Aus der Medizinischen Klinik und Poliklinik IV
Klinik der Universität München
Direktor: Prof. Dr. Martin Reincke

Identification and characterization of regulatory pathways and prognostic markers linked to the initiation and progression of diabetic kidney disease using systems biology

Dissertation
zum Erwerb des Doktorgrades der Dr. hum biol
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Viji Nair
aus Kerala, India
2022

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatt:	Prof. Dr. Peter Jon Nelson, Ph.D.
Mitberichterstatter:	Prof. Dr. med. Volker Vielhauer Prof. Dr. med. Clemens Cohen
Dekan:	Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 22.06.2022

Declaration

I hereby declare that all the presented work was carried out by me under the supervision of Prof. Dr. Peter Nelson Nephrologisches Zentrum, Medizinische Klinik und Poliklinik IV, Innenstadt, Klinikum der Universität München. This work has not been submitted in part or full to any other university or institute for any degree or diploma. This thesis work was made possible by a large international consortium of scientists and centers that provided the diverse raw data used for the development and validation of the tools described in this thesis:

- The European Renal cDNA Bank (ERCB). Technical personnel within the Nelson lab (Anke Fischer and Sylke Rohrer) and Nephrologie Zentrum LMU Munich, including critical input from Prof. Maja Lindenmeyer (current affiliation: Hamburg-Eppendorf III. Medizinische Klinik und Poliklinik, University of Hamburg).
- Anna Henger and Edgar Otto provided help with RNA preparation used for microarray and single cell analyses. Rajasree Menon performed single cell quality control processing and Prof. Dr. Matthias Kretzler was available for additional bioinformatics support from the Michigan Kidney Translational Medicine Core, University of Michigan, Ann Arbor, U.S.A.
- Robert Nelson, Jennifer Weil, Lois Jones, Jillian Loebel, Roselene Lovelace, Bernadine Waseta, and Camille Waseta from National Institutes of Health, NIDDK Bethesda, U.S.A. developed the long detailed clinical study that is at the center of this thesis in cooperation with the Pima Indian tribe of Arizona. They provided the clinical data, the renal physiology data, and the quantitative morphometry of the kidney tissue used for this study.
- Michael Mauer, Ann Palmer, and Frida Maiers from the Department of Pediatric Nephrology University of Minnesota processed and imaged the kidney tissue used for the quantitative morphometry in this study and provided expert advice concerning the quantitative morphometry of the kidney tissue. Berne Yee from Southwest Kidney Institute New Mexico, U.S.A. performed the kidney biopsies used for this study.
- Kevin Lemley from University of California Keck School of Medicine, Children's Hospital Los Angeles U.S.A. provided expert advice on the quantitative morphometry of the kidney tissue.

Date: 20. August.2021
Place: Ann Arbor, USA Signature: Viji Nair

Table of Contents

ZUSAMMENFASSUNG *1*

Summary *3*

Abbreviations *5*

1. Introduction *8*

1.1 Chronic kidney disease **8**

1.2 Prevalence of kidney disease **10**

1.3 Kidney physiology **11**

1.4 Parameters used to measure kidney function **13**

1.5 Diabetic kidney disease **16**

1.6 Kidney biopsy **19**

2. Bioinformatics *20*

2.1 Systems biology **21**

2.2 Precision medicine **22**

2.3 Enhanced characterization of disease models through systems biology **24**

2.4 Single-cell RNAseq analysis and diabetic nephropathy **26**

3. Hypothesis and goals of thesis *28*

4. Material and Methods *30*

4.1 Study cohorts **30**

4.1.1 Early DKD cohort **30**

4.1.2 Advanced DKD cohort **31**

4.2 Tissue microdissection and microarray preparation **31**

4.3 Microarray data processing **32**

4.4 Gene co-expression modules **34**

4.5 Functional gene annotation, enrichment, and pathway analysis **37**

4.6 Upstream regulator analysis approach and gene selection **37**

4.7 Tissue processing for single-cell analysis **38**

4.7.1 Single-cell isolation **38**

4.8 Next-generation sequencing of single-cell mRNA **39**

4.9 Analysis of single-cell data **40**

4.10 Statistical analysis **41**

5. Results *41*

5.1 Transcriptomic profiling **42**

5.2 Tubulointerstitial subgroup **43**

5.2.1 Clinical characteristics **43**

5.2.2 Module detection **43**

5.2.3 Functional context of VvInt-associated transcripts 44
5.2.4 Pathways associated with interstitial damage 48
5.2.5 Identification of upstream regulators 50
5.2.6 Fibrosis and kidney function decline 51
5.2.7 Findings from early DKD to advanced DKD 52
5.3 Glomerular subgroup 55
5.3.1 Clinical characteristics 55
5.3.2 Module detection 57
5.3.3 Functional network of glomerular lesion-associated transcripts 60
5.4 Mapping of transcripts into single cell 65
5.4.1 Clinical characteristics 65
5.4.2 Detection of kidney cell types 65
5.4.3 Cellular fingerprint of morphometric trait-associated genes 67
6. Discussion 70
6.1 Overview 70
6.2 Generalizability of findings beyond early DKD 71
6.3 Bioinformatic approach captures crosstalk of processes 73
6.4 Upstream regulator of clusters 74
6.5 Systems biology identifies known and new targets associated with early structural lesions of DKD 75
6.6 Cellular source of structure-associated functions 77
7. Limitations 79
8. Conclusion 79
9. Future perspectives 81
References 84
Acknowledgements 93
Affidavit 94

ZUSAMMENFASSUNG

Die diabetische Nierenerkrankung (DN) ist ein Hauptrisikofaktor für chronische Nierenerkrankungen und Nierenversagen. Die DN stellt hierbei eine heterogene Erkrankung mit multifaktorieller, komplexer Pathogenese dar. Die Interaktion von intrinsischen Nierenepithelzellen des Nephrons mit gewebsresidenten und invadierenden Immunzellen wird als entscheidender Faktor für die Erkrankungsentstehung und Progression verstanden. Die subklinische Erkrankungsentwicklung in der Initialphase der Erkrankung erschwert die frühe klinische Identifizierung und Erforschung der DN und ist damit ein Haupthindernis für eine effektive Therapieentwicklung

Dieses Hindernis stellt die Motivation füer die hier vorgestellte Arbeit dar. Mittels eines systembiologischen Ansatzes von Protokoll-Nierenbiopsien in der Frühphase der DN soll eine umfassende Erkrankungsdefinition mittels Integration von genomischen, morphologischen und klinischen Datenquellen erstellt werden. Ein besonderer Fokus der Arbeit ist hierbei die Definition der Interaktionen zwischen morphometrischer und transkriptioneller Analyse von Protokoll-Biopsien bei früher, sub-klinscher DN in Patienten mit Type 2 Diabetes des Pima Eingeborenen Stammes im Süd-Westen der USA. Die Analysen nutzen Daten einer 6-järhigen klinischen Therapiestudie die die Nierenprotektion von Losartan gegenüber Plazebo untersuchte (siehe ClinicalTrials.gov: NCT00340678). 71 Studienteilnehmer mit einer Protokoll-Biopsie am Studienende wurden in die bioinformatischen Analysen der quantitativen Morphometrie und der Biopsie-Genexpressionsanalyse eingeschlossen.

Das kortikale interstitielle Volumen (VvInt) zeigte hierbei die signifikanteste Assoziation zwischen ultrastruktureller Schädigung und der tubulären Genexpression. Transkriptionelle Ko-Expressionsanalysen identifizierten vier Module mit signifikanter Assoziation zum interstitiellen Fibrosegrad. Hierbei zeigten 930 Transkripte eine positive und 913 eine negative

Korrelation mit der Fibrose. Eine Funktionsanalyse der Transkripte zeigte eine Häufung von inflammatorischen, migratorischen und tubulär-metabolische Funktion assoziierten Genen. Im nächsten Analyseschritt wurden die Gene identifiziert, die die verschiedenen Signalwege am besten in Signalknotenpunkten verbinden konnten. Interleukin 1B stellte sich als ein solcher Hauptregulationsfaktor dar, der die verschiedene VvInt-assozierte Signaltransduktionswege ansteuern kann. VvInt-assozierte Transkripte, gemessen in der Nierenbiopsie, zeigten eine signifikante Korrelation mit der Albuminurie und der gemessen glomerulären Filtrationsrate 8.2 Jahre (Median) nach der Biopsie.

Ein paralleler Ansatz für die Assoziation der glomerulären Expressionsdaten mit der glomerulären Basalmembrandicke, der mesangialen Expansion und der podozytären Schädigung. Die assozierten Gene waren für Immunzellmigration, Komplement-Aktivierung, mTOR- und Rac-Signaluebertagung und axonale Entwicklungprozesses angereichert.

Abschließend wurde die zelluläre Lokalisation der Expressions-Signale mittels Kartierung in Einzel-Zell Genexpressionsdaten definiert. Proximale Tubulusepithelzellen stellten eine Hauptquelle für die VvInt-assoziierten Transkripte dar, Mesangialzellen für die glomerulären Strukturen und zu allen Prozessen steuerten Immunzellen Transkripte bei.

Zusammenfassend wird in der hier eingereichten Arbeit eine systembiologische Integration von intra-renalen Genexpressionsmustern mit der quantitativen Morphometrie in Nierenbiopsien vorgestellt. Es werden molekulare Mechanismen der frühen DN identifiziert, die signifikant mit der langfristigen Erkrankungsentwicklung assoziiert sind. Die hier vorgestellten Ansätze und Daten etablieren neue molekulare Therapieziele für die medikamentöse Therapie und neue Biomarker für die Früherkennung der DN, beides dringend benötigte Therapiewerkzeuge.

Summary

Diabetes is a major risk factor for the development of chronic kidney disease and a major cause of end-stage kidney failure. Diabetic kidney disease (DKD) is a multifactorial heterogenous disease with a complex pathogenesis. Crosstalk between resident kidney cells and immune cells plays a critical role in DKD initiation and progression. The long clinically silent phase of the disease makes it challenging for clinicians to identify and provide intervention to slow disease progression at an early phase when such intervention may be most beneficial. A comprehensive systems biology approach was developed in this thesis to provide an unbiased view of mechanisms behind DKD progression active in early disease. This method was derived from sources across genomics, morphogenomics, proteomics, and phenotypic data sets. Specifically, bioinformatic methods were employed to identify potential interactions between transcriptomics and morphometrics in kidney biopsies obtained from patients with early DKD in Pima Indian populations with type 2 diabetes. Participants were part of a 6-year clinical trial primarily evaluating the renoprotective efficacy of losartan versus placebo (ClinicalTrials.gov number NCT00340678). Seventy-seven participants who underwent protocol kidney biopsies at the end of the trial were included in this study. Quantitative morphometric scoring and tissue genome-wide gene expression profiling was performed on microdissected tissues from protocol biopsies. The ultrastructural lesion most strongly associated with tubular transcriptional profiles was cortical interstitial fractional volume (VvInt), a scale of interstitial fibrosis. Transcriptional co-expression network analysis revealed four gene modules significantly correlated with interstitial fibrosis. Of the 1,843 genes in these modules, 930 correlated positively and 913 correlated negatively with VvInt. Inflammatory and migratory mechanisms, along with pathways related to differentiated tubular metabolic function, were among the mechanisms enriched with greater interstitial fibrosis. Pathway network analysis was then carried out to highlight key genes that connected multiple signaling pathways. IL1B

was identified as a top upstream regulator for the core set of transcripts linked to multiple VvInt-associated signaling pathways. VvInt-associated transcripts at the time of biopsy showed a significant correlation with albuminuria and measured glomerular filtration rate a median of 8.2 years after biopsy. These results suggest a potential predictive association between early structural injury and disease progression. A similar approach was used to assess the role of early glomerular injury, characterized by increased glomerular basement membrane width, mesangial expansion, and podocyte injury, in DKD progression. Pathway functional analysis revealed gene expression fingerprints of immune cell trafficking, complement systems activation, mTOR, Rac signaling, and axonal guidance pathways associated with early glomerular injury. To further validate these results, these structural correlates were projected on transcriptional profiles at the single-cell level that showed enrichment of VvInt transcripts in proximal tubular epithelial cells and of transcripts associated with glomerular phenotypes in mesangial/smooth muscle cell types. Ample evidence of enrichment was observed on immune cell types, myeloid cells, and T cells.

In summary, we developed a novel approach integrating intra-kidney gene transcript expression with quantitative morphometry. This method can be applied to identify molecular mechanisms activated in early DKD that may help predict major health outcomes. The identified genes linked to early structural damage and long-term kidney function are potential candidates for the identification of non-invasive biomarkers for progressive kidney disease. These genes also represent promising targets for early therapeutic intervention. This systems biology approach using bioinformatics-based methods can contribute to the identification of altered mechanisms that occur early in the disease course. These methods may be applied to other types of pharmacogenetic analysis and expanded precision medicine.

Abbreviations

ACR	Albumin-to-creatinine ratio
ALDH2	Aldehyde dehydrogenase 2
ATL	Ascending thin loop of Henle
BMI	Body mass index
CCL1	C-C motif chemokine ligand 1
CCL2	C-C motif chemokine ligand 2
CCR5	C-C motif chemokine receptor 5
cDNA	Complementary DNA
CKD	Chronic kidney disease
CNT	Connecting tubule
CXCL10	C-X-C motif chemokine ligand 10
CXCR4	C-X-C motif chemokine receptor 4
DCT	Distal connecting tubule
DKD	Diabetic kidney disease
DMEM/F12	Dulbecco's Modified Eagle Medium/Nutrient Mixture F- 12
DMSO	Dimethyl Sulfoxide
DTL	Descending loop of Henle
EC	Endothelial cells
EGF	Epidermal growth factor
eQTL	Expression quantitative traits
ESKD	End stage kidney disease
FBS	Fetal bovine serum
FDA	Food and Drug Administration
FEN	Percent of endothelial fenestration
FPW	Foot process width in peripheral GBM
GBM	Glomerular basement membrane width
GFR	Glomerular filtration rate
GV	Glomerular volume
Hba1C	Glycated hemoglobin A1c
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HEY1	Hes related family bHLH transcription factor with YRPW motif 1
IC	Intercalated cells
ICAM1	Intercellular adhesion molecule 1
IL-8	C-X-C motif chemokine ligand 8

IL1B	Interleukin 1 Beta
IQR	Interquartile range
JAK1	Janus kinase 1
JAK2	Janus kinase 2
lnRNA	Long noncoding RNA
MAPK8	Mitogen-activated protein kinase 8
MC	Mesangial Cell
mGFR	Iothalamate glomerular filtration rate
MMP2	Matrix metallopeptidase 9
MMP9	Matrix metallopeptidase 2
mRNA	Messenger RNA
NFAT5	Nuclear factor of activated T cells 5
NFKB1	Nuclear factor kappa B subunit 1
NVPC	Numerical density of podocyte cell per glomerulus
PC	Principal cells
PCR	Polymerase chain reaction
PEC	Parietal epithelial cell
PHF1	PHD finger protein 1
POD	Podocyte
PRKCQ	Protein kinase C theta
PTEC	Proximal tubular cells
RAC1	Rac family small GTPase 1
RNA	Ribonucleic acid
SOX2	SRY-box transcription factor 2
STAT3	Signal transducer and activator of transcription 3
SV	Surface volume of peripheral GBM per glomerulus
TAL	Thick ascending loop
TCA	Tricarboxylic acid
TNF-A	Tumor necrosis factor
TNFR2	TNF receptor superfamily member 1B
TP53	Tumor protein p53
TREM-1	Triggering receptor expressed on myeloid cells 1
TRIM29	Tripartite motif containing 29
TSP1	Thrombospondin 1
VEGF	Vascular endothelial growth factor A

VEGFA	Vascular endothelial growth factor A
VPC	Volume of podocyte cell
VPCN	Volume of podocyte nuclei
vSMC	Vascular smooth muscle cells
VVMES	Mesangial fractional volume
VVPC	Volume fraction of podocyte cell per glomerulus
VVPCN	Volume fraction of podocyte nuclei per podocyte cell
WGCNA	Weighted gene co-expression network analysis

1. Introduction

1.1 Chronic kidney disease

In addition to excreting waste, excess water, and acid, the kidneys control mineral and salt levels in the blood. They also reabsorb and recycle important metabolites and control blood pressure. This homeostatic balance is crucial to the functioning of other organs and tissues. Acute kidney injury can inhibit or prevent the kidneys from performing these routine functions, leading to chronic kidney disease (CKD) in some instances. CKD can develop in individuals of any age, race, or gender; however, the condition is more common in women. CKD is a leading global health concern with an estimated prevalence of 11–13% worldwide. A 2010 Global Burden of Disease Study showed that CKD had risen from the 27th to the 18th leading cause of death in the world [1]. Developed and developing countries both demonstrate a high prevalence of CKD: 12% of cases occur in Europe, the Middle East, East Asia, and Latin America, while 7% are found in South Asia. Per a report from the World Congress of Nephrology (2017), Germany ranked third (17%) in CKD prevalence. General increases in the rates of diabetes and hypertension can be partly blamed for the elevated risk of CKD seen in the United States, especially in African Americans, Hispanics, American Indians, and people of South Asian origin (World Kidney Day: Chronic Kidney Disease, 2015 http://www.worldkidneyday.org/faqs/chronic-kidney-disease/). Other factors contributing to the spectrum of CKD include poor glycemic control, underlying autoimmune disease, and a family history of kidney disease (Table 1). CKD causes the kidneys to lose function over time; progression to end-stage kidney disease leads to kidney failure, which requires either dialysis or a kidney transplant [2].

Table 1: Risk factors for chronic kidney disease.

Risk Factor	Definition	Examples
Susceptibility agents	Increased susceptivity to kidney damage	Family history of kidney disease, reduced kidney mass, lower birthweight, aging race-ethnicity, socio-economic elements low education, lower income.
Initiation agents	Initiate kidney injury	Diabetes, autoimmune disease, hypertension, urinary tract infections
Progression agents	Leads to worsening kidney injury, resulting in rapid decline in kidney function	Heavy proteinuria, blood pressure, high HbA1c, smoking
End-stage agents	Increase morbidity and mortality related to kidney failure	Low serum albumin level, vascular stress, lower dialysis dose

Table 1 Legend: Categories of risk factors for chronic kidney disease (CKD), including a family history of CKD, socioeconomic and demographic factors, high blood pressure, poor glycemic control in diabetes, and vascular stress leading to end-stage kidney disease. Any of these risks (alone or in combination) can lead to both kidney and cardiovascular complications and organ damage.

Figure 1: Trajectory of kidney functions and complications leading to end-stage kidney disease

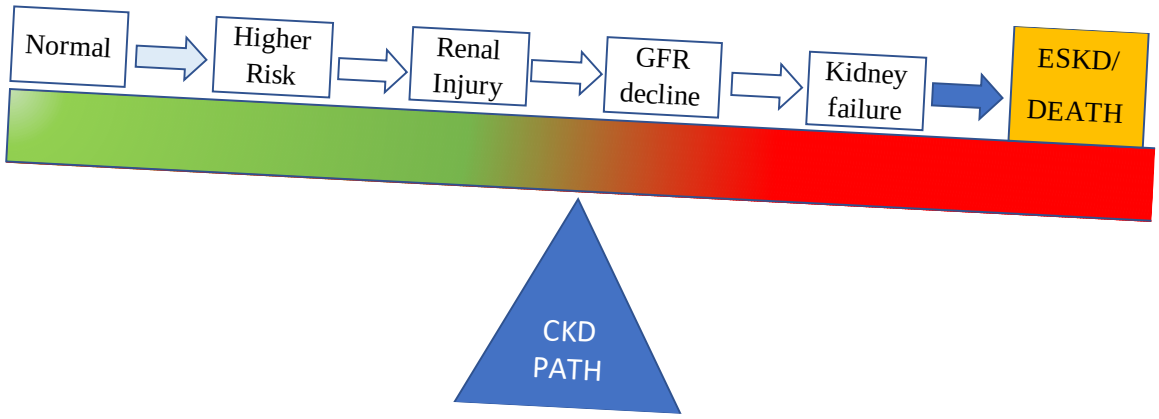


Figure 1 Legend: A framework of the course of chronic kidney disease (CKD). Risk increases towards the right extreme, tipping the balance from normal to disease progression. If left untreated, an increased risk of CKD leads to further kidney and cardiovascular complications. This mechanism can further lower glomerular filtration rate (GFR), leading to kidney failure

and requiring initiation of dialysis and to transplantation. Each rectangular box represents a phase of kidney function. The arrows between boxes depict the progression to different stages after the initial insult. The intensity of blue denotes the severity (i.e., from light blue to dark blue), indicating normal to late stages of kidney function decline. (Adapted from Lesley A et al. CHAPTER 53 - Staging and Management of Chronic Kidney Disease.(Lesley A. Stevens, Nicholas Stoycheff, Andrew S. Levey.

Many contributing factors to CKD are attributed to a lack of awareness of the disease among the general population. CKD is progressive and usually leads to end-stage kidney disease within 10 to 20 years after initial diagnosis. A rapid decline in kidney function is a significant predictor of end-stage kidney disease. Therefore, early detection and management are crucial in slowing or halting the progression to end-stage kidney disease and other adverse outcomes of CKD, including cardiovascular disease.

1.2 Prevalence of kidney disease

The estimated glomerular filtration rate (eGFR) is currently the recommended measure of kidney function, with different eGFR levels denoting various stages of kidney disease. Normal kidney function occurs at >60 mL/min/1.73m^2, mild to moderate loss of kidney function occurs at 59–45mL/min, and kidney failure occurs at less than 15 mL/min. The German Health Interview and Examination Survey for Adults included 2008–2011 data to provide a representative estimate of the prevalence of kidney dysfunction in Germans [3]; approximately 1.53 million German adults had an eGFR of <60 mL/min.1.73m^2. The prevalence of CKD in the United States reflects the global picture, with kidney disease as the 9th leading cause of mortality. The latest United States Kidney Data System report indicated the overall prevalence of CKD in the adult U.S. population to be 14.9% (2015–2018).

(https://www.usrds.org/2017/view/v1_01.aspx). Incident end-stage kidney disease cases increased by 2.3% in 2018 over 2017. Figure 2 illustrates the striking rise in the prevalence of end-stage kidney disease from 2015 to 2018 across ages, races, and sexes.

Figure 2: Prevalence of end-stage kidney disease over a 4-year period

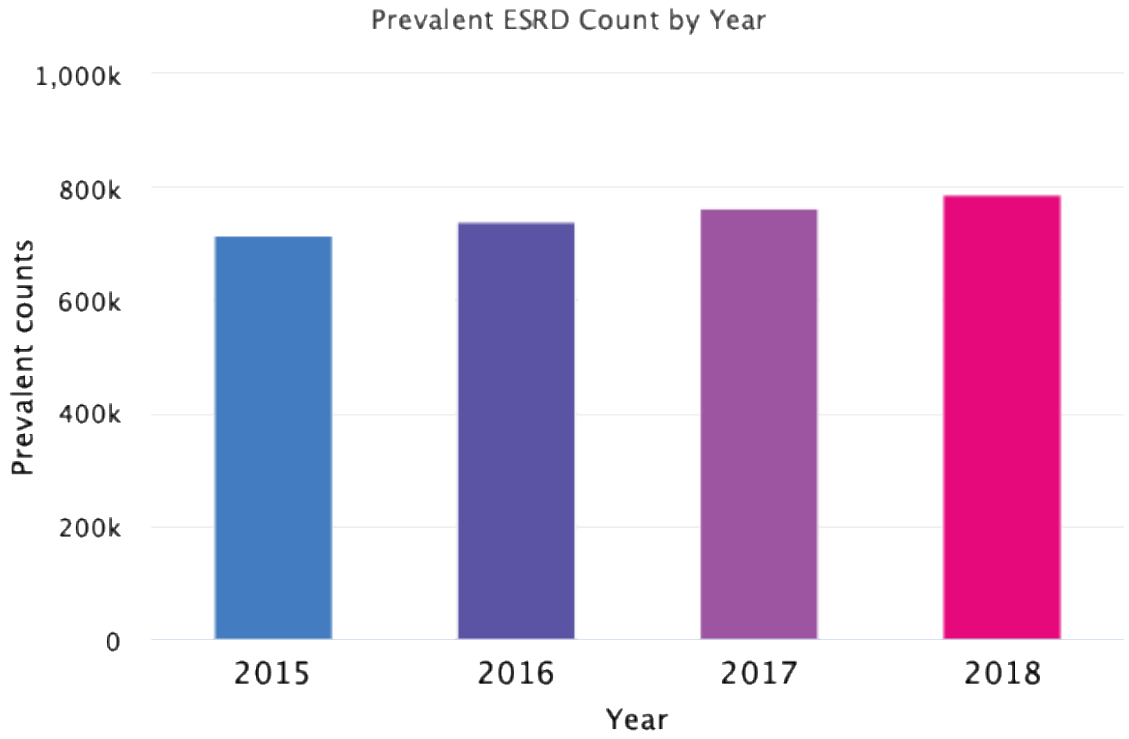


Figure 2 Legend: Adapted from https://www.usrds.org. The prevalence of end-stage kidney disease from 2015 to 2018 across ages, sexes, and races. Ethnicities included Hispanic/Latino (Non/Yes) and unknown. Source: States Kidney Data System, 2020 Annual Data Report: Epidemiology of Kidney Disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 20892.

1.3 Kidney physiology

The kidneys regulate blood volume and excrete excess water, metabolic products, and toxins from the body; the kidneys filter approximately 200 liters of fluid per day. As a highly vascularized organ, the kidneys are composed of an outer layer called the kidney cortex, a middle layer named the kidney medulla, and the inner kidney pelvis region. Each kidney

possesses roughly 1 million nephron structures that represent functional or filtering units. A single nephron comprises a kidney corpuscle, proximal convoluted tubule, loop of Henle, and distal convoluted tubule (Figure 3, created by Biorender.com). Nephrons begin in the cortex with the kidney corpuscle, proximal tubule, and distal tubule and extend into the medullary region with the loop of Henle and collecting duct. The peritubular capillaries lie next to the entire tubular length and play a critical role in solute transport. The kidney corpuscle responsible for the filtration of plasma consists of two main components: the glomerulus and Bowman's capsule. More specifically, the glomerulus is a ball of capillary tubes enclosed by Bowman's capsule. The inner layer of highly specialized epithelial cells is referred to as podocytes; the outer epithelial layer extends to proximal epithelial cells; and the space between these two layers represents Bowman's space, which collects and transports ultrafiltrated plasma. The glomerulus filters fluids, glucose, metabolites, and waste products from the glomerular capillary into Bowman's capsule. This filtration process is achieved as plasma passes through filtration barriers composed of the capillary endothelium, podocyte foot processes, and the vascular basement membrane. Mesangial cells are present within the glomerulus, where they produce basement membrane–like structures that support the glomerular complex structure. As filtered fluids pass through the proximal and distal convoluted tubules and the loop of Henle, ions and molecules are reabsorbed into the circulatory system. This process is crucial to maintaining homeostasis between glomerular filtrate and the blood. Nearly 65% of filtrate reabsorption occurs within the first tubular segment, the proximal tubules, which leads to the hair-pin structure of the loop of Henle in the medullary region. This kidney tubular section provides high osmotic pressure that facilitates the reabsorption process in later tubular segments. Further reabsorption and secretion of ions occur in the distal convoluted tubular region. Any remaining fluid gathers in the collecting duct—the terminal region of the distal convoluted tubule—for secretion to the ureter. GFR, as

detailed below, can be used to indirectly measure the total amount of filtrate formed and is proportional to fluctuations in the hydrostatic and osmotic pressures in Bowman's capsule and the glomeruli.

Figure 3 Nephron schema

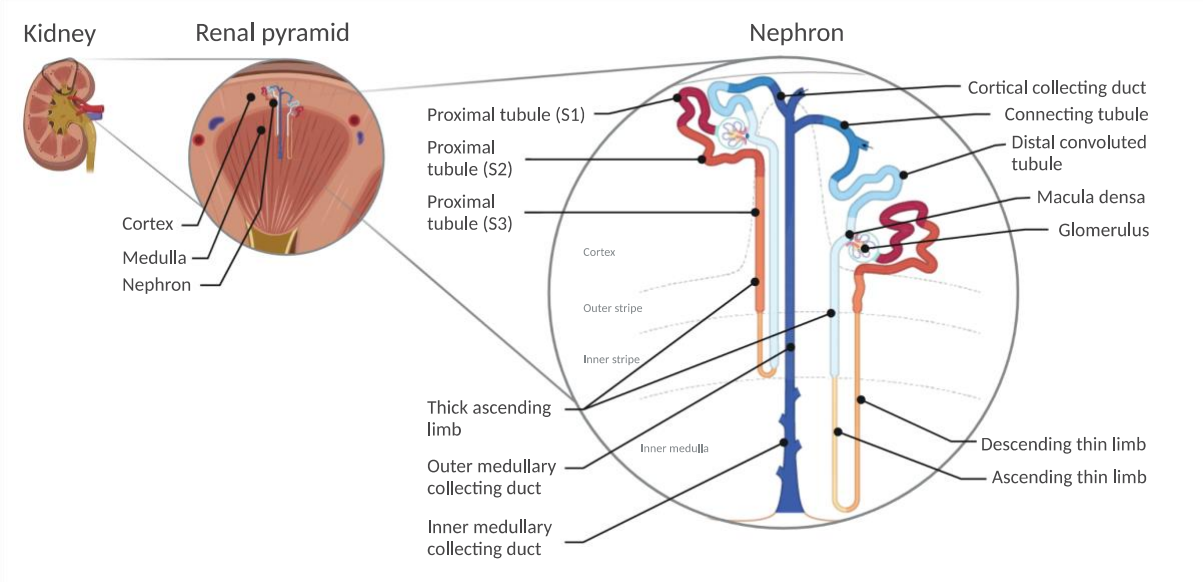


Figure 3 Legend: Schematic diagram of the kidney with an inlay figure of a single nephron showing the glomerulus and tubular segments. This figure depicts the arrangement of nephron segments within a single nephron. The round glomerulus connects to the collecting duct via tubular interstitial components of the proximal convoluted tubule, peritubular capillaries, loop of Henle, and distal convoluted tubule. Source: Created by biorender.com

1.4 Parameters used to measure kidney function

GFR is defined as the amount of blood filtered by the glomerulus per unit of time and is currently considered the best clinical indicator of overall kidney function. In addition to GFR, measures of proteinuria and albuminuria are widely used to evaluate kidney function and represent accepted parameters of kidney health[4]. Structural abnormalities detected during kidney biopsies are routinely used to determine the extent of kidney damage and to help

characterize the pathophysiologic nature of the specific disease entity. The National Kidney Foundation proposed a CKD definition and classification based on GFR (Table 2) in 2002.

Table 2: CKD stages based on kidney function

CKD stages	Description	GFR (mL/min/1.73m^2)
1	Normal kidney function	≥90
2	Mild kidney damage	60-89
3	Moderate kidney damage	30-59
4	Severe kidney damage	15-29
5	Kidney failure	>15

Table 2 Legend: CKD: chronic kidney disease; GFR: glomerular filtration rate. This table lists stages of kidney disease based on the kidney function and GFR at the time of presentation. Stages 4 and 5 are defined as progressive and end-stage kidney disease, respectively. Colors in the arrow range from low risk (green) to high risk (red).

GFR can be estimated in several ways. Methods include the Modification of Diet in Kidney Disease study approach, the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI), Berlin Initiative Study 1, and a modified version of the CKD-EPI method based on Cystatin C. However, recent studies have raised concerns about the widespread use of GFR as the main determinant of CKD staging. A major criticism of the 2002 classification scheme is the potential lack of precision of GFR estimates, particularly when estimated GFR is >60 mL/min/1.73m^2 [5]. Another confounder is the low rate of patients (~2%) who are seen to progress from Stage 3 CKD. The higher prevalence of CKD and low GFR is similarly confounded by age. These issues have led to reclassified guidelines: proteinuria and albumin

creatinine ratios are now included for more robust kidney disease staging [6-15]. The National Kidney Foundation Kidney Disease Outcomes Quality Initiative and the international guideline group, Kidney Disease: Improving Global Outcomes (KDIGO) recently proposed an updated CKD classification that incorporates albuminuria changes (Figure 3).

Figure 3: Classification of CKD by GFR and albuminuria category

			Albuminuria categories (mg/g)		
	Classification of CKD by GFR & Albuminuria		A1	A2	A3
			<30	30-300	>300
GFR categories (ml/min/1.73m²)	G1	≧90			
	G2	60-89			
	G3a	45-59			
	G3b	30-44			
	G4	15-29			
	G5	<15			

Figure 3 Legend: Modified chart of the KDIGO classification of CKD stages based on GFR and albuminuria levels. GFR categories range from 1 to 5; Category 3 is divided into 3a and 3b. Albuminuria categories of low, mild, and high are based on the spot urine albumin-to-creatinine ratio (A1, A2, and A3, respectively). Colored blocks indicate the risk of CKD by crossreferencing risk by GFR/albuminuria. Green: low risk, no CKD, no prior kidney disease; Tan: moderately increased risk with mild lowering of GFR and intermediate increase in albuminuria; Pink: persistent albuminuria increase and GFR loss; Purple: high albuminuria and severe kidney damage with possible kidney failure.

The root causes of CKD can be attributed to a diverse set of etiologies, such as diabetic kidney disease (DKD), hypertension, glomerular disease, tubulointerstitial disease, and recurrent acute kidney injury. According to the National Institute of Diabetes and Digestive and Kidney Disease, diabetes and high blood pressure are the two major contributors to CKD.

1.5 Diabetic kidney disease

As mentioned, DKD is a major global cause of CKD and end-stage kidney disease. Approximately 463 million adults have been diagnosed with diabetes to date; this figure is expected to grow to 700 million by 2045 [16]. Yet diabetes often goes undiagnosed in 1 in 2 adults with the disease, suggesting that the true number of affected individuals is higher. The 2019 International Diabetes Federation Diabetes Atlas reported that 10% of total global health expenditure is devoted to diabetes management, resulting in a substantial healthcare burden (International Diabetes Federation. IDF Diabetes Atlas, 9th edn. Brussels, Belgium: 2019. https://www.diabetesatlas.org) [17-19]. The most recent United States Kidney Data System report (https://adr.usrds.org/2020) highlighted diabetes as the top cause of end-stage kidney disease (Figure 4) followed by hypertension and glomerulonephritis.

Figure 4: Prevalence of end-stage kidney disease by cause

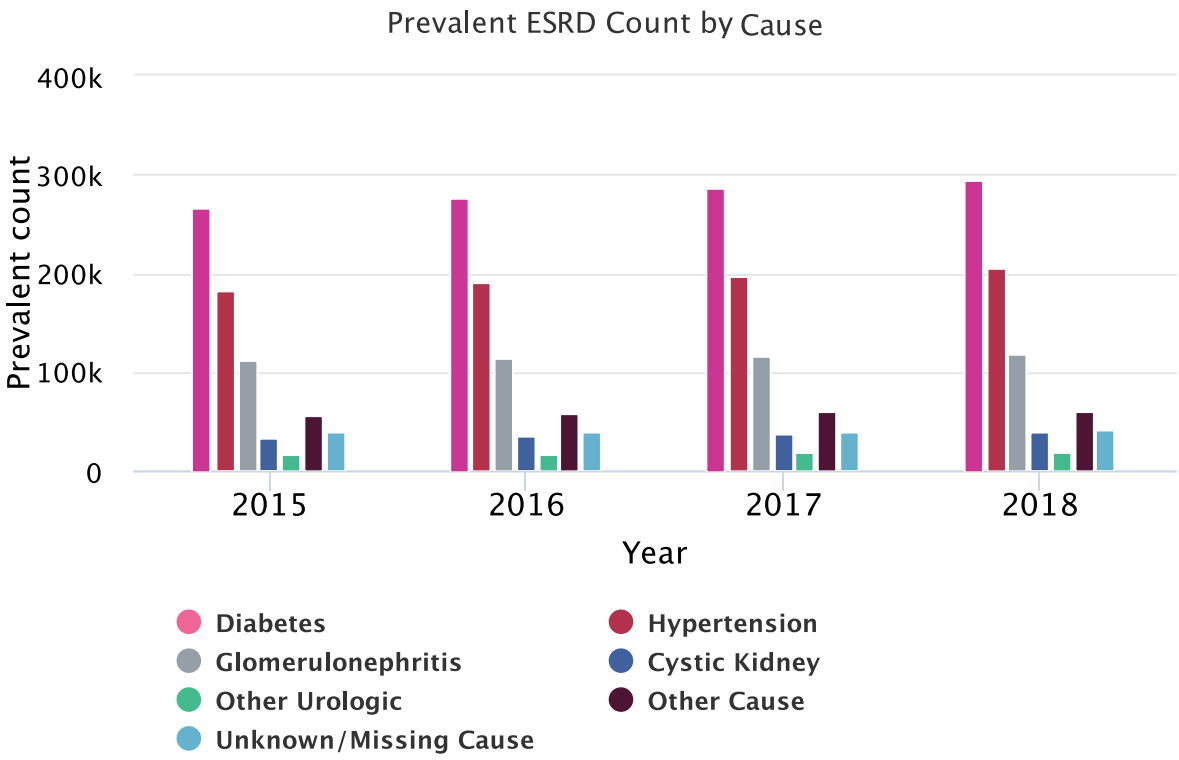


Figure 4 Legend: Modified from https://www.usrds.org. Prevalence of end-stage kidney disease by cause in 2015–2018 across all ages, sexes, and races. Ethnicities included Hispanic/Latino (Non/Yes) and unknown. Source: States Kidney Data System, 2020 Annual Data Report: Epidemiology of Kidney Disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 20892.

DKD is a complex progressive disease strongly linked with microvascular complications. High circulating levels of glucose stress both the kidney glomerulus and tubulointerstitum. If left untreated, this stress can develop into CKD and eventually end-stage kidney failure. DKD is also a significant risk predictor of cardiovascular disease. Given the health impact and financial burden of DKD, as well as its contribution to CKD, disease management is urgently needed. The current clinical regimen of diabetes management through blood pressure control and lowering hemoglobin A1c has reduced disease incidence; however, early intervention in CKD

progression remains necessary to halt or potentially reverse its course. The current focus on personalized medicine reinforces the importance of identifying subsets of patients who may respond more effectively to certain treatments.

Moderate albuminuria (microalbuminuria) is a broadly applied standard that nephrologists use for early detection of kidney complications as well as DKD progression. This standard has been widely adopted since Viberti and colleagues first described it in 1982 [20]. The diagnostic criteria have been updated repeatedly and now include new methods of urinary albumin excretion detection. Moderate albuminuria (microalbuminuria)is generally defined as a urinary albumin excretion rate of 30–140mg/min as indicated by sample type: 30–300mg/day in a 24-hr urine specimen, 20–200µg/min from a timed urine sample, or 30–300mg/g from a spot urine sample. The latest Kidney Disease Outcomes Quality Initiative guidelines labeled levels below 30 as a normal albuminuria range and those above 300 as macroalbuminuria. Non-albuminuric paths also exist in diabetes progression [21-25].

GFR estimates that are based on approaches such as the Modification of Diet in Kidney Disease study approach and the CKD Epidemiology Collaboration are largely derived from estimates of creatinine measurements. Such estimates have been criticized for their lack of precision in CKD diagnosis/staging, especially in diabetic populations. A direct measurement of GFR (i.e., iothalamate GFR) using kidney clearance of exogenous tracers is currently considered the gold standard clinical method. However, the practical measurement of iothalamate GFR renders it a challenging routine assessment. DKD is conventionally characterized based on eGFR together with the urinary albumin-to-creatinine ratio (ACR); this measure is used to stratify patients who may be susceptible to progressive kidney disease [26-30].

The variable course of DKD in its initial stages also presents a challenge to diagnostic approaches intended to identify at-risk populations with early clinical disease [31, 32]. The progression to end-stage kidney disease in the normal albuminuric stage [33] and the regression

of moderate albuminuria (microalbuminuria) in some patients further complicates diagnostic and therapeutic approaches [23, 34-36]. Additionally, scholars have recently pointed out inconsistencies in GFR and ACR in terms of characterizing DKD progression [37].

1.6 Kidney biopsy

Kidney biopsy remains a key method of assessing kidney damage. Following early hypertrophy, morphological changes develop that lead to glomerular basement thickening. Such thickening represents an early structural abnormality in DKD and is often followed by interstitial lesions and mesangial expansion in the glomerulus. These morphological changes can occur well before clinical manifestation as evidenced by eGFR or albuminuria. Several studies have indicated that applying histological features from biopsy tissues can predict disease progression much earlier than classical clinical parameters [38-40]. For other conditions such as cancer, biopsy material has been used for histological and molecular profiling to classify high-risk patients and to identify potential therapeutic targets. The molecular profiling of tissue derived from damaged organs can provide greater insight into disease staging than cell culture methods. Tissue microdissection, either via laser capture or manual dissection, can also provide valuable information about cell-specific origins of regulated gene signatures, pathways, and molecular mechanisms. General analysis of the disease process is further enhanced when these change patterns are compared with patterns in healthy tissue. With a sufficient number of samples, statistical power can detect pattern-based differences within disease groups; these differences can then be used to identify at-risk patients and novel drug targets and has been showed in several studies.

2. Bioinformatics

The field of bioinformatics emerged before the advent of DNA sequencing. In her pioneering work after World War II, Margaret Dayhoff applied state-of-the-art computational methods to create an atlas of protein sequences; she has since been deemed the mother of bioinformatics. The terms “bioinformatics” and “computational biology” are often used interchangeably. Both approaches embrace the use of computational methods to decode complex aspects of biology. Bioinformatics refers to the specialized application of computational and mathematical methods to store, organize, analyze, understand, visualize, and integrate information associated with molecular entities [41]. This domain merges specific knowledge using computer science, statistics, mathematics, and biology to identify biologic mechanisms embedded in large biologic datasets [42]. The field of bioinformatics also encompasses medical informatics, which involves the integration of clinical data and patient- or disease-specific information.

The pre- and post-genomic eras in bioinformatics can be differentiated by the publication of first complete sequence of the human genome. Methods used in bioinformatics/computational biology later expanded from computation to the storage, retrieval, and characterization of complex patterns extracted from terabytes of data derived from the data explosion accompanying recent technological advances. Biologists are now faced with the challenge of extracting knowledge from “big data.” With developments in omics data generation, more advanced bioinformatics techniques are required in statistical and computational spaces to process related information. Machine learning algorithms and artificial intelligence tools are being used to establish deep learning neural networks. Medical bioinformatics presently focuses on constructing appropriate algorithms to decode data and address questions within defined biological settings (e.g., specific disease trajectories).

2.1 Systems biology

Although the term “systems biology” is widely used, confusion persists regarding when and how to apply it. This uncertainty follows the shift from a traditional single mechanism–driven approach to a more data-driven, integrative means of characterizing biologic processes. Systems biology seeks to explain the complex dynamic interactions underlying specific processes to establish links between biomolecules and aspects of physiology. There has been a large influx of research embracing systems biology concept over the last decade rising from just two studies published in 2000 to more than 13K studies by 2020. The current systems biology is considered to evolve through three different phases: systems molecular biology, systems-mathematical biology and finally to system-medicine leading to exponential growth in drug discovery [43].

The application of bioinformatics and systems biology requires the integration of diverse scientific approaches to generate hypotheses that can be iteratively refined (Figure 5). Systems biology in medicine refers to using bioinformatics to gain disease-related knowledge at the patient level encompassing information across genome-phenome cascade from transcriptomics, proteomics, metabolomics, lipidomics. This information can reveal novel explanations about the pathophysiology of disease progression. This systems approach is especially beneficial in a multifactorial disease such as DKD as the disease pathophysiology cut across multiple organs/tissue [44].

Figure 5: Systems biology as a multidisciplinary science

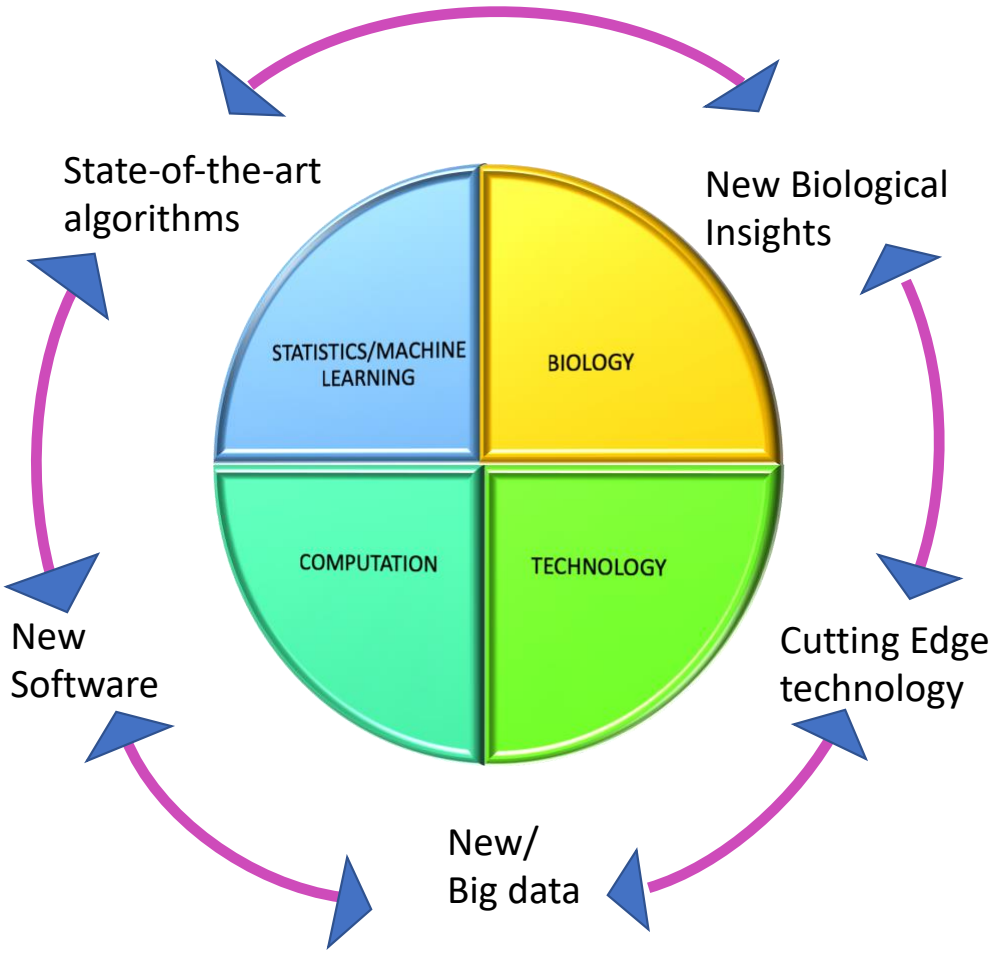


Figure 5 Legend: A pie chart displaying the major components of systems biology. Bioinformatics integrates these components. The outer circle of bidirectional arrows shows how each element leads to new ideas/concepts to be addressed by the other elements. Figure modified from the Institute for Systems Biology.

2.2 Precision medicine

Precision medicine or personalized medicine often used interchangeably refers to the same concept that changes "one size fits all" approach in medicine. Many treatment regimens and clinical identification methods focus on classic disease etiology without necessarily considering key differences within subgroups. Advances in molecular personalized medicine

have allowed for a more holistic view of patients (e.g., by integrating datasets describing transcriptomic–epigenetic modifications, proteomics, and metabolomic landscapes in addition to genetic polymorphisms). This multifaceted analysis offers several benefits: it can facilitate identification of genomic, transcriptomic, or proteomic features that predispose individuals to disease; predicts patients' responses to therapy; enhance diagnosis by identifying prognostic biomarkers; and provide potential mechanistic insights into disease processes. This general approach can also stratify patient subgroups, thereby enabling more precise treatment (Figure 6) with the expectation that the treatment can be more customized to the genetic makeup of the individual patient.

Figure 6: Precision medicine integrating multimodal networks

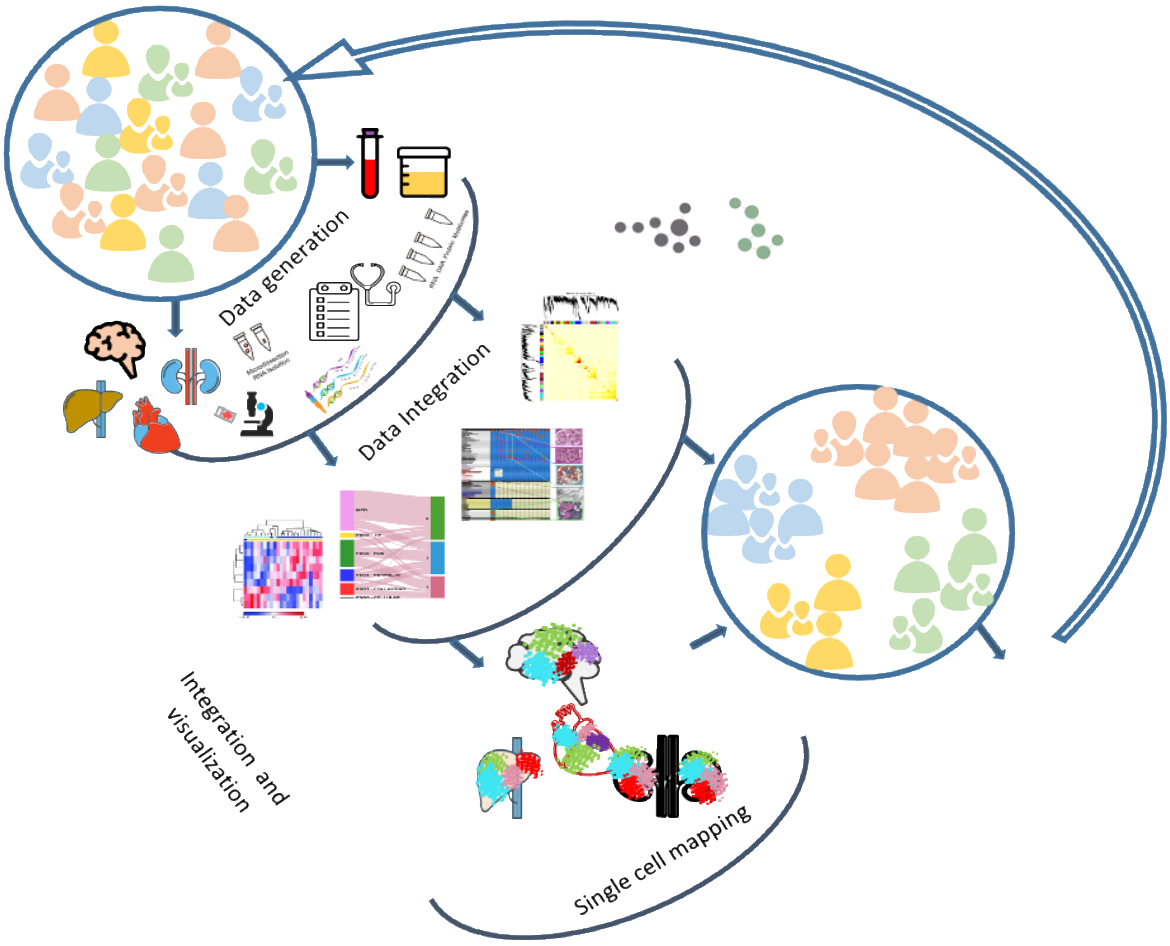


Figure 6 Legend: A schematic summary of data integration beginning with a heterogenous population. This approach is flexible enough to be adapted for any organ/tissue type. Data integration and pathway-based knowledge in specific cellular contexts facilitate patient stratification in clinical trials.

The coordinated efforts of many research groups have also inspired kidney researchers to assume a more personalized medicine approach, largely based on integrating bioinformatics and systems biology methodologies [45-47]. Cancer research has embraced the concept of disease group variations and integrated this approach of precision medicine [48]. These differences have led to subclasses of cancer types and breakthroughs in target identification and drug discovery.

As discussed, GFR and albuminuria are currently standard in the clinical characterization of DKD. However, each method has shortcomings. Variations in creatinine measurement and its inefficiency in adequately capturing tubular and vascular damage limit its utility in precisely characterizing the disease course. Patient rates of DKD progression are highly variable: some individuals with diabetes progress rapidly to the end stage while others remain in a mild to moderate stage for longer periods [49]. These variations are partially due to the molecular processes associated with disease progression. Systems biology is well positioned to identify the biology underlying this heterogeneity based on diverse omics data. Combining molecular profiling with pathologic criteria is one way to identify the underlying differences in these patients with diverse progression and thereby advance personalized medicine in kidney disease.

2.3 Enhanced characterization of disease models through systems biology

Disease model platforms have greatly advanced basic scientific research. Experimental organisms such as the mouse, rat, and zebrafish have enriched our understanding of pathophysiologic processes. Unfortunately, these models often fail to exactly recapitulate the

disease mechanisms seen in human disease [50]. Several drug targets had to be withdrawn from clinical trials due to the serious adverse effects which were not evident during the preclinical testing [51]. Systems biology approaches can be useful for identifying general pathways and molecular networks shared between species. In some cases, these approaches have also been helpful in reinterpreting animal models in the human disease setting. Our group recently employed a systems biology approach to integrate human DKD pathways across three DKD mouse models.

A novel systems biology approach, Tool for Approximate Large (TALE), was developed for network comparison [52]. This method was applied where human diabetic nephropathy (DN) networks were compared with three experimental murine DN networks via graph matching [53]. The TALE method compares network structures and extracts overlapping subnetworks between the query network (i.e., human DN in this case) and the model network. Gene expression profiling from biopsied tissue was derived from a type 2 diabetes cohort and from the three mouse models. This approach involved microdissection of kidney tissue to extract enhanced glomerular-specific RNAs to improve tissue specificity. Differentially regulated genes were identified from each dataset using bioinformatics methods to compare the diseased and control groups. The differentially regulated genes were then used to generate transcriptional networks of regulatory pathways within each species and model. These networks were produced on the basis of functional associations between individual genes described in the literature. These networks were then provided as input for the subgraph matching algorithm. First, TALE ranks genes in the query network by their degree of connectivity within the network, essentially establishing seed genes. The method relies on two parameters, namely the seed gene percentage in datasets and the mismatch percentage. The seed gene percentage defines the number or percentage of important genes needed to build the network from the query dataset; the mismatch parameter is the percentage of missing genes

allowed when generating the overlap network. The mismatch percentage introduces flexibility into the algorithm to enable identification of a set of subnetworks that may partially overlap within datasets. When comparing networks between biological systems, incomplete orthogonality can occur between species or tissues at the gene level. In TALE, functional overlapping begins with identifying seed genes that are then used to identify functionally associated genes to extend the data to a potential network. This process eventually revealed key shared processes/networks between human DN samples and the three mouse models. The networks validated pathways previously established as playing a role in the biology of DN. This approach also uncovered novel pathways linked to the pathophysiology of this disease. Essentially, this method focuses less on individual genes and more on common changes in regulatory pathways and networks between human diabetic nephropathy (DN) data samples and mouse models. Generating specific networks can more effectively characterize the general pathophysiologic processes occurring between species or models.

This example demonstrates the advantage of systems biology approaches in identifying global mechanisms or pathways altered in DN models and patient data. Such methods can be useful in evaluating optimal experimental models for mechanistic studies. Systems biology can be leveraged to discern shared and unique networks as well. A study of lupus disease–associated kidney damage further validated this method's utility across disease etiologies [54]. Computation-based approaches thus show promise in advancing our understanding of biologic mechanisms. In-silico methods in particular can reduce the financial burden and time required to identify relevant models.

2.4 Single-cell RNAseq analysis and diabetic nephropathy

Single-cell RNA sequencing (scRNAseq) is a promising analysis tool that has progressed dramatically over the past few years; it can reveal transcriptional changes occurring in a

multitude of single-cells obtained from every sample of interest. Single-cell sequencing reveals how distinct (or orchestrated) different cell populations' contributions within heterogeneous tissues are to the tissue signal as a whole. This approach is especially powerful when single-cell sequencing is available from healthy control and diseased tissue. In addition, this information represents a new means for the in-silico dissection of tissue.

This technology, along with developments in material procurement and library generation to computational resources, has shown efficacy in fields such as oncology and neuroscience [55]. When coupled with bioinformatics techniques, scRNAseq can be employed to characterize rare and complex cell types and to establish regulatory relationships between genes/cell types [56-58]. This technology can also characterize cells obtained from biofluids as noninvasive liquid biopsies [59]. The cancer field has successfully captured the tissue heterogeneity of tumors with the high throughput sequencing technology but with little success in defining the circulating tumor cells (CTC). Deploying single cell methodology, the cellular level heterogeneity of patient derived CTC is now emerging, effectively identifying targetable biomarkers with cellular precision. The FDA has approved the use of circulating tumor cells as a clinical diagnostic tool, and scRNAseq can provide detailed information in this regard. Genetic integration via expression Quantitative Traits (eQTL) is presently carried out using bulk RNA transcriptomics. Recent research integrating scRNAseq signals with candidates from genome-wide association studies has revealed new cell-specific pathways and cell-specific eQTLs [60, 61].

Kidney tissue is heterogeneous; therefore, signals should be captured from distinct cell types. The kidney contains more than 50 segments/compartments, providing a particular challenge to generate single-cell data across the highly differentiated kidney cell types. With the development of kidney specific tissue dissociation and subsequent identification of individual cell cluster identification of transcripts characteristic for known kidney cell types has become

feasible. In addition, this approach allows to identify novel cell clusters, leading to a molecular (re-) definition of kidney cells in health and disease.

3. Hypothesis and goals of thesis

Systems biology and bioinformatic methods can be applied to identify networks of genes/mechanisms linked to the pathophysiology of DKD that manifest as pathologic or clinical criteria. This information can be used to classify disease stages or severity and to determine appropriate therapeutic strategies. In this thesis, we tested the hypothesis that network analysis can generate a systemwide perspective with mechanistic information to identify early disease parameters. Such a holistic approach would be especially valuable for diseases such as DKD, which may show clinical signs of progression over 15–20 years. A comprehensive approach integrating the molecular profiling of transcriptional changes and histological analysis could be used to identify fundamental pathological mechanisms that act as therapeutic targets as well as novel biomarkers seen at an early (i.e., pre-clinical) stage of disease.

The central goals of this thesis were as follows:

- To develop a bioinformatics-based method to characterize and then link molecular changes to morphological changes and clinical parameters that occur during the initial stages and progression of DN;
- To initially use tubular interstitial tissue profiling for the analysis of changes within this compartment, that is accepted as a prognostic feature of progressive DN;
- To validate key aspects of this analysis using single-cell analysis of parallel kidney tissue;

- To validate the general approach by analyzing the corresponding glomerular compartment in order to establish viability and applicability of the method in other tissue settings.

4. Material and Methods

4.1 Study cohorts

4.1.1 Early DKD cohort

A tribe of American Indians from the Gila River Indian community in Arizona (i.e., Pima Indians) are thought to have the highest incidence and prevalence of type 2 diabetes in the world. Diabetes often develops at a young age in this population. Pima Indians generally exhibit higher insulin resistance than their Caucasian counterparts, possibly due to genetics. A large cohort of Pima Indians from this community enrolled in a longitudinal study (1965–2007) of diabetes and its complications conducted by the U.S. National Institutes of Health (NIH) [35, 62]. The NIH study involved detailed clinical follow-up for more than two decades (including kidney biopsies). Given the greater likelihood of diabetes onset and duration in this population, the NIH study offered an ideal resource for investigating DKD progression.

A group of 169 adults from the Gila River Indian Community were enrolled in a randomized, double-blinded, placebo-controlled clinical trial (ClinicalTrials.gov number, NCT00340678). The primary goal of this trial was to assess the renoprotective effect of losartan, an angiotensin receptor blocker, in relation to the standard of care [63]. The primary study outcome was a GFR decline to ≤60 mL/min or half the baseline value if the entry GFR was <120 mL/min. Exclusion criteria included type 2 diabetes for <5 years and urinary ACR of >300 mg/g. Annual iothalamate GFR measurements were taken during the trial [32, 64] and during follow up after the trial. In total, 111 of the 169 participants underwent an ultrasound-guided protocol research kidney biopsy at the end of the clinical trial [63]. An investigator who was masked to the clinical data then performed quantitative morphometric measurements using unbiased random sampling. VvInt, glomerular basement membrane width, mesangial fractional volume per glomerulus, foot process width, and percentage of glomerular endothelial cell fenestration (a measure of endothelial injury) were assessed as previously reported [63, 65].

Urinary albumin excretion was evaluated via nephelometric immunoassay and measured annually. Urine and serum creatinine were measured with a modified Jaffé reaction [63, 65]. Urine albumin excretion in this study was presented as the albumin-to-creatinine ratio (ACR) in mg/g creatinine. The study was approved by the Institutional Review Board of the National Institute of Diabetes and Digestive and Kidney Diseases. All study participants signed an informed consent document.

4.1.2 Advanced DKD cohort

Indication biopsy samples from 17 advanced DKD samples were collected from the European Kidney cDNA Bank Cohort (ERCB). The ERCB cohort is a multicenter study across 24 European centers [66]. Thirty-one living donor biopsies were used as controls. These biopsies were obtained at the time of transplantation. Study participants were predominantly Caucasian. This study captured surplus biopsy material in an RNA fixative (RNAlater, Qiagen) for transcriptomic profiling. Limited clinical information (i.e., serum creatinine, proteinuria, and demographic variables) was also collected when kidney biopsy specimens were procured. Kidney function (i.e., GFR) at the time of biopsy was expressed using the MDRD formula. Biospecimens were collected after obtaining informed consent and following approval from the local ethics committee.

4.2 Tissue microdissection and microarray preparation

Kidney biopsies were harvested using a Max-Core 16G biopsy needle and Tru-Guide from Bard Peripheral Vascular (Tempe, AZ) with ultrasound guidance. Each tissue specimen was immediately placed in a 2.0mL cryovial containing 1.0mL RNALater from Ambion (Now

ThermoFisher Scientific). Biopsies were stored at -20°C until microdissection. Tissue samples were microdissected manually using nickel-plated pin holders and 0.25mm Tungsten needles from Fine Science Tools (Foster City, CA). Specimens were dissected into RNALater using an SZX16 dissecting microscope (Olympus, Tokyo, Japan), and each compartment (i.e., glomerulus and interstitium) was placed in RNALater at -20°C until RNA extraction. Total RNA was extracted with the Qiagen (Germantown, MD) Allprep Micro kit according to the protocol using optional beta-mercaptoethanol. RNA was analyzed for concentration and quality using an Agilent Bioanalyzer 2100 (Santa Clara, CA). cDNA was synthesized and amplified from 15ng total RNA using the NuGEN Ovation PicoSL WTA System V2 Kit (NuGEN Technologies, San Carlos, CA) according to the manufacturer's instructions. Amplified cDNA was isolated over MinElute Columns (Qiagen, Venlo, Netherlands) according to recommendations in the NuGEN Ovation manual.

Amplified cDNA was fragmented and biotinylated using the NuGEN Encore Biotin Module and prepared for hybridization according to the NuGEN Ovation manual with reagents from Affymetrix (Santa Clara, CA). Next, 16 ng/ul labeled cDNA were hybridized to the Affymetrix GeneChip Array Human Genome U133A and Plus 2.0 chips which were then washed, stained, and scanned according to the Affymetrix User Guide for Expression Array Plates. Glomerular and tubular RNA underwent similar processing.

4.3 Microarray data processing

Affymetrix *.cel image files from glomerular and tubular tissues were obtained and processed as described previously [67]. The *.cel files were screened for quality control using methods implemented in the Affy package in R software, an open-source statistical platform (https://www.r-project.org). Each *.cel file represented an individual patient. The steps

outlined in the Methods section were performed separately on glomerular and tubular data. All *.cel files were then read into R software. Quality control analysis and data preprocessing were conducted next. The RNA degradation plot is a quality control measure that plots the trend of RNA degradation from 5'-3'; an array with a steep slope compared to other arrays indicates RNA degradation. Density histograms of log-intensities from all arrays were superimposed into one graph to identify arrays whose distributions differed from others'. Normalized unscaled standard error plots and relative log expression plots can also reveal outlier arrays. For this study, arrays that behaved consistently differently on all quality control measures were removed; those that passed the quality control were then normalized using the robust multichip array method. This method includes three main steps: (1) background correction to remove background noise from the probe signal on each array; (2) normalization to remove variation introduced during target preparation and hybridization (i.e., to make data from all arrays comparable); and (3) summarization of probe intensities into gene- or probe set–level expression signals depending on the downstream analysis requirement. Expression signals were then $\log_2$ transformed for computational and statistical analysis. A custom definition file from Brain Array was used to annotate and map the probes to genes. This process was carried out separately for individual batches. Ideally, all samples would have been run in one batch to avoid technical variation. Although a single run is possible for experiments from model systems, one run can be infeasible for human biopsy samples given the infrequency of biopsies. Our samples were therefore processed in several batches, which carries the risk of a potential batch effect between runs. This effect, which could be due to variations in reagents, microarray chips, other equipment, or the technician handling the processing procedure, may have systematic impacts on signals. It is thus crucial to account for this technical artifact. We employed batch correction based on the empirical Bayes framework described by Johnson et al.[68] on normalized $\log_2$-transformed data as the authors recommended. Principal component

analysis and hierarchical clustering were applied before and after batch correction to evaluate the effects of a batch and any other technical or biological separation on samples. Gene expression data from healthy living donors and DKD samples from the European Kidney cDNA Bank-Kroener Fresenius Biopsy Bank were processed accordingly [67] and are available through GEO [GSE47184]. Owing to ethical considerations, privacy protection, and to avoid identifying individual participants in this vulnerable population, the Ethics Board of the National Institute of Diabetes and Digestive and Kidney Diseases has stipulated that individual-level gene expression and genotype data from this study cannot be made publicly available [65].

4.4 Gene co-expression modules

An important research question in systems biology pertains to how genes/proteins/metabolites interact. A detailed understanding of how most genes function in a normal or disease state remains lacking particularly when dealing with high-throughput technologies including single-cell profiling. Single genes are no longer thought to drive physiological changes; rather, the concerted, orchestrated action of a set of genes is assumed to lead to alterations. Several systematic approaches can be employed to identify responsible gene sets, from clustering methods to more complex pattern recognition machine learning artificial intelligence algorithms. A common method is network construction: genes/proteins are depicted as nodes, and the connections or edges of these nodes are determined based on literature co-citation or data-driven associations. Another approach is co-expression analysis, grounded in the notion that genes which co-express or co-regulate in a system play similar functional roles. Yet another method, weighted gene co-expression network analysis (WGCNA) [69], is purely data-driven and establishes relations in a dataset to enable meaningful assessment.

Gene co-expression modules were constructed in R, based on the established statistical method of correlation that is widely accepted within the biological community. Though its basic function is to establish correlations among all genes, WGCNA enhances this analysis by implementing a "weighted" correlation and creating an adjacency matrix where coefficients are raised to a "power." This power, or β, describes disparity in the strength of correlations. A fully connected set of networks is then generated from the adjacency matrix. Given the goal of building networks that resemble the biological network, a scale-free network is adopted in this method. Nearly all real-life networks are scale-free; that is, certain nodes or hubs exhibit connections or edges within the network, whereas others have fewer connections following power law. These connections are contextualized by studying the relationships between neighbors, from which a network can grow based on functional associations. The method uses a sequential approach (Figure 7) by first calculating pairwise correlations among all gene pairs in the expression matrix to estimate the similarity between expression profiles. In other words, if there are *n* genes in a matrix, then each gene will have *n*-1 correlation values. These correlations raised to a power (β) are then computed between all gene pairs, effectively amplifying the difference between strong and weak correlations. The power is chosen based on scale-free network topology. The next step is to construct modules from a set of interconnected genes. The topological overlap matrix measures the pairwise similarity between network genes, which is then transformed into a dissimilarity measure. Subsequent hierarchical clustering of genes in the topological overlap matrix assigns them to branches in the cluster dendrogram. Gene pairs that share a high degree of topological overlap are assigned to a module. Then, a hypothetical module eigen gene is constructed; the module eigen gene is a summarized vector representing the expression profiles within each module as the first principal component of the module. Module eigen genes can be further correlated with each other to be merged into similar modules. These eigen genes are a main derivative of this method

and can be applied in many downstream steps as surrogates for genes within a given module. Module eigen genes can also be correlated with external traits, phenotypes, demographic features, histological scores, or any aggregate patient measure. The genes that constitute eigen genes can be used for granular analysis as well. Functional enrichment analysis, performed in a module-specific manner, can then reveal key underlying functions representing either module-specific or shared mechanisms.

We applied this method on tubulointerstitial and glomerular gene expression data independently and generated modules via the above steps. Finally, we correlated the module eigen genes to phenotypes and prioritized modules that were significantly associated with our variables of interest (p<0.05). Module transcripts were then used for downstream functional analysis.

Figure 7: Schema of sequential approach

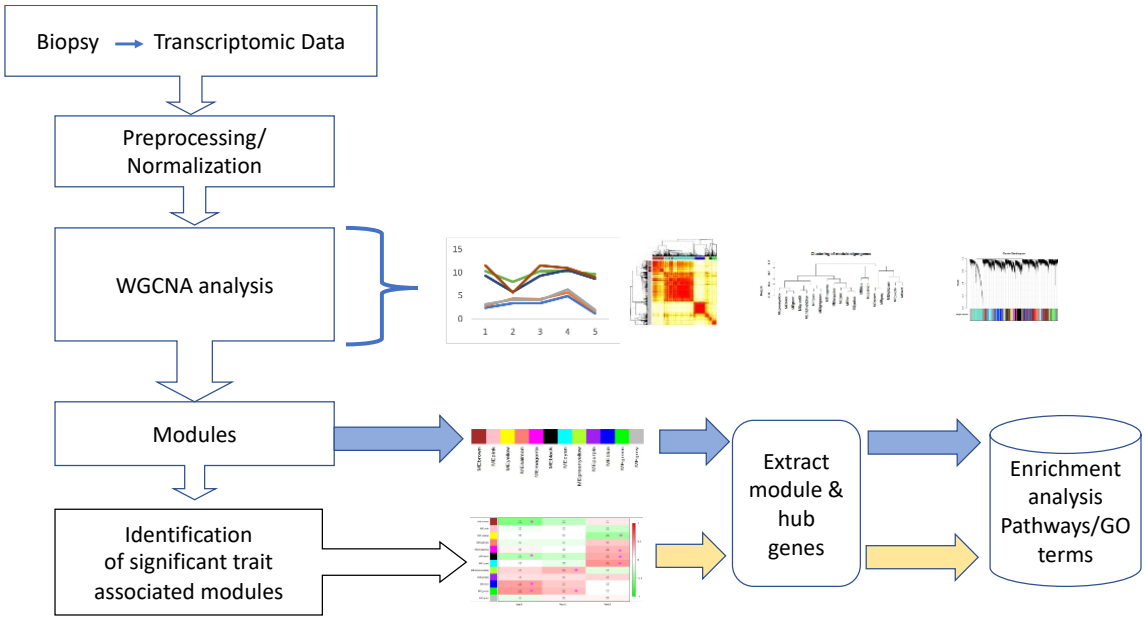


Figure 7 Legend: A schematic flowchart of the analysis illustrating the process from biopsy, data preprocessing, and analysis. WGCNA: weighted gene co-expression network analysis. WGCNA comprises several methods: co-expression patterns, module detection, module clustering based on eigen genes, and reiterative merging of the modules.

4.5 Functional gene annotation, enrichment, and pathway analysis

Several tools are available to perform functional enrichment analysis (e.g., gene set enrichment analysis from the Broad Institute, DAVID, GeneGO, and gProfiler). In this study, the Ingenuity Pathway Analysis software tool (Ingenuity Systems, Redwood City, CA) (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity- pathway-analysis) was employed to define enriched functional terms within module genes with respect to pathways. A pathway network was constructed based on the criterion that pathways should have more than two genes in common. Functional biological processes and molecular functions were inferred using the controlled vocabulary of Gene Ontology in Genomatix software (Genomatix, Munich, Germany). GePS (https://www.genomatix.de/online_help.html), Genomatix Pathway System was used to construct a network from module genes. Network nodes denoted the genes identified from modules, and network edges denoted the connections linking genes based on co-citation in the literature. The degree of co-citation can be adjusted based on user preference. In this case, the criteria for the transcriptional network in GePS indicated that genes should have functional terms (e.g., inhibition/activation) from more than two literature sources.

4.6 Upstream regulator analysis approach and gene selection

Upstream regulator network analysis effectively identifies the cascade of upstream regulators and intermediate regulators that explain downstream observed transcriptional changes. As the name implies it is a causal analysis utilizing a reverse engineering approach. This method in summary tries to learn why the genes are regulated in a certain manner in the given biological disease or condition. Upstream regulator network analysis on the IPA platform was used to construct this cascading network. IPA software queries its enriched, curated backend database to identify involved regulators and predict whether an involved upstream regulator is likely to

be activated or inhibited. Genes shared among two or more pathways were extracted from the VvInt pathway network. This gene set was then screened against the database for potential upstream regulators affecting downstream expression changes via direct or indirect connections.

4.7 Tissue processing for single-cell analysis

Similar to the bulk mRNA cohort, participating members of the American Indians from Arizona, US underwent a protocol kidney biopsy for the single cell analysis. The research core was used for quantitative morphometry and tissue expression studies. CryoStor® (Stemcell Technologies) preserved kidney tissue samples were collected from 44 American Indian participants enrolled between 2016 and 2017. Single-cell transcriptome was generated from 2–3 mg of the biopsy samples. Tissue procurement and processing protocols were adapted from the collaboration with Accelerating Medicines Partnership's lupus network, Pathway Exploration and Analysis in Kidney Disease [70]. The protocol was modified to enrich from the epithelial cell type. Kidney biospecimens in 1ml Hypothermasol™ (Stemcell Technologies) were transferred to the lab and immediately cryopreserved in DMSO containing CryoStor® CS 10 (Stemcell Technologies) solution for long-term storage in liquid nitrogen.

4.7.1 Single-cell isolation

The following protocols were applied for tissue dissociation as published previously [71, 72]. The process starts with the thawing of samples in 37˚C water bath for 1 minute. The thawed samples are then placed in a 1 mL DMEM/F12 medium (with L-Glutamine, HEPES, and high glucose) supplemented with 10% heat-inactivated FBS. After incubation at room temperature for 10 minutes, the tissue samples of 1-mm^3 pieces are cut and placed in a 1.5-mL tube containing

500 µL digestion media with LiberaseTM TL (research grade, Roche) and incubated on an orbital shaker (500 rpm) for 12 minutes at 37˚C with a mild titration step after 6 minutes using a wide-bore 1000-µL tip. At the end of 12 minutes 500µL DMEM/F12 medium containing 10% FBS was added to halt the digestion. And after a one minute of incubation at room temperature the cell suspension was then filtered through a 30-µm strainer (Milteny Biotech) into a 15-mL conical tube. The left-over solid tissue was passed through nylon mesh using the rubber end of a 3-mL syringe pestle. The filter was then washed with 10mL DMEM/F12/10% FBS medium, and cells were pelleted by centrifugation for 10 minutes at 200g at 4˚C. The supernatant was immediately removed via vacuum suction using a glass Pasteur pipet. The cell pellet was then gently re-suspended in 55 µl medium. Cell viability was analyzed with a Countess II FL automated cell counter using the Trypan Blue dye exclusion method.

4.8 Next-generation sequencing of single-cell mRNA

The single cell samples prepared using the above the protocol are transferred to the University of Michigan Advanced Genomics Core facility for sequencing on droplet-based high-throughput 10XGenomics™ Chromium™ platform.10,000 viable cells in up to 46 µl DMEM/F12/10%FBS per sample were added to each channel of a chip. The 10X platform technology allows cell lysis, individual cell barcoding, and reverse RNA transcription Specifically, the 10X Single-cell 3' GEX – version 3.1 kit was used for the processing. Library preparation at the core included emulsion breakage, PCR amplification, cDNA fragmentation, oligo adapter, and Illumina sample index addition. Libraries were pooled and sequenced on an Illumina NovaSeq6000 platform as asymmetric paired-end runs (28 x115 bases) with a median of 200 million raw sequencing reads per sample[71, 72] Sequencer output was processed in the proprietary 10X Chromium single-cell gene expression analysis software CellRanger

(https://support.10xgenomics.com/single-cell-geneexpression/software/pipelines/latest/what-is-cell-ranger).

4.9 Analysis of single-cell data

The Core facility initially processed sequence data using CellRanger. The output data files were further processed using Seurat-version 3, implemented as an R package (https://cran.r-project.org/web/packages/Seurat/index.html). Cells containing 500–5000 genes per cell were only included after initial filtering to account for cell doublet and cell viability issues. Doublet formation (i.e., two adjacent cell types captured by a single droplet) is a common concern in single-cell technology; doublets are identified as cells with a larger mRNA content and genes than a single cell. We applied the cutoff of >500 and <5000 genes per cell to account for this technical artifact. Another quality measure of cell viability in single-cell technology is the percentage of mitochondrial gene read content. A high percentage of mitochondrial gene reads suggests low cell viability, such as when cells' cytoplasmic mRNA is lost during dissociation due to plasma membrane breakage. A threshold of <50% mitochondrial reads per cell was used in this study. 10XGenomics version 3 chemistry has been found to return a higher percentage of mitochondrial reads irrespective of cell viability in certain tissue types, including kidney tissue. Deep sequencing of samples (200 million reads) to obtain more genes per cell can mitigate this issue. Downstream analysis followed the standard single-cell processing of normalization, scaling, dimensionality reduction (i.e., principal component analysis), and uniform manifold approximation and projection [73]. Sample integration was performed using the Harmony algorithm to account for batch effects and to identify differentially expressed cell type–specific markers [74]. We used the functionality embedded in the Seurat R package to normalize and scale read counts; this functionality uses a global scaling method that normalizes the gene expression measurement for each cell by the total expression, multiplies this value by

10,000 (default), and log-transforms scaled expression values [74-76]. We chose the resolution parameter of 6 for unsupervised clustering. Publicly available resources including published literature, Kidney Interactive Transcriptome (http://humphreyslab.com/SingleCell), Human Protein Atlas (https://www.proteinatlas.org), the Epithelial Systems Biology Laboratory (https://hpcwebapps.cit.nih.gov/ESBL/Database/) and Immgen (https://www.immgen.org/) were used to annotate clusters with cell-specific markers [71].

4.10 Statistical analysis

Affymetrix microarray processing, normalization, and WGCNA were carried out in R statistical platform using Affy package and WGCNA package. Single-cell processing and visualizations were performed using the Seurat package in R. Significance analysis of microarrays was conducted in the MultiExperiment Viewer within the Institute for Genomic Research suite to identify differentially regulated gene sets in advanced DKD compared to controls. The significance threshold of the false discovery rate was set at ≤ 0.05.

5. Results

The central goal of this thesis was to generate bioinformatics-based tools that could be used to identify molecular and structural fingerprints linked to disease progression in a DKD population at an early disease stage, showcasing a systematic bioinformatic approach to dataset integration. To this end, we employed a data reduction method and a weighted gene correlation network method that summarized high-dimensional data while retaining the flexibility to associate with the phenotype of interest. We focused on a type 2 diabetic cohort whose participants underwent protocol kidney biopsies. Participants exhibited clinically early kidney

disease, with most falling under a CKD stage of G1 or A1 per KDIGO guidelines (see Figure 3). Baseline characteristics of the study cohorts are described in the following sections.

5.1 Transcriptomic profiling

All kidney biopsy tissue was manually microdissected to extract glomerular and tubular tissue for RNA isolation. Samples were then sent to the sequencing core for transcriptional profiling using an Affymetrix array platform. High-throughput profiling enabled transcriptional changes to be captured on a genome-wide scale. We could therefore identify key regulatory changes and establish network-level crosstalk among molecular changes. The raw data after rigorous preprocessing (as explained in the Methods section) were quantile-normalized and log2 transformed to satisfy a Gaussian distribution. The systematic bioinformatic approach WGCNA, as explained earlier, was applied to this dataset to capture coregulated gene groups (i.e., those implying orchestrated biological functions).

Using microarray preprocessing, gene expression profiling of tubulointerstitial and glomerular tissue was generated from the protocol kidney biopsy specimens from Pima Indian samples, indication biopsies from Caucasian DN patients, and healthy living donors. A log2-transformed robust multichip array normalized gene expression matrix was constructed for three datasets (i.e., glomerular and tubular profiles from early DKD and the tubular expression matrix from advanced DKD) after preprocessing the arrays that passed quality control. A background noise signal was calculated from Affymetrix positive control intensities. These controls were included in the chip as spike-ins, which were used to maintain constant expression values. We applied the median ± 2 standard deviations of control signals and filtered the expression matrix to include only genes that were expressed above the noise threshold.

5.2 Tubulointerstitial subgroup

5.2.1 Clinical characteristics

Table 3 summarizes the baseline clinical, demographic, and morphometric characteristics of this cohort with tubulointerstitial expression profiling. Most participants had an iothalamate measured glomerular filtration rate (mGFR) of >90 mL/min with an average of 147 mL/min. The cohort's overall median ACR was 35 mg/g at the time of biopsy, with roughly half of participants in the normal albuminuric range (<30 mg/g) and the other half in the micro–macroalbuminuric range (>30 mg/g). Interstitial volume (measured as % kidney cortex) was significantly higher (29.5% vs. 11.9% in non-diabetic living kidney donor biopsies; $p<0.0001$).

Table 3: Baseline characteristics of tubulointerstitial cohort

M/F (% male)	15/34 (30.6)
Age (years)	46 ± 9.8
Duration of diabetes (years)	15.7 ± 6.8
BMI (kg/m2)	35.2 ± 8.2
HbA1c (%)	9.20 ± 2
Systolic blood pressure (mm/Hg)	124.73 ± 13.99
Diastolic blood pressure (mm/Hg)	78.02 ± 8.09
ACR (mg/g)	35.46 [90.21]
mGFR (mL/min)	147 ± 45
VvInt (%)	29.5±9.6
Follow-up post-biopsy (years)	10.1[2.0]

Data are presented as mean (±SD), median [Interquartile range], or proportions (%). M/F: male/female; BMI: body mass index; HbA1c: glycated hemoglobin A1c; ACR: urinary albumin-to-creatinine ratio; mGFR: iothalamate glomerular filtration rate; VvInt: cortical interstitial fractional volume.

5.2.2 Module detection

We used WGCNA to define co-expressed gene sets. As explained in the Methods section, this protocol first groups highly correlated gene pairs into a module. The second step iterates to integrate the modules whose eigen genes are strongly correlated. The complexity of our high-

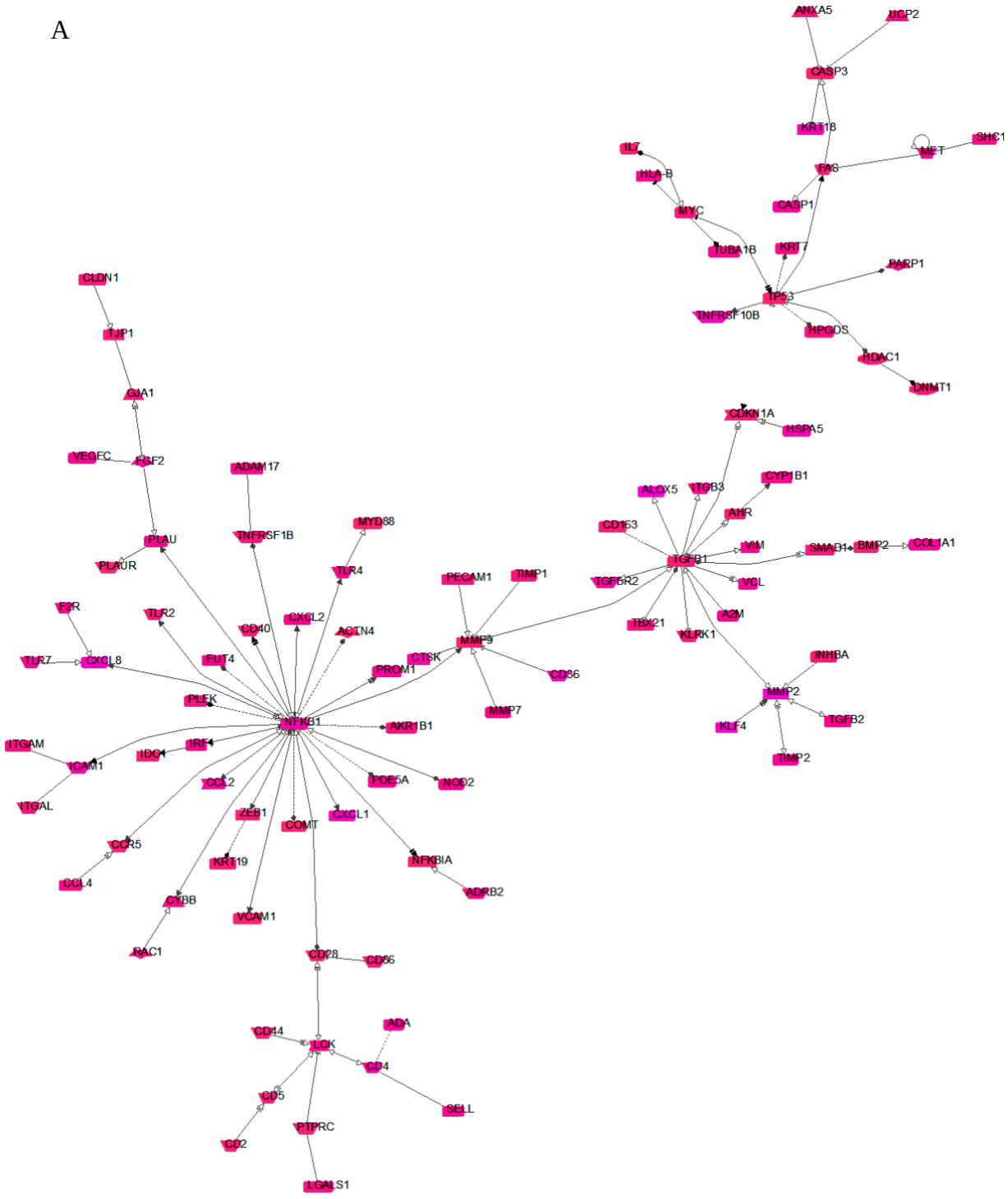
dimensional dataset was substantially reduced using this iterative approach, resulting in 11 transcriptional modules ranging from 129 to 2,378 transcripts from tubulointerstitial profiles. The eigen genes of these 11 modules were then correlated with all cross-sectional variables of interest (e.g., VvInt, GFR, ACR) as well as with GFR and ACR over the follow-up period. The reduction from ~12,000 features to 11 features minimized or eliminated the dependency on multiple correction. Four eigen genes had statistically significant associations ($p \leq 0.05$) with VvInt; these eigen genes belonged to the black, blue, brown, and green modules, the colors do not imply any significance rather for naming purpose only. Negative correlation with VvInt was captured by the eigen genes from the black and brown modules (r=-0.31 and r= -0.51, respectively) whereas a positive relationship was established by the blue and green module eigen genes (r=0.38 and r=0.43, respectively) [65].

5.2.3 Functional context of VvInt-associated transcripts

To further understand these modules' involvement in the pathophysiology of DKD, the transcripts within each module were extracted and screened for their known biological enrichments. A total of 1,843 genes were significantly correlated with VvInt (q-value≥0.05 and |r| =<=0.25) from all the four modules. This "fibrosis" signatures were then evaluated for literature-derived prior interactions using a co-citation network approach based on natural language processing in GePS. This system uses information from several public databases to compile more than 400 human pathways. The software enables to generate the co-citation networks using different criteria for establishing connection. Here the signatures were scanned for co-citations in PubMed-indexed publications by a functional term (e.g., "A induces B"). An additional filter was employed to ensure that these interactions had more than two citations on a functional level. For visual clarity, separate transcriptional networks of the top 100 most-connected transcripts were generated for genes positively and negatively correlated with VvInt

(Figure 8). Several inflammatory signaling markers (e.g., CCL2, ICAM1), proliferation mediators (e.g., TP53), and growth factor–related signaling mechanisms (e.g., EGF, VEGF) appeared on the central hub of these networks, suggesting the presence of these mechanisms in early DKD. Gene ontology enrichment demonstrated similar traits enriched with immune regulatory functions, integrin binding, cell activation, and focal adhesion with the positively correlated transcripts. Transcripts negatively correlated with VvInt were enriched for molecular functions such as oxidoreductase activity and NAD binding in addition to being highly enriched with metabolic biologic processes [65].

Figure 8: Transcriptional network of genes associated with fibrosis score (VvInt)



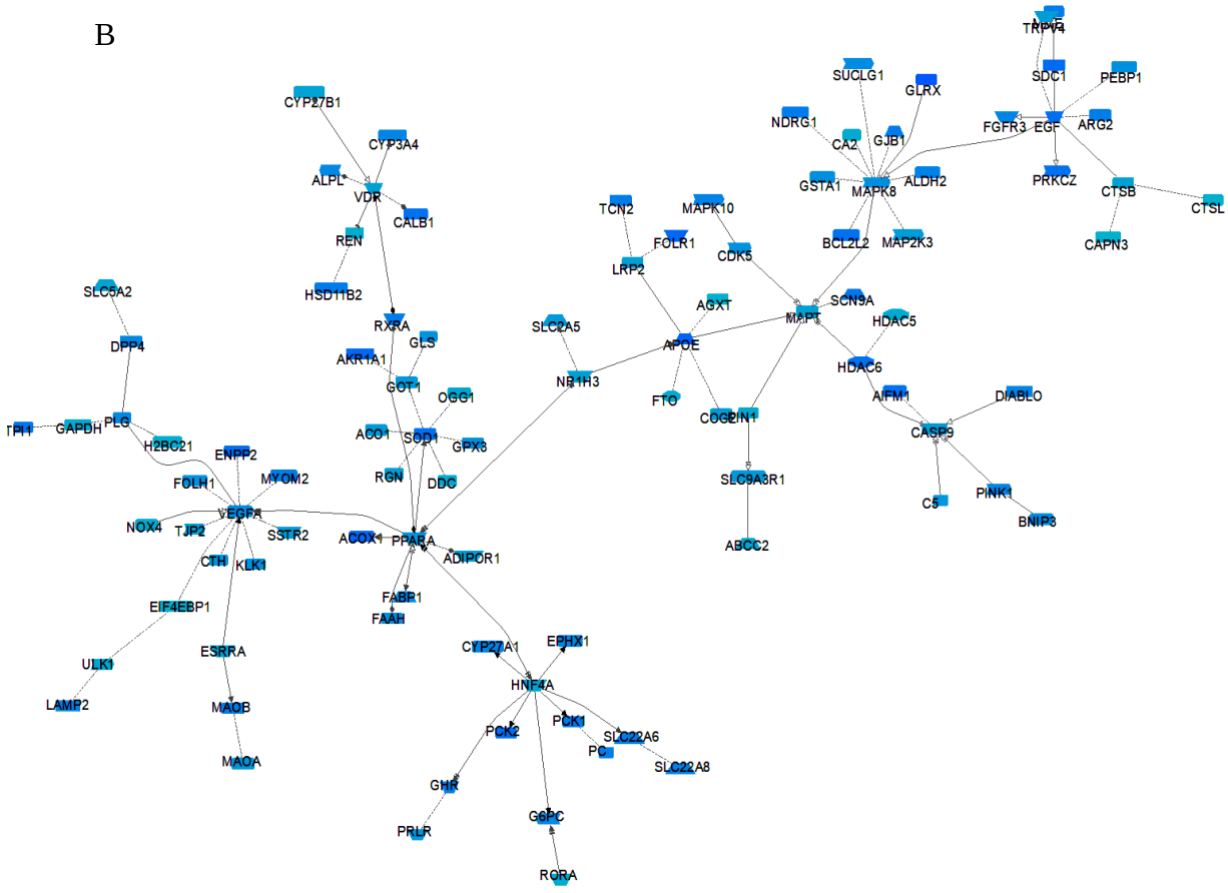


Figure 8 Legend: Transcriptional networks A (top) and B (bottom). Genes from positively and negatively correlated VvInt modules were input into the Genomatix Genome Analyzer to establish relationships among these genes. For visual clarity, the top 100 genes were selected (separately from positively and negatively correlated genes) for network generation. Edges or connections define a functional relationship between genes based on experimentally curated information in the GePS database with more than two literature citations. Connections denote a functional keyword relationship, such as the appearance of "induced" or "activated" between genes in the same sentence.

5.2.4 Pathways associated with interstitial damage

The previous section outlined relationships between previously studied genes among fibrosis-associated gene sets. Functional enrichment analysis can further uncover the mechanisms dysregulated in molecular fingerprints due to associated damage (e.g., using classic gene ontology classification analysis or pathway-level investigations). In this section, we discuss pathways challenged by interstitial damage. Pathway analysis conducted in IPA software revealed 53 significantly enriched pathways using the 1,843 VvInt-related transcripts as input. Several pathways were identified (e.g., inflammatory or cell–cell/cell–matrix interaction pathways), with PXR/RXR activation, FXR/RXR pathways, and metabolic pathways among those most enriched [65].

A network of disease-associated pathways was constructed from the 53 enriched pathways (Figure 9). Pathways with shared genes were grouped into clusters, resulting in a bipartite structure. Clear clusters of metabolic and inflammatory pathways suggested that these changes manifested in the early DKD stage and were associated with tubular damage. One metabolic cluster comprised fatty acid B oxidation, oxidative phosphorylation, and TCA cycle pathways. An inflammatory cluster composed of signaling mechanisms, including the extracellular matrix and growth factor signaling (e.g., inhibition of angiogenesis by TSP1), inflammasome pathway, and IL-8 signaling pathway, demonstrated early-stage interactions prior to clinical disease presentation. Mitochondrial dysfunction and LPS/IL-1–mediated inhibition of RXR pathways seems to be at the crossroads of this bipartite pathway network. Most pathways within each subcluster were densely interconnected (Figure 9). A small set of genes including *PRKCQ*, *NFKB1*, *MAPK8*, *ALDH2* and *RAC1* had high connectivity among VvInt-associated pathways. This pattern suggests key players' central roles in coordinating changes of potential interest [65].

Figure 9: Network of fibrosis-associated pathways

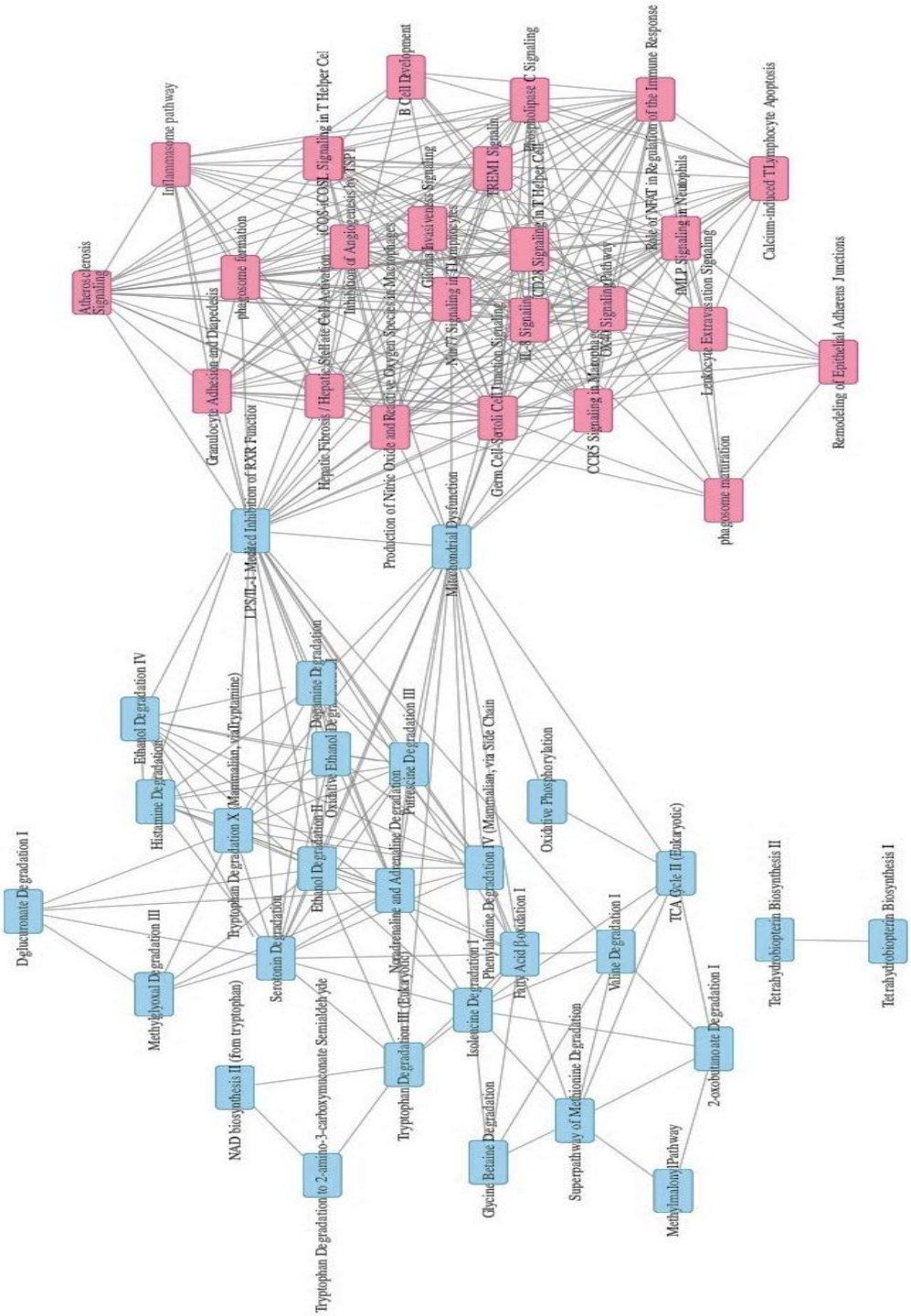


Figure 9 Legend: A network of significantly enriched canonical pathways, generated via Ingenuity Pathway Analysis. Enrichment is based on 1,843 genes from four modules associated with VvInt. The threshold for pathway significance was set at $p \leq 0.05$. The network was restricted to pathways connected by ≥ 3 genes to display connectivity or crosstalk among enriched pathways. A bowtie structure indicates enrichment of metabolic signatures predominant in the left cluster (cyan nodes) and inflammatory pathways in the right cluster (pink nodes). Edges represent more than three genes shared by the nodes.

5.2.5 Identification of upstream regulators

A dense interconnected network points towards activation of a potential causal upstream mechanism. Upstream regulators influence the expression patterns observed in downstream genes. A set of 229 genes was enriched in this interconnected pathway network. Using a causal network inference approach [77], *IL1B* along with 48 intermediate regulators was identified as the master regulator of inflammation, affecting the expression of more than 50% of shared genes (Figure 10) [65]. *IL1B* has been well studied in DKD model systems of advanced tubular dysfunction; it is essential to the interplay between tubular cell damage, linking apoptosis and innate immune activation. Activation of the downstream targets *MMP2* and *MMP9* facilitates the recruitment of inflammatory cells and fibroblasts, prompting fibrosis. Five transcriptional regulators (*PHF1*, *SOX2*, *NFAT5*, *TRIM29*, *HEY1*), connected with a common intermediate transcriptional regulator *TP53*, modulate the downstream targets of differentiated tubular function and oxidative stress [65].

Figure 10: Upstream regulator of complex fibrosis network

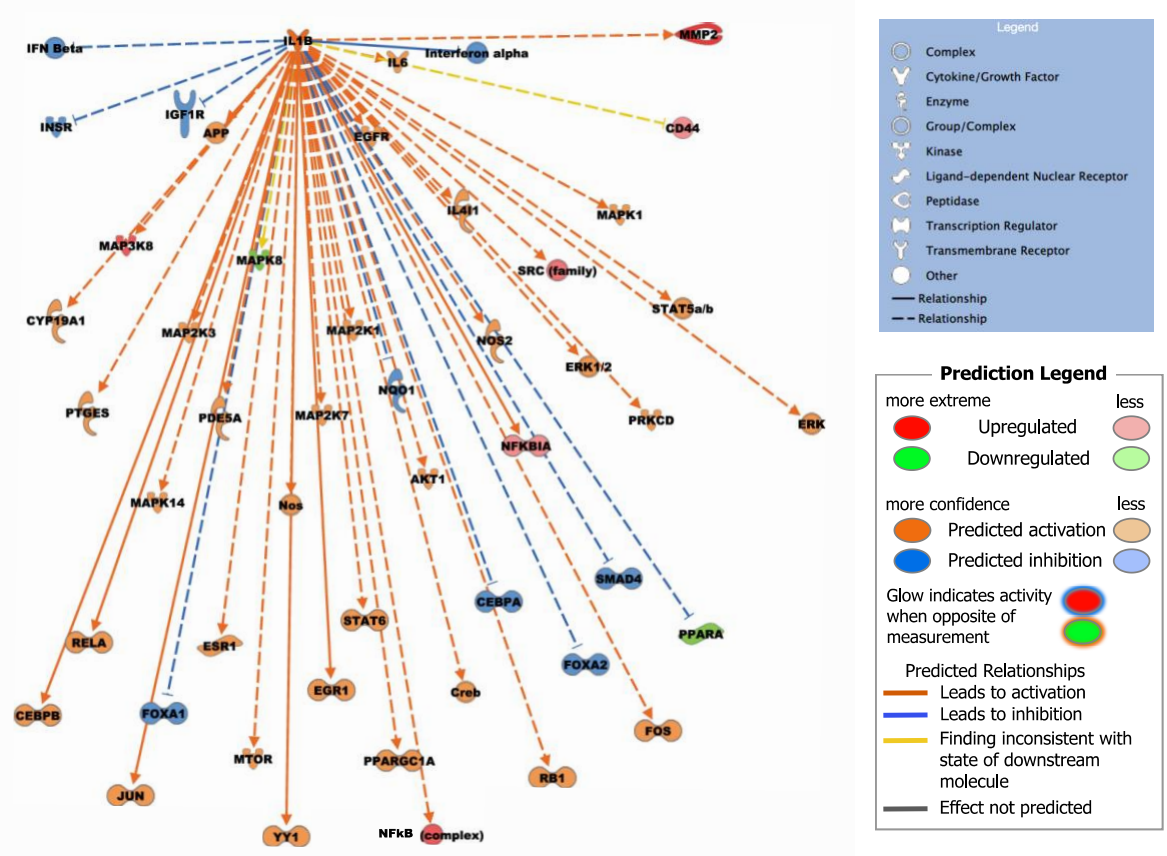


Figure 10 Legend: *IL1B* is the top upstream regulator of genes associated with VvInt and shared among more than one pathway. Based on upstream regulator analysis generated via Ingenuity Pathway Analysis, *IL1B* was predicted to be the top regulator affecting expression of the 229 genes from the four modules through 48 intermediate regulators.

5.2.6 Fibrosis and kidney function decline

The identified modules were correlated to structural damage at the time of biopsy. With regard to clinical utility, it is worthwhile to explore how these VvInt modules contribute to disease progression. The focal cohort is unique in its availability of long-term follow-up on clinical phenotypes with annual GFR and ACR measurements. The median [IQR] observation period was 15.9 [2.5] years, with 10.1 [2.0] years median follow-up length post-biopsy (Figure 11). No significant association was observed between the four modules in terms of cross-sectional

GFR and ACR. However, a significant association emerged between the module eigen genes and kidney function over time, respectively, over a median 10-year follow-up after biopsy ($p \leq 0.05$). This pattern indicates that clinical presentation might be misleading or unable to capture early-stage DKD; only over a longer diabetic duration (i.e., 20+ years) does this association start to become stronger in terms of classical clinical phenotypes [65].

Figure 11: Association of VvInt modules with GFR and ACR measurement over time

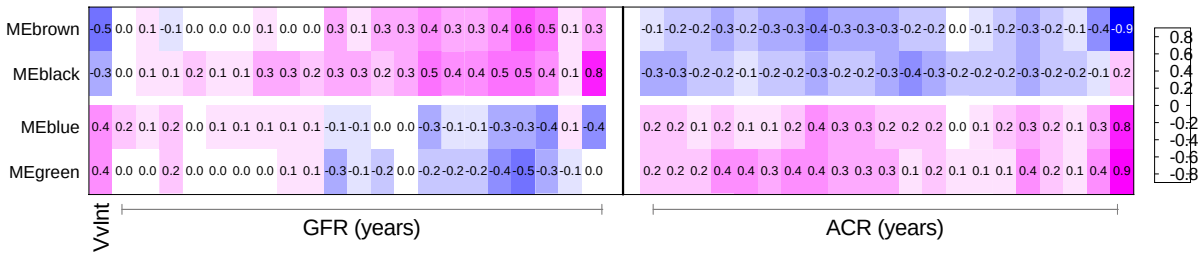


Figure 11 Legend: Eigen genes of VvInt-associated modules correlated with GFR and ACR measured over time. The first column is VvInt measured at the time of biopsy. Subsequent columns represent correlations of eigen genes with follow-up GFR and ACR. The color legend follows the legend bar on the right. Blue shades denote negative correlations, and purple shades denote positive correlations. The depth of the color indicates the correlation strength.

5.2.7 Findings from early DKD to advanced DKD

Fibrosis-associated genes were captured from a preclinical DKD cohort of American Indian ethnicity. To examine the impact of these early disease signals in advanced DKD, clinically indicated biopsies (N=17) were obtained from patients with advanced DKD and compared to healthy living donors to identify differentially regulated gene signatures. Fibrosis signatures from early DKD overlapped with differentially regulated genes in advanced DKD. Baseline characteristics of these 17 DN patients from a European (i.e., ERCB) DKD cohort are listed in Table 4. The DN patients in this cohort were mostly in CKD Stage 3–5 with an average estimated GFR of 44.3 mL/min/1.73m^2. In total, 1,302 of the 1,843 fibrosis signatures (71%)

were significantly differentially regulated in late-stage DKD compared to healthy living kidney donors (*N*=31); 98% had concordant changes with disease. Figure 12 displays preserved nodes from the VvInt-transcriptional network in advanced DKD [65].

Table 4: Baseline characteristics of DN samples in ERCB cohort (*N*=17)

Variables	
Age (years)	58.3 ± 10.7
Gender (% female)	29%
eGFR (mL/min per 1.73 m2)	44.3 ± 24.9

Table 4 Legend: Age and eGFR are presented as mean ± SD. eGFR is calculated using the MDRD equation.

Figure 12: Transcriptional network of signatures in advanced DKD

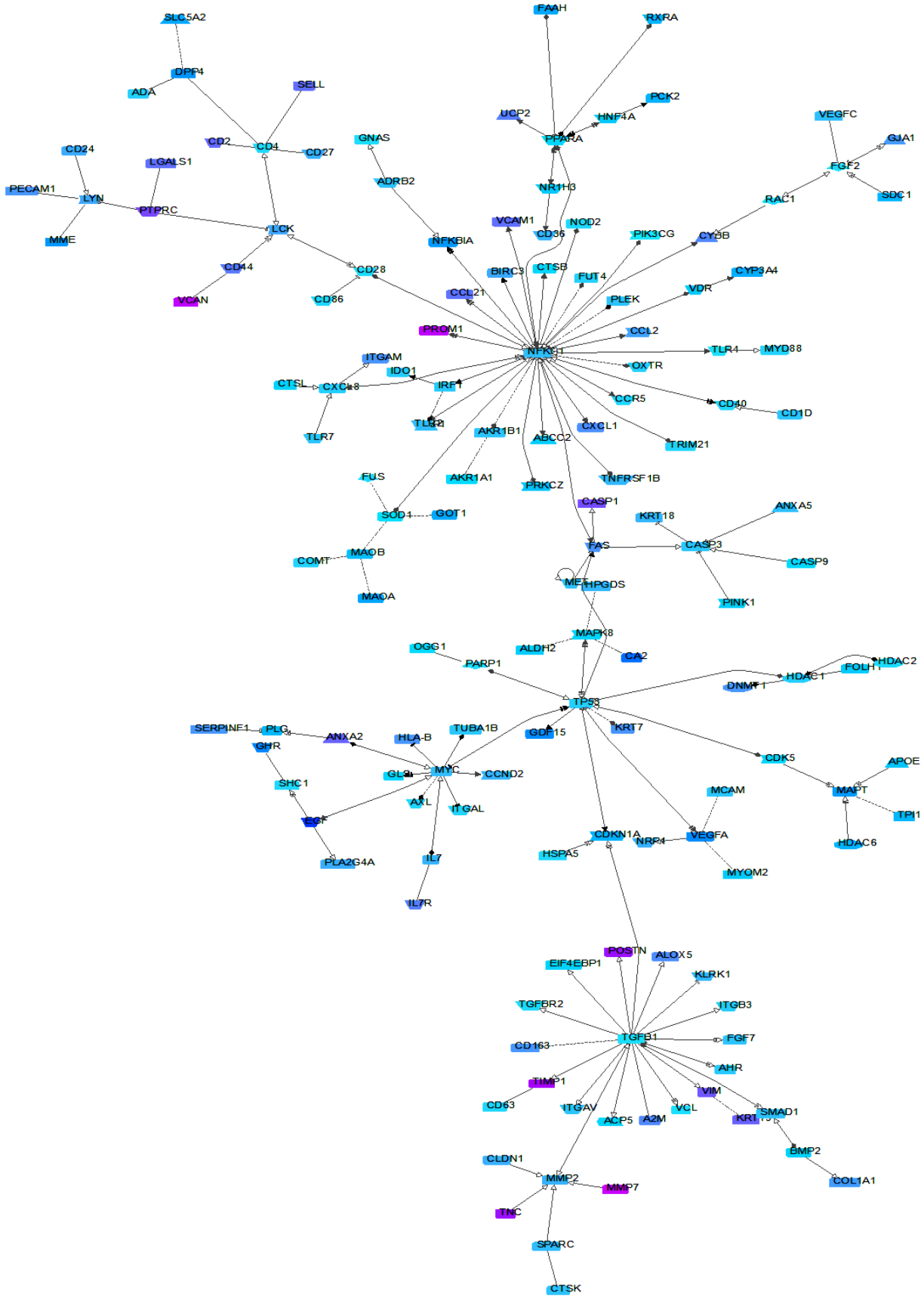


Figure 12 Legend: Transcriptional network active in the tubulointerstitial compartment of advanced diabetic nephropathy (DN) patients in the ERCB cohort. Overall, 71% of VvInt-associated genes were differentially regulated in DN patients compared to controls, indicating that genes which appeared associated with early lesions of damage remained active in later DN stages. The network presented here illustrates connections among the top 100 highly connected nodes. Purple denotes increased gene expression and blue denotes decreased gene expression compared to controls.

5.3 Glomerular subgroup

5.3.1 Clinical characteristics

While several studies have highlighted the importance of tubular involvement in DKD, the conventional mechanisms of moderate albuminuria (microalbuminuria) and hyperfiltration associated with glomerular structural and functional changes in early DKD represent key criteria of disease progression [78]. To evaluate potential glomerular changes occurring early in DKD, transcriptional associations of the glomerular expression profile were compared to observed structural changes. We used the same bioinformatic analytical pipeline to test the general validation or potential transferability of this method to other tissues or data elements. Baseline characteristics of participants in the same cohort from whom glomerular samples were obtained are described below. The weighted correlation approach (WGCNA) was used to identify modules and eigen modules associated with glomerular lesions, with output data mapped onto pathways.

Characteristics of the cohort with glomerular expression profiling are presented in Tables 5A and 5B. The 69 participants' median urine ACR was 26 mg/g with an average measured iothalamate glomerular filtration rate of 145 mL/min. Similar to the tubular cohort, more than half of these participants had normal ACR (<30 mg/g).

Table 5A: Characteristics of cohort used in glomerular analysis

Characteristic	At time of biopsy (*N*=69)
Male sex, *n* (%)	18 (30)
Age, years	45 ± 10
Diabetes duration, years	15.5 ± 6.4
HbA1c (%)	9.4 ± 2
Systolic blood pressure (mm/Hg)	122.4 ± 12.81
Diastolic blood pressure (mm/Hg)	77.79 ± 7.14
ACR (mg/g)	25.93[141.57]
mGFR (mL/min)	145 ± 52
Serum creatinine (mg/dl)	0.7 ± 0.2
Post-biopsy follow-up length, years	10.1[2.0]

Data presented as mean (±SD) or median [IQR] or proportions (%). HbA1c: glycated hemoglobin A1c; ACR: urinary albumin-to-creatinine ratio; mGFR: iothalamate glomerular filtration rate. Clinical characteristics are comparable to samples used in the tubulointerstitum analysis.

Table 5B: Summary of structural parameters from kidney biopsies of 69 participants

Characteristic	At time of biopsy
GBM Width (nm)	493.6 (107.30)
VVMES (%)	0.28 (0.08)
SV ($\mu m^2/\mu m^3$)	0.09 (0.02)
VPCN (μm^3)	150.4 (52.61)
VVPCN (%)	0.11(0.04)
VPC (μm^3)	1651(1272)
VVPC (%)	0.15 (0.05)
NVPC (N/glom)	0.0002 (0.0001)
GV ($x10^6\ \mu m^3$)	2.01(0.52)
FPW (nm)	901.3 (531.10)
FEN (%)	47.57 (18.81)

Data are presented as mean (±SD). GBM: glomerular basement membrane width; VVMES: mesangial fractional volume; SV: surface volume of peripheral GBM per glomerulus; VPCN: volume of podocyte nuclei; VVPCN: volume fraction of podocyte nuclei per podocyte cell;

VPC: volume of podocyte cell; VVPC: volume fraction of podocyte cell per glomerulus; NVPC: numerical density of podocyte cell per glomerulus; GV: glomerular volume; FPW: foot process width in peripheral GBM; FEN: percent of endothelial fenestration.

5.3.2 Module detection

The WGCNA method detailed earlier was employed to identify modules or gene sets in the 69 participants' glomerular expression profiles. Data generation and preprocessing followed the same protocol as described in the tubulointerstitial section. Fourteen functional modules of varying size (106–2,123 transcripts) were identified in the glomerular expression dataset. The eigen genes from all 14 modules were then correlated with clinical and glomerular morphometric traits measured cross-sectionally (Figure 13). The threshold for statistically significant associations was set at $p \leq 0.05$. The green-yellow module showed strong positive correlations with glomerular basement membrane width and mesangial volume and negative correlations with intact foot process width and podocyte measurements, including podocyte volume and the numerical density of podocyte per glomerulus. Out of three module eigen genes, one was positively correlated and two were negatively correlated with fenestrated endothelium. These eigen genes belonged to brown, pink, and cyan modules, respectively. The colors of module eigen genes were assigned arbitrarily. Detailed results are shown in Table 6.

Figure 13: Heatmap showing association of eigen genes and structural parameters

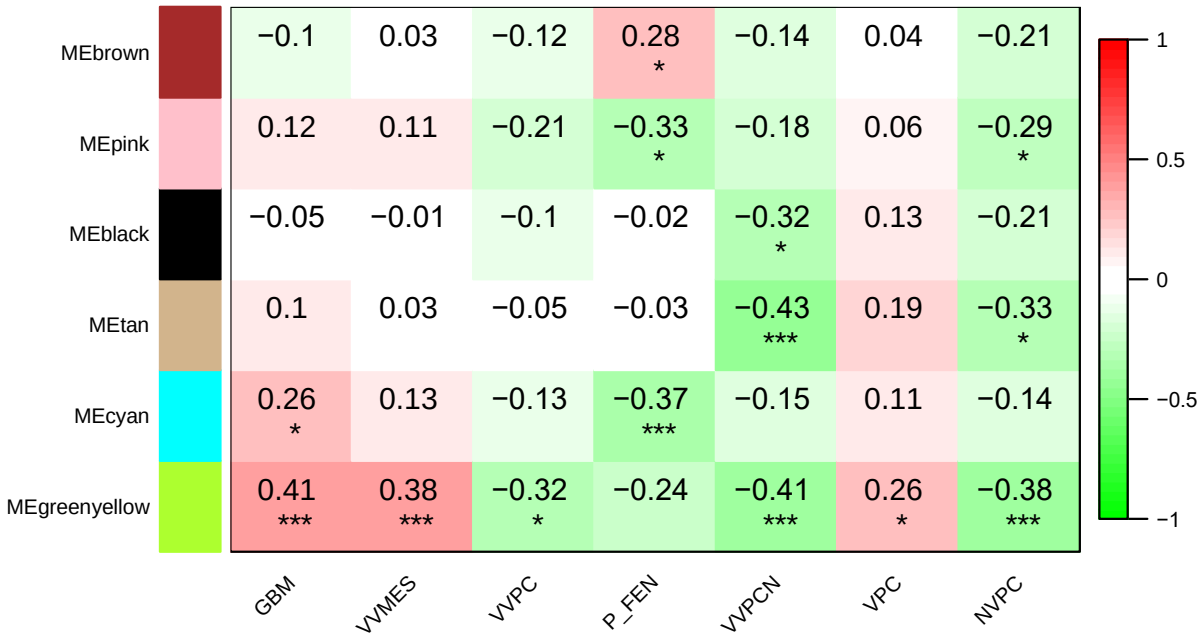


Figure 13 Legend: Each cell reports the correlation coefficient from correlating module eigen genes (rows) to traits (columns). Only modules significantly correlated with any trait are shown in the heatmap. GBM: glomerular basement membrane width; VVMES: mesangial fractional volume; VVPCN: volume fraction of podocyte nuclei per podocyte cell; VPC: volume of podocyte cell; VVPC: volume fraction of podocyte cell per glomerulus; NVPC: numerical density of podocyte cell per glomerulus; FEN: percent of endothelial fenestration. The table is color-coded by correlation according to the legend (***$p \leq 0.001$, **$p \leq 0.01$, *$p \leq 0.05$). Significance was set at $p \leq 0.05$.

Table 6: Thirteen modules from glomerular expression matrix and eigen gene–trait correlations

Modules	GBM	VVMes	GV	SV	FPW	FEN	VVPC	VVPCN	VPC	NVPC
Black								-		
Blue										
Brown						+				
Cyan	+					-				
Green										
GreenYellow	+	+					-	-	+	-
Magenta										
Midnight blue										
Pink						-				-
Red										
Salmon										
Tan								-		-
Turquoise										

Table 6 Legend: + indicates positive correlation of eigen gene with trait; - indicates negative correlation of eigen gene with trait. GBM: glomerular basement membrane width; VVMES: mesangial fractional volume; GV: glomerular volume; SV: surface volume of peripheral GBM per glomerulus; FPW: foot process width in peripheral GBM; FEN: percent of fenestrated endothelium; VVPC: volume fraction of podocyte cell per glomerulus; VVPCN: volume fraction of podocyte nuclei per podocyte cell; VPC: volume of podocyte cell; NVPC: numerical density of podocyte cell per glomerulus.

5.3.3 Functional network of glomerular lesion-associated transcripts

Transcripts which showed strong correlations with each score ($p \leq 0.05$ and |r| =>=0.25) within each significant module were extracted and assessed for active functional fingerprints. One module (green-yellow) was significantly associated with nearly all traits. Because this module demonstrated strong correlations across traits, we evaluated the underlying theme. The affiliated genes were screened for known pathophysiology using canonical pathway analysis via IPA. Ample presence of inflammation, cell signaling, cell cycle, immune cell traffic, lipid metabolism, complement system, STAT3, and dendritic cell/natural killer cell communication was observed in these transcripts. Signaling mechanisms such as TREM-1, TNFR2, Tec-kinase, CXCR4, Chemokines, and CCR5 signaling in macrophages composed the hub of these networks, again implying that these programs were activated in early stages of DKD. mTOR, Rac signaling, and axonal guidance pathways were among the significantly enriched pathways in podocyte trait-associated genes. Tables 7A–7D present detailed results on the enriched pathways for each associated trait. Thus, the general approach developed using data from the tubulointerstitial compartment could also identify unique parameters present in glomeruli.

Table 7A: Pathways enriched in gene set associated with glomerular basement membrane

GBM signature–associated pathways	**-log (B-H *p*-value)**
Hepatic Fibrosis / Hepatic Stellate Cell Activation	4.07E+00
T Helper Cell Differentiation	4.07E+00
Atherosclerosis Signaling	4.07E+00
Complement System	3.98E+00
Leukocyte Extravasation Signaling	3.96E+00
Granulocyte Adhesion and Diapedesis	3.60E+00
Th1 Pathway	3.60E+00
Agranulocyte Adhesion and Diapedesis	3.34E+00
Th1 and Th2 Activation Pathway	3.02E+00
Natural Killer Cell Signaling	2.92E+00
Th2 Pathway	2.40E+00
Dendritic Cell Maturation	2.24E+00
Phagosome Formation	2.21E+00
TREM1 Signaling	2.09E+00
GP6 Signaling Pathway	1.50E+00
MSP-RON Signaling Pathway	1.40E+00
Autoimmune Thyroid Disease Signaling	1.35E+00
Crosstalk between Dendritic Cells and Natural Killer Cells	1.35E+00
FcŒ≥ Receptor-mediated Phagocytosis in Macrophages and Monocytes	1.31E+00
Lipid Antigen Presentation by CD1	1.31E+00
Allograft Rejection Signaling	1.31E+00
Graft-versus-Host Disease Signaling	1.31E+00

Table 7A Legend: 213 genes significantly correlated with the glomerular basement membrane width at the threshold of r≥0.25| and $p \leq 0.05$ were selected for IPA. Significant canonical pathways after the Benjamin-Hochberg procedure ($p \leq 0.05$) are included in the table.

Table 7B: Pathways enriched in gene set associated with mesangial volume measurement

Mesangial expansion signature–associated pathways	**-log(B-H *p*-value)**
Granulocyte Adhesion and Diapedesis	4.37E+00
Agranulocyte Adhesion and Diapedesis	4.33E+00
Complement System	3.54E+00
Phagosome Formation	2.75E+00
Hepatic Fibrosis / Hepatic Stellate Cell Activation	2.74E+00
Natural Killer Cell Signaling	2.71E+00
Leukocyte Extravasation Signaling	2.71E+00
Autoimmune Thyroid Disease Signaling	2.71E+00
Dendritic Cell Maturation	2.68E+00
Allograft Rejection Signaling	2.68E+00
Graft-versus-Host Disease Signaling	2.68E+00
Cytotoxic T Lymphocyte-mediated Apoptosis of Target Cells	2.12E+00
Crosstalk between Dendritic Cells and Natural Killer Cells	2.01E+00
MSP-RON Signaling Pathway	2.01E+00
TREM1 Signaling	1.93E+00
iCOS-iCOSL Signaling in T Helper Cells	1.79E+00
Type I Diabetes Mellitus Signaling	1.74E+00
Th1 Pathway	1.74E+00
Neuroinflammation Signaling Pathway	1.74E+00
Th1 and Th2 Activation Pathway	1.62E+00
Atherosclerosis Signaling	1.53E+00
Tumoricidal Function of Hepatic Natural Killer Cells	1.46E+00
Th2 Pathway	1.46E+00
B Cell Development	1.33E+00
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	1.30E+00
PKCŒ∏ Signaling in T Lymphocytes	1.30E+00

Table7B Legend: 148 genes significantly correlated with mesangial volume at the threshold of $r \geq |0.25|$ and $p \leq 0.05$ were selected for IPA. Significant canonical pathways after the Benjamin-Hochberg procedure ($p \leq 0.05$) are included in the table.

Table 7C: Pathways enriched in gene set associated with all podocytes structural measurements

Podocyte signature–associated pathways	**-log(B-H *p*-value)**
EIF2 Signaling	1.43E+01
Hepatic Fibrosis / Hepatic Stellate Cell Activation	5.84E+00
Allograft Rejection Signaling	4.60E+00
mTOR Signaling	4.55E+00
Regulation of eIF4 and p70S6K Signaling	4.19E+00
OX40 Signaling Pathway	3.82E+00
CD28 Signaling in T Helper Cells	3.62E+00
Graft-versus-Host Disease Signaling	3.57E+00
Granulocyte Adhesion and Diapedesis	3.57E+00
Th1 and Th2 Activation Pathway	3.57E+00
Th2 Pathway	3.57E+00
Coronavirus Pathogenesis Pathway	3.57E+00
T Helper Cell Differentiation	3.33E+00
Cdc42 Signaling	3.33E+00
Systemic Lupus Erythematosus In T Cell Signaling Pathway	3.23E+00
Th1 Pathway	3.20E+00
Autoimmune Thyroid Disease Signaling	3.13E+00
Calcium-induced T Lymphocyte Apoptosis	2.95E+00
iCOS-iCOSL Signaling in T Helper Cells	2.95E+00
Tumoricidal Function of Hepatic Natural Killer Cells	2.92E+00
Agranulocyte Adhesion and Diapedesis	2.82E+00
Type I Diabetes Mellitus Signaling	2.73E+00
B Cell Development	2.57E+00
Cytotoxic T Lymphocyte-mediated Apoptosis of Target Cells	2.47E+00
Altered T Cell and B Cell Signaling in Rheumatoid Arthritis	2.26E+00

Table 7C Legend: 770 genes significantly correlated with any podocyte measurement at the threshold of $r \geq |0.25|$ and $p \leq 0.05$ were selected for IPA. The top 25 out of 69 significant canonical pathways after the Benjamin-Hochberg procedure ($p \leq 0.05$) are included in the table.

Table 7D: Pathways enriched in gene set associated with fenestrated endothelium damage

Pathways enriched in signatures associated with FEN	**-log(B-H p-value)**
Altered T Cell and B Cell Signaling in Rheumatoid Arthritis	3.30E+00
Hepatic Fibrosis / Hepatic Stellate Cell Activation	2.02E+00
CD40 Signaling	2.02E+00
iCOS-iCOSL Signaling in T Helper Cells	2.02E+00
HMGB1 Signaling	2.02E+00
Regulation of IL-2 Expression in Activated and Anergic T Lymphocytes	2.02E+00
TNFR2 Signaling	2.02E+00
GP6 Signaling Pathway	2.02E+00
Death Receptor Signaling	2.00E+00
Phospholipase C Signaling	2.00E+00
Role of NFAT in Regulation of the Immune Response	1.98E+00
CD27 Signaling in Lymphocytes	1.89E+00
MIF Regulation of Innate Immunity	1.74E+00
B Cell Activating Factor Signaling	1.74E+00
April Mediated Signaling	1.74E+00
Dendritic Cell Maturation	1.68E+00
CD28 Signaling in T Helper Cells	1.65E+00
Primary Immunodeficiency Signaling	1.65E+00
Systemic Lupus Erythematosus In T Cell Signaling Pathway	1.63E+00
Crosstalk between Dendritic Cells and Natural Killer Cells	1.62E+00
Atherosclerosis Signaling	1.51E+00
IL-15 Production	1.51E+00
MIF-mediated Glucocorticoid Regulation	1.45E+00
Th1 and Th2 Activation Pathway	1.45E+00
T Cell Receptor Signaling	1.43E+00

Table 7D Legend: 252 genes significantly negatively correlated with mesangial volume at the threshold of $r \geq |0.25|$ and $p \leq 0.05$ were selected for IPA. 25 significant canonical pathways after the Benjamin-Hochberg procedure ($p \leq 0.05$) are included in the table. Note: None of the pathways passed the multiple testing threshold for the gene set with a positive correlation with fenestrated endothelium (FEN).

5.4 Mapping of transcripts into single cell

5.4.1 Clinical characteristics

We used scRNAseq characterize or map the cellular sources of signatures associated with structural damage from glomerular and tubulointerstitial bulk mRNA profiles. Single-cell kidney expression profiles were obtained from the same American Indian population of early DKD used for the bulk studies, but from an independent set of patients (*N*=44). Characteristics of the DKD cohort at the time of sample collection are provided in Table 8.

Table 8: Characteristics of 44 Pima Indians with single-cell profiling

Characteristic	At time of biopsy (N=44)
Age, years	41 ± 11
Diabetes duration, years	12.2 ± 7.5
Male sex, N (%)	14 (32)
Body Mass Index (kg/m2)	36.9 (7.3)
HbA1c (%)	9.2 ± 2.4
Systolic blood pressure (mmHg)	119 ± 12
Diastolic blood pressure (mmHg)	72 ± 10
ACR, median [IQR](mg/g)	18 [9 - 53]
mGFR (mL/min)	159 ± 58

Table 8 Legend: Mean with SD are provided for continuous variables. ACR: albumin-to-creatinine ratio; mGFR: iothalamate measured glomerular filtration rate. Median [IQR] for ACR.

5.4.2 Detection of kidney cell types

The scRNAseq analysis underwent rigorous preprocessing and filtering as explained in the Methods section. The combined samples yielded 85232 cells that passed quality control parameters. Unsupervised clustering at a granularity resolution of 0.6 using all cells pass

quality control resulted in 18 distinct clusters as shown in the uniform manifold approximation and projection plot in Figure 14. Literature-derived known kidney cell type–specific markers from several public sources were used to identify clusters and annotate labels for specific cell types. Kidney cell types and tissue resident immune cells were detected from kidney biopsy specimens using this scRNAseq protocol from 44 DKD kidney biopsies [72]. The 18 clusters covered the entire kidney cell lineage along nephron and tissue resident immune cells. The size of cell clusters ranged from 12 to 17,571 cells. Most samples contributed cells to each cluster.

Figure 14: Unsupervised clustering of cells from 44 patients with diabetic kidney disease

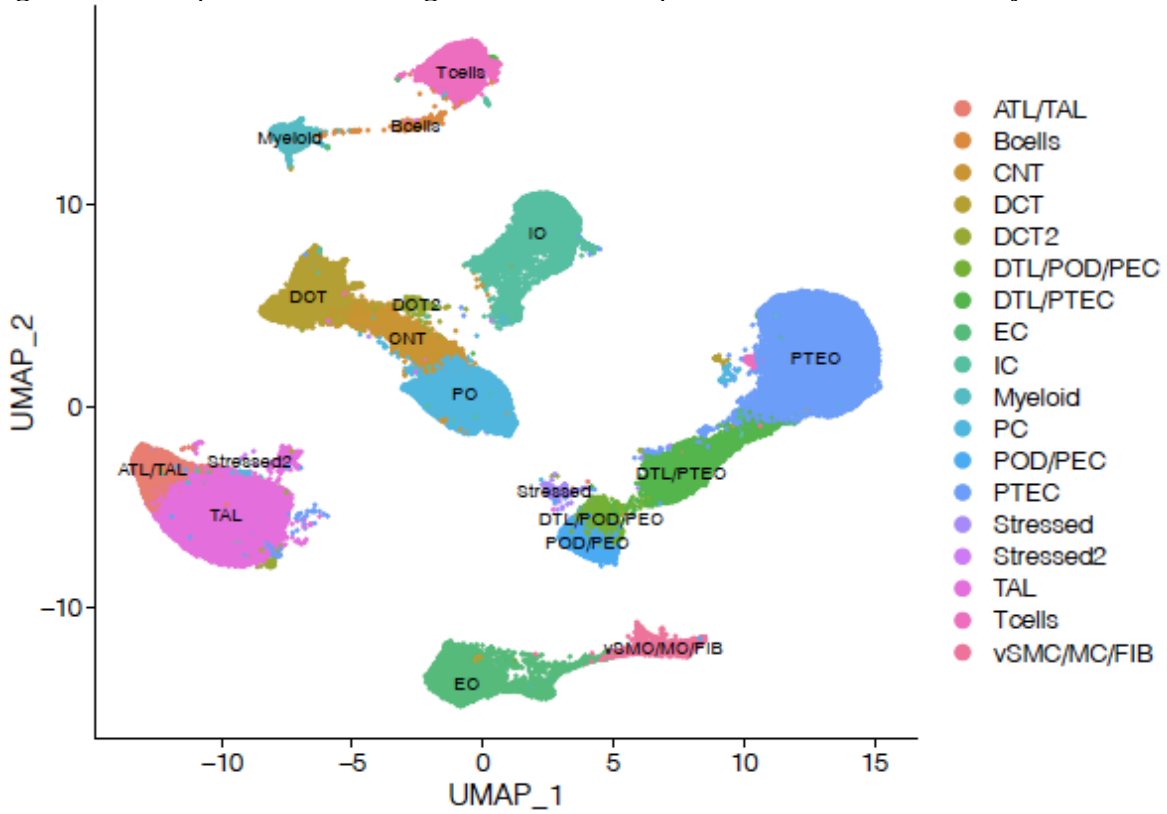


Figure 14 Legend: Uniform manifold approximation and projection plot from unsupervised clustering of 85232 cells from 44 DKD samples into 18 distinct cell type clusters. Each cluster is denoted by specific colors. Each point within each cluster is a cell. ATL: Ascending thin loop of Henle; CNT: Connecting tubule; DCT: Distal connecting tubule; DTL: Descending loop of Henle; EC: Endothelial cells; IC: intercalated cells; MC: Mesangial Cell; PC: Principal cells; PEC: Parietal epithelial cell; POD: Podocyte; PTEC: Proximal tubular cells; TAL: Thick ascending loop; vSMC: Vascular smooth muscle cells.

5.4.3 Cellular fingerprint of morphometric trait-associated genes

Fibrosis genes identified from the co-expression modules and the VvInt traits and genes associated with glomerular structural lesions were superimposed on scRNAseq data to identify the signatures' cell lineage. These genes were initially identified from the bulk RNA, which is expected to encompass signals from all cell types in each tissue compartment in varying proportions. Genes' fingerprints to their cellular localization can be determined in several ways. Here we deployed an overlay of the co-expression signatures on the expression matrix, averaged from single-cell profiles. A majority of the 1,843 fibrosis-associated genes were present in single-cell data and showed higher enrichment in the proximal tubule, immune clusters, and endothelial clusters (Figure 15). Interestingly, the proximal tubule epithelial cell exhibited a different pattern than other cell types by falling into its own cluster or branch in the dendrogram. Some genes displaying high expression in the proximal tubule epithelial cell also showed slightly higher expression in the other tubular segments of the DTL.

Figure 15: Cellular map of fibrosis signatures in DKD

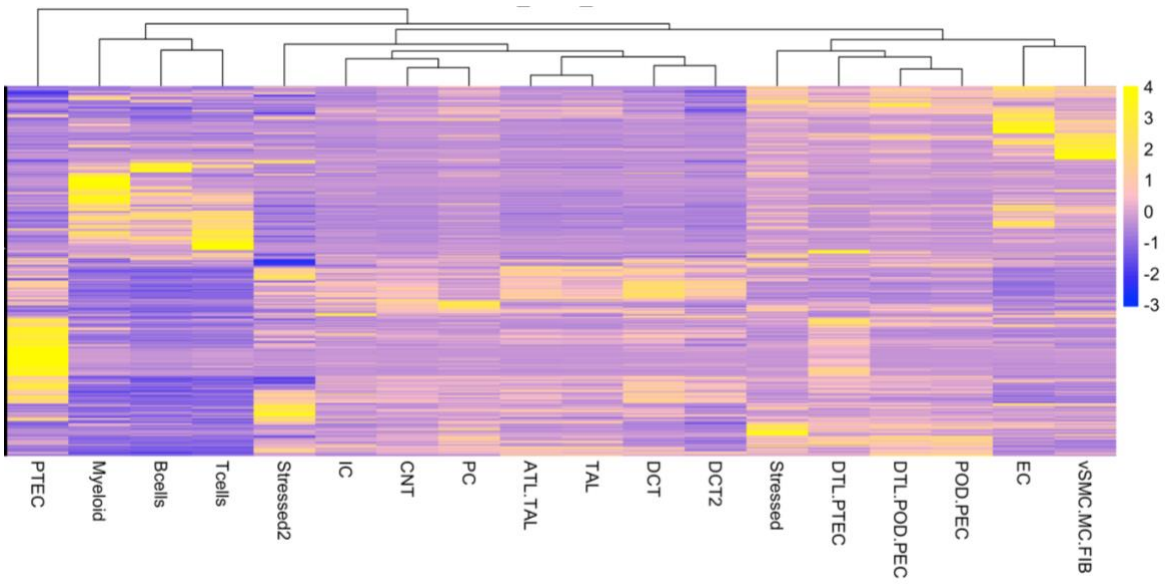


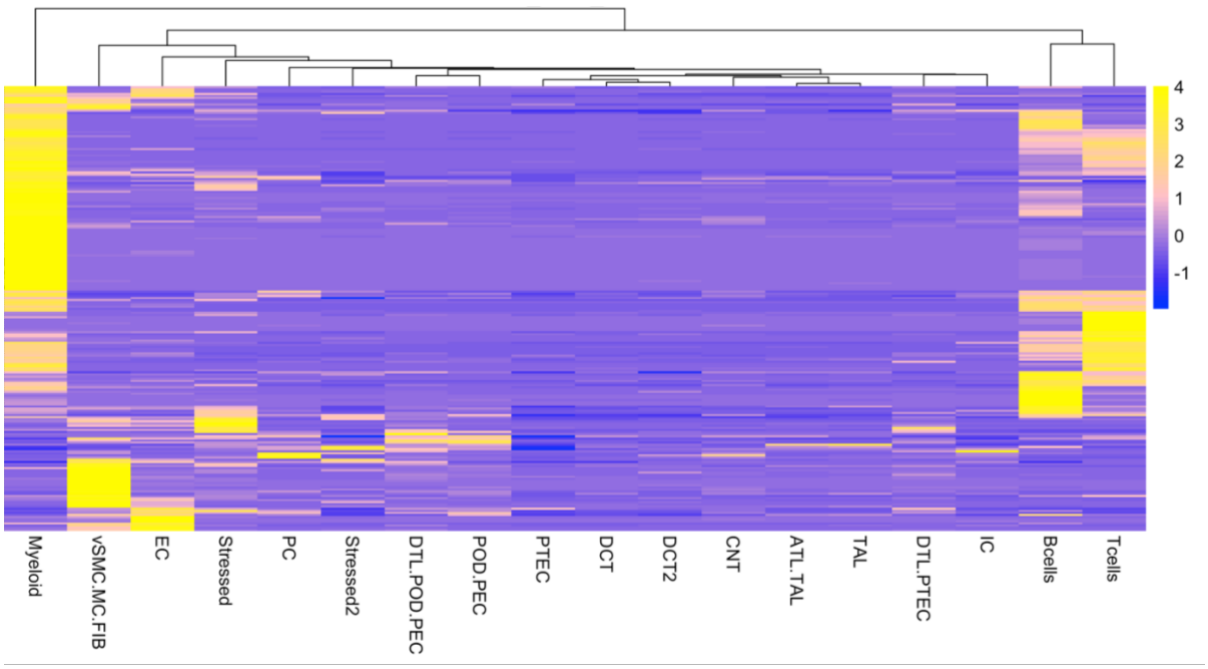
Figure 15 Legend: A heatmap showing the expression of fibrosis-associated genes across the cellular landscape of the DKD population. Raw intensities were z normalized. The color scale

follows the color legend bar on the right, with blue as the lowest signal and yellow as the highest signal.

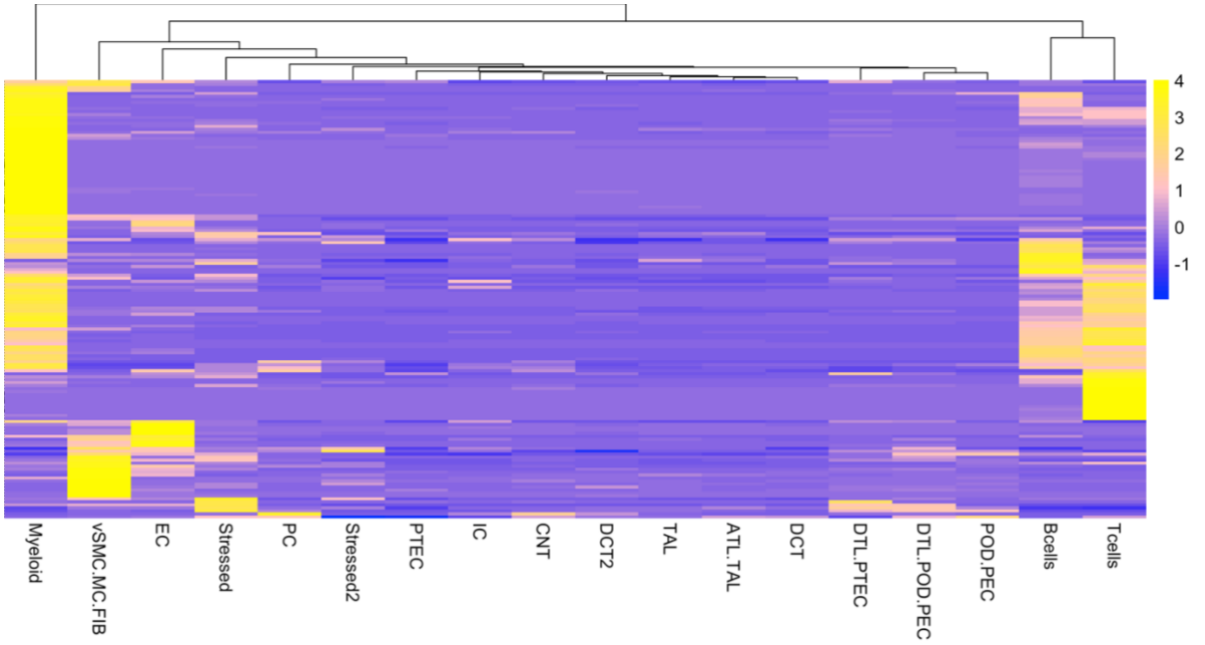
Module genes associated with the glomerular bulk RNA expression and traits were similarly evaluated for enrichment of cell specificity. Gene sets associated with each glomerular trait were mapped separately. Most genes were associated with increased mesangial expansion. The glomerular basement membrane width and podocyte loss were mapped to kidney resident endothelial cells and mesangial/smooth muscle cell clusters in addition to immune clusters (Figures 16A–16D).

Figure 16: Heatmap visual representation of cellular localization of glomerular trait–associated genes derived from bulk mRNA.

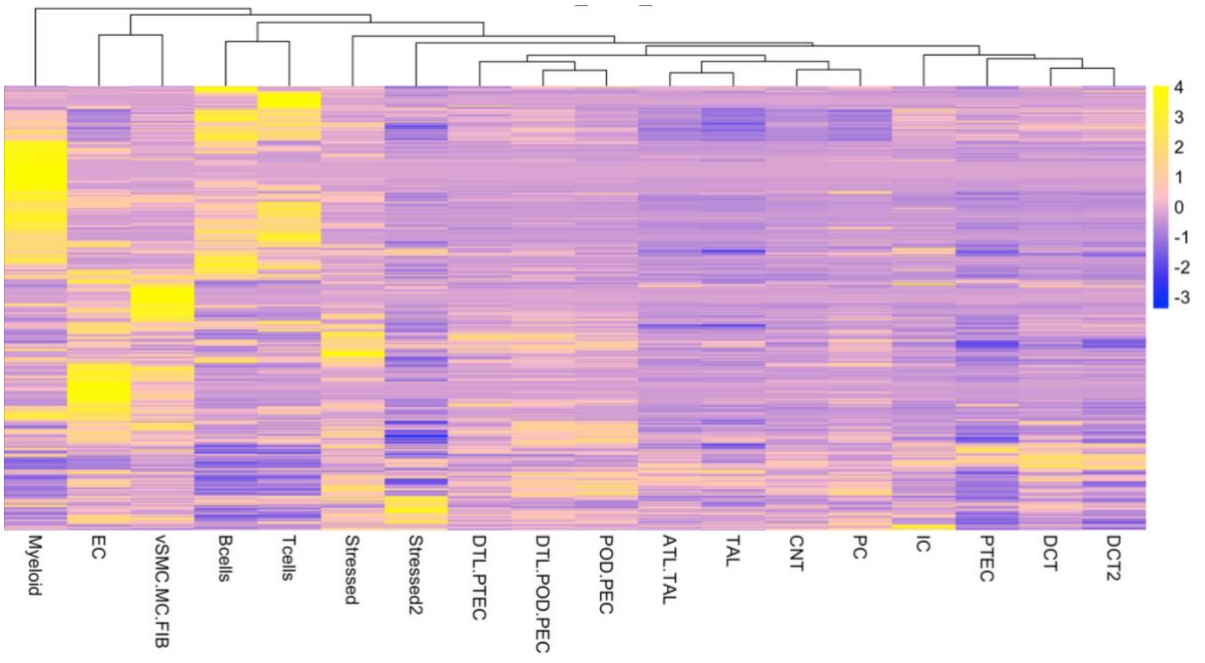
A.



B.


4
3
2
1
0
-1
Myeloid
vSMC.MC.FIB
EC
Stressed
PC
Stressed2
PTEC
IC
CNT
DCT2
TAL
ATL.TAL
DCT
DTL.PTEC
DTL.POD.PEC
POD.PEC
Bcells
Tcells

C.


4
3
2
1
0
-1
-2
-3
Myeloid
EC
vSMC.MC.FIB
Bcells
Tcells
Stressed
Stressed2
DTL.PTEC
DTL.POD.PEC
POD.PEC
ATL.TAL
TAL
CNT
PC
IC
PTEC
DCT
DCT2

D.

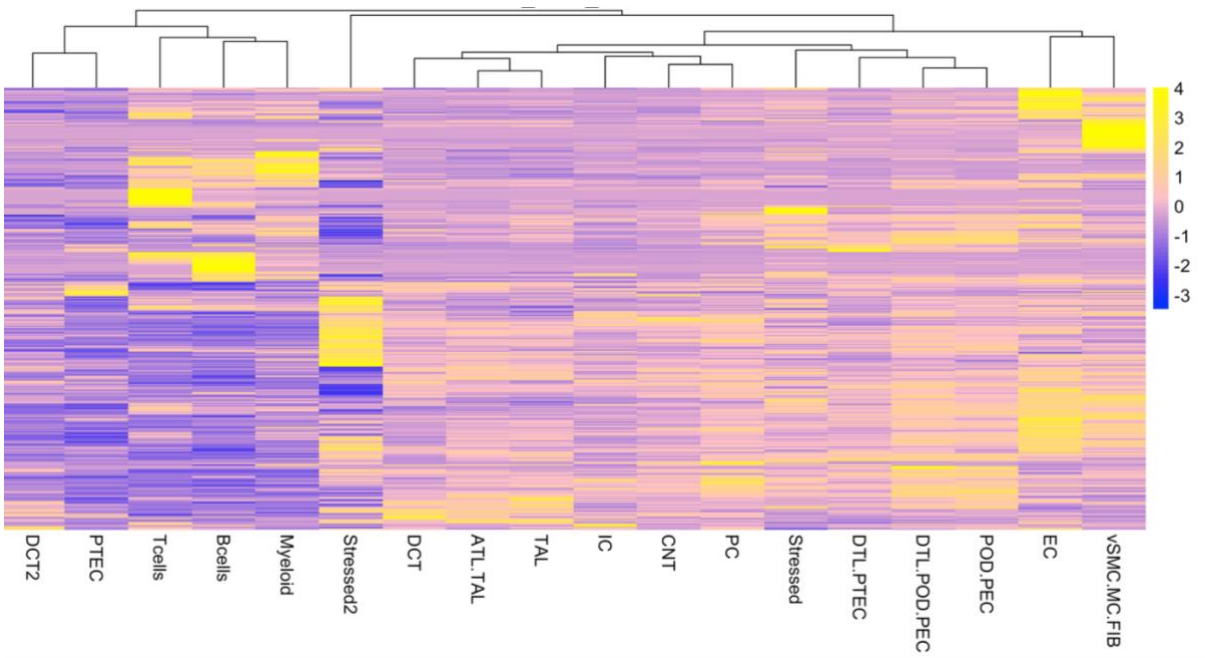


Figure 16 Legend: Single-cell localization of genes associated with structural traits [A–D] from glomerular bulk mRNA expression profiles. The color bar for expression intensity follows the scale in the legend for each plot from minimum (blue) to maximum (yellow). (A) Heatmap of the range of expression of 209 genes associated with GBM in cyan and green-yellow module in glomerular data; (B) Heatmap clustering of 147 genes from the green-yellow module associated with mesangial expansion on the single-cell landscape; (C) Heatmap of 750 genes associated with podocyte injury traits from modules (pink, black, tan, and green-yellow on the single-cell expression matrix); (D) Mapping of 645 genes from cyan, green-yellow, and brown modules associated with the fenestrated endothelium on DKD single-cell profiles.

6. Discussion

6.1 Overview

In this study, a systems biology approach sequentially combining multiomics was applied to explore mechanisms of early morphogenomic changes in DKD. Transcriptomics, quantitative morphometrics, and clinical outcome data constituted the evaluated layers. Structural damage precedes the clinical manifestation of GFR and ACR, such that nephrologists often miss these

early indications in the clinical setting. A lack of effective biomarkers for detecting early DN has hindered understanding of cellular and molecular events correlated with the earliest structural changes in DN. Here, protocol biopsies from a homogenous type 2 diabetes population enabled the detection of genome-wide signals expressed in the early stage of DKD. By leveraging structural indications of disease stages, tubulointerstitial damage and glomerular structural lesions were integrated with transcriptomic levels. This approach led to the identification of various key signaling and metabolic pathways activated in early DKD. The powerful resource of long-term outcomes enabled us to anchor these changes to kidney function over 15 years of post-biopsy follow-up. Significant enrichment in inflammatory mechanisms and metabolic processes was observed in positively and negatively correlated VvInt transcripts, respectively. The same approach was then applied to glomeruli to test this method's applicability in another pathophysiologic setting, namely by screening the associations of transcriptional levels with well-established glomerular lesions such as the glomerular basement membrane width, mesangial volume, podocyte loss, and endothelial fraction. Finally, single-cell sequencing was used to map the cellular localization of identified gene expression signatures.

6.2 Generalizability of findings beyond early DKD

This thesis describes the successful application of an approach that combines molecular and morphometric data in early DKD. The ultimate goal of integration studies such as this one is to identify early targets of therapeutic intervention (in our case, to slow the decline in kidney function before irreversible damage has occurred). Gene expression sets from tubulointerstitial, and glomerular compartments were evaluated for disease fingerprints via a "guilt-by-association" approach using a co-expression method. The VvInt-associated transcripts were further evaluated for functional disease outcomes (mGFR/ACR trajectories) over 10 years of

post-biopsy follow-up. To determine whether the activation of these signatures was specific to the disease population under study, we validated our findings in a different cohort (i.e., a European cohort). Participants of this ERCB cohort were in more advanced CKD stages with an average GFR of 44 mL/min per 1.73 m^2 compared to an mGFR of 147mL/min in Pima Indians. Roughly, 80% of implicated transcripts in Pima Indians were differentially regulated in this European cohort. These transcripts showed concordant changes of disease association as compared to healthy controls. Genes that exhibited positive/negative associations with scarring were also upregulated/downregulated, respectively, in advanced DKD. Notably, the studied cohorts differed on an array of features ranging from ethnic characteristics, environmental changes, disease stage, BMI, and other comorbidities. The concordant damage identified across cohorts therefore holds great potential clinical importance. The early onset of diabetes in the Pima Indian cohort highlights an urgent need for early intervention in the disease course. The high frequency of kidney failure (ESRD) by the age of 45 in this population with youth onset diabetes mandates further study [79]. The approach as outlined here clearly demonstrates the potential of integrating multiomics data to capture relevant regulatory events. In this instance, the option of intervening at early disease stages should be far more effective than intervening later after the development of extensive tissue damage.

Identifying gene signatures and associated functional units of pathways was here shown to facilitate the identification of dysregulated genes/functions in a tissue/disease context. Several transcriptomic studies, focusing on differential changes between DKD from indication biopsies and normal kidneys in human and in different diabetic mice models and controls, have demonstrated the utility of pathway identification in disease settings [53, 80-82]. These studies examined the role of the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) and JAK-STAT signaling and other inflammatory markers, now established as a pathogenetic contributor to DKD progression [53, 80, 81, 83, 84]. Such research has also revealed

evidence of the JAK-STAT signaling mechanism in early DKD, advanced DKD, and murine modes. These molecular whole-genome studies provided the rational for an interventional clinical trial evaluating the safety and kidney efficacy of Baricitinib, a *JAK1* and *JAK2* inhibitor, as a potential therapy in patients with advanced DKD (ClinicalTrials.gov number *NCT01683409*) [85]. This trial provides a compelling example of successfully applying bioinformatics-driven bench-side findings in clinical care. Baricitinib has been found to significantly lower the level of urine albumin excretion, the primary outcome of the trial. The dose-dependent reduction of uACR compared to placebo after six months of treatment has shown a sustained reduction in urine protein four weeks after discontinuing Baricitinib [86]. Of note, our and other transcriptomic studies predicted lower blood and urinary markers (CXCL10, VEGFA, CCL1) linked to JAK-STAT activation under treatment. Reductions in these markers were found to precede a reduction in albuminuria, underscoring the promise of this approach in identifying targets along with their engagement biomarkers.

6.3 Bioinformatic approach captures crosstalk of processes

The "translational" approach employed in the Baricitinib trial offered an example as to how to potentially discern related targets of interventional utility. However, a single gene or pathway is rarely the sole stimulus of disease progression, especially in heterogenous multifactorial diseases such as DKD where multiple cross-communicating mechanisms are thought to dictate disease progression at various disease stages. A systematic computational bioinformatic method can be used to help integrate and model these interactions to better explain the underlying disease physiology. Flexible in-silico modelling approaches can help facilitate identification of individual pathways as well as the regulatory cap of interacting pathways. These regulatory nodes can serve as potential drug targets or markers of integrated disease

activity. The tubulointerstitial fibrosis pathway network model studied here shows clear evidence of such active cross-communication among dysregulated pathways. Specifically, distinct clusters of dense networks of tubular dedifferentiation and inflammation processes (Figure8) mediated through two pathways: mitochondrial dysfunction and LPS/IL-1–mediated inhibition of RXR function. A growing body of evidence supports the suggestion of dysfunctional mitochondria as pathological mediators of DKD. Mitochondrial dysfunction has been linked to tubular injury in the kidney environment, interfering with cellular signaling and the generation of augmented reactive oxygen species. The disease state in the diabetes milieu alters this pathway, including glucose oxidation in proximal tubules. Studies of pre-clinical models have also suggested a potential pharmacological effect of targeting mitochondria to improve kidney function [87]. In particular, Zhan et al. [88] observed that a tubular-specific enzyme called myo-inositol oxygenase had increased expression in diabetic mice along with mitochondrial fragmentation. D-glucarate supplementation in these diabetic mice showed attenuated myo-inositol oxygenase expression, tubular damage, and improved kidney function.

6.4 Upstream regulator of clusters

The upstream regulator analysis detailed here effectively captured the cascade of upstream regulators influencing the transcriptional gene expression changes occurring downstream. Several studies in other modalities have suggested the effectiveness of this approach in identifying potential regulators and target genes with novel therapeutic and diagnostic utility [89-91]. The approach detailed here strongly suggests the ability to identify upstream regulators from clusters in the pathway network. Upon screening the pathways affecting the inflammation cluster, the top upstream regulator was found to be *IL1β*. A set of five master transcription factors (*PHF1*, *SOX2*, *NFAT5*, *TRIM29*, *HEY1*) through a mediator of *TP53* were activated in the dedifferentiation cluster. The regulators identified are well-studied targets in DKD. *IL1β*-

dependent mechanisms have been described in DKD model systems of advanced tubular cell dysfunction [92, 93] where these mechanisms were found to link apoptosis and innate immune activation. *IL1β*-driven genes are important downstream signaling elements of the inflammasome that becomes activated during the progressive damage associated with loss of tubular function [94].

Dedifferentiation master regulators (*PHF1*, *SOX2*, *NFAT5*, *TRIM29*, *HEY1*) were found to be connected via a common mediator *TP53*. This set should influence downstream targets involved in oxidative phosphorylation, mitochondrial dysfunction, and the production of nitric oxide and reactive oxygen species in macrophages. Intra kidney oxidative stress is a marker of both disease initiation and progression. Studies have shown that the cell-specific deletion of TP53 from proximal cells attenuates oxidative stress, macrophage infiltration, and tubular damage[95]. *NFAT5* is an osmotic stress-response transcription factor shown to have tonicity-dependent and independent modes of actions [96]. *NFAT5* provides a potential link between diabetic hyperosmotic stress and disease progression in part through modulation of pro-inflammatory cytokine expression by resident epithelial tissue and immune cells.

6.5 Systems biology identifies known and new targets associated with early structural lesions of DKD

Previous studies have identified an association between mesangial fractional volume, surface density of the glomerular basement membrane, and glomerular volume with GFR decline[40]. Importantly, these structural lesions can initiate and develop even when patients are clinically silent. No diabetes studies to date have linked molecular correlates of this early surrogate outcome of disease progression to underlying pathways, which could in theory reveal targets for early intervention. Pathways currently undergoing drug targeting, such as the GP6 signaling

pathway, were identified as “significant pathways” captured in our bioinformatic integration of structure and transcriptome. High blood glucose, vascular shear stress, and oxidative stress often coexist in patients with diabetes and are known risk factors for cardiovascular events. Signaling events downstream of the GP6 pathway are influenced by hyperglycemia, oxidative stress, and shear stress. Another well-known podocyte injury pathway, the mTOR signaling pathway, was also identified in glomerular transcriptomic associations in this study. The established role of mTOR in diabetic podocyte injury helped to validate our general analysis method [97, 98].

Inflammation is associated with progressive type 2 diabetes but has yet to be fully characterized. We also observed high enrichment and molecular fingerprints of inflammation markers and associated pathways in the network identified in VvInt-correlated tubulointerstitial transcripts. Similar inflammation signatures appeared in the glomerular data. Innate immune mediators play critical roles in the pathogenesis of type 2 diabetes. Activation of the nuclear factor-κB (NF-κB) signaling pathway via receptors for advanced glycation end products and toll-like receptors can contribute to inflammation through HMGB1[99]. HMGB1 also contributes to the production of proinflammatory cytokines including *IL-6, IL-1β, and TNF-α*, all of which were captured in DKD transcriptional signatures [100, 101]. Although inflammation is commonly associated with late events in DKD, our data demonstrate that these signatures are enriched and captured at clinically early stages of DKD. The macrophage migration inhibitory factor, another proinflammatory cytokine of the innate immune system and a main player in the induction of immunoinflammatory responses, was also enriched in the glomerular signatures studied here. These results provide convincing evidence of inflammation events in early DKD stages and strongly imply that anti-inflammatory intervention may be useful early in disease progression.

Genome-wide expression profiles represent powerful tools to capture the active tissue transcriptional response. The functional state of the transcriptional regulatory mechanism can be inferred using several approaches, such as the causal inference network method. This method has been applied successfully in kidney disease: kidney mRNA profiles were associated with disease progression to identify predictors of disease in cohorts with established CKD, including DKD. The transcripts were then screened as noninvasive targets and successfully validated as diagnostic markers across different cohorts [67]. By employing a similar approach, the signatures of early DKD reported here can be used as a framework for future experimental studies to capture an upstream regulatory mechanism of phenotype-associated signatures. Doing so may help researchers identify noninvasive markers that reflect intra–kidney disease processes occurring in patients with undetectable disease activity during preclinical presentation.

6.6 Cellular source of structure-associated functions

Single-cell sequencing was used in this study to map the expression modules identified by bioinformatics into distinct enriched cell types. The kidney is a complex organ consisting of highly specialized cell types. The sequencing technology employed in the first part of this study successfully identified global transcriptional changes occurring in the two major tissue compartments: the glomerulus and tubulointerstitial tissues. Based on the experimental constrains, these results could not be ascribed to a specific cell type. Single-cell sequencing facilitates the generation of individual transcriptomes from several thousands of cells in complex tissues, thus enabling detection of transcriptional regulation in distinct cell types. This method further allows for the identification of known or as-yet-unidentified cell subtypes or cellular states to provide evidence of unique molecular processes activated in these cell

lineages [71]. When diseased tissue can be compared to healthy control, cell types as well as molecular dynamics implicated in the disease stage can be identified.

The scRNAseq analysis of a cohort of 44 DKD kidney samples revealed 18 distinct cell clusters comprising the proximal tubule, immune cell, and endothelial clusters. This single-cell technology also indicated several unique immune clusters, including a myeloid cluster along with B cell and T cell clusters. The resultant dataset provides a reference atlas of single-cell populations in early DKD. These scRNAseq data can be used to identify similarities, differences, and relationships among cell lineages as well. Additionally, the approach can be used in conjunction with other data types involving bulk RNA[102]. Deconvolution (i.e., signal restoration) of single-cell data and bulk RNA profiles obtained from cohorts with long-term follow-up can also unveil cell-specific disease prognostic markers. The clusters identified in this study can be further dissected at different levels of granularity to allow the study of new cell types as their specific markers emerge.

For example, the mapping of fibrosis-associated genes onto the single-cell map showed an enrichment of these signatures in the proximal tubule cell cluster. A logical extension of the present study might include the evaluation of potential cross-communication between cell types of interest by investigating ligand/receptor expression patterns between epithelial cells, fibroblast, immune and endothelial cells. The characterization of cell–cell communication could provide information about normal kidney function and how it is altered by disease. Analyzing and mapping structural correlates onto a single-cell population can help identify where kidney dysfunction has impacted cellular and tissue integrity and prioritize these associations for further mechanistic studies.

7. Limitations

This study has several limitations. A lack of control subjects from the same population hindered inferences regarding how the structure and molecular fingerprints were associated with the normal condition. Another limitation is the limited sample size, which is a key obstacle in expression studies especially those on DKD. Comprehensive phenotype data were obtained over 20 years of patient follow-up, expression profiling, and pathological structural measurements. Paradoxically, the unique, extended follow-up of the Pima protocol biopsy cohort represents another limitation; it would be challenging to find another type 2 DKD cohort with such extended follow-up and rich phenotypic resources which is a key strength of this study. Single-cell sequencing was performed on a different group of participants (albeit in the same population) due to a lack of single-cell profiling techniques when biopsy tissue samples were procured for expression studies.

8. Conclusion

The innovative integration of system biology and computational approaches in this study is theoretically applicable to any domain in which disease-associated mechanisms can be identified. Conventional methods often fail to capture the vast dimensionality of features when dealing with small datasets. Novel bioinformatics methods should thus be adopted to identify mechanisms that can be rationalized biologically and statistically. The WGCNA approach adopted here is highly flexible: this method computes eigen genes that summarize biological information contained in the genes assigned to each module into a single value per sample. This aggregation greatly reduces dimensionality while maintaining the flexibility to dissect modules into individual genes as needed. Another advantage of this type of method is that the identified vectors can be further passed onto other complex algorithms (e.g., Bayesian networks or classic regression methods). A biological mechanism is not driven by individual

genes but rather involves multiple genes and their interactions. Network analysis can effectively capture subtle, coordinated changes in direct or indirect interactions between entities. This technique is particularly common in social science research, with several popular tools and apps having adopted this algorithm.

Researchers have gradually begun to identify the power and utility of these network analysis tools in medical research. A paradigm shift towards more open-source and collaborative spaces has promoted this change. Most bioinformatic and systems biology methods have borrowed concepts from other domains. Approaches are often independent of the disease setting and are instead contingent on the structure of data. This broad applicability benefits medical research, as methods can be applied to kidney disease and other domains (e.g., complex diseases such as neurological disorders and cancer). WGCNA can specifically be employed to identify candidate biomarker genes and therapeutic targets based on the associations of module genes and hub genes with disease phenotypes and disease outcomes. This method has already been adopted in cancer studies to identify biomarkers [103]. This strategy cannot only be applied on mRNA profiles, but also shows efficacy in the small RNA world, such as lnRNA where Le et. al [104] used WGCNA to analyze mRNA and lnRNA expression profiles in triple-negative breast cancers; the authors ultimately identified a set of disease-specific mRNAs and lnRNAs as potential therapeutic targets.

In addition to being applicable to different disease domains, these methods are largely technology- or platform-independent. The algorithms can therefore perform well on high-dimensional data generated from various high-throughput platforms (e.g., RNAseq, scRNAseq, or high-throughput proteomic platforms). Several scholars have implemented these methods in various omic areas, including proteomics [105] and metabolomics [106].

9. Future perspectives

This thesis describes the use of transcriptomic profiles and structural lesions to identify early disease surrogate expression signatures associated with long-term disease progression. Our work identified novel inflammatory and metabolic programs associated with fibrosis and glomerular lesions. In particular, this study highlights the complex multifactorial and heterogenous pathophysiology of DKD. Upstream regulator analysis revealed several key drivers orchestrating downstream events. The application and validation of single-cell analysis provided a cellular landscape of the altered regulation of these genes. Subsequent follow-up studies could prioritize disease-associated candidate genes for evaluation as biomarkers.

To fully harness the clinical utility of new biomarkers in precision medicine, scholars will need to integrate proteomic and metabolomic profiles preferably from the same patients. The findings of this thesis could be extended to the assessment of noninvasive biomarkers from urine/blood biofluids from enriched signatures (Figure 17). Noninvasive biomarkers hold great potential clinical utility and can enhance the applicability of research findings from bench to bedside in clinical settings.

Figure 17: Updated scheme illustrating the inclusion of biofluids to the implemented analytical pipeline

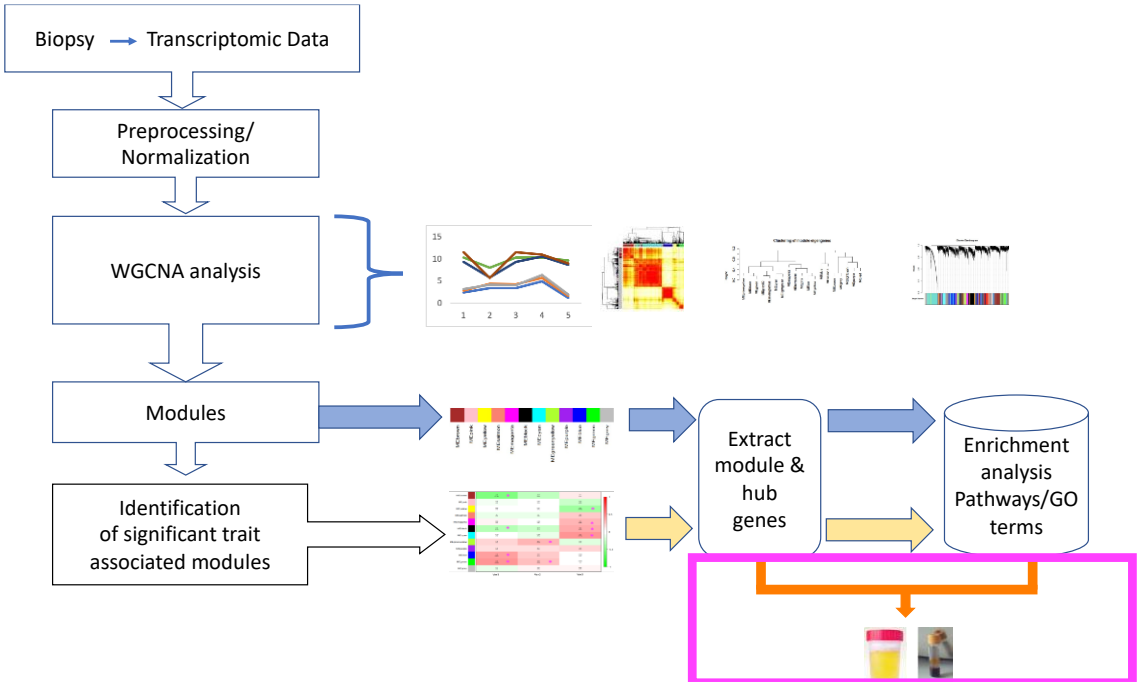


Figure 17 Legend: Schematic flowchart illustrating the process of biopsy, data preprocessing, and analysis. This figure highlights the inclusion of urine/blood samples to be integrated with module genes. WGCNA: weighted gene co-expression network analysis. WGCNA comprises several methods: co-expression patterns, module detection, module clustering based on eigen genes, and reiterative merging of modules.

A well-powered DKD cohort with multiomics data from the same patients will be necessary for a more integrated analysis and multiple activities are ongoing in this field, most notably the Kidney Precision Medicine Project (KPMP), utilizing many aspects of the studies described in this thesis. Advances in machine learning methods and artificial intelligence have led to new opportunities for data integration in bioinformatics. Precision medicine requires information coalition in a patient-centric manner. Medical research is now embracing the concept of multiomics data integration, with several methods under development. Combining clinical and histological information with multiomics profiles will provide more insight into the cellular functions at work in disease progression. This thesis highlights the potential of combining two

data domains with extension to the single-cell level. Overall, integrating proteomics and metabolomics will improve the prognostic and diagnostic predictive accuracy of disease phenotypes and enhance treatment stratification.

References

1. Jha V, Garcia-Garcia G, Iseki K, Li Z, Naicker S, Plattner B, Saran R, Wang AY, Yang CW. Chronic kidney disease: global dimension and perspectives. Lancet. 2013;382(9888):260-72. Epub 2013/06/04. doi: 10.1016/s0140-6736(13)60687-x. PubMed PMID: 23727169.
2. Levey AS, Coresh J, Balk E, Kausz AT, Levin A, Steffes MW, Hogg RJ, Perrone RD, Lau J, Eknoyan G. National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Ann Intern Med. 2003;139(2):137-47. Epub 2003/07/16. doi: 10.7326/0003-4819-139-2-200307150-00013. PubMed PMID: 12859163.
3. Girndt M, Trocchi P, Scheidt-Nave C, Markau S, Stang A. The Prevalence of Renal Failure. Results from the German Health Interview and Examination Survey for Adults, 2008-2011 (DEGS1). Dtsch Arztebl Int. 2016;113(6):85-91. Epub 2016/03/05. doi: 10.3238/arztebl.2016.0085. PubMed PMID: 26931624; PMCID: PMC4782264.
4. Keane WF, Eknoyan G. Proteinuria, albuminuria, risk, assessment, detection, elimination (PARADE): a position paper of the National Kidney Foundation. Am J Kidney Dis. 1999;33(5):1004-10. Epub 1999/04/23. doi: 10.1016/s0272-6386(99)70442-7. PubMed PMID: 10213663.
5. Inker LA, Astor BC, Fox CH, Isakova T, Lash JP, Peralta CA, Kurella Tamura M, Feldman HI. KDOQI US commentary on the 2012 KDIGO clinical practice guideline for the evaluation and management of CKD. Am J Kidney Dis. 2014;63(5):713-35. Epub 2014/03/22. doi: 10.1053/j.ajkd.2014.01.416. PubMed PMID: 24647050.
6. Clase CM, Garg AX, Kiberd BA. Classifying kidney problems: can we avoid framing risks as diseases? Bmj. 2004;329(7471):912-5. Epub 2004/10/16. doi: 10.1136/bmj.329.7471.912. PubMed PMID: 15485978; PMCID: PMC523126.
7. de Jong PE, Gansevoort RT. Fact or fiction of the epidemic of chronic kidney disease--let us not squabble about estimated GFR only, but also focus on albuminuria. Nephrol Dial Transplant. 2008;23(4):1092-5. Epub 2008/03/25. doi: 10.1093/ndt/gfn028. PubMed PMID: 18359868.
8. Eckardt KU, Berns JS, Rocco MV, Kasiske BL. Definition and classification of CKD: the debate should be about patient prognosis--a position statement from KDOQI and KDIGO. Am J Kidney Dis. 2009;53(6):915-20. Epub 2009/05/02. doi: 10.1053/j.ajkd.2009.04.001. PubMed PMID: 19406541.

9. Gansevoort RT, de Jong PE. The case for using albuminuria in staging chronic kidney disease.JAm Soc Nephrol. 2009;20(3):465-8. Epub 2009/03/04. doi: 10.1681/asn.2008111212. PubMed PMID: 19255126.
10. Glassock RJ, Winearls C. Ageing and the glomerular filtration rate: truths and consequences. Trans Am Clin Climatol Assoc. 2009;120:419-28. Epub 2009/09/22. PubMed PMID: 19768194; PMCID: PMC2744545.
11. Glassock RJ, Winearls C. An epidemic of chronic kidney disease: fact or fiction? Nephrol Dial Transplant. 2008;23(4):1117-21. Epub 2008/03/25. doi: 10.1093/ndt/gfn086. PubMed PMID: 18359870.
12. Glassock RJ, Winearls C. Screening for CKD with eGFR: doubts and dangers. Clin J Am Soc Nephrol. 2008;3(5):1563-8.Epub 2008/08/01. doi: 10.2215/cjn.00960208. PubMed PMID: 18667744; PMCID: PMC4571145.
13. Hallan SI, Orth SR. The KDOQI 2002 classification of chronic kidney disease: for whom the bell tolls. Nephrol Dial Transplant. 2010;25(9):2832-6. Epub 2010/07/06. doi: 10.1093/ndt/gfq370. PubMed PMID: 20601368.
14. Winearls CG, Glassock RJ. Dissecting and refining the staging of chronic kidney disease. Kidney Int. 2009;75(10):1009-14. Epub 2009/02/27. doi: 10.1038/ki.2009.49. PubMed PMID: 19242501.
15. Bauer C, Melamed ML, Hostetter TH. Staging of chronic kidney disease: time for a course correction. J Am Soc Nephrol. 2008;19(5):844-6. Epub 2008/04/04. doi: 10.1681/asn.2008010110. PubMed PMID: 18385419.
16. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, Colagiuri S, Guariguata L, Motala AA, Ogurtsova K, Shaw JE, Bright D, Williams R. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. Diabetes Res Clin Pract. 2019;157:107843. Epub 2019/09/14. doi: 10.1016/j.diabres.2019.107843. PubMed PMID: 31518657.
17. Saran R, Li Y, Robinson B, Ayanian J, Balkrishnan R, Bragg-Gresham J, Chen JT, Cope E, Gipson D, He K, Herman W, Heung M, Hirth RA, Jacobsen SS, Kalantar-Zadeh K, Kovesdy CP, Leichtman AB, Lu Y, Molnar MZ, Morgenstern H, Nallamothu B, O'Hare AM, Pisoni R, Plattner B, Port FK, Rao P, Rhee CM, Schaubel DE, Selewski DT, Shahinian V, Sim JJ, Song P, Streja E, Kurella Tamura M, Tentori F, Eggers PW, Agodoa LY, Abbott KC. US Renal Data System 2014 Annual Data Report: Epidemiology of Kidney Disease

in the United States. Am J Kidney Dis. 2015;66(1 Suppl 1):Svii, S1-305. Epub 2015/06/27. doi: 10.1053/j.ajkd.2015.05.001. PubMed PMID: 26111994; PMCID: PMC6643986.

18. Zhuo X, Zhang P, Barker L, Albright A, Thompson TJ, Gregg E. The lifetime cost of diabetes and its implications for diabetes prevention. Diabetes Care. 2014;37(9):2557-64. Epub 2014/08/26. doi: 10.2337/dc13-2484. PubMed PMID: 25147254.
19. Collins AJ, Foley RN, Chavers B, Gilbertson D, Herzog C, Ishani A, Johansen K, Kasiske BL, Kutner N, Liu J, St Peter W, Guo H, Hu Y, Kats A, Li S, Li S, Maloney J, Roberts T, Skeans M, Snyder J, Solid C, Thompson B, Weinhandl E, Xiong H, Yusuf A, Zaun D, Arko C, Chen SC, Daniels F, Ebben J, Frazier E, Johnson R, Sheets D, Wang X, Forrest B, Berrini D, Constantini E, Everson S, Eggers P, Agodoa L. US Renal Data System 2013 Annual Data Report.AmJKidneyDis.2014;63(1Suppl):A7.Epub2013/12/24.doi:10.1053/j.ajkd.2013.11.001.PubMed PMID: 24360288.
20. Viberti GC, Jarrett RJ, Keen H. Microalbuminuria as prediction of nephropathy in diabetics. Lancet. 1982;2(8298):611. Epub 1982/09/11. doi: 10.1016/s0140-6736(82)90688-2. PubMed PMID: 6125757.
21. Adler AI, Stevens RJ, Manley SE, Bilous RW, Cull CA, Holman RR. Development and progression of nephropathy in type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS 64). Kidney Int. 2003;63(1):225-32. Epub 2002/12/11. doi: 10.1046/j.1523-1755.2003.00712.x. PubMed PMID: 12472787.
22. Macisaac RJ, Ekinci EI, Jerums G. Markers of and risk factors for the development and progression of diabetic kidney disease. Am J Kidney Dis. 2014;63(2 Suppl 2):S39-62. Epub 2014/01/28. doi: 10.1053/j.ajkd.2013.10.048. PubMed PMID: 24461729.
23. Perkins BA, Ficociello LH, Silva KH, Finkelstein DM, Warram JH, Krolewski AS. Regression of microalbuminuria in type 1 diabetes. N Engl J Med. 2003;348(23):2285-93. Epub 2003/06/06. doi: 10.1056/NEJMoa021835. PubMed PMID: 12788992.
24. MacIsaac RJ, Tsalamandris C, Panagiotopoulos S, Smith TJ, McNeil KJ, Jerums G. Nonalbuminuric Renal Insufficiency in Type 2 Diabetes. Diabetes Care. 2004;27(1):195-200. doi: 10.2337/diacare.27.1.195.
25. Krolewski AS. Progressive renal decline: the new paradigm of diabetic nephropathy in type 1 diabetes. Diabetes Care. 2015;38(6):954-62. Epub 2015/05/23. doi: 10.2337/dc15-0184. PubMed PMID: 25998286; PMCID: PMC4439536.
26. Groop PH, Thomas MC, Moran JL, Wadèn J, Thorn LM, Mäkinen VP, Rosengård-Bärlund M, Saraheimo M, Hietala K, Heikkilä O, Forsblom C. The presence and severity of chronic

kidney disease predicts all-cause mortality in type 1 diabetes. Diabetes. 2009;58(7):1651-8. Epub 2009/04/30. doi: 10.2337/db08-1543. PubMed PMID: 19401416; PMCID: PMC2699848.

27. Orchard TJ, Secrest AM, Miller RG, Costacou T. In the absence of renal disease, 20 year mortality risk in type 1 diabetes is comparable to that of the general population: a report from the Pittsburgh Epidemiology of Diabetes Complications Study. Diabetologia. 2010;53(11):2312-9. Epub 2010/07/29. doi: 10.1007/s00125-010-1860-3. PubMed PMID: 20665208; PMCID: PMC3057031.
28. de Boer IH, Rue TC, Cleary PA, Lachin JM, Molitch ME, Steffes MW, Sun W, Zinman B, Brunzell JD, White NH, Danis RP, Davis MD, Hainsworth D, Hubbard LD, Nathan DM. Long-term renal outcomes of patients with type 1 diabetes mellitus and microalbuminuria: an analysis of the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications cohort. Arch Intern Med. 2011;171(5):412-20. Epub 2011/03/16. doi: 10.1001/archinternmed.2011.16. PubMed PMID: 21403038; PMCID: PMC3085024.
29. de Boer IH, Rue TC, Hall YN, Heagerty PJ, Weiss NS, Himmelfarb J. Temporal trends in the prevalence of diabetic kidney disease in the United States. Jama. 2011;305(24):2532-9. Epub 2011/06/23. doi: 10.1001/jama.2011.861. PubMed PMID: 21693741; PMCID: PMC3731378.
30. Fox CS, Matsushita K, Woodward M, Bilo HJ, Chalmers J, Heerspink HJ, Lee BJ, Perkins RM, Rossing P, Sairenchi T, Tonelli M, Vassalotti JA, Yamagishi K, Coresh J, de Jong PE, Wen CP, Nelson RG. Associations of kidney disease measures with mortality and end-stage renal disease in individuals with and without diabetes: a meta-analysis. Lancet. 2012;380(9854):1662-73. Epub 2012/09/28. doi: 10.1016/s0140-6736(12)61350-6. PubMed PMID: 23013602; PMCID: PMC3771350.
31. Levey AS, Cattran D, Friedman A, Miller WG, Sedor J, Tuttle K, Kasiske B, Hostetter T. Proteinuria as a surrogate outcome in CKD: report of a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. Am J Kidney Dis. 2009;54(2):205-26. Epub 2009/07/07. doi: 10.1053/j.ajkd.2009.04.029. PubMed PMID: 19577347.
32. Lemley KV, Boothroyd DB, Blouch KL, Nelson RG, Jones LI, Olshen RA, Myers BD. Modeling GFR trajectories in diabetic nephropathy. Am J Physiol Renal Physiol. 2005;289(4):F863-70. Epub 2005/05/19. doi: 10.1152/ajprenal.00068.2004. PubMed PMID: 15900022.

33. Kramer HJ, Nguyen QD, Curhan G, Hsu CY. Renal insufficiency in the absence of albuminuria and retinopathy among adults with type 2 diabetes mellitus. Jama. 2003;289(24):3273-7. Epub 2003/06/26. doi: 10.1001/jama.289.24.3273. PubMed PMID: 12824208.
34. Lemley KV, Abdullah I, Myers BD, Meyer TW, Blouch K, Smith WE, Bennett PH, Nelson RG. Evolution of incipient nephropathy in type 2 diabetes mellitus. Kidney Int. 2000;58(3):1228-37. Epub 2000/09/06. doi: 10.1046/j.1523-1755.2000.00223.x. PubMed PMID: 10972685.
35. Pavkov ME, Knowler WC, Hanson RL, Nelson RG. Diabetic nephropathy in American Indians, with a special emphasis on the Pima Indians. Curr Diab Rep. 2008;8(6):486-93. Epub 2008/11/08. doi: 10.1007/s11892-008-0083-1. PubMed PMID: 18990306; PMCID: PMC3480511.
36. Pavkov ME, Mason CC, Bennett PH, Curtis JM, Knowler WC, Nelson RG. Change in the distribution of albuminuria according to estimated glomerular filtration rate in Pima Indians with type 2 diabetes. Diabetes Care. 2009;32(10):1845-50. Epub 2009/07/14. doi: 10.2337/dc08-2325. PubMed PMID: 19592626; PMCID: PMC2752932.
37. Macisaac RJ, Jerums G. Diabetic kidney disease with and without albuminuria. Curr Opin Nephrol Hypertens. 2011;20(3):246-57. Epub 2011/03/23. doi: 10.1097/MNH.0b013e3283456546. PubMed PMID: 21422923.
38. Caramori ML, Parks A, Mauer M. Renal lesions predict progression of diabetic nephropathy in type 1 diabetes. J Am Soc Nephrol. 2013;24(7):1175-81. Epub 2013/05/21. doi: 10.1681/asn.2012070739. PubMed PMID: 23687360; PMCID: PMC3699823.
39. Gonzalez Suarez ML, Thomas DB, Barisoni L, Fornoni A. Diabetic nephropathy: Is it time yet for routine kidney biopsy? World J Diabetes. 2013;4(6):245-55. Epub 2014/01/01. doi: 10.4239/wjd.v4.i6.245. PubMed PMID: 24379914; PMCID: PMC3874483.
40. Fufaa GD, Weil EJ, Lemley KV, Knowler WC, Brosius FC, 3rd, Yee B, Mauer M, Nelson RG. Structural Predictors of Loss of Renal Function in American Indians with Type 2 Diabetes. Clin J Am Soc Nephrol. 2016;11(2):254-61. Epub 2016/01/23. doi: 10.2215/cjn.05760515. PubMed PMID: 26792530; PMCID: PMC4741038.
41. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. Methods Inf Med. 2001;40(4):346-58. Epub 2001/09/13. PubMed PMID: 11552348.
42. Can T. Introduction to bioinformatics. Methods Mol Biol. 2014;1107:51-71. Epub 2013/11/26. doi: 10.1007/978-1-62703-748-8_4. PubMed PMID: 24272431.

43. Kesić S. Systems biology, emergence and antireductionism. Saudi J Biol Sci. 2016;23(5):584-91. Epub 2016/09/01. doi: 10.1016/j.sjbs.2015.06.015. PubMed PMID: 27579007; PMCID: PMC4992115.
44. Mulder S, Hamidi H, Kretzler M, Ju W. An integrative systems biology approach for precision medicine in diabetic kidney disease. Diabetes Obes Metab. 2018;20 Suppl 3(Suppl 3):6-13. Epub 2018/10/09. doi: 10.1111/dom.13416. PubMed PMID: 30294956; PMCID: PMC6541014.
45. Schena FP, Nistor I, Curci C. Transcriptomics in kidney biopsy is an untapped resource for precision therapy in nephrology: a systematic review. Nephrol Dial Transplant. 2018;33(7):1094-102. Epub 2017/10/11. doi: 10.1093/ndt/gfx211. PubMed PMID: 28992289.
46. Kretzler M, Sedor JR. Introduction: Precision Medicine for Glomerular Disease: The Road Forward.SeminNephrol.2015;35(3):209-11.Epub2015/07/29.doi: 10.1016/j.semnephrol.2015.04.001. PubMed PMID: 26215858; PMCID: PMC5902811.
47. Sharma K, Paša-Tolić L. Toward individual glomerular phenotyping: advent of precision medicine in kidney biopsies. Kidney Int. 2018;93(6):1265-7. Epub 2018/05/25. doi: 10.1016/j.kint.2018.01.035. PubMed PMID: 29792268; PMCID: PMC6583000.
48. Krzyszczyk P, Acevedo A, Davidoff EJ, Timmins LM, Marrero-Berrios I, Patel M, White C, Lowe C, Sherba JJ, Hartmanshenn C, O'Neill KM, Balter ML, Fritz ZR, Androulakis IP, Schloss RS, Yarmush ML. The growing role of precision and personalized medicine for cancer treatment. Technology (Singap World Sci). 2018;6(3-4):79-100. Epub 2019/02/05. doi: 10.1142/s2339547818300020. PubMed PMID: 30713991; PMCID: PMC6352312.
49. Altemtam N, Russell J, El Nahas M. A study of the natural history of diabetic kidney disease (DKD). Nephrol Dial Transplant. 2012;27(5):1847-54. Epub 2011/11/08. doi: 10.1093/ndt/gfr561. PubMed PMID: 22058177.
50. Hunter P. The paradox of model organisms. The use of model organisms in research will continue despite their shortcomings. EMBO Rep. 2008;9(8):717-20. Epub 2008/08/02. doi: 10.1038/embor.2008.142. PubMed PMID: 18670440; PMCID: PMC2515201.
51. Rosenthal E. British rethinking test rules after drug trial nearly kills 6. N Y Times Web. 2006:A1, a6. Epub 2006/04/29. PubMed PMID: 16642579.
52. Tian Y, Patel J. TALE: A Tool for Approximate Large Graph Matching. 2008 IEEE 24th International Conference on Data Engineering. 2008:963-72.

53. Hodgin JB, Nair V, Zhang H, Randolph A, Harris RC, Nelson RG, Weil EJ, Cavalcoli JD, Patel JM, Brosius FC, 3rd, Kretzler M. Identification of cross-species shared transcriptional networks of diabetic nephropathy in human and mouse glomeruli. Diabetes. 2013;62(1):299-308. Epub 2012/11/10. doi: 10.2337/db11-1667. PubMed PMID: 23139354; PMCID: PMC3526018.
54. Berthier CC, Bethunaickan R, Gonzalez-Rivera T, Nair V, Ramanujam M, Zhang W, Bottinger EP, Segerer S, Lindenmeyer M, Cohen CD, Davidson A, Kretzler M. Cross-species transcriptional network analysis defines shared inflammatory responses in murine and human lupus nephritis. J Immunol. 2012;189(2):988-1001. Epub 2012/06/23. doi: 10.4049/jimmunol.1103031. PubMed PMID: 22723521; PMCID: PMC3392438.
55. Kim KT, Lee HW, Lee HO, Kim SC, Seo YJ, Chung W, Eum HH, Nam DH, Kim J, Joo KM, Park WY. Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. Genome Biol. 2015;16(1):127. Epub 2015/06/19. doi: 10.1186/s13059-015-0692-3. PubMed PMID: 26084335; PMCID: PMC4506401.
56. Moignard V, Woodhouse S, Haghverdi L, Lilly AJ, Tanaka Y, Wilkinson AC, Buettner F, Macaulay IC, Jawaid W, Diamanti E, Nishikawa SI, Piterman N, Kouskoff V, Theis FJ, Fisher J, Göttgens B. Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat Biotechnol. 2015;33(3):269-76. Epub 2015/02/11. doi: 10.1038/nbt.3154. PubMed PMID: 25664528; PMCID: PMC4374163.
57. Ocone A, Haghverdi L, Mueller NS, Theis FJ. Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. Bioinformatics. 2015;31(12):i89-96. Epub 2015/06/15. doi: 10.1093/bioinformatics/btv257. PubMed PMID: 26072513; PMCID: PMC4765871.
58. Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell. 2012;150(6):1209-22. Epub 2012/09/18. doi: 10.1016/j.cell.2012.08.023. PubMed PMID: 22980981; PMCID: PMC3457656.
59. Lim SB, Di Lee W, Vasudevan J, Lim W-T, Lim CT. Liquid biopsy: one cell at a time. npj Precision Oncology. 2019;3(1):23. doi: 10.1038/s41698-019-0095-0.
60. Qiu C, Huang S, Park J, Park Y, Ko YA, Seasock MJ, Bryer JS, Xu XX, Song WC, Palmer M, Hill J, Guarnieri P, Hawkins J, Boustany-Kari CM, Pullen SS, Brown CD, Susztak K. Renal compartment-specific genetic variation analyses identify new pathways in chronic

kidney disease. Nat Med. 2018;24(11):1721-31. Epub 2018/10/03. doi: 10.1038/s41591-018-0194-4. PubMed PMID: 30275566; PMCID: PMC6301011.

61. Gillies CE, Putler R, Menon R, Otto E, Yasutake K, Nair V, Hoover P, Lieb D, Li S, Eddy S, Fermin D, McNulty MT, Hacohen N, Kiryluk K, Kretzler M, Wen X, Sampson MG. An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. Am J Hum Genet. 2018;103(2):232-44. Epub 2018/07/31. doi: 10.1016/j.ajhg.2018.07.004. PubMed PMID: 30057032; PMCID: PMC6081280.
62. Nelson RG, Bennett PH, Beck GJ, Tan M, Knowler WC, Mitch WE, Hirschman GH, Myers BD. Development and progression of renal disease in Pima Indians with non-insulin-dependent diabetes mellitus. Diabetic Renal Disease Study Group. N Engl J Med. 1996;335(22):1636-42. Epub 1996/11/28. doi: 10.1056/nejm199611283352203. PubMed PMID: 8929360.
63. Weil EJ, Fufaa G, Jones LI, Lovato T, Lemley KV, Hanson RL, Knowler WC, Bennett PH, Yee B, Myers BD, Nelson RG. Effect of losartan on prevention and progression of early diabetic nephropathy in American Indians with type 2 diabetes. Diabetes. 2013;62(9):3224-31. Epub 2013/04/03. doi: 10.2337/db12-1512. PubMed PMID: 23545707; PMCID: PMC3749332.
64. Lemley KV, Blouch K, Abdullah I, Boothroyd DB, Bennett PH, Myers BD, Nelson RG. Glomerular permselectivity at the onset of nephropathy in type 2 diabetes mellitus. J Am Soc Nephrol. 2000;11(11):2095-105. Epub 2000/10/29. PubMed PMID: 11053486.
65. Nair V, Komorowsky CV, Weil EJ, Yee B, Hodgin J, Harder JL, Godfrey B, Ju W, Boustany-Kari CM, Schwarz M, Lemley KV, Nelson PJ, Nelson RG, Kretzler M. A molecular morphometric approach to diabetic kidney disease can link structure to function and outcome. Kidney Int. 2018;93(2):439-49. Epub 2017/10/22. doi: 10.1016/j.kint.2017.08.013. PubMed PMID: 29054530; PMCID: PMC5794609.
66. Yasuda Y, Cohen CD, Henger A, Kretzler M. Gene expression profiling analysis in nephrology: towards molecular definition of renal disease. Clin Exp Nephrol. 2006;10(2):91-8. Epub 2006/06/23. doi: 10.1007/s10157-006-0421-z. PubMed PMID: 16791393.
67. Ju W, Nair V, Smith S, Zhu L, Shedden K, Song PXK, Mariani LH, Eichinger FH, Berthier CC, Randolph A, Lai JY, Zhou Y, Hawkins JJ, Bitzer M, Sampson MG, Thier M, Solier C, Duran-Pacheco GC, Duchateau-Nguyen G, Essioux L, Schott B, Formentini I, Magnone MC, Bobadilla M, Cohen CD, Bagnasco SM, Barisoni L, Lv J, Zhang H, Wang HY, Brosius FC, Gadegbeku CA, Kretzler M. Tissue transcriptome-driven identification of

epidermal growth factor as a chronic kidney disease biomarker. Sci Transl Med. 2015;7(316):316ra193. Epub 2015/12/04. doi: 10.1126/scitranslmed.aac7071. PubMed PMID: 26631632; PMCID: PMC4861144.

68. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2006;8(1):118-27. doi: 10.1093/biostatistics/kxj037.
69. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. Epub 2008/12/31. doi: 10.1186/1471-2105-9-559. PubMed PMID: 19114008; PMCID: PMC2631488.
70. Arazi A, Rao DA, Berthier CC, Davidson A, Liu Y, Hoover PJ, Chicoine A, Eisenhaure TM, Jonsson AH, Li S, Lieb DJ, Zhang F, Slowikowski K, Browne EP, Noma A, Sutherby D, Steelman S, Smilek DE, Tosta P, Apruzzese W, Massarotti E, Dall'Era M, Park M, Kamen DL, Furie RA, Payan-Schober F, Pendergraft WF, 3rd, McInnis EA, Buyon JP, Petri MA, Putterman C, Kalunian KC, Woodle ES, Lederer JA, Hildeman DA, Nusbaum C, Raychaudhuri S, Kretzler M, Anolik JH, Brenner MB, Wofsy D, Hacohen N, Diamond B. The immune cell landscape in kidneys of patients with lupus nephritis. Nat Immunol. 2019;20(7):902-14. Epub 2019/06/19. doi: 10.1038/s41590-019-0398-x. PubMed PMID: 31209404; PMCID: PMC6726437.
71. Menon R, Otto EA, Hoover P, Eddy S, Mariani L, Godfrey B, Berthier CC, Eichinger F, Subramanian L, Harder J, Ju W, Nair V, Larkina M, Naik AS, Luo J, Jain S, Sealfon R, Troyanskaya O, Hacohen N, Hodgin JB, Kretzler M, Kpmp K. Single cell transcriptomics identifies focal segmental glomerulosclerosis remission endothelial biomarker. JCI Insight. 2020;5(6). Epub 2020/02/29. doi: 10.1172/jci.insight.133267. PubMed PMID: 32107344; PMCID: PMC7213795.
72. Menon R, Otto EA, Sealfon R, Nair V, Wong AK, Theesfeld CL, Chen X, Wang Y, Boppana AS, Luo J, Yang Y, Kasson PM, Schaub JA, Berthier CC, Eddy S, Lienczewski CC, Godfrey B, Dagenais SL, Sohaney R, Hartman J, Fermin D, Subramanian L, Looker HC, Harder JL, Mariani LH, Hodgin JB, Sexton JZ, Wobus CE, Naik AS, Nelson RG, Troyanskaya OG, Kretzler M. SARS-CoV-2 receptor networks in diabetic and COVID-19-associated kidney disease. Kidney Int. 2020;98(6):1502-18. Epub 2020/10/11. doi: 10.1016/j.kint.2020.09.015. PubMed PMID: 33038424; PMCID: PMC7543950.
73. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018. Epub 2018/12/12. doi: 10.1038/nbt.4314. PubMed PMID: 30531897.

74. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh PR, Raychaudhuri S. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289-96. Epub 2019/11/20. doi: 10.1038/s41592-019-0619-0. PubMed PMID: 31740819; PMCID: PMC6884693.
75. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. Cell. 2019;177(7):1888-902.e21. Epub 2019/06/11. doi: 10.1016/j.cell.2019.05.031. PubMed PMID: 31178118; PMCID: PMC6687398.
76. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biology. 2019;20(1):296. doi: 10.1186/s13059-019-1874-1.
77. Krämer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics. 2014;30(4):523-30. Epub 2013/12/18. doi: 10.1093/bioinformatics/btt703. PubMed PMID: 24336805; PMCID: PMC3928520.
78. Tonneijck L, Muskiet MH, Smits MM, van Bommel EJ, Heerspink HJ, van Raalte DH, Joles JA. Glomerular Hyperfiltration in Diabetes: Mechanisms, Clinical Significance, and Treatment. J Am Soc Nephrol. 2017;28(4):1023-39. Epub 2017/02/02. doi: 10.1681/asn.2016060666. PubMed PMID: 28143897; PMCID: PMC5373460.
79. Pavkov ME, Bennett PH, Knowler WC, Krakoff J, Sievers ML, Nelson RG. Effect of youth-onset type 2 diabetes mellitus on incidence of end-stage renal disease and mortality in young and middle-aged Pima Indians. Jama. 2006;296(4):421-6. Epub 2006/07/27. doi: 10.1001/jama.296.4.421. PubMed PMID: 16868300.
80. Schmid H, Boucherot A, Yasuda Y, Henger A, Brunner B, Eichinger F, Nitsche A, Kiss E, Bleich M, Gröne HJ, Nelson PJ, Schlöndorff D, Cohen CD, Kretzler M. Modular activation of nuclear factor-kappaB transcriptional programs in human diabetic nephropathy. Diabetes. 2006;55(11):2993-3003. Epub 2006/10/27. doi: 10.2337/db06-0477. PubMed PMID: 17065335.
81. Berthier CC, Zhang H, Schin M, Henger A, Nelson RG, Yee B, Boucherot A, Neusser MA, Cohen CD, Carter-Su C, Argetsinger LS, Rastaldi MP, Brosius FC, Kretzler M. Enhanced expression of Janus kinase-signal transducer and activator of transcription pathway members in human diabetic nephropathy. Diabetes. 2009;58(2):469-77. Epub 2008/11/20. doi: 10.2337/db08-1328. PubMed PMID: 19017763; PMCID: PMC2628622.

82. Woroniecka KI, Park AS, Mohtat D, Thomas DB, Pullman JM, Susztak K. Transcriptome analysis of human diabetic kidney disease. Diabetes. 2011;60(9):2354-69. Epub 2011/07/15. doi: 10.2337/db10-1181. PubMed PMID: 21752957; PMCID: PMC3161334.
83. Ortiz-Muñoz G, Lopez-Parra V, Lopez-Franco O, Fernandez-Vizarra P, Mallavia B, Flores C, Sanz A, Blanco J, Mezzano S, Ortiz A, Egido J, Gomez-Guerrero C. Suppressors of cytokine signaling abrogate diabetic nephropathy. J Am Soc Nephrol. 2010;21(5):763-72. Epub 2010/02/27. doi: 10.1681/asn.2009060625. PubMed PMID: 20185635; PMCID: PMC2865742.
84. Navarro-González JF, Mora-Fernández C, Muros de Fuentes M, García-Pérez J. Inflammatory molecules and pathways in the pathogenesis of diabetic nephropathy. Nat Rev Nephrol. 2011;7(6):327-40. Epub 2011/05/04. doi: 10.1038/nrneph.2011.51. PubMed PMID: 21537349.
85. Brosius FC, Tuttle KR, Kretzler M. JAK inhibition in the treatment of diabetic kidney disease. Diabetologia. 2016;59(8):1624-7. Epub 2016/06/24. doi: 10.1007/s00125-016-4021-5. PubMed PMID: 27333885; PMCID: PMC4942738.
86. Tuttle KR, Brosius FC, III, Adler SG, Kretzler M, Mehta RL, Tumlin JA, Tanaka Y, Haneda M, Liu J, Silk ME, Cardillo TE, Duffin KL, Haas JV, Macias WL, Nunes FP, Janes JM. JAK1/JAK2 inhibition by baricitinib in diabetic kidney disease: results from a Phase 2 randomized controlled clinical trial. Nephrology Dialysis Transplantation. 2018;33(11):1950-9. doi: 10.1093/ndt/gfx377.
87. Forbes JM, Thorburn DR. Mitochondrial dysfunction in diabetic kidney disease. Nature Reviews Nephrology. 2018;14(5):291-312. doi: 10.1038/nrneph.2018.9.
88. Zhan M, Usman IM, Sun L, Kanwar YS. Disruption of renal tubular mitochondrial quality control by Myo-inositol oxygenase in diabetic kidney disease. J Am Soc Nephrol. 2015;26(6):1304-21. Epub 2014/10/02. doi: 10.1681/asn.2014050457. PubMed PMID: 25270067; PMCID: PMC4446875.
89. Alimadadi A, Aryal S, Manandhar I, Joe B, Cheng X. Identification of Upstream Transcriptional Regulators of Ischemic Cardiomyopathy Using Cardiac RNA-Seq Meta-Analysis. Int J Mol Sci. 2020;21(10). Epub 2020/05/20. doi: 10.3390/ijms21103472. PubMed PMID: 32423033; PMCID: PMC7278960.
90. Araujo RP, Petricoin EF, Liotta LA. A mathematical model of combination therapy using the EGFR signaling network. Biosystems. 2005;80(1):57-69. Epub 2005/03/03. doi: 10.1016/j.biosystems.2004.10.002. PubMed PMID: 15740835.

91. Jiang XW, Lu HY, Xu Z, Liu TY, Wu Q, Yang Y, Zhao QC, Gao HY. In Silico Analyses for Key Genes and Molecular Genetic Mechanism in Epilepsy and Alzheimer's Disease. CNS Neurol Disord Drug Targets. 2018;17(8):608-17. Epub 2018/07/27. doi: 10.2174/1871527317666180724150839. PubMed PMID: 30047339.
92. Lorenz G, Darisipudi MN, Anders HJ. Canonical and non-canonical effects of the NLRP3 inflammasome in kidney inflammation and fibrosis. Nephrol Dial Transplant. 2014;29(1):41-8. Epub 2013/09/13. doi: 10.1093/ndt/gft332. PubMed PMID: 24026244.
93. Wada J, Makino H. Innate immunity in diabetes and diabetic nephropathy. Nat Rev Nephrol. 2016;12(1):13-26. Epub 2015/11/17. doi: 10.1038/nrneph.2015.175. PubMed PMID: 26568190.
94. Anders HJ. Of Inflammasomes and Alarmins: IL-1β and IL-1α in Kidney Disease. J Am Soc Nephrol. 2016;27(9):2564-75. Epub 2016/08/16. doi: 10.1681/asn.2016020177. PubMed PMID: 27516236; PMCID: PMC5004665.
95. Ying Y, Kim J, Westphal SN, Long KE, Padanilam BJ. Targeted deletion of p53 in the proximal tubule prevents ischemic renal injury. J Am Soc Nephrol. 2014;25(12):2707-16. Epub 2014/05/24. doi: 10.1681/asn.2013121270. PubMed PMID: 24854277; PMCID: PMC4243356.
96. Neuhofer W. Role of NFAT5 in inflammatory disorders associated with osmotic stress. Curr Genomics. 2010;11(8):584-90. Epub 2011/06/02. doi: 10.2174/138920210793360961. PubMed PMID: 21629436; PMCID: PMC3078683.
97. Li Q, Zeng Y, Jiang Q, Wu C, Zhou J. Role of mTOR signaling in the regulation of high glucose-induced podocyte injury. Exp Ther Med. 2019;17(4):2495-502. Epub 2019/03/25. doi: 10.3892/etm.2019.7236. PubMed PMID: 30906437; PMCID: PMC6425130.
98. Inoki K, Mori H, Wang J, Suzuki T, Hong S, Yoshida S, Blattner SM, Ikenoue T, Rüegg MA, Hall MN, Kwiatkowski DJ, Rastaldi MP, Huber TB, Kretzler M, Holzman LB, Wiggins RC, Guan KL. mTORC1 activation in podocytes is a critical step in the development of diabetic nephropathy in mice. J Clin Invest. 2011;121(6):2181-96. Epub 2011/05/25. doi: 10.1172/jci44771. PubMed PMID: 21606597; PMCID: PMC3104745.
99. Wang Y, Zhong J, Zhang X, Liu Z, Yang Y, Gong Q, Ren B. The Role of HMGB1 in the Pathogenesis of Type 2 Diabetes. J Diabetes Res. 2016;2016:2543268. Epub 2017/01/20. doi: 10.1155/2016/2543268. PubMed PMID: 28101517; PMCID: PMC5215175.
100. Klune JR, Dhupar R, Cardinal J, Billiar TR, Tsung A. HMGB1: endogenous danger signaling. Mol Med. 2008;14(7-8):476-84. Epub 2008/04/24. doi: 10.2119/2008-00034.Klune. PubMed PMID: 18431461; PMCID: PMC2323334.

101. Mantell LL, Parrish WR, Ulloa L. Hmgb-1 as a therapeutic target for infectious and inflammatory disorders. Shock. 2006;25(1):4-11. Epub 2005/12/22. doi: 10.1097/01.shk.0000188710.04777.9e. PubMed PMID: 16369179.
102. Tang X, Huang Y, Lei J, Luo H, Zhu X. The single-cell sequencing: new developments and medical applications. Cell Biosci. 2019;9:53. Epub 2019/08/09. doi: 10.1186/s13578-019-0314-y. PubMed PMID: 31391919; PMCID: PMC6595701.
103. Pan S, Zhan Y, Chen X, Wu B, Liu B. Identification of Biomarkers for Controlling Cancer Stem Cell Characteristics in Bladder Cancer by Network Analysis of Transcriptome Data Stemness Indices. Front Oncol. 2019;9:613. Epub 2019/07/25. doi: 10.3389/fonc.2019.00613. PubMed PMID: 31334127; PMCID: PMC6620567.
104. Le K, Guo H, Zhang Q, Huang X, Xu M, Huang Z, Yi P. Gene and lncRNA co-expression network analysis reveals novel ceRNA network for triple-negative breast cancer. Scientific Reports. 2019;9(1):15122. doi: 10.1038/s41598-019-51626-7.
105. Gibbs DL, Baratt A, Baric RS, Kawaoka Y, Smith RD, Orwoll ES, Katze MG, McWeeney SK. Protein co-expression network analysis (ProCoNA). J Clin Bioinforma. 2013;3(1):11. Epub 2013/06/04. doi: 10.1186/2043-9113-3-11. PubMed PMID: 23724967; PMCID: PMC3695838.
106. Ding M, Li F, Wang B, Chi G, Liu H. A comprehensive analysis of WGCNA and serum metabolomics manifests the lung cancer-associated disordered glucose metabolism. J Cell Biochem. 2019;120(6):10855-63. Epub 2019/02/21. doi: 10.1002/jcb.28377. PubMed PMID: 30784104.

Acknowledgements

I cannot end my thesis without expressing my gratitude to the wonderful group of people around me. Today I feel proud and solaced myself as I fulfill my promise to my mother who sadly is not in this world now to see me in this podium but showering her blessing from "there"! She must be feeling proud on her daughter's great achievement.

But this would not be possible without the support and motivation of the great team of scientists I worked with. Foremost, I would like to express my sincere gratitude to my advisor Prof. Peter Nelson for taking me in as his Dr. hum biol mentee and supporting my Dr. hum biol research. His guidance, motivation, enthusiasm, and immense knowledge helped me throughout my research and writing of this thesis. Peter along with the lab colleagues here at LMU greatly supported my Dr. hum biol journey.

I would like to pay my special regards to Prof. Dr. Matthias Kretzler currently Professor, University of Michigan and former LMU researcher for providing me the opportunity to conduct my research at University of Michigan and his brilliant bioinformatic guidance. I extend my thanks to the wonderful scholastic friendly team there as well.

The whole experience provided a holistic approach to my scientific and professional growth and personal outlook which I'm sure to stay with me forever. I consider myself to be truly blessed to get this great opportunity to enhance my scientific knowledge and career.

Finally, a very special thanks to my family where words come short to express my love and gratitude. The thank you note starts with my husband who motivated and believed in me that I can do this even when I was in doubt. My sincere appreciation and gratitude to my father, grandmother, brother and of course my son who always encouraged and supported me. A big thank you for all your love and support!

Affidavit

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN | Dean's Office Faculty of Medicine | MMRS

Affidavit

Viji Nair
Surname, first name

7970 Oxford Drive
Street

48187, Canton
Zip code, town

USA
Country

I hereby declare, that the submitted thesis entitled

Identification and characterization of regulatory pathways and prognostic markers linked to the initiation and progression of diabetic kidney disease using systems biology

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

Ann Arbor, 20.August 2021
Place, date

Viji Nair
Signature doctoral candidate