

Determining the risk profile for chronic kidney disease (CKD) in rural South Africans using genetic risk scores and protein markers

by

Melanie Ann Govender



Supervisors: Prof. Michèle Ramsay, Dr. June Fabian, Dr. Jean-Tristan Brandenburg

A Thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) in Human Genetics.

Johannesburg, 2024.



DECLARATION

I, Melanie Ann Govender, student number 1487683, declare that this PhD thesis is my own work, and that work that has been performed by others are mentioned in the acknowledgments. It is being submitted for the degree of PhD in Human Genetics at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University (See Appendices A.1 Plagiarism report).

I confirm that I performed the scoping review, conducted the relevant genomic and proteomic analyses and data interpretation, wrote the first draft of the papers, and incorporated comments and feedback, with direction from my supervisors. All co-authors agreed for content to be used in this thesis (See Appendix A.2 Agreement of co-authors).

A handwritten signature in black ink, appearing to read 'M. Govender', with a long horizontal flourish extending to the left.

Melanie Govender

Date: 19 August 2024

DEDICATION

This is dedicated to my parents – your sacrifices, selflessness, and belief in me have encouraged me to pursue my dreams.

OUTPUTS FROM THE PHD

1. PUBLICATIONS

The first page of article outputs, including abstracts, can be seen in Appendix A.3 PhD Article Outputs.

1. The use of 'omics for diagnosing and predicting progression of chronic kidney disease: A scoping review

Status: Published in *Frontiers in Genetics* (8 Nov 2021)

Details: Govender MA, Brandenburg JT, Fabian J, Ramsay M. The use of 'omics for diagnosing and predicting progression of chronic kidney disease: A scoping review.

Frontiers in Genetics. 2021;12:682929. [DOI: 10.3389/fgene.2021.682929](https://doi.org/10.3389/fgene.2021.682929)

2. Proteomic insights into the pathophysiology of hypertension-associated albuminuria: Pilot study in a South African cohort

Status: Published in *Clinical Proteomics* (24 Feb 2024)

Details: Govender MA, Stoychev SH, Brandenburg JT, Ramsay M, Fabian J, Govender IS. Proteomic insights into the pathophysiology of hypertension-associated albuminuria: Pilot study in a South African cohort. *Clinical Proteomics*. 2024;21(1):1-2. [DOI:](https://doi.org/10.1186/s12014-024-09458-9)

[10.1186/s12014-024-09458-9](https://doi.org/10.1186/s12014-024-09458-9)

2. OTHER PUBLICATIONS DURING THE PHD

Yusuf AA, **Govender MA**, Brandenburg JT, Winkler CA. Kidney disease and *APOL1*. *Human Molecular Genetics*. 2021 Mar 1;30(R1):R129-37.

Brandenburg JT, **Govender MA**, Winkler CA, Boua PR, Agongo G, Fabian J, Ramsay M. Apolipoprotein L1 High-Risk Genotypes and Albuminuria in sub-Saharan African Populations. *Clinical Journal of the American Society of Nephrology*. 2022 Jun 1;17(6):798-808.

Brandenburg JT, Chen WC, Boua PR, **Govender MA**, Agongo G, Micklesfield LK, Sorgho H, Tollman S, Asiki G, Mashinya F, Hazelhurst S. Genetic Association and Transferability for Urinary Albumin-Creatinine Ratio as a Marker of Kidney Disease in four sub-Saharan African Populations and non-continental Individuals of African Ancestry. *Frontiers in Genetics*. 2024;5:1372042.

ABSTRACT

Chronic kidney disease (CKD) is a global public health concern, with disproportionate morbidity and mortality in low- and middle-income regions, including Sub-Saharan Africa. Recent advancements in multi-omics approaches have explored disease risk indicators and contributed to the understanding of the pathophysiology of CKD in high-income populations. The overall aim of this research was to assess and understand CKD in a Sub-Saharan population using genetic risk models for kidney disease and to evaluate the proteomic profile of individuals with hypertension-associated albuminuria, with a view to detecting indicators of CKD and disease progression in a Sub-Saharan cohort.

The first objective was a scoping review that was undertaken to evaluate existing literature for potential biomarkers for CKD and to identify gaps in literature. Key literature gaps identified included the lack of studies that focus on HT in the context of kidney disease and only one study on African individuals residing in Africa. In this work, two research studies were developed based on existing data from the African Research Kidney Disease (ARK) study, a well characterised population-based cohort study of black individuals from Agincourt in the rural Mpumalanga Province, South Africa. The second objective was a genomics study which aimed to examine the potential of using summary statistics from three discovery datasets to assess the predictive accuracy of polygenic scores (PGSs) for CKD and kidney function markers. Limited transferability was observed, explaining <1% of the variability in kidney disease markers in this African cohort. A PGS model derived from the transethnic cohort for estimated glomerular filtration rate (eGFR) explained the highest variability (0.8%) in African individuals and was significantly associated with HT ($P<0.001$), diabetes ($P=0.007$), and HIV ($P=0.001$). The third objective was a proteomics study which aimed to compare proteomic profiles of cases with both HT and albuminuria to controls (neither condition) to identify proteins and pathways involved in hypertension-associated albuminuria. Pathways including immune system ($q=1.4\times 10^{-45}$) and innate immune system ($q=1.1\times 10^{-32}$) were linked with hypertension-associated albuminuria. Proteins including angiotensinogen, apolipoprotein L1, and uromodulin had the highest disease scores (76–100% confidence). A machine learning approach was able to identify a set of 20 proteins that contributed to classifying disease status (ie, hypertension-associated albuminuria).

The assessment of PGSs for kidney function markers contribute to understanding of CKD genetic risk prediction in Africa, while the proteomics research added new knowledge to understanding the role of proteins and pathways involved in hypertension-associated albuminuria in Africa. This research addresses the gap of a lack of 'omics research in resident African populations. It also contributes to the understanding of risk prediction for CKD and identifies potential proteomic markers for hypertension-associated albuminuria that may inform the development of personalised treatment strategies in Africa.

Key words: ARK, African genomics, chronic kidney disease, polygenic scores, hypertension, machine learning.

ACKNOWLEDGEMENTS

I would like to acknowledge the following people:

My supervisors Prof. Michèle Ramsay, Dr June Fabian, and Dr Jean-Tristan Brandenburg, for offering their expertise and supervision during this journey, and for going above and beyond to give me the support I needed.

Dr Ireshyn Govender and Dr Stoyan Stoychev for allowing me to conduct the proteomics research at their lab, and for providing valuable expertise and feedback.

Dr Cheryl Winkler from the National Cancer Institute, USA for the genotyping.

The members of the Agincourt community who volunteered to participate in the ARK study, and without whom we would not have been able to conduct this research.

The National Research Foundation (NRF) for financial support.

My family and friends for their love and support, and for continuously motivating me throughout this journey.

TABLE OF CONTENTS

DECLARATION.....	I
DEDICATION.....	II
OUTPUTS FROM THE PHD	III
1. PUBLICATIONS.....	III
2. OTHER PUBLICATIONS DURING THE PHD	IV
ABSTRACT	V
ACKNOWLEDGEMENTS	VII
TABLE OF CONTENTS	VIII
FIGURES AND TABLES.....	XI
1. FIGURES.....	XI
2. TABLES	XIII
LIST OF ABBREVIATIONS	XV
CHAPTER 1	1
1. BACKGROUND AND PURPOSE OF STUDY.....	1
1.1. Epidemiology of CKD and HT.....	1
1.2. Causes of CKD.....	4
1.3. Guidelines for defining CKD and HT	8
1.4. The complex pathophysiology and interplay between CKD and hypertension.....	12
1.5. The importance of omics' studies in kidney disease	17
1.6. Genomic approaches and examples.....	19
1.7. Polygenic scores (PGSs) and related steps	21
1.8. Proteomic approaches and examples	25
1.9. The African Research on Kidney Disease (ARK) Study.....	29
1.10. The Africa Wits-INDEPTH partnership for Genomic Studies (AWI-Gen)	30

1.11.	Study rationale, aims and objectives	31
1.12.	Thesis outline	32
1.13.	Ethical Considerations.....	32
CHAPTER 2		34
2. THE USE OF 'OMICS FOR DIAGNOSING AND PREDICTING PROGRESSION OF CHRONIC KIDNEY DISEASE: A SCOPING REVIEW		34
2.1.	Abstract	34
2.2.	Introduction	35
2.3.	Methods	36
2.4.	Results	39
2.5.	Discussion	48
2.6.	Author Contributions.....	54
2.7.	Funding.....	54
CHAPTER 3		55
3. POLYGENIC SCORES FOR KIDNEY FUNCTION IN A SOUTH AFRICAN COHORT		55
3.1.	Abstract	55
3.2.	Introduction	55
3.3.	Methods.....	57
3.4.	Results	66
3.5.	Discussion	74
3.6.	Conclusions	76
CHAPTER 4		77
4. PROTEOMIC INSIGHTS INTO THE PATHOPHYSIOLOGY OF HYPERTENSION- ASSOCIATED ALBUMINURIA: PILOT STUDY IN A SOUTH AFRICAN COHORT. 77		
4.1.	Abstract	77
4.2.	Introduction	78

4.3.	Methods	80
4.4.	Results	85
4.5.	Discussion	93
4.6.	Conclusions	96
4.7.	Funding declaration	96
4.8.	Declarations	97
4.9.	Author contributions.....	97
4.10.	Data availability.....	97
4.11.	Supplementary material.....	98
CHAPTER 5		112
5. INTEGRATED DISCUSSION AND CONCLUSIONS		112
5.1.	Study overview.....	112
5.2.	Implications from the genomics study.....	116
5.3.	Implications from the proteomics study	117
5.4.	Genetic and proteomic associations with CKD risk, and use of multi-omics approaches.....	118
5.5.	Strengths and limitations	119
5.6.	Future considerations.....	123
5.7.	Conclusions	125
REFERENCES		127
APPENDICES		152
APPENDIX A		152
A.1.	Plagiarism report	152
A.2.	Agreement of co-authors	155
A.3.	PhD Article Outputs	158
A.4.	Ethics Approval Certificate	160
APPENDIX B.....		161

B.1.	List of 123 full-text papers assessed and included in the review	161
------	--	-----

FIGURES AND TABLES

1. FIGURES

CHAPTER 1

Figure 1.1: Prevalence of CKD in the general population in Africa by region.....	3
Figure 1.2: KDIGO 2012: CKD classification and prognosis by GFR and albuminuria categories	10
Figure 1.3: Schematic presentation of cellular events involved in CKD	14
Figure 1.4: Mechanisms involved in the pathophysiology of hypertension-associated kidney disease.....	16

CHAPTER 2

Figure 2.1: Article selection process according to PRISMA-ScR guidelines	38
Figure 2.2: Bar graph showing number of papers with each of the comorbidities associated with CKD	42

CHAPTER 3

Figure 3.1: Workflow summary for datasets and analyses.....	57
Figure 3.2: Distribution of kidney function markers in the ARK genetic sub-study.....	67
Figure 3.3: Proportion of participants with markers of kidney function	68
Figure 3.4: Transferability of polygenic scores generated from summary statistics of external cohort studies	70
Figure 3.5: Odds ratios for eGFR levels across PGS deciles	71

Figure 3.6: Variance in eGFR explained by PGS in the presence or absence of CKD risk factors (HT, diabetes, and HIV).....	72
--	----

CHAPTER 4

Figure 4.1: Flow diagram for this case-control study	80
Figure 4.2: Summary of quantified proteins.....	86
Figure 4.3: PC (1 and 2) analysis based on protein abundance features.....	87
Figure 4.4: Differential protein abundances between cases and controls illustrated in a volcano plot	88
Figure 4.5: Top 20 features identified from the classifier using the XGBoost algorithm	89
Figure 4.6: Common proteins between this case-control pilot study and other studies on CKD and hypertension, as imported from public databases).....	91
Supplementary Figure 4.1: Project specific system suitability-quality control	97
Supplementary Figure 4.2: Dynamic range of proteins quantified.....	98
Supplementary Figure 4.3: Confusion matrix following machine learning using the XGBoost algorithm.....	99
Supplementary Figure 4.4: Differentially abundant proteins associated with the immune system	100
Supplementary Figure 4.5: Differentially abundant proteins associated with the innate immune system	101
Supplementary Figure 4.6: Differentially abundant proteins associated with extracellular matrix organisation (A) and activation of matrix metalloproteinases (B)....	101

2. TABLES

CHAPTER 1

Table 1.1: Global leading causes of CKD.....	4
Table 1.2: Risk factors for CKD in Sub-Saharan Africa.....	6
Table 1.3: Ranges for normal and elevated blood pressure according to various guidelines.....	11
Table 1.4: Summary of key GWASs for kidney function traits.....	20

CHAPTER 2

Table 2.1: Search strategy used in database searches	36
Table 2.2: Classification of CKD	37
Table 2.3: Participant and study characteristics	39
Table 2.4: The different ‘omics technologies and techniques reported in this scoping review for detection of CKD and associated risk in 123 studies.....	40
Table 2.5: Table 2.4: The different ‘omics technologies and techniques reported in this scoping review for detection of CKD and associated risk in 123 studies.....	45

CHAPTER 3

Table 3.1: Description of discovery GWAS cohorts	65
Table 3.2: Baseline demographics and disease characteristics of ARK participants comparing the full study cohort to the sub-cohort used for the PGS analyses	66
Table 3.3: GLM models for Stanzick eGFR PGS stratified by CKD risk factors in the target ARK cohort.....	72

CHAPTER 4

Table 4.1: Baseline characteristics of participants in the ARK study and this case-control study.....	85
--	----

Supplementary Table 4.1: Significantly different proteins in the volcano plot.....	102
---	-----

Supplementary Table 4.2: Top 20 proteins identified from the classifier	110
--	-----

CHAPTER 5

Table 5.1: Key findings with regards to the objectives of the thesis	112
---	-----

LIST OF ABBREVIATIONS

ACC	American College of Cardiology
AHA	American Heart Association
AKI	Acute kidney injury
ARIC	Atherosclerosis Risk in Communities
ARK	African Research Kidney Disease study
AWI-Gen	Africa Wits-INDEPTH Partnership for Genomics Research
BMI	Body mass index
BP	Blood pressure
CE	Capillary electrophoresis
CKD	Chronic kidney disease
DBP	Diastolic blood pressure
DIA	Data-independent acquisition
DM	Diabetes mellitus
DN	Diabetic nephropathy
ECM	Extracellular matrix
eGFR	Estimated Glomerular filtration rate
ESC	European Society of Cardiology
ESH	European Society of Hypertension
ESKD	End stage kidney disease
FDR	False discovery rate
GBD	Global Burden of Disease
H3Africa	Human Heredity and Health in Africa
HT	Hypertension
HIV	Human immunodeficiency virus
HR	Hazard ratio
HWE	Hardy-Weinberg equilibrium
iRT	Normalised retention time
JNC7	The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High BP
KDIGO	Kidney Disease Improving Global Outcomes
LC	Liquid chromatography
LC-MS	Liquid chromatography–mass spectrometry
LD	Linkage disequilibrium
MAF	Minor allele frequency
MeCN	Acetonitrile
MS	Mass spectrometry
OR	Odds ratio
PCA	Principal component analysis
PGS	Polygenic score
QC	Quality control

RAAS	Renin-angiotensin-aldosterone system
SBP	Systolic blood pressure
SD	Standard deviation
SGLT2i	Sodium-glucose cotransporter 2 inhibitors
SNP	Single nucleotide polymorphism
SRA	Serum resistance associated
SSA	sub-Saharan Africa
SWATH	Sequential window acquisition of all theoretical mass spectra
UACR	Urine albumin-to-creatinine ratio
uHLC	urine-Hydrophilic interaction liquid chromatography
UPLC	Ultra-performance liquid chromatography

CHAPTER 1

1. BACKGROUND AND PURPOSE OF STUDY

Chronic kidney disease (CKD) is emerging as a global public health problem, with disproportionate morbidity and mortality in low- and middle-income countries in regions such as sub-Saharan Africa (SSA) (Crews et al., 2019, Ashuntantang et al., 2017). Common risk factors for CKD include hypertension (HT) and diabetes, with HT being the most common in regions such as Western and SSA (Bikbov et al., 2020). Infections such as human immunodeficiency virus (HIV) also contribute to additional risk in high prevalence regions like South Africa (Couser et al., 2011, Ekrikpo et al., 2018). Recently, the introduction of ‘omics technologies have led to the identification of various biomarkers for the early detection of CKD (Provenzano et al., 2021). Polygenic scores (PGSs) estimate an individual's genetic risk for a trait or disease and is calculated by summarising aggregate effects many genetic variants in the genome (Chatterjee et al., 2013, Collister et al., 2022, Khera et al., 2018a). Proteomic approaches permit classification of structure and concentration of unknown molecules through mass spectrometry (MS) combined with ultra-performance liquid chromatography (UPLC) and capillary electrophoresis (CE) (Provenzano et al., 2021). Genetic risk and urinary proteomic profiles for CKD and are understudied in Africa and South Africa. This PhD thesis aims to address this knowledge gap by determining genetic risk models for kidney disease (taking into account risk factors such as HT, diabetes, and HIV infection), and investigating urinary proteomic profiles in a rural South African setting.

1.1. Epidemiology of CKD and HT

1.1.1. Global prevalence and burden of CKD

CKD is a leading cause of premature disability and mortality, globally (Bulto et al., 2017, Rhee and Kovesdy, 2015). In a meta-analysis of 100 studies of 6,908,440 participants, the worldwide prevalence for CKD categories G1–G5 was 13.4% and 10.6% for CKD categories G3–G5 (Hill et al., 2016). A more recent study by the Global Burden of Disease (GBD) Chronic Kidney Disease Collaboration using data from 195 countries from 1990–

2017, reported the worldwide prevalence of CKD to be 9.1% (95% uncertainty interval (UI): 8.5–9.8), which is approximately 700 million cases (Bikbov et al., 2020).

The 2013 Global Burden of Disease (GBD) report showed that even though relative mortality rates declined for the majority of communicable and non-communicable diseases, CKD was one of a few diseases that showed an increase in deaths since 1990 (Aboyans and Collaborators, 2015, Rhee and Kovesdy, 2015). From 1990 to 2017, the global all-age mortality rate for CKD increased by 41.5% relative to previous years (Bikbov et al., 2020). Apart from being one of the leading causes of mortality, CKD is the 12th highest cause of years of life lost (estimated from the number of deaths attributed to CKD and life expectancy of people in a variety of age groups at the time of their death from CKD) (Aboyans and Collaborators, 2015). CKD is predicted to be the 5th leading global cause of years of life lost by 2040 (Foreman et al., 2018).

1.1.2. Prevalence of CKD in Africa, sub-Saharan Africa, and South Africa

In population-based data from 54 African countries originating from the five African sub-regions, the overall pooled population prevalence of CKD was 15.8% (95% CI: 12.1–19.9) for CKD categories G1–G5 in Africa (Kaze et al., 2018). Compared to North Africa (6.1%, 95% CI: 3.6–9.3), the population prevalence of CKD in SSA was significantly higher (17.7%, 95% CI: 13.7–22.1; $P < 0.001$) (Kaze et al., 2018). However, there are marked regional differences as shown in **Figure 1.1**. In a systematic review of the 152 CKD studies, the prevalence of CKD ranged from 2–41% in Africa (pooled prevalence: 10.1%; 95% CI: 9.8–10.5) (Abd ElHafeez et al., 2018).

In population-based cohort data from 2,021 adults from Bushbuckridge, Mpumalanga, South Africa, the WHO age-standardised prevalence for CKD (low eGFR and/or albuminuria) was 6.7% (95% CI: 5.4–7.9), dominated by persistent albuminuria (Fabian et al., 2022). In the Africa Wits-INDEPTH Partnership for Genomics Research (AWI-Gen) study on 10 702 participants from West, East and South Africa, the age-standardised population prevalence for CKD (low eGFR and/or albuminuria) was 10.7% (95% CI: 9.9–11.7) (George et al., 2019). In a community-based cross sectional study of 1,999 participants from the North West Province of South Africa, the crude prevalence of

albuminuria (urine albumin-to-creatinine ratio [UACR] ≥ 3.0 mg/mmol) was 13.0% (Navise et al., 2023).

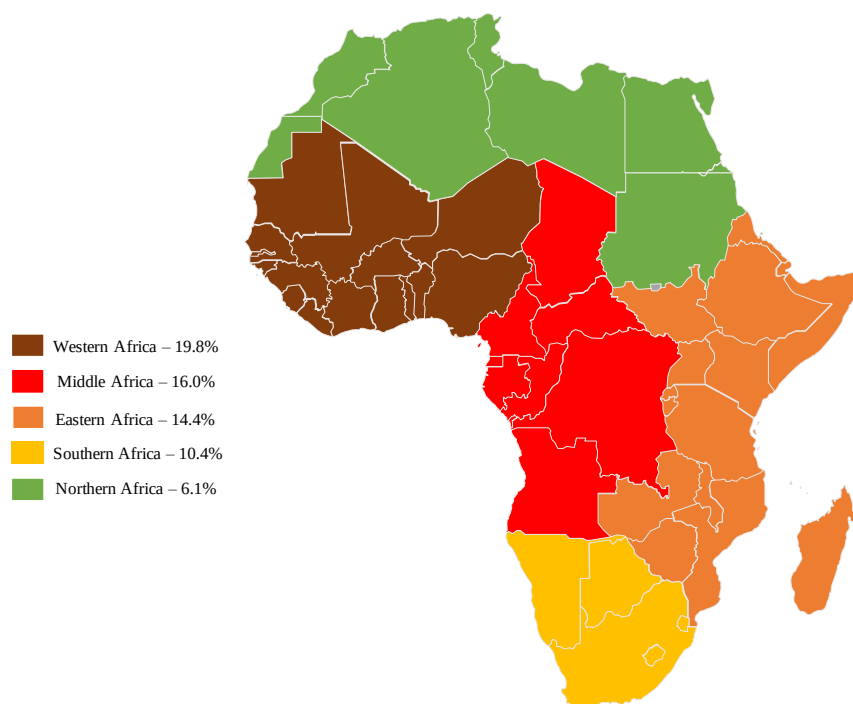


Figure 1.1: Prevalence of CKD in the general population in Africa by region

Adapted from (Kaze et al., 2018): <https://www.pptmaps.com/Editable-map-of-Africa.html>. (Terms of use – Permitted for personal and educational use only).

CKD, chronic kidney disease.

1.1.3. Prevalence and burden of HT

In a global study using data from population-representative studies from Europe, Asia, Latin America, Caribbean, Oceania, Africa, and SSA, the age-standardised prevalence of HT in adults aged 30–79 years was 32% in females (95% credible intervals: 30–34) and 34% in males (95% credible intervals: 32–37) (Zhou et al., 2021). Although the global age-standardised prevalence remained stable, the absolute number of people with HT doubled from 1990 to 2019 (331 to 626 million females and 317 to 652 million males) (Zhou et al., 2021). While the prevalence of HT in high-income countries has decreased over time, it has increased in low- and middle-income countries (Schutte et al., 2023, Mills et al., 2016). Low- and middle-income countries in regions such as SSA have seen little improvement in treatment and control rates of HT since 1990, with treatment rates of <25%, and control rates of <10% (Zhou et al., 2021).

In the past few decades, HT has become a major public health concern in SSA. Current epidemiological trends indicate that the disease burden in SSA has surpassed many European and North American regions. The prevalence of HT ranged from 30–50% in large urban-dwelling areas in Nigeria, and 15–25% in rural communities of Burkina Faso, Ghana, and Nigeria (Seeley et al., 2020). This differences in prevalence may in part be due to the lack of awareness and poor detection rates in rural areas in SSA (Ferdinand, 2020). In a nationally representative study including 13,217 participants from South Africa, the prevalence of HT increased from 27–31% in 1998 to 45–48% in 2016 (Peer et al., 2021). In the AWI-Gen study on 10,695 participants from six sites in four African countries (Burkina, Ghana, Kenya, and South Africa), the prevalence of HT was 33.3% (95% CI: 32.4–34.2) (Gómez-Olivé et al., 2017). The 2020 International Society of HT guidelines proposes that the high prevalence of HT across African-ancestry populations could be attributed to genetic differences and an interplay between environmental and lifestyle factors (Ferdinand, 2020, Unger et al., 2020).

1.2. Causes of CKD

The causes of CKD vary globally, with the most common causes shown in **Table 1.1** (Vaidya and Aeddula, 2023, Webster et al., 2017).

Table 1.1: Global leading causes of CKD (Vaidya and Aeddula, 2023, Webster et al., 2017)

Primary disease	Prevalence
Diabetes mellitus type 2	30–50%
Hypertension	27.2%
Primary glomerulonephritis	8.2%
Diabetes mellitus type 1	3.9%
Chronic tubulointerstitial nephritis	3.6%
Hereditary or cystic diseases	3.1%
Secondary glomerulonephritis or vasculitis	2.1%
Plasma cell dyscrasias or neoplasm	2.1%
Sickle cell nephropathy	<1%

Table from Vaidya SR, Aeddula NR. StatPearls Publishing; 2023. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK535404/> and Webster AC, et al. The lancet. 2017;389(10075):1238-52.
CKD, chronic kidney disease.

1.2.1. Causes and risk factors for CKD vary by region

The causes of CKD (usually biopsy-proven) vary by region, with HT predominating in east Asia, tropical Latin America, and Western and Southern SSA, whereas diabetes is the most common cause in all other regions of the world (Bikbov et al., 2020). CKD arising from glomerulonephritis and unknown causes are more common in Asia, India, and SSA (Vaidya and Aeddula, 2023).

In low- and middle-income settings such as South Africa and SSA, there are many potential causes of CKD. Most individuals with CKD will not have primary kidney disease but kidney damage secondary to HT, diabetes, and vascular disease (Esmail et al., 2023). According to the 2021 South African Renal Registry report, the most common causes of kidney failure were hypertension associated kidney disease (36.6%), kidney failure due to unknown causes (30.4%), diabetes associated kidney disease (14.3%), and glomerular disease (11.0%) (Davids et al., 2023). In settings like South Africa, infections such as HIV also contribute to additional risk for kidney disease (Couser et al., 2011, Ekrikpo et al., 2018).

In the absence of a biopsy, or other definitive diagnoses, the risk for developing CKD is estimated using association analyses. The most common risk factors for CKD in SSA are displayed in **Table 1.2**. In a population-based study from West, East and South Africa, CKD was associated with HIV, HT, and diabetes (George et al., 2019). In a retrospective study from Tanzania, the development of CKD was associated with obesity (body mass index [BMI] ≥ 30 kg/m²), diabetes, and HT (Meremo et al., 2018). In cross-sectional study from Malawi in individuals without diabetes, HT, and/or proteinuria, the risk associated with reduced estimated glomerular filtration rate (eGFR; <90 ml/min per 1.73 m²) were age per ten-year increase, and BMI per 5 kg/m² weight increase (Hamilton et al., 2020). In Uganda, the major risk factors associated with CKD were age (>60 years) and HIV (Muiru et al., 2020). In a cross-sectional study from South Africa, CKD was associated with the presence of diabetes, abdominal obesity, and higher diastolic blood pressure (Adeniyi et al., 2017).

Table 1.2: Risk factors for CKD in sub-Saharan Africa (George et al., 2019, Meremo et al., 2018, Hamilton et al., 2020, Muiru et al., 2020, Adeniyi et al., 2017)

Country/region	Sample size	Risk factors	Risk measurement* (OR, 95% CI)	P value
West, East and South Africa	8,110	HIV	1.65 (1.36–1.99)	<0.0001
		HT	1.97 (1.68–2.30)	<0.0001
		Diabetes	2.22 (1.76–2.78)	<0.0001
Tanzania	1,244	Obesity	8.7 (3.5–15.9)	<0.001
		Diabetes	9.2 (4.6–19.6)	<0.001
		HT	7.5 (3.9–11.7)	<0.001
Malawi	821	Age	3.6 (2.6–5.2)	-
		BMI	2.0 (1.3–3.4)	-
Uganda and Kenya	3,686	Age (> 60 years)	3.5 (1.9–6.5)	-
		HIV infection	1.6 (1.1–2.2)	-
South Africa	489	Diabetes	2.8 (1.1–6.9)	0.03
		Abdominal obesity	2.5 (1.0–5.9)	0.046
		High diastolic blood pressure	1.0 (1.0–1.1)	0.03

Risk measurements included odds ratio and adjusted prevalence ratio.

BMI, body mass index; CI, confidence interval; CKD, chronic kidney disease; HIV, human immunodeficiency virus; HT, hypertension; OR, odds ratio.

1.2.2. CKD heritability

The heritability of CKD depends on the genetic contribution to the disease. On the one hand, it can be very high, for example when attributed to monogenic (also known as “Mendelian”) causes, and lower on the other hand when attributed to complex, polygenic and multi-factorial factors in combination (Köttgen et al., 2022). There are currently more than 600 genes implicated in monogenic causes of CKD (Rasouly et al., 2019, Köttgen et al., 2022). In monogenic CKD, the disease arises as a result of rare, pathogenic variants in a single gene; with a strong genotype-to-phenotype relationship and limited influence from environmental factors (Köttgen et al., 2022). These single gene diseases include autosomal dominant polycystic kidney disease, Alport syndrome, and autosomal recessive disorders such as nephronophthisis (Aung et al., 2021, Soraru et al., 2020, Mallett et al., 2014). Monogenic causes of CKD account for 30–50% of paediatric cases, and 10–30% of adult cases (Mann et al., 2019, van der Ven et al., 2018, Groopman et al., 2019, Domingo-

Gallego et al., 2022, Bullich et al., 2018, Connaughton et al., 2019). In complex or polygenic causes of CKD, there is a lack of a simple pattern of inheritance (i.e., dominant, recessive, or sex-linked inheritance), and it is the result of the aggregate effects of common variants (defined as minor allele frequency $\geq 1\%$) in the genome and environmental factors (Ghosh and Collins, 1996). The last 10 years has seen significant progress in identification of many common variants associated with CKD and kidney function traits through genome-wide association studies (GWASs) and whole exome- or whole genome-sequencing studies, with sample sizes increasing from 100,000 to $\sim 1,000,000$ facilitating the detection of more than 250 genetic loci strongly associated with CKD in different ancestral groups (Tin and Köttgen, 2020) (more information in section 1.6.2).

A major development regarding susceptibility to CKD involved the use of ancestry studies. These ancestry studies took advantage of mixed African and European ancestry in African Americans to identify potentially causal variants in the myosin-9 (*MYH9*) gene that were strongly associated with kidney disease (Kao et al., 2008, Kopp et al., 2011). Subsequent studies identified coding variants in the adjacent apolipoprotein L1 (*APOL1*) gene that explained a substantial fraction of this risk difference for kidney diseases such as CKD and end stage kidney disease (Tzur et al., 2010, Yusuf et al., 2021, Genovese et al., 2010). The *APOL1* risk variants comprise of two non-synonymous variants collectively referred to as the G1 risk allele and a 6 bp in-frame deletion referred to as the G2 risk allele (Limou et al., 2014, Yusuf et al., 2021).

Individuals carrying the *APOL1* risk variants have been shown to be protected from trypanosomiasis, also known as sleeping sickness, a condition that affects millions of Africans (Kasembeli et al., 2015). The G1 and G2 variants are distinctly distributed among individuals with African ancestry (Kasembeli et al., 2015). The *APOL1* risk alleles arose by chance and were subjected to positive selection, therefore, increased in frequency in trypanosomiasis endemic regions (Kasembeli et al., 2015). There are 3 *Trypanosoma* species; *Trypanosoma brucei brucei*, *Trypanosoma brucei rhodesiense*, and *Trypanosoma brucei gambiense* (Brun et al., 2010). T.b. *brucei* species is incapable of infecting humans due to serum containing apolipoprotein A1, high-density lipoprotein, and trypanolytic factor containing apolipoprotein L1 (Brun et al., 2010). However, the other two

Trypanosoma species (*T.b. rhodesiense* and *T.b. gambiense*) evade this host defence mechanism by expressing a serum resistance associated (SRA) protein. The SRA protein binds to and inactivates the wild type *APOL1* protein. However, the ability of *APOL1* to lyse *T.b. rhodesiense* is restored by the presence of the G1 and G2 risk variants, which occur near and in the SRA binding site (Brun et al., 2010).

The *APOL1* renal risk alleles (G1 and G2) evolved to protect carriers from trypanosomiasis in West Africa and were subsequently spread to other regions of the world after the out-of-Africa migrations. Because of the evolutionary selection, the highest G1 and G2 risk allele frequency is observed in West Africa, with a combined allele prevalence of over 40% (Daneshpajouhnejad et al., 2022). The combined frequency (G1 and G2) is approximately 34% in African Americans and absent in individuals without African ancestry (Daneshpajouhnejad et al., 2022). In the AWI-Gen study the allele frequencies showed some differences from previous studies, South and West Africa had the highest G1 allele frequencies (approximately 7–13%). The frequencies for the G2 allele were also highest in South Africa (15–24%), followed by West Africa (9–12%) (Brandenburg et al., 2022a). The lower frequencies in the West African populations in AWI-Gen point to differences among West African populations and ethnic groups. African American individuals with both *APOL1* risk alleles tend to have a 5-fold higher risk of incident albuminuria, 10-fold higher risk of ESKD and 7-fold higher risk of ESKD due to HT (Cañadas-Garre et al., 2019, Peralta et al., 2016). In East, West and South Africans from the AWI-Gen study, *APOL1* risk variants were significantly associated with albuminuria (OR, 1.63; 95% CI: 1.25–2.12) in the recessive model (G1/G1, G2/G2, or G1/G2), but not with eGFR levels (Brandenburg et al., 2022a).

1.3. Guidelines for defining CKD and HT

1.3.1. Classification of CKD

The Kidney Disease Improving Global Outcomes (KDIGO) 2012 Clinical Practice Guideline for the Evaluation and Management of CKD is widely accepted as a reference for clinical care (Levin et al., 2013). Using the KDIGO guidelines, CKD is determined by measuring GFR and urine albumin excretion (albuminuria). Normal kidney function is defined as (i) eGFR ≥ 90 ml/min per 1.73 m^2 and (ii) UACR of < 3 mg/mmol. The diagnosis

of CKD is based on (i) the presence of $\text{eGFR} < 60 \text{ ml/min/1.73 m}^2$ determined using the Chronic Kidney Disease Epidemiology Collaboration equation (CKD-EPI) equation, or (ii) one or more markers of kidney damage: albuminuria ($\text{UACR} \geq 3.0 \text{ mg/mmol}$), abnormal urine sediment, tubular disorders, histology or imaging abnormalities, or history of kidney transplant (Levin et al., 2013).

To define CKD, evidence of chronicity is required, either from a patient's history, prior investigations, or with repeat measurements at least three months apart. CKD is defined using specified cut-offs for both eGFR and albuminuria, as each biomarker independently predicts increased risk for premature death and adverse cardiovascular outcomes (Go et al., 2004). If CKD is confirmed, it is further classified by GFR and albuminuria category for stratification into risk categories (**Figure 1.2**) (Levin et al., 2013). It is important to note that these cut-offs are defined from pooled analyses in population cohorts and none of the studies included African populations. Therefore, while these cut-offs are commonly used, there have not been validated as appropriate for African populations. CKD progression is assessed by measuring eGFR and albuminuria annually in individuals with confirmed CKD. CKD progression is determined by: (i) a decline in GFR category (G1, ≥ 90 ; G2, 60–89; G3a, 45–59; G3b, 30–44; G4, 15–29; G5, $< 15 \text{ ml/min/1.73 m}^2$); or (ii) a more than 5 $\text{ml/min/1.73 m}^2/\text{year}$ decline in eGFR (i.e., rapid progression) (Levin et al., 2013).

				Persistent albuminuria categories Description and range		
				A1	A2	A3
				Normal to mildly increased	Moderately increased	Severely increased
				<30 mg/g <3 mg/mmol	30-300 mg/g 3-30 mg/mmol	>300 mg/g >30 mg/mmol
GFR categories (ml/min/ 1.73 m ²) Description and range	G1	Normal or high	≥90			
	G2	Mildly decreased	60-89			
	G3a	Mildly to moderately decreased	45-59			
	G3b	Moderately to severely decreased	30-44			
	G4	Severely decreased	15-29			
	G5	Kidney failure	<15			

Figure 1.2: KDIGO 2012: CKD classification and prognosis by GFR and albuminuria categories (Levin and Stevens, 2014)

Green: low risk (if no other markers of kidney disease, no CKD); Yellow: moderately increased risk; Orange: high risk; Red, very high risk.

Diagram from Levin A, Stevens PE. (2014) (Under an Elsevier user license; see Creative Commons licence: <http://creativecommons.org/licenses/by/4.0/>).

CKD, chronic kidney disease; GFR, glomerular filtration rate

The 2022 KDIGO Clinical Practice Guideline for Diabetes Management in Chronic Kidney Disease recommends treatment with sodium-glucose cotransporter 2 inhibitors (SGLT2is) for improving kidney function outcomes in individuals with type 2 diabetes, CKD, and eGFR ≥30 ml/min per 1.73 m² (Rossing et al., 2022).

1.3.2. Classification of HT

There are various guidelines (including guidelines from South Africa) for determining ranges for normal and elevated blood pressure (BP) as shown in **Table 1.3**: (i) the Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High BP (JNC7) guidelines (Chobanian et al., 2003); (ii) the 2017 American College of Cardiology/American Heart Association (ACC/AHA) Clinical Practice Guideline for High Blood Pressure in Adults (Whelton and Carey, 2018); (iii) 2020

International Society of Hypertension guidelines; (iv) 2023 European Society of Hypertension (ESH) guidelines for the Management of Arterial Hypertension (Mancia et al., 2023), and (v) South African hypertension practice 2014 guidelines (Seedat et al., 2014).

Table 1.3: Ranges for normal and elevated blood pressure according to various guidelines

BP category	Systolic blood pressure (mmHg)	And/or	Diastolic blood pressure (mmHg)
JNC7 (Chobanian et al., 2003)			
Normal	<120	And	<80
Prehypertension	120–139	Or	80–89
Hypertension, stage 1	140–159	Or	90–99
Hypertension, stage 2	>160	Or	>100
2017 ACC/AHA (Whelton and Carey, 2018)			
Normal	<120	And	<80
Elevated	120–129	And	<80
Hypertension, stage 1	130–139	Or	80–89
Hypertension, stage 2	≥140	Or	≥90
ISH 2020 (Unger et al., 2020)			
Normal	<130	And	<85
High-normal	130–139	and/or	85–89
Hypertension, stage 1	140–159	and/or	90–99
Hypertension, stage 2	≥160	and/or	≥100
2023 ESH (Mancia et al., 2023)			
Optimal	<120	And	<80
Normal	120–129	And	80–84
High-normal	130–139	and/or	85–89
Hypertension, grade 1	140–159	and/or	90–99
Hypertension, grade 2	160–179	and/or	100–109
Hypertension, grade 3	≥180	and/or	≥110
Isolated systolic hypertension	≥140	And	<90
Isolated diastolic hypertension	<140	And	≥90
2014 South African hypertension practice 2014 (Seedat et al., 2014)			
Normal	<120	And	<80
Optimal	120–129	and/or	80–84
High normal	130–139	and/or	85–89
Hypertension, grade 1	140–159	and/or	90–99
Hypertension, grade 2	160–179	and/or	100–109
Hypertension, grade 3	≥180	and/or	≥110
Isolated systolic hypertension	≥140	And	<90

The ESC/ESH guidelines does not focus on prevention and specific ethnic/racial groups (i.e., black, Hispanic etc) (Bakris et al., 2019). The ACC/AHA guidelines have lower cut-off points for hypertension, resulting in an increased projected hypertension prevalence at a population level and have been criticised for mislabelling many patients as hypertensive, with no clear indication of increased CVD risk (Egan et al., 2018, Egan et al., 2019, Poulter et al., 2018, Khera et al., 2018b, Muntner et al., 2018). The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High BP (JNC7) guidelines were the most up to date at the time the African Research on Kidney Disease (ARK) study was initiated in 2016 and the choice to abide by them was made apriori, as part of the study design. For the purposes of this PhD, the JNC7 guidelines were used as it is used globally and provides an evidence-based approach for the prevention and management of high BP (Chobanian et al., 2003). The classification of HT in adults (≥ 18 years) is based on the average of two or more BP readings.

Since high BP is one of the most common risk factors for adverse kidney outcomes, the KDIGO published guidelines for the management of BP in CKD (Cheung et al., 2021). In adults with elevated BP and CKD, standardised office BP measurement can be used to target systolic blood pressure (SBP) of <120 mm Hg, when tolerated. In patients with very limited life expectancy, clinicians may propose less intensive BP-lowering therapy (Cheung et al., 2021). In addition, in patients with elevated BP, CKD, and increased albuminuria (A3), renin-angiotensin-system inhibitors, angiotensin-converting enzyme inhibitor, or angiotensin II receptor blocker are recommended (Cheung et al., 2021).

1.4. The complex pathophysiology and interplay between CKD and hypertension

1.4.1. The molecular pathways in the pathophysiology of CKD

The pathogenesis of CKD is multifactorial, involving renal inflammation, tubular atrophy, glomerulosclerosis, and tubulointerstitial fibrosis (**Figure 1.3**) (Gusev et al., 2021, Yang and He, 2019). In the setting of high BP, dyslipidaemia, injury or stress stimuli,

glomerulus-resident cells such as mesangial cells and podocytes become damaged and activated and release various cytokines (e.g., interferons and tumour necrosis factor) chemokines (e.g., monocyte chemoattractant protein-1 and RANTES), which permits the migration and accumulation of multiple types of inflammatory cells to the site of injury (Yang and He, 2019, Yuan et al., 2022). These infiltrating immune cells together with intrinsic cells produce various cytokines and growth factors such as angiotensin, transforming growth factor β 1, and fibroblast growth factor which lead to collapse obliteration of glomerular capillaries and loss of podocytes (Yang and He, 2019, López-Novoa et al., 2011). Consequently, extracellular matrix (ECM) components are accumulated to replace the tissue that has been damaged. Glomerulosclerosis is characterised by an increase in mesangial matrix accumulation and collapse of glomerular capillaries and leads to loss of kidney function (Yang and He, 2019).

In the setting of high glucose concentration, hypoxemia, or sustained kidney injury, glomerulus-resident cells are injured and release chemokines that provides a signal for inflammatory cells to immigrate to the site of injury (Yang and He, 2019, Lin and Hsu, 2020). These invading inflammatory cells produce multiple compounds and protein factors such as reactive oxygen species, interleukin 1, and tumour necrosis factor α , which leads to exacerbation of the renal-cell injury (Yang and He, 2019). This series of events activates fibroblasts, myofibroblasts and tubular epithelial cells, resulting in the overproduction of ECM components. The excessive deposition of ECM components in the renal interstitial space causes renal tubular damage and atrophy, which in turn leads to fibrosis, a key characteristic of CKD (Yang and He, 2019, Cho, 2010, Huang et al., 2023).

HT can cause kidney injury in tubular cells, beyond its well-established effects on the renal vasculature and glomeruli (Sievers and Eckardt, 2019). The mechanisms of hypertension-induced tubular injury include ischemia and hypoxia, increased intraluminal pressure, proteinuria, oxidative stress and inflammatory responses, and activation of RAAS. In the setting of chronic hypertension, blood vessels can narrow and stiffen, resulting in decreased blood flow to the kidneys (Safar, 2018). This diminished blood flow can lead to ischemia (reduced blood supply) and hypoxia (reduced oxygen supply) of the tubular cells (Wang et al., 2022). Over time, this can result in tubular injury and contribute to the progression of

CKD (Ullah and Basile, 2019). HT has been shown to cause raised glomerular pressure, that can increase the pressure in the tubules (Wang et al., 2013). This increased pressure can directly damage the tubular epithelial cells and lead to kidney dysfunction (Hall, 2016). Raised glomerular pressure has been shown to result in filtration of proteins into the tubular fluid (Palatini, 2012). This excess proteins in the tubules can be toxic to the tubular cells, leading to injury through processes like oxidative stress and inflammation (Gyurászová et al., 2020). HT can induce oxidative stress through the generation of reactive oxygen species (ROS) (Rodrigo et al., 2011). These ROS can damage tubular cells directly or indirectly by activating inflammatory pathways, which further contribute to tubular injury and fibrosis. (Rodrigo et al., 2011). Over activation of the RAAS is strongly linked to hypertension, and can exacerbate kidney damage (Moon, 2013, Muñoz-Durango et al., 2016). Components of RAAS, particularly angiotensin II, has been shown to directly induce tubular cell injury and fibrosis (Gelen et al., 2021). Angiotensin II has been shown to activate many inflammatory and fibrotic pathways which contribute to tubular cell apoptosis and fibrosis (Navar et al., 1996, Rüster and Wolf, 2006, Siragy and Carey, 2010).

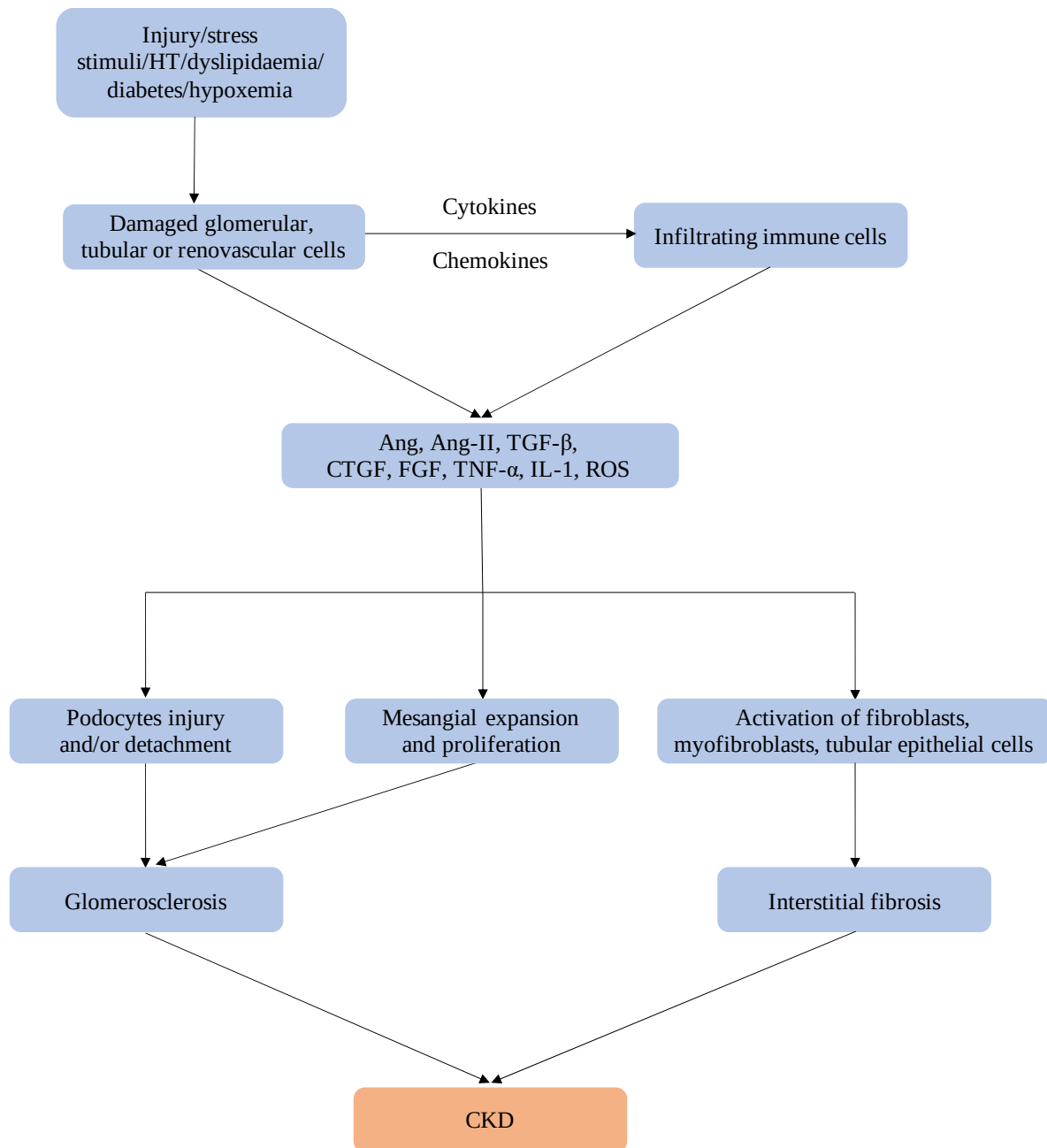


Figure 1.3: Schematic presentation of cellular events involved in CKD

Adapted from Yang J, He W. Springer, 2019 (Yang and He, 2019).

Ang, angiotensin; CKD, chronic kidney disease; CTGF, connective tissue growth factor; FGF, fibroblast growth factor; HT, hypertension; IL-1, interleukin 1; ROS, reactive oxygen species; TGFβ, transforming growth factor beta; TNF-α, tumour Necrosis Factor alpha.

1.4.2. The mechanisms underlying hypertension-associated kidney disease

HT can be both a cause and a consequence of kidney disease, and the interaction between HT and CKD is complex, making it difficult to establish which disease precedes the other, as both share similar risk factors including age and comorbidities such as diabetes, cardiovascular disease and obesity (Hamrahan and Falkner, 2016). There are a number of mechanisms underlying hypertension-associated kidney disease including endothelial dysfunction, renin-angiotensin-aldosterone system (RAAS), and glomerular hypertension (**Figure 1.4**). HT causes changes in endothelial cells by altering the balance of the various endothelial cell factors such as reduced nitric oxide and increased endothelin-1 (ET-1), which in turn leads to endothelial dysfunction (Konukoglu and Uzun, 2017, Drożdż et al., 2022, Kostov, 2021). A reduction of nitric oxide synthesis causes damage to the arteries, arterioles, and capillaries of the kidney; and enhances progression of kidney disease in the setting of hypertension (Agarwal, 2004, Vaziri et al., 2002). Increased levels of ET-1 have been shown to cause an increase in oxidative stress and contribute to renal vascular inflammation and fibrosis, which are well known features of kidney disease (Elmarakby et al., 2005, Pollock, 2000, Panizo et al., 2021, Dhaun et al., 2012).

The RAAS plays an important role in regulating blood pressure, with over activation of the RAAS strongly linked to hypertension (Moon, 2013, Muñoz-Durango et al., 2016). Angiotensin II has been shown to directly affect vascular smooth muscle cells in the kidneys resulting in the vasoconstriction of both afferent and efferent arterioles, causing glomerular hypertension and a reduction of blood flow in the kidneys (Mennuni et al., 2014). Angiotensin II has been shown to activate many inflammatory and fibrotic pathways which contribute to CKD and ESKD (Navar et al., 1996, Rüster and Wolf, 2006, Siragy and Carey, 2010). Angiotensin II also stimulates the production of ET-1 resulting in an increase in oxidative stress (Pollock, 2000, Sasser et al., 2002).

In the setting of sustained elevated BP, afferent arterioles (blood vessels which carry blood to the glomeruli) undergo structural adaptations and develop hypertrophy leading to glomerular injury (Bidani and Griffin, 2004, Shankland, 2006, Folkow et al., 1977). In the glomerulus, raised capillary pressure results in increased glomerular protein filtration and

scarring which leads to abnormally high amounts of protein in the urine. In many cases, the first sign of CKD is the presence of albuminuria (Bidani and Griffin, 2004, Shankland, 2006, Folkow et al., 1977, Buffet and Ricchetti, 2012, Heerspink and Gansevoort, 2015).

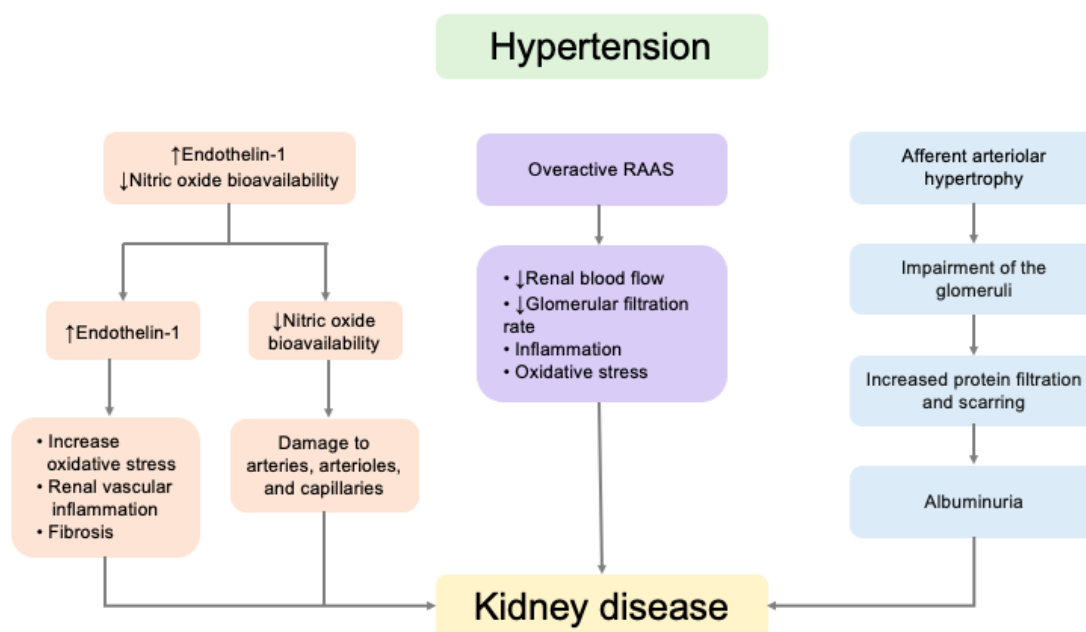


Figure 1.4: Mechanisms involved in the pathophysiology of hypertension-associated kidney disease

Created own image.

RAAS, renin-angiotensin-aldosterone system.

1.5. The importance of omics' studies in kidney disease

CKD is diagnosed based on changes in eGFR and/or albuminuria which indicate deterioration of kidney function (Levin et al., 2013). However, eGFR is relatively insensitive, as there can be a large decline in kidney function with no/little change in eGFR. In addition, eGFR has limited usefulness in predicting the risk of progression to ESKD, unless it is greatly reduced (Argiles et al., 2013). Albuminuria can be an early urinary marker of kidney disease, and may predict decline in eGFR. However, kidney disease progresses in some patients regardless of albuminuria and/or proteinuria being absent (Miller et al., 2009). Therefore, there remains a need for other kidney disease markers that permit early detection and prediction of disease progression (Govender et al., 2021). Such markers would facilitate interventions to slow the progression of kidney disease, which is

useful in settings like South Africa with limited resources, where there is little/no access to dialysis and kidney transplants for patients with ESKD (Govender et al., 2021).

The introduction of omics approaches has permitted the discovery of various biomarkers for the early detection of CKD (Provenzano et al., 2021). Omics techniques allow high-throughput, comprehensive exploration of the genomic, proteomic, epigenomic, transcriptomic, and metabolomic biomarkers, through various approaches such as next-generation sequencing, mass spectrometry, and mRNA arrays (Papadopoulos et al., 2016, Provenzano et al., 2021). These omics biomarkers generated from urine, blood and kidney biopsy tissues may be more strongly associated with pathophysiology of disease in comparison to conventional markers including eGFR and albuminuria (Provenzano et al., 2021).

One of the key advantages of utilising genomics is that if a causal genetic variation is identified, unnecessary treatments can be avoided, and tailored therapies can be proposed (Grobe et al., 2023). Genomic sequencing has successfully identified monogenic causes of CKD, attributing up to 30% of all adult kidney disease cases as inherited (Mann et al., 2019, van der Ven et al., 2018, Groopman et al., 2019, Domingo-Gallego et al., 2022, Bullich et al., 2018, Connaughton et al., 2019). A noteworthy example of translating genomics into the clinic came with the discovery of renal risk variants in the *APOL1* gene in individuals with African ancestry (Genovese et al., 2010, Tzur et al., 2010). Individuals with African ancestry and high-risk *APOL1* variants are at increased risk for developing ESKD and hypertensive-associated ESKD (Cañadas-Garre et al., 2019). In addition, many studies have shown that donor kidneys from patients with high-risk *APOL1* variants fail at higher rates than non-risk donor kidneys (Freedman et al., 2015, Freedman et al., 2016). This has allowed for identification of individuals at risk for kidney disease and current discussions for the assessment of risk stratification in a kidney transplant setting (Friedman and Pollak, 2021). Relatively inexpensive genomics-informed methods that predict kidney disease risk present a great opportunity for low- and middle-income countries like South Africa, where the morbidity and mortality of noncommunicable conditions such as CKD are disproportionately high (Govender et al., 2021).

Recently, the use of proteomic approaches has gained a lot of interest since it is able to use urine as the biological sample (Provenzano et al., 2021). Urine samples are easily accessible and can be often requested by healthcare workers for assessing treatment response and stratification of risk for future outcomes in CKD patients (Provenzano et al., 2020, Heerspink and Gansevoort, 2015). Proteomic approaches that permit classification of structure and concentration of unknown molecules are MS combined with UPLC and CE (Provenzano et al., 2021) are used to identify markers as early indicators of kidney disease and disease mechanisms (Mischak et al., 2010, Fan et al., 2021, Pontillo and Mischak, 2017a). However, these biomarkers are mostly used to identify CKD and predict disease progression in populations with European ancestry, emphasizing the necessity for representative proteomics studies in populations with African ancestry (Govender et al., 2021).

1.6. Genomic approaches and examples

1.6.1. GWAS definition and concept

Since the completion of the Human Genome Project in 2003, GWASs have evolved as an important approach in detecting genetic associations owing to its non-hypothesis driven approach (Lander et al., 2001). A GWAS approach tests hundreds of thousands to millions of variants in the genome in a large group of individuals to identify associations between the genotype and phenotype (Tam et al., 2019). The GWAS approach depends on the linkage disequilibrium (LD) concept, where genotype-phenotype associations are identified using non-random association between genetic variants and traits within a population (Visscher et al., 2012). The association between genomic variants and a diseases or trait is investigated by the detection of markers in the gene that tag the variants that are causal (Schaid et al., 2018). These associations have provided an understanding of the genetic architecture of disease susceptibility of numerous diseases leading to significant developments in clinical care (identification of disease biomarkers and drug targets) and precision medicine (risk prediction and tailored therapies based on genotype).

1.6.2. GWAS for CKD

In the last few years, there have been significant progress in GWASs of kidney function traits with sample sizes increasing from 100,000 to over 1,000,000 permitting the

identification of more than 250 highly reproducible genetic loci, with inclusion of participants with European and non-European ancestries and utilisation of advanced computational approaches to identify potentially causal genes and variants (Tin and Köttgen, 2020). To understand the genetic architecture of complex diseases, continuous outcomes offers higher statistical power in comparison to binary outcomes, defined as the presence or absence of a disease (Bush and Moore, 2012). As such, continuous measures used for CKD diagnosis including eGFR and UACR have become the main outcomes for most CKD-related GWASs (Levin et al., 2013, Tin and Köttgen, 2020).

A summary of the main GWASs for CKD and kidney function traits is shown in **Table 1.4**. Kidney functions traits investigated included eGFR, UACR, albuminuria and Cystatin C. The sample sizes ranged from 3,288 to 1.5 million, with the largest study including only 5.5% of individuals with African ancestry. The heritability for the various kidney function traits ranged from 0.4–21%, with the lowest heritability in a GWAS on individuals from continental Africa.

The first GWAS for CKD (eGFR based on creatinine) in continental Africa was conducted for 3,288 participants of East African descent from Uganda General Population Cohort (Fatumo et al., 2021). They identified two loci for eGFR (single nucleotide polymorphism [SNP] rs2433603 in the glycine amidinotransferase [*GATM*] locus; SNP rs141845179 in the haemoglobin beta [*HBB*] locus), at the genome-wide significance level of $P < 5 \times 10^{-8}$. The association signal in the *GATM* locus with eGFR was distinct from signals that were previously reported at this locus in European and transethnic ancestry studies (Hellwege et al., 2019, Wuttke et al., 2019). In another study, a GWAS for eGFR (based on cystatin-C) on 5,877 individuals from the Uganda Genomic Resource identified three SNPs at genome-wide significance ($P < 5 \times 10^{-8}$); rs59288815 (*ANK3*), rs4277141 (*OR51B5*), and rs911119 (*CST3*) (Mayanja et al., 2023).

Although there have been increased efforts to improve genetic diversity in GWASs, the majority of the SNPs linked to CKD or kidney function are derived from studies on European individuals, indicating sample selection bias in current GWASs. Larger studies

are needed to detect genetic variants that are exclusive to African populations that may influence kidney function.

Table 1.4: Summary of key GWASs for kidney function traits

First author and year	Sample size	Ancestry	% African ancestry	Trait	Heritability
(Wuttke et al., 2019)	1,046,070	European, African American, South and East Asian, Hispanic	1.3%	eGFR	0.20
(Teumer et al., 2019)	564,257	European, African American, Asian, Hispanic	1.0%	UACR	0.04
				Albuminuria	-
(Casanova et al., 2019)	437,027	European	0%	UACR	0.006
(Stanzick et al., 2021)	765,348	European, African American, Asian	1.8%	eGFR	0.098
(Sinnott-Armstrong et al., 2021)	363,228	European	0%	eGFR	0.20
				Urea	0.12
				Cystatin C	0.21
				Albuminuria	0.03
(Liu et al., 2022)	1,500,000	European, African American, African, East and South Asian, Hispanic or Latino, Admixed/Native/Asian American, Middle Eastern.	African (4.0%) African American (1.5%)	eGFR	0.07
(Fatumo et al., 2021)	3,288	Continental African	100%	eGFR	0.0004
(Mayanja et al., 2023)	5,877	Continental African	100%	Cystatin C	-

eGFR, estimated glomerular filtration rate; UACR, urine albumin-to-creatinine ratio.

1.7. Polygenic scores (PGSs) and related steps

1.7.1. Overview of PGSs

The contribution of genetic risk for complex diseases is based on the heritability of the trait and the calculation of a polygenic score. A PGS, also called a polygenic risk score or genetic risk score, is a calculation of an individual's risk for a certain disease or condition

based on their genetic makeup, attained by capturing the aggregate effects of common variants (minor allele frequency at least 1%) in the genome, that individually contribute a small fraction of overall risk, but collectively account for a higher proportion of risk for a specific trait or disease (Chatterjee et al., 2013, Collister et al., 2022, Khera et al., 2018a).

A basic PGS equation for an individual is:

$$PGS = \sum_{i=1}^n w_i \cdot X_i$$

where n is the number of SNPs, w_i is the effect size (beta value) of the i^{th} variant for a particular outcome, and X_i is the effect of the allele (e.g., 0= non-effect allele homozygote, 0.5=heterozygote, 1=effect allele homozygote) (Reisberg et al., 2017). The effect size or beta value is attained from a GWAS, known as the “discovery” data, where each variant is evaluated for association with the particular trait or disease of interest, and the effect sizes are estimated (Collister et al., 2022). If an individual’s genetic score is greater than the control or reference group, that individual is thought to be at increased risk for the given trait or condition (Osterman et al., 2021).

1.7.2. SNP genotyping

Genotyping is the process of analysing genetic variations in an individual’s genome (Kockum et al., 2023). A genotype refers to the total sum of alleles an individual carries for a specific gene, inheriting one allele from each parent (Kockum et al., 2023). A low-cost approach to genotyping selected variants arose with the development of SNP genotyping arrays. The H3Africa (Human Heredity and Health in Africa) Consortium together with Illumina Inc., designed a custom array that includes ~ 2.3 million SNPs (Mulder et al., 2018). This array was specifically developed to capture common variants in African populations and is based on >3000 African whole genome sequences (Mulder et al., 2018). The H3African SNP array comprise of genetic variants that were previously described to be associated with several diseases (Mulder et al., 2018). The LD efficiency in African populations was highest in the H3Africa SNP array compared to other arrays (Omni, Panaf, and Mega), making it the most appropriate array for GWASs performed in African populations (Mulder et al., 2018).

1.7.3. Quality control

The quality of the discovery (GWAS summary statistics [e.g. P and beta and values]) and target (genotypes and phenotypes of individuals in the target sample) datasets determines the power and validity of imputation and PGS analysis. Thus, both datasets need to go through stringent quality control (QC) to ensure the data is of high standard (Choi et al., 2020). In this doctoral study, the QC process involving a stepwise removal of SNPs and individuals using the H3Africa Consortium Pan-African Bioinformatics Network (H3ABioNet H3AGWAS) workflow (<https://github.com/h3abionet/h3agwas>) will be used (Baichoo et al., 2018, Brandenburg et al., 2022b). Individual QC includes the removal of samples with discordant sex information (genotype versus documented) and those with a SNP missingness greater than 0.02. Heterozygosity cut-offs are used to remove individuals with excessive heterozygosity (>0.343) which indicates genotyping errors, and low heterozygosity (< 0.15) which indicates inbreeding. Relatedness cut-offs ($\pi^{\wedge}$ - π hat between 0.1 and 0.7) are applied to account for closely related individuals in a study i and keep as many samples as possible. The SNP QC process involves exclusion of duplicate and non-autosomal SNPs from the dataset. SNPs with a minor allele frequency (MAF) below 0.01, missingness greater than 0.01, and Hardy-Weinberg equilibrium (HWE) $P < 0.0005$ are removed during the QC process.

1.7.4. Imputation

Imputation is a statistical approach that uses a reference panel to infer haplotypes, thereby predicting unobserved genotypes in an array (Das et al., 2018). This technique increases the coverage for GWAS, and enhances meta-analysis of independent GWASs by increasing the number of overlapping SNPs between different studies (Das et al., 2018). Some of these reference panels are publicly available, while others can be accessed through particular imputation services (Sengupta et al., 2023). Imputation performance has been shown to be positively impacted by reference panel size and its genetic closeness to the targeted population (Das et al., 2018). As such, larger panels with greater representation from different geographical regions have been developed. Examples of these panels include the 1000 Genomes Project ($n=2,504$) (Auton et al., 2015), the Haplotype Reference Consortium ($n=32,470$) (McCarthy et al., 2016), Trans-Omics for Precision Medicine (TOPMed; $n=97,256$) (Taliun et al., 2021), and the African Genome Resource ($n=4,956$) at

the Sanger Imputation Server (<https://imputation.sanger.ac.uk/>) (McCarthy et al., 2016, Gurdasani et al., 2019). The African Genome Resource is a promising panel for SSA datasets based on size, geographic region, and representation of African individuals (Sengupta et al., 2023).

1.7.5. Techniques and software for calculating PGS

Various techniques such as clumping (LD adjustment), beta shrinkage, and P value thresholding are used to calculate a PGS using external summary statistics and genotypes. Since all SNPs do not influence the trait or phenotype and SNP effects are not determined with certainty, using SNPs with unadjusted effect size estimates in a PGS could result in an estimate score with high standard error (Choi et al., 2020). Two shrinkage approaches have been created to address this issue: (i) effect size shrinkage of all SNPs, and (ii) P value thresholding (Choi et al., 2020). The effect estimates shrinkage of all SNPs are performed using conventional or custom statistical approaches including Least Absolute Shrinkage and Selection Operator (LASSO) or ridge regression, or Bayesian techniques that perform shrinkage using prior distribution specification (Mak et al., 2017, Ge et al., 2019). In the P value thresholding approach, SNPs with a P value below a specific threshold from a GWAS (e.g. $P < 1 \times 10^{-5}$) are selected to be incorporated in the PGS, resulting in shrinking all SNPs that are excluded to an effect size estimate of zero and not shrinking any of the included SNPs (Purcell et al., 2009, Euesden et al., 2015). There are two options for LD adjustment: the first is LD-driven clumping, which involves retaining SNPs with a weak correlation with each other (a clump comprises of all SNPs within 250 kb of the index SNP), therefore, their effects can be aggregated. The second option is inclusion of all SNPs and considering LD between the SNPs (Choi et al., 2020). The usual approach to calculation of a PGS includes a combining LD-driven clumping with P value thresholding (“C+T” method) (Choi et al., 2020).

Although there are various software programs that can be used to generate a PGS, PRSice-2 was used for this doctoral work. PRSice-2, is known for its innovative features and improvements from earlier versions, and its ability to automate and simplify PGS analyses for large data sets (Choi and O'Reilly, 2019). PRSice-2 uses the “C+T” standard approach

to calculating a PGS and has a user-friendly interface which makes it widely accessible to researchers from a various backgrounds (Choi and O'Reilly, 2019).

1.7.6. PGSs for kidney disease

Teumer et al. (2019) performed a large trans-ethnic meta-analysis study on 564,257 participants (**Table 1.4**) and examined the association between a PGS and CKD. A PGS derived from 64 UACR associated-loci was significantly linked to CKD ($P=5.0 \times 10^{-6}$). Additional analyses showed that this PGS was associated with an increased risk of HT ($P=2.4 \times 10^{-21}$) and diabetes ($P=1.0 \times 10^{-10}$) (Teumer et al., 2019). A genome-wide PGS for eGFR was developed and assessed for its association with incident CKD, kidney failure, and acute kidney injury (AKI) in 8,866 individuals from the Atherosclerosis Risk in Communities (ARIC) study (Yu et al., 2021). The PGS was significantly associated with all outcomes of kidney disease: hazard ratios (HRs; 95% CI) were 1.33 (1.28–1.37), 1.18 (1.08–1.29), and 1.06 (1.01–1.11) for incident CKD, kidney failure, and AKI, respectively (Yu et al., 2021).

A PGS for eGFR was examined for its association with adverse events (including overall mortality) in 4,873 European ancestry individuals from German Chronic Kidney Disease study with CKD (Steinbrenner et al., 2022). A high PGS corresponded to an increased genetic risk for lower eGFR, and associated with increased risk of mortality (HR; 95% CI: 1.12; 1.04–1.22) (Steinbrenner et al., 2022). In a more recent study, 11,813 healthy elderly participants from the Australian ASPREE trial were genotyped and a PGS was developed for eGFR (Bakshi et al., 2023). The PGS was associated with a 3.9 mL/min/1.73m² (95% CI: –4.17 to –3.68) decrease in eGFR per standard deviation (SD) increase of the PGS, and also associated with increased risk for CKD stage 3 (OR, 1.75; 95% CI: 1.66–1.85) (Bakshi et al., 2023).

1.8. Proteomic approaches and examples

1.8.1. Urine as a source of proteins and extraction by precipitation

The proteome is the complete set of proteins in a cell, tissue or organism at a specific time (Chandramouli and Qian, 2009). Proteomics is the study and characterisation of the function and structure of the complete set of proteins present in a cell to understand the expression of proteins, protein-protein interactions and post-translational alterations

(Chandramouli and Qian, 2009, Al-Amrani et al., 2021, Aslam et al., 2017). A biomarker is an objective measure that is used to denote usual biological processes, pathogenic processes, or the effect of drug responses (Califf, 2018). Urine has become one of the leading biological samples in proteomic analyses because it can be easily collected and is relatively stable in comparison to other biological samples (Barratt and Topham, 2007, Lin et al., 2018). Urine contains a range of components from low molecular weight peptides and metabolites to large complexes such as exosomes, extracellular vesicles, and nucleic acids (Lin et al., 2018, Hildonen et al., 2016).

Proteins can be obtained from fresh or frozen urine samples using numerous techniques, that depends on the objective of the research (Shantaram, 2021). The urinary proteome is dynamic in nature as it changes significantly over time (Kalantari et al., 2015). The appropriate collection of urine samples is important to prevent contamination or deterioration of urine components (Delanghe and Speeckaert, 2014). Therefore, urine need to be collected and stored in manner that preserves its components to allow for downstream analysis (Bosschieter et al., 2018).

Proteomic analyses require reliable sample preparation, in which the sample contains a high protein concentration that is free from other components such as salt and nucleic acids (Antonioli et al., 2009, Shantaram, 2021). There are a wide range of approaches that can be used to effectively isolate proteins from biological samples. However, precipitation techniques are the most commonly used process to concentrate and fractionate protein from biological fluids such as urine (Shantaram, 2021). Studies comparing different precipitation methods have shown that acetone precipitation generated better results including greater protein recoveries, in comparison to other precipitation methods (Shantaram, 2021, Jiang et al., 2004, Saito et al., 2019). This is likely due to the ability of acetone to remove components such as salts, lipids, and pigments that hinder proteomic analyses (Rastegari et al., 2011).

1.8.2. Protein clean up and digestion

Proteomic approaches often require the digestion of complex protein molecules into peptides, that are readily analysed by CE/liquid chromatography (LC) coupled with MS

(Tanca et al., 2013). The ability to effectively solubilise and subsequently identify how much proteins are in a particular sample is the main requirement for most proteomic studies (Tanca et al., 2013). There are various methods for urine processing including the 96DRA-Urine, on-membrane protein capture (MStern Blot), and filter-aided sample preparation methods (Tang et al., 2022, Dupree et al., 2020, Saito et al., 2019). However, these current approaches are limited by their ability to analyse larger sample sizes due to constraints generated by stages including precipitation, centrifugation, and buffer exchange (Govender et al., 2023).

Recently, a new approach has been created for the preparation and processing of urinary samples (Govender et al., 2023). The method is called urine-Hydrophilic interaction liquid chromatography (uHLC), and is based on direct, on-bead protein capture (from 100 L of urine), clean-up, and protein digestion. The uHLC sample preparation workflow uses standard sample collection practices in clinics or hospitals, is automated (can process up to 96 samples in a run), and can be easily executed in a MS laboratory (Govender et al., 2023). The uHLC method is based on the previously established HILIC-based workflow for tissue and plasma samples, and has been modified for human urine samples (Nweke et al., 2020, Baichan et al., 2023). When compared against the established MStern Blot approach, the uHLC workflow was more rapid and performed better (Govender et al., 2023).

1.8.3. Liquid chromatography–mass spectrometry (LC-MS) for proteomic analysis

Proteomic studies usually involve high throughput MS-based strategies to identify and quantify proteins in a sample (Cui et al., 2022). However, highly sensitive and effective MS analysis of proteomic samples require suitable separations (Yu et al., 2023). MS in combination with various separation and pre-fractionation approaches can be used to detect and quantify proteins (Peng et al., 2018). For example, the standard approach for proteomics research is two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) which separates mixtures of protein by the electrical charges and molecular weight. However, this technique is laborious and inefficient, and not appropriate for large-scale analyses (Cui et al., 2022). By comparison, LC or UHPLC has excellent separation capacity, permitting continuous separation of a larger number of proteins that can be

combined with MS for increased automation (Corbett et al., 2020, Cupp-Sutton and Wu, 2020).

In LC-MS-based proteomics, a sample solute that contains the specific analytes of interest (peptides) are pushed through a stationary phase (chromatography column) by a mobile phase (liquid phase) flowing through at high pressure (Clarke et al., 2014, Thomas et al., 2022). The chemical interaction between the elements of the sample, the stationary and mobile phase influences the various migration rates through the chromatography column, allowing separation of analytes of interest (Clarke et al., 2014). The effluent is then directed to the mass spectrometer after elution from the chromatography column. In the mass spectrometer system, the chromatography column effluent is nebulised, desolvated, and ionised by an ionisation source creating charged particles (Clarke et al., 2014). Electromagnetic fields are applied to these charged particles, allowing them to migrate under high vacuum conditions through a series of mass analysers (Clarke et al., 2014).

Currently, the two main approaches for data acquisition to generate MS proteomic data are data-independent acquisition (DIA) and data-dependent acquisition (DDA). (Hu et al., 2016). The DIA approach includes fragmenting all peptides that are present in a sample, while the DDA approach only fragments certain peptides created through the first cycle of MS (Hu et al., 2016). DIA is superior over DDA for analysing complex proteomic samples due to its many advantageous qualities (Guan et al., 2020). These include its capacity to identify low-abundance peptides, increased specificity, and higher reproducibility (Guan et al., 2020).

Recently, the use of proteomic techniques to detect biomarkers associated with early CKD and disease mechanisms have gained a lot of interest. For example, the CE-MS approach was used to create the CKD273 proteomic biomarker classifier that was able to differentiate CKD cases and healthy controls, using urine samples from participants in Europe, America, and Australia (Good et al., 2010). This classifier permits early detection of kidney disease, that is more sensitive than albuminuria in predicting eGFR decline (Good et al., 2010). The classifier included endogenous peptides from proteins such α 1-antitrypsin (A1AT), serum albumin, fibrinogen alpha chain, collagen alpha-1 (I, II, III) chain, collagen alpha-2 (I)

chain, and uromodulin, all of which demonstrate significant changes in urine concentration in several types of CKD, including those arising from diabetes and hypertension (Good et al., 2010). Proteins such as A1AT and fibrinogen are typically involved in inflammation, while collagen fragments and uromodulin are associated with extracellular matrix remodelling and interstitial fibrosis or tubular atrophy, respectively. Uromodulin is distinctly synthesised in the thick ascending limb epithelium of the kidney; reduced levels of uromodulin likely reflect decreased tubular function indicating tubular injury (Roscioni et al., 2013).

In a more recent study on 3,235 participants from the Chronic Renal Insufficiency Cohort Study (comprising of African/African American, Asian/Pacific Islander/Native American, Hispanic and non-Hispanic White participants), 100 circulating proteins were significantly associated (after Bonferroni adjustment) with a decline in eGFR or kidney failure over 10 years (Dubin et al., 2023). Biological pathway analyses revealed significant associations with pathways such as matrix metalloprotease inhibition, metabolic processes, inflammation and fibrosis (Dubin et al., 2023). In another study on 229 CKD patients of European and African ancestry, a combination of five proteins (neutrophil gelatinase-associated lipocalin chemokine ligand 2, epidermal growth factor, transforming growth factor alpha, and kidney injury molecule-1) improved the prediction of GFR decline compared with albuminuria (Bienaimé et al., 2023). The mean area under the curve (AUC) for these five proteins were 0.72 (95% CI: 0.652–0.795) (Bienaimé et al., 2023).

In a cross-sectional analysis of 956 White and Black participants from the North-West, South Africa, eGFR was positively associated with the UMOD protein ($P=0.002$), and negatively associated with the CKD273 classifier ($P<0.001$) (Degenaar et al., 2023). Individuals with lower levels of eGFR were present in the upper 25th percentile of the CKD273 classifier, indicating kidney dysfunction (Degenaar et al., 2023).

1.9. The African Research on Kidney Disease (ARK) Study

The African Research Kidney Disease (ARK) study is a cohort study of adults of self-identified black ancestry from Bushbuckridge, Mpumalanga, South Africa (Fabian et al.,

2022). The research of this thesis analysed DNA and urine samples from a subset of participants from the South African cohort.

Participants aged 20–79 years and who were able to give informed consent were included in the ARK study. Demographic factors included age, sex, employment and livelihood, and place of residence. A comprehensive CKD risk questionnaire was used to assess medical history and treatment for communicable and non-communicable diseases) such as HIV, diabetes, HT, cardiovascular disease and CKD, as well as prior and existing use of traditional medicines. Physical examination included height, weight, waist and hip circumference, and calculation of BMI. Blood pressure was classified according to the 7th Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (Chobanian et al., 2003). Blood pressure was measured in triplicate and the average of the last two readings were defined as the true blood pressure (Fabian et al., 2022). The isotope dilution mass spectrometry traceable Jaffe method was used to measure serum and urine creatinine (Craik et al., 2023). GFR (ml/min/1.73 m²) was determined using the 2009 CKD-EPI creatinine equation without adjustment for African American ethnicity (George et al., 2019). Immunoturbidimetry was used to measure urine albumin concentration (Craik et al., 2023). This aforementioned measurement was used to calculate UACR (mg/mmol) (Craik et al., 2023). Individuals were screened twice to confirm CKD diagnosis, at visit 1 and visit 2 that were approximately five months apart.

In the ARK study of 2,021 black adults, the WHO age-standardised prevalence for CKD (defined as albuminuria and/or low eGFR) was 6.7% (95% CI: 5.4–7.9) (Fabian et al., 2022). The prevalence for low eGFR (defined as <60ml/min/1.73m²) was 0.9% (95% CI: 0.4–1.4), and albuminuria (UACR ≥3.0mg/mmol) was 6.2% (95% CI: 5.0–7.4) (Fabian et al., 2022).

1.10. The Africa Wits-INDEPTH partnership for Genomic Studies (AWI-Gen)

AWI-Gen is a Collaborative Centre of the H3Africa Consortium that aims to understand genomic and environmental factors influencing cardiometabolic diseases in African populations (Ali et al., 2018, Ramsay et al., 2016). In this thesis, we used genetic data from

the AWI-Gen study cohort as the discovery data set to compute PGSs for eGFR, UACR, and albuminuria in CHAPTER 3.

The AWI-Gen cohort has participants aged between 40–60 years from AWI-Gen centres in three different regions within Africa: West Africa (Nanoro and Navrongo) East Africa (Nairobi), and South Africa (Agincourt, Dikgale and Soweto) (Ali et al., 2018, Ramsay et al., 2016). In the AWI-Gen study on 10,702 black participants, the age-standardised population prevalence for CKD was 10.7% (95% CI: 9.9–11.7) (George et al., 2019). The prevalence for low eGFR was 2.4% (95% CI: 2.1–2.8), and albuminuria was 10.7% (95% CI: 9.9–11.7) (George et al., 2019).

1.11. Study rationale, aims and objectives

CKD is a global public health issue, with high prevalence and mortality in low- and middle-income countries in regions such as SSA. Genetic risk for CKD and the impact of gene–environment interactions remain understudied in SSA and South Africa. Risk profiles vary by region, with HT predominating as a risk factor in SSA. In SSA, resources are scarce for treating individuals with CKD who progress to kidney failure, resulting in fatality for untreated individuals. Hence, public health policy needs to focus on identifying CKD early and implementing cost-effective and appropriate interventions to delay progression, for example, advocating for the use of SGLT-2 inhibitors which have been shown to improve kidney function in diabetic and CKD patients (Zoungas and de Boer, 2021).

1.11.1. Hypothesis

Genetic factors (taking into consideration HT, diabetes and HIV) and urinary proteomic markers will be useful in determining a risk profile for CKD in a rural South African setting.

1.11.2. Aim

To determine genetic risk for CKD (taking into account HT, diabetes and HIV), and to compare urinary proteomic markers of those with combined HT and albuminuria, with controls who have neither, in a South African rural setting.

1.11.3. Objectives

The objectives were as follows:

1. Conduct a systematic review on the use of omics for diagnosing and predicting progression of CKD (**CHAPTER 2**)
 - Evaluate existing literature for potential biomarkers (apart from eGFR and albuminuria) that could facilitate early detection of CKD, and predict its progression
2. Assess the predictive accuracy of PGS for CKD and kidney function markers in an African setting (**CHAPTER 3**)
 - Explore the ability of using summary statistics from three discovery datasets (two from multi-ethnic GWAS studies and one was from a continental African GWAS study) to develop and test the predictive power of a PGS for CKD and kidney function markers in an African setting
3. Compare urinary proteomic profiles between cases (having both hypertension and albuminuria) and controls (having neither condition) (**CHAPTER 4**)
 - Identify potential proteins and pathways involved in hypertension-associated albuminuria by assessing urinary proteomic profiles in black South African participants with combined hypertension and albuminuria compared to those who have neither condition

1.12. Thesis outline

This PhD thesis follows the University of Witwatersrand, Faculty of Health Sciences, guidelines in the ‘divided block’ format. This format comprises a detailed background (CHAPTER 1), followed by 3 blocks, each written as a chapter and/or in the format of a paper (1. Systematic review on the omics and CKD [published, CHAPTER 2]; 2. Genetic study of PGS [CHAPTER 3] and 3. Proteomics study [published, CHAPTER 4]), followed by a final chapter on an integrated discussion and main conclusions (CHAPTER 5).

1.13. Ethical Considerations

This study was approved by the Human Research Ethics Committee (HREC) (Medical) of the University of the Witwatersrand (clearance certificate number: M200331). This was a sub-study to the ARK study (clearance certificate number: M170583). AWI-Gen GWAS

data were used in accordance with the H3Africa policies and guidelines (www.h3africa.org) (See Appendix A.4 Ethics Approval Certificate).

CHAPTER 2

2. THE USE OF 'OMICS FOR DIAGNOSING AND PREDICTING PROGRESSION OF CHRONIC KIDNEY DISEASE: A SCOPING REVIEW

This work was raised as a paper publication (Govender et al., 2021). This review was published in the journal *Frontiers in Genetics* in 2021.

2.1. Abstract

Globally, chronic kidney disease (CKD) is an emerging public health problem. Various 'omics platforms have been used to provide insights into the molecular basis of kidney dysfunction. This scoping review is a synthesis of the current literature on the use of different 'omics platforms to identify pathophysiological processes leading to CKD and to detect biomarkers that could be used to detect early-stage CKD and predict disease progression. This review was conducted according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews protocol. Seventy-six studies published from January 2007 to May 2020 were included following a structured selection process. The proteomic approach was the most common 'omic platform (n=41), followed by metabolomics (n=22), genomics (n=5), epigenomics (n=3), transcriptomics (n=2) and three studies used a combination of 'omics platforms. The majority of studies (n=58) reported on CKD associated with diabetes mellitus. Fourteen studies that provided information on the pathogenic mechanisms of CKD and biomarkers that play a role in early detection are discussed. These biomarkers were mainly used to detect CKD and predict progression in European populations, highlighting the need for 'omics research in African populations since disease etiologies may differ across ethnic groups. To assess the transferability of findings, it is essential to do research in diverse populations.

Keywords: CKD, 'omics, biomarkers, early detection, diagnosis

2.2. Introduction

The global prevalence of chronic kidney disease (CKD) is estimated to be 13.4%, and thought to be higher in low- and middle-income settings, such as sub-Saharan Africa (SSA) (15.8%) (Hill et al., 2016, Kaze et al., 2018). End stage kidney disease (ESKD) is the most severe stage of CKD, at which point the damage is irreversible and support in the form of chronic dialysis or a kidney transplant is required to sustain life. Major risk factors for CKD vary by setting, with hypertension (HT) being the most common in east Asia, tropical Latin America, and Western and Southern SSA, whereas diabetes predominates as a risk factor in the remaining regions of the world (Bikbov et al., 2020). Endemic and other infectious diseases, for example schistosomiasis, malaria, and human immunodeficiency virus (HIV) still contribute substantial additional risk for CKD in high prevalence countries (Couser et al., 2011, Ekrikpo et al., 2018). The diagnosis of CKD is currently based on estimated glomerular filtration rate (eGFR) and/or albuminuria (Levey et al., 2002). However, low GFR defined as $eGFR < 60 \text{ ml/min/1.73 m}^2$, is a relatively late marker of kidney disease, as organ damage may precede functional changes such as filtration impairment. Even though albuminuria is suggested to be an early marker of kidney disease, kidney disease may still progress despite the absence of albuminuria (Miller et al., 2009). Thus, there is a need for alternative markers of kidney disease that identify early disease and predict risk for progression. Such markers would enable interventions to prevent or slow progression of CKD, which is particularly relevant in resource-limited settings where access to kidney replacement therapy for ESKD remains severely restricted.

‘Omics technologies refer to collective analyses of cell populations, tissues, organs or the whole organism at the molecular level (Horgan and Kenny, 2011). In the last decade, there have been significant advances in use of ‘omics platforms to investigate susceptibility to, and detection of CKD, and related risk factors (Abbiss et al., 2019, Cañadas-Garre et al., 2019, Good et al., 2010). Omics technologies permit high-throughput, comprehensive exploration of the genome (DNA sequence), epigenome (epigenetic modifications), proteome (proteins), transcriptome (transcribed RNA) and metabolome (metabolites), through a wide range of platforms including next-generation sequencing, protein and mRNA arrays and mass spectrometry (Papadopoulos et al., 2016). The past decade has seen significant advances in genome-wide association studies (GWAS) for CKD and kidney

function traits, with study sample sizes increasing from >100,000 to 1,000,000 facilitating the identification of >250 genetic loci robustly associated with CKD in different ethnic groups (Tin and Köttgen, 2020). Proteomic approaches to identify urinary biomarkers for early detection and prediction of progression of CKD are also on the rise, attractive (in part) for their non-invasive use in clinical settings (Mischak et al., 2010). For example, capillary electrophoresis coupled with mass spectrometry (CE-MS) has been used to develop a proteome-based urine biomarker panel of 273 peptides with profiles that differed significantly between individuals with CKD and healthy controls (Good et al., 2010). These peptides were combined to develop a single score known as the CKD273 risk score (Good et al., 2010). Previous studies defined a CKD273 risk score cut-off of 0.343 to predict diabetic nephropathy (DN) (Alkhalaf et al., 2010). This threshold was lowered to 0.154 to accommodate the early detection of CKD (Lindhardt et al., 2016).

The overall aim of this scoping review was to evaluate existing literature for potential biomarkers—aside from those currently used, namely eGFR and albuminuria—that might facilitate early detection of CKD, and predict its progression. The specific objectives were to 1) describe potential biomarkers identified through different ‘omics platforms (genomics, proteomics, metabolomic, transcriptomics and epigenomics) that are associated with CKD, and 2) to explore potential pathways involved in CKD that could inform novel treatment strategies for CKD. This is pertinent for developing affordable diagnostic, prevention, and therapeutic strategies for countries with high prevalence of CKD and limited resources, such as are found in SSA.

2.3. Methods

2.3.1. Protocol and registration

This review was conducted using the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) approach (de Oliveira, 2018). The background and methods were documented in a protocol in the Open Science Framework (OSF) database (Doi:10.17605/OSF.IO/QM64P, <https://osf.io/sw8ne/>).

2.3.2. Search Strategy

Relevant studies published from January 2007 to 18 May 2021 (Search date: 18 May 2021) were identified through a comprehensive search of electronic databases including Embase,

Web of Science, PubMed, SCOPUS and African Index Medicus. The electronic database search was conducted using medical subject heading (MeSH) terms, keywords pertinent to the topic, and Boolean operators such as AND/OR to obtain the maximum number of studies (Table 2.1).

Table 2.1: Search strategy used in database searches

Search no.	Search query
#1	“chronic kidney disease” OR “CKD” OR “GFR” OR “glomerular filtration rate” OR “nephropathy” OR “serum creatinine” OR “urinary albumin” OR “ACR” OR “albumin creatinine ratio” OR “albuminuria” OR “urinary albumin creatinine ratio” OR “UACR” OR “proteinuria” OR “protein creatinine ratio” OR “urinary protein creatinine ratio” OR “UPCR”
#2	“diabetes” OR “hypertension” OR “infection” OR “infectious disease” OR “HIV” OR “human immunodeficiency virus” OR “malaria” OR “schistosomiasis” OR “cytomegalovirus” OR “CMV” OR “JC virus” OR “John Cunningham virus” OR “polyoma virus” OR “tuberculosis” OR “TB” OR “hepatitis B virus” OR “hepatitis C virus”
#3	“genomic” OR “metabolomic” OR “proteomic” OR “transcriptomic” OR “epigenomic”
#4	1# AND #2 AND #3

2.3.3. Screening and eligibility criteria

Database search records were combined and all duplicate records removed. One investigator (MG) screened all records by title and abstract for inclusion, and articles with irrelevant topics were excluded. Investigators (J-TB and JF) independently screened records by title and abstract for verification. Any inconsistencies or disagreements were resolved by review with investigator MR. This was followed by screening full-text articles

according to the eligibility criteria. The main inclusion criterion was that a study had to be an original study containing ‘omics data used to detect CKD and related risk factors. The studies had to report on the characteristics of the study population, type of ‘omics technique used, CKD or kidney function traits, relevant risk factors (defined in **Table 2.2**), and potential biomarker or indicator of early detection of CKD and prediction of progression (methylation profile, genes/genetic variants, transcripts, proteins and/or metabolites). Since the first large GWAS was conducted in 2007 (Craddock and Jones, 2007), we chose to include only studies published from January 2007 to May 2021. The age limit was restricted to participants 18 years and older and only studies published in English were included. Studies were excluded if the study was a review, meta-analysis, conference abstract or presentation, book chapter or response to treatment. Studies involving animals and paediatric participants were also excluded. The reasons for excluding each article were recorded.

Table 2.2: Classification of CKD

Class	Definition
Normal kidney function	UACR <30 mg/g and GFR>60 mL/ min/1.73 m ²
Microalbuminuria	UACR 30-300 mg/g
Macroalbuminuria	UACR >300 mg/g
CKD stage 1 and 2	eGFR ≥60 ml/min per 1.73 m ²
CKD stage 3	eGFR <60 ml/min per 1.73 m ²
CKD stage 4 and 5	eGFR <30 ml/min per 1.73 m ²
ESKD	eGFR <15 ml/min per 1.73 m ²

CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; ESKD, end stage kidney disease; UACR, urine albumin-to-creatinine ratio.

2.3.4. Data extraction

A data extraction sheet was created on Microsoft Excel (version 16.42, available at: <https://office.microsoft.com/excel>) to ensure all relevant data were extracted. Data extraction was performed by MG, and any discrepancies were resolved by J-TB, JF and MR. Data extracted for each article included the digital object identifier (DOI), name of first author, year of publication and title, participant/population characteristics (sample size, ethnicity, stage of CKD, presence of risk factors and/or infections), type of ‘omics technology, ‘omics technique, biological sample and potential biomarkers (genes, variants,

miRNAs, mRNAs, proteins, and peptides) associated with early detection of CKD and prediction of progression. As PRISMA-ScR guidelines recommend, we conducted a qualitative synthesis to summarise the main components of the CKD ‘omics field.

2.4. Results

The database searches identified 4,793 articles, of which 1,291 were duplicates. Through title and abstract screening, we excluded 3,351 articles mainly due to the following reasons: irrelevant topic, year of publication outside the period for this review, and animal studies. A further 28 articles were excluded after assessing the full-text and applying the eligibility criteria. Thus, one hundred and twenty-three full-text articles were included in this review (**Figure 2.1**, see Appendix B.1 List of 123 articles).

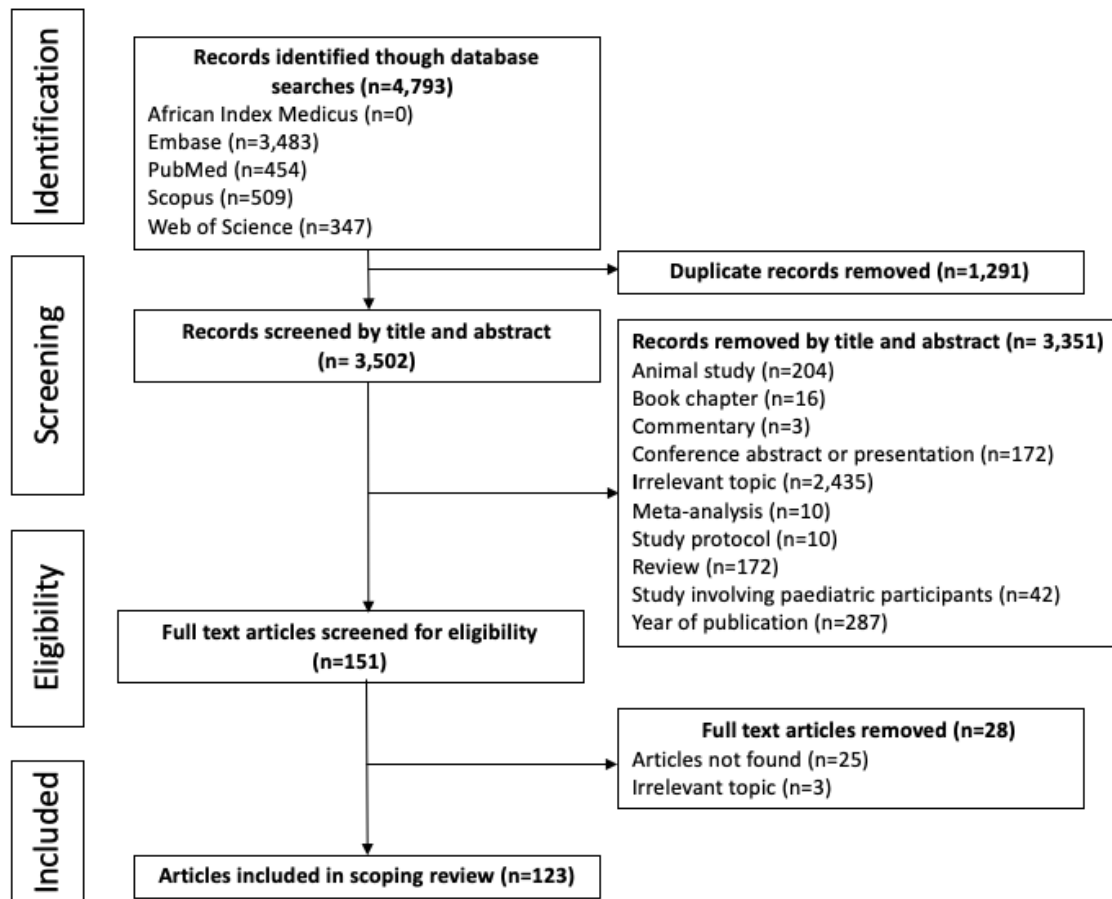


Figure 2.1: Article selection process according to PRISMA-ScR guidelines

PRISMA-ScR, Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews.

2.4.1. Participant and study characteristics

The participant and study characteristics of all studies (n=123) included in this scoping review are shown in **Table 2.3**. There was an increase in publications from 2015. Majority of studies (28.5%) had sample sizes greater than 500. Most of the participants (42.3%) were between 51 and 60 years of age. Most of the most common ethnic group represented was European ancestry (15.4%), followed by Asian (12.2%), African ancestry (6.5%), Indian (3.3%), and studies reporting on multiple ethnicities (10.6%). It is important to note that ethnicity was not reported in over half of the studies (52.0%). Of eight studies that included participants of self-reported African ethnicity, only one included African individuals residing in Africa (Chen et al., 2007).

Table 2.3: Participant and study characteristics

Characteristic	Number of studies, n (%)
Year of publication	
2007–2010	22 (17.9)
2011–2015	28 (22.8)
2016–May 2021	73 (59.3)
Number of participants per study	
1–10	2 (2.6)
11–50	22 (17.9)
51–100	20 (16.3)
101–200	26 (21.1)
201–500	18 (14.6)
>500	35 (28.5)
Mean age of participants	
18–30 years	0
31–40 years	11 (8.9)
41–50 years	11 (8.9)
51–60 years	52 (42.3)
61–70 years	32 (26.1)
>70 years	2 (1.6)
Unknown	15 (12.2)
Sex	
Males only	5 (4.0)
Females only	0
Both	98 (79.7)
Not stated	20 (16.3)
Participant ethnicity	
Asian only	15 (12.2)
African ancestry only	8 (6.5)
European ancestry only	19 (15.4)

Indian only	4 (3.3)
Studies including multiple ethnicities (European ancestry, African ancestry, Hispanic, Indian and Asian)	13 (10.6)
Unknown	64 (52.0)

2.4.2. Omics' platforms and techniques

We reported on five different 'omics technologies including genomic, epigenomic, proteomic, metabolomic and transcriptomic (**Table 2.4**). The proteomic approach was the most common 'omics platform (n=53), followed by metabolomics (n=30), genomics (n=17), epigenomics (n=10), transcriptomics (n=5) and eight studies used a combination of 'omics platforms. A wide range of techniques were used to analyse the 'omics data (**Table 2.4**). Genomic analysis was mainly performed by arrays including Illumina's High-throughput Bead array, Human Core Exome Bead array, HumanHap550-Duo Bead Chip and Omni Express Exome array. Proteomic analysis was mainly done by liquid chromatography coupled with mass spectrometry (LC-MS) and CE-MS. Gas chromatography-mass spectrometry (GC-MS) and LCMS were important analytical techniques for metabolomic analysis. Methylation arrays (including Illumina's Infinium 27, 450 and 850K methylation arrays) were key for epigenomic analysis and transcriptomic analysis was mainly performed by RNA sequencing.

Table 2.4: The different 'omics technologies and techniques reported in this scoping review for detection of CKD and associated risk in 123 studies

Characteristic	Number of studies, n (%)
Genomic Affymetrix 370K and 500k arrays Illumina High-throughput Bead array Illumina HumanCoreExome Bead array Illumina HumanHap550-Duo BeadChip Illumina's IV panel Illumina OmniExpressExome array Panel of 372 autosomal short tandem repeat markers Variety of GWAS arrays	17 (13.8)
Epigenomic Illumina Infinium 27K methylation array Illumina Infinium HumanMethylation450 Beadchip Illumina Infinium MethylationEPIC BeadChip (HM850K) Reverse phase high pressure liquid chromatography (RP-HPLC) analysis and bisulfite sequencing	10 (8.1)

SuperTAG methylation-specific digital karyotyping	
Proteomic 1 and 2-dimensional gel electrophoresis (1 and 2DE) and mass spectrometry (MS) Capillary electrophoresis coupled with mass spectrometry (CE-MS) Fluorescence two-dimensional (2-D) differential in-gel electrophoresis (DIGE) Isobaric tag for relative and absolute quantitation iTRAQ MS Liquid chromatography coupled with mass spectrometry (LC-MS) Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) Olink cardiovascular panel Protein chip arrays Proximity Extension Assay (PEA) Surface-enhanced laser desorption/ionization mass spectrometry (SELDI-MS)	53 (43.1)
Metabolomic Commercial kits CE-MS Gas chromatography coupled with mass spectroscopy (GC-MS) LC-MS Nuclear magnetic resonance (NMR)	30 (24.4)
Transcriptomic GeneChip human genome series U133A and Plus 2.0 Array RNA sequencing	5 (4.1)
Genomic and metabolomic Infinium Multi-Ethnic Global BeadChip array Exome BeadChip array LC-MS	2 (1.6)
Genomic and epigenomic Illumina Human MethylationEPIC BeadChip	1 (0.8)
Proteomic and metabolomic ELISA Assays LC-MS SOMAscan	2 (1.6)
Transcriptomic and epigenomic Illumina 450K methylation array RNA sequencing	1 (0.8)
Transcriptomic and metabolomic Affymetrix transcriptomics arrays NMR	2 (1.6)

2.4.3. CKD risk categories

The most common risk factors associated with CKD were diabetes mellitus (DM) (either type 1 diabetes, type 2 diabetes or both) (n=91), followed by hypertension (n=4), cardiovascular disease (n=2), and infection (HIV) (n=2). Sixteen studies reported on more than one risk factor and eight studies did not report on any risk factors (**Figure 2.2**).

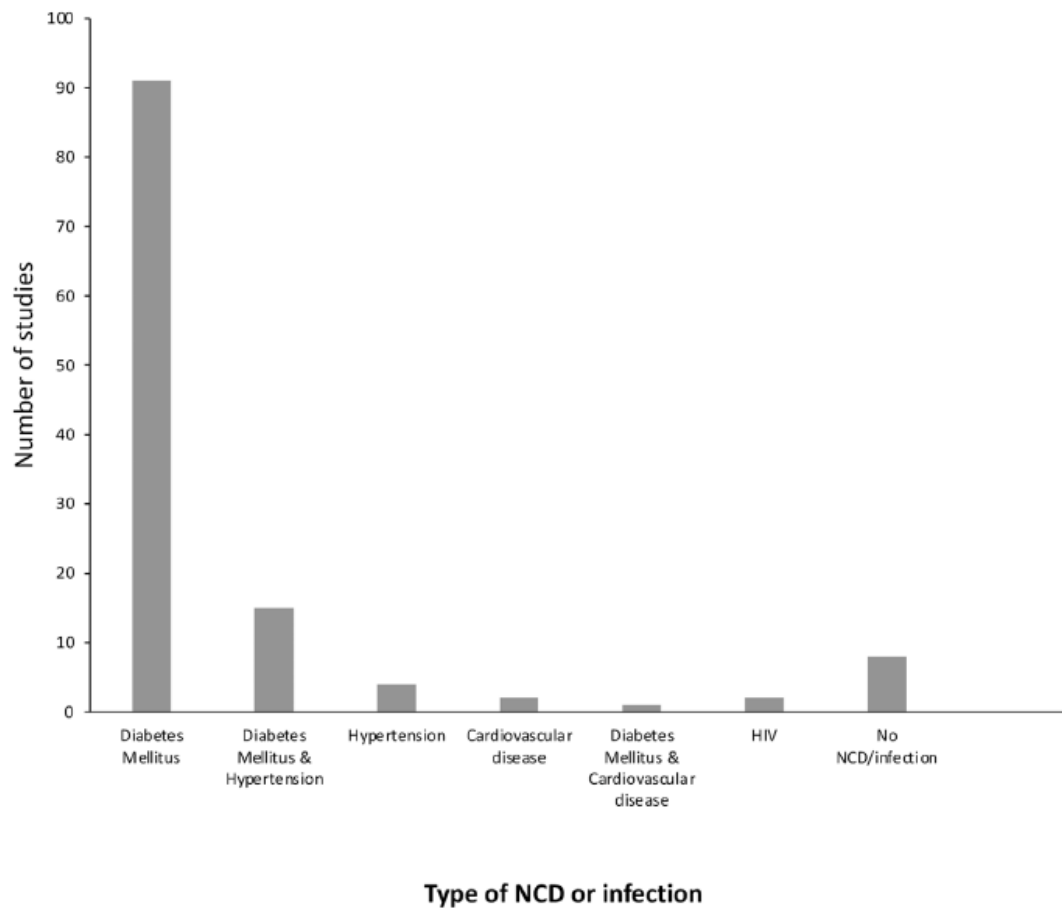


Figure 2.2: Bar graph showing number of papers with each of the comorbidities associated with CKD

CKD, chronic kidney disease, NCD, non-communicable disease

2.4.4. CKD risk categories

Whilst it was not feasible to discuss each biomarker identified for each study across all ‘omics platforms, a list of selected studies that identified biomarkers/indicators involved in risk, early detection and prediction of progression, and those that explored potential disease pathways are shown in Table 2.5.

2.4.4.1. Genetic variants associated with risk for CKD

The genes implicated in risk for CKD were non-muscle myosin heavy chain 9 (*MYH9*)/apolipoprotein L1 (*APOL1*) and uromodulin (*UMOD*). Kao et al. (2008) performed a genome-wide admixture scan in 2,178 individuals with European and African ancestry. They identified three SNPs (rs16996674, rs16996677, and rs5756152) in the *MYH9* gene that were associated with a ~3 times greater risk of nondiabetic ESKD in African Americans

compared to European Americans. Similarly, Shlush et al. (2010) identified two SNPs (rs7286127 and rs5756133) associated with nondiabetic ESKD on chromosome 22 in the region that contains the *MYH9* gene, as well as the neighbouring *APOL1* gene. In addition, Igo Jr et al. (2011) identified one SNP (rs5769116) 10 Mb distal to *MYH9/APOL1* that was associated with DN. Morris et al. (2019) performed an association study of eGFR, a measure of kidney function, in 312,468 individuals of diverse ancestry. They identified a highly significant SNP (rs77924615; $P=1.5\times 10^{-54}$) in the *UMOD* gene associated with eGFR.

2.4.4.2. Biomarkers associated with early detection of CKD

The proteomic CKD273 biomarker panel and metabolite pantothenic acid (PA) could potentially play a role in the early detection of CKD. Three proteomic studies by Siwy et al. (2014), Øvrehus et al. (2015), and Lindhardt et al. (2017) assessed the ability of a proteomic CKD273 biomarker panel to predict DN with diagnostic accuracy, measured as area under receiver operating characteristic curve, ranging from 0.84 to 1.0. All three proteomic studies used the CE-MS approach with sample sizes ranging from 35 to 737. In addition, Gonzalez-Calero et al. (2016) identified the metabolite PA as a potential predictor of albuminuria development in 118 hypertensive patients.

2.4.4.3. Biomarkers associated with progression of CKD

The proteomic CKD273 biomarker panel, metabolite PA, and methylation probe cg17944885 were suggested to be involved in progression of CKD. Two proteomic studies, Roscioni et al. (2013) and Tofte et al. (2020) found that a high CKD273 risk score was able to predict progression of microalbuminuria with an odds ratio (OR) of 1.35 and P value of $<1.0\times 10^{-4}$, respectively. In addition to the metabolite PA being a potential predictor of albuminuria development in hypertensive patients, it was also associated with DN progression ($P=5.0\times 10^{-3}$) (Ma et al., 2020). Lecamwasam et al. (2020) compared methylation profiles between early and late stages of DN. They found that the methylation probe cg17944885 was associated with progressive renal dysfunction ($P=0.01$ after adjustment for multiple testing).

2.4.4.4. Biomarkers exploring potential disease pathways

Genes of the retinoic acid pathway, *UMOD* and the metabolite PA provided insight into potential pathways involved in CKD. *UMOD* was specifically expressed in epithelial cells of the ascending loop of Henle in the kidney. This points to the role of *UMOD* in kidney

physiology (Morris et al., 2019). Genes of the retinoic acid pathway are essential for retinoic acid (RA) synthesis and impairment of RA synthesis leads to progression of kidney disease (Li et al., 2014). Pathway analysis of PA revealed that altered PA synthesis and PA-related metabolism could lead to DN (Ma et al., 2020).

Table 2.5: Studies that identified biomarkers/indicators for early detection of CKD, prediction of CKD progression, and disease pathways

Author and year	Type of 'omic platform	'Omic technique	Sample size	Biomarker/indicator of disease	Effect estimate OR/P value	Clinical phenotype/outcome	What does the biomarker inform on?
(Kao et al., 2008)	Genomic	1,354 SNP panel	2,178	rs16996674 (<i>MYH9</i>) rs16996677 (<i>MYH9</i>) rs5756152 (<i>MYH9</i>)	3.10 3.03 2.59	Non-diabetic ESKD	Susceptibility/risk
(Shlush et al., 2010)	Genomic	2,016 SNP panel	833	rs7286127 (chromosome 22) rs5756133 (chromosome 22)	$\leq 10^{-3}$ $\leq 10^{-3}$	Non-diabetic ESKD	Susceptibility/risk
(Igo Jr et al., 2011)	Genomic	Illumina IV panel of 5,500 SNPs	3,972	rs5769116 (10 Mb distal to <i>MYH9/APOL1</i>)	1.5×10^{-3}	DN	Susceptibility/risk
(Morris et al., 2019)	Genomic	A variety of GWAS arrays	312468	rs77924615 (<i>UMOD</i>)	1.5×10^{-54}	CKD	Susceptibility/risk /disease pathway
(Fan et al., 2020)	Transcriptomic	RNA-sequencing	37	<i>RDH8</i> <i>RDH12</i> <i>RBP4</i>	0.02 0.03 0.03	DN	Disease pathway

(Øvrehus et al., 2015)	Proteomic	CE-MS	35	CKD273	AUC of 0.977*	DN and Hypertension associated kidney disease	Early detection
(Siwy et al., 2014)	Proteomic	CE-MS	165	CKD273	AUC of 0.95 to 1.0* <0.05	DN	Early detection
(Lindhardt et al., 2017)	Proteomic	CE-MS	737	CKD273	AUC of 0.84* <1.0x10 ⁻⁴	DN	Early detection
(Roscioni et al., 2013)	Proteomic	CE-MS	88	CKD273	1.35	DN	Progression
(Tofte et al., 2020)	Proteomic	CE-MS	1,775	CKD273	<1.0x10 ⁻⁴	DN	Progression
(Gonzalez-Calero et al., 2016)	Metabolomic	Nuclear magnetic resonance (NMR)	118	Pantothenate	<1.0x10 ⁻⁴	Hypertension associated kidney disease	Early detection
(Ma et al., 2020)	Metabolomic	Liquid chromatography-mass spectrometry	68	Pantothenate	5.0x10 ⁻³	DN	Progression/ disease pathway
(Lecamwasam et al., 2020)	Epigenomic	Illumina Infinium Methylation EPIC BeadChip	119	cg17944885	0.01	DN	Progression

*Diagnostic accuracy measured as area under receiver operating characteristic curve.

CKD, chronic kidney disease; DN, diabetic nephropathy ; ESKD, end stage kidney disease; OR, odds ratio.

2.5. Discussion

In this scoping review the first objective was to identify studies on potential biomarkers associated with CKD. The biomarkers that were most commonly associated with risk or early detection of CKD are genes and genetic variants in chromosomes 16 and 22 (Kao et al., 2008, Shlush et al., 2010, Igo Jr et al., 2011, Liao et al., 2014, Taira et al., 2018, Morris et al., 2019), the CKD273 biomarker panel (consisting of proteins such as collagen fragments, uromodulin [UMOD], α -1-antitrypsin, transthyretin and β -2-microglobulin) (Lindhardt et al., 2017, Øvrehus et al., 2015, Siwy et al., 2014), and the metabolite PA (also known as pantothenate or vitamin B5) (Gonzalez-Calero et al., 2016). Majority of these potential biomarkers are exploratory, having been identified in small discovery studies and have not been validated, with the exception of the CKD273 biomarker panel.

It is well known from studies in the USA that individuals with African ancestry are at greater risk for developing incident CKD and progressing to ESKD, than other population groups (Tarver-Carr et al., 2002, Perneger et al., 1995, Saran et al., 2020). Much of this risk was attributed to genetic variants in the *MYH9/APOL1* gene region on chromosome 22 that were strongly associated with non-diabetic ESKD (Kao et al., 2008, Shlush et al., 2010). The *MYH9* and *APOL1* genes are tightly linked on chromosome 22, and fine mapping studies revealed that the risk variants were actually in the *APOL1* gene and not in the *MYH9* gene, as originally thought (Tzur et al., 2010, Genovese et al., 2010). Zhang et al. (2016) developed and validated an *APOL1* genotyping assay as a precision medicine genetic test for non-diabetic CKD risk prediction (Zhang et al., 2016). There are still ongoing clinical trials (<http://www.clinicaltrials.gov>; identifiers NCT02234063, NCT04191824 and NCT04910867), assessing the clinical utility of *APOL1* genetic testing. Implementation of an *APOL1* genotyping assay in a clinical setting could lead to a cost-effective screening program for African Americans who at high risk for developing CKD and progressing to ESKD.

Studies have suggested the uromodulin (*UMOD*) gene located on chromosome 16 as a candidate gene for CKD (Köttgen et al., 2009, Köttgen et al., 2010, Pattaro et al., 2016). Not only does it have a large effect on eGFR decline and CKD risk, the consistency of the effect is seen across different ethnicities (Devuyst et al., 2017). Morris et al. (2019)

identified rs77924615 ($P=1.5 \times 10^{-54}$), located in *UMOD*, as the main variant driving eGFR association. This variant was associated with decreased eGFR and increased *UMOD* expression (Morris et al., 2019).

Due to the multifactorial nature of CKD, a combination approach of several biomarkers is likely to be more predictive than a single biomarker. Compared to currently available markers such as reduced eGFR and albuminuria, the CKD273 biomarker panel enables more accurate diagnosis and earlier detection of CKD. For example, in one study, CKD273 enabled detection of CKD with 95% sensitivity, 100% specificity and an excellent diagnostic accuracy of 0.977 (95% confidence interval (CI) 0.930–1.000) (Øvrehus et al., 2015). In another study, CKD273 had the ability to detect DN in individuals with type 2 diabetes, with diagnostic accuracy ranging from 0.95 to 1.0 across 9 study sites (Siwy et al., 2014). This same biomarker panel was able to predict the development of microalbuminuria in normoalbuminuric type 2 diabetic patients, with a diagnostic accuracy of 0.79 (95% CI 0.75–0.84; $P < 1.0 \times 10^{-4}$) (Lindhardt et al., 2017). The CKD273 biomarker panel includes clinically relevant proteins such as uromodulin and β -2-microglobulin, which are suggested to be markers of early renal damage (Torffvit et al., 1999, Torffvit and Agardh, 1993, Schlatzer et al., 2012, Alaje Abiodun et al., Zeng et al., 2014). These findings demonstrate the utility of CKD273 in early non-invasive renal risk assessment and may help to prevent unnecessary biopsies. While CKD273 has been validated for DN, its utility for hypertension-associated CKD is yet to be established. The CKD273 biomarker can assist clinicians in decision-making for more frequent check-ups in high-risk patients and guide therapeutic options to prevent long-term chronic conditions. Although the CKD273 biomarker panel has been supported by the FDA (https://www.fda.gov/files/drugs/published/Biomarker-Letter-of-Support_Mischak.pdf), its utility has yet to be established in African populations and the African diaspora. Furthermore, the current cost of screening using the CKD273 biomarker panel is €3,053 per patient making it unaffordable in resource-poor settings like SSA (Critselis et al., 2018).

CKD273, PA, and methylation probe cg17944885 are suggested to be potential biomarkers associated with CKD progression. In addition to PA being used to predict albuminuria in hypertensive normoalbuminuric patients, it is suggested to be a promising biomarker for

predicting diabetes induced kidney impairment (Ma et al., 2020, Gonzalez-Calero et al., 2016). Methylation marker cg17944885 showed an increasing gradient of methylation with progressive renal dysfunction (Lecamwasam et al., 2020). A high CKD273 risk score was independently associated with faster progression of albuminuria in two separate studies (Tofte et al., 2020, Roscioni et al., 2013). The ability to identify individuals at high-risk for progression has important clinical implications. For those patients at a high-risk for progression, it is suggested that blood pressure be controlled to less than 130/80 mmHg in hypertensive patients. There are many therapeutic options to control blood pressure including angiotensin-converting enzyme (ACE) inhibitors, angiotensin-receptor blockers (ARBs) or a combination of ACE inhibitors or ARBs with a diuretic (Nguyen et al., 2010). Meanwhile, management of glucose and lipid levels are also suggested to delay the development of chronic complications in DN patients.

The second objective of the scoping review was to explore potential pathways involved in CKD. Using expression quantitative trait loci (eQTL) analysis, UMOD expression was mapped to epithelial cells lining the thick ascending limb of the loop of Henle (Morris et al., 2019). Studies on transgenic mice suggested that overexpression of UMOD causes abnormal activation of the sodium-potassium-chloride transporter (NKCC2) involved in sodium chloride (NaCl) reabsorption by the thick ascending limb segment, leading to salt-sensitive hypertension and renal lesions (Trudu et al., 2013). Mice treated with a single dose of furosemide, a diuretic that targets NKCC2, exhibited a significant reduction in blood pressure (Trudu et al., 2013). This finding suggests that NKCC2 is a potential target for hypertension-induced CKD intervention.

Fan et al. (2020) compared gene expression profiles in three groups; controls, early DN and advanced DN) (Fan et al., 2020). They identified genes of the retinoic acid pathway (including retinol dehydrogenase 8 (*RDH8*), retinol dehydrogenase 12 (*RDH12*), and retinol binding protein 4 (*RBP4*)) with increased expression in early DN but decreased expression in advanced DN. These findings suggest that RA might have kidney protective effects in the early stage of DN that slow the progression to advanced DN. A previous study suggested that downregulation of enzymes (retinol dehydrogenase 9 (*RDH9*) and retinol dehydrogenase 1 (*RDH1*)) essential for RA synthesis causes impaired local synthesis of RA

in the kidney that contributes to the progression of kidney disease (Li et al., 2014). These results point to restoration of retinoic acid synthesis as a potential therapy to slow the progression of kidney disease. There is an ongoing clinical trial assessing the efficacy of RA isotretinoin for treatment of various podocyte diseases including collapsing glomerulopathy (<http://www.clinicaltrials.gov>; identifier [NCT00098020](#)). However, the side effects of RA therapy pose a challenge for its use in the clinic (Montesinos et al., 2009).

The metabolite PA is suggested to be a predictive biomarker for development of CKD and progression in hypertensive and diabetic patients (Ma et al., 2020, Gonzalez-Calero et al., 2016). PA (also known as vitamin B5) is an essential vitamin and dietary factor, and has long been recognised as an essential cofactor of biochemical reactions in numerous physiological processes (Patassini et al., 2019). Pathway analysis results showed PA synthesis and PA-related metabolism were disturbed in individuals with DN (Ma et al., 2020). Since mammalian cells are unable to synthesize vitamin B5, they require it to be generated by the intestinal microbiome or to be part of the diet, resulting in a potential therapeutic target for DN treatment (Daugherty et al., 2002). A previous study in a type 2 diabetes rat model has shown that a vitamin B5 supplement restored body weight and blood pressure whilst also decreasing blood glucose levels (Demirci et al., 2014). In addition, this study also reported that a vitamin B5 supplement may reduce diabetes-related oxidative stress and prevent endothelial cell injury (Demirci et al., 2014). However, further studies are required to determine the efficacy of a vitamin B5 supplement in reducing blood pressure and glucose levels in humans.

The translation of ‘omics research into disease-related biomarkers for use in clinical practice is a complicated, time-consuming, and expensive process. Some challenges include clinical heterogeneity of individuals with a particular diagnosis, different standards of evidence to evaluate the validity of the biomarker, and a lengthy developmental pathway from biomarker discovery to clinical utility (Omenn et al., 2012). The diagnostic and therapeutic potential of the biomarker has to be carefully assessed before being considered for introduction into the clinic, and additional uses should be considered. For example, the CKD273 biomarker panel is available as a commercial *in vitro* diagnostic test and is a good example of a “bench-to-bedside” approach (Pontillo and Mischak, 2017a). In addition to its

diagnostic and prognostic value, CKD273 also has the potential to assess the impact (outcome) of therapeutic interventions (Haubitz et al., 2009). For example, in one study, CKD273 risk scores were significantly reduced (from 0.721 at visit 2 to 0.277 at visit 9), below the previously defined threshold (0.343) after 2 years of treatment with irbesartan (an antihypertensive and antiproteinuric drug) (Andersen et al., 2010).

Low-cost genomics-informed approaches to disease risk prediction pose great promise for low-and middle-income countries, where the morbidity and mortality of non-communicable diseases such as CKD are disproportionately high. However, there are many factors that hinder the implementation of a genomic medicine healthcare model for CKD in resource limited settings. These include population-specific research to validate predictive accuracy and clinical utility, a low critical mass of skilled scientists, insufficiently equipped diagnostic research laboratories and a lagging public health policy framework (Tekola-Ayele and Rotimi, 2015). For example, the paucity of genomics research in individuals from Africa has limited knowledge of clinically relevant CKD genetic variants in these populations. The clinical implementation of genomic medicine requires more than just developing a diagnostic or screening test. To have a functional genomic medicine healthcare model for CKD, costly technology and skill-strengthening among scientists in the field of medical genetics, genetic counselling, genetic epidemiology, bioinformatics and computer biology are required. In addition, legal and ethical frameworks also need to be established in the local context (Pereira et al., 2021, Tekola-Ayele and Rotimi, 2015).

The strengths of the review include the novelty of the topic and the assessment of the literature on candidate biomarkers for early detection of CKD and prediction of progression. Several potential biomarkers are highlighted and may be useful in developing screening programs for CKD, however, they need validation in different clinical settings and ethnic backgrounds. In addition, this review has shown that ‘omics technologies can be used to explore potential disease mechanisms and therapeutic targets for treatment. There were some limitations in this scoping review. Despite internationally accepted guidelines, studies used various metrics to define CKD and potential sex differences in the manifestation of CKD were not examined. The sample sizes were usually small and results may not be transferable to populations that were not included in the studies, for example

African populations. Furthermore, it was difficult to compare the outcomes across studies because each study design was unique.

A few knowledge gaps were identified. While DM is a major risk factor for CKD in some regions of the world, reflected in the number of DN studies reported in the ‘omics literature, other regions such as Asia and Africa have other dominant risk factors such as hypertension and others factors that remain to be identified. The major underlying causes of kidney disease in Africa are poorly understood and often attributed to hypertension which is likely to be secondary hypertension as a consequence of numerous insults. In this regard, there are relatively few ‘omics studies that focus on hypertension in the context of CKD in African populations (Alicic et al., 2017). There was a single study with individuals from West Africa, highlighting the lack of ‘omics research in continental Africans (Chen et al., 2007), despite many studies documenting the well-established genetic link to CKD in African Americans (*APOL1*- associated nephropathy) (George et al., 2018).

In sub-Saharan Africa, there are limited resources for treating individuals with CKD who progress to ESKD, rendering this condition fatal for those who are not treated. At least half of the patients who require ESKD treatment do not have access to treatment, and only 1.5% are predicted to receive renal replacement therapy (Ashuntantang et al., 2017).

Approximately 75% of patients who begin dialysis later die due to poor quality of dialysis or termination of dialysis based on its cost (Ashuntantang et al., 2017). This makes screening, early detection and the implementation of preventative interventions for CKD particularly relevant. Clinically relevant biomarkers will be useful to develop screening programs and potential therapeutic approaches to treat CKD in sub-Saharan Africa. Ideally the goal is to identify individuals at risk for developing CKD in the future, and developing preventative intervention strategies that would delay onset and slow progression, by treating the underlying aetiologies.

The reported ‘omics studies identified potential biomarkers associated with CKD. These biomarkers provide insight into the pathways involved in CKD and could possibly be used to guide diagnostic and treatment options. The biomarkers that have emerged through synthesis of existing literature provide evidence for effective approaches for early detection

of CKD and prediction of progression in European populations. Their transferability to Africans and other ethnicities needs to be verified through research in relevant populations and assessment of diagnostic implementation in low-resourced healthcare settings. There is extensive regional and ethnic diversity in Africa, and the major causes of CKD appear to differ substantively from those in high-income settings, requiring population representative studies that address the local context and causes of CKD. In addition, many factors need to be addressed to facilitate effective genomic medicine healthcare models for CKD.

2.6. Author Contributions

MG screened all records by title and abstract for inclusion. J-TB and JF independently screened records by title and abstract for verification. Any inconsistencies or disagreements were resolved by mutual agreement including investigator MR. Writing of the paper was done by MG. MR, JF, and J-TB contributed to editing the paper.

2.7. Funding

MG has a bursary from the South African National Research Foundation (NRF); J-TB's postdoctoral fellowship was funded by the NRF and MR is a South African Research Chair in Genomics and Bioinformatics of African populations hosted by the University of the Witwatersrand, funded by the Department of Science and Technology, and administered by the NRF.

CHAPTER 3

3. POLYGENIC SCORES FOR KIDNEY FUNCTION IN A SOUTH AFRICAN COHORT

3.1. Abstract

Chronic kidney disease (CKD) is a public health concern with disproportionate morbidity and mortality seen in low- and-middle income countries, such as those from sub-Saharan Africa (SSA). Genome-wide association studies (GWASs) have identified more than 250 loci linked to disease risk. Polygenic scores (PGSs) have the ability to enhance the power of GWAS findings by aggregating the effect of common and rare variants that are associated with the diseases of interest. In this study we aimed to assess the predictive power of PGSs in an African setting using summary statistics from three discovery datasets. PGSs developed from different ancestries showed poor predictability in African individuals in this study, explaining less than 1% of the variability in kidney disease markers (urine albumin-to-creatinine ratio [UACR], estimated glomerular filtration rate [eGFR], albuminuria). A PGS model based on a transethnic discovery data set accounted for the highest variance in eGFR (0.8%). Our results showed that the PGS was significantly associated with eGFR in the presence of hypertension (HT), diabetes, and HIV, compared to the absence of these CKD risk factors. The assessment of PGSs for kidney function markers contributes to the understanding of CKD risk prediction in African populations, on the African continent.

3.2. Introduction

In the past decade, GWASs have identified >250 genetic loci robustly associated with CKD and measures of kidney function (UACR, eGFR, serum creatinine, and cystatin C) (Tin and Köttgen, 2020). However, accurate quantification of genetic risk for most complex disorders, including CKD, is challenging. This is because these traits are multifactorial and include genetic, environmental and behavioural risk components, and interactions between them (Liu and Kiryluk, 2018). The contribution of genetic risk for complex diseases is based on the heritability of the trait, and one measure for assessing this is the calculation of polygenic scores (PGSs) based on GWAS (Cañadas-Garre et al., 2019). GWASs have

identified many genetic variants associated with disease risk, that individually contribute to only a small fraction of overall risk, but collectively account for a higher proportion of risk, with the actual contribution varying between traits (Khera et al., 2018a). PGSs have the ability to harness the predictive power of GWAS findings by capturing the aggregate effects of common and rare variants, with a range of effect sizes on the phenotype (Chatterjee et al., 2013).

Zu et al. (2021) developed a PGS for eGFR and examined its association with outcomes defined as incident CKD, kidney failure, and acute kidney injury (AKI) in 8,866 individuals of African American and European ancestry. The base data for the PGS included a meta-analysis of summary statistics from the UK Biobank GWAS for eGFR and summary statistics from the CKDGen Consortium GWAS for eGFR (Yu et al., 2021). The PGS was significantly associated with all the aforementioned outcomes of kidney disease, however, it was not highly predictive for these traits as the hazard ratios (HRs) were small (Yu et al., 2021). The HRs (95% confidence interval [CI]) were 1.33 (1.28–1.37), 1.18 (1.08–1.29), and 1.06 (1.01–1.11) for incident CKD, kidney failure, and AKI, respectively (Yu et al., 2021). In another study on 4,873 participants of European ancestry, a PGS for lower eGFR (defined as eGFR 30–60 mL/min per 1.73 m²) was associated with an increased risk of kidney failure (HR 1.23; 95% CI: 1.12–1.35) and mortality (HR 1.12; 95% CI: 1.04–1.22) (Steinbrenner et al., 2022). Similarly, the base data for the PGS was a meta-analysis to combine the summary statistics from the UK Biobank and CKDGen Consortium GWASs for eGFR (Steinbrenner et al., 2022).

Studies have shown that PGSs derived from continental African populations or transethnic populations including African ancestry participants, perform significantly better in SSA than when using PGSs derived from only European ancestry populations (Duncan et al., 2019, Cavazos and Witte, 2021, Adam et al., 2022, Choudhury et al., 2020). This is because PGS accuracy for continental African populations is impacted by various factors including population differences in LD and allele frequencies, and inclusion of variants unique to individuals with African ancestry explained by specific genetic structure due to genetic drift, migration and selection patterns (Wang et al., 2020, Cavazos and Witte, 2021). The inclusion of African-specific variants in PGSs have the ability to enhance risk

prediction and achieve more unbiased genetic risk estimates across global populations (Cavazos and Witte, 2021). With limited transferability of PGSs derived from European ancestry populations to continental African populations (Fatumo et al., 2023), poor accuracy of PGSs in individuals with African ancestry is a key barrier in achieving precise risk stratification (Fatumo et al., 2023).

Using data from the South African site of the African Research Kidney Disease (ARK) study as the target cohort (Fabian et al., 2022), we explored the potential of using summary statistics from three discovery datasets to develop and to test the predictive power of PGSs in an African setting. One set of summary statistics was generated in another continental African study, the AWI-Gen (Africa Wits-INDEPTH partnership for Genomic Studies) study (Ramsay et al., 2016), and the other two were from transethnic GWASs for eGFR (Stanzick et al., 2021), UACR, and albuminuria (Teumer et al., 2019).

3.3. Methods

The summary workflow is shown in **Figure 3.1**.

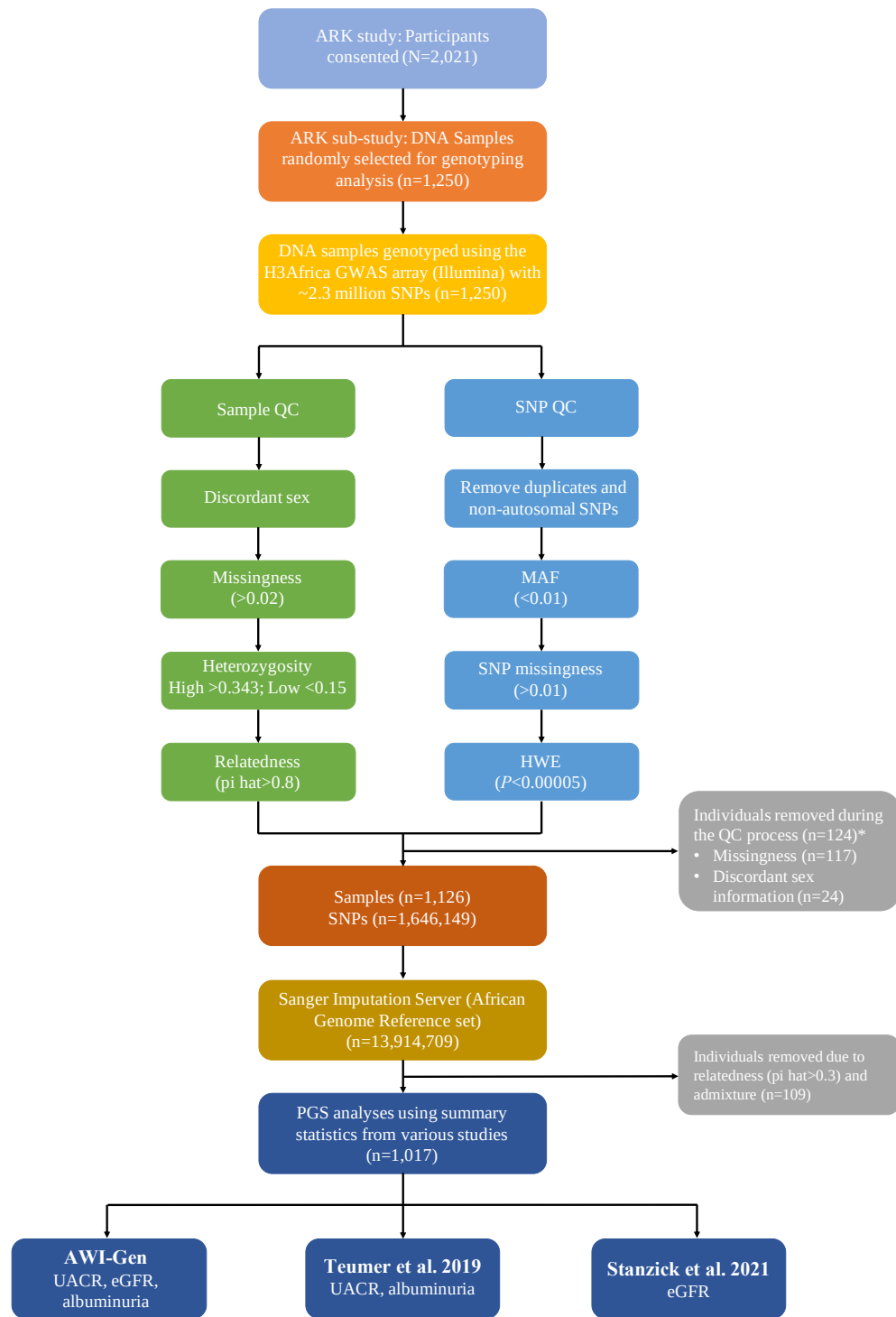


Figure 3.1: Workflow summary for datasets and analyses

*The total number may not add up due to some individuals having both discordant sex information and missingness.

ARK, African Research on Kidney Disease; AWI-Gen, Africa Wits-INDEPTH partnership for Genomic Studies; DNA, Deoxyribonucleic acid; eGFR, estimated glomerular filtration rate; GWAS, genome-wide association study; HWE, Hardy-Weinberg equilibrium; MAF, minor allele frequency; PGS, polygenic score; QC, quality control; SNP, single nucleotide polymorphism; UACR, urine albumin-creatinine ratio.

3.3.1. Study participants

This is a sub-cohort of the African Research Kidney Disease (ARK) study, comprising 2,021 Black African adults aged 20 to 80 years from Agincourt, South Africa. In the ARK study, all participants provided demographic, health, and family history information; urine and blood samples were collected. This sub-study consisted of 1,250 participants randomly selected for genotyping analysis.

3.3.2. Kidney function measurement

An isotope dilution mass spectrometry traceable Jaffe method was used to measure serum and urine creatinine (Craik et al., 2023). The 2009 CKD-EPI creatinine equation without adjustment for African American ethnicity was used to calculate GFR (ml/min/1.73 m²) (George et al., 2019). Urine albumin concentration was measured using immunoturbidimetry, and this measurement was used to calculate UACR (mg/mmol) (Craik et al., 2023).

3.3.3. Genotyping, quality control and imputation

3.3.3.1. Genotyping

DNA samples from the ARK study (n=1,250 participants) were genotyped in two batches at the Frederick National Laboratory, USA, using the H3Africa genotyping array designed as an African-common-variant-enriched GWAS array (Illumina) with ~2.3 million SNPs.

3.3.3.2. Calling

Calling of raw data was performed using IAAP Genotyping Command Line Interface (<https://emea.support.illumina.com/downloads/iaap-genotyping-cli.html>) for each batch independently. Duplicate samples were removed from each batch, and both batches were merged in PLINK. For each position, a Fisher's exact test was performed on PLINK to compare distribution of both alleles and assess homogeneity between the batches. Four positions with $P < 0.000001$ were discarded.

3.3.3.3. Quality control

Pre-imputation quality control (QC) was performed using the H3Africa Consortium Pan-African Bioinformatics Network (H3ABioNet H3AGWAS) workflow (<https://github.com/h3abionet/h3agwas>) (Brandenburg et al., 2022b, Baichoo et al., 2018).

The standard QC process involved a stepwise removal of individuals and SNP markers. Sample QC included removing individuals with discordant sex information (genotype versus documented) and SNP missingness greater than 0.02. Heterozygosity cut-offs were used to remove individuals with excessive heterozygosity (>0.343) indicative of genotyping errors, and low heterozygosity (<0.15) which implies inbreeding. Relatedness cut-off ($\pi^{\wedge} - \pi \text{ hat} > 0.8$) was applied to account for mislabelling of duplicates in the dataset.

The SNP QC process involved the removal of duplicate and non-autosomal SNPs from the dataset. SNPs with a minor allele frequency (MAF) less than 0.01 and missingness greater than 0.01 were removed. SNPs that did not conform to genotype distributions in Hardy-Weinberg equilibrium (HWE, $P < 0.0005$) were also removed as the assumption is that genetic variation within this dataset has remained constant. To increase the number of SNPs in the final analysis, imputation was performed on the QC'd dataset (with 1,646,149 SNPs and 1,126 individuals).

3.3.3.4. *Admixture and outliers*

PCASmart, a feature of the EIGENSOFT software (Price et al., 2006) was used to exclude outliers; on the first four principal components (PC), any individual outlier at more than 5 standard deviations of each PC was excluded.

The ARK dataset was merged with data from the 1000 Genomes (KGP) and African genome variation project (AGVP). African ancestry was estimated for each individual using the ADMIXTURE software (<https://dalexander.github.io/admixture/>). The ADMIXTURE software is a model-based estimation of ancestry in unrelated individuals (Alexander et al., 2009). African ancestry for each individual in this study was estimated using AGVP, African individuals from KGP, and other ancestries (South Asian, East Asian, and European American) from KGP as a reference. Individuals from the ARK dataset where African ancestry was less than 90% were removed. After the initial QC, 109 individuals were removed due to admixture and relatedness ($\pi^{\wedge} - \pi \text{ hat} > 0.3$).

3.3.3.5. *Imputation*

Imputation was performed on the 1.65 million SNPs using the African Genome Resources reference panel at the Sanger Imputation Server (<https://imputation.sanger.ac.uk/>)

(McCarthy et al., 2016). The African Genome Resource is a promising panel for sub-Saharan African datasets based on sample size, geographic origin, and ancestry representation of the reference panel and target population (Sengupta et al., 2023). EAGLE2 (Loh et al., 2016) was used for pre-phasing and the positional Burrows Wheeler transformation (PBWT) algorithm (Durbin, 2014) was used for imputation. This method uses a reference-based phasing approach to statistically estimate haplotypes from genotype data leveraging data from large reference panels (Loh et al., 2016). After imputation, poorly imputed SNPs with info scores less than 0.3 were removed. The final dataset included ~13.9 million SNPs.

3.3.4. Phenotype variables

This study included two quantitative markers and one dichotomous biomarker marker of kidney function: (1) UACR – calculated as the urine albumin-creatinine ratio, (2) eGFR – calculated using the CKD-EPI equation without the African American coefficient (Levey et al., 2009), and (3) albuminuria – defined as UACR of at least 3.0mg/mmol; control - UACR <3.0mg/mmol. Since there were only a few participants with low eGFR in this study, it was not included as a variable in the PGS analysis. In most instances, only one measurement was available for eGFR and UACR for each participant. In the absence of repeated measures chronicity could not be confirmed, therefore, we defined single screen measures as “markers of kidney function” instead of CKD.

3.3.5. Polygenic score (PGS) analyses

The PGS analyses were performed using phenotype-specific effect-weighted variants obtained from summary statistics in previous studies using PRSice-2 V2.3.5 (Euesden et al., 2015). This ARK sub-study imputed data was used for the target cohort. Summary statistics from previous GWASs for the different phenotypic outcomes were used for the discovery datasets.

PGSs are calculated as the sum of genotypes for variants genome-wide, generally weighted by effect sizes that are estimated from a GWAS study and the basic mathematic formula is:

$$PGS = \sum_{i=1}^n w_i \cdot X_i$$

where X_i is the allele count (0=non-effect allele homozygote, 0.5=heterozygote, 1=effect allele homozygote), w_i is the weight of the i^{th} SNP for a particular outcome and n denotes the number of SNPs that will be included in the PGS, depending on the disease/trait (Reisberg et al., 2017).

The PGS analyses included the following steps: (1) P value thresholding to exclude SNPs with very low significance/not significant by carrying out several PGS analyses with varying P values (P value thresholds for 0.05 to 5×10^{-8}), (2) clumping (linkage disequilibrium [LD] adjustment) to identify and choose the significant independent SNPs found in (1) in each LD region for additional analyses (Clumping distance=250kb, $R^2=0.1$); for each P value threshold, PGS results were generated using beta value, and (3) to select the PGS closest to the phenotype.

To account for overfitting data, an adjusted R^2 was used. The highest adjusted R^2 was used to estimate the best predictive PGS. In PRSice-2, an empirical P value was calculated for each PGS to determine the P value threshold (Euesden et al., 2015). The P value threshold was determined using the formula: $Empirical - P = \frac{\sum Nn=1I(P_{null} < P_o)+1}{N+1}$, where P_o indicates the optimum P value threshold; P_{null} refers to the P value of association of the optimum P value threshold under the null; and I is the indicator function. For each phenotype, different P value thresholds were identified and evaluated using the R^2 , where the optimum P value threshold was identified by the largest R^2 .

The PGS software PRSice-2 V2.3.5 (Choi and O'Reilly, 2019) was used to calculate PGSs for ARK participants using the summary statistics from previous GWAS discovery studies. The PGS prediction models were adjusted for age and sex.

3.3.6. Previous genome-wide association study cohorts

A description of the discovery GWAS cohorts used in the PGS analyses are shown in **Table 3.1**.

3.3.6.1. Africa Wits-INDEPTH partnership for Genomic Studies (AWI-Gen)

The Africa Wits-INDEPTH partnership for Genomic Studies (AWI-Gen) is a Human Heredity and Health in Africa (H3Africa) Consortium study that consists of 10,702

individuals (aged 40–60 years) from six study sites in East, West and South Africa (Ali et al., 2018, Ramsay et al., 2016). Peripheral blood and urine sample collections were described in detail previously (Ali et al., 2018, George et al., 2019). Immunoturbidimetry measurements were used to quantify urine albumin concentration, and serum and urine creatinine was quantified by the Jaffe’s kinetic technique; UACR (mg/mmol) was calculated from these measurements eGFR was calculated using the CKD-EPI equation without the African American coefficient (George et al., 2019).

DNA samples were genotyped using the H3Africa genotyping array, and similar QC and imputation was performed, as previously described (Choudhury et al., 2022, Singh et al., 2023). Analyses were performed for all AWI-Gen participants, with adjustment for age and sex (Brandenburg et al., 2024). Linear regression of variables with adjustment of covariates were performed in R: eGFR, logUACR and albuminuria, and extraction of residuals. Association analyses were performed using residuals of the variables (eGFR, UACR and albuminuria) with imputed genotype (best guessed) data using GEMMA (Version 0.98.1) and the H3ABioNet H3AGWAS pipeline (<https://github.com/h3abionet/h3agwas>) (Brandenburg et al., 2024). Population sub-structure and relatedness between individuals are accounted for using a relatedness matrix in the GEMMA software (Brandenburg et al., 2024). A sub-set of pre-imputed SNPs were used to build the relatedness matrix (Brandenburg et al., 2024). The set was pruned for LD using PLINK and an LD (r^2) threshold of 0.6, with windows of 100 kb and 20 kb for step size. Summary statistics from the AWI-Gen GWAS were obtained for eGFR (n=10,476), UACR (n=8,850), and albuminuria (n=8,850) (Brandenburg et al., 2024).

3.3.6.2. *Teumer et al. GWAS meta-analysis for UACR and albuminuria*

A GWAS meta-analysis of UACR and albuminuria was performed in 564,257 individuals of predominantly European ancestry (Teumer et al., 2019). Various methods were used to measure urinary albumin and creatinine in each study. Values for urinary albumin that were below the lower limit of the assay were set to the lower limit of detection. UACR was calculated as urinary albumin (mg/l)/urinary creatinine (mg/dl) $\times$ 100 and assessed in mg/g. Albuminuria was defined as UACR >30 mg/g and controls were defined as UACR <10 mg/g (Teumer et al., 2019).

Samples were genotyped using genome-wide arrays and study-specific QC filters were applied in each study (Teumer et al., 2019). Genome-wide data were imputed using the Haplotype Reference Consortium and 1000 Genomes Project phase 1 and phase 3 reference panels at the Sanger Imputation Server (<https://imputation.sanger.ac.uk/>) and Michigan Imputation Server (<https://imputationserver.sph.umich.edu/>). Linear regression models using inverse normal transformed age and sex residuals of logUACR as the dependent variable were performed in each GWAS analysis. A logistic regression model with adjustment for sex and age was performed for albuminuria.

The meta-analysis included 54 GWAS summary statistics files for UACR which consisted of 547,361 Europeans, 6,324 East Asians, 6,795 African Americans, 2,335 South Asians and 1,442 Hispanics. The meta-analysis for albuminuria included 38 GWAS summary files which consisted of 348,954 individuals of transethnic ancestry (European, African America, Asian, and Hispanic). Fixed-effect inverse-variance weighted meta-analyses were performed on METAL, and adapted for higher precision to obtain effects and standard errors up to seven decimal places (Willer et al., 2010). Variants with a low minor allele count (<400) were excluded resulting in 8,034,757 variants for UACR and 8,326,000 variants for albuminuria in the final results. The LD-score regression intercept was 0.95, implying that any residual inflation was probably due to polygenicity instead of confounding. Summary statistics from the GWAS meta-analysis for UACR were obtained for participants with transethnic (n=564,257), European (n=547,361), and African American (n=6,795) ancestry (Teumer et al., 2019). In addition, summary statistics from the GWAS meta-analysis for albuminuria were only obtained for participants with transethnic ancestry (n=348,954).

3.3.6.3. Stanzick et al. GWAS meta-analysis for eGFR

A GWAS meta-analysis of eGFR was performed in ~1.2 million individuals of predominantly European ancestry from the UK Biobank and CKDGen Consortium (Stanzick et al., 2021). The primary outcome of the GWAS meta-analysis was log-transformed eGFR based on creatinine (eGFR_{crea}). Creatinine concentration was measured in serum and GFR was estimated using the CKD-EPI equation (for participants >18 years) or the Schwartz formula (for individuals <18 years) (Stanzick et al., 2021).

In each CKDGen study, GWAS for eGFR_{crea} was conducted with adjustment for age, sex and other covariates that were specific to the study (Stanzick et al., 2021). The CKDGen meta-analysis included 121 GWAS summary statistics files for eGFR_{crea} which consisted of 567,460 individuals of European ancestry, 165,726 of East Asian ancestry, 13,842 of African American ancestry, 13,359 of South Asian ancestry and 4,961 of Hispanic ancestry. Data was QC'd using EasyQC (Winkler et al., 2014). In the UK Biobank study, a linear mixed model GWAS for eGFR was conducted on 436,765 participants of European using the fastGWA tool (Jiang et al., 2019). Covariates included in the association analyses were age, age², sex, age×sex, age²×sex, and 20 principal components. A fixed-effect inverse-variance weighted meta-analysis of CKDGen and UK Biobank association results was performed on METAL (Willer et al., 2010). Variants with a low minor allele count (<400) were excluded resulting in 13,633,840 variants in the final GWAS meta-analysis results for eGFR_{crea}. The meta-analysis results for the GC lambda inflation factor (λ) was $\lambda=1.28$, and LD-score regression intercept was 0.90, indicating conservative GC correction and absence of confounding by population sub-structure, respectively (Stanzick et al., 2021).

Table 3.1: Description of discovery GWAS cohorts

Study	Ethnicity	Sample size	Phenotype
AWI-Gen (Brandenburg et al., 2024)	Continental African	10,476	eGFR
AWI-Gen (Brandenburg et al., 2024)	Continental African	8,850	UACR
AWI-Gen (Brandenburg et al., 2024)	Continental African	8,850	Albuminuria
(Teumer et al., 2019)	Transethnic (European, African America, Asian, Hispanic)	564,257	UACR
(Teumer et al., 2019)	European	547,361	UACR
(Teumer et al., 2019)	African American	6,795	UACR
(Teumer et al., 2019)	Transethnic (European, African America, Asian, Hispanic)	348,954	Albuminuria
(Stanzick et al., 2021)	Transethnic (European, African American, Asian)	1,202,113	eGFR

AWI-Gen, Africa Wits-INDEPTH partnership for Genomic Studies; eGFR, estimated glomerular filtration rate; GWAS, genome-wide association study; UACR, urine albumin-to-creatinine ratio.

3.3.7. Stratification by CKD risk factors

Using the transethnic (Stanzick) eGFR PGS computed in PRSice-2 , we stratified the analysis by CKD risk factors (HT, diabetes, and HIV) in R software (<https://github.com/jeantristanb/pgs/tree/main/PRSice-2/plotresult>). GLM models computed in R were used to assess the association between the PGS and eGFR when stratified by risk factors (HT, diabetes, and HIV). The GLM models were adjusted for age and sex.

3.4. Results

3.4.1. Baseline demographics and disease characteristics

Between 2nd November 2017 and 5th September 2018, 2021 adults of self-identified black ethnicity were recruited from Agincourt in the rural Mpumalanga Province, South Africa.

Baseline characteristics of the overall ARK study and the 1,017 participants whose genetic data passed quality control are displayed in **Table 3.2**. Baseline characteristics were similar in the overall ARK study and this genetic sub-study. There were no significant difference between the full ARK study cohort and this genetic sub-cohort. Overall, there were many risk factors for CKD present in the full study cohort and genetic sub-cohort, including HT, HIV, and diabetes.

Table 3.2: Baseline demographics and disease characteristics of ARK participants comparing the full study cohort to the sub-cohort used for the PGS analyses

Variable	Overall ARK study (N=2,021)	Genetic sub-study (n=1,017)	P value
Mean age ($\pm$ SD) years	37.9 ($\pm$ 14.1)	38.8 ($\pm$ 14.7)	0.1138
Sex			
Male	851 (42.1)	423 (41.6)	0.7861
Female	1,170 (57.9)	594 (58.4)	
BMI ($\pm$ SD) kg/m ²	26.6 (6.3)	26.7 (6.4)	0.5618
Hypertension^a			
No	1,424 (70.5)	693 (68.1)	0.1896
Yes	597 (29.5)	324 (31.9)	
Diabetes^b			
No	1,933 (95.6)	972 (95.6)	0.9286
Yes	88 (4.4)	45 (4.4)	
HIV infection			
No	1,641(81.2)	764 (80.0)	0.4389
Yes	380 (18.8)	191 (20.0)	
UACR (mg/mmol), median (IQR)	1.0 (0.2–1.3)	0.4 (0.2–1.2)	0.6154
eGFR (ml/min/1.73m²),^c mean (SD)	111.6 (18.3)	110.9 (19.3)	0.7564
Markers of kidney function^d			
Albuminuria (UACR >3 mg/mmol)	241 (11.9)	136 (13.4)	0.2618
Low eGFR (eGFR <60 ml/min per 1.73 m²)	32 (1.6)	21 (2.1)	0.3398

T-tests and a Chi-squared analysis using STATA were used to compute associations between the overall ARK study and genetic sub-study, as appropriate.

^aHypertension (SBP $\geq$ 140mmHg or DBP $\geq$ 90mmHg).

^bDiabetes: non-fasting glucose $\geq$ 11.1mmol/L.

^cMeasured using the Chronic Kidney Disease Epidemiology (CKD-EPI) equation without adjustment for race (Levey et al., 2011).

^dOverall ARK study: 13 participants had both UACR >3 mg/mmol and eGFR <60 ml/min per 1.73 m²; Sub-study: 7 participants with both UACR >3 mg/mmol and eGFR <60 ml/min per 1.73 m². eGFR, estimated glomerular filtration rate (CKD-EPI equation); SD, standard deviation; UACR, urine albumin-creatinine ratio.

3.4.2. Distribution of phenotypes

The distribution of eGFR is shown in **Figure 3.2A**, and was not transformed. UACR was log₁₀ transformed for the PGS analysis due to a skewed distribution (**Figure 3.2B & C**). There were no significant differences in the proportion of participants with low eGFR and albuminuria between the two groups (**Table 3.2; Figure 3.3**).

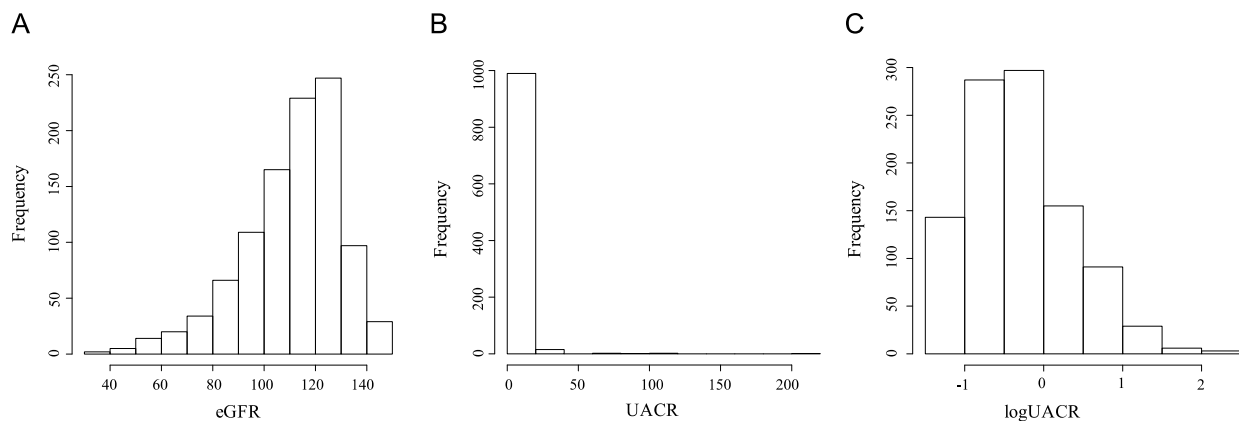


Figure 3.2: Distribution of kidney function markers in the ARK genetic sub-study
eGFR, estimated glomerular filtration rate; UACR, urine albumin-to-creatinine ratio.

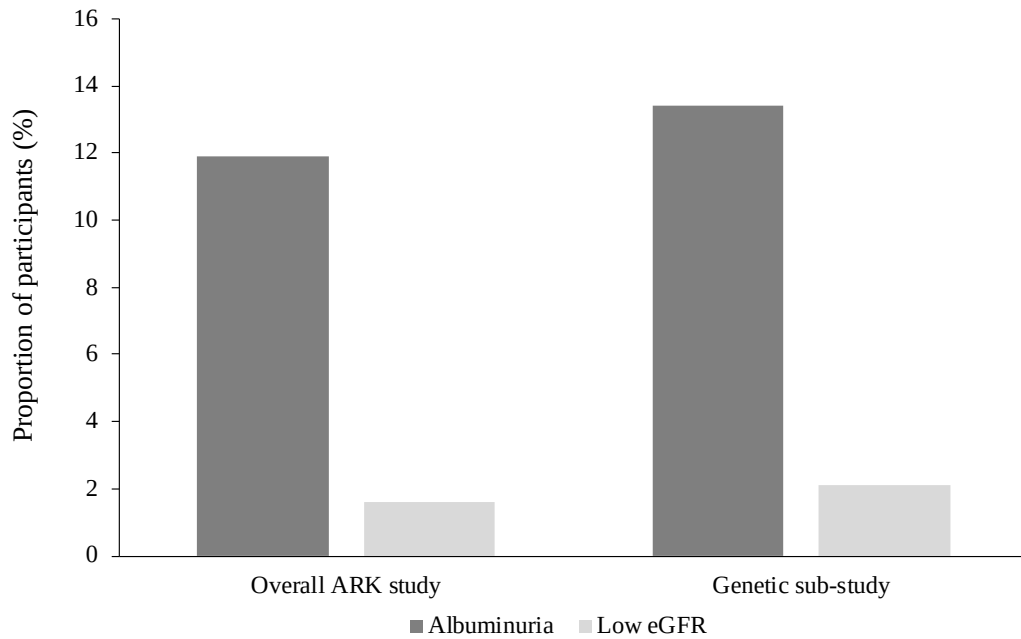


Figure 3.3: Proportion of participants with markers of kidney function

Albuminuria was defined as UACR >3 mg/mmol.

Low eGFR was defined as GFR <60 ml/min per 1.73 m².

eGFR, estimated glomerular filtration rate; UACR, urine albumin-to-creatinine ratio.

3.4.3. Polygenic score and association analyses

Eight PGS were developed for three markers of kidney function (eGFR, UACR and albuminuria) using summary statistics from previous GWAS in participants of diverse ancestries and applied to the individuals in this ARK genetic sub-study: (1) AWI-Gen African ancestry cohort for eGFR; (2) AWI-Gen African ancestry cohort for UACR; (3) AWI-Gen African ancestry cohort for albuminuria; (4) Stanzick transethnic cohort for eGFR; (5) Teumer transethnic cohort for UACR; (6) Teumer European cohort for UACR; (7) Teumer African American cohort for UACR; and (8) Teumer transethnic cohort for albuminuria.

The variance in kidney function markers explained by the eight PGSs was low; below 1% for all phenotypes (range 0.1–0.8%) (**Figure 3.4**). The highest R^2 was for the Stanzick transethnic PGS for eGFR (0.8%). The PGS for eGFR generated using summary statistics from Stanzick transethnic cohort was significant when applied to individuals in this genetic sub-study (Empirical P value=0.003); likely due to the high number of SNPs ($n=199,836$) used to generate the PGS and large sample size ($n=765,348$) consisting of 13,842 African

Americans. In comparison, the PGS for UACR generated using summary statistics from the Teumer transethnic cohort (n=564,257, consisting of 6,795 African Americans) and applied to individuals in this study had a R^2 of 0.1%, was not significant (Empirical P value=0.99), and used 9,332 SNPs to generate the PGS. The PGS for eGFR, UACR and albuminuria generated using summary statistics from AWI-Gen of individuals with African ancestry was not significant when applied to individuals in this genetic sub-study (Empirical P values=0.11, 0.27, and 0.71, respectively); likely due to the lower number of SNPs (n=137,136, 11,504, and 8,960, respectively) used to generate the PGS and smaller sample sizes (n=10,476, 8,850, and 8,850, respectively) of the discovery cohorts.

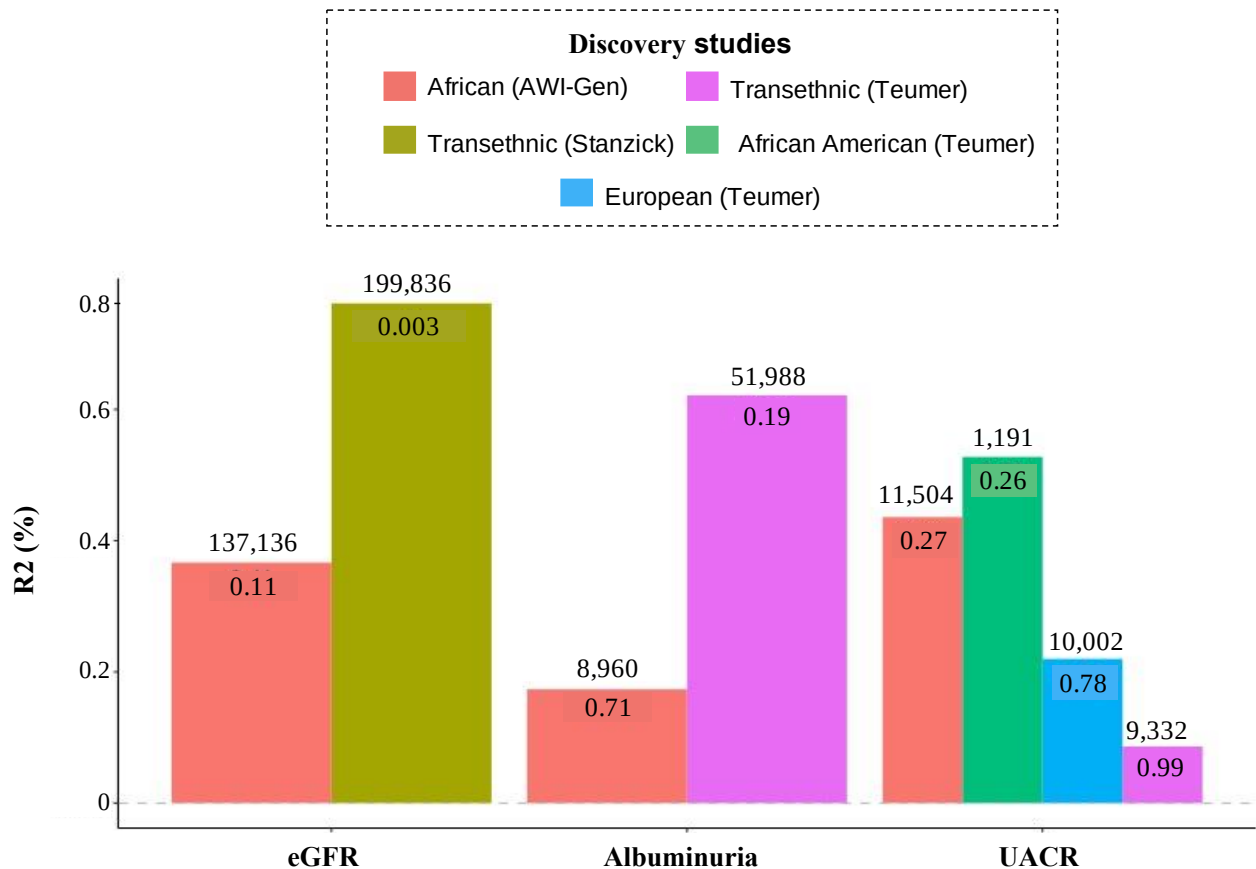


Figure 3.4: Transferability of polygenic scores generated from summary statistics of external cohort studies

Bar chart showing adjusted variance explained (R^2 , %) for each PGS derived from discovery GWASs for eGFR, albuminuria, and UACR: P value (within the bar) and number of SNPs (above the bar) are stated for each kidney function marker. PGS derived from discovery GWASs and applied to the ARK sub-study (target): (1) AWI-Gen African ancestry cohort for eGFR (N=10,476); (2) AWI-Gen African ancestry cohort for UACR (N=8,850); (3) AWI-Gen African ancestry cohort for albuminuria (N=8,850); (4) Stanzick transeethnic cohort for eGFR (N=765,348); (5) Teumer transeethnic cohort for UACR (N=564,257); (6) Teumer European cohort for UACR (N=547,361); (7) Teumer African American cohort for UACR (N=6,795); (8) Teumer transeethnic cohort for albuminuria (N=348,954).

ARK, African Research Kidney Disease; eGFR, estimated glomerular filtration rate; GWAS, genome-wide association study; PGS, polygenic score; SNP, single nucleotide polymorphism; UACR, urine albumin-to-creatinine ratio.

As only the PGS developed from the Stanzick transeethnic cohort for eGFR was significant, the odds ratio for eGFR levels by PGS deciles were plotted. Across PGS deciles of the eGFR, participants in the highest decile (10), corresponded to >2-fold increased genetic risk for lower eGFR levels. (**Figure 3.5**). In addition, the association between PGS score

and eGFR when stratified by risk factors (HT, diabetes, and HIV) were assessed. The PGS was significantly associated with eGFR in individuals with HT, diabetes, and HIV, compared to those without (**Table 3.3**). Most of the variance in the PGS for eGFR can be explained by the presence of HT ($R^2=0.12\%$) and diabetes ($R^2=0.20\%$), in comparison to the absence of the HT ($R^2=0.07\%$) and diabetes ($R^2=0.04\%$) (**Figure 3.6**). The variance in the PGS for eGFR was similar in individuals with and without HIV ($R^2=0.05\%$ and 0.04% , respectively).

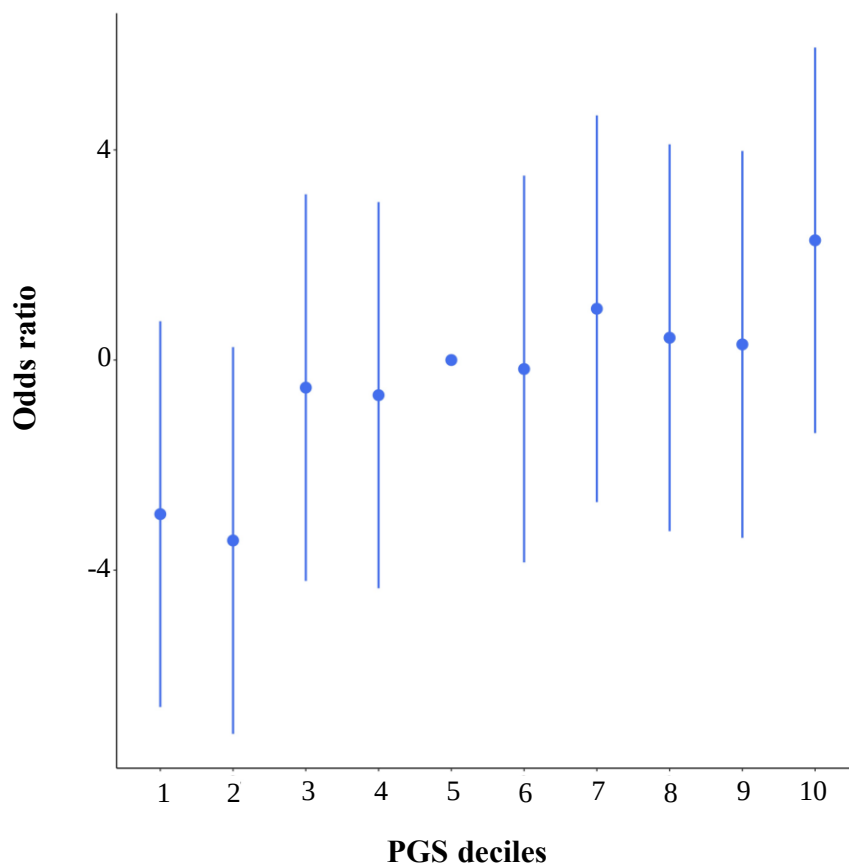


Figure 3.5: Odds ratios for eGFR levels across PGS deciles

PGS derived from Stanzick transethnic cohort for eGFR (N=765,348) GWAS and applied to the ARK sub-study (target). Point range-plots comparing the difference in eGFR levels of the upper/lower PGS decile from the middle PGS decile.

ARK, African Research Kidney Disease; eGFR, estimated glomerular filtration rate; GWAS, genome-wide association study; PGS, polygenic score.

Table 3.3: GLM models for Stanzick eGFR PGS stratified by CKD risk factors in the target ARK cohort

	Estimate	Std error	P value
All individuals	2.2×10^4	7.8×10^3	0.006
Hypertension			
Yes (n=324)	3.2×10^4	8.0×10^3	<0.001
No (n=693)	2.4×10^3	1.7×10^4	0.88
Diabetes			
Yes (n=45)	2.1×10^4	7.7×10^3	0.007
No (n=972)	2.7×10^4	7.2×10^4	0.715
HIV infection			
Yes (n=191)	2.4×10^4	9.0×10^3	0.001
No (n=764)	2.5×10^4	2.0×10^4	0.209

GLM models were adjusted for age and sex.

eGFR, estimated glomerular filtration rate; GLM, general linear model; PGS, polygenic score.

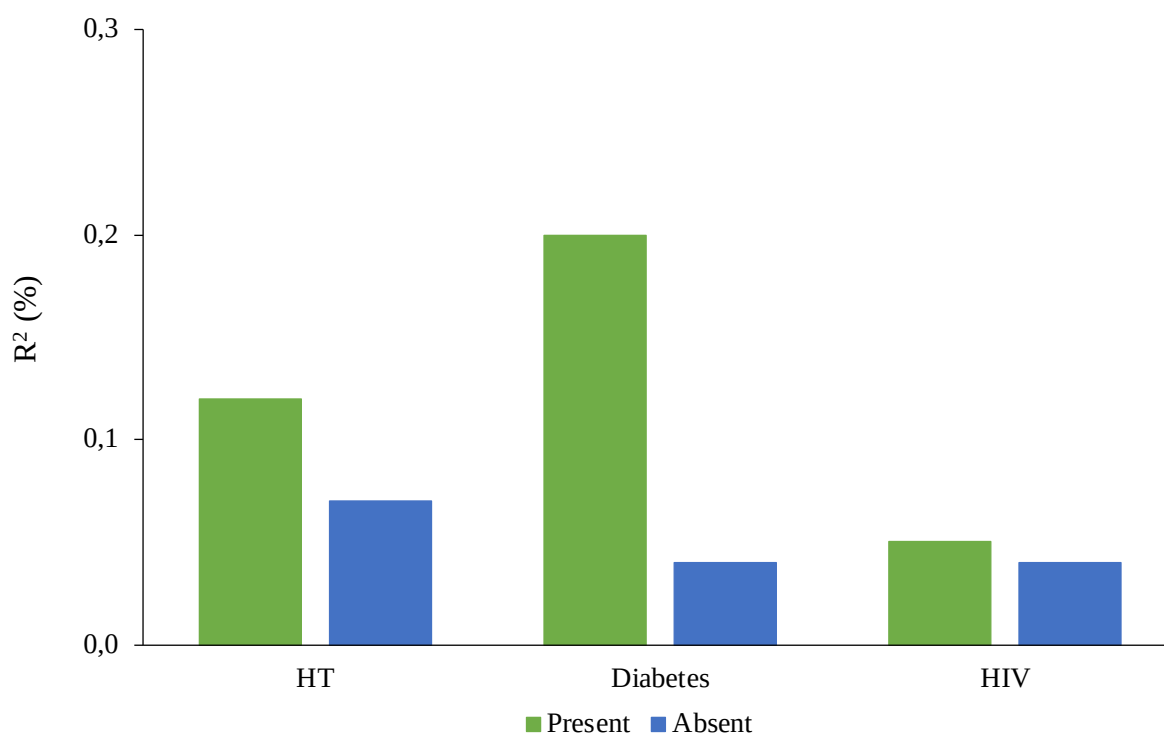


Figure 3.6: Variance in eGFR explained by PGS in the presence or absence of CKD risk factors (HT, diabetes, and HIV)

eGFR, estimated glomerular filtration rate; HIV, human immunodeficiency virus; HT, hypertension; PGS, polygenic score.

3.5. Discussion

CKD is a complex multifactorial disease that involves genetic and environmental risk factors, making accurate identification of genetic risk challenging (Friedman, 2019). This study aimed to develop and assess the predictive power of PGSs derived from discovery GWASs from transethnic, European, continental African and African American studies for kidney function markers in a black South African population, represented in this ARK sub-study. PGSs derived from continental African, African American, European and transethnic populations GWAS summary statistics explained less than 1% of the variability in eGFR, UACR, and albuminuria in this cohort.

In this study, a PGS for eGFR derived from transethnic GWAS summary statistics explained the highest variability in eGFR (0.8%) in comparison to a PGS developed from a continental African population (0.4%). In a PGS for eGFR (based on creatinine) on 2,900 European individuals, the PGS explained 9.6% of the variance in eGFR; much higher than this study (Herold et al., 2023). In another study on 8,886 participants of European and African ancestry (82% Europeans), an eGFR PGS explained 5.7% of the variance in the total population, and only 1.8% in the African ancestry population (Yu et al., 2021). Previous studies have shown that a PGS derived from a continental African population or transethnic cohort performs significantly better in SSA than a PGS derived from only European ancestry populations (Duncan et al., 2019, Cavazos and Witte, 2021, Adam et al., 2022, Choudhury et al., 2020). A study by Hughes et al. (2023) supports the findings in this study, by demonstrating increased predictive power with a transethnic PGS for eGFR (Hughes et al., 2023). The ability of PGSs is limited by genetic contribution to the trait, sample size, and genotype and phenotype diversity (Wang et al., 2023). Multiple methods have been shown to improve PGS accuracy. These include the PolyPred method (Weissbrod et al., 2022, Fatumo et al., 2023), and Bayesian and “P+T” approaches together with prior knowledge of trait-specific genetic architecture estimates (Wang et al., 2023). Previous studies have shown that the genetic variance for eGFR ranges from 2.0–9.8% in studies with large sample sizes ($n=47,219$ – $436,581$) (Stanzick et al., 2021). In this doctoral thesis, the low eGFR variance explained by the PGS may have limited by the small sample size ($n=1,017$).”

The low phenotypic variance explained by the continental African derived PGS could be the result of regional variability across Africa. For example, PGS variance for low-density lipoprotein cholesterol greatly varied with SSA between a Zulu-speaking South African population (8.1%) and a Ugandan population (0.03%) (Kamiza et al., 2022). The poor transferability within SSA may be because of genetic and environmental differences between populations (Kamiza et al., 2022).

The low variance for UACR (0.1–0.6%) in this study could be due to the low genetic heritability of UACR (4.3%) (Teumer et al., 2019). Similarly, PGSs for UACR using CKDGen summary statistics explained 0–0.8% variance of UACR in another continental African cohort (Brandenburg et al., 2024). In addition, a PGS for UACR demonstrated a variance of 0.01% in the total population (consisting of 8,886 European and African ancestry participants), and 0.1% in the African ancestry (n=2,871) population (Yu et al., 2021). There was also low transferability between the PGS derived from the African American discovery GWAS for UACR and this study population. The low phenotypic variance explained by the African American derived PGS could be because the African American population was from an admixed group that did not reflect the genetic diversity and LD patterns of continental Africans (Choudhury et al., 2020).

PGS transferability relies on similarities between the discovery and target datasets. However, the low transferability of the continental African ancestry derived PGS to black African individuals in this study could be attributed the small sample size, participants from a different geographic region in Africa, differences in LD, allele frequency, and gene-environmental factors. Furthermore, African kidney disease cohorts are substantially underpowered (<10,000), in comparison to European cohorts with over a million individuals (Stanzick et al., 2021). The genetic architecture of kidney disease in Africans could also be influenced by environmental factors and variability in the prevalence of risk factors such as hypertension, diabetes, and HIV (George et al., 2019, Brandenburg et al., 2022a, Fatumo et al., 2021, Limou et al., 2014).

It has been previously established that the relative risk for CKD is strongly associated with diabetes, HT, and HIV (George et al., 2019, Xie et al., 2023). As such, we assessed the association between the PGS derived from a transethnic discovery GWAS for eGFR when

stratified by risk factors including HT, diabetes, and HIV. The results indicated that the PGS was significantly associated with eGFR in participants with HT, diabetes, and HIV, compared to those without. This is consistent with studies from East, West and South Africa. In the overall ARK cohort of 2,021 individuals from Agincourt, Mpumalanga, South Africa, HIV infection (odds ratio [OR] 1.8; 95% CI: 1.1–2.8); HT (OR 2.8; 95% CI: 1.8–4.3), and diabetes (OR 4.1; 95% CI: 2.0–8.4) were risk factors for CKD (Fabian et al., 2022). In the AWI-Gen study on 8,110 participants from West, East and Southern Africa, the relative risk for CKD was also strongly associated with HT (OR 1.97; 95% CI: 1.68–2.30; $P<0.0001$), diabetes (OR 2.22; 95% CI: 1.76–2.78; $P<0.0001$), and HIV (OR 1.65; 95% CI: 1.36–1.99; $P<0.0001$) (George et al., 2019).

A limitation of the study included the small sample sizes for GWASs for kidney disease in continental Africans ($n=10,476$) and African Americans ($n=6,795$) that were used to develop the PGSs. In addition, it is important to note that the number of individuals with diabetes in the stratified analysis may be too low ($n=45$) to make any meaningful conclusion. It is important for future research to focus on continental African populations from different geographic regions with increased sample sizes to improve PGS prediction. Another limitation of the study was bias due to differing genotyping and GWAS study designs used by the various studies. No additional tools, except for PRScise-2, were used to compute and compare PGS estimates. The arrays used for the other studies did not include many of the common African variants that were present in the H3Africa array, and therefore were not included in this study. This could be addressed by using discovery GWASs that follow the same protocol and study design as the target population.

3.6. Conclusions

Polygenic scores derived from studies in different ancestries showed limited predictability, explaining less than 1% of the variability in kidney function markers in African individuals in this study. A PGS derived from a transethnic cohort for eGFR explained the highest variability in this cohort. The eGFR PGS was significantly associated with the presence of CKD risk factors, including HT, HIV, and diabetes. The poor transferability for kidney function markers highlights the need for more research that focuses on continental African populations with large sample sizes to enhance CKD risk prediction.

CHAPTER 4

4. PROTEOMIC INSIGHTS INTO THE PATHOPHYSIOLOGY OF HYPERTENSION-ASSOCIATED ALBUMINURIA: PILOT STUDY IN A SOUTH AFRICAN COHORT

This work was raised as a paper publication (Govender et al., 2024). This review was published in the *Clinical Proteomics* journal in February 2024.

4.1. Abstract

Background: Hypertension is an important public health priority with a high prevalence in Africa. It is also an independent risk factor for kidney outcomes. We aimed to identify potential proteins and pathways involved in hypertension-associated albuminuria by assessing urinary proteomic profiles in black South African participants with combined hypertension and albuminuria compared to those who have neither condition.

Methods: The study included 24 South African cases with both hypertension and albuminuria and 49 control participants who had neither condition. Protein was extracted from urine samples and analysed using ultra-high-performance liquid chromatography coupled with mass spectrometry. Data was generated using data-independent acquisition (DIA) and processed using Spectronaut™ 15. Statistical and functional data annotation were performed on Perseus and Cytoscape to identify and annotate differentially abundant proteins. Machine learning was applied to the dataset using the OmicLearn platform.

Results: Overall, a mean of 1,225 and 915 proteins were quantified in the control and case groups, respectively. Three hundred and thirty-two differentially abundant proteins were constructed into a network. Pathways associated with these differentially abundant proteins included the immune system (q -value [false discovery rate]= 1.4×10^{-45}), innate immune system ($q=1.1 \times 10^{-32}$), extracellular matrix (ECM) organisation ($q=0.03$) and activation of matrix metalloproteinases ($q=0.04$). Proteins with high disease scores (76–100% confidence) for both hypertension and CKD included angiotensinogen (AGT), albumin (ALB), apolipoprotein L1 (*APOL1*), and uromodulin (UMOD). A machine learning

approach was able to identify a set of 20 proteins, differentiating between cases and controls.

Conclusions: The urinary proteomic data combined with the machine learning approach was able to classify disease status and identify proteins and pathways associated with hypertension and albuminuria.

Keywords: Proteomics, albuminuria, hypertension, ultra-high-performance liquid chromatography, machine learning.

4.2. Introduction

Globally, the World Health Organization estimates that more than a billion people have hypertension, a risk factor for adverse cardiovascular and cerebrovascular events and kidney disease. Hypertension is a major cause of premature death worldwide, with a disproportionate burden in low- and middle-income countries, largely due to a rising prevalence of risk factors in recent decades (Stanzick et al., 2021).

As an upper middle-income country, South Africa has a high prevalence of hypertension, and hypertensive kidney disease is the most common cause of kidney failure documented by the South African Renal Registry (Hariparshad et al., 2023). Hypertension and kidney disease are inextricably linked, with common pathophysiological denominators that begin with foetal programming in-utero (De Bhailis and Kalra, 2022, Zandi-Nejad et al., 2006). There is strong epidemiological evidence linking low birth weight, as a marker of adverse intrauterine circumstances, to adult hypertension and kidney disease (Zandi-Nejad et al., 2006, Gjerde, 2022). Potential mechanisms include a congenital deficit in nephron number which may arise from changes in DNA methylation, accelerated apoptosis in the developing kidney, changes in renal renin–angiotensin system activity, and an increase in foetal glucocorticoid exposure (Zandi-Nejad et al., 2006). A decrease in nephron number is associated with compensatory glomerular hypertrophy and an increased susceptibility to the progression of kidney disease (Zandi-Nejad et al., 2006).

Angiotensin, an endocrine hormone peptide, is a vital part of the renin-angiotensin-aldosterone system, an inter-related endocrine system that plays a significant role in volume

and blood pressure control (Xie et al., 2023). In response to a drop in blood pressure, or sympathetic nerve activity in the kidney, renin is released and cleaves off two amino acids enzymatically to form angiotensin I (ATI), which is cleaved by the angiotensin converting enzyme (ACE) to form angiotensin II (ATII) (Xie et al., 2023). ATII is the main effector molecule of this system, increasing blood pressure, enhancing renal tubular reabsorption of sodium and water, and stimulating aldosterone release from the adrenal gland (Benigni et al., 2010). In addition to being an effective vasoconstrictor, ATII has also been shown to activate proliferative, pro-inflammatory and pro-fibrotic pathways and stimulates the production of ET-1 resulting in an increase in oxidative stress (Benigni et al., 2010, Pollock, 2000, Sasser et al., 2002). These combined effects of ATII contribute to the development of kidney disease (Navar et al., 1996, Rüster and Wolf, 2006, Siragy and Carey, 2010).

With sustained elevations of blood pressure, the afferent arterioles in the kidney undergo structural changes and hypertrophy, and intraglomerular pressure results in glomerular hypertrophy, hyperfiltration, and albuminuria (Bidani and Griffin, 2004, Shankland, 2006, Folkow et al., 1977). Sustained glomerular hyperfiltration can result in glomerular scarring and irreversible kidney injury (Bidani and Griffin, 2004, Shankland, 2006, Folkow et al., 1977, Buffet and Ricchetti, 2012, Heerspink and Gansevoort, 2015).

Kidney disease is diagnosed using estimated glomerular filtration rate (eGFR) and albuminuria, as each independently confer an increased risk of cardiovascular and all-cause mortality (Levin et al., 2013). Chronic kidney disease is defined as eGFR <60 mL/min/1.73 m²; and/ or albuminuria, defined as an albumin-to-creatinine ratio ≥ 3.0 mg/mmol (Levin et al., 2013). Albuminuria may precede the onset of reduced eGFR, which makes it appealing as a potential for early detection and treatment strategies for kidney disease, but this is not a consistent, replicable finding (Lopez-Giacoman and Madero, 2015, Heerspink and Gansevoort, 2015). The search for more sensitive markers of early kidney disease has extended to exploration of urinary proteomics with significant progress in the proteomic analysis of biological fluids, in this case, urine.

The past decade has seen omics studies contributing to the diagnosis, therapeutic intervention and prognosis of kidney disease (Wu et al., 2022, Cisek et al., 2016, Good et al., 2010). Samples derived from urine are non-invasive and easy to collect compared to other body fluids, and can be used to identify genomic, metabolomic, transcriptomic, and proteomic biomarkers that are strongly associated with pathophysiologic mechanisms of disease (Provenzano et al., 2021). Proteomic strategies such as mass spectrometry combined with liquid chromatography and capillary electrophoresis are used to identify urine biomarkers involved in early detection of kidney disease and potential disease pathways that provide insight into the pathogenic mechanisms of disease (Mischak et al., 2010, Fan et al., 2021, Pontillo and Mischak, 2017a). One such example is a classifier based on 273 urinary peptides (CKD273) which reliably allows for early detection of kidney disease and is more sensitive than albuminuria in predicting a decline in eGFR (Pontillo and Mischak, 2017a).

To date, there is one study from South Africa that has investigated the urinary proteome in young adults with hypertension (De Beer et al., 2023). This study found that combining 20 peptides into a single classifier resulted in the separation of normotensive and hypertensive groups with an area under the curve of 0.85 ($P<0.001$) (De Beer et al., 2023). There are no published studies from South Africa investigating the urinary proteome in the setting of hypertension and albuminuria – as an early indicator of kidney injury. In this exploratory study, we hypothesise that urinary proteomic analysis will identify urinary markers of kidney disease in hypertensive individuals with albuminuria and well preserved eGFR.

4.3. Methods

4.3.1. Study participants

This case-control pilot study is a sub-study of the African Research Kidney Disease (ARK) study, a well characterised population-based cohort study of 2021 adults (20–80 years) of self-identified black ethnicity from Agincourt in the rural Mpumalanga Province, South Africa (Kalyesubula et al., 2020). Demographic, health, and family history information were obtained for all participants, with collection of stored urine and blood samples. Participants did not disclose food or beverage consumption prior to urine collection. From

the cohort, ninety participants with stored urine samples were selected for proteomic analysis (**Figure 4.1**).

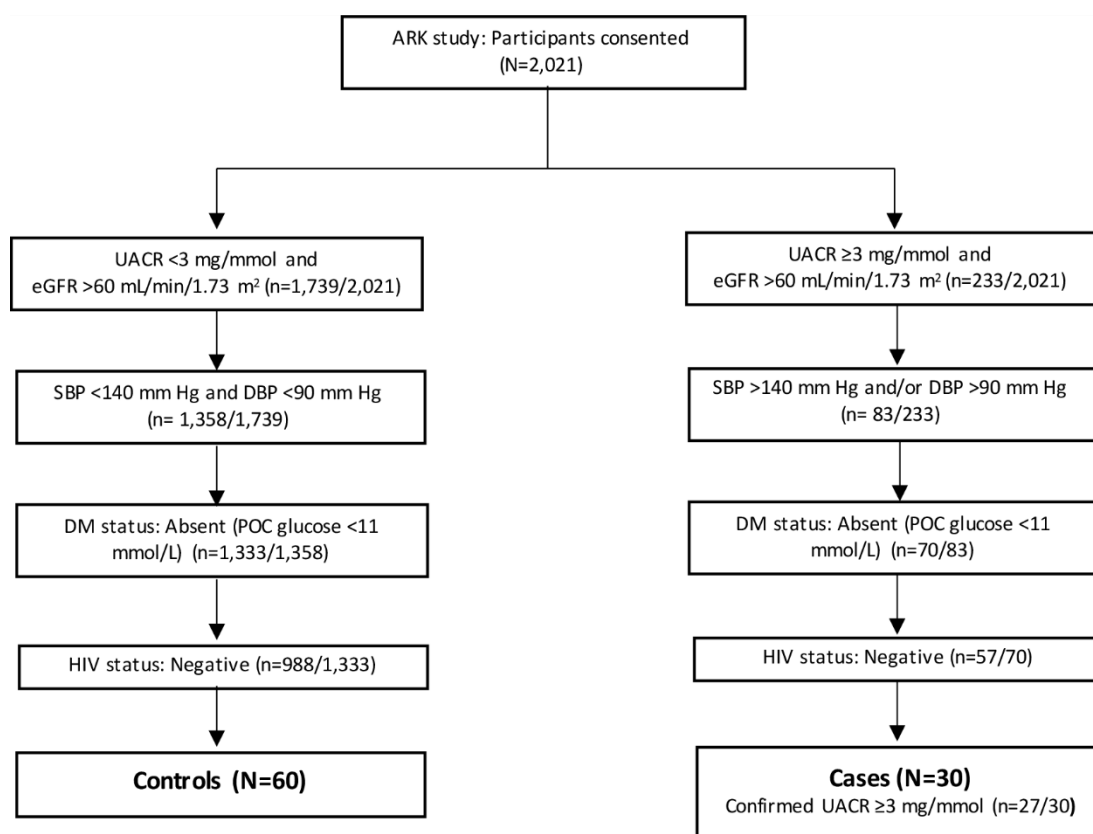


Figure 4.1: Flow diagram for this case-control study

ARK, African Research Kidney Disease; DBP, diastolic blood pressure; DM, diabetes mellitus; HIV; human immunodeficiency virus; POC, point of care; SBP, systolic blood pressure; UACR, urine albumin-to-creatinine ratio.

The inclusion criteria of this sub-study were participants (cases) with albuminuria (urine albumin-to-creatinine ratio [UACR] ≥ 3 mg/mmol) and hypertension (SBP ≥ 140 mm Hg and/or diastolic blood pressure [DBP] ≥ 90 mm Hg, JNC7 criteria). The control group included participants with UACR < 3 mg/mmol, SBP < 140 mm Hg, and DBP < 90 mm Hg. Individuals with other potential causes or consequences of kidney disease, including diabetes mellitus and HIV infection were excluded as cases and controls. Diet was not considered in this study. The control group was matched by age and sex.

4.3.2. Urine collection and Measurement of eGFR and UACR

Approximately 20–30 ml urine was collected from ARK study participants. Urine samples were processed and stored at -80°C prior to shipping to the Council for Scientific and Industrial Research laboratory for proteomic testing (Craik et al., 2023). Serum and urine creatinine was measured using an isotope dilution mass spectrometry traceable Jaffe method (Craik et al., 2023). The 2009 Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine equation without adjustment for African American ethnicity was used to calculate eGFR ($\text{ml}/\text{min}/1.73 \text{ m}^2$) (George et al., 2019). Immunoturbidimetry was used to measure urine albumin concentration (Craik et al., 2023). The UACR was calculated (mg/mmol) based on this measurement (Craik et al., 2023).

4.3.3. Urinary protein extraction by precipitation

An in-house urinary proteome sample preparation method was used. For each urine sample, protein precipitation was achieved by adding 1,600 μL ice-cold acetone to 400 μL crude urine, stored at -32°C for 1 h. The samples were centrifuged at 12,000 g at 4°C for 1 h. Thereafter, the supernatant was discarded and the pellet was allowed to air dry. The precipitated protein pellet was resuspended in 100 μL 2% sodium dodecyl sulfate solution and sonicated for 5 min (Elmi, Riga, Latvia). Dithiothreitol was added to the solution (final 10 mM), and placed on a 70°C heating block for 15 min, and then on a 40°C heating block for 15 min. The tubes were cooled to room temperature (RT) and iodoacetamide was added for alkylation (final 30 mM) and kept in the dark at RT for 30 min. The protein sample was mixed with an equal volume HILIC binding buffer (30% acetonitrile [MeCN]/200mM Ammonium acetate [NH_4Ac]) and kept at RT before it was added to a KingFisher™ deep well plate.

4.3.4. Automated on-bead digestion using MagReSyn® HILIC

Protein samples were digested on-bead using multimode magnetic microparticles (MagReSyn® HILIC, ReSyn Biosciences) in a KingFisher Duo™ system (Thermo Fisher Scientific), as previously described (Nweke et al., 2020, Govender et al., 2023), with minor modifications. Briefly, magnetic hydrophilic affinity microparticles (20 μL , 200 μg) were equilibrated in 200 μL 100 mM NH_4Ac pH 4.5, 15% MeCN. The microparticles were transferred to a well containing the protein-binding buffer solution and mixed for 30 min at

RT. The captured proteins were washed twice in 200µl 95% MeCN and transferred to 200µl 25mM ammonium bicarbonate containing 1 µg sequencing grade modified trypsin (Promega, Madison, USA) and mixed for 4 h at 37°C. Finally, beads were washed in 1% trifluoroacetic acid to elute any remaining bound peptides. The resulting peptides (pool of digest and eluate) were vacuum dried, resuspended in 2% MeCN, 0.2% FA and quantified using the Pierce™ Peptide Quantification (Thermo Fisher Scientific) assay as per the manufacturer's instructions.

A project specific system suitability-quality control (PQC) sample was prepared by pooling an equal volume of 16 urinary peptide samples. Additionally, a complex proteome digest was used as a general system suitability assessment. These PQC samples were injected at least once with each batch and analysed and processed together with the study samples.

4.3.5. Low pH reverse phase liquid chromatography with mass spectrometry (LCMS/MS) data acquisition

Individual participant peptide samples (500 ng, single shot) were analysed using a Dionex UltiMate™ 3000 UHPLC in nanoflow configuration. Samples were inline desalted on an Acclaim PepMap C18 trap column (75 µm × 2 cm; 2 min at 5 µl/min using 2% MeCN/0.2% FA). Trapped peptides were gradient eluted and separated on a nanoEase M/Z Peptide CSH C18 Column (130 Å, 1.7 µm, 75 µm X 250 mm) (Waters) at a flow-rate of 300 nl/min with a gradient of 6–35% over 30 min (A: 0.1% FA; B: 80% MeCN/0.1% FA).

Data was acquired using DIA - or Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH) using a TripleTOF® 6600 mass spectrometer (SCIEX, Massachusetts, USA) (Röst et al., 2014). Eluted peptides were delivered into the mass spectrometer via a Nanospray® III ion source equipped with a 20 µm Sharp Singularity emitter (Fossil Ion Technology, Madrid, Spain). Source settings were: Curtain gas - 20, Gas 1 - 16, Gas 2 - 0, temperature – 0 (off) and ion spray voltage – 2 600 V.

Data was acquired using 64 MS/MS scans of overlapping sequential precursor isolation windows (variable m/z isolation width, 1 m/z overlap, high sensitivity mode), with a precursor MS scan for each cycle. The accumulation time was 100 ms for the MS1 scan

(from 400 to 900 m/z) and 15 ms for each product ion scan (100 to 1 800 m/z) for a 1.06 sec cycle time.

4.3.6. DIA (SWATH) library generation and data extraction

A spectral library was built (from all patient DIA files) in Spectronaut™ 15 software using default settings with minor adjustments as follows: segmented regression was used to determine the normalised retention time (iRT) in each run; iRTs were calculated as median for all runs; the digestion rule was set as “Trypsin” and modified peptides were allowed; fragment ions between 300 and 1 800 m/z and ions with greater than 3 amino acids were considered; peptides with a minimum 3 and maximum 6 (most intense) fragment ions were accepted. This study specific spectral library was concatenated with an in-house generated urinary proteome spectral library (using Spectronaut™ 15’s “Search Archives” feature).

Raw SWATH (.wiff) data files were converted into Spectronaut™ HTRMS format and analysed using Spectronaut™ 15. The default settings that were used for targeted analysis were: dynamic iRT prediction with correction factor for window 1; mass calibration was set to local; decoy method set as scrambled; the FDR, based on the mProphet approach (Reiter et al., 2011), set at 1% on the precursor peptide and protein group levels; protein inference set to “default” which is based on the ID picker algorithm (Zhang et al., 2007) and global cross-run normalisation on median. The concatenated urinary proteome spectral library (peptides – 20 616, protein groups – 2 604) was used as a reference for targeted data extraction.

Following acquisition, data was curated and filtered for sample preparation and/or LCMS related failures.

4.3.7. Statistical and functional data annotation

For the table of descriptive data, t-tests and a Chi-squared analysis using STATA were used to compute associations between the case and control group, as appropriate.

Protein abundances were imported into SRplot and log₂ transformed. Principal component analysis (PCA) based on protein abundances was performed to assess stratification within the cohort. A non-parametric t-test was performed to assess the number of proteins that

differed between the case and control group. Statistically significant differentially abundant proteins between the cases and controls were calculated by a two-sided t-test, with a cut-off minimum of 2-fold difference and *P* values adjusted for multiple testing by FDR at 1%.

The differentially abundant proteins were constructed into networks and annotated using Cytoscape and STRING functional enrichment (Shannon et al., 2003). Enrichment was considered statistically significant when corrected for multiple testing by FDR with adjusted q-values <0.05. Additional networks were imported on Cytoscape using public databases (including Pathway Commons, IntAct, BioMart, and NCBI Entrez Gene) for hypertension and CKD with a high confidence score of 0.8 (i.e., 80% confidence) and maximum 200 proteins.

4.3.8. Machine learning

OmicLearn (v1.4) was used for data analysis, model execution, and creation of plots and charts (Torun et al., 2022). Machine learning was performed in Python (3.10.12). Spectronaut feature tables were imported via the Pandas package (1.5.3) and manipulated using the Numpy package (1.24.2). The machine learning pipeline was employed using the scikit-learn package (1.2.2). The Plotly (5.9.0) library was used to generate plots and charts. No normalisation on the data was performed. To impute missing values, a Zero-imputation strategy was used. Features were selected using a ExtraTrees (n_trees=100) strategy with a maximum number of 20 features. During training, normalisation and feature selection was individually performed using the data of each split. A repeated (RepeatedStratifiedKFold, n_repeats=10), stratified cross-validation (n_splits=5) approach to classify the groups was used. This resulted in total 50 iterations for training the model, each time with 20 features. The XGBoost-Classifer (version: 1.7.4, random_state=23, learning_rate=0.3, min_split_loss=0, max_depth=6, min_child_weight=1) was used for classification.

4.4. Results

4.4.1. Study participants

Of the initial 90 participants selected for proteomic analyses, nine were removed due to poor peptide/protein recoveries and eight were removed due to poor liquid chromatography separation possibly due to incomplete removal of sample specific contamination during

sample preparation. Baseline characteristics of the overall ARK study and the 24 case and 49 control participants whose data passed quality control are displayed in **Table 4.1**.

Table 4.1: Baseline characteristics of participants in the ARK study and this case-control study

Variable	Overall ARK study (N=2,021)	Cases (n=24)	Controls (n=49)	P value ^a
Age (years), mean (SD)	38 (14)	38 (13)	42 (14)	0.255
Sex - Male, n (%)	851 (42)	16 (67)	31 (63)	0.776
BMI Kg/m ² , mean (SD)	27 (6)	31 (7)	26 (7)	0.005
eGFR (ml/min/1.73m ²), ^b mean (SD)	112 (18)	101 (23)	116 (14)	0.044
UACR (mg/mmol), median (IQR)	0.4 (0.2–1.2)	9.4 (4.5–18.7)	0.4 (0.2–0.7)	<0.001 ^c
Systolic blood pressure mm Hg, mean (SD)	128 (17)	162 (20)	123 (9)	<0.001
Diastolic blood pressure mm Hg, mean (SD)	78 (11)	98 (14)	74 (7)	<0.001
Hypercholesterolaemia, ^d n (%)	465 (23)	9 (38)	13 (27)	0.337
APOL1 low-risk (zero or one risk allele), n (%)	1776 (89)	19 (79)	37 (76) ^e	0.965
APOL1 high-risk (two risk alleles), n (%)	226 (11)	5 (21)	10 (20) ^e	0.965

^aComparison between cases and controls.

^bMeasured using the Chronic Kidney Disease Epidemiology (CKD-EPI) equation.

^cT-test performed on log-transformed data.

^dNon-fasting total cholesterol >5.0mmol/L.

^eUnable to interpret *APOL1* status in two controls.

APOL1, Apolipoprotein L1; BMI, body mass index; eGFR, estimated glomerular filtration rate; IR, interquartile range; SD, standard deviation; UACR, urinary albumin-to-creatinine ratio.

The age and sex distribution of participants was similar in the case and control groups. The case group had a higher mean body mass index compared to controls, and as expected, blood pressure and UACR differed between the groups. In addition, *apolipoprotein L1* (*APOL1*) allele distribution was similar between cases and controls.

4.4.2. Project specific system suitability-quality control

The study specific and commercial system suitability quality controls are shown in (Supplementary Figure 4.1A–F). The protein co-efficient of variation (CV) was 11.6% for the HeLa digest system suitability assessment and 18.0% for the urine peptide pool over the seven days of analysis (Supplementary Figure 4.1A and B). Protein, peptide and precursor counts were stable throughout the data acquisition process (Supplementary Figure 4.1C and D). The correlation plots indicate the system was stable with minimal drift over the course of data acquisition (Supplementary Figure 4.1E and F).

4.4.3. Identification of proteins by LC-MS analysis

The spectral library represents all possible proteins that are detectable using the LCMS configuration that was setup for this study. Only a subset of all these are detectable in each sample hence the average numbers of 915 (± 264) proteins in the case group and 1,225 (± 227) in the control group (Figure 4.2, Supplementary Figure 4.2). The number of quantified proteins significantly differed between the case and control groups ($P < 0.001$).

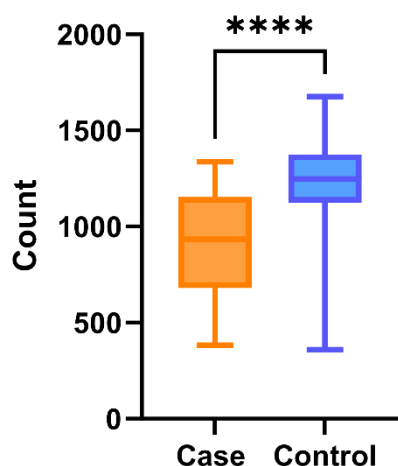


Figure 4.2: Summary of quantified proteins

The number of proteins quantified post SWATH LC-MS analysis for each sample and grouped per condition.

LC-MS, liquid chromatography with mass spectrometry.

****Denotes $P < 0.001$ performed using a non-parametric t-test.

4.4.4. Principal Component (PC) analysis

PC analysis using protein abundances highlighted differential clustering of the two groups with some overlap (**Figure 4.3**).

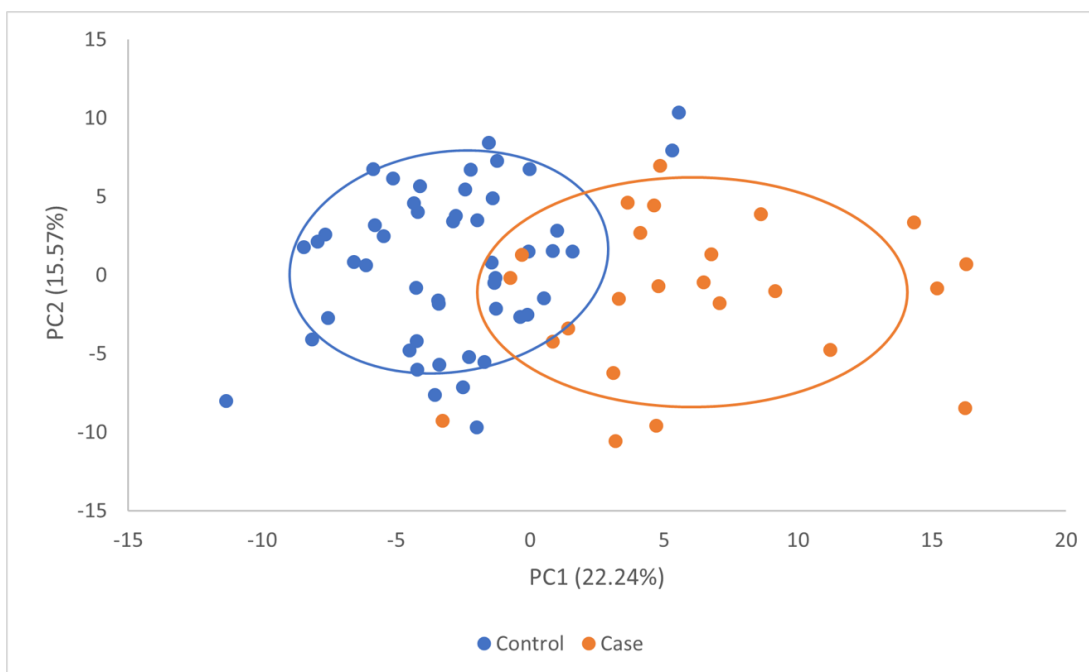


Figure 4.3: PC (1 and 2) analysis based on protein abundance features

PC analysis using log-transformed protein abundance values indicates clustering of cases and controls. Orange and blue circles represent cases and controls, respectively. PC, principal component.

4.4.5. Identification of proteins with significantly different abundances between cases and controls

Differentially abundant proteins between the cases and controls are shown in **Figure 4.4**. Majority of the proteins deemed statistically different in abundances between cases and controls, were found to have lower abundance in the cases. The full list of differential protein abundances is shown in **Supplementary Table 4.1**.

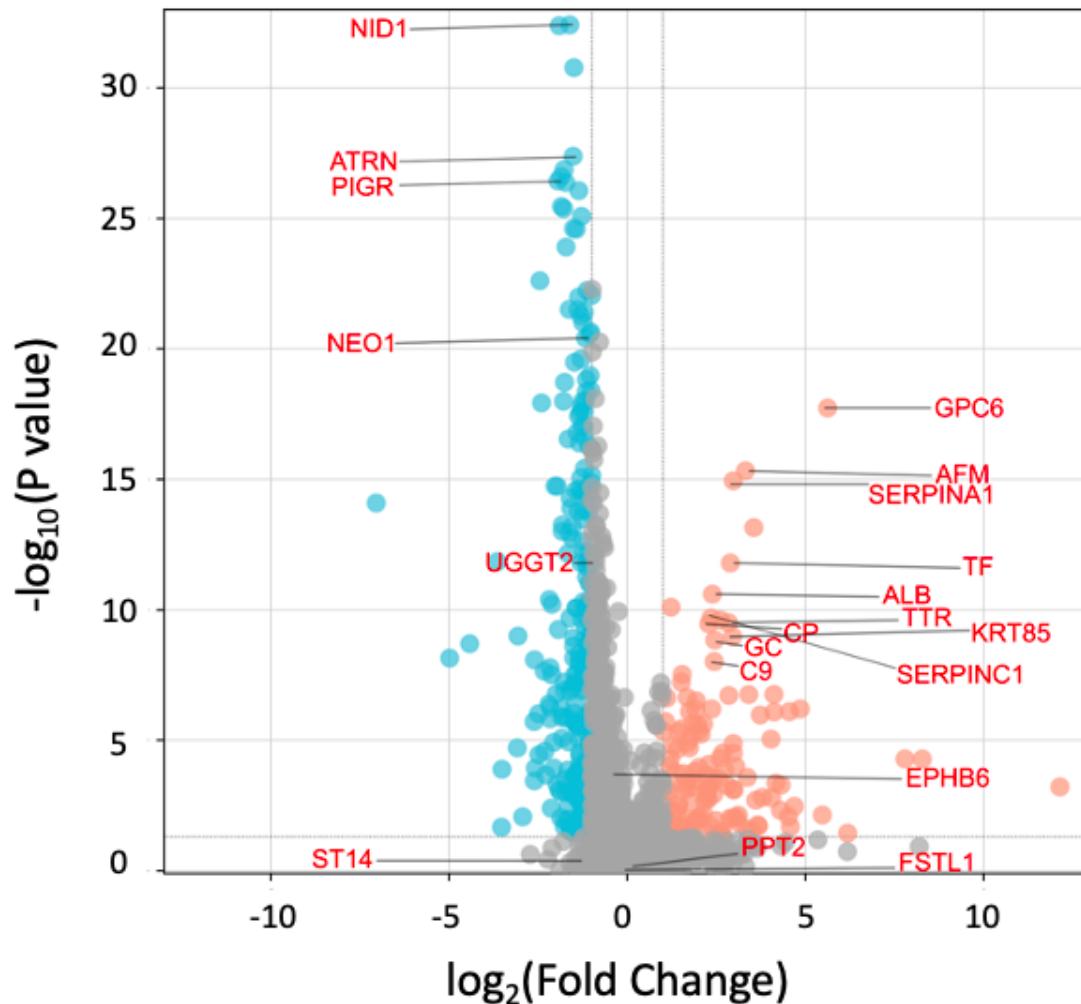


Figure 4.4: Differential protein abundances between cases and controls illustrated in a volcano plot

Proteins were graphed by fold change (difference) on the x-axis and significance ($-\log_{10} P$) on the y-axis using an FDR of 0.01 and a fold-change of 2. Statistical analysis was performed in Perseus. Orange dots indicate proteins with higher abundances in the cases. Blue dots indicate proteins with lower abundances in the cases. The top 20 features of the machine learning algorithm are annotated in red.

FDR, false discovery rate.

4.4.6. Machine learning model

The performance of the machine learning algorithm was evaluated by receiver operating characteristics analysis with an area under the curve of 0.98 and confusion matrix analysis where predicted and actual classification closely matched (**Supplementary Figure 4.3**).

The top 20 features identified from the classifier are shown in **Figures 4.4 and 4.5**. When

comparing the top 20 features to the list of 332 differentially abundant proteins identified (using a Student's t-test), 16 were found to overlap and four were unique to the XgBoost algorithm. (**Supplementary Table 4.2**).

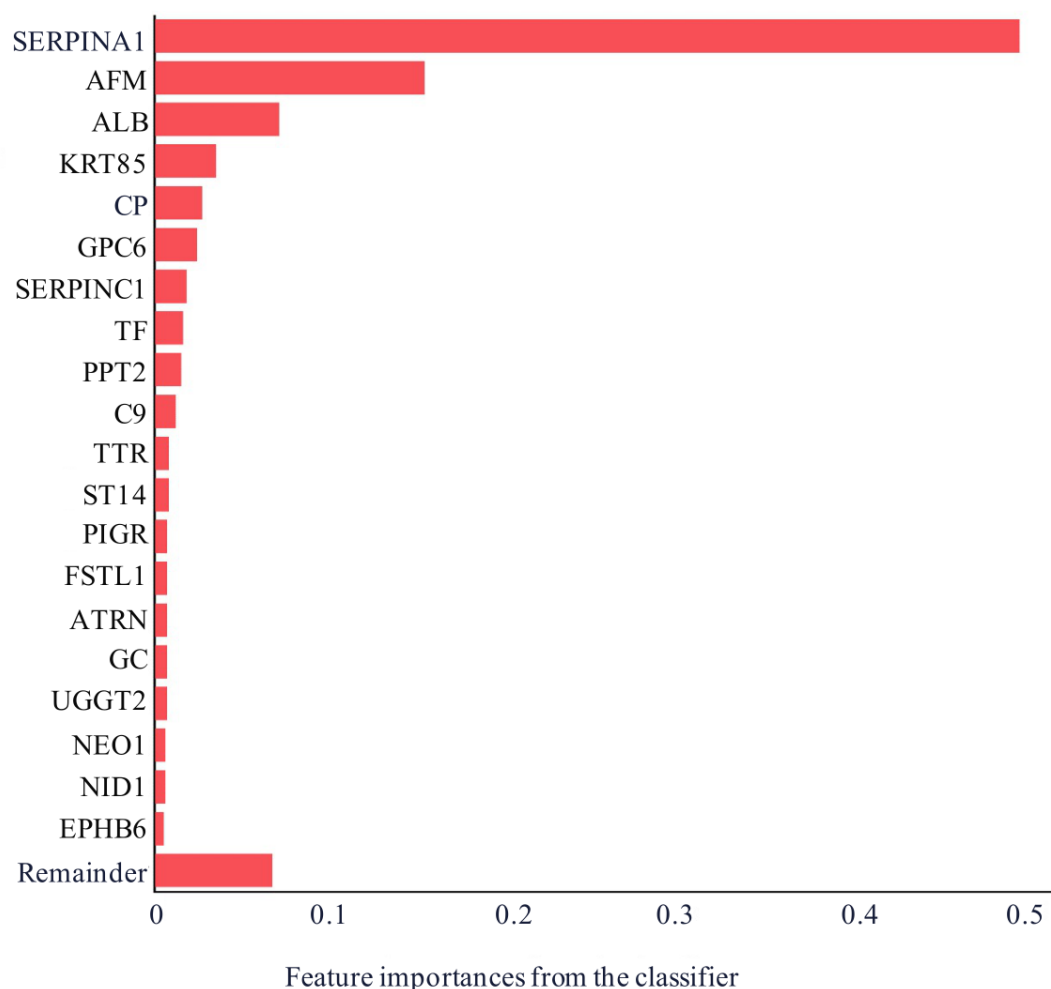


Figure 4.5: Top 20 features identified from the classifier using the XGBoost algorithm
Gene names for each protein are shown.

4.4.7. Functional annotation of proteins with significantly different abundances between cases and controls

Three hundred and thirty-two proteins with significantly different abundances between cases and controls were constructed into a network and annotated using Cytoscape. To gain more insight on the biological significance of these proteins, STRING enrichment was

performed. The network was split into sub clusters based on relevant pathways related to CKD. We identified 112 proteins associated with the immune system (q-value [FDR]= 1.4×10^{-45}), and 89 proteins associated with the innate immune system ($q=1.1 \times 10^{-32}$) (**Supplementary Figures 4.4 and 4.5**). Additionally, 17 and six proteins were associated with extracellular matrix (ECM) organisation ($q=0.03$) and activation of matrix metalloproteinases ($q=0.04$), respectively (**Supplementary Figure 4.6A and B**).

Using Cytoscape public databases, 14 common proteins were identified between this pilot case-control study and other studies on CKD and hypertension, as imported from public databases (**Figure 4.6**). The size of the circle represents the disease score, where a score of 0 indicates that the protein is not related to the disease and a score of 5 indicates that the protein is related to the disease. Proteins angiotensinogen (AGT), albumin (ALB), APOL1, and uromodulin (UMOD) were associated with CKD and hypertension with high disease scores of 5.0 (100% confidence), 4.0 (80%), 3.9 (78%), and 3.8 (76%), respectively. The proteins matrix metalloproteinase 9 (MMP9) and TIMP Metallopeptidase Inhibitor 1 (TIMP1) were associated with pathways involved in degradation of the ECM, interleukin (IL)-4 and IL-13 signalling, and activation of matrix metalloproteinases.

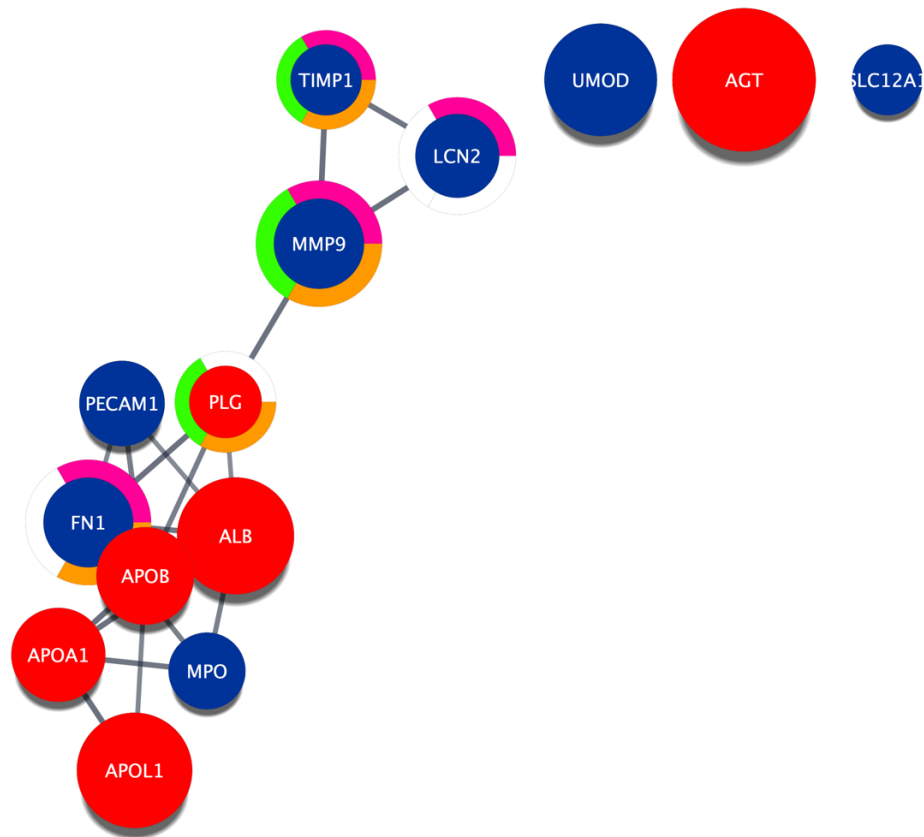


Figure 4.6: Common proteins between this case-control pilot study and other studies on CKD and hypertension, as imported from public databases

The gene names for each protein in the network are shown. Red indicates proteins with higher abundance in cases while blue represents proteins with lower abundance.

Size of the circle represents the disease score (0 [not related to disease] – 5 [related to disease]). The closer to 5, the higher the confidence that the protein is associated with CKD and hypertension.

AGT=5.0; ALB=4.0; APOL1=3.9; UMOD=3.8; APOB=3.2; APOA1=3.1; MMP9=2.9; FN1=2.9; PECAM1=2.7; LCN2=2.7; MPO=2.4; PLG=2.2; TIMP1=2.2; SLC12A1=2.2.

Edges represent protein-to-protein interactions that were selected by default in Cytoscape. Pathways are annotated on the network using a split donut ring. Orange indicates proteins associated with degradation of the extracellular matrix; pink indicates proteins associated with IL-4 and IL-13 signalling; green indicates proteins associated with activation of matrix metalloproteinases.

CKD, chronic kidney disease; IL, interleukin.

4.5. Discussion

This case-control pilot study aimed to identify potential proteins and pathways involved in hypertension-associated albuminuria by assessing urinary proteomic profiles in black South African participants with both hypertension and albuminuria compared to those who have neither condition.

Peptide and protein CVs showed that the methods were reproducible and that the system and workflow contributed low technical variability. The instrumental drift observed throughout the data acquisition period was minimal, as shown in the commercial HeLa digest system suitability data that were acquired from a fresh sample on each day. A larger drift was observed in the study-specific peptide pool system suitability data; however, this can be attributed to autosampler-associated degradation as the sample was kept in the autosampler throughout the data acquisition process. This suggests that most of the variation observed in the study can be attributed to biological variation. On average, more proteins (1,225) were quantified in the control group than in the case group (915). Possible explanations include that a minor difference in the missed cleavage rate (~3%) was observed between the two groups and several high abundance proteins, such as ALB, *APOL1*, apolipoprotein A1 (*APOA1*), and apolipoprotein B (*APOB*) are significantly higher in abundance in the cases ‘masking’ identification of lower abundance proteins, resulting in fewer quantifications overall.

The majority of differentially abundant proteins were associated with pathways involving the immune system, innate immune system, extracellular matrix organisation and activation of matrix metalloproteinases. In the setting of sustained hypertension, high blood pressure can cause permanent damage to the nephrons through arterial injury and glomerular ischaemia (Benson et al., 2023). Damage from renal ischaemia can lead to the production of various inflammatory cytokines that lead to immune cell infiltration and tubular atrophy (Benson et al., 2023). The latter is supported by Mattson (2014), (Mattson, 2014) who showed an infiltration of inflammatory mononuclear cells in the arterioles and small arteries in kidney tissue from patients with hypertension. Enrichment analysis has shown that pathways related to inflammation and immune response are associated with the progression of kidney disease (Trudu et al., 2017).

We identified six proteins associated with the activation of matrix metalloproteinase (MMPs) pathways. Of these, five had a lower abundance in participants with hypertension-associated albuminuria. It has been suggested that decreased activity of matrix metalloproteases may cause an accumulation of proteins in the extracellular matrix leading to fibrosis, which is one of the features of hypertension-associated kidney disease (Hughson et al., 2014, Bazzi et al., 2020, Cheng et al., 2006). Most of the proteins (71%; n=12/17) associated with the ECM organisation pathway were also reduced in participants with hypertension-associated albuminuria. It has been suggested that the reduced abundance of ECM organisation proteins may reflect reduced ECM turnover (decreased degradation), subsequently causing an increased deposit in ECM, resulting in fibrosis (Nkuipou-Kenfack et al., 2014). Currently, kidney biopsy is the only validated approach to evaluate fibrosis. The identification of ECM organisation proteins may provide a non-invasive approach to detect early modifications in the ECM that lead to renal fibrosis.

Fourteen common proteins were identified between this pilot study and public databases on CKD and hypertension. Proteins including AGT, ALB, APOL1, and UMOD had the highest disease scores (76–100% confidence) for CKD and hypertension. The PLG protein has been shown to regulate macrophage migration in inflammation through activation of the MMP9 protein, which regulates the ability of the cells to migrate across the ECM (Gong et al., 2008). Therefore, targeting the PLG/MMP9 pathways may lead to regulation of inflammatory responses in CKD.

In this pilot study, UMOD had a lower abundance in cases with hypertension-associated albuminuria. Similarly, Nqebelele et al. 2019 found reduced levels of UMOD in black African individuals with hypertensive-attributed CKD (Nqebelele et al., 2019). Navise et al. (2023) also found lower levels of UMOD in black African individuals with CKD (Navise et al., 2023). In addition to its role in ECM remodelling and fibrosis, UMOD has been involved in regulating ion transport in the kidney (Renigunta et al., 2011). Reduced levels of UMOD have been postulated to be a result of decreased secretion from damaged tubules (Prajczar et al., 2010). Results from a mouse model with altered UMOD expression have shown that the induction of pro-inflammatory signalling is one of the first events that occur in the kidneys (Trudu et al., 2017). The animal study also showed that the initiation of

inflammatory signals precedes fibrosis and kidney damage, and possibly plays a vital role in disease onset (Trudu et al., 2017).

It is important to note that hypercholesterolaemia was present in 38% of the cases in this study. Several studies have linked apolipoproteins (*APOA1*, *APOB* and *APOL1*) with CKD (Goek et al., 2012, Zhao and Alberto, 2020, Ma et al., 2019) and hypertension (Nayak et al., 2016). Moreover, a Mendelian-randomization study identified a causal association between lipids (including low-density lipoprotein and triglycerides) and kidney function in individuals of African ancestry (Kintu et al., 2023). It has been postulated that the association between apolipoproteins and kidney disease may be mediated by the impact of these lipoprotein particles on the kidney. Apolipoproteins and their associated lipids may have toxic effects on glomerular cells leading to glomerulosclerosis, which is scarring of the glomerulus (Kwon et al., 2021). As a result, Kidney Disease: Improving Global Outcomes (KDIGO) developed guidelines for lipid management in patients with CKD (Wanner and Tonelli, 2014).

Machine learning enabled us to train a classifier that on average correctly identified cases with an 88% true positive rate. Additionally, it identified controls with a 97% true negative rate. This analysis revealed that alpha-1-antitrypsin (P01009) and afamin (P43652) were the two most important features of the classifier. Studies have shown that alpha-1-antitrypsin and afamin are potential biomarkers for the diagnosis of early diabetic kidney disease (DKD) and can predict the decline in renal function (Ning et al., 2020, Kaburagi et al., 2019). Interestingly, in this pilot study, we have identified DKD markers in normoglycemia participants who have hypertension and albuminuria, which suggests the generalisability of the findings.

This is the first study to characterise the urinary proteome in South African individuals with hypertension-associated albuminuria. This case-control study identified common proteins with previous studies on CKD and hypertension from other ethnicities, and proteins associated with DKD, indicating generalisability of the findings to other populations and diseases. This study contributes to the understanding of pathways involved in

hypertension-associated albuminuria and suggest opportunities for translation into the clinical setting.

The results from this study must be interpreted in context of its limitations, which included a relatively small sample size, consisting of only 73 individuals. All participants are from a single centre, therefore, results may not be generalisable. The machine learning models applied in the current study show preliminary data and will require validation in a larger dataset with blinded data to assess overfitting.

Further research is essential to validate these outcomes in larger cohorts. In addition, the investigation of other diseases such as DKD and acute kidney injury would help to further explore generalisability of findings.

4.6. Conclusions

In this pilot study, pathways associated with hypertension-associated albuminuria included the immune system, innate immune system, ECM organisation and activation of matrix metalloproteinases. These pathways contribute insights into the pathophysiology (e.g., immune cell infiltration, tubular atrophy, and ECM remodelling) of hypertension-related albuminuria. Additionally, proteins such as AGT, ALB, *APOL1* and UMOD had the highest disease scores (76–100% confidence) for hypertension and CKD. The urinary proteomic data combined with the machine learning approach classified disease status and identified proteins and pathways associated with hypertension and albuminuria.

4.7. Funding declaration

The ARK study was also jointly funded by the South African MRC, MRC UK (via the Newton Fund), and GSK Africa Non-Communicable Disease Open Lab (via a supporting grant; project number 074); the National Health Laboratory Services, South Africa; and Faculty of Health Sciences Research Incentive Grant (number 0012838434203512110500000000000000004550), University of the Witwatersrand, South Africa. Funding supported implementation, execution, and completion of the study in South Africa. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. JTB is funded by the Cancer Association of South Africa and the Science for African Foundation - REACCT-CAN Grant (Del-22-008).

MR is a South African Research Chair in Genomics and Bioinformatics of African populations hosted by the University of the Witwatersrand, funded by the Department of Science and Technology, and administered by the NRF. The remaining authors have no funding.

4.8. Declarations

The study was approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (certificate number, M200331). All participants provided written informed consent

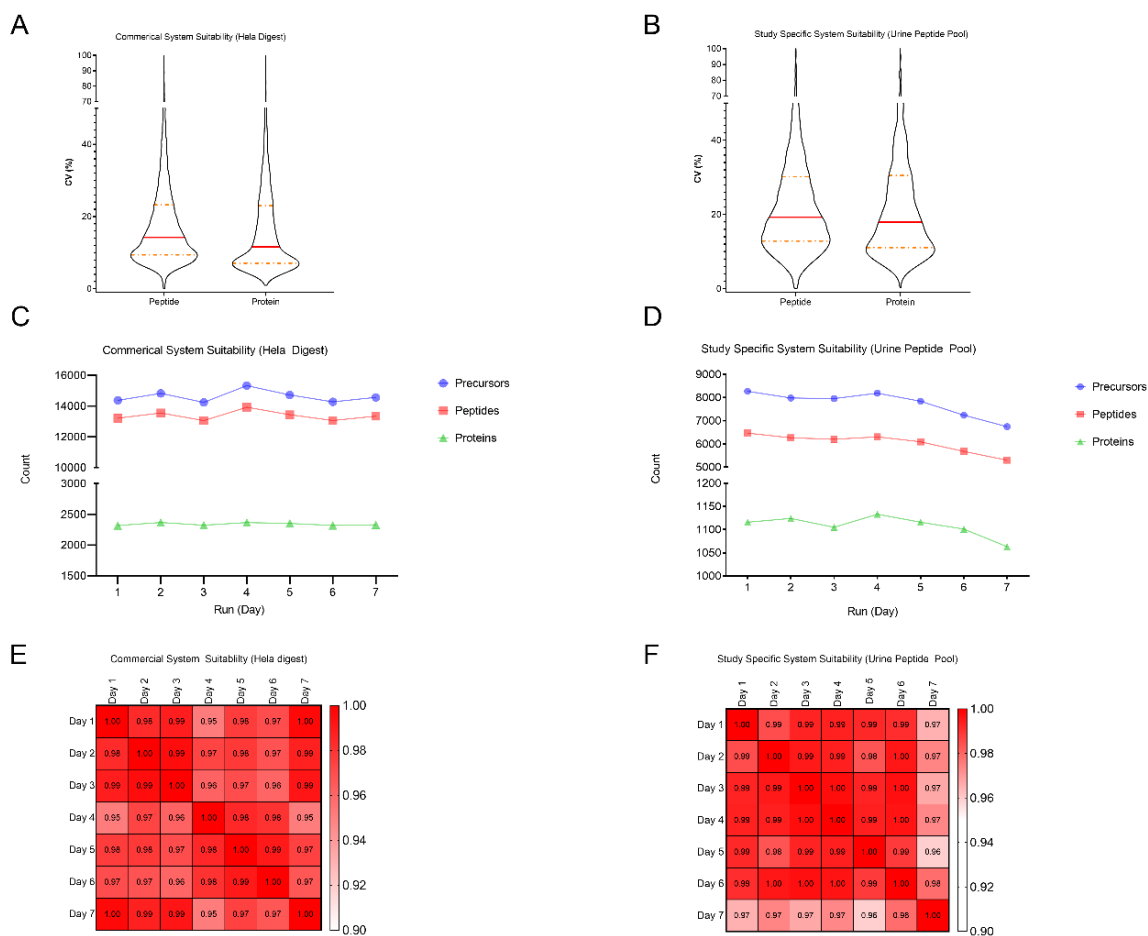
4.9. Author contributions

All authors have contributed to the design of the study. MAG and ISG wrote the main manuscript text and prepared figures. ISG, SHS, MR, JF, and J-TB critically reviewed the manuscript.

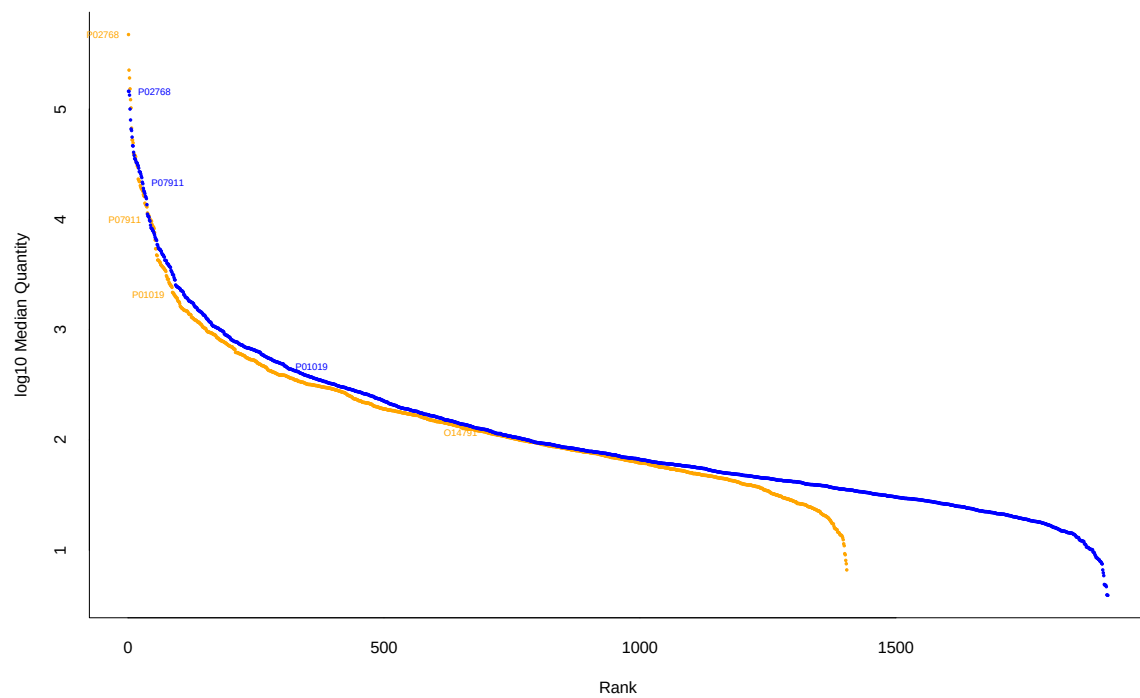
4.10. Data availability

The proteomic data files described in this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier (PXD046441) (Perez-Riverol et al., 2022).

4.11. Supplementary material

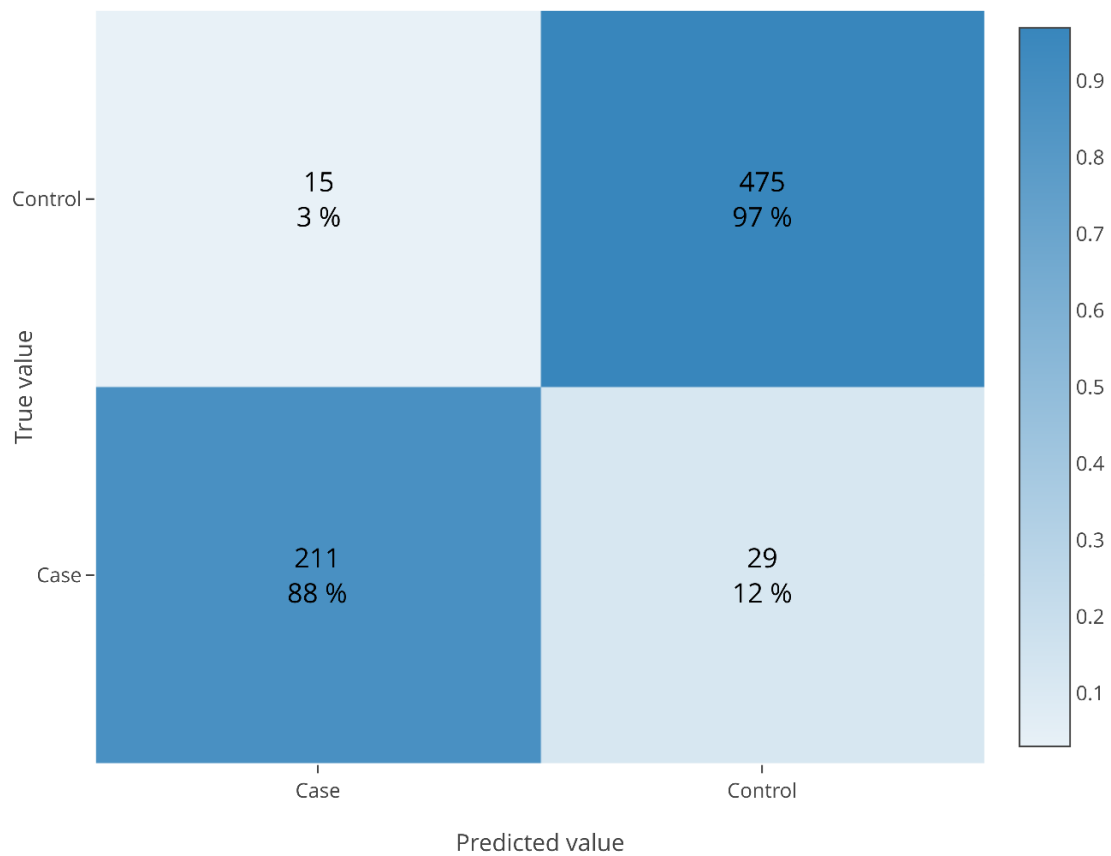


Supplementary Figure 4.1: Project specific system suitability-quality control
CV, co-efficient of variation.



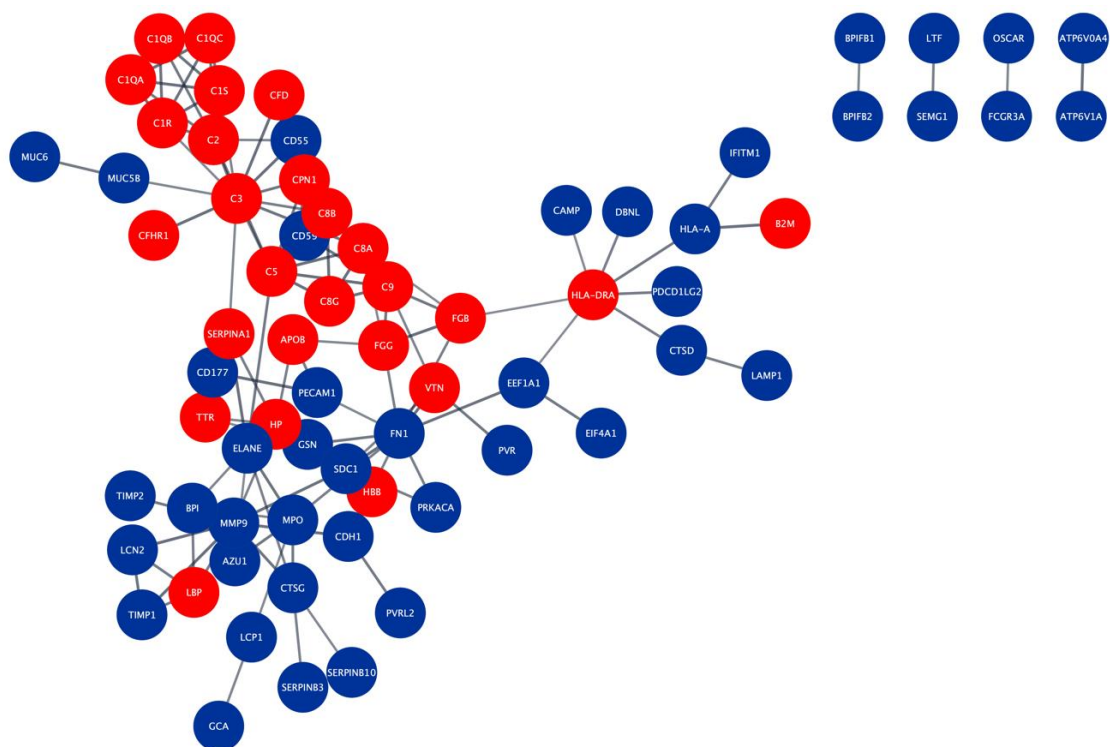
Supplementary Figure 4.2: Dynamic range of proteins quantified

Orange dots indicate cases and blue dots indicate controls. The high disease scoring proteins are annotated.



Supplementary Figure 4.3: Confusion matrix following machine learning using the XGBoost algorithm

Confusion matrix refers to the percentage of samples assigned to each group.



Supplementary Figure 4.4: Differentially abundant proteins associated with the immune system

The gene names for each protein in the network are shown. Differentially abundant proteins were detected at an FDR of 0.01 and a fold-change of 2. Red indicates proteins with higher abundance in cases while blue represents proteins with lower abundance.

FDR, false discovery rate.

Supplementary Table 4.1: Significantly different proteins in the volcano plot

Protein Groups	Difference	Direction	Log q-value
P43652	4,01	Positive	14,79
P25691; P78386	3,65	Positive	10,24
Q9Y625	3,03	Positive	9,85
P02787	3,00	Positive	13,76
P02144	2,87	Positive	8,61
P01024	2,58	Positive	7,00
P01009	2,66	Positive	18,70
P02647	2,51	Positive	5,45
Q8WYP5	2,49	Positive	7,59
P02766	2,43	Positive	7,54
P01008	2,38	Positive	10,34
P19823	2,35	Positive	5,40
Q9UGM5	2,23	Positive	6,23
P69905	2,15	Positive	2,55
Q9NYQ6	2,10	Positive	2,80
P06727	2,07	Positive	4,21
P00450	2,00	Positive	8,21
P02774	1,99	Positive	9,26
P02748	1,99	Positive	7,57
P02768	1,98	Positive	10,99
P00747	1,93	Positive	5,55
P36955	1,89	Positive	7,18
P27169	1,85	Positive	4,34
P00738	1,81	Positive	2,89
P01019	1,80	Positive	5,35
P01860	1,77	Positive	3,08
P08697	1,77	Positive	7,77
P06681	1,76	Positive	6,70
P00739	1,73	Positive	3,73
P18428	1,66	Positive	4,77
P04217	1,64	Positive	5,17

Q96KN2	1,50	Positive	5,10
P07357	1,48	Positive	5,68
P07360	1,47	Positive	3,36
A0A0C4DH38	1,46	Positive	3,17
P08185	1,37	Positive	3,76
P28072	-1,15	Negative	4,44
Q9UN73; Q9Y5H7	-1,16	Negative	4,15
P28070	-1,17	Negative	4,09
P06733	-1,18	Negative	4,79
P60174	-1,21	Negative	5,01
Q12907	-1,21	Negative	4,45
P04275	-1,22	Negative	4,13
P05362	-1,22	Negative	3,88
P27930	-1,22	Negative	4,15
Q8TER0	-1,22	Negative	6,23
Q9HAT2	-1,22	Negative	3,79
O00462	-1,23	Negative	5,53
P05067	-1,24	Negative	4,84
O75083	-1,24	Negative	4,40
P16083	-1,26	Negative	3,82
A0A0C4DH25	-1,27	Negative	3,92
Q99497	-1,27	Negative	4,76
O00182	-1,28	Negative	3,70
Q15113	-1,29	Negative	4,94
Q08257	-1,29	Negative	4,93
Q99536	-1,29	Negative	3,69
P21709	-1,30	Negative	4,12
P50395	-1,30	Negative	4,63
Q04760	-1,31	Negative	3,57
O00241	-1,31	Negative	3,57
P68104; Q5VTE0	-1,31	Negative	4,58
P40121	-1,32	Negative	6,02

P41222	-1,32	Negative	3,49
Q9P2B2	-1,32	Negative	5,32
Q16651	-1,32	Negative	4,51
Q14894	-1,33	Negative	4,81
Q96J84	-1,34	Negative	5,58
P51148	-1,34	Negative	4,85
Q14393	-1,34	Negative	4,11
O94919	-1,34	Negative	3,39
P11021	-1,34	Negative	4,88
Q96JQ0	-1,34	Negative	4,58
P09619	-1,35	Negative	3,77
Q15223	-1,35	Negative	4,84
Q06830	-1,36	Negative	5,40
P07204	-1,36	Negative	4,16
P07711	-1,37	Negative	6,11
P13639	-1,38	Negative	4,76
P11233	-1,39	Negative	4,37
Q9H8L6	-1,39	Negative	4,22
P05090	-1,39	Negative	3,94
P38606	-1,39	Negative	3,90
Q96AP7	-1,39	Negative	3,15
Q86T13	-1,40	Negative	5,49
P23526	-1,41	Negative	5,12
Q13332	-1,41	Negative	4,95
P30086	-1,41	Negative	4,22
Q13621	-1,41	Negative	3,45
P23284	-1,42	Negative	4,55
P17174	-1,42	Negative	4,20
Q8WVQ1	-1,43	Negative	6,00
P61970	-1,44	Negative	5,76
P08582	-1,45	Negative	5,80
Q8N3J6	-1,45	Negative	6,53

P11597	-1,45	Negative	3,13
P14174	-1,46	Negative	3,53
P52758	-1,46	Negative	3,49
O95967	-1,46	Negative	5,46
P55285	-1,47	Negative	4,18
P35555	-1,48	Negative	3,55
Q7Z7M0	-1,48	Negative	4,99
Q14118	-1,48	Negative	3,66
Q9HCU0	-1,49	Negative	4,99
P04156	-1,50	Negative	4,36
Q9Y624	-1,51	Negative	4,05
O00757	-1,51	Negative	2,98
P05937	-1,51	Negative	3,67
Q16769	-1,52	Negative	3,44
P12109	-1,52	Negative	5,28
P15941	-1,53	Negative	3,22
O60888	-1,53	Negative	3,06
O00391	-1,54	Negative	6,36
P0DJD8	-1,55	Negative	4,96
Q6UX71	-1,55	Negative	6,07
P01133	-1,55	Negative	4,05
P12111	-1,55	Negative	4,38
P06396	-1,56	Negative	3,64
Q10588	-1,56	Negative	4,48
Q99983	-1,56	Negative	4,85
P42785	-1,56	Negative	4,36
P19022	-1,57	Negative	3,82
P14618	-1,57	Negative	3,55
O75882	-1,57	Negative	5,56
P13645	-1,57	Negative	2,95
A6NI73	-1,57	Negative	5,77
Q53RD9	-1,57	Negative	4,44

Q01459	-1,58	Negative	4,02
P29323	-1,58	Negative	6,59
Q92692	-1,58	Negative	4,22
P08572	-1,58	Negative	3,66
P15121	-1,59	Negative	6,18
P00966	-1,59	Negative	2,82
Q9HCN6	-1,60	Negative	3,51
O00115	-1,60	Negative	6,66
Q6UXB4	-1,60	Negative	3,35
O95865	-1,61	Negative	3,90
P98160	-1,61	Negative	3,82
Q12860	-1,61	Negative	8,78
Q8WW52	-1,62	Negative	3,89
P16035	-1,62	Negative	8,09
P10153	-1,62	Negative	3,36
O75144	-1,63	Negative	3,72
Q01973	-1,63	Negative	5,42
P35908	-1,63	Negative	3,20
Q9UNN8	-1,64	Negative	4,34
Q07507	-1,64	Negative	4,28
P04899	-1,65	Negative	5,41
P55287	-1,65	Negative	4,26
P56537	-1,66	Negative	5,07
P19440	-1,66	Negative	3,75
P63000	-1,66	Negative	4,48
Q9UKU9	-1,67	Negative	5,25
P26992	-1,67	Negative	6,09
P24821	-1,68	Negative	5,45
O00592	-1,68	Negative	3,19
P10253	-1,68	Negative	6,72
Q7Z3B1	-1,68	Negative	4,13
P07858	-1,70	Negative	4,99

P07911	-1,71	Negative	5,28
P08294	-1,71	Negative	4,92
P19835	-1,72	Negative	4,07
Q9HD42	-1,72	Negative	2,98
P08571	-1,72	Negative	3,03
Q15375	-1,72	Negative	7,75
Q9Y2E5	-1,72	Negative	5,31
Q92859	-1,73	Negative	6,10
Q12805	-1,73	Negative	4,70
P07998	-1,74	Negative	3,02
Q15907	-1,74	Negative	7,36
P01591	-1,74	Negative	3,88
Q15746	-1,74	Negative	3,74
Q9HBB8	-1,75	Negative	3,18
P05091	-1,76	Negative	4,74
O43451	-1,76	Negative	3,86
P05452	-1,76	Negative	4,84
P63092; Q5JWF2	-1,76	Negative	8,03
Q9UBX5	-1,76	Negative	5,63
Q99574	-1,77	Negative	6,23
O15197	-1,77	Negative	7,21
P98164	-1,78	Negative	5,99
A0A075B6I9	-1,79	Negative	3,21
P08519	-1,79	Negative	2,74
P14543	-1,80	Negative	8,06
P12429	-1,81	Negative	3,26
Q8IUL8	-1,82	Negative	5,10
Q08174	-1,83	Negative	5,27
Q9UIB8	-1,83	Negative	5,14
P43121	-1,84	Negative	7,49
Q8N6C8	-1,86	Negative	6,47
Q8WZ75	-1,86	Negative	5,74

Q9UBD6	-1,86	Negative	3,71
P58499	-1,87	Negative	7,58
Q9BXP8	-1,87	Negative	6,15
Q9UBI6	-1,88	Negative	2,42
Q9Y6R7	-1,88	Negative	2,47
Q8IV08	-1,89	Negative	7,72
P40189	-1,89	Negative	8,51
Q5T011	-1,89	Negative	2,96
P61916	-1,90	Negative	5,15
O75339	-1,90	Negative	5,41
Q03403	-1,91	Negative	3,91
P06703	-1,94	Negative	3,30
O43278	-1,94	Negative	10,56
Q9UJX4	-1,95	Negative	2,71
P05451	-1,95	Negative	2,74
Q7Z3E2	-1,96	Negative	4,84
P54760	-1,97	Negative	9,31
P98172	-1,98	Negative	6,59
Q8IYS5	-1,99	Negative	4,65
P50895	-1,99	Negative	6,86
P12830	-2,01	Negative	5,57
P11279	-2,02	Negative	5,66
Q14956	-2,02	Negative	4,90
Q15828	-2,04	Negative	2,66
P01833	-2,04	Negative	10,27
P80188	-2,05	Negative	4,43
P02751	-2,05	Negative	6,26
Q96RW7	-2,09	Negative	6,41
Q6UVK1	-2,11	Negative	7,77
Q9GZX9	-2,13	Negative	3,20
P14384	-2,14	Negative	8,54
O00533	-2,15	Negative	7,84

Q969P0	-2,16	Negative	6,44
Q7KYR7; Q8WVV5	-2,17	Negative	5,58
P19961	-2,18	Negative	4,81
Q9P121	-2,22	Negative	8,78
Q6UXG3	-2,22	Negative	3,94
P08174	-2,22	Negative	5,44
Q9BX67	-2,22	Negative	5,98
O60494	-2,23	Negative	8,65
Q9UMR5	-2,26	Negative	8,14
P04745; P0DTE7; P0DTE8; P0DUB6	-2,27	Negative	7,08
Q96NY8	-2,30	Negative	7,71
O95716	-2,32	Negative	6,10
P08637	-2,33	Negative	5,30
P06870	-2,33	Negative	5,05
Q8NFZ8	-2,34	Negative	7,58
P0DP57	-2,35	Negative	5,12
P54753	-2,37	Negative	12,14
Q8N6Q3	-2,37	Negative	6,42
Q9UN70	-2,39	Negative	8,09
Q14982	-2,42	Negative	8,03
P15328	-2,44	Negative	6,89
P62879	-2,46	Negative	8,12
P78324	-2,47	Negative	8,97
P13987	-2,51	Negative	6,63
P30530	-2,52	Negative	7,16
Q9BQ51	-2,54	Negative	7,30
O75594	-2,54	Negative	8,31
Q8NBJ4	-2,85	Negative	7,86
Q9NYU1; Q9NYU2	-3,09	Negative	9,70
Q9H299	-3,10	Negative	4,12

Supplementary Table 4.2: Top 20 proteins identified from the classifier

Common^a	Unique^b
P01009 (SERPINA1)	Q9UMR5 (PPT2)
P43652 (AFM)	Q9Y5Y6 (ST14)
P02768 (ALB)	Q12841 (FSTL1)
P78386 (KRT85)	O15197 (EPHB6)
P00450 (CP)	
Q9Y625 (GPC6)	
P01008 (SERPINC1)	
P02787 (TF)	
P02748 (C9)	
P02766 (TTR)	
P01833 (PIGR)	
O75882 (ATRN)	
P02774 (GC)	
Q9NYU1 (UGGT2)	
Q92859 (NEO1)	
P14543 (NID1)	

Gene names are shown in parentheses.

^aOverlap with the list differentially abundant proteins.

^bUnique to the machine learning algorithm.

CV, co-efficient of variation.

CHAPTER 5

5. INTEGRATED DISCUSSION AND CONCLUSIONS

5.1. Study overview

In Agincourt, South Africa, there is an existing burden of infections and an emerging burden of non-communicable diseases (NCDs) that differs from high-income countries. Currently, only a few ‘omics research cohorts have been studied in African populations, and even fewer relating to kidney disease despite the high prevalence and increasing risk for incident disease with infections and other emerging NCDs. The overall aim of this doctoral thesis was to assess genetic risk models for kidney disease, and to evaluate the urinary proteomic profile of individuals with combined hypertension-associated albuminuria and those who have neither condition in a South African rural setting. The specific study objectives included (1) conducting a scoping review on the use of ‘omics’ for early detection of chronic kidney disease (CKD) and prediction of progression; (2) examining the potential of using summary statistics from three discovery datasets to assess the predictive accuracy of polygenic scores (PGSs) for CKD and kidney function markers in an African setting (using ARK as the target cohort); and (3) comparing urinary proteomic profiles of cases with hypertension and albuminuria to controls to identify potential proteins and pathways involved in hypertension-associated albuminuria.

For the first objective, a scoping review was undertaken to evaluate existing literature for potential biomarkers for CKD and to identify gaps in literature. As there are only a few genomic studies evaluating polygenic risk for CKD in African populations, the second objective was to assess genetic risk models for kidney disease in individuals from a South African rural setting. Because of the high prevalence of hypertension (HT) and that HT is the most common risk factor for CKD in sub-Saharan Africa (SSA), the third objective was to identify potential proteins and pathways involved in hypertension-associated albuminuria. The key findings with regards to these objectives are presented in **Table 5.1**.

Table 5.1: Key findings with regards to the objectives of the thesis

Objective 1: To evaluate existing literature for potential biomarkers that facilitate early detection of CKD and predict its progression, and identify gaps in literature.	• Previous studies, predominantly in African Americans, had highlighted variants in the non-muscle myosin heavy chain 9 (<i>MYH9</i>) and apolipoprotein L1 (<i>APOL1</i>) genes that were significantly associated with increased risk of CKD and non-diabetic end stage kidney disease. • There was good evidence that variants in the <i>uromodulin</i> (<i>UMOD</i>) gene and the UMOD protein were associated with risk of CKD. • The CKD273 proteomic biomarker panel was identified as a tool for early detection of CKD and to predict progression. • Literature gaps identified in the scoping review: ○ While diabetes is a key risk factor for CKD in most regions of the world, other regions such as Africa have other risk factors such as HT and additional contributing factors such as infections. In this regard, there are relatively few ‘omics studies that focus on HT in the context of kidney disease in African individuals residing in Africa. • Only one study on African individuals residing in Africa.
Objective 2: To explore the potential of using summary statistics from three discovery datasets to develop and to test the predictive power of PGSs in an African setting	• PGSs developed from different ancestries showed poor predictability in African individuals in this study, explaining less than 1% of the variability in kidney disease markers (eGFR, UACR, and albuminuria). • A PGS model derived from the transethnic cohort for eGFR explained the highest variability in eGFR (0.8%) in African individuals in this study. • This genetic risk model for eGFR was significantly associated with HT ($P<0.001$), diabetes ($P=0.007$), and HIV ($P=0.001$).
Objective 3: To identify potential proteins and pathways involved in hypertension-associated albuminuria by assessing urinary proteomic profiles in black South African participants with	• Overall, an average of 1,225 and 915 proteins were quantified in the control and case groups, respectively. • Pathways associated with hypertension-associated albuminuria were immune system ($q\text{-value}=1.4\times 10^{-45}$), innate immune system ($q=1.1\times 10^{-32}$), extracellular matrix (ECM) organisation ($q=0.03$), and activation of matrix metalloproteinases ($q=0.04$).

combined hypertension and albuminuria compared to those who have neither condition.	• Proteins including angiotensinogen (AGT), albumin (ALB), APOL1, and UMOD had the highest disease scores (76–100% confidence) for both HT and CKD • A machine learning algorithm was able to identify a set of 20 proteins that differentiated between cases and controls. The top protein in the algorithm was alpha-1-antitrypsin
---	---

CKD, chronic kidney disease; HT, hypertension.

The scoping review (CHAPTER 2) identified research gaps after reviewing the existing literature (Govender et al., 2021). There was only one genomics study on individuals from continental Africa, indicating little research in continental African populations despite a well-established genetic link to kidney disease in individuals with African ancestry (Apolipoprotein L1 [*APOL1*]-associated nephropathy). Therefore, in CHAPTER 3, we assessed genetic risk models for kidney disease in a South African rural setting. The review identified only a few ‘omics studies that focussed on hypertension-associated kidney disease, and none in continental Africans. Since HT is most common risk factor in regions such as Western and SSA, we assessed proteomic profiles in individuals with hypertension-associated albuminuria in CHAPTER 4.

This doctoral thesis is a sub-study of the South African component of the African Research Kidney Disease (ARK) study, a population-based cohort of 2021 adults aged 20–80 years of self-identified black ethnicity from Agincourt in the rural Mpumalanga Province, South Africa. Genotypic and phenotypic risk factors for CKD are well-characterised in the Agincourt Health and Socio-Demographic Surveillance System site, making it perfectly positioned to assess CKD risk in black African individuals. Blood and urine samples from ARK were used for the genomic and proteomic analyses, respectively.

In CHAPTER 3, PGSs were computed for estimated glomerular filtration rate (eGFR), urine albumin-to-creatinine ratio (UACR) and albuminuria, assessing the predictive accuracy of CKD risk based on current algorithms, using GWAS summary statistics (*P* values and weighted effect sizes) from three other external studies. Limited transferability was observed, explaining less than 1% of the variability in kidney disease markers in this cohort. A PGS model derived from the transethnic cohort for eGFR explained the highest variability in eGFR (0.8%) in African individuals in this study, and was significantly associated with HT, diabetes, and HIV.

In CHAPTER 4, the urinary proteomic profiles between cases with HT and albuminuria and controls who had neither condition were assessed by ultra-high-performance liquid chromatography coupled with mass spectrometry. We identified the immune system, innate immune system, extracellular matrix (ECM) organisation, and activation of matrix metalloproteinases pathways associated with HT and albuminuria. Proteins such as

angiotensinogen, albumin, APOL1, and UMOD had the highest disease scores for those individuals with combined HT and albuminuria. A machine learning approach using the OmicLearn platform was able to identify a set of 20 proteins to classify disease status.

5.2. Implications from the genomics study

The assessment of PGSs for kidney function markers, as investigated in this study (CHAPTER 3), contribute to the understanding of CKD risk prediction in African populations, specifically in continental Africa. While we were unable to demonstrate predictability of kidney function traits using summary statistics derived from continental Africans, likely due to the small discovery sample size, we showed that a PGS derived from a large transethnic discovery cohort (~1.2 million individuals) improved predictability power. This was confirmed by a recent study which showed that a PGS for eGFR derived from a large transethnic population demonstrated better predictive power in African Americans and Africans residing in Africa (Hughes et al., 2023). This has implications for researches actively seeking to determine the genetic risk of CKD in Africa – that large cohorts are required to accurately capture the complexity of the genetic architecture of CKD traits in African populations. Such large-scale studies have the capacity to identify novel genes and variants not found in studies conducted largely on European participants and provide an opportunity for developing regional PGSs within Africa to address high genetic diversity within different regions in Africa. Our findings also emphasise the need for further considerations before PGSs can be integrated into clinical practice, including validation in African populations, careful tailoring and possibly integrating traditional risk factors such as age, sex, and comorbidities in the overall predictive score.

The associations between the PGS for eGFR and presence of CKD risk factors (HT, diabetes, and HIV) also have important implications. Because in the absence of a cure for CKD, management of risk factors are vital to prevent or delay progression. Treatment with sodium-glucose cotransporter 2 inhibitors (SGLT-2) and non-steroidal mineralocorticoid receptor antagonists have been shown to improve kidney function and cardiovascular outcomes, indicating that blood pressure lowering may play a key role in mediating the cardiorenal protection (Georgianos and Agarwal, 2023). Furthermore, the use of SGLT-2

inhibitors have also been shown to improve kidney function in individuals with type 2 diabetes (Rossing et al., 2022).

5.3. Implications from the proteomics study

Apart from our study, there is only one other South African study (from the North-West Province) that used urinary proteomics to assess kidney function in individuals resident in Africa (Degenaar et al., 2023). This study contributes new knowledge to the understanding of proteins and pathways involved in hypertension-associated albuminuria in African populations and addresses the knowledge gap identified by the published review in CHAPTER 2, that there are relatively few ‘omics studies focusing on HT in the context of kidney disease in African populations residing in Africa. We demonstrate that while a large-scale, hypothesis-free approach as conducted in this thesis is useful for identifying relevant protein markers in certain populations or disease states, it is unable to evaluate kidney disease risk. This has implications for researchers investigating the pathogenesis of hypertension-associated kidney disease. The proteins identified in this study can be used to develop a hypothesis-driven targeted proteomics approach to further characterise disease pathways that may lead to the development of novel therapeutic approaches and diagnostic biomarkers.

The study also has implications for public health policy. We have reported that there is an existing burden of infections (including HIV) and NCDs (including HT and diabetes) in Agincourt, South Africa. As resources in SSA are limited for treating individuals with CKD who progress to kidney failure and health care facilities are unable match the need for renal replacement therapy, public health policy should focus on primary prevention strategies. With the lack of infrastructure and lack of validated cost-effective screening assays for CKD in SSA, focus should be on educating health care workers to routinely check blood pressure and urine dipsticks, which are cost-effective ways of identifying at risk individuals. Following this, emphasis should be placed on managing the underlying conditions to prevent progression to end stage kidney disease.

5.4. Genetic and proteomic associations with CKD risk, and use of multi-omics approaches

Multiple genetic and environmental factors play a role in CKD, making it difficult to identify the underlying pathophysiologic mechanisms (Köttgen et al., 2022). Although not examined in this study, other studies have assessed associations between polygenic scores for kidney function and the proteome. For example, a PGS derived using GWAS summary statistics from the CKDGen Consortium (N=558,423) and UK Biobank (N=289,432) was assessed in 8,886 participants of European ancestry in the Atherosclerosis Risk in Communities study (Yu et al., 2021). This PGS explained 7% of the variance in eGFR, and was significantly associated (Adjusted $P < 1.02 \times 10^{-5}$) with 108 plasma proteins. Proteins including cystatin-C (a marker of kidney function), collagen alpha-1(XV) chain, and desmocollin-2 had the strongest associations with the PGS (Yu et al., 2021).

In another study, a PGS for CKD derived from GWAS summary statistics in participants of European ancestry was assessed in 3,087 healthy European individuals (Ritchie et al., 2021). Eight proteins were significantly associated with the PGS for CKD. The PGS-protein associations consisted of proteins such as cystatin-C and beta2-microglobulin, which were previously identified as biomarkers for CKD (Ritchie et al., 2021, Jovanović et al., 2003). Explanations for these PGS-protein associations could be that the protein has a causal role in disease, the protein itself is not causal but the levels of the protein changes in response to disease processes, or protein levels are associated with a confounding factor (Ritchie et al., 2021).

A recent large meta-analysis of GWASs of eGFR was conducted in 145,732 individuals of predominantly African ancestry (~72%) (Hughes et al., 2024). PGSs were derived from four base GWASs in continental Africans and African Americans (sample size ranging from 4,246–8,270) and applied to the target population consisting of only African ancestry individuals (Hughes et al., 2024). The PGSs explained 0.71–8.6% of the eGFR variance in the target population (Hughes et al., 2024). Furthermore, epigenomic and transcriptomic analyses were integrated with fine-mapped loci to identify several kidney pathways, including renal system development, regulation of metanephric glomerular mesangial cell

proliferation, and kidney development (Hughes et al., 2024). In another study on 16,474 African American participants, a genome-wide eGFR PGS explained 0.24% of the variance (Vy et al., 2023). In a separate CKD cohort comprising 6,118 individuals with high-risk *APOL1* genotypes, the PGS was significantly associated with CKD (OR=1.44; 95% CI: 1.16–1.80; $P=8.8 \times 10^{-4}$) (Vy et al., 2023).

Multi-omics techniques are now being utilised to enhance our understanding on the pathophysiology of CKD and account for some of the “missing heritability” (Hill et al., 2022). However, missing data and correct and timely interpretation through integration with clinical data are the main challenges in multi-omics integration due to the all biomolecules not being measured in all samples (Flores et al., 2023). Missing data is a common challenge in multi-omics approaches and can arise due to budget restrictions, poor tissue quality, instrument sensitivity, or inadequate sample volume (Flores et al., 2023). Recently, advancements in artificial intelligence and machine learning have significantly facilitated the analyses of multi-omics data. A combination of ‘omics with artificial intelligence have been applied to assess kidney transplant rejection and sub-types of renal cell carcinoma (Reeve et al., 2017, Kang et al., 2017, Clark et al., 2019, Wei et al., 2021). Another study used a proteomics-based machine learning approach to separate healthy participants from patients with CKD with high confidence (97.8%) (Glazyrin et al., 2020, Grobe et al., 2023).

Over the past decade, technological advancements in next-generation DNA sequencing, SNP genotyping, and mass spectrometry-based proteomics approaches have significantly decreased in cost, making large scale ‘omics studies available for the development of powerful predictive algorithms (Grobe et al., 2023, Pinu et al., 2019). With large-scale ‘omics data generation becoming increasingly accessible and affordable, there has been an increase in multi-omics research (Pinu et al., 2019), although this is not yet common in Africa.

5.5. Strengths and limitations

The advent of ‘omics technologies and bioinformatics approaches has enabled the generation and analysis of large-scale data sets, which has led into insights into the

mechanistic pathways of kidney disease. For example, a study leveraged multi-omics data to identify cell types, gene targets and mechanisms for kidney disease (Liu et al., 2022). They identified >500 target genes associated with ~800 GWAS loci for eGFR by employing a single-cell chromatin co-accessibility analysis. They used single-nucleus transposase-accessible chromatin with sequencing data and identified that most GWAS variants for eGFR were located in proximal tubular-specific regions. GWAS analysis identified the solute carrier family 47, member 1 (*SLC47A1*) gene as a kidney disease risk gene. Using single-cell RNA-sequencing data, *SLC47A1* expression was restricted to proximal tubules in the kidney. The causal role of *SLC47A1* in kidney disease was further characterised by phenotypically normal knockout (KO) *SLC47A1* mice, highlighting potential new target for treatment of kidney disease (Liu et al., 2022).

Insights into the potential mechanisms of *APOL1*-mediated kidney disease have enabled exploration of therapies for *APOL1*-associated kidney disease. For example, inaxaplin, a small molecule inhibitor of the *APOL1* channel, has been shown to reduce proteinuria in patients with *APOL1*-associated focal segmental glomerulosclerosis in a phase 2 study (Egbuna et al., 2023). Participants demonstrated a mean decrease of ~48% (95% CI: 60–31) in protein-to-creatinine ratio at week 13 following administration of the drug (Egbuna et al., 2023). Antisense oligonucleotide (ASO) analogues have been shown to alter the expression of particular RNAs and can modify protein synthesis (Vasquez-Rios et al., 2023). The safety and pharmacokinetics of an *APOL1* ASO (ION532) is currently being assessed in a first-in-human, single ascending dose, phase 1 study (NCT04269031) (Vasquez-Rios et al., 2023).

While PGSs are an exciting novel approach that offers insights into the genetic architecture of complex diseases and correlates well at a population level, it has limited potential in informing individual risk (Jefferis et al., 2023). CKD is a complex disease, and although a PGS may be used to detect high-risk genes associated with common pathways in CKD such as fibrosis, it does not take into account gene-gene interactions or environmental and lifestyle factors (Jefferis et al., 2023). Studies have shown that the diagnostic and prognostic performance of PGSs alone are consistently low (Koch et al., 2023). A major

barrier preventing the translation of PGSs into clinical practice is the difficult interpretation of an actual PGS value (Koch et al., 2023). For example, a high PGS value does not mean a high absolute disease risk and vice versa (Jefferis et al., 2023, Koch et al., 2023). Currently, there is limited evidence on the use of PGSs to guide medical management (Abu-El-Haija et al., 2023). The PRSice-2 software, which utilises the standard “C+T” approach, was used for this doctoral work (Choi and O'Reilly, 2019). A key limitation of this approach is overfitting of the PGS analysis. As the “C+T” approach is based on testing many different *P* value thresholds, there is a concern that using the most predictive PGS overfits the target data, thereby generating inflated results and false conclusions (Choi et al., 2020).

The CKD273 proteomics biomarker panel has been validated as a tool for early detection of CKD and to predict progression (Pontillo and Mischak, 2017b). The CKD273 classifier consists of 273 peptides (including collagen fragments, uromodulin [UMOD], α -1-antitrypsin, transthyretin and β -2-macroglobulin) that are combined to create a single score (Good et al., 2010). A CKD273 risk score above the cut-off of 0.154 is used as an indicator for the early detection of CKD (Lindhardt et al., 2017). The CKD273 proteomics biomarker panel is obtainable as a commercial in vitro diagnostic test and received a letter of support from the US-based Food and Drug Administration (Pontillo and Mischak, 2017b). However, the high cost with the necessary specialised equipment, lack of validity in continental Africans, and limited infrastructure prevents its implementation in clinical practice in SSA.

A key theme that emerged during the writing of this thesis was lack of ‘omics studies to compare findings with in continental African populations. For example, there have only been three other GWAS for kidney disease/kidney function traits in continental African populations (Mayanja et al., 2023, Fatumo et al., 2021, Brandenburg et al., 2024). The relatively modest sample sizes of these three GWAS studies ($n=3,288-8,970$) may have impacted the statistical power to detect small-effect associations (Brandenburg et al., 2024, Fatumo et al., 2021, Mayanja et al., 2023), and may be the reason for the low phenotypic variance observed in the PGS analyses (Fatumo et al., 2021, Brandenburg et al., 2024). In addition, there is one study from South Africa that has investigated the urinary proteome in

young adults with hypertension and no studies investigating the urinary proteomics of hypertensive-associated nephropathy in continental Africa populations. The sample size of the urinary proteomics study in young adults with hypertension was also relatively low (n=157) and not adequate to perform the adjustment for multiple testing, or adjustment for potential confounding factors (e.g., BMI) (De Beer et al., 2023).

African populations are largely understudied mainly due to lack of resources and a dearth of skills, and less importance being placed on precision medicine approaches, especially for NCDs, in Africa. Thus, the lack of infrastructure and equipment to support research and clinical translation is also contribute to the lack of 'omics studies in Africa (Omotoso et al., 2022). Objectives 2 and 3 of this PhD thesis contribute to the generation of genomic and proteomic data in African individuals, residing in Africa.

The following are required to bridge the genomic data gap in Africa; (1) build a larger genomic and 'omic data repositories, (2) develop research capabilities across the continent by improving research facilities and technical expertise through training, and (3) funding and global partnerships (Omotoso et al., 2022). Many programs have been established to enhance scientific research in Africa. These include the H3Africa consortium and The Developing Excellence in Leadership, Training, and Science in Africa (DELTA) program. The H3Africa consortium has established its presence in 30 countries in Africa through a \$176 million grant provided by the U.S. National Institute of Health (NIH) and Wellcome Trust (H3Africa, 2024). DELTA is a \$100 million US initiative to support the development of world class researchers and scientific leaders in Africa (Hamdi et al., 2021). To promote the advancement of proteomics in Africa, several initiatives have emerged. For example, the Centre for Proteomic and Genomic Research (CPGR) is an initiative supported by the South African Department of Science and Innovation Technology Innovation Agency that aims to increase 'omics efforts in South Africa (CPGR, 2024).

African-ancestry (including African Americans and the continental African diaspora) individuals are represented by only ~10,000 participants in research studies, while European ancestry individuals are represented by more than 500,000 participants in the two

international GWAS discovery cohorts. The small sample size of African ancestry cohorts in GWAS limits the detection of moderate to low effect variants and transferability of findings. Biobanks are important for genomics research. Consequently, several national biobanks have been established to assist scientific research and advance precision medicine. One such example is the UKBioBank, a large-scale biomedical database that contains health and genetic information from 500,000 individuals in the UK including ~8,000 with African ancestry (Bycroft et al., 2018). However, such biobanks are lacking in African representation. In order to improve African representation in genomic data, some initiatives have commenced in Africa such as the Uganda Genome Resource and the H3Africa consortium collaborative centres (Fatumo et al., 2022, Adoga et al., 2014). The main aim of the African Population Cohorts Consortium (APCC) initiative is to focus on lack of large population-based data in Africa, by collecting medical, social and environmental information to assess the health and wellness of individual with African-ancestry s through an established network of cohorts across Africa (APCC, 2021).

There are many barriers that limit the progress of ‘omics research in Africa. Lack of infrastructure and a critical mass of skills are major hindrances of research in Africa, with power/electricity and more recently, water supply of the biggest infrastructural challenges (Laurance et al., 2015). In addition, the ability to conduct ‘omics research is restricted by the availability and high cost of the required equipment (Inzaule et al., 2021). Many countries in Africa lack sequencing equipment and capacity. Despite these challenges, many programs have been established to enhance scientific research in Africa. They include the H3Africa consortium, the APCC and The Developing Excellence in Leadership, Training, and Science in Africa (DELTAS).

5.6. Future considerations

There is high genetic diversity and regional variability in allele frequencies and variants across Africa (Choudhury et al., 2022, Pereira et al., 2021, Kamiza et al., 2022). Therefore, future PGS analyses need to include individuals from various regions in Africa to understand CKD risk across different regions of continental Africa. The Agincourt study site in Bushbuckridge in northern South Africa, together with the other ARK study sites

and the AWI-Gen sites in West, East and South Africa can be used to assess CKD risk across various regions in Africa. Previous GWAS for kidney-related traits in African-ancestry populations include predominantly African American participants, therefore, the reported association signals are not simply transferable to African populations, residing in Africa, and could provide inaccurate risk estimations. It is important for future research to include larger sample sizes, use genotyping arrays with common African variants, and larger imputation reference panels from the target populations, to enhance CKD risk prediction in African populations. The proposed African Population Cohorts Consortium (APCC) initiative has the potential to address the gap in large African population and make important contributions to improve health and well-being in Africa (APCC, 2021).

There have been significant advancements in PGS methods to address the transferability gap. These methods are mainly based on ensembling and leveraging multi-ancestry and multi-trait data to detect more likely causal variants (Xiang et al., 2024). Ensembling-based techniques incorporate several sets of GWAS data from multiple diseases/traits or multiple ancestries to generate a better set of variants and weights for the PGS calculation (Xiang et al., 2024). An example of this technique is PRS-CSx, which is an extension of the continuous shrinkage models (Ruan et al., 2022). PGSs based on causal variants may result in improved prediction performance, as these variants are similar across ancestries (Xiang et al., 2024). Given that causal variants may have relevant biological functions, this data can be utilised as biological priors to generate a better set of variants and weights for the PGS analysis (Xiang et al., 2024). A variety of methods have been shown to improve the transferability across ancestries using annotation/biological information. These methods include LDpred-funct (Márquez-Luna et al., 2021), PolyPred (Weissbrod et al., 2022), and BayesRC (MacLeod et al., 2016). Future work could assess the aforementioned methods to improve the prediction across ancestries. Furthermore, a meta-analysis using summary statistics from the trans-ethnic and AWI-Gen eGFR GWASs in CHAPTER 3 could be performed to potentially increase the variance for eGFR.

Gene-protein interactions should also be investigated to provide a comprehensive understanding of CKD pathophysiology. A combination of ‘omics approaches (e.g.

genomics and proteomics) have been previously shown to provide insights into understanding the pathophysiology of CKD (Yu et al., 2021, Ritchie et al., 2021). In addition, future work can assess the value of combining a PGS together with the proteomic profile in risk prediction for CKD.

The integration of multi-omics technologies can be utilised to identify cell types, prioritise disease-risk genes, and understand pathophysiology of disease. For example, a combination of genomics and transcriptomics provides a way characterise transcriptomic changes in specific cell types that are linked to risk variants in specific genes, and could provide insight to the underlying mechanisms of diseases (Liu et al., 2022). Another multi-omics study identified 1,038 genes and 479 GWAS association signals associated with blood pressure through an integrated analysis of genomic, transcriptomic and epigenomic data (n=430) (Zhou et al., 2023). They used kidney-associated SNPs as genetic instruments in Mendelian Randomisation studies to identify casual effects of blood pressure on kidney disease. This provided evidence that genetic effects on blood pressure is mediated by the kidney (Zhou et al., 2023).

Recently, single-cell multi-omics technologies have been introduced to simultaneously assess one or more omics applications in a single cell (Wen et al., 2022). The development of a single cell sequencing approach initially focused on transcriptome profiling, however, there have been a rapid growth in adopting sequencing assays and proteome profiling at the single cell level (Tang et al., 2009, Pregizer et al., 2023). There have been significant developments in single-cell sequencing approaches which have permitted the study of specific cells obtained from kidney biopsies, urine and blood, highlighting cellular signatures of inflammation, cell injury, and fibrosis in numerous kidney diseases (Pregizer et al., 2023). These cellular signatures provide a more comprehensive understanding of disease pathophysiology and may enable development of therapeutic targets (Pregizer et al., 2023).

5.7. Conclusions

The findings from this doctoral thesis contribute to a better understanding of kidney disease risk prediction; and proteins and pathways involved in hypertension-associated albuminuria

in African populations, and may inform the development of personalised treatment strategies. While Africa is advancing in the field of ‘omics, there is still significant progress to be made. The small samples sizes are one of the major limitations of ‘omics research in African populations. However, several initiatives have commenced to increase African representation in ‘omics data and address the gap in large population-based data from the African continent.

REFERENCES

- Abbiss, H., Maker, G. L. & Trengove, R. D. 2019. Metabolomics approaches for the diagnosis and understanding of kidney diseases. *Metabolites*, 9, 34.
- Abd Elhafeez, S., Bolignano, D., D'arrigo, G., Dounousi, E., Tripepi, G. & Zoccali, C. 2018. Prevalence and burden of chronic kidney disease among the general population and high-risk groups in Africa: a systematic review. *BMJ open*, 8, e015069.
- Aboyans, V. & Collaborators, C. O. D. 2015. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet (British edition)*, 385, 117-71.
- Abu-El-Haija, A., Reddi, H. V., Wand, H., Rose, N. C., Mori, M., Qian, E. & Murray, M. F. 2023. The clinical application of polygenic risk scores: A points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 25, 100803.
- Adam, Y., Sadeeq, S., Kumuthini, J., Ajayi, O., Wells, G., Solomon, R., Ogunlana, O., Adetiba, E., Iweala, E., Brors, B. & Adebisi, E. 2022. Polygenic Risk Score in African populations: progress and challenges. *F1000Research*, 11, 175.
- Adeniyi, A. B., Laurence, C. E., Volmink, J. A. & Davids, M. R. 2017. Prevalence of chronic kidney disease and association with cardiovascular risk factors among teachers in Cape Town, South Africa. *Clinical kidney journal*, 10, 363-69.
- Adoga, M. P., Fatumo, S. A. & Agwale, S. M. 2014. H3Africa: a tipping point for a revolution in bioinformatics, genomics and health research in Africa. *Source code for biology and medicine*, 9, 10.
- Agarwal, R. 2004. Chronic kidney disease is associated with oxidative stress independent of hypertension. *Clinical Nephrology*, 61, 377-83.
- Al-Amrani, S., Al-Jabri, Z., Al-Zaabi, A., Alshekaili, J. & Al-Khabori, M. 2021. Proteomics: Concepts and applications in human medicine. *World journal of biological chemistry*, 12, 57-69.
- Alaje Abiodun, K., Benedict, N. & Idogun Sylvester, E. β 2 Microglobulin as A Marker of Early Renal Damage In Patients With Sickle Cell Nephropathy.
- Alexander, D. H., Novembre, J. & Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*, 19, 1655-64.
- Ali, S. A., Soo, C., Agongo, G., Alberts, M., Amenga-Etego, L., Boua, R. P., Choudhury, A., Crowther, N. J., Depuur, C. & Gómez-Olivé, F. X. 2018. Genomic and environmental risk factors for cardiometabolic diseases in Africa: methods used for Phase 1 of the AWI-Gen population cross-sectional study. *Global health action*, 11, 1507133.
- Alicic, R. Z., Rooney, M. T. & Tuttle, K. R. 2017. Diabetic kidney disease: challenges, progress, and possibilities. *Clinical Journal of the American Society of Nephrology*, 12, 2032-45.
- Alkhalaf, A., Zürlbig, P., Bakker, S. J., Bilo, H. J., Cerna, M., Fischer, C., Fuchs, S., Janssen, B., Medek, K. & Mischak, H. 2010. Multicentric validation of proteomic biomarkers in urine specific for diabetic nephropathy. *PloS one*, 5, e13421.
- Andersen, S., Mischak, H., Zürlbig, P., Parving, H.-H. & Rossing, P. 2010. Urinary proteome analysis enables assessment of renoprotective treatment in type 2 diabetic patients with microalbuminuria. *BMC nephrology*, 11, 1-9.
- Antonioli, P., Bachi, A., Fasoli, E. & Righetti, P. G. 2009. Efficient removal of DNA from proteomic samples prior to two-dimensional map analysis. *Journal of chromatography A*, 1216, 3606-12.
- Apcc 2021. African Population Cohorts Consortium (APCC) Scoping Report.
- Argiles, A., Siwy, J., Duranton, F., Gayraud, N., Dakna, M., Lundin, U., Osaba, L., Delles, C., Mourad, G. & Weinberger, K. M. 2013. CKD273, a new proteomics classifier assessing CKD and its prognosis. *Plos one*, 8, e62837.

- Ashuntantang, G., Osafo, C., Olowu, W. A., Arogundade, F., Niang, A., Porter, J., Naicker, S. & Luyckx, V. A. 2017. Outcomes in adults and children with end-stage kidney disease requiring dialysis in sub-Saharan Africa: a systematic review. *The Lancet Global Health*, 5, e408-e17.
- Aslam, B., Basit, M., Nisar, M. A., Khurshid, M. & Rasool, M. H. 2017. Proteomics: Technologies and Their Applications. *Journal of chromatographic science*, 55, 182-96.
- Aung, T. T., Bhandari, S. K., Chen, Q., Malik, F. T., Willey, C. J., Reynolds, K., Jacobsen, S. J. & Sim, J. J. 2021. Autosomal dominant polycystic kidney disease prevalence among a racially diverse United States population, 2002 through 2018. *Kidney360*, 2, 2010.
- Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., Mcvean, G. A. & Abecasis, G. R. 2015. A global reference for human genetic variation. *Nature*, 526, 68-74.
- Baichan, P., Naicker, P., Augustine, T. N., Smith, M., Candy, G., Devar, J. & Nweke, E. E. 2023. Proteomic analysis identifies dysregulated proteins and associated molecular pathways in a cohort of gallbladder cancer patients of African ancestry. *Clinical proteomics*, 20, 8.
- Baichoo, S., Souilmi, Y., Panji, S., Botha, G., Meintjes, A., Hazelhurst, S., Bendou, H., Beste, E., Mpangase, P. T., Souiai, O., Alghali, M., Yi, L., O'connor, B. D., Crusoe, M., Armstrong, D., Aron, S., Joubert, F., Ahmed, A. E., Mbiyavanga, M., Heusden, P. V., Magosi, L. E., Zermeno, J., Mainzer, L. S., Fadlilmola, F. M., Jongeneel, C. V. & Mulder, N. 2018. Developing reproducible bioinformatics analysis workflows for heterogeneous computing environments to support African genomics. *BMC bioinformatics*, 19, 457.
- Bakris, G., Ali, W. & Parati, G. 2019. ACC/AHA versus ESC/ESH on hypertension guidelines: JACC guideline comparison. *Journal of the American College of Cardiology*, 73, 3018-26.
- Bakshi, A., Jefferis, J., Wolfe, R., Wetmore, J. B., Mcneil, J. J., Murray, A. M., Polkinghorne, K. R., Mallett, A. J. & Lacaze, P. 2023. Association of polygenic scores with chronic kidney disease phenotypes in a longitudinal study of older adults. *Kidney International*, 103, 1156-66.
- Barratt, J. & Topham, P. 2007. Urine proteomics: the present and future of measuring urinary protein components in disease. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, 177, 361-8.
- Bazzi, C., Seccia, T. M., Napodano, P., Campi, C., Caroccia, B., Cattarin, L. & Calò, L. A. 2020. High blood pressure is associated with tubulointerstitial damage along with glomerular damage in glomerulonephritis. A large cohort study. *Journal of Clinical Medicine*, 9, 1656.
- Benigni, A., Cassis, P. & Remuzzi, G. 2010. Angiotensin II revisited: new roles in inflammation, immunology and aging. *EMBO molecular medicine*, 2, 247-57.
- Benson, L. N., Guo, Y., Deck, K., Mora, C., Liu, Y. & Mu, S. 2023. The link between immunity and hypertension in the kidney and heart. *Frontiers in Cardiovascular Medicine*, 10.
- Bidani, A. K. & Griffin, K. A. 2004. Pathophysiology of hypertensive renal damage: implications for therapy. *Hypertension*, 44, 595-601.
- Bienaimé, F., Muorah, M., Metzger, M., Broeuilh, M., Houiller, P., Flamant, M., Haymann, J.-P., Vonderscher, J., Mizrahi, J. & Friedlander, G. 2023. Combining robust urine biomarkers to assess chronic kidney disease progression. *Ebiomedicine*, 93.
- Bikbov, B., Purcell, C. A., Levey, A. S., Smith, M., Abdoli, A., Abebe, M., Adebayo, O. M., Afarideh, M., Agarwal, S. K. & Agudelo-Botero, M. 2020. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The lancet*, 395, 709-33.
- Boschieter, J., Bach, S., Bijnsdorp, I. V., Segerink, L. I., Rurup, W. F., Van Splunter, A. P., Bahce, I., Novianti, P. W., Kazemier, G., Van Moorselaar, R. J. A., Steenbergen, R. D. M. & Nieuwenhuijzen, J. A. 2018. A protocol for urine collection and storage prior to DNA methylation analysis. *PLoS One*, 13, e0200906.

- Brandenburg, J.-T., Govender, M. A., Winkler, C. A., Boua, P. R., Agongo, G., Fabian, J. & Ramsay, M. 2022a. Apolipoprotein L1 High-Risk Genotypes and Albuminuria in Sub-Saharan African Populations. *Clinical Journal of the American Society of Nephrology*, 17, 798-808.
- Brandenburg, J. T., Chen, W. C., Boua, P. R., Govender, M. A., Agongo, G., Micklesfield, L. K., Sorgho, H., Tollman, S., Asiki, G., Mashinya, F., Hazelhurst, S., Morris, A. P., Fabian, J. & Ramsay, M. 2024. Genetic Association and Transferability for Urinary Albumin-Creatinine Ratio as a Marker of Kidney Function in four Sub-Saharan African Populations and non-continental Individuals of African Ancestry. *medRxiv : the preprint server for health sciences*.
- Brandenburg, J. T., Clark, L., Botha, G., Panji, S., Baichoo, S., Fields, C. & Hazelhurst, S. 2022b. H3AGWAS: a portable workflow for genome wide association studies. *BMC bioinformatics*, 23, 498.
- Brun, R., Blum, J., Chappuis, F. & Burri, C. 2010. Human african trypanosomiasis. *The Lancet*, 375, 148-59.
- Buffet, L. & Ricchetti, C. 2012. Chronic kidney disease and hypertension: A destructive combination.
- Bullich, G., Domingo-Gallego, A., Vargas, I., Ruiz, P., Lorente-Grandoso, L., Furlano, M., Fraga, G., Madrid, Á., Ariceta, G. & Borregán, M. 2018. A kidney-disease gene panel allows a comprehensive genetic diagnosis of cystic and glomerular inherited kidney diseases. *Kidney international*, 94, 363-71.
- Bulto, L. N. B., Defo, B. K., Islam, S. M. S. & Collaborators, G. C. O. D. 2017. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, 390, 1151-210.
- Bush, W. S. & Moore, J. H. 2012. Chapter 11: Genome-wide association studies. *PLoS computational biology*, 8, e1002822.
- Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., Mcvean, G., Leslie, S., Allen, N., Donnelly, P. & Marchini, J. 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562, 203-09.
- Califf, R. M. 2018. Biomarker definitions and their applications. *Experimental biology and medicine (Maywood, N.J.)*, 243, 213-21.
- Cañadas-Garre, M., Anderson, K., Cappa, R., Skelly, R., Smyth, L. J., Mcknight, A. J. & Maxwell, A. P. 2019. Genetic susceptibility to chronic kidney disease—some more pieces for the heritability puzzle. *Frontiers in Genetics*, 10, 453.
- Casanova, F., Tyrrell, J., Beaumont, R. N., Ji, Y., Jones, S. E., Hattersley, A. T., Weedon, M. N., Murray, A., Shore, A. C., Frayling, T. M. & Wood, A. R. 2019. A genome-wide association study implicates multiple mechanisms influencing raised urinary albumin-creatinine ratio. *Human molecular genetics*, 28, 4197-207.
- Cavazos, T. B. & Witte, J. S. 2021. Inclusion of variants discovered from diverse populations improves polygenic risk score transferability. *HGG advances*, 2.
- Chandramouli, K. & Qian, P. Y. 2009. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Human genomics and proteomics : HGP*, 2009.
- Chatterjee, N., Wheeler, B., Sampson, J., Hartge, P., Chanock, S. J. & Park, J. H. 2013. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nature genetics*, 45, 400-5, 05e1-3.
- Chen, G., Adeyemo, A. A., Zhou, J., Chen, Y., Doumatey, A., Lashley, K., Huang, H., Amoah, A., Agyenim-Boateng, K. & Eghan Jr, B. A. 2007. A genome-wide search for linkage to renal function phenotypes in West Africans with type 2 diabetes. *American journal of kidney diseases*, 49, 394-400.

- Cheng, S., Pollock, A. S., Mahimkar, R., Olson, J. L., Lovett, D. H., Cheng, S., Pollock, A. S., Mahimkar, R., Olson, J. L. & Lovett, D. H. 2006. Matrix metalloproteinase 2 and basement membrane integrity: a unifying mechanism for progressive renal injury. *The FASEB journal*, 20, 1898-900.
- Cheung, A. K., Chang, T. I., Cushman, W. C., Furth, S. L., Hou, F. F., Ix, J. H., Knoll, G. A., Muntner, P., Pecoits-Filho, R. & Sarnak, M. J. 2021. KDIGO 2021 clinical practice guideline for the management of blood pressure in chronic kidney disease. *Kidney International*, 99, S1-S87.
- Cho, M. H. 2010. Renal fibrosis. *Korean journal of pediatrics*, 53, 735.
- Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo Jr, J. L., Jones, D. W., Materson, B. J., Oparil, S. & Wright Jr, J. T. 2003. Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. *hypertension*, 42, 1206-52.
- Choi, S. W., Mak, T. S. & O'reilly, P. F. 2020. Tutorial: a guide to performing polygenic risk score analyses. *Nature protocols*, 15, 2759-72.
- Choi, S. W. & O'reilly, P. F. 2019. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*, 8, giz082.
- Choudhury, A., Aron, S., Botigué, L. R., Sengupta, D., Botha, G., Bensellak, T., Wells, G., Kumuthini, J., Shriner, D., Fakim, Y. J., Ghoorah, A. W., Dareng, E., Odia, T., Falola, O., Adebisi, E., Hazelhurst, S., Mazandu, G., Nyangiri, O. A., Mbiyavanga, M., Benkahla, A., Kassim, S. K., Mulder, N., Adebamowo, S. N., Chimusa, E. R., Muzny, D., Metcalf, G., Gibbs, R. A., Rotimi, C., Ramsay, M., Adeyemo, A. A., Lombard, Z. & Hanchard, N. A. 2020. High-depth African genomes inform human migration and health. *Nature*, 586, 741-48.
- Choudhury, A., Brandenburg, J. T., Chikowore, T., Sengupta, D., Boua, P. R., Crowther, N. J., Agongo, G., Asiki, G., Gómez-Olivé, F. X., Kisiangani, I., Maimela, E., Masemola-Maphutha, M., Micklesfield, L. K., Nonterah, E. A., Norris, S. A., Sorgho, H., Tinto, H., Tollman, S., Graham, S. E., Willer, C. J., Hazelhurst, S. & Ramsay, M. 2022. Meta-analysis of sub-Saharan African studies provides insights into genetic architecture of lipid traits. *Nature communications*, 13, 2578.
- Cisek, K., Krochmal, M., Klein, J. & Mischak, H. 2016. The application of multi-omics and systems biology to identify therapeutic targets in chronic kidney disease. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*, 31, 2003-11.
- Clark, D. J., Dhanasekaran, S. M., Petralia, F., Pan, J., Song, X., Hu, Y., Da Veiga Leprevost, F., Reva, B., Lih, T. M., Chang, H. Y., Ma, W., Huang, C., Ricketts, C. J., Chen, L., Krek, A., Li, Y., Rykunov, D., Li, Q. K., Chen, L. S., Ozbek, U., Vasaikar, S., Wu, Y., Yoo, S., Chowdhury, S., Wyczalkowski, M. A., Ji, J., Schnaubelt, M., Kong, A., Sethuraman, S., Avtonomov, D. M., Ao, M., Colaprico, A., Cao, S., Cho, K. C., Kalayci, S., Ma, S., Liu, W., Ruggles, K., Calinawan, A., Gümüş, Z. H., Geiszler, D., Kawaler, E., Teo, G. C., Wen, B., Zhang, Y., Keegan, S., Li, K., Chen, F., Edwards, N., Pierorazio, P. M., Chen, X. S., Pavlovich, C. P., Hakimi, A. A., Brominski, G., Hsieh, J. J., Antczak, A., Omelchenko, T., Lubinski, J., Wiznerowicz, M., Linehan, W. M., Kinsinger, C. R., Thiagarajan, M., Boja, E. S., Mesri, M., Hiltke, T., Robles, A. I., Rodriguez, H., Qian, J., Fenyö, D., Zhang, B., Ding, L., Schadt, E., Chinnaiyan, A. M., Zhang, Z., Omenn, G. S., Cieslik, M., Chan, D. W., Nesvizhskii, A. I., Wang, P. & Zhang, H. 2019. Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell*, 179, 964-83.e31.
- Clarke, W., Molinaro, R., Bachmann, L., Botelho, J., Cao, Z., French, S., Garg, S., Gawoski, J., Grant, R. & Ghoofignagle, A. 2014. Liquid chromatography-mass spectrometry methods; approved guideline. *C62-A*, 34, 1-62.

- Collister, J. A., Liu, X. & Clifton, L. 2022. Calculating Polygenic Risk Scores (PRS) in UK Biobank: A Practical Guide for Epidemiologists. *Front Genet*, 13, 818574.
- Connaughton, D. M., Kennedy, C., Shril, S., Mann, N., Murray, S. L., Williams, P. A., Conlon, E., Nakayama, M., Van Der Ven, A. T. & Ityel, H. 2019. Monogenic causes of chronic kidney disease in adults. *Kidney international*, 95, 914-28.
- Corbett, J. R., Robinson, D. E. & Patrie, S. M. 2020. Robustness and ruggedness of isoelectric focusing and superficially porous liquid chromatography with Fourier transform mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 32, 346-54.
- Couser, W. G., Remuzzi, G., Mendis, S. & Tonelli, M. 2011. The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. *Kidney international*, 80, 1258-70.
- Cpgr. 2024. Centre for Proteomics and Genomics Research. Available at: <https://www.tia.org.za/blog/2018/01/30/centre-for-proteomics-and-genomics-research-cpgr/>. Accessed January 2024. [Online]. [Accessed].
- Craddock, N. J. & Jones, I. R. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447, 661-78.
- Craik, A., Gondwe, M., Mayindi, N., Chipungu, S., Khoza, B., Gómez-Olivé, X., Tollman, S., Frean, J., Tomlinson, L. A. & Fabian, J. 2023. Forgotten but not gone in rural South Africa: Urinary schistosomiasis and implications for chronic kidney disease screening in endemic countries. *Wellcome Open Res*, 8, 68.
- Crews, D. C., Bello, A. K. & Saadi, G. 2019. Burden, access, and disparities in kidney disease. *Brazilian Journal of Medical and Biological Research*, 52.
- Critselis, E., Vlahou, A., Stel, V. S. & Morton, R. L. 2018. Cost-effectiveness of screening type 2 diabetes patients for chronic kidney disease progression with the CKD273 urinary peptide classifier as compared to urinary albumin excretion. *Nephrology Dialysis Transplantation*, 33, 441-49.
- Cui, M., Cheng, C. & Zhang, L. 2022. High-throughput proteomics: a methodological mini-review. *Laboratory Investigation*, 102, 1170-81.
- Cupp-Sutton, K. A. & Wu, S. 2020. High-throughput quantitative top-down proteomics. *Molecular omics*, 16, 91-99.
- Daneshpajouhnejad, P., Kopp, J. B., Winkler, C. A. & Rosenberg, A. Z. 2022. The evolving story of apolipoprotein L1 nephropathy: the end of the beginning. *Nature Reviews Nephrology*, 18, 307-20.
- Das, S., Abecasis, G. R. & Browning, B. L. 2018. Genotype Imputation from Large Reference Panels. *Annual review of genomics and human genetics*, 19, 73-96.
- Daugherty, M., Polanuyer, B., Farrell, M., Scholle, M., Lykidis, A., De Crécy-Lagard, V. & Osterman, A. 2002. Complete reconstitution of the human coenzyme A biosynthetic pathway via comparative genomics. *Journal of Biological Chemistry*, 277, 21431-39.
- Davids, M. R., Marais, N., Sebastian, S., Jardine, T. & Jacobs, J. C. 2023. South African Renal Registry Annual Report 2021. *African Journal of Nephrology*, 26, 83-94.
- De Beer, D., Mels, C. M., Schutte, A. E., Delles, C., Mary, S., Mullen, W., Latosinska, A., Mischak, H. & Kruger, R. 2023. Identifying a urinary peptidomics profile for hypertension in young adults: The African-PREDICT study: Urinary peptidomics and hypertension. *Proteomics*, 23, 2200444.
- De Bhailis, Á. M. & Kalra, P. A. 2022. Hypertension and the kidneys. *British Journal of Hospital Medicine*, 83, 1-11.
- De Oliveira, B. G. R. B. 2018. Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist.

- Degenaar, A., Jacobs, A., Kruger, R., Delles, C., Mischak, H. & Mels, C. 2023. Cardiovascular risk and kidney function profiling using conventional and novel biomarkers in young adults: the African-PREDICT study. *BMC nephrology*, 24, 1-9.
- Delanghe, J. & Speeckaert, M. 2014. Preanalytical requirements of urinalysis. *Biochemia medica*, 24, 89-104.
- Demirci, B., Demir, O., Dost, T. & Birincioglu, M. 2014. Protective effect of vitamin B5 (dexpantenol) on cardiovascular damage induced by streptozocin in rats. *Bratisl Lek Listy*, 115, 190-96.
- Devuyst, O., Olinger, E. & Rampoldi, L. 2017. Uromodulin: from physiology to rare and complex kidney disorders. *Nature Reviews Nephrology*, 13, 525-44.
- Dhaun, N., Webb, D. J. & Kluth, D. C. 2012. Endothelin-1 and the kidney—beyond BP. *British journal of pharmacology*, 167, 720-31.
- Domingo-Gallego, A., Pybus, M., Bullich, G., Furlano, M., Ejarque-Vila, L., Lorente-Grandoso, L., Ruiz, P., Fraga, G., López González, M. & Piñero-Fernández, J. A. 2022. Clinical utility of genetic testing in early-onset kidney disease: seven genes are the main players. *Nephrology Dialysis Transplantation*, 37, 687-96.
- Drożdż, D., Drożdż, M. & Wójcik, M. 2022. Endothelial dysfunction as a factor leading to arterial hypertension. *Pediatric Nephrology*, 1-13.
- Dubin, R. F., Deo, R., Ren, Y., Wang, J., Zheng, Z., Shou, H., Go, A. S., Parsa, A., Lash, J. P. & Rahman, M. 2023. Proteomics of CKD progression in the chronic renal insufficiency cohort. *Nature communications*, 14, 6340.
- Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., Peterson, R. & Domingue, B. 2019. Analysis of polygenic risk score usage and performance in diverse human populations. *Nature communications*, 10, 3328.
- Dupree, E. J., Jayathirtha, M., Yorkey, H., Mihasan, M., Petre, B. A. & Darie, C. C. 2020. A critical review of bottom-up proteomics: the good, the bad, and the future of this field. *Proteomes*, 8, 14.
- Durbin, R. 2014. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics (Oxford, England)*, 30, 1266-72.
- Egan, B., Wozniak, G., Yang, J., Davis, R., Sutherland, S. & Rakotz, M. 2018. USING THE NEW DEFINITION OF HYPERTENSION FROM THE 2017 US GUIDELINE FOR HIGH BLOOD PRESSURE (BP) IN ADULTS: CHANGES IN CONTROL OF HYPERTENSION IN A SAMPLE OF OUTPATIENT CLINICS. *Journal of hypertension*, 36, e203.
- Egan, B. M., Kjeldsen, S. E., Grassi, G., Esler, M. & Mancia, G. 2019. The global burden of hypertension exceeds 1.4 billion people: should a systolic blood pressure target below 130 become the universal standard? *Journal of hypertension*, 37, 1148-53.
- Egbuna, O., Zimmerman, B., Manos, G., Fortier, A., Chirieac, M. C., Dakin, L. A., Friedman, D. J., Bramham, K., Campbell, K., Knebelmann, B., Barisoni, L., Falk, R. J., Gipson, D. S., Lipkowitz, M. S., Ojo, A., Bunnage, M. E., Pollak, M. R., Altshuler, D. & Chertow, G. M. 2023. Inaxaplin for Proteinuric Kidney Disease in Persons with Two APOL1 Variants. *The New England journal of medicine*, 388, 969-79.
- Ekrikpo, U. E., Kengne, A. P., Bello, A. K., Effa, E. E., Noubiap, J. J., Salako, B. L., Rayner, B. L., Remuzzi, G. & Okpechi, I. G. 2018. Chronic kidney disease in the global adult HIV-infected population: a systematic review and meta-analysis. *PloS one*, 13, e0195443.
- Elmarakby, A. A., Loomis, E. D., Pollock, J. S. & Pollock, D. M. 2005. NADPH oxidase inhibition attenuates oxidative stress but not hypertension produced by chronic ET-1. *Hypertension*, 45, 283-87.

- Esmail, A. M., Bates, W. D., Amirali, M. H., Jardine, T. & Davids, M. R. 2023. Patterns of biopsy-proven kidney disease amongst South African adults from 1995 to 2017. *African Journal of Nephrology*, 26, 9-16.
- Euesden, J., Lewis, C. M. & O'reilly, P. F. 2015. PRSice: Polygenic Risk Score software. *Bioinformatics (Oxford, England)*, 31, 1466-8.
- Fabian, J., Gondwe, M., Mayindi, N., Chipungu, S., Khoza, B., Gaylard, P., Wade, A. N., Gómez-Olivé, F. X., Tomlinson, L. A. & Ramsay, M. 2022. Chronic kidney disease (CKD) and associated risk in rural South Africa: a population-based cohort study. *Wellcome open research*, 7.
- Fan, G., Gong, T., Lin, Y., Wang, J., Sun, L., Wei, H., Yang, X., Liu, Z., Li, X. & Zhao, L. 2021. Urine proteomics identifies biomarkers for diabetic kidney disease at different stages. *Clinical Proteomics*, 18, 1-12.
- Fan, Y., Yi, Z., D'agati, V. D., Sun, Z., Zhong, F., Zhang, W., Wen, J., Zhou, T., Li, Z. & He, L. 2020. Erratum. Comparison of Kidney Transcriptomic Profiles of Early and Advanced Diabetic Nephropathy Reveals Potential New Mechanisms for Disease Progression. *Diabetes* 2019; 68: 2301–2314. *Diabetes*, 69, 797-97.
- Fatumo, S., Chikowore, T., Kalyesubula, R., Nsubuga, R. N., Asiki, G., Nashiru, O., Seeley, J., Crampin, A. C., Nitsch, D., Smeeth, L., Kaleebu, P., Burgess, S., Nyirenda, M., Franceschini, N., Morris, A. P., Tomlinson, L. & Newton, R. 2021. Discovery and fine-mapping of kidney function loci in first genome-wide association study in Africans. *Human molecular genetics*, 30, 1559-68.
- Fatumo, S., Mugisha, J., Soremekun, O. S., Kalungi, A., Mayanja, R., Kintu, C., Makanga, R., Kakande, A., Abaasa, A., Asiki, G., Kalyesubula, R., Newton, R., Nyirenda, M., Sandhu, M. S. & Kaleebu, P. 2022. Uganda Genome Resource: A rich research database for genomic studies of communicable and non-communicable diseases in Africa. *Cell genomics*, 2, None.
- Fatumo, S., Sathan, D., Samtal, C., Isewon, I., Tamuhla, T., Soremekun, C., Jafali, J., Panji, S., Tiffin, N. & Fakim, Y. J. 2023. Polygenic risk scores for disease risk prediction in Africa: current challenges and future directions. *Genome medicine*, 15, 87.
- Ferdinand, K. C. 2020. Uncontrolled hypertension in sub-Saharan Africa: Now is the time to address a looming crisis. *Journal of clinical hypertension (Greenwich, Conn.)*, 22, 2111-13.
- Flores, J. E., Claborne, D. M., Weller, Z. D., Webb-Robertson, B. M., Waters, K. M. & Bramer, L. M. 2023. Missing data in multi-omics integration: Recent advances through artificial intelligence. *Frontiers in artificial intelligence*, 6, 1098308.
- Folkow, B., Göthberg, G., Lundin, S. & Ricksten, S. E. 1977. Structural “resetting” of the renal vascular bed in spontaneously hypertensive rats (SHR). *Acta physiologica Scandinavica*, 100, 270-72.
- Foreman, K. J., Marquez, N., Dolgert, A., Fukutaki, K., Fullman, N., Mcgaughey, M., Pletcher, M. A., Smith, A. E., Tang, K. & Yuan, C.-W. 2018. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories. *The Lancet*, 392, 2052-90.
- Freedman, B. I., Julian, B. A., Pastan, S. O., Israni, A. K., Schladt, D., Gautreaux, M. D., Hauptfeld, V., Bray, R. A., Gebel, H. M., Kirk, A. D., Gaston, R. S., Rogers, J., Farney, A. C., Orlando, G., Stratta, R. J., Mohan, S., Ma, L., Langefeld, C. D., Hicks, P. J., Palmer, N. D., Adams, P. L., Palanisamy, A., Reeves-Daniel, A. M. & Divers, J. 2015. Apolipoprotein L1 gene variants in deceased organ donors are associated with renal allograft failure. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, 15, 1615-22.
- Freedman, B. I., Pastan, S. O., Israni, A. K., Schladt, D., Julian, B. A., Gautreaux, M. D., Hauptfeld, V., Bray, R. A., Gebel, H. M., Kirk, A. D., Gaston, R. S., Rogers, J., Farney, A. C., Orlando,

- G., Stratta, R. J., Mohan, S., Ma, L., Langefeld, C. D., Bowden, D. W., Hicks, P. J., Palmer, N. D., Palanisamy, A., Reeves-Daniel, A. M., Brown, W. M. & Divers, J. 2016. APOL1 Genotype and Kidney Transplantation Outcomes From Deceased African American Donors. *Transplantation*, 100, 194-202.
- Friedman, D. J. 2019. Genes and environment in chronic kidney disease hotspots. *Current opinion in nephrology and hypertension*, 28, 87-96.
- Friedman, D. J. & Pollak, M. R. 2021. APOL1 Nephropathy: From Genetics to Clinical Applications. *Clinical journal of the American Society of Nephrology : CJASN*, 16, 294-303.
- Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. 2019. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature communications*, 10, 1776.
- Gelen, V., Kükürt, A. & Şengül, E. 2021. Role of the renin-angiotensin-aldosterone system in various disease processes: An overview. *Renin-Angiotensin Aldosterone System*, 55.
- Genovese, G., Friedman, D. J., Ross, M. D., Lecordier, L., Uzureau, P., Freedman, B. I., Bowden, D. W., Langefeld, C. D., Oleksyk, T. K. & Uscinski Knob, A. L. 2010. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science*, 329, 841-45.
- George, C., Yako, Y. Y., Okpechi, I. G., Matsha, T. E., Folefack, F. J. K. & Kengne, A. P. 2018. An African perspective on the genetic risk of chronic kidney disease: a systematic review. *BMC medical genetics*, 19, 187.
- George, J. A., Brandenburg, J.-T., Fabian, J., Crowther, N. J., Agongo, G., Alberts, M., Ali, S., Asiki, G., Boua, P. R. & Gómez-Olivé, F. X. 2019. Kidney damage and associated risk factors in rural and urban sub-Saharan Africa (AWI-Gen): a cross-sectional population study. *The Lancet Global Health*, 7, e1632-e43.
- Georgianos, P. I. & Agarwal, R. 2023. Hypertension in chronic kidney disease-treatment standard 2023. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*, 38, 2694-703.
- Ghosh, M., Ph. D, Soumitra & Collins, M., Ph. D, Francis S 1996. The geneticist's approach to complex disease. *Annual review of medicine*, 47, 333-53.
- Gjerde, A. 2022. Low birth weight, intrauterine growth restriction and risk of chronic kidney disease in adult age.
- Glazyrin, Y. E., Veprintsev, D. V., Ler, I. A., Rossovskaya, M. L., Varygina, S. A., Glizer, S. L., Zamay, T. N., Petrova, M. M., Minic, Z., Berezovski, M. V. & Kichkailo, A. S. 2020. Proteomics-Based Machine Learning Approach as an Alternative to Conventional Biomarkers for Differential Diagnosis of Chronic Kidney Diseases. *Int J Mol Sci*, 21.
- Go, A. S., Chertow, G. M., Fan, D., McCulloch, C. E. & Hsu, C.-Y. 2004. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *New England Journal of Medicine*, 351, 1296-305.
- Goek, O.-N., Köttgen, A., Hoogeveen, R. C., Ballantyne, C. M., Coresh, J. & Astor, B. C. 2012. Association of apolipoprotein A1 and B with kidney function and chronic kidney disease in two multiethnic population samples. *Nephrology Dialysis Transplantation*, 27, 2839-47.
- Gómez-Olivé, F. X., Ali, S. A., Made, F., Kyobutungi, C., Nonterah, E., Micklesfield, L., Alberts, M., Boua, R., Hazelhurst, S., Debpuur, C., Mashinya, F., Dikotope, S., Sorgho, H., Cook, I., Muthuri, S., Soo, C., Mukomana, F., Agongo, G., Wandabwa, C., Afolabi, S., Oduro, A., Tinto, H., Wagner, R. G., Haregu, T., Wade, A., Kahn, K., Norris, S. A., Crowther, N. J., Tollman, S., Sankoh, O. & Ramsay, M. 2017. Regional and Sex Differences in the Prevalence and Awareness of Hypertension: An H3Africa AWI-Gen Study Across 6 Sites in Sub-Saharan Africa. *Global heart*, 12, 81-90.
- Gong, Y., Hart, E., Shchurin, A. & Hoover-Plow, J. 2008. Inflammatory macrophage migration requires MMP-9 activation by plasminogen in mice. *The Journal of clinical investigation*, 118, 3012-24.

- Gonzalez-Calero, L., Martin-Lorenzo, M., Martínez, P. J., Baldan-Martin, M., Ruiz-Hurtado, G., Segura, J., De La Cuesta, F., Barderas, M. G., Ruilope, L. M. & Vivanco, F. 2016. Hypertensive patients exhibit an altered metabolism. A specific metabolite signature in urine is able to predict albuminuria progression. *Translational Research*, 178, 25-37. e7.
- Good, D. M., Zürlbig, P., Argilés, A., Bauer, H. W., Behrens, G., Coon, J. J., Dakna, M., Decramer, S., Delles, C., Dominiczak, A. F., Ehrich, J. H., Eitner, F., Fliser, D., Frommberger, M., Ganser, A., Girolami, M. A., Golovko, I., Gwinner, W., Haubitz, M., Herget-Rosenthal, S., Jankowski, J., Jahn, H., Jerums, G., Julian, B. A., Kellmann, M., Kliem, V., Kolch, W., Krolewski, A. S., Luppi, M., Massy, Z., Melter, M., Neusüss, C., Novak, J., Peter, K., Rossing, K., Rupperecht, H., Schanstra, J. P., Schiffer, E., Stolzenburg, J. U., Tarnow, L., Theodorescu, D., Thongboonkerd, V., Vanholder, R., Weissinger, E. M., Mischak, H. & Schmitt-Kopplin, P. 2010. Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. *Molecular & cellular proteomics : MCP*, 9, 2424-37.
- Govender, I. S., Mokoena, R., Stoychev, S. & Naicker, P. 2023. Urine-HILIC: Automated sample preparation for bottom-up urinary proteome profiling in clinical proteomics. *Proteomes*, 11, 29.
- Govender, M. A., Brandenburg, J. T., Fabian, J. & Ramsay, M. 2021. The Use of 'Omics for Diagnosing and Predicting Progression of Chronic Kidney Disease: A Scoping Review. *Front Genet*, 12, 682929.
- Govender, M. A., Stoychev, S. H., Brandenburg, J.-T., Ramsay, M., Fabian, J. & Govender, I. S. 2024. Proteomic insights into the pathophysiology of hypertension-associated albuminuria: Pilot study in a South African cohort. *Clinical proteomics*, 21, 1-12.
- Grobe, N., Scheiber, J., Zhang, H., Garbe, C. & Wang, X. 2023. Omics and Artificial Intelligence in Kidney Diseases. *Advances in Kidney Disease and Health*, 30, 47-52.
- Groopman, E. E., Marasa, M., Cameron-Christie, S., Petrovski, S., Aggarwal, V. S., Milo-Rasouly, H., Li, Y., Zhang, J., Nestor, J. & Krithivasan, P. 2019. Diagnostic utility of exome sequencing for kidney disease. *New England Journal of Medicine*, 380, 142-51.
- Guan, S., Taylor, P. P., Han, Z., Moran, M. F. & Ma, B. 2020. Data dependent-independent acquisition (DDIA) proteomics. *Journal of proteome research*, 19, 3230-37.
- Gurdasani, D., Carstensen, T., Fatumo, S., Chen, G., Franklin, C. S., Prado-Martinez, J., Bouman, H., Abascal, F., Haber, M., Tachmazidou, I., Mathieson, I., Ekoru, K., Degorter, M. K., Nsubuga, R. N., Finan, C., Wheeler, E., Chen, L., Cooper, D. N., Schiffels, S., Chen, Y., Ritchie, G. R. S., Pollard, M. O., Fortune, M. D., Mentzer, A. J., Garrison, E., Bergström, A., Hatzikotoulas, K., Adeyemo, A., Doumatey, A., Elding, H., Wain, L. V., Ehret, G., Auer, P. L., Kooperberg, C. L., Reiner, A. P., Franceschini, N., Maher, D., Montgomery, S. B., Kadie, C., Widmer, C., Xue, Y., Seeley, J., Asiki, G., Kamali, A., Young, E. H., Pomilla, C., Soranzo, N., Zeggini, E., Pirie, F., Morris, A. P., Heckerman, D., Tyler-Smith, C., Motala, A. A., Rotimi, C., Kaleebu, P., Barroso, I. & Sandhu, M. S. 2019. Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. *Cell*, 179, 984-1002.e36.
- Gusev, E., Solomatina, L., Zhuravleva, Y. & Sarapultsev, A. 2021. The pathogenesis of end-stage renal disease from the standpoint of the theory of general pathological processes of inflammation. *International journal of molecular sciences*, 22, 11453.
- Gyurászová, M., Gurecká, R., Bábíčková, J. & Tóthová, Ľ. 2020. Oxidative stress in the pathophysiology of kidney disease: implications for noninvasive monitoring and identification of biomarkers. *Oxidative medicine and cellular longevity*, 2020, 5478708.
- H3africa 2024. Available from: <https://h3africa.org/index.php/h3africa-success-stories-collaborative-centers-research-projects/> (accessed March 2024).

- Hall, J. E. 2016. Kidney dysfunction mediates salt-induced increases in blood pressure. *Circulation*, 133, 894.
- Hamdi, Y., Zass, L., Othman, H., Radouani, F., Allali, I., Hanachi, M., Okeke, C. J., Chaouch, M., Tendwa, M. B., Samtal, C., Mohamed Sallam, R., Alsayed, N., Turkson, M., Ahmed, S., Benkahla, A., Romdhane, L., Souiai, O., Tastan Bishop, Ö., Ghedira, K., Mohamed Fadlelmola, F., Mulder, N. & Kamal Kassim, S. 2021. Human OMICs and Computational Biology Research in Africa: Current Challenges and Prospects. *Omics : a journal of integrative biology*, 25, 213-33.
- Hamilton, S. A., Nakanga, W. P., Prynne, J. E., Crampin, A. C., Fecht, D., Vineis, P., Caplin, B., Pearce, N. & Nyirenda, M. J. 2020. Prevalence and risk factors for chronic kidney disease of unknown cause in Malawi: a cross-sectional analysis in a rural and urban population. *BMC nephrology*, 21, 1-12.
- Hamrahian, S. M. & Falkner, B. 2016. Hypertension in chronic kidney disease. *Hypertension: from basic research to clinical practice*, 307-25.
- Hariparshad, S., Bhimma, R., Nandlal, L., Jembere, E., Naicker, S. & Assounga, A. 2023. The prevalence of chronic kidney disease in South Africa-limitations of studies comparing prevalence with sub-Saharan Africa, Africa, and globally. *BMC nephrology*, 24, 62.
- Haubitz, M., Good, D. M., Woywodt, A., Haller, H., Rupprecht, H., Theodorescu, D., Dakna, M., Coon, J. J. & Mischak, H. 2009. Identification and validation of urinary biomarkers for differential diagnosis and evaluation of therapeutic intervention in anti-neutrophil cytoplasmic antibody-associated vasculitis. *Molecular & Cellular Proteomics*, 8, 2296-307.
- Heerspink, H. J. L. & Gansevoort, R. T. 2015. Albuminuria is an appropriate therapeutic target in patients with CKD: the pro view. *Clinical journal of the American Society of Nephrology: CJASN*, 10, 1079.
- Hellwege, J. N., Velez Edwards, D. R., Giri, A., Qiu, C., Park, J., Torstenson, E. S., Keaton, J. M., Wilson, O., Robinson-Cohen, C. & Chung, C. P. 2019. Mapping eGFR loci to the renal transcriptome and phenome in the VA Million Veteran Program. *Nature communications*, 10, 3842.
- Herold, J. M., Nano, J., Gorski, M., Winkler, T. W., Stanzick, K. J., Zimmermann, M. E., Brandl, C., Peters, A., Koenig, W., Burkhardt, R., Gessner, A., Heid, I. M., Gieger, C. & Stark, K. J. 2023. Polygenic scores for estimated glomerular filtration rate in a population of general adults and elderly - comparative results from the KORA and AugUR study. *BMC genomic data*, 24, 28.
- Hildonen, S., Skarpen, E., Halvorsen, T. G. & Reubsaet, L. 2016. Isolation and mass spectrometry analysis of urinary extraexosomal proteins. *Sci Rep*, 6, 36331.
- Hill, C., Avila-Palencia, I., Maxwell, A. P., Hunter, R. F. & Mcknight, A. J. 2022. Harnessing the Full Potential of Multi-Omic Analyses to Advance the Study and Treatment of Chronic Kidney Disease. *Frontiers in nephrology*, 2, 923068.
- Hill, N. R., Fatoba, S. T., Oke, J. L., Hirst, J. A., O'callaghan, C. A., Lasserson, D. S. & Hobbs, F. R. 2016. Global prevalence of chronic kidney disease—a systematic review and meta-analysis. *PloS one*, 11, e0158765.
- Horgan, R. P. & Kenny, L. C. 2011. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist*, 13, 189-95.
- Hu, A., Noble, W. S. & Wolf-Yadlin, A. 2016. Technical advances in proteomics: new developments in data-independent acquisition. *F1000Research*, 5.
- Huang, R., Fu, P. & Ma, L. 2023. Kidney fibrosis: from mechanisms to therapeutic medicines. *Signal Transduction and Targeted Therapy*, 8, 129.
- Hughes, O., Bentley, A. R., Breeze, C. E., Aguet, F., Xu, X., Nadkarni, G., Sun, Q., Lin, B. M., Gilliland, T., Meyer, M. C., Du, J., Raffield, L. M., Kramer, H., Morton, R. W., Gouveia, M.

- H., Atkinson, E. G., Valladares-Salgado, A., Wacher-Rodarte, N., Dueker, N. D., Guo, X., Hai, Y., Adeyemo, A., Best, L. G., Cai, J., Chen, G., Chong, M., Doumatey, A., Eales, J., Goodarzi, M. O., Ipp, E., Irvin, M. R., Jiang, M., Jones, A. C., Kooperberg, C., Krieger, J. E., Lange, E. M., Lanktree, M. B., Lash, J. P., Lotufo, P. A., Loos, R. J. F., Ha My, V. T., Peralta-Romero, J., Qi, L., Raffel, L. J., Rich, S. S., Rodriguez, E. J., Tarazona-Santos, E., Taylor, K. D., Umans, J. G., Wen, J., Young, B. A., Yu, Z., Zhang, Y., Ida Chen, Y. D., Rundek, T., Rotter, J. I., Cruz, M., Fornage, M., Lima-Costa, M. F., Pereira, A. C., Paré, G., Natarajan, P., Cole, S. A., Carson, A. P., Lange, L. A., Li, Y., Perez-Stable, E. J., Do, R., Charchar, F. J., Tomaszewski, M., Mychaleckyj, J. C., Rotimi, C., Morris, A. P. & Franceschini, N. 2023. Genome-wide study investigating effector genes and polygenic prediction for kidney function in persons with ancestry from Africa and the Americas. *Cell genomics*, 4, 100468.
- Hughes, O., Bentley, A. R., Breeze, C. E., Aguet, F., Xu, X., Nadkarni, G., Sun, Q., Lin, B. M., Gilliland, T., Meyer, M. C., Du, J., Raffield, L. M., Kramer, H., Morton, R. W., Gouveia, M. H., Atkinson, E. G., Valladares-Salgado, A., Wacher-Rodarte, N., Dueker, N. D., Guo, X., Hai, Y., Adeyemo, A., Best, L. G., Cai, J., Chen, G., Chong, M., Doumatey, A., Eales, J., Goodarzi, M. O., Ipp, E., Irvin, M. R., Jiang, M., Jones, A. C., Kooperberg, C., Krieger, J. E., Lange, E. M., Lanktree, M. B., Lash, J. P., Lotufo, P. A., Loos, R. J. F., Ha My, V. T., Peralta-Romero, J., Qi, L., Raffel, L. J., Rich, S. S., Rodriguez, E. J., Tarazona-Santos, E., Taylor, K. D., Umans, J. G., Wen, J., Young, B. A., Yu, Z., Zhang, Y., Ida Chen, Y. D., Rundek, T., Rotter, J. I., Cruz, M., Fornage, M., Lima-Costa, M. F., Pereira, A. C., Paré, G., Natarajan, P., Cole, S. A., Carson, A. P., Lange, L. A., Li, Y., Perez-Stable, E. J., Do, R., Charchar, F. J., Tomaszewski, M., Mychaleckyj, J. C., Rotimi, C., Morris, A. P. & Franceschini, N. 2024. Genome-wide study investigating effector genes and polygenic prediction for kidney function in persons with ancestry from Africa and the Americas. *Cell genomics*, 4, 100468.
- Hughson, M. D., Puellas, V. G., Hoy, W. E., Douglas-Denton, R. N., Mott, S. A. & Bertram, J. F. 2014. Hypertension, glomerular hypertrophy and nephrosclerosis: the effect of race. *Nephrology Dialysis Transplantation*, 29, 1399-409.
- Igo Jr, R. P., Iyengar, S. K., Nicholas, S. B., Goddard, K. A., Langefeld, C. D., Hanson, R. L., Duggirala, R., Divers, J., Abboud, H. & Adler, S. G. 2011. Genomewide linkage scan for diabetic renal failure and albuminuria: the FIND study. *American journal of nephrology*, 33, 381-89.
- Inzaule, S. C., Tessema, S. K., Kebede, Y., Ogwel Ouma, A. E. & Nkengasong, J. N. 2021. Genomic-informed pathogen surveillance in Africa: opportunities and challenges. *The Lancet. Infectious diseases*, 21, e281-e89.
- Jefferis, J., Hudson, R., Lacaze, P., Bakshi, A., Hawley, C., Patel, C. & Mallett, A. 2023. Monogenic and polygenic concepts in chronic kidney disease (CKD). *Journal of nephrology*.
- Jiang, L., He, L. & Fountoulakis, M. 2004. Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. *Journal of chromatography. A*, 1023, 317-20.
- Jiang, L., Zheng, Z., Qi, T., Kemper, K. E., Wray, N. R., Visscher, P. M. & Yang, J. 2019. A resource-efficient tool for mixed model association analysis of large-scale data. *Nature genetics*, 51, 1749-55.
- Jovanović, D., Krstivojević, P., Obradović, I., Durdević, V. & Dukanović, L. 2003. Serum cystatin C and beta2-microglobulin as markers of glomerular filtration rate. *Renal failure*, 25, 123-33.
- Kaburagi, Y., Takahashi, E., Kajio, H., Yamashita, S., Yamamoto-Honda, R., Shiga, T., Okumura, A., Goto, A., Fukazawa, Y. & Seki, N. 2019. Urinary afamin levels are associated with the progression of diabetic nephropathy. *Diabetes research and clinical practice*, 147, 37-46.
- Kalantari, S., Jafari, A., Moradpoor, R., Ghasemi, E. & Khalkhal, E. 2015. Human Urine Proteomics: Analytical Techniques and Clinical Applications in Renal Diseases. *International journal of proteomics*, 2015, 782798.

- Kalyesubula, R., Fabian, J., Nakanga, W., Newton, R., Ssebunnya, B., Prynn, J., George, J., Wade, A. N., Seeley, J. & Nitsch, D. 2020. How to estimate glomerular filtration rate in sub-Saharan Africa: design and methods of the African Research into Kidney Diseases (ARK) study. *BMC nephrology*, 21, 1-12.
- Kamiza, A. B., Toure, S. M., Vujkovic, M., Machipisa, T., Soremekun, O. S., Kintu, C., Corpas, M., Pirie, F., Young, E., Gill, D., Sandhu, M. S., Kaleebu, P., Nyirenda, M., Motala, A. A., Chikowore, T. & Fatumo, S. 2022. Transferability of genetic risk scores in African populations. *Nature medicine*, 28, 1163-66.
- Kang, T., Ding, W., Zhang, L., Ziemek, D. & Zarringhalam, K. 2017. A biological network-based regularized artificial neural network model for robust phenotype prediction from gene expression data. *BMC bioinformatics*, 18, 565.
- Kao, W. L., Klag, M. J., Meoni, L. A., Reich, D., Berthier-Schaad, Y., Li, M., Coresh, J., Patterson, N., Tandon, A. & Powe, N. R. 2008. MYH9 is associated with nondiabetic end-stage renal disease in African Americans. *Nature genetics*, 40, 1185-92.
- Kasembeli, A. N., Duarte, R., Ramsay, M., Mosiane, P., Dickens, C., Dix-Peek, T., Limou, S., Sezgin, E., Nelson, G. W. & Fogo, A. B. 2015. APOL1 risk variants are strongly associated with HIV-associated nephropathy in black South Africans. *Journal of the American Society of Nephrology: JASN*, 26, 2882.
- Kaze, A. D., Ilori, T., Jaar, B. G. & Echouffo-Tcheugui, J. B. 2018. Burden of chronic kidney disease on the African continent: a systematic review and meta-analysis. *BMC nephrology*, 19, 1-11.
- Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T. & Kathiresan, S. 2018a. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics*, 50, 1219-24.
- Khera, R., Lu, Y., Lu, J., Saxena, A., Nasir, K., Jiang, L. & Krumholz, H. M. 2018b. Impact of 2017 ACC/AHA guidelines on prevalence of hypertension and eligibility for antihypertensive treatment in United States and China: nationally representative cross sectional study. *bmj*, 362.
- Kintu, C., Soremekun, O., Kamiza, A. B., Kalungi, A., Mayanja, R., Kalyesubula, R., Bagaya, B., Jjingo, D., Fabian, J. & Gill, D. 2023. The causal effects of lipid traits on kidney function in Africans: bidirectional and multivariable Mendelian-randomization study. *EBioMedicine*, 90.
- Koch, S., Schmidtke, J., Krawczak, M. & Caliebe, A. 2023. Clinical utility of polygenic risk scores: a critical 2023 appraisal. *Journal of community genetics*, 14, 471-87.
- Kockum, I., Huang, J. & Stridh, P. 2023. Overview of Genotyping Technologies and Methods. *Current protocols*, 3, e727.
- Konukoglu, D. & Uzun, H. 2017. Endothelial dysfunction and hypertension. *Hypertension: from basic research to clinical practice*, 511-40.
- Kopp, J. B., Nelson, G. W., Sampath, K., Johnson, R. C., Genovese, G., An, P., Friedman, D., Briggs, W., Dart, R. & Korbet, S. 2011. APOL1 genetic variants in focal segmental glomerulosclerosis and HIV-associated nephropathy. *Journal of the American Society of Nephrology: JASN*, 22, 2129.
- Kostov, K. 2021. The causal relationship between endothelin-1 and hypertension: Focusing on endothelial dysfunction, arterial stiffness, vascular remodeling, and blood pressure regulation. *Life*, 11, 986.
- Köttgen, A., Cornec-Le Gall, E., Halbritter, J., Kiryluk, K., Mallett, A. J., Parekh, R. S., Rasouly, H. M., Sampson, M. G., Tin, A. & Antignac, C. 2022. Genetics in chronic kidney disease: Conclusions from a kidney disease: Improving global outcomes (KDIGO) controversies conference. *Kidney International*, 101, 1126-41.

- Köttgen, A., Glazer, N. L., Dehghan, A., Hwang, S.-J., Katz, R., Li, M., Yang, Q., Gudnason, V., Launer, L. J. & Harris, T. B. 2009. Multiple loci associated with indices of renal function and chronic kidney disease. *Nature genetics*, 41, 712-17.
- Köttgen, A., Pattaro, C., Böger, C. A., Fuchsberger, C., Olden, M., Glazer, N. L., Parsa, A., Gao, X., Yang, Q. & Smith, A. V. 2010. New loci associated with kidney function and chronic kidney disease. *Nature genetics*, 42, 376-84.
- Kwon, S., Kim, D. K., Oh, K.-H., Joo, K. W., Lim, C. S., Kim, Y. S. & Han, S. S. 2021. Apolipoprotein B is a risk factor for end-stage renal disease. *Clinical Kidney Journal*, 14, 617-23.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczy, J., Levine, R., Mcewan, P., Mckernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., Mcpherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chisoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.
- Laurance, W. F., Sloan, S., Weng, L. & Sayer, J. A. 2015. Estimating the Environmental Costs of Africa's Massive "Development Corridors". *Current biology : CB*, 25, 3202-8.
- Lecamwasam, A., Novakovic, B., Meyer, B., Ekinci, E. I., Dwyer, K. M. & Saffery, R. 2020. DNA methylation profiling identifies epigenetic differences between early versus late stages of diabetic chronic kidney disease. *Nephrology Dialysis Transplantation*.
- Levey, A. S., Coresh, J., Bolton, K., Culeton, B., Harvey, K. S., Ikizler, T. A., Johnson, C. A., Kausz, A., Kimmel, P. L. & Kusek, J. 2002. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *American Journal of Kidney Diseases*, 39.
- Levey, A. S., De Jong, P. E., Coresh, J., El Nahas, M., Astor, B. C., Matsushita, K., Gansevoort, R. T., Kasiske, B. L. & Eckardt, K. U. 2011. The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report. *Kidney Int*, 80, 17-28.
- Levey, A. S., Stevens, L. A., Schmid, C. H., Zhang, Y., Castro Iii, A. F., Feldman, H. I., Kusek, J. W., Eggers, P., Van Lente, F. & Greene, T. 2009. A new equation to estimate glomerular filtration rate. *Annals of internal medicine*, 150, 604-12.
- Levin, A. & Stevens, P. E. 2014. Summary of KDIGO 2012 CKD Guideline: behind the scenes, need for guidance, and a framework for moving forward. *Kidney international*, 85, 49-61.
- Levin, A., Stevens, P. E., Bilous, R. W., Coresh, J., De Francisco, A. L., De Jong, P. E., Griffith, K. E., Hemmelgarn, B. R., Iseki, K. & Lamb, E. J. 2013. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney international supplements*, 3, 1-150.

- Li, X., Dai, Y., Chuang, P. Y. & He, J. C. 2014. Induction of retinol dehydrogenase 9 expression in podocytes attenuates kidney injury. *Journal of the American Society of Nephrology*, 25, 1933-41.
- Liao, L.-N., Chen, C.-C., Wu, F.-Y., Lin, C.-C., Hsiao, J.-H., Chang, C.-T., Kardia, S. L., Li, T.-C. & Tsai, F.-J. 2014. Identified single-nucleotide polymorphisms and haplotypes at 16q22. 1 increase diabetic nephropathy risk in Han Chinese population. *BMC genetics*, 15, 1-11.
- Limou, S., Nelson, G. W., Kopp, J. B. & Winkler, C. A. 2014. APOL1 kidney risk alleles: population genetics and disease associations. *Advances in chronic kidney disease*, 21, 426-33.
- Lin, L., Yu, Q., Zheng, J., Cai, Z. & Tian, R. 2018. Fast quantitative urinary proteomic profiling workflow for biomarker discovery in kidney cancer. *Clinical proteomics*, 15, 42.
- Lin, T.-Y. & Hsu, Y.-H. 2020. IL-20 in acute kidney injury: role in pathogenesis and potential as a therapeutic target. *International journal of molecular sciences*, 21, 1009.
- Lindhardt, M., Persson, F., Currie, G., Pontillo, C., Beige, J., Delles, C., Von Der Leyen, H., Mischak, H., Navis, G. & Noutsou, M. 2016. Proteomic prediction and Renin angiotensin aldosterone system Inhibition prevention Of early diabetic nephRopathy in TYpe 2 diabetic patients with normoalbuminuria (PRIORITY): essential study design and rationale of a randomised clinical multicentre trial. *BMJ open*, 6, e010310.
- Lindhardt, M., Persson, F., Zürgb, P., Stalmach, A., Mischak, H., De Zeeuw, D., Lambers Heerspink, H., Klein, R., Orchard, T., Porta, M., Fuller, J., Bilous, R., Chaturvedi, N., Parving, H. H. & Rossing, P. 2017. Urinary proteomics predict onset of microalbuminuria in normoalbuminuric type 2 diabetic patients, a sub-study of the DIRECT-Protect 2 study. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*, 32, 1866-73.
- Liu, H., Doke, T., Guo, D., Sheng, X., Ma, Z., Park, J., Vy, H. M. T., Nadkarni, G. N., Abedini, A., Miao, Z., Palmer, M., Voight, B. F., Li, H., Brown, C. D., Ritchie, M. D., Shu, Y. & Susztak, K. 2022. Epigenomic and transcriptomic analyses define core cell types, genes and targetable mechanisms for kidney disease. *Nature genetics*, 54, 950-62.
- Liu, L. & Kiryluk, K. 2018. Genome-wide polygenic risk predictors for kidney disease. *Nature Reviews Nephrology*, 14, 723-24.
- Loh, P. R., Danecek, P., Palamara, P. F., Fuchsberger, C., Y, A. R., H, K. F., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G. R., Durbin, R. & A, L. P. 2016. Reference-based phasing using the Haplotype Reference Consortium panel. *Nature genetics*, 48, 1443-48.
- Lopez-Giacoman, S. & Madero, M. 2015. Biomarkers in chronic kidney disease, from kidney function to kidney damage. *World journal of nephrology*, 4, 57.
- López-Novoa, J. M., Rodríguez-Peña, A. B., Ortiz, A., Martínez-Salgado, C. & López Hernández, F. J. 2011. Etiopathology of chronic tubular, glomerular and renovascular nephropathies: clinical implications. *Journal of translational medicine*, 9, 1-26.
- Ma, L., Divers, J. & Freedman, B. I. 2019. Mechanisms of injury in APOL1-associated kidney disease. *Transplantation*, 103, 487.
- Ma, T., Liu, T., Xie, P., Jiang, S., Yi, W., Dai, P. & Guo, X. 2020. UPLC-MS-based urine nontargeted metabolic profiling identifies dysregulation of pantothenate and CoA biosynthesis pathway in diabetic kidney disease. *Life Sciences*, 258, 118160.
- Macleod, I., Bowman, P., Vander Jagt, C., Haile-Mariam, M., Kemper, K., Chamberlain, A., Schrooten, C., Hayes, B. J. & Goddard, M. 2016. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC genomics*, 17, 1-21.
- Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X. & Sham, P. C. 2017. Polygenic scores via penalized regression on summary statistics. *Genetic epidemiology*, 41, 469-80.

- Mallett, A., Patel, C., Salisbury, A., Wang, Z., Healy, H. & Hoy, W. 2014. The prevalence and epidemiology of genetic renal disease amongst adults with chronic kidney disease in Australia. *Orphanet Journal of Rare Diseases*, 9, 1-9.
- Mancia, G., Kreutz, R., Brunström, M., Burnier, M., Grassi, G., Januszewicz, A., Muiesan, M. L., Tsioufis, K., Agabiti-Rosei, E., Algharably, E. a. E., Azizi, M., Benetos, A., Borghi, C., Hitij, J. B., Cifkova, R., Coca, A., Cornelissen, V., Cruickshank, J. K., Cunha, P. G., Danser, A. H. J., Pinho, R. M., Delles, C., Dominiczak, A. F., Dorobantu, M., Doumas, M., Fernández-Alfonso, M. S., Halimi, J. M., Járαι, Z., Jelaković, B., Jordan, J., Kuznetsova, T., Laurent, S., Lovic, D., Lurbe, E., Mahfoud, F., Manolis, A., Miglinas, M., Narkiewicz, K., Niiranen, T., Palatini, P., Parati, G., Pathak, A., Persu, A., Polonia, J., Redon, J., Sarafidis, P., Schmieder, R., Spronck, B., Stabouli, S., Stergiou, G., Taddei, S., Thomopoulos, C., Tomaszewski, M., Van De Borne, P., Wanner, C., Weber, T., Williams, B., Zhang, Z. Y. & Kjeldsen, S. E. 2023. 2023 ESH Guidelines for the management of arterial hypertension The Task Force for the management of arterial hypertension of the European Society of Hypertension: Endorsed by the International Society of Hypertension (ISH) and the European Renal Association (ERA). *Journal of hypertension*, 41, 1874-2071.
- Mann, N., Braun, D. A., Amann, K., Tan, W., Shril, S., Connaughton, D. M., Nakayama, M., Schneider, R., Kitzler, T. M. & Van Der Ven, A. T. 2019. Whole-exome sequencing enables a precision medicine approach for kidney transplant recipients. *Journal of the American Society of Nephrology: JASN*, 30, 201.
- Márquez-Luna, C., Gazal, S., Loh, P.-R., Kim, S. S., Furlotte, N., Auton, A. & Price, A. L. 2021. Incorporating functional priors improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. *Nature communications*, 12, 6052.
- Mattson, D. L. 2014. Infiltrating immune cells in the kidney in salt-sensitive hypertension and renal injury. *American Journal of Physiology-Renal Physiology*, 307, F499-F508.
- Mayanja, R., Machipisa, T., Soremekun, O., Kamiza, A. B., Kintu, C., Kalungi, A., Kalyesubula, R., Sande, O. J., Jjingo, D., Fabian, J., Robinson-Cohen, C., Franceschini, N., Nitsch, D., Nyirenda, M., Zeggini, E., Morris, A. P., Chikowore, T. & Fatumo, S. 2023. Genome-wide association analysis of cystatin-C kidney function in continental Africa. *EBioMedicine*, 95, 104775.
- Mccarthy, S., Das, S., Kretschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., Veldink, J., Peters, U., Pato, C., Van Duijn, C. M., Gillies, C. E., Gandin, I., Mezzavilla, M., Gilly, A., Cocca, M., Traglia, M., Angius, A., Barrett, J. C., Boomsma, D., Branham, K., Breen, G., Brummett, C. M., Busonero, F., Campbell, H., Chan, A., Chen, S., Chew, E., Collins, F. S., Corbin, L. J., Smith, G. D., Dedoussis, G., Dorr, M., Farmaki, A. E., Ferrucci, L., Forer, L., Fraser, R. M., Gabriel, S., Levy, S., Groop, L., Harrison, T., Hattersley, A., Holmen, O. L., Hveem, K., Kretzler, M., Lee, J. C., Mcgue, M., Meitinger, T., Melzer, D., Min, J. L., Mohlke, K. L., Vincent, J. B., Nauck, M., Nickerson, D., Palotie, A., Pato, M., Pirastu, N., Mcinnis, M., Richards, J. B., Sala, C., Salomaa, V., Schlessinger, D., Schoenherr, S., Slagboom, P. E., Small, K., Spector, T., Stambolian, D., Tuke, M., Tuomilehto, J., Van Den Berg, L. H., Van Rheenen, W., Volker, U., Wijmenga, C., Toniolo, D., Zeggini, E., Gasparini, P., Sampson, M. G., Wilson, J. F., Frayling, T., De Bakker, P. I., Swertz, M. A., Mccarroll, S., Kooperberg, C., Dekker, A., Altshuler, D., Willer, C., Iacono, W., Ripatti, S., et al. 2016. A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics*, 48, 1279-83.
- Mennuni, S., Rubattu, S., Pierelli, G., Tocci, G., Fofi, C. & Volpe, M. 2014. Hypertension and kidneys: unraveling complex molecular mechanisms underlying hypertensive renal damage. *Journal of human hypertension*, 28, 74-79.

- Meremo, A. J., Masalu, M. B., Sabi, I., Ngilangwa, D. P., Kapinga, J., Tagalile, R., Munyogwa, M. J. & Mwashambwa, M. Y. 2018. Prevalence and risk factors associated with chronic kidney disease among patients presenting at a haemodialysis unit in dodoma, Tanzania. *The East African Health Research Journal*, 2, 53.
- Miller, W. G., Bruns, D. E., Hortin, G. L., Sandberg, S., Aakre, K. M., Mcqueen, M. J., Itoh, Y., Lieske, J. C., Seccombe, D. W., Jones, G., Bunk, D. M., Curhan, G. C. & Narva, A. S. 2009. Current issues in measurement and reporting of urinary albumin excretion. *Clinical chemistry*, 55, 24-38.
- Mills, K. T., Bundy, J. D., Kelly, T. N., Reed, J. E., Kearney, P. M., Reynolds, K., Chen, J. & He, J. 2016. Global disparities of hypertension prevalence and control: a systematic analysis of population-based studies from 90 countries. *Circulation*, 134, 441-50.
- Mischak, H., Delles, C., Klein, J. & Schanstra, J. P. 2010. Urinary proteomics based on capillary electrophoresis-coupled mass spectrometry in kidney disease: discovery and validation of biomarkers, and clinical application. *Advances in chronic kidney disease*, 17, 493-506.
- Montesinos, P., Bergua, J. M., Vellenga, E., Rayón, C., Parody, R., De La Serna, J., León, A., Esteve, J., Milone, G. & Debén, G. 2009. Differentiation syndrome in patients with acute promyelocytic leukemia treated with all-trans retinoic acid and anthracycline chemotherapy: characteristics, outcome, and prognostic factors. *Blood*, 113, 775-83.
- Moon, J.-Y. 2013. Recent update of renin-angiotensin-aldosterone system in the pathogenesis of hypertension. *Electrolytes & Blood Pressure: E & BP*, 11, 41.
- Morris, A. P., Le, T. H., Wu, H., Akbarov, A., Van Der Most, P. J., Hemani, G., Smith, G. D., Mahajan, A., Gaulton, K. J. & Nadkarni, G. N. 2019. Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nature communications*, 10, 1-14.
- Muiru, A. N., Charlebois, E. D., Balzer, L. B., Kwarisiima, D., Elly, A., Black, D., Okiror, S., Kabami, J., Atukunda, M. & Snyman, K. 2020. The epidemiology of chronic kidney disease (CKD) in rural East Africa: A population-based study. *PLoS One*, 15, e0229649.
- Mulder, N., Abimiku, A. L., Adebamowo, S. N., De Vries, J., Matimba, A., Olowoyo, P., Ramsay, M., Skelton, M. & Stein, D. J. 2018. H3Africa: current perspectives. *Pharmacogenomics and personalized medicine*, 59-66.
- Muñoz-Durango, N., Fuentes, C. A., Castillo, A. E., González-Gómez, L. M., Vecchiola, A., Fardella, C. E. & Kalergis, A. M. 2016. Role of the renin-angiotensin-aldosterone system beyond blood pressure regulation: molecular and cellular mechanisms involved in end-organ damage during arterial hypertension. *International journal of molecular sciences*, 17, 797.
- Muntner, P., Carey, R. M., Gidding, S., Jones, D. W., Taler, S. J., Wright Jr, J. T. & Whelton, P. K. 2018. Potential US population impact of the 2017 ACC/AHA high blood pressure guideline. *Circulation*, 137, 109-18.
- Navar, L., Inscho, E., Majid, S., Imig, J., Harrison-Bernard, L. & Mitchell, K. 1996. Paracrine regulation of the renal microcirculation. *Physiological reviews*, 76, 425-536.
- Navise, N. H., Mokwatsi, G. G., Gafane-Matemané, L. F., Fabian, J. & Lammertyn, L. 2023. Kidney dysfunction: prevalence and associated risk factors in a community-based study from the North West Province of South Africa. *BMC nephrology*, 24, 1-8.
- Nayak, P., Panda, S., Thatoi, P. K., Rattan, R., Mohapatra, S. & Mishra, P. K. 2016. Evaluation of lipid profile and apolipoproteins in essential hypertensive patients. *Journal of Clinical and Diagnostic Research: JCDR*, 10, BC01.
- Nguyen, Q., Dominguez, J., Nguyen, L. & Gullapalli, N. 2010. Hypertension management: an update. *American health & drug benefits*, 3, 47.

- Ning, J., Xiang, Z., Xiong, C., Zhou, Q., Wang, X. & Zou, H. 2020. Alpha1-antitrypsin in urinary extracellular vesicles: A potential biomarker of diabetic kidney disease prior to microalbuminuria. *Diabetes, Metabolic Syndrome and Obesity*, 2037-48.
- Nkuipou-Kenfack, E., Duranton, F., Gayraud, N., Argiles, A., Lundin, U., Weinberger, K. M., Dakna, M., Delles, C., Mullen, W. & Husi, H. 2014. Assessment of metabolomic and proteomic biomarkers in detection and prognosis of progression of renal function in chronic kidney disease. *PLoS one*, 9, e96955.
- Nqebelele, N. U., Dickens, C., Dix-Peek, T., Duarte, R. & Naicker, S. 2019. Urinary Uromodulin Levels and UMOD Variants in Black South Africans with Hypertension-Attributed Chronic Kidney Disease. *International journal of nephrology*, 2019, 8094049.
- Nweke, E. E., Naicker, P., Aron, S., Stoychev, S., Devar, J., Tabb, D. L., Omoshoro-Jones, J., Smith, M. & Candy, G. 2020. SWATH-MS based proteomic profiling of pancreatic ductal adenocarcinoma tumours reveals the interplay between the extracellular matrix and related intracellular pathways. *Plos one*, 15, e0240453.
- Omenn, G. S., Nass, S. J. & Micheel, C. M. 2012. Evolution of translational omics: lessons learned and the path forward.
- Omotoso, O. E., Teibo, J. O., Atiba, F. A., Oladimeji, T., Adebesein, A. O. & Babalghith, A. O. 2022. Bridging the genomic data gap in Africa: implications for global disease burdens. *Globalization and health*, 18, 103.
- Osterman, M. D., Kinzy, T. G. & Bailey, J. N. C. 2021. Polygenic Risk Scores. *Current protocols*, 1, e126.
- Øvrehus, M. A., Zürlbig, P., Vikse, B. E. & Hallan, S. I. 2015. Urinary proteomics in chronic kidney disease: diagnosis and risk of progression beyond albuminuria. *Clinical proteomics*, 12, 21.
- Palatini, P. 2012. Glomerular hyperfiltration: a marker of early renal damage in pre-diabetes and pre-hypertension. Oxford University Press.
- Panizo, S., Martínez-Arias, L., Alonso-Montes, C., Cannata, P., Martín-Carro, B., Fernández-Martín, J. L., Naves-Díaz, M., Carrillo-López, N. & Cannata-Andía, J. B. 2021. Fibrosis in chronic kidney disease: pathogenesis and consequences. *International Journal of Molecular Sciences*, 22, 408.
- Papadopoulos, T., Krochmal, M., Cisek, K., Fernandes, M., Husi, H., Stevens, R., Bascands, J.-L., Schanstra, J. P. & Klein, J. 2016. Omics databases on kidney disease: where they can be found and how to benefit from them. *Clinical kidney journal*, 9, 343-52.
- Patassini, S., Begley, P., Xu, J., Church, S. J., Kureishy, N., Reid, S. J., Waldvogel, H. J., Faull, R. L., Snell, R. G. & Unwin, R. D. 2019. Cerebral vitamin B5 (D-pantothenic acid) deficiency as a potential cause of metabolic perturbation and neurodegeneration in Huntington's disease. *Metabolites*, 9, 113.
- Pattaro, C., Teumer, A., Gorski, M., Chu, A. Y., Li, M., Mijatovic, V., Garnaas, M., Tin, A., Sorice, R. & Li, Y. 2016. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nature communications*, 7, 1-19.
- Peer, N., Uthman, O. A. & Kengne, A. P. 2021. Rising prevalence, and improved but suboptimal management, of hypertension in South Africa: A comparison of two national surveys. *Global epidemiology*, 3, 100063.
- Peng, L., Cantor, D. I., Huang, C., Wang, K., Baker, M. S. & Nice, E. C. 2018. Tissue and plasma proteomics for early stage cancer detection. *Molecular omics*, 14, 405-23.
- Peralta, C. A., Bibbins-Domingo, K., Vittinghoff, E., Lin, F., Fornage, M., Kopp, J. B. & Winkler, C. A. 2016. APOL1 Genotype and Race Differences in Incident Albuminuria and Renal Function Decline. *Journal of the American Society of Nephrology : JASN*, 27, 887-93.
- Pereira, L., Mutesa, L., Tindana, P. & Ramsay, M. 2021. African genetic diversity and adaptation inform a precision medicine agenda. *Nature reviews. Genetics*, 22, 284-306.

- Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D. J., Prakash, A., Frericks-Zipper, A. & Eisenacher, M. 2022. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic acids research*, 50, D543-D52.
- Perneger, T. V., Whelton, P. K. & Klag, M. J. 1995. Race and end-stage renal disease: socioeconomic status and access to health care as mediating factors. *Archives of internal medicine*, 155, 1201-08.
- Pinu, F. R., Beale, D. J., Paten, A. M., Kouremenos, K., Swarup, S., Schirra, H. J. & Wishart, D. 2019. Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. *Metabolites*, 9.
- Pollock, D. M. 2000. Renal endothelin in hypertension. *Current opinion in nephrology and hypertension*, 9, 157-64.
- Pontillo, C. & Mischak, H. 2017a. Urinary peptide-based classifier CKD273: towards clinical application in chronic kidney disease. *Clinical kidney journal*, 10, 192-201.
- Pontillo, C. & Mischak, H. 2017b. Urinary peptide-based classifier CKD273: towards clinical application in chronic kidney disease. *Clin Kidney J*, 10, 192-201.
- Poulter, N. R., Castillo, R., Charchar, F. J., Schlaich, M. P., Schutte, A. E., Tomaszewski, M., Touyz, R. M. & Wang, J.-G. 2018. Are the American Heart Association/American College of Cardiology high blood pressure guidelines fit for global purpose?: Thoughts from the International Society of Hypertension. *Hypertension*, 72, 260-62.
- Prajczek, S., Heidenreich, U., Pfaller, W., Kotanko, P., Lhotta, K. & Jennings, P. 2010. Evidence for a role of uromodulin in chronic kidney disease progression. *Nephrology Dialysis Transplantation*, 25, 1896-903.
- Pregizer, S., Vreven, T., Mathur, M. & Robinson, L. N. 2023. Multi-omic single cell sequencing: Overview and opportunities for kidney disease therapeutic development. *Frontiers in molecular biosciences*, 10, 1176856.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, 38, 904-09.
- Provenzano, M., Garofalo, C., Chiodini, P., Mancuso, C., Barbato, E., De Nicola, L. & Andreucci, M. 2020. [Role of proteinuria in clinical research: for each old-answer, a new key-question.]. *Recenti progressi in medicina*, 111, 74-81.
- Provenzano, M., Serra, R., Garofalo, C., Michael, A., Crugliano, G., Battaglia, Y., Ielapi, N., Bracale, U. M., Faga, T. & Capitoli, G. 2021. Omics in chronic kidney disease: Focus on prognosis and prediction. *International Journal of Molecular Sciences*, 23, 336.
- Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'donovan, M. C., Sullivan, P. F. & Sklar, P. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 460, 748-52.
- Ramsay, M., Crowther, N., Tambo, E., Agongo, G., Baloyi, V., Dikotope, S., Gómez-Olivé, X., Jaff, N., Sorgho, H., Wagner, R., Khayeka-Wandabwa, C., Choudhury, A., Hazelhurst, S., Kahn, K., Lombard, Z., Mukomana, F., Soo, C., Soodyall, H., Wade, A., Afolabi, S., Agorinya, I., Amenga-Etego, L., Ali, S. A., Bognini, J. D., Boua, R. P., Debpuur, C., Diallo, S., Fato, E., Kazienga, A., Konkobo, S. Z., Kouraogo, P. M., Mashinya, F., Micklesfield, L., Nakanabo-Diallo, S., Njamwea, B., Nonterah, E., Ouedraogo, S., Pillay, V., Somande, A. M., Tindana, P., Twine, R., Alberts, M., Kyobutungi, C., Norris, S. A., Oduro, A. R., Tinto, H., Tollman, S. & Sankoh, O. 2016. H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Global health, epidemiology and genomics*, 1, e20.

- Rasouly, H. M., Groopman, E. E., Heyman-Kantor, R., Fasel, D. A., Mitrotti, A., Westland, R., Bier, L., Weng, C., Ren, Z. & Copeland, B. 2019. The burden of candidate pathogenic variants for kidney and genitourinary disorders emerging from exome sequencing. *Annals of internal medicine*, 170, 11-21.
- Rastegari, E., Ahmad, Z., Spencer, D. F. & Ismai, M. 2011. Two-dimensional profiling of proteins from *Curculigo latifolia* fruit by three different extraction protocols. *J Med Plant Res*, 5, 3719-24.
- Reeve, J., Böhmig, G. A., Eskandary, F., Einecke, G., Lefaucheur, C., Loupy, A. & Halloran, P. F. 2017. Assessing rejection-related disease in kidney transplant biopsies based on archetypal analysis of molecular phenotypes. *JCI insight*, 2.
- Reisberg, S., Iljasenko, T., Läll, K., Fischer, K. & Vilo, J. 2017. Comparing distributions of polygenic risk scores of type 2 diabetes and coronary heart disease within different populations. *PLoS One*, 12, e0179238.
- Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., Beck, M., Brusniak, M.-Y., Hengartner, M. O. & Aebersold, R. 2011. mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nature methods*, 8, 430-35.
- Renigunta, A., Renigunta, V., Saritas, T., Decher, N., Mutig, K. & Waldegger, S. 2011. Tamm-Horsfall glycoprotein interacts with renal outer medullary potassium channel ROMK2 and regulates its function. *Journal of Biological Chemistry*, 286, 2224-35.
- Rhee, C. M. & Kovesdy, C. P. 2015. Spotlight on CKD deaths—increasing mortality worldwide. *Nature Reviews Nephrology*, 11, 199-200.
- Ritchie, S. C., Lambert, S. A., Arnold, M., Teo, S. M., Lim, S., Scepanovic, P., Marten, J., Zahid, S., Chaffin, M., Liu, Y., Abraham, G., Ouwehand, W. H., Roberts, D. J., Watkins, N. A., Drew, B. G., Calkin, A. C., Di Angelantonio, E., Soranzo, N., Burgess, S., Chapman, M., Kathiresan, S., Khera, A. V., Danesh, J., Butterworth, A. S. & Inouye, M. 2021. Integrative analysis of the plasma proteome and polygenic risk of cardiometabolic diseases. *Nature metabolism*, 3, 1476-83.
- Rodrigo, R., González, J. & Paoletto, F. 2011. The role of oxidative stress in the pathophysiology of hypertension. *Hypertension Research*, 34, 431-40.
- Roscioni, S., De Zeeuw, D., Hellemons, M. E., Mischak, H., Zürbig, P., Bakker, S., Gansevoort, R. T., Reinhard, H., Persson, F. & Lajer, M. 2013. A urinary peptide biomarker set predicts worsening of albuminuria in type 2 diabetes mellitus. *Diabetologia*, 56, 259-67.
- Rossing, P., Caramori, M. L., Chan, J. C., Heerspink, H. J., Hurst, C., Khunti, K., Liew, A., Michos, E. D., Navaneethan, S. D. & Olowu, W. A. 2022. KDIGO 2022 clinical practice guideline for diabetes management in chronic kidney disease. *Kidney International*, 102, S1-S127.
- Röst, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinović, S. M., Schubert, O. T., Wolski, W., Collins, B. C., Malmström, J. & Malmström, L. 2014. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nature biotechnology*, 32, 219-23.
- Ruan, Y., Lin, Y.-F., Feng, Y.-C. A., Chen, C.-Y., Lam, M., Guo, Z., He, L., Sawa, A. & Martin, A. R. 2022. Improving polygenic prediction in ancestrally diverse populations. *Nature genetics*, 54, 573-80.
- Rüster, C. & Wolf, G. 2006. Renin-angiotensin-aldosterone system and progression of renal disease. *Journal of the American Society of Nephrology*, 17, 2985-91.
- Safar, M. E. 2018. Arterial stiffness as a risk factor for clinical hypertension. *Nature Reviews Cardiology*, 15, 97-105.
- Saito, S., Hirao, Y., Quadery, A. F., Xu, B., Elguoshy, A., Fujinaka, H., Koma, S., Yamamoto, K. & Yamamoto, T. 2019. The optimized workflow for sample preparation in LC-MS/MS-based urine proteomics. *Methods and Protocols*, 2, 46.

- Saran, R., Robinson, B., Abbott, K. C., Bragg-Gresham, J., Chen, X., Gipson, D., Gu, H., Hirth, R. A., Hutton, D. & Jin, Y. 2020. US renal data system 2019 annual data report: epidemiology of kidney disease in the United States.
- Sasser, J. M., Pollock, J. S. & Pollock, D. M. 2002. Renal endothelin in chronic angiotensin II hypertension. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 283, R243-R48.
- Schaid, D. J., Chen, W. & Larson, N. B. 2018. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature reviews. Genetics*, 19, 491-504.
- Schlatzer, D., Maahs, D. M., Chance, M. R., Dazard, J.-E., Li, X., Hazlett, F., Rewers, M. & Snell-Bergeon, J. K. 2012. Novel urinary protein biomarkers predicting the development of microalbuminuria and renal function decline in type 1 diabetes. *Diabetes care*, 35, 549-55.
- Schutte, A. E., Jafar, T. H., Poulter, N. R., Damasceno, A., Khan, N. A., Nilsson, P. M., Alsaied, J., Neupane, D., Kario, K. & Beheiry, H. 2023. Addressing global disparities in blood pressure control: perspectives of the International Society of Hypertension. *Cardiovascular research*, 119, 381-409.
- Seedat, Y., Rayner, B. & Veriava, Y. 2014. South African hypertension practice guideline 2014. *South African Journal of Diabetes and Vascular Disease*, 11, 139-44.
- Seeley, A., Prynne, J., Perera, R., Street, R., Davis, D. & Etyang, A. O. 2020. Pharmacotherapy for hypertension in Sub-Saharan Africa: a systematic review and network meta-analysis. *BMC medicine*, 18, 75.
- Sengupta, D., Botha, G., Meintjes, A., Mbiyavanga, M., Hazelhurst, S., Mulder, N., Ramsay, M. & Choudhury, A. 2023. Performance and accuracy evaluation of reference panels for genotype imputation in sub-Saharan African populations. *Cell genomics*, 3, 100332.
- Shankland, S. 2006. The podocyte's response to injury: role in proteinuria and glomerulosclerosis. *Kidney international*, 69, 2131-47.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13, 2498-504.
- Shantaram, M. 2021. Assessment of Different Urinary Protein Precipitation Methods for Protein Profiling by Mass Spectrometry. *Asian Journal of Pharmaceutics (AJP)*, 15.
- Shlush, L. I., Bercovici, S., Wasser, W. G., Yudkovsky, G., Templeton, A., Geiger, D. & Skorecki, K. 2010. Admixture mapping of end stage kidney disease genetic susceptibility using estimated mutual information ancestry informative markers. *BMC medical genomics*, 3, 1-12.
- Sievers, L. K. & Eckardt, K. U. 2019. Molecular Mechanisms of Kidney Injury and Repair in Arterial Hypertension. *Int J Mol Sci*, 20.
- Singh, S., Choudhury, A., Hazelhurst, S., Crowther, N., Boua, P., Sorgho, H., Agongo, G., Nonterah, E., Micklesfield, L., Norris, S., Kisiangani, I., Mohamed, S., Gomez-Olive, F., Tollman, S., Choma, S., Brandenburg, J. T. & Ramsay, M. 2023. Genome-wide Association Study Meta-analysis of Blood Pressure Traits and Hypertension in Sub-Saharan African Populations: An AWI-Gen Study. *Research square*.
- Sinnott-Armstrong, N., Tanigawa, Y., Amar, D., Mars, N., Benner, C., Aguirre, M., Venkataraman, G. R., Wainberg, M., Ollila, H. M., Kiiskinen, T., Havulinna, A. S., Pirruccello, J. P., Qian, J., Shcherbina, A., Rodriguez, F., Assimes, T. L., Agarwala, V., Tibshirani, R., Hastie, T., Ripatti, S., Pritchard, J. K., Daly, M. J. & Rivas, M. A. 2021. Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nature genetics*, 53, 185-94.
- Siragy, H. M. & Carey, R. M. 2010. Role of the intrarenal renin-angiotensin-aldosterone system in chronic kidney disease. *American journal of nephrology*, 31, 541-50.
- Siwy, J., Schanstra, J. P., Argiles, A., Bakker, S. J., Beige, J., Boucek, P., Brand, K., Delles, C., Duranton, F., Fernandez-Fernandez, B., Jankowski, M. L., Al Khatib, M., Kunt, T., Lajer, M.,

- Lichtinghagen, R., Lindhardt, M., Maahs, D. M., Mischak, H., Mullen, W., Navis, G., Noutsou, M., Ortiz, A., Persson, F., Petrie, J. R., Roob, J. M., Rossing, P., Ruggenenti, P., Rychlik, I., Serra, A. L., Snell-Bergeon, J., Spasovski, G., Stojceva-Taneva, O., Trillini, M., Von Der Leyen, H., Winklhofer-Roob, B. M., Zürlbig, P. & Jankowski, J. 2014. Multicentre prospective validation of a urinary peptidome-based classifier for the diagnosis of type 2 diabetic nephropathy. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*, 29, 1563-70.
- Soraru, J., Isbel, N., Wong, G., Coates, P. T., Mantha, M., Abraham, A., Juneja, R., Hsu, D., Brown, F. & Bose, B. 2020. Baseline characteristics of patients with atypical haemolytic uraemic syndrome (aHUS): the Australian cohort in a global aHUS registry. *Nephrology*, 25, 683-90.
- Stanzick, K. J., Li, Y., Schlosser, P., Gorski, M., Wuttke, M., Thomas, L. F., Rasheed, H., Rowan, B. X., Graham, S. E., Vanderweff, B. R., Patil, S. B., Robinson-Cohen, C., Gaziano, J. M., O'donnell, C. J., Willer, C. J., Hallan, S., Åsvold, B. O., Gessner, A., Hung, A. M., Pattaro, C., Köttgen, A., Stark, K. J., Heid, I. M. & Winkler, T. W. 2021. Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. *Nature communications*, 12, 4350.
- Steinbrenner, I., Yu, Z., Jin, J., T. Schultheiß, U., Kinga Kotsis, F., Grams, M., Coresh, J., Eckardt, K.-U., Sekula, P. & Chatterjee, N. 2022. MO517: A Polygenic Risk Score for Reduced EGFR is Associated With Adverse Events in a Chronic Kidney Disease Cohort—the German Chronic Kidney Disease Study. *Nephrology Dialysis Transplantation*, 37, gfac071. 48.
- Taira, M., Imamura, M., Takahashi, A., Kamatani, Y., Yamauchi, T., Araki, S.-I., Tanaka, N., Van Zuydam, N. R., Ahlqvist, E. & Toyoda, M. 2018. A variant within the FTO confers susceptibility to diabetic nephropathy in Japanese patients with type 2 diabetes. *PloS one*, 13, e0208654.
- Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., Taliun, S. a. G., Corvelo, A., Gogarten, S. M., Kang, H. M., Pitsillides, A. N., Lefaive, J., Lee, S. B., Tian, X., Browning, B. L., Das, S., Emde, A. K., Clarke, W. E., Loesch, D. P., Shetty, A. C., Blackwell, T. W., Smith, A. V., Wong, Q., Liu, X., Conomos, M. P., Bobo, D. M., Aguet, F., Albert, C., Alonso, A., Ardlie, K. G., Arking, D. E., Aslibekyan, S., Auer, P. L., Barnard, J., Barr, R. G., Barwick, L., Becker, L. C., Beer, R. L., Benjamin, E. J., Bielak, L. F., Blangero, J., Boehnke, M., Bowden, D. W., Brody, J. A., Burchard, E. G., Cade, B. E., Casella, J. F., Chalazan, B., Chasman, D. I., Chen, Y. I., Cho, M. H., Choi, S. H., Chung, M. K., Clish, C. B., Correa, A., Curran, J. E., Custer, B., Darbar, D., Daya, M., De Andrade, M., Demeo, D. L., Dutcher, S. K., Ellinor, P. T., Emery, L. S., Eng, C., Fatkin, D., Fingerlin, T., Forer, L., Fornage, M., Franceschini, N., Fuchsberger, C., Fullerton, S. M., Germer, S., Gladwin, M. T., Gottlieb, D. J., Guo, X., Hall, M. E., He, J., Heard-Costa, N. L., Heckbert, S. R., Irvin, M. R., Johnsen, J. M., Johnson, A. D., Kaplan, R., Kardia, S. L. R., Kelly, T., Kelly, S., Kenny, E. E., Kiel, D. P., Klemmer, R., Konkole, B. A., Kooperberg, C., Köttgen, A., Lange, L. A., Lasky-Su, J., Levy, D., Lin, X., Lin, K. H., Liu, C., Loos, R. J. F., et al. 2021. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, 590, 290-99.
- Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G. & Meyre, D. 2019. Benefits and limitations of genome-wide association studies. *Nature reviews. Genetics*, 20, 467-84.
- Tanca, A., Biosa, G., Pagnozzi, D., Addis, M. F. & Uzzau, S. 2013. Comparison of detergent-based sample preparation workflows for LTQ-O rbitrap analysis of the *Escherichia coli* proteome. *Proteomics*, 13, 2597-607.
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K. & Surani, M. A. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature methods*, 6, 377-82.

- Tang, X., Xiao, X., Sun, H., Zheng, S., Xiao, X., Guo, Z., Liu, X. & Sun, W. 2022. 96DRA-Urine: A high throughput sample preparation method for urinary proteome analysis. *Journal of Proteomics*, 257, 104529.
- Tarver-Carr, M. E., Powe, N. R., Eberhardt, M. S., Laveist, T. A., Kington, R. S., Coresh, J. & Brancati, F. L. 2002. Excess risk of chronic kidney disease among African-American versus white subjects in the United States: a population-based study of potential explanatory factors. *Journal of the American Society of Nephrology*, 13, 2363-70.
- Tekola-Ayele, F. & Rotimi, C. N. 2015. Translational Genomics in Low- and Middle-Income Countries: Opportunities and Challenges. *Public health genomics*, 18, 242-7.
- Teumer, A., Li, Y., Ghasemi, S., Prins, B. P., Wuttke, M., Hermle, T., Giri, A., Sieber, K. B., Qiu, C., Kirsten, H., Tin, A., Chu, A. Y., Bansal, N., Feitosa, M. F., Wang, L., Chai, J. F., Cocca, M., Fuchsberger, C., Gorski, M., Hoppmann, A., Horn, K., Li, M., Marten, J., Noce, D., Nutile, T., Sedaghat, S., Sveinbjornsson, G., Tayo, B. O., Van Der Most, P. J., Xu, Y., Yu, Z., Gerstner, L., Ärnlöv, J., Bakker, S. J. L., Baptista, D., Biggs, M. L., Boerwinkle, E., Brenner, H., Burkhardt, R., Carroll, R. J., Chee, M. L., Chee, M. L., Chen, M., Cheng, C. Y., Cook, J. P., Coresh, J., Corre, T., Danesh, J., De Borst, M. H., De Grandi, A., De Mutsert, R., De Vries, A. P. J., Degenhardt, F., Dittrich, K., Divers, J., Eckardt, K. U., Ehret, G., Endlich, K., Felix, J. F., Franco, O. H., Franke, A., Freedman, B. I., Freitag-Wolf, S., Gansevoort, R. T., Giedraitis, V., Gögele, M., Grundner-Culemann, F., Gudbjartsson, D. F., Gudnason, V., Hamet, P., Harris, T. B., Hicks, A. A., Holm, H., Foo, V. H. X., Hwang, S. J., Ikram, M. A., Ingelsson, E., Jaddoe, V. W. V., Jakobsdottir, J., Josyula, N. S., Jung, B., Kähönen, M., Khor, C. C., Kiess, W., Koenig, W., Körner, A., Kovacs, P., Kramer, H., Krämer, B. K., Kronenberg, F., Lange, L. A., Langefeld, C. D., Lee, J. J., Lehtimäki, T., Lieb, W., Lim, S. C., Lind, L., Lindgren, C. M., Liu, J., Loeffler, M., et al. 2019. Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nature communications*, 10, 4130.
- Thomas, S. N., French, D., Jannetto, P. J., Rappold, B. A. & Clarke, W. A. 2022. Liquid chromatography–tandem mass spectrometry for clinical diagnostics. *Nature Reviews Methods Primers*, 2, 96.
- Tin, A. & Köttgen, A. 2020. Genome-Wide Association Studies of CKD and Related Traits. *Clinical journal of the American Society of Nephrology : CJASN*, 15, 1643-56.
- Tofte, N., Lindhardt, M., Adamova, K., Bakker, S. J., Beige, J., Beulens, J. W., Birkenfeld, A. L., Currie, G., Delles, C. & Dimos, I. 2020. Early detection of diabetic kidney disease by urinary proteomics and subsequent intervention with spironolactone to delay progression (PRIORITY): a prospective observational study and embedded randomised placebo-controlled trial. *The Lancet Diabetes & Endocrinology*.
- Torffvit, O. & Agardh, C.-D. 1993. Tubular secretion of Tamm-Horsfall protein is decreased in type 1 (insulin-dependent) diabetic patients with diabetic nephropathy. *Nephron*, 65, 227-31.
- Torffvit, O., Agardh, C.-D. & Thulin, T. 1999. A study of Tamm-Horsfall protein excretion in hypertensive patients and type 1 diabetic patients. *Scandinavian journal of urology and nephrology*, 33, 187-91.
- Torun, F. M., Virreira Winter, S., Doll, S., Riese, F. M., Vorobyev, A., Mueller-Reif, J. B., Geyer, P. E. & Strauss, M. T. 2022. Transparent exploration of machine learning for biomarker discovery from proteomics and omics data. *Journal of Proteome Research*, 22, 359-67.
- Trudu, M., Janas, S., Lanzani, C., Debaix, H., Schaeffer, C., Ikehata, M., Citterio, L., Demaretz, S., Trevisani, F. & Ristagno, G. 2013. Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression. *Nature medicine*, 19, 1655-60.

- Trudu, M., Schaeffer, C., Riba, M., Ikehata, M., Brambilla, P., Messa, P., Martinelli-Boneschi, F., Rastaldi, M. P. & Rampoldi, L. 2017. Early involvement of cellular stress and inflammatory signals in the pathogenesis of tubulointerstitial kidney disease due to UMOD mutations. *Scientific reports*, 7, 7383.
- Tzur, S., Rosset, S., Shemer, R., Yudkovsky, G., Selig, S., Tarekegn, A., Bekele, E., Bradman, N., Wasser, W. G. & Behar, D. M. 2010. Missense mutations in the APOL1 gene are highly associated with end stage kidney disease risk previously attributed to the MYH9 gene. *Human genetics*, 128, 345-50.
- Ullah, M. M. & Basile, D. P. 2019. Role of Renal Hypoxia in the Progression From Acute Kidney Injury to Chronic Kidney Disease. *Seminars in nephrology*, 39, 567-80.
- Unger, T., Borghi, C., Charchar, F., Khan, N. A., Poulter, N. R., Prabhakaran, D., Ramirez, A., Schlaich, M., Stergiou, G. S., Tomaszewski, M., Wainford, R. D., Williams, B. & Schutte, A. E. 2020. 2020 International Society of Hypertension global hypertension practice guidelines. *Journal of hypertension*, 38, 982-1004.
- Vaidya, S. R. & Aeddula, N. R. 2023. Chronic Kidney Disease. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; Available from: <https://www.ncbi.nlm.nih.gov/books/NBK535404/>.
- Van Der Ven, A. T., Connaughton, D. M., Ityel, H., Mann, N., Nakayama, M., Chen, J., Vivante, A., Hwang, D.-Y., Schulz, J. & Braun, D. A. 2018. Whole-exome sequencing identifies causative mutations in families with congenital anomalies of the kidney and urinary tract. *Journal of the American Society of Nephrology: JASN*, 29, 2348.
- Vasquez-Rios, G., De Cos, M. & Campbell, K. N. 2023. Novel Therapies in APOL1-Mediated Kidney Disease: From Molecular Pathways to Therapeutic Options. *Kidney international reports*, 8, 2226-34.
- Vaziri, N. D., Ni, Z., Oveisi, F., Liang, K. & Pandian, R. 2002. Enhanced nitric oxide inactivation and protein nitration by reactive oxygen species in renal insufficiency. *Hypertension*, 39, 135-41.
- Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. 2012. Five years of GWAS discovery. *American journal of human genetics*, 90, 7-24.
- Vy, H. M. T., Coca, S. G., Sawant, A., Sakhuja, A., Gutierrez, O. M., Cooper, R., Loos, R. J., Horowitz, C. R., Do, R. & Nadkarni, G. N. 2023. Genome-Wide Polygenic Risk Score for CKD in Individuals with APOL1 High-Risk Genotypes. *Clinical Journal of the American Society of Nephrology*, 10.2215.
- Wang, B., Li, Z. L., Zhang, Y. L., Wen, Y., Gao, Y. M. & Liu, B. C. 2022. Hypoxia and chronic kidney disease. *EBioMedicine*, 77, 103942.
- Wang, H., D'ambrosio, M. A., Garvin, J. L., Ren, Y. & Carretero, O. A. 2013. Connecting tubule glomerular feedback in hypertension. *Hypertension*, 62, 738-45.
- Wang, Y., Guo, J., Ni, G., Yang, J., Visscher, P. M. & Yengo, L. 2020. Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nature communications*, 11, 3865.
- Wang, Y., Namba, S., Lopera, E., Kerminen, S., Tsuo, K., Läll, K., Kanai, M., Zhou, W., Wu, K.-H. & Favé, M.-J. 2023. Global Biobank analyses provide lessons for developing polygenic risk scores across diverse cohorts. *Cell genomics*, 3.
- Wanner, C. & Tonelli, M. 2014. KDIGO Clinical Practice Guideline for Lipid Management in CKD: summary of recommendation statements and clinical approach to the patient. *Kidney international*, 85, 1303-09.
- Webster, A. C., Nagler, E. V., Morton, R. L. & Masson, P. 2017. Chronic kidney disease. *The lancet*, 389, 1238-52.

- Wei, B., Yu, M., Yao, J., Jiang, M., An, J., Yang, J., Lin, J., Zhao, Y. & Zhu, Y. 2021. Multidimensional Analyses of Tumor Immune Microenvironment Reveal the Possible Rationality of Immunotherapy and Identify High Immunotherapy Response Subtypes for Renal Papillary Cell Carcinoma. *Frontiers in immunology*, 12, 657951.
- Weissbrod, O., Kanai, M., Shi, H., Gazal, S., Peyrot, W. J., Khera, A. V., Okada, Y., Martin, A. R., Finucane, H. K. & Price, A. L. 2022. Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nature genetics*, 54, 450-58.
- Wen, L., Li, G., Huang, T., Geng, W., Pei, H., Yang, J., Zhu, M., Zhang, P., Hou, R., Tian, G., Su, W., Chen, J., Zhang, D., Zhu, P., Zhang, W., Zhang, X., Zhang, N., Zhao, Y., Cao, X., Peng, G., Ren, X., Jiang, N., Tian, C. & Chen, Z. J. 2022. Single-cell technologies: From research to application. *Innovation (Cambridge (Mass.))*, 3, 100342.
- Whelton, P. K. & Carey, R. M. 2018. The 2017 American College of Cardiology/American Heart Association clinical practice guideline for high blood pressure in adults. *JAMA cardiology*, 3, 352-53.
- Willer, C. J., Li, Y. & Abecasis, G. R. 2010. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)*, 26, 2190-1.
- Winkler, T. W., Day, F. R., Croteau-Chonka, D. C., Wood, A. R., Locke, A. E., Mägi, R., Ferreira, T., Fall, T., Graff, M., Justice, A. E., Luan, J., Gustafsson, S., Randall, J. C., Vedantam, S., Workalemahu, T., Kilpeläinen, T. O., Scherag, A., Esko, T., Kutalik, Z., Heid, I. M. & Loos, R. J. 2014. Quality control and conduct of genome-wide association meta-analyses. *Nature protocols*, 9, 1192-212.
- Wu, I. W., Tsai, T. H., Lo, C. J., Chou, Y. J., Yeh, C. H., Chan, Y. H., Chen, J. H., Hsu, P. W., Pan, H. C., Hsu, H. J., Chen, C. Y., Lee, C. C., Shyu, Y. C., Lin, C. L., Cheng, M. L., Lai, C. C., Sytwu, H. K. & Tsai, T. F. 2022. Discovering a trans-omics biomarker signature that predisposes high risk diabetic patients to diabetic kidney disease. *NPJ digital medicine*, 5, 166.
- Wuttke, M., Li, Y., Li, M., Sieber, K. B., Feitosa, M. F., Gorski, M., Tin, A., Wang, L., Chu, A. Y., Hoppmann, A., Kirsten, H., Giri, A., Chai, J. F., Sveinbjornsson, G., Tayo, B. O., Nutile, T., Fuchsberger, C., Marten, J., Cocca, M., Ghasemi, S., Xu, Y., Horn, K., Noce, D., Van Der Most, P. J., Sedaghat, S., Yu, Z., Akiyama, M., Afaq, S., Ahluwalia, T. S., Almgren, P., Amin, N., Ärnlöv, J., Bakker, S. J. L., Bansal, N., Baptista, D., Bergmann, S., Biggs, M. L., Biino, G., Boehnke, M., Boerwinkle, E., Boissel, M., Bottinger, E. P., Boutin, T. S., Brenner, H., Brumat, M., Burkhardt, R., Butterworth, A. S., Campana, E., Campbell, A., Campbell, H., Canouil, M., Carroll, R. J., Catamo, E., Chambers, J. C., Chee, M. L., Chee, M. L., Chen, X., Cheng, C. Y., Cheng, Y., Christensen, K., Cifkova, R., Ciullo, M., Concas, M. P., Cook, J. P., Coresh, J., Corre, T., Sala, C. F., Cusi, D., Danesh, J., Daw, E. W., De Borst, M. H., De Grandi, A., De Mutsert, R., De Vries, A. P. J., Degenhardt, F., Delgado, G., Demirkan, A., Di Angelantonio, E., Dittrich, K., Divers, J., Dorajoo, R., Eckardt, K. U., Ehret, G., Elliott, P., Endlich, K., Evans, M. K., Felix, J. F., Foo, V. H. X., Franco, O. H., Franke, A., Freedman, B. I., Freitag-Wolf, S., Friedlander, Y., Froguel, P., Gansevoort, R. T., Gao, H., Gasparini, P., Gaziano, J. M., Giedraitis, V., Gieger, C., et al. 2019. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nature genetics*, 51, 957-72.
- Xiang, R., Kelemen, M., Xu, Y., Harris, L. W., Parkinson, H., Inouye, M. & Lambert, S. A. 2024. Recent advances in polygenic scores: translation, equitability, methods and FAIR tools. *Genome Medicine*, 16, 33.
- Xie, D., Ma, T., Cui, H., Li, J., Zhang, A., Sheng, Z. & Xie, Y. 2023. Global burden and influencing factors of chronic kidney disease due to type 2 diabetes in adults aged 20–59 years, 1990–2019. *Scientific Reports*, 13, 20234.
- Yang, J. & He, W. 2019. *Chronic kidney disease: Diagnosis and treatment*, Springer.

- Yu, H., Tai, Q., Yang, C., Gao, M. & Zhang, X. 2023. Technological development of multidimensional liquid chromatography-mass spectrometry in proteome research. *Journal of Chromatography A*, 1700, 464048.
- Yu, Z., Jin, J., Tin, A., Kötting, A., Yu, B., Chen, J., Surapaneni, A., Zhou, L., Ballantyne, C. M., Hoogeveen, R. C., Arking, D. E., Chatterjee, N., Grams, M. E. & Coresh, J. 2021. Polygenic Risk Scores for Kidney Function and Their Associations with Circulating Proteome, and Incident Kidney Diseases. *Journal of the American Society of Nephrology : JASN*, 32, 3161-73.
- Yuan, Q., Tang, B. & Zhang, C. 2022. Signaling pathways of chronic kidney diseases, implications for therapeutics. *Signal transduction and targeted therapy*, 7, 182.
- Yusuf, A. A., Govender, M. A., Brandenburg, J.-T. & Winkler, C. A. 2021. Kidney disease and APOL1. *Human molecular genetics*, 30, R129-R37.
- Zandi-Nejad, K., Luyckx, V. A. & Brenner, B. M. 2006. Adult hypertension and kidney disease: the role of fetal programming. *Hypertension*, 47, 502-08.
- Zeng, X., Hossain, D., Bostwick, D. G., Herrera, G. A. & Zhang, P. L. 2014. Urinary β 2-microglobulin is a good indicator of proximal tubule injury: a correlative study with renal biopsies. *Journal of biomarkers*, 2014.
- Zhang, B., Chambers, M. C. & Tabb, D. L. 2007. Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *Journal of proteome research*, 6, 3549-57.
- Zhang, J., Fedick, A., Wasserman, S., Zhao, G., Edelmann, L., Bottinger, E. P., Kornreich, R. & Scott, S. A. 2016. Analytical Validation of a Personalized Medicine APOL1 Genotyping Assay for Nondiabetic Chronic Kidney Disease Risk Assessment. *The Journal of molecular diagnostics : JMD*, 18, 260-6.
- Zhao, W.-B. & Alberto, D. L. P. S. M. 2020. Serum apolipoprotein B/apolipoprotein A1 ratio is associated with the progression of diabetic kidney disease to renal replacement therapy. *International Urology and Nephrology*, 52, 1923-28.
- Zhou, B., Carrillo-Larco, R. M., Danaei, G., Riley, L. M., Paciorek, C. J., Stevens, G. A., Gregg, E. W., Bennett, J. E., Solomon, B. & Singleton, R. K. 2021. Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. *The Lancet*, 398, 957-80.
- Zhou, X.-J., Zhong, X.-H. & Duan, L.-X. 2023. Integration of artificial intelligence and multi-omics in kidney diseases. *Fundamental Research*, 3, 126-48.
- Zoungas, S. & De Boer, I. H. 2021. SGLT2 inhibitors in diabetic kidney disease. *Clinical Journal of the American Society of Nephrology: CJASN*, 16, 631.

APPENDICES

APPENDIX A

A.1. Plagiarism report



PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I MELANIE GOVENDER (Student number: 1487683) am a student registered for Doctor of Philosophy (PhD) in Human Genetics in the year 2024. I hereby declare the following:

- ☐ I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- ☐ I confirm that ALL the work submitted for assessment for the above course is my own unaided work except where I have explicitly indicated otherwise.
- ☐ I have followed the required conventions in referencing the thoughts and ideas of others.
- ☐ I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

A handwritten signature in black ink, appearing to read 'M Govender', is written over a horizontal line.

Signature: _____ Date: 19 February 2024

ORIGINALITY REPORT

47%

SIMILARITY INDEX

43%

INTERNET SOURCES

39%

PUBLICATIONS

4%

STUDENT PAPERS

PRIMARY SOURCES

1	www.frontiersin.org Internet Source	12%
2	clinicalproteomicsjournal.biomedcentral.com Internet Source	10%
3	assets.researchsquare.com Internet Source	5%
4	www.researchsquare.com Internet Source	4%
5	www.ncbi.nlm.nih.gov Internet Source	3%
6	Submitted to University of Witwatersrand Student Paper	2%
7	www.medrxiv.org Internet Source	1%
8	Melanie A. Govender, Stoyan H. Stoychev, Jean-Tristan Brandenburg, Michèle Ramsay, June Fabian, Ireshyn S. Govender. "Proteomic insights into the pathophysiology of hypertension-associated albuminuria: Pilot	1%



University of the Witwatersrand
Faculty of Health Sciences

25 March 2024

RE: Turnitin Originality report

The thesis of PhD candidate, Melanie Govender (student number: 1487683) was submitted to Turnitin to check for plagiarism. The originality report from Turnitin provided a similarity score of 47% and this is mostly attributed to the two published papers (CHAPTERS 2 and 4).

All sources for the published papers add up to 35%:

- Sources 1 and 5 are two versions of the review paper (CHAPTER 2) on Frontiers in Genetics journal and on NCBI
- Sources 2, 3, 4 and 7 refer to the published proteomics paper (CHAPTER 4)
 - Please note that the proteomics paper was submitted to a pre-print server prior to publication

The remainder of the similarities are commonly used phrases and constitute <1% for most of the sources referred to. We have examined the originality report and agree that this is not an indication of plagiarism.

Kind Regards,

Professor Michèle Ramsay
Supervisor
SBIMB, University of the
Witwatersrand
Tel: 011 717 6635
Email:
Michele.Ramsay@wits.ac.za

Dr. June Fabian
Supervisor
Wits Donald Gordon
Medical Centre
Tel: 011 356 6395
Email:
june.fabian@mweb.co.za

Dr. Jean-Tristan Brandenburg
Supervisor
SBIMB, University of the
Witwatersrand
Tel: 011 717 6630
Email: jean-tristan.brandenburg@wits.ac.za

A.2. Agreement of co-authors

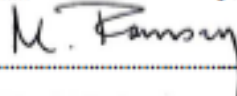
Declaration: Student's contribution to articles and agreement of co-authors

I, Melanie Govender, student number 1487683, declare that this Thesis is my own work and that I contributed adequately towards research findings published in the article(s) stated below, which are included in my Thesis.



Signature of Student Date 25 February 2024

Name of Supervisor Michèle Ramsay

Signature of Supervisor  Date ... 26 February 2024

Name of Supervisor June Fabian

Signature of Supervisor  Date 27th February 2024

Name of Supervisor Jean-Tristan Brandenburg

Signature of Supervisor  Date

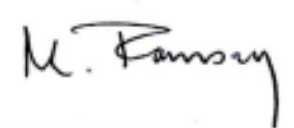
Agreement by co-authors: By signing this declaration, the co-authors listed below agree to the use of the articles by Melanie Govender as part of her Thesis.

Article 1: Title: The Use of 'Omics for Diagnosing and Predicting Progression of Chronic Kidney Disease: A Scoping Review

Status: Published (8 November 2021)

Journal: *Frontiers in Genetics*

Details: DOI: [10.3389/fgene.2021.682929](https://doi.org/10.3389/fgene.2021.682929)

Authors	Name	Title	Signature	Date
Primary	Melanie Govender	Ms		25 Feb. 24
Co-author	Jean-Tristan Brandenburg	Dr		27 th February 2024
Co-author	June Fabian	Dr		27 th February 2024
Senior	Michèle Ramsay	Prof		26 Feb 2024

Comments by Supervisors

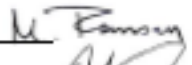
We confirm that Melanie did the bulk of the work for the review. She collected the information and wrote the first draft which was refined under our guidance.

Article 2: Title: Proteomic insights into the pathophysiology of hypertension-associated albuminuria: Pilot study in a South African cohort

Status: Published (24 February 2024)

Journal: *Clinical Proteomics*

Details: DOI: [10.1186/s12014-024-09458-9](https://doi.org/10.1186/s12014-024-09458-9)

Authors	Name	Title	Signature	Date
Primary	Melanie Govender			25 Feb. 24
Co-author	Stoyan H. Stoychev			25 Feb. 24
Co-author	Jean-Tristan Brandenburg			27 th February 2024
Co-author	Michèle Ramsay	Prof		26 Feb 2024
Senior	June Fabian			27 th February 2024
Senior	Ireshyn S. Govender			25 Feb. 24

Comments by Supervisors

We confirm that this is the work of Melanie as part of her PhD thesis. She collected the data, performed the analysis under to guidance of the co-authors, wrote the first draft and did the extensive edits and refinements under guidance.

A.3. PhD Article Outputs



The Use of 'Omics for Diagnosing and Predicting Progression of Chronic Kidney Disease: A Scoping Review

Melanie A. Govender^{1,2*}, Jean-Tristan Brandenburg², June Fabian³ and Michèle Ramsay^{1,2*}

¹Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, ²Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, ³Wits Donald Gordon Medical Centre, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

OPEN ACCESS

Edited by:

Martha J. Kotsis,
Stellenbosch University, South Africa

Reviewed by:

Romy Lesmana,
Universitas Padjadjaran, Indonesia
Jianxiang Shi,
Zhengzhou University, China

*Correspondence:

Melanie A. Govender
melanieangovender@gmail.com
Michèle Ramsay
Michele.Ramsay@wits.ac.za

Specialty section:

This article was submitted to
Human and Medical Genetics,
a section of the journal
Frontiers in Genetics

Received: 19 March 2021

Accepted: 18 October 2021

Published: 08 November 2021

Globally, chronic kidney disease (CKD) contributes substantial morbidity and mortality. Recently, various 'omics platforms have provided insight into the molecular basis of kidney dysfunction. This scoping review is a synthesis of the current literature on the use of different 'omics platforms to identify biomarkers that could be used to detect early-stage CKD, predict disease progression, and identify pathways leading to CKD. This review includes 123 articles published from January 2007 to May 2021, following a structured selection process. The most common type of 'omic platform was proteomics, appearing in 55 of the studies and two of these included a metabolomics component. Most studies ($n = 91$) reported on CKD associated with diabetes mellitus. Thirteen studies that provided information on the biomarkers associated with CKD and explored potential pathways involved in CKD are discussed. The biomarkers that are associated with risk or early detection of CKD are SNPs in the *MYH9/APOL1* and *UMOD* genes, the proteomic CKD273 biomarker panel and metabolite pantothenic acid. Pantothenic acid and the CKD273 biomarker panel were also involved in predicting CKD progression. Retinoic acid pathway genes, *UMOD*, and pantothenic acid provided insight into potential pathways leading to CKD. The biomarkers were mainly used to detect CKD and predict progression in high-income, European ancestry populations, highlighting the need for representative 'omics research in other populations with disparate socio-economic strata, including Africans, since disease etiologies may differ across ethnic groups. To assess the transferability of findings, it is essential to do research in diverse populations.

Keywords: 'omics, biomarkers, early detection, diagnosis, Sub-Saharan Africa, diabetes, chronic kidney disease, hypertension

INTRODUCTION

(Govender et al., 2021). Published in *Frontiers in Genetics*:
<https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2021.682929/full>

RESEARCH

Open Access



Proteomic insights into the pathophysiology of hypertension-associated albuminuria: Pilot study in a South African cohort

Melanie A. Govender^{1,2*}, Stoyan H. Stoychev^{3,4}, Jean-Tristan Brandenburg^{2,5}, Michèle Ramsay^{1,2}, June Fabian^{6,7†} and Ireshyn S. Govender^{3,4†}

Abstract

Background Hypertension is an important public health priority with a high prevalence in Africa. It is also an independent risk factor for kidney outcomes. We aimed to identify potential proteins and pathways involved in hypertension-associated albuminuria by assessing urinary proteomic profiles in black South African participants with combined hypertension and albuminuria compared to those who have neither condition.

Methods The study included 24 South African cases with both hypertension and albuminuria and 49 control participants who had neither condition. Protein was extracted from urine samples and analysed using ultra-high-performance liquid chromatography coupled with mass spectrometry. Data were generated using data-independent acquisition (DIA) and processed using Spectronaut™ 15. Statistical and functional data annotation were performed on Perseus and Cytoscape to identify and annotate differentially abundant proteins. Machine learning was applied to the dataset using the OmicLearn platform.

Results Overall, a mean of 1,225 and 915 proteins were quantified in the control and case groups, respectively. Three hundred and thirty-two differentially abundant proteins were constructed into a network. Pathways associated with these differentially abundant proteins included the immune system (q-value [false discovery rate] = 1.4×10^{-45}), innate immune system (q = 1.1×10^{-32}), extracellular matrix (ECM) organisation (q = 0.03) and activation of matrix metalloproteinases (q = 0.04). Proteins with high disease scores (76–100% confidence) for both hypertension and chronic kidney disease included angiotensinogen (AGT), albumin (ALB), apolipoprotein L1 (APOL1), and uromodulin (UMOD). A machine learning approach was able to identify a set of 20 proteins, differentiating between cases and controls.

Conclusions The urinary proteomic data combined with the machine learning approach was able to classify disease status and identify proteins and pathways associated with hypertension-associated albuminuria.

[†]June Fabian and Ireshyn S. Govender contributed equally to this work.

*Correspondence:
Melanie A. Govender
Melanieanngovender@gmail.com
Ireshyn S. Govender
IGovender@csir.co.za

A.4. Ethics Approval Certificate

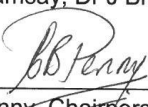
UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG



R14/49 Miss Melanie Govender

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M200331

NAME: Miss Melanie Govender
(Principal Investigator)
DEPARTMENT: Sydney Brenner Institute for Molecular Bioscience (SBIMB)
Human Genetics
PROJECT TITLE: Determining the risk profile for chronic kidney disease (CKD)
in rural South Africans using genetic risk scores and protein
markers
DATE CONSIDERED: 27/03/2020
DECISION: Approved unconditionally
CONDITIONS:
SUPERVISOR: Prof M Ramsay, Dr J Brandenburg and Dr J Fabian
APPROVED BY: 
Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL: 29/05/2020

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report.** The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in **March** and will therefore be due in the month of **March** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).


Principal Investigator Signature

31 May 2020

Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

APPENDIX B

B.1. List of 123 full-text papers assessed and included in the review

1. Gil, R. B. et al. Increased urinary osmolyte excretion indicates chronic kidney disease severity and progression rate. *Nephrology Dialysis Transplantation* 33, 2156-2164 (2018).
2. Hallan, S. et al. Metabolomics and gene expression analysis reveal down-regulation of the citric acid (TCA) cycle in non-diabetic CKD patients. *EBioMedicine* 26, 68-77 (2017).
3. Fan, Y. et al. Erratum. Comparison of Kidney Transcriptomic Profiles of Early and Advanced Diabetic Nephropathy Reveals Potential New Mechanisms for Disease Progression. *Diabetes* 2019; 68: 2301–2314. *Diabetes* 69, 797-797 (2020).
4. Perez-Hernandez, J. et al. Urinary-and Plasma-Derived Exosomes Reveal a Distinct MicroRNA Signature Associated With Albuminuria in Hypertension. *Hypertension* 77, 960-971 (2021).
5. Monteiro, M. B. et al. Urinary Sediment Transcriptomic and Longitudinal Data to Investigate Renal Function Decline in Type 1 Diabetes. *Frontiers in Endocrinology* 11, 238 (2020).
6. Roux, M. et al. Plasma levels of hsa-miR-152-3p are associated with diabetic nephropathy in patients with type 2 diabetes. *Nephrology Dialysis Transplantation* 33, 2201-2207 (2018).
7. Nair, V. et al. A molecular morphometric approach to diabetic kidney disease can link structure to function and outcome. *Kidney international* 93, 439-449 (2018).
8. Li, H. et al. Epigenetic regulation of RCAN1 expression in kidney disease and its role in podocyte injury. *Kidney international* 94, 1160-1176 (2018).
9. Liu, J.-J. et al. Profiling of plasma metabolites suggests altered mitochondrial fuel usage and remodeling of sphingolipid metabolism in individuals with type 2 diabetes and kidney disease. *Kidney international reports* 2, 470-480 (2017).
10. Gordin, D. et al. Characterization of glycolytic enzymes and pyruvate kinase M2 in type 1 and 2 diabetic nephropathy. *Diabetes Care* 42, 1263-1273 (2019).
11. Kammer, M. et al. Integrative analysis of prognostic biomarkers derived from multiomics panels helps discrimination of chronic kidney disease trajectories in people with type 2 diabetes. *Kidney International* 96, 1381-1388 (2019).
12. Smith, A. et al. Detecting Proteomic Indicators to Distinguish Diabetic Nephropathy from Hypertensive Nephrosclerosis by Integrating Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging with High-Mass Accuracy Mass Spectrometry. *Kidney and Blood Pressure Research* 45, 233-248 (2020).
13. Baldan-Martin, M. et al. Plasma Molecular Signatures in Hypertensive Patients With Renin–Angiotensin System Suppression: New Predictors of Renal Damage and De Novo Albuminuria Indicators. *Hypertension* 68, 157-166 (2016).
14. Edfors, R. et al. Use of proteomics to identify biomarkers associated with chronic kidney disease and long-term outcomes in patients with myocardial infarction. *Journal of Internal Medicine* 288, 581-592 (2020).
15. Øvrehus, M. A., Zürlbig, P., Vikse, B. E. & Hallan, S. I. Urinary proteomics in chronic kidney disease: diagnosis and risk of progression beyond albuminuria. *Clinical proteomics* 12, 1-9 (2015).
16. Glazyrin, Y. E. et al. Proteomics-based machine learning approach as an alternative to conventional biomarkers for differential diagnosis of chronic kidney diseases. *International Journal of Molecular Sciences* 21, 4802 (2020).
17. Dihazi, H. et al. Characterization of diabetic nephropathy by urinary proteomic analysis: identification of a processed ubiquitin form as a differentially excreted protein in diabetic nephropathy patients. *Clinical chemistry* 53, 1636-1645 (2007).

18. Rajesh, K., Dheeba, B., Sampathkumar, P. & Sivakumar, R. Proteomic analysis of human blood and urine in diabetic nephropathy.
19. Carlsson, A. C. et al. Use of a proximity extension assay proteomics chip to discover new biomarkers associated with albuminuria. *European Journal of Preventive Cardiology* 24, 340-348 (2017).
20. Verbeke, F. et al. The urinary proteomics classifier chronic kidney disease 273 predicts cardiovascular outcome in patients with chronic kidney disease. *Nephrology Dialysis Transplantation* 36, 811-818 (2021).
21. Subasi, E. et al. A classification model to predict the rate of decline of kidney function. *Frontiers in Medicine* 4, 97 (2017).
22. Ngo, D. et al. Circulating testican-2 is a podocyte-derived marker of kidney health. *Proceedings of the National Academy of Sciences* 117, 25026-25035 (2020).
23. Gu, Y.-M. et al. The urinary proteome as correlate and predictor of renal function in a population study. *Nephrology Dialysis Transplantation* 29, 2260-2268 (2014).
24. So, E. J., Kim, H. J. & Kim, C. W. Proteomic analysis of human proximal tubular cells exposed to high glucose concentrations. *PROTEOMICS–Clinical Applications* 2, 1118-1126 (2008).
25. Pontillo, C. et al. Prediction of chronic kidney disease stage 3 by CKD273, a urinary proteomic biomarker. *Kidney international reports* 2, 1066-1075 (2017).
26. Caseiro, A. et al. Pursuing type 1 diabetes mellitus and related complications through urinary proteomics. *Translational Research* 163, 188-199 (2014).
27. Merchant, M. L. et al. Plasma kininogen and kininogen fragments are biomarkers of progressive renal decline in type 1 diabetes. *Kidney international* 83, 1177-1184 (2013).
28. Millionini, R. et al. Abnormal cytoskeletal protein expression in cultured skin fibroblasts from type 1 diabetes mellitus patients with nephropathy: A proteomic approach. *PROTEOMICS–Clinical Applications* 2, 492-503 (2008).
29. Overgaard, A. J. et al. Quantitative iTRAQ-based proteomic identification of candidate biomarkers for diabetic nephropathy in plasma of type 1 diabetic patients. *Clinical proteomics* 6, 105-114 (2010).
30. Overgaard, A. J. et al. Plasma proteome analysis of patients with type 1 diabetes with diabetic nephropathy. *Proteome science* 8, 4 (2010).
31. Schlatter, D. et al. Novel urinary protein biomarkers predicting the development of microalbuminuria and renal function decline in type 1 diabetes. *Diabetes care* 35, 549-555 (2012).
32. Vitova, L. et al. Early urinary biomarkers of diabetic nephropathy in type 1 diabetes mellitus show involvement of kallikrein-kinin system. *BMC nephrology* 18, 1-10 (2017).
33. Gianazza, E. et al. Different expression of fibrinopeptide A and related fragments in serum of type 1 diabetic patients with nephropathy. *Journal of proteomics* 73, 593-601 (2010).
34. Wu, J., Chen, Y.-D., Yu, J.-K., Shi, X.-L. & Gu, W. Analysis of urinary proteomic patterns for type 2 diabetic nephropathy by ProteinChip. *diabetes research and clinical practice* 91, 213-219 (2011).
35. Zhang, S., Li, X., Luo, H., Fang, Z.-Z. & Ai, H. Role of aromatic amino acids in pathogenesis of diabetic nephropathy in Chinese patients with type 2 diabetes. *Journal of Diabetes and its Complications* 34, 107667 (2020).
36. Lu, H., Deng, S., Zheng, M. & Hu, K. iTRAQ plasma proteomics analysis for candidate biomarkers of type 2 incipient diabetic nephropathy. *Clinical proteomics* 16, 1-11 (2019).
37. Lindhardt, M. et al. Urinary proteomics predict onset of microalbuminuria in normoalbuminuric type 2 diabetic patients, a sub-study of the DIRECT-Protect 2 study. *Nephrology Dialysis Transplantation* 32, 1866-1873 (2017).
38. Guillén-Gómez, E. et al. Urinary proteome analysis identified neprilysin and VCAM as proteins involved in diabetic nephropathy. *Journal of diabetes research* 2018 (2018).
39. Jiang, H. et al. Identification of urinary soluble E-cadherin as a novel biomarker for diabetic nephropathy. *Diabetes/metabolism research and reviews* 25, 232-241 (2009).
40. Lim, S. et al. Adipocytokine zinc α 2 glycoprotein (ZAG) as a novel urinary biomarker for normo-albuminuric diabetic nephropathy. *Diabetic Medicine* 29, 945-949 (2012).

41. Yang, J.-K. et al. Urine proteome specific for eye damage can predict kidney damage in patients with type 2 diabetes: a case-control and a 5.3-year prospective cohort study. *Diabetes care* 40, 253-260 (2017).
42. Yang, Y. et al. Predicting diabetic nephropathy by serum proteomic profiling in patients with type 2 diabetes. *Wiener klinische Wochenschrift* 127, 669-674 (2015).
43. Alkhalaf, A. et al. Multicentric validation of proteomic biomarkers in urine specific for diabetic nephropathy. *PloS one* 5, e13421 (2010).
44. Roscioni, S. et al. A urinary peptide biomarker set predicts worsening of albuminuria in type 2 diabetes mellitus. *Diabetologia* 56, 259-267 (2013).
45. Ahn, J. M., Kim, B. G., Yu, M. H., Lee, I. K. & Cho, J. Y. Identification of diabetic nephropathy-selective proteins in human plasma by multi-lectin affinity chromatography and LC-MS/MS. *PROTEOMICS—Clinical Applications* 4, 644-653 (2010).
46. Rao, P. V. et al. Proteomic identification of urinary biomarkers of diabetic nephropathy. *Diabetes care* 30, 629-637 (2007).
47. Marikanty, R. et al. Identification of urinary proteins potentially associated with diabetic kidney disease. *Indian journal of nephrology* 26, 434 (2016).
48. Yeh, S.-H. et al. Differentiation of type 2 diabetes mellitus with different complications by proteomic analysis of plasma low abundance proteins. *Journal of Diabetes & Metabolic Disorders* 15, 1-7 (2015).
49. Tofte, N. et al. Characteristics of high-and low-risk individuals in the PRIORITY study: urinary proteomics and mineralocorticoid receptor antagonism for prevention of diabetic nephropathy in Type 2 diabetes. *Diabetic Medicine* 35, 1375-1382 (2018).
50. Currie, G. E. et al. Urinary proteomics for prediction of mortality in patients with type 2 diabetes and microalbuminuria. *Cardiovascular diabetology* 17, 50 (2018).
51. Bellei, E. et al. Proteomic analysis of early urinary biomarkers of renal changes in type 2 diabetic patients. *PROTEOMICS—Clinical Applications* 2, 478-491 (2008).
52. Kim, M.-R., Yu, S.-A., Kim, M.-Y., Choi, K. M. & Kim, C.-W. Analysis of glycosylated serum proteins in type 2 diabetes patients with nephropathy. *Biotechnology and bioprocess engineering* 19, 83-92 (2014).
53. Petrica, L. et al. Glycosylated peptides are associated with proximal tubule dysfunction in type 2 diabetes mellitus. *International journal of clinical and experimental medicine* 8, 2516 (2015).
54. Siwy, J. et al. Multicentre prospective validation of a urinary peptidome-based classifier for the diagnosis of type 2 diabetic nephropathy. *Nephrology Dialysis Transplantation* 29, 1563-1570 (2014).
55. Hung, P.-H. et al. Proteomic identification of plasma biomarkers in type 2 diabetic nephropathy. *Journal of Integrated OMICS* 1, 151-156 (2010).
56. Bellei, E. et al. Urinary proteomics in biomarker discovery of kidney-related disorders: diabetic nephropathy and drug-induced nephrotoxicity in chronic headache. *Ejifcc* 29, 290 (2018).
57. Kaburagi, Y. et al. Urinary afamin levels are associated with the progression of diabetic nephropathy. *Diabetes research and clinical practice* 147, 37-46 (2019).
58. Papale, M. et al. Urine proteome analysis may allow noninvasive differential diagnosis of diabetic nephropathy. *Diabetes care* 33, 2409-2415 (2010).
59. Patel, D. N. & Kalia, K. Characterization of low molecular weight urinary proteins at varying time intervals in type 2 diabetes mellitus and diabetic nephropathy patients. *Diabetology & metabolic syndrome* 11, 39 (2019).
60. Otu, H. H. et al. Prediction of diabetic nephropathy using urine proteomic profiling 10 years prior to development of nephropathy. *Diabetes care* 30, 638-643 (2007).
61. Lindhardt, M. et al. Predicting albuminuria response to spironolactone treatment with urinary proteomics in patients with type 2 diabetes and hypertension. *Nephrology Dialysis Transplantation* 33, 296-303 (2018).
62. Pena, M. J. et al. Plasma proteomics classifiers improve risk prediction for renal disease in patients with hypertension or type 2 diabetes. *Journal of hypertension* 33, 2123-2132 (2015).

63. Lewandowicz, A. et al. Changes in urine proteome accompanying diabetic nephropathy progression. *Polskie Archiwum Medycyny Wewnętrznej* 125, 27-38 (2015).
64. Tofte, N. et al. Early detection of diabetic kidney disease by urinary proteomics and subsequent intervention with spironolactone to delay progression (PRIORITY): a prospective observational study and embedded randomised placebo-controlled trial. *The Lancet Diabetes & Endocrinology* 8, 301-312 (2020).
65. Li, Y. et al. Genome-wide association studies of metabolites in patients with CKD identify multiple loci and illuminate tubular transport mechanisms. *Journal of the American Society of Nephrology* 29, 1513-1524 (2018).
66. Luo, S. et al. Genome-wide association study of serum metabolites in the African American Study of Kidney Disease and Hypertension. *Kidney international* (2021).
67. Hu, J.-R. et al. Serum metabolites are associated with all-cause mortality in chronic kidney disease. *Kidney international* 94, 381-389 (2018).
68. Luo, S. et al. Serum metabolomic alterations associated with proteinuria in CKD. *Clinical Journal of the American Society of Nephrology* 14, 342-353 (2019).
69. Kwan, B. et al. Metabolomic markers of kidney function decline in patients with diabetes: Evidence From the Chronic Renal Insufficiency Cohort (CRIC) study. *American Journal of Kidney Diseases* 76, 511-520 (2020).
70. Huang, J. et al. Machine Learning Approaches Reveal Metabolic Signatures of Incident Chronic Kidney Disease in Individuals with Prediabetes and Type 2 Diabetes. *Diabetes* 69, 2756-2765 (2020).
71. Yu, B. et al. Serum metabolomic profiling and incident CKD among African Americans. *Clinical Journal of the American Society of Nephrology* 9, 1410-1417 (2014).
72. Duranton, F. et al. Plasma and urinary amino acid metabolomic profiling in patients with different levels of kidney function. *Clinical Journal of the American Society of Nephrology* 9, 37-45 (2014).
73. Feng, Q., Li, Y., Yang, Y. & Feng, J. Urine metabolomics analysis in patients with normoalbuminuric diabetic kidney disease. *Frontiers in Physiology* 11 (2020).
74. Kimura, T. et al. Identification of biomarkers for development of end-stage kidney disease in chronic kidney disease by metabolomic profiling. *Scientific reports* 6, 1-8 (2016).
75. Lee, J. et al. Changes in serum metabolites with the stage of chronic kidney disease: Comparison of diabetes and non-diabetes. *Clinica Chimica Acta* 459, 123-131 (2016).
76. Toyohara, T. et al. Metabolomic profiling of uremic solutes in CKD patients. *Hypertension research* 33, 944-952 (2010).
77. Gonzalez-Calero, L. et al. Hypertensive patients exhibit an altered metabolism. A specific metabolite signature in urine is able to predict albuminuria progression. *Translational Research* 178, 25-37. e27 (2016).
78. Rhee, E. P. et al. Metabolomics of chronic kidney disease progression: A case-control analysis in the chronic renal insufficiency cohort study. *American journal of nephrology* 43, 366-374 (2016).
79. Masania, J. et al. Urinary Metabolomic Markers of Protein Glycation, Oxidation, and Nitration in Early-Stage Decline in Metabolic, Vascular, and Renal Health. *Oxidative medicine and cellular longevity* 2019 (2019).
80. Manca, M. L. et al. Differential metabolomic signatures of declining renal function in types 1 and 2 diabetes. *Nephrology Dialysis Transplantation* (2020).
81. Sharma, K. et al. Metabolomics reveals signature of mitochondrial dysfunction in diabetic kidney disease. *Journal of the American Society of Nephrology* 24, 1901-1912 (2013).
82. Moon, S. et al. Circulating short and medium chain fatty acids are associated with normoalbuminuria in type 1 diabetes of long duration. *Scientific reports* 11, 1-13 (2021).
83. Haukka, J. K. et al. Metabolomic profile predicts development of microalbuminuria in individuals with type 1 diabetes. *Scientific reports* 8, 1-10 (2018).
84. Niewczas, M. A. et al. Circulating modified metabolites and a risk of ESRD in patients with type 1 diabetes and chronic kidney disease. *Diabetes Care* 40, 383-390 (2017).

85. Tofte, N. et al. Metabolomic assessment reveals alteration in polyols and branched chain amino acids associated with present and future renal impairment in a discovery cohort of 637 persons with type 1 diabetes. *Frontiers in endocrinology* 10, 818 (2019).
86. Ma, T. et al. UPLC-MS-based urine nontargeted metabolic profiling identifies dysregulation of pantothenate and CoA biosynthesis pathway in diabetic kidney disease. *Life Sciences* 258, 118160 (2020).
87. Niewczas, M. A. et al. Uremic solutes and risk of end-stage renal disease in type 2 diabetes: metabolomic study. *Kidney international* 85, 1214-1224 (2014).
88. Ng, D. et al. A metabolomic study of low estimated GFR in non-proteinuric type 2 diabetes mellitus. *Diabetologia* 55, 499-508 (2012).
89. Tao, S. et al. Analysis of serum metabolomics among biopsy-proven diabetic nephropathy, type 2 diabetes mellitus and healthy controls. *RSC advances* 9, 18713-18719 (2019).
90. Lin, H.-T. et al. ¹H nuclear magnetic resonance (NMR)-based cerebrospinal fluid and plasma metabolomic analysis in type 2 diabetic patients and risk prediction for diabetic microangiopathy. *Journal of clinical medicine* 8, 874 (2019).
91. Ibarra-González, I. et al. Optimization of kidney dysfunction prediction in diabetic kidney disease using targeted metabolomics. *Acta diabetologica* 55, 1151-1161 (2018).
92. Chen, C.-J., Liao, W.-L., Chang, C.-T., Lin, Y.-N. & Tsai, F.-J. Identification of urinary metabolite biomarkers of type 2 diabetes nephropathy using an untargeted metabolomic approach. *Journal of proteome research* 17, 3997-4007 (2018).
93. Solini, A. et al. Prediction of declining renal function and albuminuria in patients with type 2 diabetes by metabolomics. *The Journal of Clinical Endocrinology & Metabolism* 101, 696-704 (2016).
94. Devi, S. et al. Acyl ethanolamides in Diabetes and Diabetic Nephropathy: Novel targets from untargeted plasma metabolomic profiles of South Asian Indian men. *Scientific reports* 9, 1-11 (2019).
95. Rhee, E. P. et al. A combined epidemiologic and metabolomic approach improves CKD prediction. *Journal of the American Society of Nephrology* 24, 1330-1338 (2013).
96. Sheng, X. et al. Systematic integrated analysis of genetic and epigenetic variation in diabetic kidney disease. *Proceedings of the National Academy of Sciences* 117, 29013-29024 (2020).
97. Zawada, A. M. et al. SuperTAG methylation-specific digital karyotyping reveals uremia-induced epigenetic dysregulation of atherosclerosis-related genes. *Circulation: Cardiovascular Genetics* 5, 611-620 (2012).
98. Smyth, L. J., Patterson, C. C., Swan, E. J., Maxwell, A. P. & McKnight, A. J. DNA methylation associated with diabetic kidney disease in Blood-Derived DNA. *Frontiers in cell and developmental biology* 8 (2020).
99. Qiu, C. et al. Cytosine methylation predicts renal function decline in American Indians. *Kidney international* 93, 1417-1431 (2018).
100. Chen, J. et al. Epigenetic associations with estimated glomerular filtration rate among men with human immunodeficiency virus infection. *Clinical Infectious Diseases* 70, 667-673 (2020).
101. Lecamwasam, A. et al. DNA methylation profiling identifies epigenetic differences between early versus late stages of diabetic chronic kidney disease. *Nephrology Dialysis Transplantation* (2020).
102. Smyth, L. J. et al. Assessment of differentially methylated loci in individuals with end-stage kidney disease attributed to diabetic kidney disease: an exploratory study. *Clinical epigenetics* 13, 1-19 (2021).
103. Bell, C. G. et al. Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. *BMC medical genomics* 3, 1-11 (2010).
104. Maghbooli, Z. et al. Aberrant DNA methylation patterns in diabetic nephropathy. *Journal of Diabetes & Metabolic Disorders* 13, 1-8 (2014).
105. Zhang, H. et al. Correlation of CTGF gene promoter methylation with CTGF expression in type 2 diabetes mellitus with or without nephropathy. *Molecular medicine reports* 9, 2138-2144 (2014).

106. VanderJagt, T. A., Neugebauer, M. H., Morgan, M., Bowden, D. W. & Shah, V. O. Epigenetic profiles of pre-diabetes transitioning to type 2 diabetes and nephropathy. *World journal of diabetes* 6, 1113 (2015).
107. Shlush, L. I. et al. Admixture mapping of end stage kidney disease genetic susceptibility using estimated mutual information ancestry informative markers. *BMC medical genomics* 3, 1-12 (2010).
108. Iyengar, S. K. et al. Genome-wide scans for diabetic nephropathy and albuminuria in multiethnic populations: the family investigation of nephropathy and diabetes (FIND). *Diabetes* 56, 1577-1585 (2007).
109. Thomson, R. J. et al. New genetic loci associated with chronic kidney disease in an indigenous Australian population. *Frontiers in genetics* 10, 330 (2019).
110. Hishida, A. et al. Genome-wide association study of renal function traits: results from the Japan Multi-Institutional Collaborative Cohort Study. *American journal of nephrology* 47, 304-316 (2018).
111. Lin, B. M. et al. Whole genome sequence analyses of eGFR in 23,732 people representing multiple ancestries in the NHLBI trans-omics for precision medicine (TOPMed) consortium. *EBioMedicine* 63, 103157 (2021).
112. Van Zuydam, N. R. et al. A genome-wide association study of diabetic kidney disease in subjects with type 2 diabetes. *Diabetes* 67, 1414-1427 (2018).
113. Haukka, J. et al. Novel linkage peaks discovered for diabetic nephropathy in individuals with type 1 diabetes. *Diabetes* 70, 986-995 (2021).
114. Palmer, N. D. et al. A genome-wide association search for type 2 diabetes genes in African Americans. *PloS one* 7, e29202 (2012).
115. Huang, Y. et al. SNPs in PRKCA-HIF1A-GLUT1 are associated with diabetic kidney disease in a Chinese Han population with type 2 diabetes. *European journal of clinical investigation* 50, e13264 (2020).
116. Liao, L.-N. et al. Identified single-nucleotide polymorphisms and haplotypes at 16q22. 1 increase diabetic nephropathy risk in Han Chinese population. *BMC genetics* 15, 1-11 (2014).
117. Taira, M. et al. A variant within the FTO confers susceptibility to diabetic nephropathy in Japanese patients with type 2 diabetes. *PloS one* 13, e0208654 (2018).
118. Kao, W. L. et al. MYH9 is associated with nondiabetic end-stage renal disease in African Americans. *Nature genetics* 40, 1185-1192 (2008).
119. Igo Jr, R. P. et al. Genomewide linkage scan for diabetic renal failure and albuminuria: the FIND study. *American journal of nephrology* 33, 381-389 (2011).
120. Morris, A. P. et al. Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nature communications* 10, 1-14 (2019).
121. Pezzolesi, M. G. et al. Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes. *Diabetes* 58, 1403-1410 (2009).
122. Rogus, J. J. et al. High-density single nucleotide polymorphism genome-wide linkage scan for susceptibility genes for diabetic nephropathy in type 1 diabetes: discordant sibpair approach. *Diabetes* 57, 2519-2526 (2008).
123. Chen, G. et al. A genome-wide search for linkage to renal function phenotypes in West Africans with type 2 diabetes. *American journal of kidney diseases* 49, 394-400 (2007).