five rather than three mature PTC phenotypes were identified in mouse kidneys, in which transitional phenotypes between S1-S2 and S2-S3 were also identified. In addition, four new classes of PTCs were discovered. The new PTC clusters were validated using confocal microscopy, including the proliferative PTCs and three dedifferentiated PTC phenotypes. The new-PT clusters were found at significantly increased abundance in the fibrotic kidney. These rare but distinct PTC phenotypes are proposed to be the key node of recovery versus fibrosis of proximal tubular damage.

Using trajectory and RNA velocity analysis, two major processes of PTC transition and differentiation were described, the path toward cellular senescence and the path toward tubular regeneration (Figure 7.1). The injured PTCs might dedifferentiate to New-PT1 and enter either the “NewPT1 - NewPT2 - proliferative PT - normal PTCs” path toward PTC regeneration or the “NewPT1 - NewPT3” path toward failed repair. Persistent existence of the senescent PTCs in the kidney was associated with CKD progression, which was also described in ageing kidney and

post-AKI status (Humphreys et al. 2013). Inhibition of the senescent PTCs restores the ability of tubular regeneration and improves renal fibrosis after injury (Mylonas et al. 2021). Ameliorating the cellular senescence or enhancing the tubular regeneration by targeting the new PTC phenotypes can be a potential treatment for AKI and CKD.

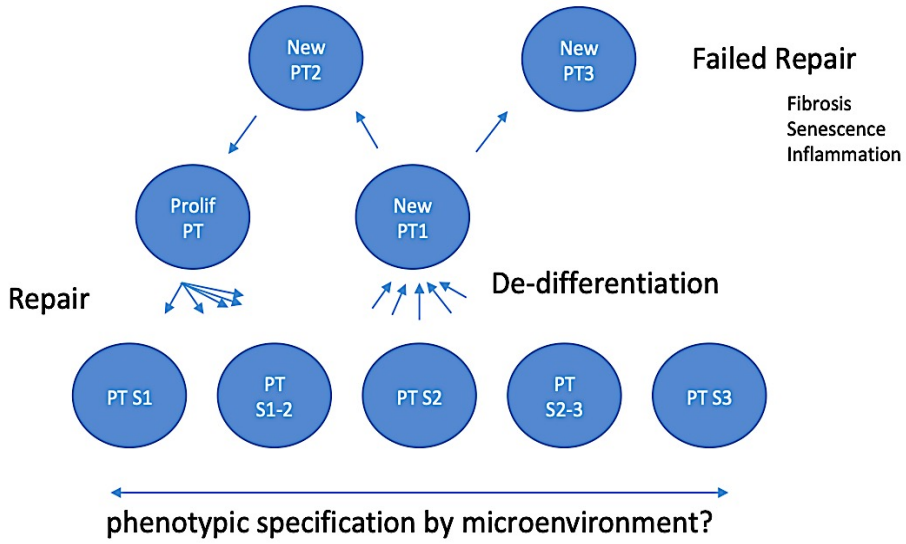


Figure 7.1 Proposed cellular transitions of PTCs following injury. The trajectory and velocity analysis following AA exposure provided a potential path of PTC transition. The injured mature PTCs may dedifferentiate to New-PT1 and enter either the "NewPT1 - NewPT2 - proliferative PT - normal PTCs" path toward PTC regeneration. Alternatively, they may enter the "NewPT1 - NewPT3" path culminating in failed repair, fibrosis, senescence, and/or inflammation.

The characteristics of the proliferative PTC and the dedifferentiated regenerating PTC phenotype helped to understand the PTC regeneration process. My experiments have confirmed the PTC heterogeneity and have pointed to PTCs differentiation pathways and the presence of dedifferentiated cells linked through trajectory analysis to proliferating ones in normal and injured kidneys. However, limited number of dedifferentiated or proliferating cells were seen in mature healthy kidney. The mouse nephron number was fixed within 2 days after birth but the kidneys still exhibit significant growth over the period to maturity. Tubular elongation mainly contributes to the kidney size enlargement. Tubular elongation in the growing kidneys and PTC regeneration may share a common mechanism of PTC proliferation. For further investigation of PTC proliferation and regeneration, I performed snRNA-Seq using 2 and 4-week-old growing kidneys from juvenile healthy mice.

In this experiment, I included both male and female mice to investigate the sex difference and the potential mechanism protecting females from kidney injury. Proliferative cells of the major cell types were identified. The proliferative PTCs were segregated into three distinct clusters by the cell cycle status (S or M phase) and the status of cellular transition toward S1/S2 or S3. An increased proportion of PTCs in the 4-week-old kidneys suggested that PTCs contribute the majority of renal size enlargement during growth. The sex difference of PTCs was not significant in the 2-week-old kidneys but became significant in the 4-week-old kidneys. My future work will focus on PTC proliferation and differentiation in

growing mouse kidneys, in order to validate the PTC phenotypes identified in the growing kidney snRNA-Seq. Furthermore, snRNA-Seq of 1 and 12-week-old kidney will be carried out, to complete a data repository of cells found in healthy mouse kidneys from young age to adult. This can also provide a comprehensive view of cellular differentiation during PTC elongation.

Overall, my work identifies new PT phenotypes and links different phenotypes to kidney recovery versus fibrosis. These new PTC phenotypes were associated with physiological tubular proliferation in growing mice and senescent PTCs in aging kidneys, which built up a hypothesis of PTCs differentiation in both physiological and pathological conditions. Bioinformatic analysis of intracellular pathways and receptor-ligand interaction generated hypotheses of biological changes of a heterogeneous cellular population in fibrosis, that will require further study (Baek and Lee 2020). The differentially expressed genes in male and female PT might be linked to the renoprotective effect of female sex from AKI and CKD (Neugarten and Golestaneh 2018,2022). Male or female specific PTC phenotypes identified in the growing mice dataset could be referred to the AAN experiment in the Chapter 4 and predict candidate genes / pathways that associated with fibrosis progression or recovery.

PCT heterogeneity has been confirmed using single-cell sequencing in human study (Wilson et al. 2019; Muto et al. 2021). The resolution of PTC phenotypes in human single-cell sequencing was limited, which might be related to using samples from nephrectomy rather than renal biopsy. Prolonged ischemic time of

the resected kidney may impair RNA quality. With ethical approval in place, use of human renal biopsy samples for scRNA-Seq / snRNA-Seq experiments may improve data quality. Also, in my work to date, the path of PTC differentiation was predicted using RNA expression profiles in animal models. However, the cellular and nuclear RNA expression found in a differentiating cell may change rapidly. The velocity analysis that I performed used the ratio of spliced and unspliced RNA to predict the speed and direction of cellular transition. An important limitation of this approach is that it extrapolates from splicing events occurring over hours (La Manno et al. 2018). Animal models are more constrained in timescale, but human CKD develops over years to decades. Further studies are therefore required to investigate the link between PTC phenotypes and CKD development in patients.

There are several other limitations of snRNA-Seq. The lack of spatial information, as mentioned in chapter 5, can be improved by microscopic validation. Alternatively, spatial transcription technology provides regional expression profiles and localisation, which addresses this key deficiency of single-cell experiments (Dixon et al. 2022). However, current state of the art spatial approaches are limited in their resolution -as technology improves, further gains in spatial resolution towards the single cell level can be anticipated. Depth of coverage and bioinformatics analysis may remain as challenges for the application of spatial transcriptomics. Sequencing depth for single-cell approaches is typically a few thousand genes. This is a limitation for single-cell approaches generally when compared to bulk sequencing. Lack of sequencing depth may lead single-

cell sequencing to bias against rare but important transcripts, e.g. transcription factors. Combining scRNA-Seq / snRNA-Seq with assay for transposase-accessible chromatin using sequencing (ATAC-Seq), which investigates genome-wide chromatin accessibility, may be a key advance in this regard.

Taken together, my studies have identified and validated new PTC phenotypes associated with PTC regeneration and renal fibrosis. PTC proliferation was further investigated using growing mouse kidneys. My future research work will focus on identification and validation of proliferation-associated PTC phenotypes, and further study of the PTC regeneration pathway.

Reference

Mouse Genome Informatics Gene Expression Data (http://www.informatics.jax.org/assay/MGI:5542047).

10x Genomics, I. 2018. Isolation of Nuclei for Single Cell RNA Sequencing. (Rev D), Available at: https://support.10xgenomics.com/permalink/1dIB6Z91VqCIgUmSC2OM8k.

Abbate, M. et al. 1999. Expression of NCAM recapitulates tubulogenic development in kidneys recovering from acute ischemia. *Am J Physiol* 277(3), pp. F454-463. doi: 10.1152/ajprenal.1999.277.3.F454

Baek, S. and Lee, I. 2020. Single-cell ATAC sequencing analysis: From data preprocessing to hypothesis generation. *Comput Struct Biotechnol J* 18, pp. 1429-1439. doi: 10.1016/j.csbj.2020.06.012

Bankhead, P. et al. 2017. QuPath: Open source software for digital pathology image analysis. *Sci Rep* 7(1), p. 16878. doi: 10.1038/s41598-017-17204-5

Berg JM, T. J., Stryer L. 2002. *Biochemistry. 5th edition. New York: W H Freeman; 2002. Section 30.2, Each Organ Has a Unique Metabolic Profile. Available from: https://www.ncbi.nlm.nih.gov/books/NBK22436/*.

Bertram, J. F. et al. 2011. Human nephron number: implications for health and disease. *Pediatr Nephrol* 26(9), pp. 1529-1533. doi: 10.1007/s00467-011-1843-8

Blighe, K., S Rana, and M Lewis. 2018. *EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling*. Available at: https://github.com/kevinblighe [Accessed.

Bonventre, J. V. 2014. Primary proximal tubule injury leads to epithelial cell cycle arrest, fibrosis, vascular rarefaction, and glomerulosclerosis. *Kidney Int Suppl*

(2011) 4(1), pp. 39-44. doi: 10.1038/kisup.2014.8

Breljak, D. et al. 2016. Distribution of organic anion transporters NaDC3 and OAT1-3 along the human nephron. *Am J Physiol Renal Physiol* 311(1), pp. F227-238. doi: 10.1152/ajprenal.00113.2016

Broer, S. and Palacin, M. 2011. The role of amino acid transporters in inherited and acquired diseases. *Biochem J* 436(2), pp. 193-211. doi: 10.1042/BJ20101912

Bruno, S. and Darzynkiewicz, Z. 1992. Cell cycle dependent expression and stability of the nuclear protein detected by Ki-67 antibody in HL-60 cells. *Cell Prolif* 25(1), pp. 31-40. doi: 10.1111/j.1365-2184.1992.tb01435.x

Butler, A. et al. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 36(5), pp. 411-420. doi: 10.1038/nbt.4096

Buzhor, E. et al. 2013. Reactivation of NCAM1 defines a subpopulation of human adult kidney epithelial cells with clonogenic and stem/progenitor properties. *Am J Pathol* 183(5), pp. 1621-1633. doi: 10.1016/j.ajpath.2013.07.034

Canaud, G. et al. 2019. Cyclin G1 and TASCC regulate kidney epithelial cell G2-M arrest and fibrotic maladaptive repair. *Sci Transl Med* 11(476), doi: 10.1126/scitranslmed.aav4754

Cao, J. et al. 2018. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* 361(6409), pp. 1380-1385. doi: 10.1126/science.aau0730

Cao, J. et al. 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566(7745), pp. 496-502. doi: 10.1038/s41586-019-0969-x

Castrop, H. 2019. The Role of Renal Interstitial Cells in Proximal Tubular

Regeneration. *Nephron* 141(4), pp. 265-272. doi: 10.1159/000496278

Ceovic, S. et al. 1992. Epidemiology of Balkan endemic nephropathy. *Food Chem Toxicol* 30(3), pp. 183-188. doi: 10.1016/0278-6915(92)90031-f

Chang-Panesso, M. et al. 2019. FOXM1 drives proximal tubule proliferation during repair from acute ischemic kidney injury. *J Clin Invest* 129(12), pp. 5501-5517. doi: 10.1172/JCI125519

Chang-Panesso, M. et al. 2018. Meis1 is specifically upregulated in kidney myofibroblasts during aging and injury but is not required for kidney homeostasis or fibrotic response. *Am J Physiol Renal Physiol* 315(2), pp. F275-F290. doi: 10.1152/ajprenal.00030.2018

Charlton, J. R. et al. 2021. Nephron number and its determinants: a 2020 update. *Pediatr Nephrol* 36(4), pp. 797-807. doi: 10.1007/s00467-020-04534-2

Chen, C. H. et al. 2012. Aristolochic acid-associated urothelial cancer in Taiwan. *Proc Natl Acad Sci U S A* 109(21), pp. 8241-8246. doi: 10.1073/pnas.1119920109

Chen, L. et al. 2019a. Renal-Tubule Epithelial Cell Nomenclature for Single-Cell RNA-Sequencing Studies. *J Am Soc Nephrol* 30(8), pp. 1358-1364. doi: 10.1681/ASN.2019040415

Chen, S. et al. 2019b. Tenascin-C protects against acute kidney injury by recruiting Wnt ligands. *Kidney Int* 95(1), pp. 62-74. doi: 10.1016/j.kint.2018.08.029

Chen, S. J. et al. 2020. Crosstalk between tubular epithelial cells and glomerular endothelial cells in diabetic kidney disease. *Cell Prolif* 53(3), p. e12763. doi: 10.1111/cpr.12763

Chen, Y. et al. 2021. Emerging role of air pollution in chronic kidney disease.

Environ Sci Pollut Res Int 28(38), pp. 52610-52624. doi: 10.1007/s11356-021-16031-6

Chevalier, R. L. 2016. The proximal tubule is the primary target of injury and progression of kidney disease: role of the glomerulotubular junction. *Am J Physiol Renal Physiol* 311(1), pp. F145-161. doi: 10.1152/ajprenal.00164.2016

Cianciolo Cosentino, C. et al. 2013. Histone deacetylase inhibitor enhances recovery after AKI. *J Am Soc Nephrol* 24(6), pp. 943-953. doi: 10.1681/ASN.2012111055

Clark, J. Z. et al. 2019. Representation and relative abundance of cell-type selective markers in whole-kidney RNA-Seq data. *Kidney Int* 95(4), pp. 787-796. doi: 10.1016/j.kint.2018.11.028

Combes, A. N. et al. 2019. Single cell analysis of the developing mouse kidney provides deeper insight into marker gene expression and ligand-receptor crosstalk. *Development* 146(12), doi: 10.1242/dev.178673

Conrad, L. et al. 2021. The biomechanical basis of biased epithelial tube elongation in lung and kidney development. *Development* 148(9), doi: 10.1242/dev.194209

D'Agati, V. D. and Shankland, S. J. 2019. Recognizing diversity in parietal epithelial cells. *Kidney Int* 96(1), pp. 16-19. doi: 10.1016/j.kint.2019.02.036

Dantzler, W. H. and Wright, S. H. 2003. The molecular and cellular physiology of basolateral organic anion transport in mammalian renal tubules. *Biochim Biophys Acta* 1618(2), pp. 185-193. doi: 10.1016/j.bbamem.2003.08.015

Das, N. A. et al. 2020. Empagliflozin reduces high glucose-induced oxidative stress and miR-21-dependent TRAF3IP2 induction and RECK suppression, and inhibits human renal proximal tubular epithelial cell migration and epithelial-to-

mesenchymal transition. *Cell Signal* 68, p. 109506. doi: 10.1016/j.cellsig.2019.109506

Davidson, A. J. 2008. Mouse kidney development.*StemBook*. Cambridge (MA).

Denic, A. et al. 2017. The Substantial Loss of Nephrons in Healthy Human Kidneys with Aging. *J Am Soc Nephrol* 28(1), pp. 313-320. doi: 10.1681/ASN.2016020154

Desjardins, P. and Conklin, D. 2010. NanoDrop microvolume quantitation of nucleic acids. *J Vis Exp* (45), doi: 10.3791/2565

Detrisac, C. J. et al. 1984. Tissue culture of human kidney epithelial cells of proximal tubule origin. *Kidney Int* 25(2), pp. 383-390. doi: 10.1038/ki.1984.28

Dixon, E. E. et al. 2022. Spatially Resolved Transcriptomic Analysis of Acute Kidney Injury in a Female Murine Model. *J Am Soc Nephrol* 33(2), pp. 279-289. doi: 10.1681/ASN.2021081150

Docherty, M. H. et al. 2019. Cellular Senescence in the Kidney. *J Am Soc Nephrol* 30(5), pp. 726-736. doi: 10.1681/ASN.2018121251

Duffield, J. S. 2014. Cellular and molecular mechanisms in kidney fibrosis. *J Clin Invest* 124(6), pp. 2299-2306. doi: 10.1172/JCI72267

Eickelberg, O. et al. 2002. Betaglycan inhibits TGF-beta signaling by preventing type I-type II receptor complex formation. Glycosaminoglycan modifications alter betaglycan function. *J Biol Chem* 277(1), pp. 823-829. doi: 10.1074/jbc.M105110200

Eleftheriadis, T. et al. 2021. A Role for Human Renal Tubular Epithelial Cells in Direct Allo-Recognition by CD4+ T-Cells and the Effect of Ischemia-Reperfusion. *Int J Mol Sci* 22(4), doi: 10.3390/ijms22041733

Falconer, D. S. et al. 1978. Cell numbers and cell sizes in organs of mice selected for large and small body size. *Genet Res* 31(3), pp. 287-301. doi: 10.1017/s0016672300018061

Fattah, H. and Vallon, V. 2018. Tubular Recovery after Acute Kidney Injury. *Nephron* 140(2), pp. 140-143. doi: 10.1159/000490007

Ferenbach, D. A. and Bonventre, J. V. 2015. Mechanisms of maladaptive repair after AKI leading to accelerated kidney ageing and CKD. *Nat Rev Nephrol* 11(5), pp. 264-276. doi: 10.1038/nrneph.2015.3

Ferrara, N. 2004. Vascular endothelial growth factor: basic science and clinical progress. *Endocr Rev* 25(4), pp. 581-611. doi: 10.1210/er.2003-0027

Forbes, M. S. et al. 2011. Proximal tubular injury and rapid formation of atubular glomeruli in mice with unilateral ureteral obstruction: a new look at an old model. *Am J Physiol Renal Physiol* 301(1), pp. F110-117. doi: 10.1152/ajprenal.00022.2011

Fujigaki, Y. et al. 2006. Kinetics and characterization of initially regenerating proximal tubules in S3 segment in response to various degrees of acute tubular injury. *Nephrol Dial Transplant* 21(1), pp. 41-50. doi: 10.1093/ndt/gfi035

Garg, P. 2018. A Review of Podocyte Biology. *Am J Nephrol* 47 Suppl 1, pp. 3-13. doi: 10.1159/000481633

Gewin, L. and Zent, R. 2012. How does TGF-beta mediate tubulointerstitial fibrosis? *Semin Nephrol* 32(3), pp. 228-235. doi: 10.1016/j.semnephrol.2012.04.001

Ghezzi, C. et al. 2018. Physiology of renal glucose handling via SGLT1, SGLT2 and GLUT2. *Diabetologia* 61(10), pp. 2087-2097. doi: 10.1007/s00125-018-4656-5

Gilbert, R. E. 2017. Proximal Tubulopathy: Prime Mover and Key Therapeutic Target in Diabetic Kidney Disease. *Diabetes* 66(4), pp. 791-800. doi: 10.2337/db16-0796

Grande, M. T. et al. 2015. Snail1-induced partial epithelial-to-mesenchymal transition drives renal fibrosis in mice and can be targeted to reverse established disease. *Nat Med* 21(9), pp. 989-997. doi: 10.1038/nm.3901

Gu, Z. et al. 2014. circlize Implements and enhances circular visualization in R. *Bioinformatics* 30(19), pp. 2811-2812. doi: 10.1093/bioinformatics/btu393

He, L. et al. 2017. AKI on CKD: heightened injury, suppressed repair, and the underlying mechanisms. *Kidney Int* 92(5), pp. 1071-1083. doi: 10.1016/j.kint.2017.06.030

Hegde, A. 2020. Diuretics in Acute Kidney Injury. *Indian J Crit Care Med* 24(Suppl 3), pp. S98-S99. doi: 10.5005/jp-journals-10071-23406

Heung, M. et al. 2016. Acute Kidney Injury Recovery Pattern and Subsequent Risk of CKD: An Analysis of Veterans Health Administration Data. *Am J Kidney Dis* 67(5), pp. 742-752. doi: 10.1053/j.ajkd.2015.10.019

Higashi, A. Y. et al. 2019. Expression Profiling of Fibroblasts in Chronic and Acute Disease Models Reveals Novel Pathways in Kidney Fibrosis. *J Am Soc Nephrol* 30(1), pp. 80-94. doi: 10.1681/ASN.2018060644

Hirst, J. A. et al. 2020. Prevalence of chronic kidney disease in the community using data from OxRen: a UK population-based cohort study. *Br J Gen Pract* 70(693), pp. e285-e293. doi: 10.3399/bjgp20X708245

Hosszu, A. et al. 2020. Sex differences in renal ischemia-reperfusion injury. *Am J Physiol Renal Physiol* 319(2), pp. F149-F154. doi: 10.1152/ajprenal.00099.2020

Hoy, W. E. et al. 2003. A stereological study of glomerular number and volume: preliminary findings in a multiracial study of kidneys at autopsy. *Kidney Int Suppl* (83), pp. S31-37. doi: 10.1046/j.1523-1755.63.s83.8.x

Huang, L. et al. 2013. Development of a chronic kidney disease model in C57BL/6 mice with relevance to human pathology. *Nephron Extra* 3(1), pp. 12-29. doi: 10.1159/000346180

Humphreys, B. D. et al. 2013. Chronic epithelial kidney injury molecule-1 expression causes murine kidney fibrosis. *J Clin Invest* 123(9), pp. 4023-4035. doi: 10.1172/JCI45361

Hwang, J. S. et al. 2010. Expression of OAT1 and OAT3 in differentiating proximal tubules of the mouse kidney. *Histol Histopathol* 25(1), pp. 33-44. doi: 10.14670/HH-25.33

Jacobson, H. R. 1981. Functional segmentation of the mammalian nephron. *Am J Physiol* 241(3), pp. F203-218. doi: 10.1152/ajprenal.1981.241.3.F203

Jadot, I. et al. 2017. An Integrated View of Aristolochic Acid Nephropathy: Update of the Literature. *Int J Mol Sci* 18(2), doi: 10.3390/ijms18020297

Jankowski, J. et al. 2018. Epithelial and Endothelial Pannexin1 Channels Mediate AKI. *J Am Soc Nephrol* 29(7), pp. 1887-1899. doi: 10.1681/ASN.2017121306

Jenkins, R. H. et al. 2014. miR-192 induces G2/M growth arrest in aristolochic acid nephropathy. *Am J Pathol* 184(4), pp. 996-1009. doi: 10.1016/j.ajpath.2013.12.028

Johnson, R. J. et al. 1992. The activated mesangial cell: a glomerular "myofibroblast"? *J Am Soc Nephrol* 2(10 Suppl), pp. S190-197.

Kalluri, R. and Weinberg, R. A. 2009. The basics of epithelial-mesenchymal transition. *J Clin Invest* 119(6), pp. 1420-1428. doi: 10.1172/JCI39104

Kamiyama, M. et al. 2012. The establishment of a primary culture system of proximal tubule segments using specific markers from normal mouse kidneys. *Int J Mol Sci* 13(4), pp. 5098-5111. doi: 10.3390/ijms13045098

Kanai, Y. et al. 1994. The human kidney low affinity Na+/glucose cotransporter SGLT2. Delineation of the major renal reabsorptive mechanism for D-glucose. *J Clin Invest* 93(1), pp. 397-404. doi: 10.1172/JCI116972

Kang, K. P. et al. 2014. Effect of gender differences on the regulation of renal ischemia-reperfusion-induced inflammation in mice. *Mol Med Rep* 9(6), pp. 2061-2068. doi: 10.3892/mmr.2014.2089

Karaiskos, N. et al. 2018. A Single-Cell Transcriptome Atlas of the Mouse Glomerulus. *J Am Soc Nephrol* 29(8), pp. 2060-2068. doi: 10.1681/ASN.2018030238

Karbach, U. et al. 2000. Localization of organic cation transporters OCT1 and OCT2 in rat kidney. *Am J Physiol Renal Physiol* 279(4), pp. F679-687. doi: 10.1152/ajprenal.2000.279.4.F679

Kaverina, N. V. et al. 2019. Dual lineage tracing shows that glomerular parietal epithelial cells can transdifferentiate toward the adult podocyte fate. *Kidney Int* 96(3), pp. 597-611. doi: 10.1016/j.kint.2019.03.014

KDIGOAcuteKidneyInjuryWorkGroup. 2012. KDIGO Clinical Practice Guideline for Acute Kidney Injury. *Kidney Int Suppl (2011)* 2(1), pp. 19-36. doi: 10.1038/kisup.2011.32

Kefaloyianni, E. et al. 2019. Proximal Tubule-Derived Amphiregulin Amplifies and Integrates Profibrotic EGF Receptor Signals in Kidney Fibrosis. *J Am Soc Nephrol*

30(12), pp. 2370-2383. doi: 10.1681/ASN.2019030321

Kefaloyianni, E. et al. 2016. ADAM17 substrate release in proximal tubule drives kidney fibrosis. *JCI Insight* 1(13), doi: 10.1172/jci.insight.87023

Kirita, Y. et al. 2020. Cell profiling of mouse acute kidney injury reveals conserved cellular responses to injury. *Proc Natl Acad Sci U S A* 117(27), pp. 15874-15883. doi: 10.1073/pnas.2005477117

Kirk, R. 2013. Renal fibrosis: KIM-1 expression links kidney injury with CKD in mice. *Nat Rev Nephrol* 9(11), p. 627. doi: 10.1038/nrneph.2013.194

Kitada, K. et al. 2014. Hyperglycemia causes cellular senescence via a SGLT2- and p21-dependent pathway in proximal tubules in the early stage of diabetic nephropathy. *J Diabetes Complications* 28(5), pp. 604-611. doi: 10.1016/j.jdiacomp.2014.05.010

Kobayashi, A. et al. 2014. Identification of a multipotent self-renewing stromal progenitor population during mammalian kidney organogenesis. *Stem Cell Reports* 3(4), pp. 650-662. doi: 10.1016/j.stemcr.2014.08.008

Korsunsky, I. et al. 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 16(12), pp. 1289-1296. doi: 10.1038/s41592-019-0619-0

Krishnaswami, S. R. et al. 2016. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc* 11(3), pp. 499-524. doi: 10.1038/nprot.2016.015

Kurz, D. J. et al. 2000. Senescence-associated (beta)-galactosidase reflects an increase in lysosomal mass during replicative ageing of human endothelial cells. *J Cell Sci* 113 (Pt 20), pp. 3613-3622.

Kusaba, T. et al. 2014. Differentiated kidney epithelial cells repair injured proximal tubule. *Proc Natl Acad Sci U S A* 111(4), pp. 1527-1532. doi: 10.1073/pnas.1310653110

La Manno, G. et al. 2018. RNA velocity of single cells. *Nature* 560(7719), pp. 494-498. doi: 10.1038/s41586-018-0414-6

Lamouille, S. et al. 2014. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* 15(3), pp. 178-196. doi: 10.1038/nrm3758

Lebeau, C. et al. 2005. Early proximal tubule injury in experimental aristolochic acid nephropathy: functional and histological studies. *Nephrol Dial Transplant* 20(11), pp. 2321-2332. doi: 10.1093/ndt/gfi042

Lee, J. H. and Berger, J. M. 2019. Cell Cycle-Dependent Control and Roles of DNA Topoisomerase II. *Genes (Basel)* 10(11), doi: 10.3390/genes10110859

Lee, J. W. et al. 2015. Deep Sequencing in Microdissected Renal Tubules Identifies Nephron Segment-Specific Transcriptomes. *J Am Soc Nephrol* 26(11), pp. 2669-2677. doi: 10.1681/ASN.2014111067

Lee, S. et al. 2004. Fanconi's syndrome and subsequent progressive renal failure caused by a Chinese herb containing aristolochic acid. *Nephrology (Carlton)* 9(3), pp. 126-129. doi: 10.1111/j.1440-1797.2003.00232.x

Levin, A., Stevens, P. E., Bilous, R. W., Coresh, J., De Francisco, A. L. M., De Jong, P. E., Griffith, K. E., Hemmelgarn, B. R., Iseki, K., Lamb, E. J., Levey, A. S., Riella, M. C., Shlipak, M. G., Wang, H., White, C. T., & Winearls, C. G. . 2013. Chapter 1: Definition and classification of CKD. *Kidney Int Suppl (2011)* 3(1), pp. 19-62. doi: 10.1038/kisup.2012.64

Li, C. et al. 2020a. Potent inhibitors of organic anion transporter 1 and 3 from natural compounds and their protective effect on aristolochic acid nephropathy.

Toxicol Sci, doi: 10.1093/toxsci/kfaa033

Li, J. et al. 2020b. Renal protective effects of empagliflozin via inhibition of EMT and aberrant glycolysis in proximal tubules. *JCI Insight* 5(6), doi: 10.1172/jci.insight.129034

Liao, Y. et al. 2019. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* 47(W1), pp. W199-W205. doi: 10.1093/nar/gkz401

Lima-Posada, I. et al. 2017. Gender Differences in the Acute Kidney Injury to Chronic Kidney Disease Transition. *Sci Rep* 7(1), p. 12270. doi: 10.1038/s41598-017-09630-2

Lin, F. et al. 2005. Intrarenal cells, not bone marrow-derived cells, are the major source for regeneration in postischemic kidney. *J Clin Invest* 115(7), pp. 1756-1764. doi: 10.1172/JCI23015

Lindstrom, N. O. et al. 2019. Single-Cell RNA Sequencing of the Adult Mouse Kidney: From Molecular Cataloging of Cell Types to Disease-Associated Predictions. *Am J Kidney Dis* 73(1), pp. 140-142. doi: 10.1053/j.ajkd.2018.07.002

Linke, A. et al. 2022. Antigen Cross-Presentation by Murine Proximal Tubular Epithelial Cells Induces Cytotoxic and Inflammatory CD8(+) T Cells. *Cells* 11(9), doi: 10.3390/cells11091510

Little, M. H. and McMahon, A. P. 2012. Mammalian kidney development: principles, progress, and projections. *Cold Spring Harb Perspect Biol* 4(5), doi: 10.1101/cshperspect.a008300

Livingston, M. J. et al. 2016. Persistent activation of autophagy in kidney tubular cells promotes renal interstitial fibrosis during unilateral ureteral obstruction. *Autophagy* 12(6), pp. 976-998. doi: 10.1080/15548627.2016.1166317

Lombardi, D. and Lasagni, L. eds. 2016. *Cell-cycle Alterations in Post-mitotic Cells and Cell Death by Mitotic Catastrophe*.

Long, K. R. et al. 2022. Cubilin-, megalin-, and Dab2-dependent transcription revealed by CRISPR/Cas9 knockout in kidney proximal tubule cells. *Am J Physiol Renal Physiol* 322(1), pp. F14-F26. doi: 10.1152/ajprenal.00259.2021

Lovisa, S. et al. 2015. Epithelial-to-mesenchymal transition induces cell cycle arrest and parenchymal damage in renal fibrosis. *Nat Med* 21(9), pp. 998-1009. doi: 10.1038/nm.3902

Lovisa, S. et al. 2016. Partial Epithelial-to-Mesenchymal Transition and Other New Mechanisms of Kidney Fibrosis. *Trends Endocrinol Metab* 27(10), pp. 681-695. doi: 10.1016/j.tem.2016.06.004

Lu, Y. et al. 2017. Single-cell RNA-sequence analysis of mouse glomerular mesangial cells uncovers mesangial cell essential genes. *Kidney Int* 92(2), pp. 504-513. doi: 10.1016/j.kint.2017.01.016

McGinnis, C. S. et al. 2019. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* 8(4), pp. 329-337 e324. doi: 10.1016/j.cels.2019.03.003

Mei, N. et al. 2016. Chapter 20 - Gene Expression Profiling in Evaluating the Safety and Toxicity of Nutraceuticals**The information in this chapter is not a formal dissemination of information by the US Food and Drug Administration and does not represent agency position or policy. In: Gupta, R.C. ed. *Nutraceuticals*. Boston: Academic Press, pp. 249-262.

Meinild, A. K. et al. 2000. Water transport by the renal Na(+)-dicarboxylate cotransporter. *Am J Physiol Renal Physiol* 278(5), pp. F777-783. doi: 10.1152/ajprenal.2000.278.5.F777

Melincovici, C. S. et al. 2018. Vascular endothelial growth factor (VEGF) - key factor in normal and pathological angiogenesis. *Rom J Morphol Embryol* 59(2), pp. 455-467.

Meran, S. and Steadman, R. 2011. Fibroblasts and myofibroblasts in renal fibrosis. *Int J Exp Pathol* 92(3), pp. 158-167. doi: 10.1111/j.1365-2613.2011.00764.x

Moeller, M. J. and Smeets, B. 2014. Role of parietal epithelial cells in kidney injury: the case of rapidly progressing glomerulonephritis and focal and segmental glomerulosclerosis. *Nephron Exp Nephrol* 126(2), p. 97. doi: 10.1159/000360677

Moujaber, O. et al. 2019. Cellular senescence is associated with reorganization of the microtubule cytoskeleton. *Cell Mol Life Sci* 76(6), pp. 1169-1183. doi: 10.1007/s00018-018-2999-1

Murawski, I. J. et al. 2010. The relationship between nephron number, kidney size and body weight in two inbred mouse strains. *Organogenesis* 6(3), pp. 189-194. doi: 10.4161/org.6.3.12125

Muto, Y. et al. 2021. Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. *Nat Commun* 12(1), p. 2190. doi: 10.1038/s41467-021-22368-w

Mylonas, K. J. et al. 2021. Cellular senescence inhibits renal regeneration after injury in mice, with senolytic treatment promoting repair. *Sci Transl Med* 13(594), doi: 10.1126/scitranslmed.abb0203

Nagamori, S. et al. 2016. Novel cystine transporter in renal proximal tubule identified as a missing partner of cystinuria-related plasma membrane protein rBAT/SLC3A1. *Proc Natl Acad Sci U S A* 113(3), pp. 775-780. doi:

10.1073/pnas.1519959113

Narita, M. et al. 2011. Spatial coupling of mTOR and autophagy augments secretory phenotypes. *Science* 332(6032), pp. 966-970. doi: 10.1126/science.1205407

Neal, B. et al. 2017. Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. *The New England journal of medicine* 377(7), pp. 644-657. doi: 10.1056/NEJMoa1611925

Nespoux, J. et al. 2020. Gene knockout of the Na(+)-glucose cotransporter SGLT2 in a murine model of acute kidney injury induced by ischemia-reperfusion. *Am J Physiol Renal Physiol* 318(5), pp. F1100-F1112. doi: 10.1152/ajprenal.00607.2019

Neugarten, J. and Golestaneh, L. 2018. Female sex reduces the risk of hospital-associated acute kidney injury: a meta-analysis. *BMC Nephrol* 19(1), p. 314. doi: 10.1186/s12882-018-1122-z

Neugarten, J. and Golestaneh, L. 2022. Sex Differences in Acute Kidney Injury. *Semin Nephrol* 42(2), pp. 208-218. doi: 10.1016/j.semnephrol.2022.04.010

Nigam, S. K. et al. 2015. Handling of Drugs, Metabolites, and Uremic Toxins by Kidney Proximal Tubule Drug Transporters. *Clin J Am Soc Nephrol* 10(11), pp. 2039-2049. doi: 10.2215/CJN.02440314

O'Sullivan, E. D. et al. 2019. Complementary Roles for Single-Nucleus and Single-Cell RNA Sequencing in Kidney Disease Research. *J Am Soc Nephrol* 30(4), pp. 712-713. doi: 10.1681/ASN.2019020112

Oliver, J. A. et al. 2009. Proliferation and migration of label-retaining cells of the kidney papilla. *J Am Soc Nephrol* 20(11), pp. 2315-2327. doi: 10.1681/ASN.2008111203

Parekh, S. et al. 2018. zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. *Gigascience* 7(6), doi: 10.1093/gigascience/giy059

Park, J. et al. 2018. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* 360(6390), pp. 758-763. doi: 10.1126/science.aar2131

Park, K. M. et al. 2004. Testosterone is responsible for enhanced susceptibility of males to ischemic renal injury. *J Biol Chem* 279(50), pp. 52282-52292. doi: 10.1074/jbc.M407629200

Petronic, V. J. et al. 1991. Balkan endemic nephropathy and papillary transitional cell tumors of the renal pelvis and ureters. *Kidney Int Suppl* 34, pp. S77-79.

Piwnicka, M. et al. 1983. RNA and DNA content of isolated cell nuclei measured by multiparameter flow cytometry. *Cytometry* 3(4), pp. 269-275. doi: 10.1002/cyto.990030407

Poudel, N. and Okusa, M. D. 2019. Pannexins in Acute Kidney Injury. *Nephron* 143(3), pp. 158-161. doi: 10.1159/000501278

Poulsom, R. and Little, M. H. 2009. Parietal epithelial cells regenerate podocytes. *J Am Soc Nephrol* 20(2), pp. 231-233. doi: 10.1681/ASN.2008121279

Pozdzik, A. A. et al. 2008. Aristolochic acid induces proximal tubule apoptosis and epithelial to mesenchymal transformation. *Kidney Int* 73(5), pp. 595-607. doi: 10.1038/sj.ki.5002714

Qiu, C. et al. 2018. Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease. *Nat Med* 24(11), pp. 1721-1731. doi: 10.1038/s41591-018-0194-4

Ramilowski, J. A. et al. 2015. A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat Commun* 6, p. 7866. doi: 10.1038/ncomms8866

Ransick, A. et al. 2019. Single-Cell Profiling Reveals Sex, Lineage, and Regional Diversity in the Mouse Kidney. *Dev Cell* 51(3), pp. 399-413 e397. doi: 10.1016/j.devcel.2019.10.005

Romanov, V. et al. 2015. Aristolochic acid-induced apoptosis and G2 cell cycle arrest depends on ROS generation and MAP kinases activation. *Arch Toxicol* 89(1), pp. 47-56. doi: 10.1007/s00204-014-1249-z

Rudnicki, M. et al. 2007. Gene expression profiles of human proximal tubular epithelial cells in proteinuric nephropathies. *Kidney Int* 71(4), pp. 325-335. doi: 10.1038/sj.ki.5002043

Ryan, M. J. et al. 1994. HK-2: an immortalized proximal tubule epithelial cell line from normal adult human kidney. *Kidney Int* 45(1), pp. 48-57. doi: 10.1038/ki.1994.6

Sagrinati, C. et al. 2006. Isolation and characterization of multipotent progenitor cells from the Bowman's capsule of adult human kidneys. *J Am Soc Nephrol* 17(9), pp. 2443-2456. doi: 10.1681/ASN.2006010089

Schiessl, I. M. et al. 2018. Renal Interstitial Platelet-Derived Growth Factor Receptor-beta Cells Support Proximal Tubular Regeneration. *J Am Soc Nephrol* 29(5), pp. 1383-1396. doi: 10.1681/ASN.2017101069

Schnaper, H. W. 2017. The Tubulointerstitial Pathophysiology of Progressive Kidney Disease. *Adv Chronic Kidney Dis* 24(2), pp. 107-116. doi: 10.1053/j.ackd.2016.11.011

Schutte-Nutgen, K. et al. 2019. Getting a Notch closer to renal dysfunction:

activated Notch suppresses expression of the adaptor protein Disabled-2 in tubular epithelial cells. *FASEB J* 33(1), pp. 821-832. doi: 10.1096/fj.201800392RR

Seron, D. et al. 1991. Expression of VCAM-1 in the normal and diseased kidney. *Nephrol Dial Transplant* 6(12), pp. 917-922. doi: 10.1093/ndt/6.12.917

Shimizu, H. et al. 2011. NF-kappaB plays an important role in indoxyl sulfate-induced cellular senescence, fibrotic gene expression, and inhibition of proliferation in proximal tubular cells. *Am J Physiol Cell Physiol* 301(5), pp. C1201-1212. doi: 10.1152/ajpcell.00471.2010

Slane, D. et al. 2015. Profiling of embryonic nuclear vs. cellular RNA in Arabidopsis thaliana. *Genom Data* 4, pp. 96-98. doi: 10.1016/j.gdata.2015.03.015

Smeets, B. et al. 2013. Proximal tubular cells contain a phenotypically distinct, scattered cell population involved in tubular regeneration. *J Pathol* 229(5), pp. 645-659. doi: 10.1002/path.4125

Stark, R. et al. 2019. RNA sequencing: the teenage years. *Nat Rev Genet* 20(11), pp. 631-656. doi: 10.1038/s41576-019-0150-2

Stevens, P. E. et al. 2013. Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. *Ann Intern Med* 158(11), pp. 825-830. doi: 10.7326/0003-4819-158-11-201306040-00007

Stewart, B. J. et al. 2019. Spatiotemporal immune zonation of the human kidney. *Science* 365(6460), pp. 1461-1466. doi: 10.1126/science.aat5031

Strober, W. 2015. Trypan Blue Exclusion Test of Cell Viability. *Curr Protoc Immunol* 111, pp. A3 B 1-A3 B 3. doi: 10.1002/0471142735.ima03bs111

Stuart, T. et al. 2019. Comprehensive Integration of Single-Cell Data. *Cell* 177(7), pp. 1888-1902 e1821. doi: 10.1016/j.cell.2019.05.031

Sturmlechner, I. et al. 2017. Cellular senescence in renal ageing and disease. *Nat Rev Nephrol* 13(2), pp. 77-89. doi: 10.1038/nrneph.2016.183

Svensson, V. et al. 2017. Power analysis of single-cell RNA-sequencing experiments. *Nat Methods* 14(4), pp. 381-387. doi: 10.1038/nmeth.4220

Takaori, K. et al. 2016. Severity and Frequency of Proximal Tubule Injury Determines Renal Prognosis. *J Am Soc Nephrol* 27(8), pp. 2393-2406. doi: 10.1681/ASN.2015060647

Tan, R. J. et al. 2016. Signaling Crosstalk between Tubular Epithelial Cells and Interstitial Fibroblasts after Kidney Injury. *Kidney Dis (Basel)* 2(3), pp. 136-144. doi: 10.1159/000446336

Tang, F. et al. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 6(5), pp. 377-382. doi: 10.1038/nmeth.1315

Thiagarajan, R. D. et al. 2011. Identification of anchor genes during kidney development defines ontological relationships, molecular subcompartments and regulatory pathways. *PLoS One* 6(2), p. e17286. doi: 10.1371/journal.pone.0017286

Tran, H. T. N. et al. 2020. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol* 21(1), p. 12. doi: 10.1186/s13059-019-1850-9

Uhlen, M. et al. 2015. Proteomics. Tissue-based map of the human proteome. *Science* 347(6220), p. 1260419. doi: 10.1126/science.1260419

Urate, S. et al. 2021. Aristolochic Acid Induces Renal Fibrosis and Senescence in

Mice. *Int J Mol Sci* 22(22), doi: 10.3390/ijms222212432

Viall, A. K. and LeVine, D. N. 2020. Performance of a Nageotte hemocytometer method and a flow cytometric assay for residual leukocyte quantification in leukoreduced canine packed red blood cells. *J Vet Emerg Crit Care (San Antonio)* 30(3), pp. 272-278. doi: 10.1111/vec.12947

Vinas, J. L. et al. 2020. Sex diversity in proximal tubule and endothelial gene expression in mice with ischemic acute kidney injury. *Clin Sci (Lond)* 134(14), pp. 1887-1909. doi: 10.1042/CS20200168

Wang, E. T. et al. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456(7221), pp. 470-476. doi: 10.1038/nature07509

Wang, Y. J. et al. 2019. Comparative analysis of commercially available single-cell RNA sequencing platforms for their performance in complex human tissues. *bioRxiv*, p. 541433. doi: 10.1101/541433

Wanner, C. et al. 2016. Empagliflozin and Progression of Kidney Disease in Type 2 Diabetes. *The New England journal of medicine* 375(4), pp. 323-334. doi: 10.1056/NEJMoa1515920

Wilson, P. C. et al. 2019. The single-cell transcriptomic landscape of early human diabetic nephropathy. *Proc Natl Acad Sci U S A* 116(39), pp. 19619-19625. doi: 10.1073/pnas.1908706116

Wu, C. F. et al. 2013. Transforming growth factor beta-1 stimulates profibrotic epithelial signaling to activate pericyte-myofibroblast transition in obstructive kidney fibrosis. *Am J Pathol* 182(1), pp. 118-131. doi: 10.1016/j.ajpath.2012.09.009

Wu, H. et al. 2019. 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(1), pp. 23-32. doi: 10.1681/ASN.2018090912

Yang, H. C. et al. 2010a. Models of chronic kidney disease. *Drug Discov Today Dis Models* 7(1-2), pp. 13-19. doi: 10.1016/j.ddmod.2010.08.002

Yang, L. et al. 2010b. Epithelial cell cycle arrest in G2/M mediates kidney fibrosis after injury. *Nat Med* 16(5), pp. 535-543, 531p following 143. doi: 10.1038/nm.2144

Ye, J. et al. 2019. Aristolochic acid I aggravates renal injury by activating the C3a/C3aR complement system. *Toxicol Lett* 312, pp. 118-124. doi: 10.1016/j.toxlet.2019.04.027

Young, M. D. et al. 2018. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* 361(6402), pp. 594-599. doi: 10.1126/science.aat1699

Yu, L. et al. 2003. TGF-beta isoforms in renal fibrogenesis. *Kidney Int* 64(3), pp. 844-856. doi: 10.1046/j.1523-1755.2003.00162.x

Yu, W. et al. 2016. Transcriptomic changes in human renal proximal tubular cells revealed under hypoxic conditions by RNA sequencing. *Int J Mol Med* 38(3), pp. 894-902. doi: 10.3892/ijmm.2016.2677

Zaghlool, A. et al. 2021. Characterization of the nuclear and cytosolic transcriptomes in human brain tissue reveals new insights into the subcellular distribution of RNA transcripts. *Sci Rep* 11(1), p. 4076. doi: 10.1038/s41598-021-83541-1

Zanetti, F. 2020. Chapter 7 - Kidney-on-a-chip. In: Hoeng, J. et al. eds. *Organ-on-a-chip*. Academic Press, pp. 233-253.

Zeisberg, E. M. et al. 2007. Endothelial-to-mesenchymal transition contributes

to cardiac fibrosis. *Nat Med* 13(8), pp. 952-961. doi: 10.1038/nm1613

Zhuo, J. L. and Li, X. C. 2013. Proximal nephron. *Compr Physiol* 3(3), pp. 1079-1123. doi: 10.1002/cphy.c110061

Appendix

The top 100 genes with lowest adjusted P value of each cluster in the AAN snRNA-seq experiment (Chapter 4).

Podocyte	Endothelial cell	Mesangial cell	JG cell	PT-S1	PT-S1S2	PT-S2	PT-S2-cortical S3	PT-medullary S3	New-PT1	New-PT2	New-PT3	Proliferative cell	DTL	ATL	TAL	DCT1	DCT2	CNT	PC-OMCD	PC-IMCD	IC-A	IC-B	Transitional epithelium	Immune cell	Fibroblast-1	Fibroblast-2	Podocyte
Rab3b	Ltbp1	Lrriq1	Shc3	Slc7a7	Slc13a3	Wwox	Mlxipl	Slc22a18	Ngf	Wt1	Ccnd2	Kif20a	Fst	Akr1b3	Cdh1	Oprm1	Slc2a9	Scnn1g	Scnn1g	Aldh1a3	Pde1c	Insrr	Fer1l4	Spi1	Rem1	Fhl2	Rab3b
Nphs1	Mef2c	Alx1	Rgs5	Folh1	Pah	Bicc1	Slc47a1	Haao	Vcam1	Ms4a2	Hk2	Cdc20	Bst1	Scube2	Egfl6	Prkd1	Klhl3	Slc2a9	Pde1c	Wnt9b	Kit	Atp6v1b1	Itga2	Hck	Lgals9	Lama2	Nphs1
Clic3	Tek	Kcnd2	Serpini1	Ndrg1	Slc34a1	Gpx3	Hnf4g	Pxmp4	Cxcl10	Tnc	Itga3	Pole	Parm1	Col4a6	Tom1l2	Abca13	Sall3	Kcnq1	Col26a1	Tbx3	Atp6v1b1	Syn2	Eya2	Cyth4	Col1a1	Cfh	Clic3
Wt1	Elk3	Adamtsl1	Stac	Slc34a3	Slc6a19	Wwc1	Slc13a3	Slc13a2	Dock10	Wnt16	Peg3	Kif11	Cdh6	Nrip3	Slc1a5	Atp4a	Trpm6	Gata3	Fxyd4	Aqp2	Slc4a1	Atp6v1c2	Tbx3	Ikzf1	Grk5	Tshz2	Wt1
Sema3g	Exoc3l2	6530403H02Rik	Pcp4l1	Fam163a	Cryl1	Utrn	Slc5a8	Cyp51	Ankrd6	Kcnmb2	Axl	Ncapg	Pgm5	Ndrg4	Tiam1	Hmgcll1	Atp2b4	Abr	Gata3	Aqp4	Syn2	Slc26a4	Abcc3	Dock2	Cp	4930578G10Rik	Sema3g
Sncaip	Cd300lg	Pdgfra	Mannr	Slc16a10	Dab2	Sim1	Ntn1	Tmem25	Gm8126	Ncam1	Cd44	Birc5	Tshz2	Igfbp2	Scin	Kcnq1	Calb1	Kif26b	Crybg1	Elf5	Epb41l2	Tshr	Krt19	Myo1g	Nid1	Pde3a	Sncaip
Podxl	Fyn	Nav3	Irag1	Rhobtb1	Slc16a14	Pde7b	4933417A18Rik	Hsf2bp	G930009F23Rik	Wt1os	Klf5	Cdk1	Tbpl2	Pgm5	Prkd1	Klhl3	Sfrp1	Calb1	Mgat4c	Rasal1	Atp6v1c2	Exoc3l4	Upk3a	Syk	Irag1	Robo2	Podxl
Nphs2	Lama4	Tnn	Bvht	Pah	Bnc2	Cdkl1	Slc6a18	Napsa	Gm32357	Foxd1	Epha2	E2f7	Slc4a11	Upk3b	Pla1a	Arnt2	Slc12a3	Tmem52b	Pde8b	Gcnt3	Fgf12	Slc4a9	Rab27b	Fyb	Gata6	B3galt1	Nphs2
Sgip1	Ptprb	Frem1	Ano1	Adap2	Cyp2j13	Rgs6	Acot12	Pdxk-ps	Gm49146	Gm45004	Cpt1c	Hmmr	Platr22	Sptssb	Edar	Cacnb4	Trpv5	Atp2b2	Rasgrf2	Muc20	Alcam	Syt7	Il18r1	Nckap1l	Fblim1	Dlc1	Sgip1
Gabra4	Flt4	Gm44691	Rgs7bp	Abcd4	Alpl	Aopep	Dab2	Cd36	Gm49141	Gm28730	Tbx1	Prr11	Stk32a	Rbm20	Grik5	Cadps2	Tox3	Rhcg	Aqp2	Adgrf1	Slc4a9	Rp1	Clca3a2	Runx1	Ggt5	Cald1	Gabra4
Gabrb1	Rsad2	Gm29683	Kcnk3	Slc34a1	Cit	Slc34a1	Zbtb20	Crot	Gm49873	Cacna1g	Spaca7	Rrm2	Slc14a2	Proser2	Abca13	Lhx1	Phactr1	Scnn1b	Kif26b	A830021F12Rik	Atp6v1g3	Tmem163	Plch2	Fgd2	Fhl2	Gm53189	Gabrb1
Ptpro	Rnf144a	Ltbp1	Nr2f1	Slc6a19	Auts2	Mbnl2	Ldc1	Dnah2	Stx11	Mir351	Wnt2	Top2a	Creb5	Ube2ql1	Ndrg1	Slc16a7	Acss3	Hsd11b2	Rhcg	Itga2	Atp6v0d2	Sctr	Bmp3	Myo1f	Adam19	Abca8a	Ptpro
Myom2	Pecam1	Colec12	Pde3a	Pde4d	Slc13a1	Fth1	Kcnk5	Mlxipl	C3	Nkain3	Kdelr3	Efcab11	Cp	Atp10b	Clcnkb	Srgap1	Tsc22d1	Trpv5	Scnn1b	Wnt4	Asb15	Atp6v1g3	Cers3	Pik3ap1	Aldh1a2	Pdgfrb	Myom2
Ntng2	Prkch	Lepr	Carmn	Cryl1	Pik3c2g	Keg1	Galnt14	Por	Birc3	Mir503	Csf1	Diaph3	Slc7a12	Plpp4	Slc47a1	Rgs6	AI838599	Sdk1	Hsd11b2	Slc14a2	Rhcg	Pear1	Syt8	Inpp5d	Eya2	Gucy1a2	Ntng2
Lmx1b	Plpp1	Prrx1	4930511M06Rik	Dab2	Slc5a2	Farp1	Ttc39c	Hyi	Kcnh8	C430049B03Rik	Gadd45b	Esco2	Ptpn14	Tldc2	Prox1	Esrrb	Pde10a	Tox3	Kcnh5	Ptprh	Pdlim3	Col24a1	Styk1	Ptprc	Ksr1	Ak5	Lmx1b
Synpo	Rgcc	Rerg	Gpc6	Slc7a8	Car12	Stk39	Cndp2	Runx1t1	Cp	Adam33	Adamtsl4	Shcbp1	Thsd7a	Gm14546	Slc5a1	Tsc22d1	Trim43a	Frmpd4	Mcoln3	Slc38a11	Plcg2	Atp6v0d2	Pof1b	Slc9a9	Myo1b	Gpc6	Synpo
Mgat5b	Ednrb	Lama2	A830082K12Rik	Pdzd2	Glis1	Slc27a2	Smarca2	Slc38a3	Gm7233	Mark1	Etv4	Atad2	Gm45886	NA	Bicc1	Tdrd3	Sgms2	Adamts16	Tmem45b	Fxyd4	Slc35f1	Pde4b	Foxa1	Arhgap45	Sparc	Prkg1	Mgat5b
Nebl	Shank3	Srpx2	Piezo2	Angpt1	Prodh2	Pck1	Abcc2	Slc47a1	Prickle2	Sorcs3	Nipal4	Racgap1	Serpina10	4930546K05Rik	Map2	Ppp2r2b	Tmem52b	Bmpr1b	Frmpd4	NA	Clnk	Plcg2	Tmprss13	Prex1	Lama2	Cdh11	Nebl
Ntrk3	Cd200	Itga8	Adgrl3	Cyp2d26	Rab11fip3	Mecom	Slc6a20b	Mogat1	Creb5	Apba1	Btg2	Cdca3	D030045P18Rik	Scart1	Cacna1d	Sall3	4933406I18Rik	Slc8a1	Lypd6	Pde1c	Rcan2	Ripor2	Grhl3	Stab1	Dcn	Rbms3	Ntrk3
Ddn	Arhgap31	Grem2	Ren1	Igsf11	Phyhipl	Nbea	Dhtkd1	Eps8	Havcr1	Eya4	Itgb4	Sgo1	Ppp4r4	Pdgfa	Fli1	Trpm6	Scnn1b	Phactr1	Tmem150c	2700089I24Rik	Slc26a7	Clnk	Kcnk2	Arhgap30	Epb41l2	Colec12	Ddn
Tdrd5	Zeb1	Pde3a	Speg	Epb41l3	Ppp1r16b	Cyp4b1	Col19a1	Stard13	Icam1	Wfdc16	Edn1	Ndc80	Scel	Tiam1	Cacnb4	Grb14	Slc8a1	Egfem1	Bmpr1b	Frmpd4	Pgm5	Rcan2	Itprid1	Prkcb	Meis1	Rerg	Tdrd5
Sema3e	Arap3	Atp2a3	Trpc6	Gna14	Slc5a12	Shroom3	Lrp2	Mpped2	Cftr	Kirrel	Adcy2	Kifc5b	Proser2	Gm5296	Dnah11	Stk39	9330159	AI83859	Cacnb2	Rasgrf2	Aqp6	Trpc5	Foxq1	5430437	Nav3	Lhfp	Sema3e

														9			M07Rik	9						J10Rik			
Wt1os	Sncaip	Cped1	Ebf1	Gldc	Col27a1	Foxp1	Slc27a2	Cmah	Spaca7	Arhgap28	F2rl1	Incenp	Rbms3	Abi3bp	Acsl1	Pamr1	Cdk14	Oprm1	Phactr1	Plet1os	Oxgr1	Hepacam2	Abcc8	Mrc1	Cpm	Fbln5	Wt1os
Gm13817	Podxl	Lhfp	Slit3	Asl	Adra1a	Vgll4	Ak4	Hnf4g	Gxylt2	Scrn1	Plk2	Cep55	Atp10a	Robo1	Wwc1	Trpm7	Cnnm2	Nr3c2	Spock3	Tnn	Ccbe1	Oxgr1	Bace2	Arhgap25	Fam20a	Csmd1	Gm13817
9130410C08Rik	Nrp2	Gpc6	Rspo3	Dock9	Hnf4aos	Tspan9	Cyp2j13	Hnf4a	Relb	Adgrv1	Fst	Bard1	Chn2	Scel	Utrn	Mecom	Oprm1	Tmem45b	St6gal1	Apod	Dmrt2	Ccbe1	Capn13	Epsti1	Mrc2	Slco2b1	9130410C08Rik
Foxd2os	Sp100	Cfh	Sgip1	Slc16a14	Fut9	Sipa1l3	Cyp4b1	Ndel1	Gm35330	Lsp1	Wnt7b	Ube2t	Slc34a2	D030045P18Rik	Sim1	Tsc22d2	Wnk1	Defb1	Sptbn2	Styk1	Gm9871	Hmx2	Ttc6	Ly86	Mxra7	Cped1	Foxd2os
Wipf3	Inpp5d	C1qtnf7	Rnf150	Lrp2	Nhs	Me3	Cda	Etnppl	Gm5799	C1qtnf7	Cldn1	Cenpf	Cdh13	Clcnka	Rhobtb1	Ptgfr	Kitl	Col26a1	Kcnc2	Bmpr1b	Bmpr1b	Bmpr1b	Ankfn1	Pou2f2	Fbln5	Gucy1a1	Wipf3
Gm29266	Esrrg	Spon1	Col12a1	Gatm	Slc4a4	Fgfr2	Mtor	Esr1	Dpysl3	Scn5a	Ccdc80	Nuf2	Chst8	Bsnd	Cabcoco1	Cdk14	Ltc4s	Gpr39	Kcne1	Nyap1	Adgrf5	9130008F23Rik	Depp1	Ctss	Tshz3	Kcnt2	Gm29266
Gm3772	Eng	Cntn1	Ngf	Slc2a2	Tnfaip8	Acsm2	Aacs	Slc6a15	Klf6	Tafa1	Adamts5	Knl1	Sorcs3	Ppl	Pde7b	Stk32b	Tdrd3	4933406I18Rik	Lypd6b	Hacd4	Tmem117	Lsamp	Upk1b	Elmo1	Gli3	Kcnd3	Gm3772
Arhgap28	Egfl7	Negr1	Hip1	Kif12	Snx29	Nr3c2	Aass	Enpp3	Jazf1	Akap12	Runx1	Bub1	Sema5a	Spon1	Ntn4	Cwh43	Defb1	Gm15581	Tmem117	Spink8	Wscd2	Tmem117	Sytl5	Pid1	Nkd2	Col1a2	Arhgap28
Gbx1	Rbms1	NA	Dgkb	Ptprd	Ano3	Maml2	Gas2	Slc5a8	Samd4	Ptchd4	Cdkn1a	Tpx2	Arhgap28	Rbfox3	Nudt4	Tmem52b	Cwh43	Tmem117	Nav1	Tbx3os1	Gm12121	Wscd2	Gsdmc4	Wdfy4	Pde8b	Daam2	Gbx1
Tmem150c	Meis2	Cxcl12	Cald1	Bnc2	Ass1	Paqr5	Acsm3	Slc1a4	Pdgfb	Brinp1	Zfp51	Aurka	Spon1	Soga1	Slc16a7	Fgf13	D630023O14Rik	Crybg1	Cdk14	Fam241a	Rhbg	Rtl4	Gsdmc3	Art2a	Osmr	Ebf1	Tmem150c
Tspan2	Ehd4	Ripor3	Gm20125	Fggy	Gm15802	Kank1	Atp11a	Osbpl8	Fyb2	Rasl11a	Cdh2	Ect2	Epha7	Platr22	Osbpl8	Slit2	Tsc22d2	Peli2	4933406I18Rik	Gprc5b	Chst8	Malrd1	Gsdmc2	Otulinl	Zfpm2	Adam12	Tspan2
C1qtnf7	Ccm2l	Vwf	Acta2	Alpl	Pakap	Kap	Ces1f	Slc5a4a	Mark1	Robo1	Ankrd1	Iqgap3	Tbc1d4	Prox1	Cyfip2	Sorbs2	Kcnq1	Slc7a15	Ptger1	Camkk2	Tshr	Gm12121	Tmem40	Dock10	Col14a1	Ldb2	C1qtnf7
Gm11747	Trpc3	Ntf3	Daam2	Prkag2	Gm31718	Ptprm	Hsd17b2	Nopchap1	Myof	Amph	Rin1	Depdc1a	Jag1	Rasal1	Adcy1	Slc12a3	Cadps2	Cdk14	Pappa	Tmem45b	Calcr	Gm37381	Gm13219	Raet1e	Tmem45a	C7	Gm11747
Gm13814	Npr1	Adam12	Fhl2	Gpat3	Gm31641	Fhod3	Acsm1	Ntn1	Masp1	Clstn2	Fosl1	Smc2	Cald1	Chsy3	Shmt1	Rab27a	Abca13	Tbck	Gata2	Apobec3	Susd1	Gm47795	Mir205hg	Pik3r5	Fstl1	Gm16332	Gm13814
Kcng2	Ifi44	Hgf	Trpc4	Auts2	Pde4d	Ptpn13	Pank1	Serpina1f	Bmp6	Eya1	Atp10a	Kif2c	Ncam1	Fam241a	Stxbp4	Mgll	Emx1	Pde3b	NA	Akr1b3	Cpsf4l	B830017H08Rik	Gm28286	Cd180	Robo1	Mrc2	Kcng2
Usp13	Spaar	Cacna1c	Tbx3os1	Slc13a1	Lpar3	Mitf	Nrxn2	Ccdc88c	Col18a1	Gm32357	Adam8	Stil	Ccnjl	Sorcs3	Mboat2	Emx1	Adamts3	Gata2	Apobec3	Tmtc2	Cacnb2	Rhcg	Gm36640	Tgfbi	Adamts5	Pcdh9	Usp13
Mafb	Ptpru	Ddr2	B3galt1	Frmd4b	Slc7a7	Tanc1	Dgkh	Dglucy	Patl2	Gpc6	Ptprn	Cdca8	Satb2	Gabre	Ston2	Uroc1	Scube3	Rhbg	Adgrg3	Epor	Rapgef4	Klhl29	Sprr1a	Mir142hg	Serping1	Fbn2	Mafb
Asic2	Cd38	Rbms3	Adora1	Pik3c2g	Nox4	Myo1d	Glis1	Slc17a1	Parp14	9130410C08Rik	Serpine2	Cenpa	Rcan2	Mboat2	Cdkl1	Wnk4	Scnn1g	St6gal1	Atp2b2	Samd12	Pde4b	Asb15	Junos	Nlrp1b	Lbh	C1qtnf7	Asic2
Gja3	D5Ertd615e	Cdh11	Ndst3	Itpr2	Gna14	Kctd1	Slc25a21	Tpmt	Csf1	Cdh6	Prrg4	Kntc1	Chsy3	Scin	Rgs6	Acss3	Gata3	Mecom	Fam110b	Heg1	Parm1	Klkb1	Arhgap40	Rhoh	Cyp1b1	Adamts10	Gja3
Ephb1	Flt1	Dlc1	Prkg1	Clec2h	Gatm	Shank2	Gc	4933417A18Rik	Sema3c	Cp	Ptpn22	Rad51ap1	Fam78b	Plxdc1	Esrrb	Kctd1	Gm27239	Phactr2	Arhgef37	Gpr39	Sytl5	Fam243	Slc14a1	Card11	Adamts10	Rftn1	Ephb1
Thsd7a	Tmtc1	Gm16332	1500009L16Rik	Slc7a9	Rhobtb1	Limch1	Hykk	Eci2	Cxcl1	Plce1	Ngf	Kif22	Stxbp6	Psca	Slc34a1	Rap1gap2	Syt17	Prdm16	Sdk1	Plet1	Slc44a3	Kctd16	Tmc7	Clec12a	Fbn2	Meis2	Thsd7a
Gm44129	Nova2	Ldb2	Akap6	Nox4	Ppara	Stox2	Slc17a3	Slc6a18	Smpdl3b	Rasl11b	Gask1b	Mki67	Gprc5a	Prox1os	Dapk1	Prdm16	Kl	Kcne1	Pde3b	Atp10b	Kctd16	Parm1	Klf5	Gm49339	Pdgfrb	Olfr1033	Gm44129
Foxd1	Gm45837	Gm5834	Tafa1	Slc5a2	Slc22a28	P3h2	Csgalnact1	Acot12	Cd44	Proser2	Ndst3	Cenpi	Plekhh2	Rbms3	Ocln	Lrrc52	Rhcg	Pde8b	Kcnq1	Myzap	Kcns3	Pkib	Upk1a	Nfam1	Il33	Dapk2	Foxd1
Robo2	Dlc1	Pdgfrb	Meis1	Acsm2	Pdzd2	Rap1gap2	Marchf1	Hmgcr	Nkap	Tshz2	Kcnq5	Zwilch	Gpc6	Elovl6	Pde4d	Cables1	Trim43c	Kitl	Rbfox1	Vtcn1	Eya2	Aldh1l2	Acer2	Irag2	Nrp2	Pdgfra	Robo2
Cldn5	Cdh5	Ust	Myh11	Tinag	Atxn1	Prdm16	Aspg	Pde4d	Sbno2	3100003L05Rik	Ccn1	Troap	Pkhd1	Srgap3	Map3k1	Kcnj1	Mapk10	Plppr1	Camkk2	Gm43305	Malrd1	Car12	Ly6d	4930469K13Rik	Raph1	Lrp1	Cldn5
Gm4804	Robo4	Rasgef1b	Abcc9	Car12	Aspa	Dnm3	Ppm1k	Fam107a	Nfkbiz	Slc24a3	Smpdl3b	Cip2a	Prrg1	Clstn2	Dnah12	Tox3	Sorbs2	Rap1gap	Nr3c2	NA	Kcnq1	Chl1	Msln	Cybb	Col3a1	Gm17167	Gm4804
Plce1	Ephb1	Tshz2	Lhfp	Ablim3	Ndrg1	Plekhg1	Camk4	Acox2	Nlrc5	Grid1	Abcb1b	Aspm	Naaladl2	Zg16	Cryl1	Wnk1	Kctd1	Tfcp2l1	Mpped2	Ehf	Ica1	Foxi1	Grhl1	P2rx7	Il1r1	Lbh	Plce1

2810433D01Rik	Lmo2	Meis2	Myl9	Dpyd	Pitpnc1	Sox5	Trim7	Cspg4b	Mtss2	Zfp536	Bst1	Fancd2	Jazf1	Tbc1d4	Fgf9	Mapk10	Adamts16	Stox2	Edaradd	Gnao1	Itpr2	Irs1	Ankdd1a	Pik3cd	Fn1	Osmr	2810433D01Rik
Sulf1	Ccdc85a	Prkg1	Cap2	Gabrb3	Miox	Arl15	Erc2	Thrb	Nrg1	Gm49141	Kcnip4	Kif4	Pax8	Gm30725	Enox1	Tmem728	Tmcc3	Tmem178	Phactr2	Slc9a4	Arhgef374	Gm20574	Upk2	Themis2	Speg	Cacna2d1	Sulf1
Il6ra	Gas7	Daam2	Dlc1	Glis1	Frmd4b	Pkhd1	Slc7a13	Bmp4	Il34	Gm29683	Cxcl1	Trim59	Rtn1	Epha7	Scel	Efna5	Dach1	Scube3	Plppr1	Celf3	Tfcp2l1	Nbea	Sidt1	Adgre1	Sp100	NA	Il6ra
Cdkn1c	Tie1	Adgrl3	Mark1	Lpar3	Pdzk1	Efna5	Slco1a1	Lcp1	Actn1	Npr3	Anxa3	Ncaph	Wwc1	Tshz2	Dab2	Sgms2	Grb14	Tfap2a	Tpd52l1	Gata3	Serpinb9	Plekhd1	Entpd3	Csf2rb	Cfh	Gm29683	Cdkn1c
Plekhh2	Ptprm	Myo1b	Rapgef4	Prodh2	Nr1h4	Dock1	Slc5a10	Erich6b	Pakap	Rnf150	Gxylt2	Parpbp	G930009F23Rik	Prrg1	Pdzd2	Phactr1	Cables1	Dach1	Srgap3	Kif26b	Junos	Ralgapa2	Pparg	Apobec1	Prrx1	Esrrg	Plekhh2
Aff3	Dysf	Fhl2	Gabrb1	Rab11fip3	C330002G04Rik	Pde1a	Agps	Slc39a14	Zeb1	Plppr4	Bhlhe41	Melk	Kcnh8	Wnt7b	Grhl2	Dach1	Gm1604a	Arl15	Myzap	Lypd6b	Bmp8a	Gm35576	Plet1os	Itgam	Mark1	Rhoj	Aff3
Pde3a	Fgd5	Kcnt2	Ntrk3	Sirpa	Enpp2	Maml3	Nxph1	Ankrd33b	C430002N11Rik	Pgm5	Col4a6	Ska1	9130019P16Rik	Hs6st3	Zfpm2	Med12l	Malrd1	Kl	Gm45886	Pcdh7	Scnn1b	Stap1	Krt23	H2-Eb1	Lamb3	Zeb2	Pde3a
Opcml	Edil3	Lamb3	Mef2c	Slc22a23	Gldc	Ano6	Hnf4aos	Tef	Sytl2	Kcnma1	Pak3	Kif15	Gm32357	Nupr1	Deptor	Syt17	Syne1	Atp2b4	Tfap2a	Acer2	Hepacam2	Thsd7a	Lyzl4	Cd84	Ddr2	Nrp2	Opcml
Srgap1	Plvap	Clstn2	Cacna2d1	Bmp6	Slc22a8	mt-Rnr2	Glt1d1	Parvb	Tgtp2	Dkk3	Plat	Anln	Gramd1b	Pcdh7	Ppp1r1a	Kng2	Mecom	Grb14	Gprc5b	Nav1	Plekhd1	Pros1	Adam28	Gm35154	Itga8	Slit3	Srgap1
Astn2	Ushbp1	Atoh8	Serpine2	Magi2	Prkag2	Tacc1	Slc22a30	Bcl6	Sema5a	Adamts5	Slco2a1	Tacc3	Susd4	Hs3st3a1	Zbtb20	Maml3	Kcnj1	A4galt	Styk1	Scamp5	Pam	Emb	Sptssb	Cd86	Eng	Col14a1	Astn2
Plat	Fmnl2	Rarb	Dkk2	Cfap69	Igsf11	Ank3	Sugct	Mpv17l	Sash1	Nav3	Styk1	Dtl	Bcat1	Slc7a1	St3gal6	Syt9	4930550C14Rik	Casz1	Tfcp2l1	AU021092	Tmem61	Grb10	A830008E24Rik	Arhgap15	Col5a1	Frem1	Plat
Zfp385b	Inpp4b	Tnc	Fry	Grid1	Msra	Msi2	Them7	Zbtb20	Nek6	St3gal5	Kcnk10	Dlgap5	Cobl	Sgcz	Kng1	Cnnm2	Tmem65	Tmcc3	Slc44a3	Gm14546	Rasef	Tmem61	Gm14066	Btk	Zeb2	Adamts5	Zfp385b
St3gal1	Acer2	Rem1	4933431K23Rik	Slc5a12	Gm6999	Gm28437	Ghr	Hgd	H2-Q7	Gm47416	Cpne2	Cdkn3	Adgrg6	Dsp	Ildr1	Nrxn3	Slc4a3	Aqp2	Ctnnd2	Pik3c2b	Pros1	Gm12216	Gata3	Gm26740	Fbn1	Col6a3	St3gal1
Crb2	Adgrl4	Robo2	Ntng1	Npl	Xylt1	Gm29216	Fto	Slc15a2	Grip1	Sdk1	Eda2r	Ckap2	Veph1	Atp8b1	Enah	Scara5	Arl15	Slit2	Stox2	Mal2	Rorb	Bmp8a	Parm1	Selplg	Meis2	Rnf150	Crb2
Eya4	Ldb2	Gem	Msrb3	Tmem108	Lrp2	Iqschfp	Cndp1	Csad	Plaur	Mical2	Celf5	Ncapd2	Slc1a3	Crlf1	Kcnk5	Magi1	Rap1gap2	Tpd52l1	Npnt	Ociad2	Abr	Lrrc3	Gm16083	Adcy7	Dclk1	Adamts2	Eya4
Amph	Hip1	Gm32834	Hopx	Wdfy3	Tmem108	Tmcc3	Skint11	Slc10a2	Cdk6	Gm49146	Kcnh8	Ttk	Tinagl1	Sim2	Galnt14	Gata3	Plekhd1	Adgrg3	Prdm16	Tiam1	Anxa4	Tbc1d24	Hopx	Pirb	Ngf	Egflam	Amph
Raver2	Sema3d	Cdc42ep3	Pcdh9	Snhg11	4921539H07Rik	Pax2	Anks1b	Ldc1	Kcnip4	Zbtb7c	Krt20	Prc1	Plce1	Chst9	Dtna	Sfrp1	Slit2	Hdac7	Mecom	Ros1	Tmod1	Nipal2	Crybg1	Plbd1	Gucy1b1	Pkhd1	Raver2
St6galnac3	Rcsd1	Igfbp5	Gucy1a2	Adra1a	Gm42397	Zmiz1	Ksr2	Acy1	Ccn1	Cntn3	Ggta1	Fanci	Utrn	Megf11	Zfp521	Tfap2a	Myzap	Pitpnc1	Grip1	Muc1	Rnf152	Rnf152	Col4a6	Cd74	Col15a1	Gabra3	St6galnac3
Fyn	Erg	Pid1	Cabp1	Atxn1	Acsm2	Cyfip2	Slc22a12	Mep1a	Adgrg2	Lbp	Ripor2	Neil3	Myo5b	Cacnb3	Ppp2r2b	Gm15848	Lypd6b	Myzap	L1cam	Bcam	Nav1	Sh2d4b	Htr4	Ciita	Tnfsf8	Col3a1	Fyn
Colec12	Elmo1	Grm7	Jph2	Agmo	Acnat2	Efhd1	Ctnna2	Cbs	Mapk4	Tspan18	Rap2b	Bub1b	Akap12	Dnm3	Keg1	Lnx1	Esrrb	Srgap3	Gm15581	Me2	Bcar3	Sytl5	Efnb2	Rgs10	Tnc	Pla2r1	Colec12
Asb15	Cyyr1	Aox3	Foxp2	Fut9	Galnt14	Sema5a	Slc22a28	Galnt14	Sh3d19	Gli3	Neto2	Arhgap11a	Lypd6b	Cldn19	Pcsk5	Gm13052	Gm26618	Lypd6	Naaladl2	E330013P04Rik	Gm20574	Jag1	Macc1	Gng2	Svep1	Lpar1	Asb15
Flrt2	Rapgef5	Kirrel	Sulf1	Maf	Gm32461	Specc1	Cyp4a32	Slc3a1	Adamts1	Srgap1	Serpine1	Ccnb1	Slc6a6	Tmem184b	Chka	Kcnj10	Scara5	Cdkl1	Arhgef4	Map6	Pde1a	Susd1	9530036O11Rik	Pld4	Astn2	Tenm3	Flrt2
Cdc14a	Kank3	Palld	Kcnt2	Nhs	Timp3	Dst	Cntnap5a	Mep1b	Map3k13	Cald1	Arhgef4	Ncapg2	Eva1c	Gm28175	Tcf7l2	Sgsm1	Nudt4	Plcb1	Erc1	Rgs11	Col18a1	Naaladl2	Gm49246	Grap2	Ptch2	Serping1	Cdc14a
Elmo1	Fbxl7	Arhgap28	Pla2r1	Unc5c	Dao	Tmem116	Snx29	Slc23a1	Slc22a15	Sncaip	Tnfrsf23	Clspn	Ext1	Sox5	Gk	Tacc1	Pde3b	Plcl1	Ehf	Cpne8	Ralgapa2	Alcam	Wnt4	Aim2	Hgf	Irag1	Elmo1
Coro2b	Nxpe4	Gm13470	Slc38a11	Slc4a4	Aoah	Lrch1	Slco1a6	Proc	Cdh6	Tbc1d4	Inf2	Cenph	Litaf	Col4a5	Casp7	Erbb4	Nr3c2	Maml2	Ccser1	Best3	Cpeb2	Aph1b	Vtcn1	Nrros	Fam114a1	Mill2	Coro2b
C1qtnf1	Rasip1	Mrc2	Rbms3	Tnfaip8	Ces1f	Nckap5	Nat8f6	Nedd4l	Sema6a	Adgrg6	Cpe	Cenpe	Tspan5	Hspg2	Gcgr	Syne1	Flvcr1	Nfia	Abr	Slco4a1	Zfhx3	Tafa1	Adamts17	Abcg3	Pdgfra	Ggt5	C1qtnf1
Npr3	Lrrc3b	Gm30524	Hdac9	Col8a1	Fggy	Fli1	Alms1-ps2	Csf1r	Birc2	Cpe	Zfp365	Ankle1	Prickle2	Capn2	Abcc2	Pde10a	Maml2	Cdh1	Adgrg6	Coro2a	Exoc3l4	Itpr2	Piezo2	Esrrg	Wdr66	Dclk1	Npr3

H2-Q6	Gvin1	Lrp1	Filip1	Acnat1	Arhgap10	Magi1	Vmn1r18	Keg1	Itgav	2610307P16Rik	Atg9b	Ticrr	Lhfp	Ptpn14	Nnt	Sgpp2	Tfap2a	Zbtb20	Samd12	Mycbpap	Mme	Tfcp2l1	Fstl4	Gm36161	Col1a2	Meis1	H2-Q6
St6gal1	Chrm3	Ncam1	Sparcl1	Snx29	Gm37245	Tmem164	Kcnb2	Pcsk5	Pvr	Fhl2	Dennd2a	Ckap2l	Gls	Wipf3	Farp1	Nfia	Cacnb4	Blnk	Grb14	9130008F23Rik	Nedd4l	Ncald	Apobec3	Pltp	Eln	Mir100hg	St6gal1
Gm35048	Rhoj	Pde8b	Prkg2	Ano3	Idua	Tom1l2	4930533I22Rik	Slc22a19	Ptprn2	Crb2	Gdf15	Cdca2	Bicc1	Rnf43	Rps6ka6	Pde7b	Kng2	Iqgap2	Atp10b	Srgap3	Plcb4	Ociad2	Adgrg6	Ccr2	Cald1	Cacna1c	Gm35048
Fam78b	Gimap6	Hoxd11	Emid1	Ass1	Slc2a2	Kcnq1ot1	0610031O16Rik	Kif20b	H2-K1	Cdh4	Akap12	Gen1	Sptb	Gm28172	Tfap2b	Grin3a	Tmem117	Fut9	Rbms3	Gm2682	Plagl1	Abr	Homer2	Lpcat2	Rerg	Galnt17	Fam78b
Zbtb7c	Tshz2	Abca6	Meis2	Gm15802	Slc47a1	Sorbs2	Gm28905	Pcgf5	Tgfbr2	Rcan2	Ppl	Ccdc18	Samd4	Myzap	Lactb2	Pip5k1b	Efhd1	Nhs	Cdkl1	Rtn1	Tmcc3	Ckmt1	Alcam	Kcnk13	Slco2b1	Carmn	Zbtb7c
Galnt10	Entpd1	Adamts5	Smoc2	Gm12962	Slc22a23	Slit2	Gm6300	Slc1a1	Nfkb1	Gxylt2	Havcr1	Espl1	Bcam	Hacd4	Ptpn18	Acss1	Sdk1	Trpm6	Tox3	Boc	Casz1	Sh3pxd2b	Tmem45b	Gm21188	Dkk3	Vstm4	Galnt10
Lingo2	Prex2	Pla2g4a	Agtr1a	Pakap	Acsm1	Hk1	Gm43158	Gk	4930414F18Rik	Gm49873	Clcf1	Ccnf	Dnm3	Myo3b	Dst	Sim1	Efhc2	Sorbs2	Tmtc2	Myh10	NA	Car15	Atp8b1	Ccr5	Stim1	Mgp	Lingo2
Nes	Rapgef4	Meis1	Gng11	Gm906	Syn3	Nfat5	Gm45193	Vwa2	Irf1	Gm20488	Inhba	Gas2l3	Hnmt	Fry	Efhd1	Msi2	Vdr	Col4a4	Gpr39	Ptger1	Cpne8	Myo5a	Liph	Apobec3	Denn2b	Spon1	Nes
Tmem108	Ntf3	B130024G19Rik	Runx1t1	Gm28845	Kcnk5	Map3k1	Gm36011	Afap1l2	Snx10	Eno2	Phlda3	Rad54b	Mtcl1	Utrn	Rgl1	Cdh3	Gpr39	Pde7b	Grk5	Atp8b1	Rabep2	Gulp1	Myzap	Ctsc	Ano1	Rora	Tmem108
Mmp23	Zfp366	Msrb3	Rhoj	Gm37245	Trabd2b	Setbp1	Gm42397	Ablim1	Acat3	Ehd2	Svop	Kifc1	Myo9a	Reln	Esrrg	Dapk1	Acss1	Cacnb2	Nckap5	Camk2b	Racgap1	A630001G21Rik	Atp10b	Unc93b1	Gabra3	Casp4	Mmp23
Ptchd4	Pbx1	Kcnip3	Tenm3	4921539H07Rik	Gria3	Dock8	BC024386	Gfra1	Msn	Plekhh2	Ptchd4	Eldr	Glis3	Frmpd4	Frmd4a	Fcamr	E230016K23Rik	Wnk1	Pde7b	Fam174b	Ybx2	Slc44a3	Trp63	Il6ra	Bgn	Col15a1	Ptchd4
Rhbdl3	St6galnac3	Fbln5	Sntb1	Gm31718	Errfi1	Ssh2	Gm8189	Hoga1	Kap	Col4a6	Ikbke	Lockd	St6gal1	Bcl6	Pter	Dcdc2a	Megf9	Nckap5	Dock7	Atp8a1	Sh3bgrl2	Rabep2	Ddx43	Vav1	Rasa3	Neat1	Rhbdl3
Agrp	Prkg1	Ebf1	Ism1	Gm47564	Slc7a8	Iqgap1	AA536875	Abcc2	Rgs20	Tmtc2	Rhbdl2	Traip	Ankrd1	Cab39l	Prkcq	Malat1	Gm26685	Slc44a3	Sox5	Fgf12	Ncald	Kcnq1	Eef2k	Cd48	Tnfsf13b	Grm7	Agrp
Pag1	Arhgef15	Tshz3	Gm32834	Gm31641	Maf	Exoc6b	Vmn1r20	Slc6a20b	Abcc4	Gm4876	Zfp536	Nusap1	Gmds	Hmga2	Cytip	Col5a2	Cntfr	Akr1c19	4930550C14Rik	Bcat1	Gpr39	Adgb	Dapp1	Lcp1	Dlc1	Rasgrp2	Pag1
Speg	Tcf4	Gm48727	Epas1	Gm32461	Agxt2	Pard3b	Vmn1r28	Itgb8	Gm20559	Chst9	Bcl2l15	Kif18a	Rbfox1	Arl15	Pkp4	Nr3c2	Rap1gap	Adamts3	Plcb4	Otulinl	Eps8	Cpeb2	Cbr2	Fgr	Tenm3	Grk5	Speg
Arhgef26	Tll1	Mmd	Ntrk2	Tfpi2	Pter	Map4k3	B4galt5	Map3k7cl	Olfm2	Styk1	Hs3st5	Aunip	Dcdc2a	Chn2	Slc25a12	Ccdc141	Papss1	Gm44120	Pak3	Gstm2	Sh3pxd2b	Dmxl1	Slc9a4	Zbtb20	Slc7a2	Rem1	Arhgef26
Clic5	Emcn	Zdhhc14	Gucy1b1	Cpn1	Iqgap2	Fry	Rhobtb1	Itih5	Anxa3	Hspa12a	Basp1	Nsl1	Spns2	Pax2	Stk39	Plekhd1	Gm21847	Magi1	Sorbs2	Megf9	Nr3c2	Prkg1	AU022754	Ank3	Nkd1	Denn2b	Clic5
Gm13936	Lrrc8c	Bmpr1b	St6gal1	Chrna4	Rgl1	Kng2	Mfsd4b5	Dhtkd1	Lhfpl2	Anxa3	Ccdc9b	Cenpp	Tns3	Slco4a1	Lrp2	Pappa	Prkd1	Nudt4	Blnk	Pde8b	Blnk	Tmcc3	Ctnnd2	P2ry6	Cdh11	Kirrel	Gm13936
Kirrel	Adgrf5	Serping1	Slco2a1	Syn3	Lrmda	Ptprj	Agbl4	Myo5b	Plekhg2	Meis2	Cd109	Mybl1	Adgrg3	Ophn1	Slc12a1	Ank2	Rgs6	Aif1l	Serinc3	Agmo	B830017H08Rik	Spink8	Robo1	Pkhd1	Ets1	Pknox2	Kirrel
Zfp423	Ebf1	Nrp2	Vstm4	Slc43a2	Them7	Btbd9	Gm31641	Lactb2	Grid1	Pdpn	Plaur	Brca1	Ppl	Ehf	Rasgrp1	Efhd1	Scg5	Sim1	Hoxd3	Sptbn2	Nkain3	Cyp2s1	Syt16	Pip4k2a	Col12a1	Cp	Zfp423
H2-Q7	Samd12	Osmr	Mir100hg	Arhgap10	Slc7a9	Rora	Pdzrn3	B3gat2	Rbpms	Efemp1	Gprc5a	Wdr62	Aqp1	Stard13	Slc27a2	Cabcoco1	Uroc1	Gm13652	Oprm1	Bspry	Dmxl1	Gm47790	Snx31	Rbpj	Loxl1	Abi3bp	H2-Q7
Tmod3	Calcrl	Kcnd3	Rcsd1	Cmip	Bcl11b	Chka	Mtss1	Ralgps2	Rhbdf2	Heyl	Frmd6	Fignl1	Rora	Mroh4	Pex5l	Neat1	Ccdc141	Kcnj1	Frk	Olfr373	Sh3pxd2a	Clcnkb	Tns4	Gvin-ps7	Dapk2	Wdr66	Tmod3
Fzd2	Kdr	Col6a3	Osmr	Gm7652	Zfp697	Gm28661	Cd36	Lpgat1	H2-D1	Cap2	Lhfp	Cenpm	Acsm2	Rasl11b	Mecom	Atp1a1	Atp1a1	Frk	Sim1	Slc4a7	9130008F23Rik	Pde1a	Nfe2l3	Tns1	Ttc28	Slfn5os	Fzd2
C430049B03Rik	Rasgrp3	Grk5	Rapgef5	Lipa	Atp6v0a4	Acss1	Gm39079	Cubn	Dennd2a	Frmpd1	Inka2	Cdc7	Mical3	Anxa2	Nbea	Itga6	Gm6277	Klhl3	Sort1	Nek10	Magi1	Rhbg	Pdlim1	Pik3r6	Casp4	Fn1	C430049B03Rik