

https://dspace.org/

Investigating potential plasma nephrotoxicity biomarkers associated with first-line antiretroviral therapy regimens in people living with HIV in South Africa

Mokoena, Rethabile

2024-08

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/48710

Mokoena, Rethabile. (2024). Investigating potential plasma nephrotoxicity biomarkers associated with first-line antiretroviral therapy regimens in people living with HIV in South Africa. [PhD thesis, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/48710

Downloaded from a DSpace repository. https://dspace.org/



Investigating potential plasma nephrotoxicity biomarkers associated with first-line antiretroviral therapy regimens in people living with HIV in South Africa.

by

Rethabile Mokoena

(1076599)

Thesis

Submitted in fulfilment of the requirements for the degree

Doctor of Philosophy

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Dr S. Fanucchi (University of the Witwatersrand)

Co-supervisor: Dr P. Naicker (CSIR, Pretoria)

Co-supervisor: Dr I.S. Govender (CSIR, Pretoria)

August 2024

Declaration

I, **Rethabile Mokoena** (**1076599**), am a student registered for the degree of Doctor of Philosophy in Molecular and Cell Biology in the academic year of 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 the research thesis submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- 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.

Signature: 

Date: 8th August 2024

Abstract

***Introduction*:** South Africa contributes to 20% of the global HIV burden and hosts the largest national antiretroviral therapy program worldwide. Antiretroviral drugs (ARVs) used for HIV treatment can cause nephrotoxicity, which can progress into acute kidney injury (AKI) in patients. With early detection, this side-effect is reversible. However, current serum creatinine tests are unreliable for early detection in the South African population. Therefore, identifying potential protein biomarkers in plasma using mass spectrometry-based proteomics, can help with early detection of ARV-associated AKI in people living with HIV (PLHIV) in South Africa. Plasma is a favourable source of novel biomarkers as it is easily accessible, with minimal invasiveness, and is present throughout the body, serving as an indicator of a patient's health state.

***Aim*:** This study aims to identify and quantify differentially abundant plasma proteins that can distinguish PLHIV with and without AKI in the South African population.

***Methodology*:** One hundred and thirty-five plasma samples (53 AKI; 82 non-AKI) were compared with an in-house sample preparation protocol, using magnetic microspheres for automated on-bead sample clean-up and protein digestion. The peptides were analysed with an Evosep One LC system coupled to a Sciex TripleTOF® 5600 mass spectrometer, using a SWATH-data acquisition method. Data were processed using Spectronaut™ 17. An unpaired t-test was used to determine differentially abundant proteins identified between the AKI and non-AKI groups. Differentially abundant proteins with a fold-change ≥ 1.68, at an FDR of 1% were considered candidate markers. Enrichr was used to identify enriched gene ontology (GO) biological processes (BP), molecular functions (MF), cellular compartments (CC), and enriched pathways associated with the candidate markers (ranked by p-value ≤ 0.05).

***Results and discussion*:** Thirty-four candidate protein markers were identified. Three candidate markers showed increased abundance, while 31 candidate markers showed decreased abundance in the AKI group compared to the non-AKI group. The enrichment analysis revealed that the candidate markers participated in enriched biological processes, functions, and pathways of immune response, programmed cell death, cholesterol/lipid metabolism, and regulation of cell survival, repair, and regeneration.

***Conclusion*:** Candidate markers participate in enriched biological processes that correspond to known pathophysiological and cellular mechanisms contributing to AKI, suggesting that the candidate marker identified can be further verified for distinguishing the AKI group from the non-AKI group and validated as biomarkers for AKI.

Dedication

Lord Jesus, thank you for being the light along my path, for renewing my strength (when I **regularly** wanted to give up), for encouraging my heart with your word, for consistently making ways in the wilderness, for giving me understanding, for provision, and teaching me to trust in you and your sovereignty. It is my joy to dedicate this work to you, because without you and those you had placed to help me along my way, this would not have come to fruition.

To God be the glory. Kganya.

You have given me your shield of victory; your help has made me great. You have made a wide path for my feet to keep them from slipping. – 2 Samuel 22: 36 – 37 (NLT)

Acknowledgements

My sincerest gratitude goes to the various organisations and individuals that contributed to this study and the successful completion of this thesis.

I'd especially like to thank my supervisors Dr Previn Naicker, Dr Ireshyn Govender, and Dr Sylvia Fanucchi for their guidance, encouragement, understanding and tolerance. To Drs Naicker and Govender, thank you for your mentorship and the opportunity to pursue my research at the CSIR. I'm grateful for the knowledge you've both shared in guiding the completion of my project and formed an important part of my scientific foundation. Thank you for affording me the space to make mistakes and learn at my own pace, as well as the exposure I've received within the proteomics space. To Dr Fanucchi, thank you for your encouragement in taking on this PhD opportunity and many other opportunities that presented themselves along the way, as well as supporting me throughout this project. I'd also like to extend my gratitude to Dr Sindisiwe Buthelezi for her mentorship, encouragement, and assistance, in and outside of the lab. I've truly appreciated her guidance and friendship.

To my colleagues, Sinegugu Dubazana and Pateswana Mokwena, as well as other academic peers, thank you for your support, help and all the enjoyable experiences we've had these past 4 years. My gratitude also goes to the scientific and non-scientific community at the CSIR and the opportunity to integrate into the community and collaborate with clients. To my Wits community at the PSFRU, thank you for your support throughout this project and your helpful input in all my presentation and conference preparations.

To my parents and siblings, thank you for all your help, your encouragement, immense support, prayers, and persistent belief that I would reach this point successfully. To the rest of my family, thank you for your support and encouragement. To my friends, thank you for ensuring I had balance in my life during this time, for your support, grace, and faith in me. To my partner, thank you for the role you played over the past year, I appreciate your understanding, support, prayers, for pushing me and for being my sounding board.

I am grateful to the NFR and the CSIR for the funding I was awarded towards completing this qualification. To the clinical team at the PHRU, thank you for the work done in patient recruitment and management, as well as the work done with Dr Ireshyn Govender in sample collection and management. Additionally, I'd like to thank the patients that agreed to participate in this study, without them this study would not have andomized d.

Ke leboha ho menahane.

Research Outputs

Publications

R. J. Mokoena, S. Fanucci, F. Nabeemeeah, E. Variava, N. Martinson, I. S. Govender and P. Naicker. Plasma proteome profiling for candidates for AKI biomarkers associated with first-line ART in people living with HIV in South Africa. Journal of Proteomics. Submitted (JPROT-D-24-00352). August 2024.

I. S. Govender, **R. J. Mokoena**, S. Stoychev and P. Naicker. Urine-HILIC: Automated sample preparation for bottom-up urinary proteome profiling in clinical proteomics. Proteomes. 2023, 11, 29.

Oral Presentation

R. J. Mokoena, I. S. Govender, P. Naicker and S. Fanucchi. Investigating potential plasma nephrotoxicity biomarkers associated with first-line antiretroviral therapy regimens in people living with HIV in South Africa. 14th Wits Cross-Faculty Postgraduate Symposium. University of the Witwatersrand, Johannesburg, South Africa. August 2023.

Poster Presentations

R. J. Mokoena, I. S. Govender, P. Naicker and S. Fanucchi. Investigating potential plasma nephrotoxicity biomarkers associated with first-line antiretroviral therapy regimens in people living with HIV (PLHIV) in South Africa. Molecular Biosciences Research Trust (MBRT) Postgraduate Research Day. University of the Witwatersrand, Johannesburg, South Africa. December 2023.

R. J. Mokoena, I. S. Govender, P. Naicker and S. Fanucchi. Investigating potential plasma nephrotoxicity biomarkers associated with first-line antiretroviral therapy regimens in HIV positive South African patients. 22nd Human Proteome Organization World Congress. Busan, South Korea. September 2023.

R. J. Mokoena, I. S. Govender, P. Naicker and S. Fanucchi. Investigating plasma nephrotoxicity biomarkers associated with antiretroviral therapy regimens in HIV positive South African patients. Ubuntu Proteomics Summer School. Stellenbosch, South Africa. February 2023.

R. J. Mokoena, I. S. Govender, P. Naicker and S. Fanucchi. Investigating plasma nephrotoxicity biomarkers associated with antiretroviral therapy regimens in HIV positive South African patients. Molecular Biosciences Research Trust (MBRT) Postgraduate Research Day. University of the Witwatersrand, Johannesburg, South Africa. December 2023.

Gradflash Presentation (video/animation)

R. J. Mokoena, I. S. Govender, P. Naicker and S. Fanucchi. Investigating plasma nephrotoxicity biomarkers associated with antiretroviral therapy regimens in HIV positive South African patients. 13th Wits Cross-Faculty Symposium. University of the Witwatersrand, Johannesburg, South Africa. July 2022.

Table of Contents

Titles:	Page
Declaration	**i**
Abstract	**ii**
Dedication	**iii**
Acknowledgements	**iv**
Research Outputs	**v**
List of Figures	**xii**
List of Tables	**xv**
List of Abbreviations	**xvii**
Chapter 1: Introduction	**1**
1.1) Origin and overview of HIV/AIDS	1
1.1.1) Current statistics and impact of HIV/AIDS pandemic globally and in South Africa	2
1.1.2) HIV classification	4
1.1.3) HIV genome structure, life cycle and disease progression	5
1.1.3.1) HIV genome structure	5
1.1.3.2) HIV-1 life cycle	7
1.1.3.3) Disease progression	9
1.2) HIV prevention and treatment	10
1.2.1) Antiretroviral Therapy	10
1.2.2) Administration of HIV regimens	12
1.2.3) ART in South Africa	13
1.2.4) ARV-related toxicity	15

1.3) Justification for study in South African cohort 18

1.3.1) TDF-associated nephrotoxicity in South Africa 18
1.3.2) Mechanism of TDF-associated nephrotoxicity 20

1.4) Protein biomarkers for AKI detection 26

1.5) Mass spectrometry-based proteomics for biomarker discovery in plasma 30

1.5.1) Background on proteomics 30

1.5.2) Mass spectrometry-based proteomics 31

1.5.2.1) Mass spectrometry and related technologies 31

1.5.2.2) Other proteomic techniques and related technologies 32

1.5.2.3) Top-down vs bottom-up MS-based proteomic workflows 34

1.5.2.4) Targeted vs untargeted MS-based proteomic workflows 37

1.5.3) MS-based quantification strategies 41

1.5.3.1) Label-based quantification strategies 41

1.5.3.2) Labe-free quantification strategies 42

1.5.4) Considerations for MS-based proteomics biomarker discovery 43

1.5.4.1) Study design 43

1.5.4.2) Sample collection and preparation 44

1.5.4.2.1) Sample collection 44

1.5.4.2.2) Sample Preparation 45

1.5.4.3) Quality control measures for LC-MS/MS analysis 45

1.5.4.4) Statistical analysis of data 46

1.5.5) The utility of plasma in clinical proteomics studies 48

1.6) Prior ARV-associated AKI urinary biomarker discovery study in a South African Cohort 49

1.6.4) Brief overview of results from previous urinary proteome profiling of AKI study 50

1.7) Project Aim 52

1.8) Project Objectives 52

Chapter 2: Materials and Methods 53

2.1) Study Design 53

2.2) Materials 54

2.3) Overview of experimental procedures 56

2.4) Patient recruitment and sample collection 58

2.4.1) Patient recruitment 58

2.4.2) Sample collection 58

2.5) Use of randomised block design and quality control for sample analysis 60

2.5.1) Randomised block design 60

2.5.2) Quality control measures for sample analysis 60

2.6) Protein quantification using the BCA assay 62

2.7) Sample preparation 63

2.8) Automated HILIC sample clean-up and on-bead protein digestion 64

2.9) Peptide quantification using Pierce Quantitative Colorimetric Peptide Assay 67

2.10) Acquisition of LC-MS/MS data for the AKI study 68

2.11) Processing of LC-MS/MS patient data using Spectronaut™ 17 70

2.12) Retrospective power analysis 70

2.13) Gene ontology functional enrichment and network interaction analysis of candidate markers 71

2.14) Statistical analysis and Machine Learning 71

Chapter 3: Results 73

3.1) Patient demographic details and clinical characteristics of AKI 73

3.2) Preliminary analysis of the AKI study 75

3.3) Quality control for AKI study analysis 79

3.4) Identified candidate markers and gene ontology enrichment analysis of AKI study 84
3.5) Using machine learning to identify features that distinguish the conditions of the AKI study 93
3.6) Comparing identified candidate markers from the plasma and urinary AKI cohorts 101
3.7) Summary of Results 103
Chapter 4: Discussion **105**
4.1) Statistical assessment of patient demographic and AKI clinical details 106
4.2) Preliminary results of the AKI study 106
4.3) Quality control for AKI study 109
4.4) Differentially abundant proteins identified in AKI study 111
4.4.1) Immune Response 112
4.4.2) Programmed cell death: apoptosis and necrosis 115
4.4.3) Plasma lipoprotein assembly and cholesterol/lipid metabolism 118
4.4.4) Cellular processes in AKI cell survival, repair, and regeneration 121
4.5) Proteins distinguishing the AKI patient group from the non-AKI group identified by machine learning 126
4.5.1) Exploring machine learning models for distinguishing study groups in AKI study 126
4.5.2) Top five protein features used to distinguish the AKI study groups by selected ML model 129
4.5.3) Comparing ML-derived top protein features with statistically derived candidate markers 133
4.6) Comparing plasma and urinary proteome profiles of the same study groups 134
4.6.1) Benefits of analysing plasma and urine proteomic profiles of the

same study groups 134

4.6.2) Comparing plasma and urinary proteome-derived AKI candidate markers 137

Chapter 5: Conclusion **140**

Chapter 6: Appendix **143**

6.1) Plasma separation protocol 143

6.2) Technical reproducibility of protocol with KingFisher™ Flex 144

6.3) List of 65 candidate markers identified in preliminary data 146

6.4) Hela and Peptide_Pool QC replicate signal acquired during five days of analysis 148

6.5) List of 34 candidate markers identified in the AKI study 150

6.6) Feature importance rank values for distinguishing AKI group from non-AKI group using selected ML model 151

6.7) List of identified protein groups in the plasma AKI cohort 152

6.8) List of identified protein groups in the urine AKI cohort 154

Chapter 7: References **159**

List of Figures

Figure 1.1: Map of percentage global distribution of adult (15 – 49 years) HIV prevalence**.** **3**

Figure 1.2: Diagram of HIV-1 RNA genome structure. **6**

Figure 1.3: HIV-1 life cycle. **8**

Figure 1.4: Mechanism of action of TDF against HIV reverse transcriptase**.** **22**

Figure 1.5: Movement of tenofovir diphosphate (TFV) into lumen of renal proximal tubule. **23**

Figure 1.6: Schematic image of top-down(A) and bottom-up (B) proteomics workflows**.** **36**

Figure 1.7: Diagram of overview of targeted and untargeted (discovery) proteomics (bottom-up) workflow in plasma. **39**

Figure 1.8: Schematic diagram of DDA and DIA (SWATH) acquisition methods during untargeted tandem mass spectrometry analysis. **40**

Figure 2.1: Schematic diagram of randomized block design of AKI study. **55**

Figure 2.2: Overview of experimental procedures of the AKI study. **57**

Figure 2.3: Diagram of 96-well plate positioning in KingFisher Flex™ for sample clean-up and on-bead protein digestion. **66**

Figure 2.4: Diagram of 48 variable window (VW) schema used for acquiring data with the SWATH acquisition method and an overlap of 1 m/z between the windows over range of 399 – 1100 m/z. **69**

Figure 3.1: Volcano plot of differentially abundant proteins generated from the preliminary analysis**.** **76**

Figure 3.2: Bar graphs indicating the number of protein groups and modified peptides identified in the Hela_QC (A) and Peptide_Pool_QC (B) replicates over 5 days. **80**

Figure 3.3: Multi-scatter plots showing Pearson correlation of identified protein groups and peptides in Hela_QC (A, C) and Peptide_Pool_QC (B, D) replicates analysed over 5 days. **81**

Figure 3.4: Boxplot showing the median percentage coefficient of variation (% CV) for the Hela (A) in blue and Peptide_Pool (B) in green QC replicates. **82**

Figure 3.5: Bar graph indicating the average number of protein groups (A) and peptides (B) identified between the non-AKI (pink) and AKI (orange) study conditions. **85**

Figure 3.6: Line graph indicating the estimated power of the AKI study and the corresponding desired fold-change (FC) defining the candidate markers. **86**

Figure 3.7: Volcano plot of differentially abundant proteins generated from the AKI study**.** **87**

Figure 3.8: Interactions and networks of candidate markers identified with the pathways, processes, and tissues in which they are associated with. **92**

Figure 3.9: Confusion matrix of the selected machine learning (ML) model. **97**

Figure 3.10: Principal component analysis plot for features contributing to machine learning (ML) classification of study groups. **99**

Figure 3.11: Venn diagram comparing differentially abundant proteins identified by an unpaired t-test and a machine learning (ML) model of the same study group. **100**

Figure 3.12: Venn diagram comparing candidate markers identified in the urinary and plasma proteomes of the same study groups. **102**

Figure 4.1: Immune responses associated with injury to the kidney proximal tubule. **114**

Figure 4.2: Simplified illustration of main intrinsic and extrinsic signalling events and associated proteins known to promote apoptosis in renal proximal tubule epithelial cells (PTECs) under cellular stress. **117**

Figure 4.3: Changes in the mature HDL structural protein composition (A) and the associated functional changes of HDL (B) as well as interaction of ABCA-1 and mature HDL particle during cholesterol efflux suppression (C). **120**

Figure 4.4: Phases of AKI progression due to nephrotoxic or ischaemic-AKI and the associated percentage glomerular filtration rate. **124**

Figure 4.5: Illustration indicating the result of adaptive and maladaptive repair responses to AKI. **125**

Figure 4.6: Simplified diagram indicating glomerular filtration, tubular reabsorption, and secretion in the kidney nephron. **136**

Figure 6.1: Bar graph (A) and boxplot (B) indicating technical reproducibility of the plasma sample preparation workflow between and across Day 1 and 2 using the KingFisher Flex**.** **145**

Figure 6.2: Boxplots indicating the original and normalised (log10) response signal for the Hela_QC replicates analysed over 5 days. **148**

Figure 6.3: Boxplots indicating the original and normalised (log10) response signal for the Peptide_Pool_QC replicates analysed over 5 days**.** **149**

List of Tables

Table 1.1: List of available ARV drugs in southern Africa and related adverse effects**.** **17**

Table 1.2: List of published markers of AKI. **28**

Table 3.1: Demographic details and AKI clinical characteristics of study participants in the AKI and non-AKI groups. **74**

Table 3.2: Overview of the average number of identifications determined in the preliminary analysis. **75**

Table 3.3: GO biological processes and pathway analysis of candidate markers. **77**

Table 3.4: Table of LC derived peak parameters. **83**

Table 3.5: Overview of the average number of identifications determined in the preliminary analysis. **84**

Table 3.6: GO biological processes, molecular functions, and pathway analysis of candidate markers. **88**

Table3.7A: Mean performance values for ML models to evaluate feature selection methods. **95**

Table 3.7B: Mean performance values for ML models to evaluate ML classifiers**.** **95**

Table 3.7C: Mean performance values for ML models to evaluate imputation methods. **96**

Table 3.8: List of top 20 proteins differentiating the two study conditions ranked in order of importance using machine learning (ML). **98**

Table 3.9: List of candidate markers overlapping the urinary and plasma proteome of the study groups. **102**

Table 6.1: List of differentially abundant proteins generated from the preliminary study. **146**

Table 6.2: List of differentially abundant proteins generated from the AKI study. **150**

Table 6.3: List of top 20 proteins distinguishing the two study conditions ranked in order of importance using machine learning (ML). **151**

Table 6.4: List of identified protein groups generated from the plasma AKI cohort. **152**

Table 6.5: List of identified protein groups generated from the urine AKI cohort. **154**

List of Abbreviations

2D-DIGE	Two dimensional-differential gel electrophoresis
2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
ABC	Ammonium bicarbonate
ACN	Acetonitrile
ADPKD	Autosomal dominant polycystic kidney disease
AIDS	Acquired Immunodeficiency Syndrome
AKI	Acute kidney injury
ALAT	Alanine aminotransferase
AMP	Adenosine monophosphate
ART	Antiretroviral therapy
ARVs	Antiretrovirals
ASAT	Aspartate aminotransferase
ATP	Adenosine triphosphate
AUC	Area under the curve
AZT	Azidothymidine
BCA	Bicinchoninic acid assay
BGS	Biognosys
C-18	Octadecyl (18-carbon chain) bonded stationary phase
CCR5	C-C chemokine receptor type 5
$CD4^+$	Cluster of differentiation 4
CDC	Centers for Disease Control and Prevention
cDNA	Complementary deoxyribonucleic acid
CKD	Chronic Kidney Disease

CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
COVID-19	Coronavirus disease (2019)
CRF	Circulating Recombinant Form
CTD	Comparative Toxicogenomics Database
CV	Coefficient of variation
CXCR4	C-X-C chemokine receptor type 4
DAMP	Damage-associated molecular pattern
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
eGFR	Estimated glomerular filtration rate
ELISA	Enzyme-linked immunosorbent assay
EMT	Epithelial-mesenchymal transition
ERK	Extracellular-regulated protein kinase
ER	Endoplasmic reticulum
ESI	Electrospray ionisation
EV	Extracellular vesicle
FA	Formic acid
FC	Fold-change
FDA	Food and Drug Administration
FDR	False Discovery Rate
FLC	Free light chains
FPR	False positive rate

FWHM	Full width half maximum
g	Gravitational force (RCF)
GluFib	Glu-fibrinopeptide β
GO	Gene Ontology
gp160/120/41	Glycoprotein 160/120/41
HAART	Highly Active Antiretroviral Therapy
HDL	High Density Lipoprotein
HIT	Heparin-induced thrombocytopenia
HILIC	Hydrophilic Interaction Liquid Chromatography
HIV	Human Immunodeficiency Virus
HPLC	High Performance Liquid Chromatography
IAA	Iodoacetamide
IGFBP-7	Insulin-like Growth Factor Binding Protein 7
IL-18	Interleukin-18
InSTI	Integrase strand transfer inhibitors
I/RI	Ischaemia/Reperfusion Injury
KEGG	Kyoto Encyclopedia of Genes and Genomes
KDIGO	Kidney Disease Improving Guidelines Outcome
KIM-1	Kidney injury molecule-1
LAV	Lymphadenopathy-associated virus
LC	Liquid Chromatography
LC-MS	Liquid Chromatography-mass spectrometry
LDL	Low-Density Lipoproteins
MALDI	Matrix-assisted laser desorption/ionization
MAPK	Mitogen-activated protein kinase
MDRD	Modification of diet in renal disease

MHC	Major histocompatibility class
ML	Machine Learning
MMP	Matrix Metalloproteinase
MRM	Multiple Reaction Monitoring
mRNA	messenger ribonucleic acid
MRP	Multi-drug proteins
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
mtDNA	mitochondrial deoxyribonucleic acid
m/z	mass-to-charge
NAG	N-acetyl-β-D-glucosaminidase
NGAL	Neutrophil gelatinase-associated lipocalin
NH_4Ac	Ammonium acetate
NNRTI	Non-nucleoside reverse transcriptase inhibitors
NRTI	Nucleoside reverse transcriptase inhibitors
NtRTI	Nucleotide reverse transcriptase inhibitors
OAT	Organic anion transporter
PCD	Programmed cell death
PEP	Post-exposure prophylaxis
PI	Protease inhibitors
PLHIV	People living with HIV
PR	Precision-Recall
PrEP	Pre-exposure prophylaxis
PRM	Parallel Reaction Monitoring
PTC	Proximal tubular cells

PTEC	Proximal tubular epithelial cells
PTM	Post-translational modifications
QC	Quality control
RAGE	Receptor for advanced glycation end-product
RBP	Retinol-binding protein
RCC	Renal cell carcinoma
RNA	Ribonucleic acid
ROC	Receiver operating characteristics
ROS	Reactive oxygen species
RPLC	Reverse phase liquid chromatography
RPTEC	Renal proximal tubular epithelial cells
SAPK	Stress-activated protein kinase
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SIV	Simian Immunodeficiency Virus
SOP	Standard operating procedure
SPD	Samples per day
srCr	serum creatinine
SRM	Selected Reaction Monitoring
STI	Sexually Transmitted Infection
SWATH	Sequential Window Acquisition of all Theoretical fragment-ion spectra
TB	Tuberculosis
TDF	Tenofovir disoproxil fumarate
TFA	Trifluoroacetic acid
TFV	Tenofovir
TFV-DP	Tenofovir-diphosphate

TIMP-2	Tissue inhibitor of metalloproteinase-2
TLR	Toll-like receptor
TOF	Time-of-Flight
Tris-HCl	Tris(hydroxymethyl)aminomethane-hydrochloric acid
TPR	True Positive Rate
UNAIDS	United Nations Programme on HIV/AIDS
UUO	Unilateral ureteral obstruction
UV-Vis	Ultra violet-visible
VEGF	Vascular endothelial growth factor
VLDL	Very Low-Density Lipoprotein
VW	Variable Window
vWF	von Willebrand Factors
WHO	World Health Organization

Chapter 1

Introduction

1.1) Origin and overview of HIV/AIDS

Human Immunodeficiency Virus (HIV) is a retrovirus that infects immune cells, such as $CD4^+$ T-lymphocytes, and hinders the immune system from adequately fighting infections (World Health Organization, 2022a). In the 1980s, 2 types of HIV were identified: HIV-1 and HIV-2 (Barré-Sinoussi *et al*., 1983; Clavel *et al*., 1986). Both types of HIV have similar morphologies but are genetically distinct from each other (Motomura *et al*., 2008) and possess different mechanisms that are used to avoid the host's immune response, with HIV-1 being the most common and virulent HIV type (Sharp and Hahn, 2011).

HIV-1 was first isolated in France, in 1983, and was previously referred to as lymphadenopathy-associated virus (LAV) (Barré-Sinoussi *et al*., 1983). LAV was isolated from a patient with generalized lymphadenopathy and it was later found that similar viruses were isolated from various patients who were diagnosed with acquired immunodeficiency syndrome (AIDS) (Barré-Sinoussi *et al*., 1983).

AIDS was first considered a disease in 1981, with the rise in unusual deaths of young homosexual men in the USA (Sharp and Hahn, 2011). It was found that HIV-1 and -2 infection were AIDS-causing agents (Barré-Sinoussi *et al*., 1983; Ja *et al*., 1984; Popovic *et al*., 1984). AIDS is a fatal chronic condition that arises due to a progressively weakened immune system, that leaves an infected individual open to lethal opportunistic infections and reduces life expectancy (Sharp and Hahn, 2011).

HIV-2 was later identified in 1986 in West Africa, and phylogenetic analysis showed that HIV-2 is distantly related to HIV-1 but is closely related to a Simian Immunodeficiency Virus (SIV) found in Sooty Mangabeys monkeys (*Cercocebus atys*) (Clavel *et al*., 1986). It was later found that the precursor for HIV-1 originated from a SIV in chimpanzees (*Pan troglodytes troglodytes*) (Gao *et al*., 1999). This provided evidence indicating that HIV arose through crossing of the species barrier from primates as SIV to humans via bloodborne transmission (Peeters *et al*., 1989).

HIV-1 and -2 share the same modes of transmission, which is primarily through sexual intercourse, however it can also be transmitted from mother to child during birth and while

breastfeeding, as well as through the sharing of needles, blood transfusions and direct contact with the bodily fluids of an infected individual (Campbell-Yesufu and Gandhi, 2011; Cohen *et al*., 2011). Although both HIV types are similarly transmitted, HIV-2 has lower transmissibility than HIV-1 and is characterized by a slower progression to AIDS, a higher CD4 cell count and lower viral loads (Campbell-Yesufu and Gandhi, 2011) and as a result, HIV-1 is known to be the main cause for the global AIDS pandemic (Sharp and Hahn, 2011; Barré-Sinoussi *et al*., 2013).

With the emergence of AIDS, the HIV/AIDS pandemic has since been a prime global health concern, which has claimed the lives of approximately 40.1 million people by the end of 2021 (World Health Organization, 2022a). Although numerous lives have been lost as a result of this pandemic, the implementation of antiretroviral therapy (ART) has reduced AIDS-related deaths by 68% between 2004 (recorded the highest number of AIDS-related deaths) and 2021, and it has decreased new infections by 54% between 1996 (recorded the largest number of new infections) and 2021; this has enabled people living with HIV to lead longer and relatively healthy lives (UNAIDS, 2022).

1.1.1) Current statistics and impact of HIV/AIDS pandemic globally and in South Africa

In 2021, approximately 38.4 million people were living with HIV worldwide (World Health Organization, 2022a), with about 1.5 million people being newly infected and 650 000 infected people succumbing to AIDS-related deaths (UNAIDS, 2022). The sub-Saharan region of Africa is especially affected, and accounts for approximately 67% of all people living with HIV globally (World Health Organization, 2022b).

South Africa, in particular, is the epicenter of the HIV/AIDS epidemic in this region, with 8.45 million people living with HIV in 2022, translating to about 13.9% of the overall population and 19.6% of the adult population (15-49 years) as shown in Figure 1.1 (Statistics South Africa, 2022). Furthermore, South Africa had 230 000 newly infected people and a loss of 63 000 people due to AIDS-related illnesses in 2021 (UNAIDS, 2021a). These figures confirm that the impact and burden of HIV on the South African population, is still unnervingly high after 20 years of combatting the epidemic, even with access to free HIV treatment.

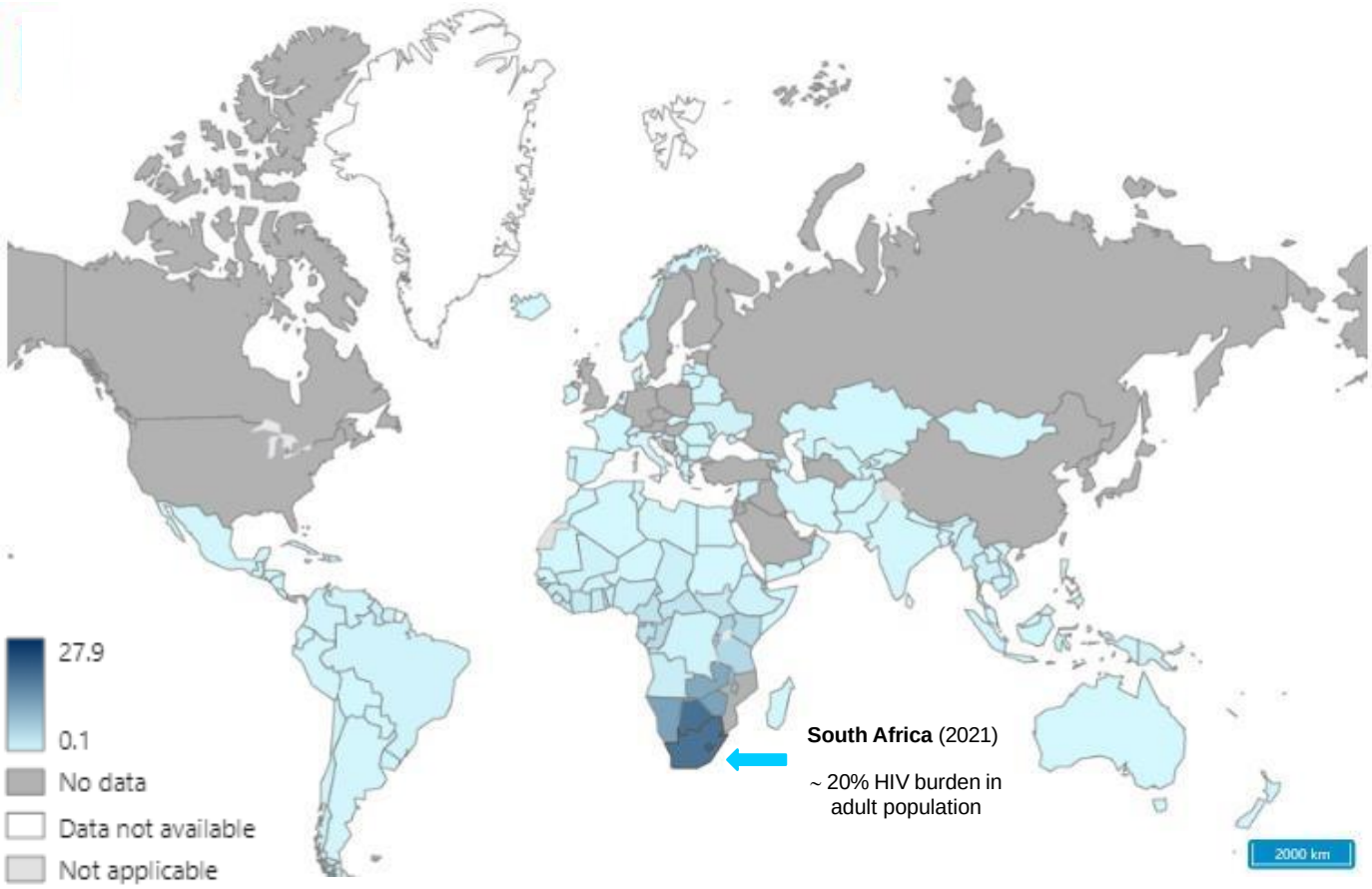


Figure 1.1: Map of percentage global distribution of adult (15 – 49 years) HIV prevalence*. Map showing the percentage HIV prevalence in the adult population distributed across the world, with approximately 20% of the South African population living with HIV in 2021. Figure adapted from World Health Organization, 2023.*

1.1.2) HIV classification

HIV-1 and -2 underwent multiple cross-species transmission events at various times in their history and these events resulted in different viral groups (Peeters *et al*., 1989; Barré-Sinoussi *et al*., 2013). HIV-2 is classified into 9 groups that are genetically distinct, group A-I, with only group A and B circulating in the human population and both groups have mainly been confined to the West African region (Santiago *et al*., 2005; Visseaux *et al*., 2016); the other groups were only identified in single individuals (Sharp and Hahn, 2011). HIV-1, however, consists of 4 genetically distinct groups, group M, N, O and P (Gürtler *et al*., 1994; Mauclère *et al*., 1997; Simon *et al*., 1998; Plantier *et al*., 2009; Sharp and Hahn, 2011). Group M or the main group, is the most prevalent group and is responsible for about 90% of HIV-1 infections; group M is further divided into 9 subtypes (genetically related HIV-1 strains): A-D, F-H, J and K, with subtypes A-D having a higher global prevalence than the rest (Taylor *et al*., 2008).

Subtype C affects approximately 50% of all people living with HIV-1 and is the most dominant strain in southern Africa and India (Hemelaar *et al*., 2006; Taylor *et al*., 2008; Bbosa *et al*., 2019). Subtype B, which is less prevalent and accounts for 12% of global infections, is the most researched subtype and is found in Europe, Australia, the Americas, the Middle East, and Northern Africa (Hemelaar *et al*., 2006; Taylor *et al*., 2008; Bbosa *et al*., 2019). Subtype A is also less prevalent and is responsible for 10% of all HIV-1 infections, it is mainly found in Russia and is dominant in eastern Africa (Hemelaar *et al*., 2006; Taylor *et al*., 2008; Bbosa *et al*., 2019). Asia, West and West-Central Africa predominantly face recombinant forms of HIV-1 subtypes, such as circulating recombinant forms (CRFs), which are hybrid forms of different HIV-1 subtypes (Hemelaar *et al*., 2006; Taylor *et al*., 2008; Bbosa *et al*., 2019). The various subtypes in group M indicate the high genetic variability of HIV-1, which is a result of the rapid and continuous evolution of HIV (Taylor *et al*., 2008; Sharp and Hahn, 2011).

Currently, the majority of available antiretrovirals (ARVs) and clinical research accommodate people living with HIV-1 subtype B, which caters for approximately 12% of the global HIV burden; whereas optimised HIV treatment and research for HIV-1 subtype C, which plagues 50% of people living with HIV, is not as extensive (Taylor *et al.,* 2008). This contributes to the challenges experienced with regards to drug resistance in the context of sub-Saharan Africa, as different HIV-1 subtypes present different responses to ARVs and vaccines and have a different disease progression, as well as different transmission rates (Taylor *et al.,* 2008).

Unlike group M, group O (outlier group), group N (non-main and non-outlier) and group P are much less prevalent (Taylor *et al*., 2008; Sharp and Hahn, 2011). Group O accounts for < 1% of global infections and is concentrated in West-Central Africa, and group N and P were

identified only in a few individuals (Taylor *et al*., 2008; Sharp and Hahn, 2011). Although these viral groups are genetically different from each other and are distributed differently across the human population, they all have the potential to cause AIDS (Campbell-Yesufu and Gandhi, 2011).

1.1.3) HIV genome structure, life cycle and disease progression

1.1.3.1) HIV genome structure

As a retrovirus, HIV uses ribonucleic acid (RNA) as its genetic material in order to replicate itself within the host cells (Hu and Hughes, 2012; German Advisory Committee Blood, 2016). A retrovirus uses reverse transcriptase, a viral enzyme, to convert RNA into viral complementary deoxyribonucleic acid (cDNA), which is then integrated into the host genome to make more copies of itself using host cellular machinery (Hu and Hughes, 2012; German Advisory Committee Blood, 2016). The full HIV genome is encoded in 2 separate single-stranded RNA molecules encapsulated with viral enzymes, namely reverse transcriptase, and integrase, within the viral capsid (Frankel and Young, 1998; German Advisory Committee Blood, 2016).

Upon infection, the RNA molecules are reverse transcribed into double-stranded viral cDNA, which is incorporated into the host genome (Frankel and Young, 1998; German Advisory Committee Blood, 2016) and as shown in Figure 1.2, the HIV-1 genome encodes for 3 precursor proteins: gag, gag/pol and env. The *gag* gene encodes for proteins responsible for the formation of the viral capsid (capsid protein – p24, matrix protein – p17, nucleocapsid protein – p 7 and a p6 protein), *gag/pol* encodes for the viral enzymes (reverse transcriptase, protease, integrase, and RNase H) and *env* encodes for proteins associated with the HIV-1 lipid bilayer membrane (glycoprotein 120 and 41) to facilitate budding (Frankel and Young, 1998; Kuiken *et al*., 2008; German Advisory Committee Blood, 2016). The HIV-1 genome also encodes for 7 other viral proteins, namely tat, rev, nef, vif, vpr, vpu, and asp; these proteins are responsible for initiating and regulating protein expression and viral replication for the formation of mature virions (Frankel and Young, 1998; Kuiken *et al*., 2008; German Advisory Committee Blood, 2016).

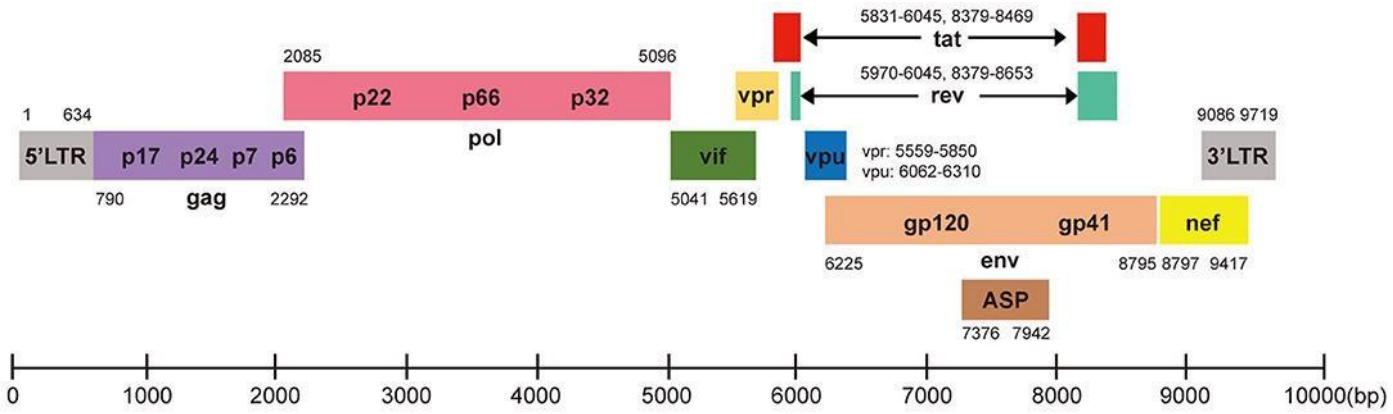


Figure 1.2: Diagram of HIV-1 RNA genome structure*. Diagram shows open reading frames of structural and regulatory HIV-1 genes encoded between a 5' and 3' long term repeats. Figure taken from Xiao et al., 2019.*

1.1.3.2) HIV-1 life cycle

In order to produce mature virions, HIV undergoes a complex replication cycle displayed in Figure 1.3. There are 7 stages of the HIV-1 life cycle: **attachment/binding, cell fusion, reverse transcription, integration, replication, assembly** and **budding** (German Advisory Committee Blood, 2016). During the attachment stage, glycoprotein gp120, found on the outer surface of HIV, binds to the CD4 receptor on the surface of the host $CD4^+$ immune cell and this protein-protein interaction initiates the infection of the host immune cell (Wilen *et al*., 2012; German Advisory Committee Blood, 2016). This binding process causes conformational changes to occur to both bound receptors, enabling gp120 to also bind to host cell chemokine co-receptor, CCR5 and/or CXCR4, as this is necessary for entry into the host cell (Wilen *et al*., 2012; German Advisory Committee Blood, 2016). This interaction allows glycoprotein gp41, found on the inside of the viral envelope, to induce fusion of the host cell membrane with the viral membrane, thereby releasing the viral capsid into the cytoplasm of the host cell (Wilen *et al*., 2012; German Advisory Committee Blood, 2016).

Once in the cytoplasm, the viral capsid releases its contents and viral RNA is converted into cDNA by HIV reverse transcriptase (Hu and Hughes, 2012; German Advisory Committee Blood, 2016). Thereafter, the viral DNA, HIV integrase and some regulatory proteins form the pre-integration complex, which is transported into the cell nucleus where HIV integrase incorporates the viral DNA into the host genome (Craigie and Bushman, 2012; German Advisory Committee Blood, 2016). At this point in the life cycle, the infection of the host cell is established, and viral DNA is replicated with the host genome during cell division (Craigie and Bushman, 2012; German Advisory Committee Blood, 2016). Subsequently, the host cell's transcription machinery is used to transcribe viral mRNA, which is then translated in the cytoplasm to produce gag, gag/pol and env precursor polypeptides, as well as the regulatory proteins (Craigie and Bushman, 2012; German Advisory Committee Blood, 2016). These 3 precursor polypeptides, the regulatory proteins and genomic RNA then migrate to the cell membrane to prepare for the assembly of a non-infectious HIV particle (Sundquist and Kräusslich, 2012; German Advisory Committee Blood, 2016). The env precursor polypeptide is then complexed with carbohydrates to form glycoprotein gp160, which is cleaved into gp120 and gp41 by cellular enzymes; these surface proteins then enable the budding process by pushing out the immature HIV particle from the host cell surface (Sundquist and Kräusslich, 2012; German Advisory Committee Blood, 2016).

After the budding process, HIV protease then cleaves the remaining polypeptide precursors gag and gag/pol, therefore initiating the formation of the viral capsid, enabling the HIV particle to mature into an infectious virion (Sundquist and Kräusslich, 2012; German Advisory

Committee Blood, 2016). The mature virion can now attach to another $CD4^+$ immune cell and repeat this process to produce more of the virus within the host.

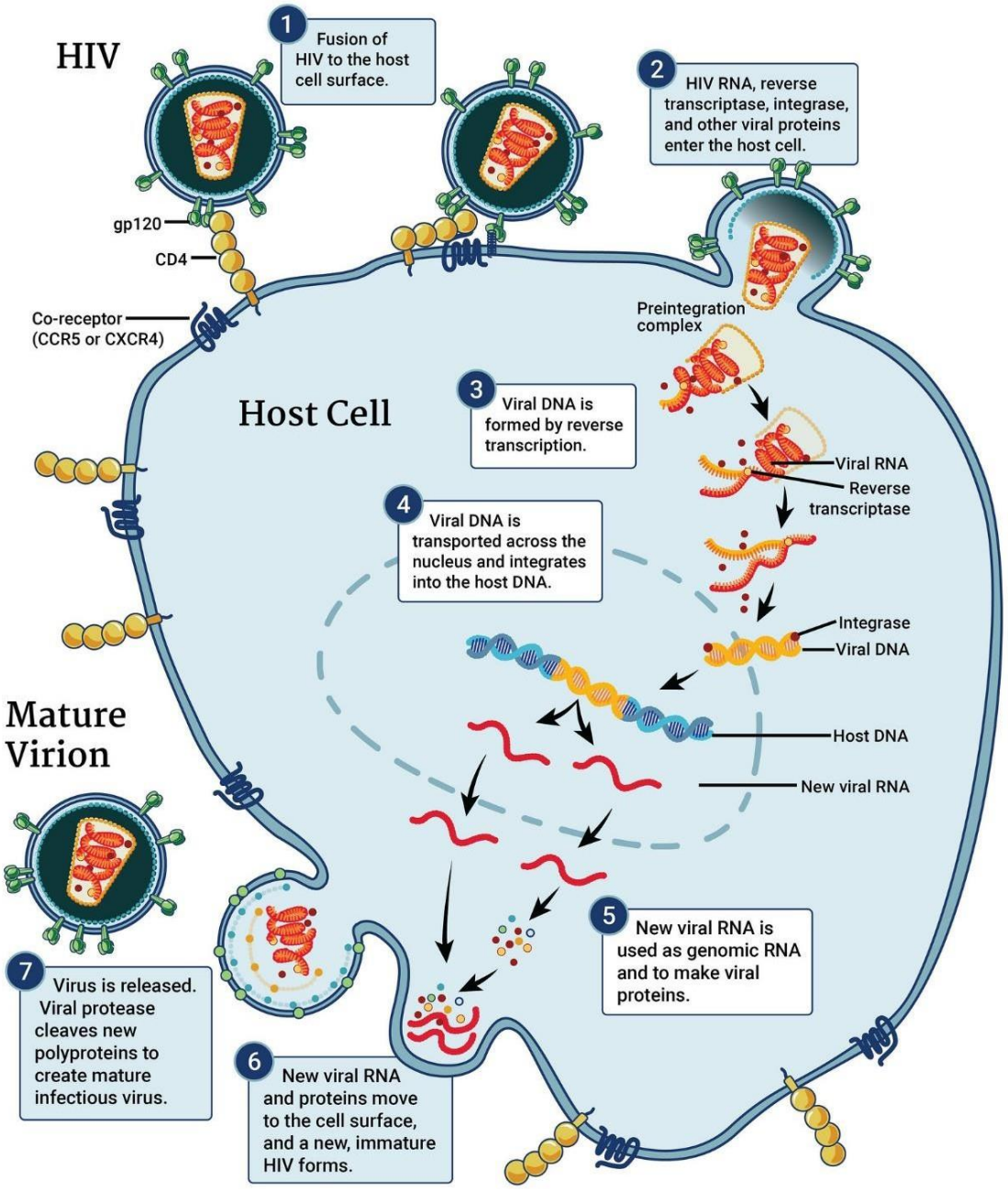


Figure 1.3: HIV-1 life cycle*. The life cycle of HIV-1 is illustrated from the point of entry and fusion of HIV-1 into the host cell to the budding and formation of a mature HIV-1 virion. Figure taken from National Institute of Health: NIAID, 2018.*

1.1.3.3) Disease progression

HIV infection progresses in 3 main phases: acute infection, chronic infection, and AIDS (Pantaleo and Fauci, 1996). Upon infection, the virus quickly replicates and spreads within the host, destroying $CD4^+$ immune cells and causing flu-like symptoms, such as headaches, fever, etc (Pantaleo and Fauci, 1996; Weinberg and Kovarik, 2010; Deeks *et al*., 2015). This phase continues for approximately 2 - 4 weeks and the host is at their most infectious (Weinberg and Kovarik, 2010; Deeks *et al*., 2015). The infection then proceeds into the chronic phase, where the virus continues to reproduce at a slower rate than in the acute phase and the host is generally asymptomatic (Pantaleo and Fauci, 1996; Weinberg and Kovarik, 2010). This period can last up to 10 years (or less, depending on the individual) and thereafter, the infected individual can start to experience symptoms as their immune system has progressively weakened (Weinberg and Kovarik, 2010; Deeks *et al*., 2015). This stage then advances to AIDS, which is the most severe stage of infection, with a 3-year survival rate (Weinberg and Kovarik, 2010). This final phase is highly transmissible and is characterised by a CD4 count that is less than 200 cells/mm^3 and/or the presence of opportunistic infections and cancers (Weinberg and Kovarik, 2010; Deeks *et al*., 2015).

When left untreated, HIV infection is fatal but with the appropriate treatment, the viral infection can slow down and even hinder disease progression (Egger *et al*., 2002). To ensure that the chances of a longer and healthier life are available, it is important for treatment to begin as soon as possible post infection (Egger *et al*., 2002).

1.2) HIV prevention and treatment

There is no effective cure for HIV/AIDS and as a result, there has been immense emphasis on prevention and strict adherence to treatment (CDC: HIV Prevention Division, 2023). Various prevention strategies for HIV infection have been implemented in the affected regions globally, such as programs educating the public about HIV prevention and transmission (Gupta *et al*., 2008), promotion of safer sexual practices, avoiding the sharing of needles and regular testing for HIV and other sexually transmitted infections (STIs) (Weller and Davis, 2002; Rodger *et al*., 2016). Additionally, voluntary medical male circumcision has shown that it can reduce heterosexually acquired HIV by 50% in men (Cohen *et al*., 2011a). As a means to prevent mother-to-child transmission, mothers need to maintain strict adherence to their treatment during and after pregnancy and to avoid breastfeeding their infant(s) (McIntyre and Gray, 2002).

Furthermore, medications such as pre-exposure prophylaxis (PrEP) and post-exposure prophylaxis (PEP) can be administered to people at a high risk of HIV infection and those who need to prevent infection after possible HIV exposure, respectively (Young *et al*., 2007; Buchbinder and Liu, 2011; Krakower *et al*., 2015; Riddell *et al*., 2018). Although these prevention strategies are in place, the global burden of HIV persists. Underlying social, economic, political and environmental factors that contribute to HIV risk and vulnerability need to also be addressed for the long-term success of HIV prevention programs (Gupta *et al*., 2008). As a result, current HIV treatment regimens have been paramount in minimizing the transmission of HIV globally (Broder, 2010).

1.2.1) Antiretroviral Therapy

Antiretroviral therapy (ART) is treatment recommended by the World Health Organization (WHO) for persons infected with HIV (World Health Organization, 2021). There are various HIV treatment regimens available, and they are meant to be taken daily to reduce the amount of HIV in the blood (i.e., the viral load) of the infected individual (World Health Organization, 2021). Decreasing the viral load provides the immune system with the opportunity to recover, produce more $CD4^+$ immune cells and regain the capacity to fight infections (Meintjes *et al.,* 2017). The main objective of ART is to reduce the viral load to undetectable amounts (< 200 copies of HIV per ml of blood) (Autran *et al*., 1997; Meintjes *et al*., 2017) in the body and this can only be accomplished with strict adherence to the prescribed HIV treatment regimen (Meintjes *et al.,* 2017).

The first antiretroviral (ARV) drug, azidothymidine (AZT), which was originally synthesised as an anti-cancer agent, was approved by the FDA in 1987 after a clinical trial showed that it can suppress HIV replication without harming uninfected cells *in vitro* and can delay the onset of AIDS in HIV positive patients (Fischl *et al*., 1987; Broder, 2010). This discovery led to the design and development of various ARV drugs between the late 1980s and early 1990s (Broder, 2010). Currently, there are 7 classes of ARVs, and each type inhibits a specific stage of the HIV replication cycle (Figure 1.3). ARVs are prescribed according to treatment guidelines, the health state and lifestyle of the patient, as well as possible ARV drug resistance (Broder, 2010; Arts and Hazuda, 2012).

The various classes of ARVs are listed: **nucleoside/nucleotide reverse transcriptase inhibitors** (NRTIs/NtRTIs), **non-nucleoside reverse transcriptase inhibitors** (NNRTIs), **protease inhibitors** (PIs), **integrase strand transfer inhibitors** (InSTIs), **fusion inhibitors**, **CCR5 antagonists** and **post-attachment inhibitors** (Broder, 2010; Arts and Hazuda, 2012). NRTIs/NtRTIs and NNRTIs bind to reverse transcriptase as competitive and non-competitive substrate inhibitors, respectively, thus inhibiting the conversion of RNA to DNA (Broder, 1990; De Clercq, 2004). The PIs prevent the maturation of virions by inhibiting protease activity (Robins and Plattner, 1993), and InSTIs hinder the integration of viral DNA into the host genome by inhibiting integrase activity (Evering and Markowitz, 2007). The CCR5 antagonist blocks the CCR5 co-receptor from binding to HIV glycoproteins and the fusion inhibitors impede HIV from entering the host cell (Kuritzkes, 2009; Tilton and Doms, 2010); post-attachment inhibitors block infected cells from transmitting the virus to a healthy cell (Kuritzkes, 2009).

Even with the increasing development of ARVs, no single ARV can suppress HIV for long periods of time due to the error-prone nature of HIV reverse transcriptase and host DNA-dependent RNA-polymerase, leading to mutant versions of HIV that can cause drug resistance (Abram *et al*., 2010; Arts and Hazuda, 2012). This initiated the use of combination therapy in the early 1990s, where mainly 2 NTRIs were used as an HIV treatment regimen; this resulted in more effective and lasting suppression of HIV replication, and reduced drug resistance by creating more barriers for HIV replication.

This led to the highly active antiretroviral therapy or HAART, which came into effect in 1996 (Hogg *et al*., 1999; Egger *et al*., 2002). This HIV regimen made use of a combination of 3 or more ARVs (consisting of at least 2 different ARV classes) to ensure the suppression of HIV replication to minimal levels; HAART has become the most effective regimen to date (Hogg *et al*., 1999; Egger *et al*., 2002). Despite this breakthrough, the drawbacks of HAART include

drug toxicities, drug-related side effects and resultant challenges with patient adherence (which is vital in preventing drug resistance) (Ferguson *et al*., 2002; Mills *et al*., 2006).

1.2.2) Administration of HIV regimens

The preferred first-line ART is the recommended HIV regimen used to initiate ART in all persons diagnosed with HIV, irrespective of their CD4 cell count or the clinical stage of their infection (World Health Organization, 2021). According to the 2016 WHO ART Guidelines, the preferred first-line regimen was a fixed-dose combination of tenofovir disoproxil fumarate and emtricitabine/lamivudine (NRTI drugs), as well as efavirenz (NNRTI drug) (World Health Organization, 2018). The preferred first-line ART can fail to effectively suppress a patient's viral load usually due to pre-treatment NNRTI drug resistance. This would require the patient to adopt an alternate first-line or second-line ART regimen that excludes NNRTI drugs (World Health Organization, 2018).

The alternative first-line regimen consisted of tenofovir disoproxil fumarate and emtricitabine/lamivudine (NRTI drugs) and dolutegravir (InSTI drug), and although shown to be more effective than the preferred first-line regimen in 2016, there was insufficient data on the safety and efficacy of this regimen amongst pregnant women, children, and HIV positive individuals with comorbidities (World Health Organization, 2018). Furthermore, the high cost and lack of generic fixed-dose combinations were other limiting aspects for prior recommendation of dolutegravir as a preferred regimen (World Health Organization, 2018). Currently, tenofovir disoproxil fumarate and emtricitabine/lamivudine (NRTI drugs) and dolutegravir (InSTI drug) have become the preferred first-line regimen for ART initiation; the safety and efficacy profile of this regimen has been established in response to the increasing demand for access to alternative ARV drugs, because of the rising levels of pre-treatment NNRTI drug resistance, especially in middle- and low-income countries (World Health Organization, 2021). The combination of tenofovir disoproxil fumarate and emtricitabine/lamivudine (NRTI drugs), as well as a lower dose of efavirenz (NNRTI drug) has become the current alternate first-line regimen (World Health Organization, 2021).

These first-line regimens have recently been formulated into fixed-dose tablets and generic formulations for convenient administration to improve patient adherence and reduce HIV drug resistance (Chesney, 2000; Davies, 2013). A second-line ART regimen would be adopted if the first-line regimens fail. These regimens consist of 2 NRTI drugs and a PI drug that is in combination with a low dose pharmacokinetic booster (World Health Organization, 2021). In the event that a second-line regimen fails, then a third-line regimen should be used, in accordance with national health guidelines. Policies and guidelines for third-line regimens are

at the discretion of individual national health departments and should include the use of new ARV drugs (InSTIs or second-generation PIs and NNRTIs) that will not cause cross-resistance with previous HIV regimens; if no new ARV drugs are accessible, then a tolerable regimen should be established instead (World Health Organization, 2021).

It is important to note that treatment efficacy does not only rely on adequate patient adherence to treatment and diminished drug resistance, but also on the access and cost of treatment (Ivers *et al*., 2005). The cost of ARV drugs increases as classes move from NRTIs and NNRTIs to PIs, InSTIs and entry inhibitors, and with each change of the HIV regimen, the cost is higher and the accessibility to more costly regimens become a challenge in resource-deprived regions (Meintjes *et al*., 2017; Moorhouse *et al*., 2019).

The inaccessibility of second- and third-line ART means that there can be delays in initiating patients into a new regimen promptly and this increases the risk of drug resistance, transmission of drug-resistant HIV, as well as increasing the mortality in these patients (Mtisi *et al.,* 2019). It is imperative that HIV regimens are planned adequately prior to patient administration and that the patient's health profile is taken into consideration in order to minimize side effects experienced and promote patient adherence, in addition to curbing drug resistance (Moorhouse *et al.,* 2019).

1.2.3) ART in South Africa

An estimated 20% of the adult South African population (15 – 49 years) lives with HIV and in an effort to combat the epidemic, South Africa also supports the largest HIV treatment program in the world, with approximately 5.5 million people having access to ART (Rees and Bisnauth, 2022; Statistics South Africa, 2022). South Africa's national ART program began in 2004 and is freely accessible to all South Africans infected with HIV (Meyer-Rath *et al.,* 2017). The South African government has committed to achieving the updated UNAIDS targets for HIV testing and treatment by 2025 and aim to: have 95% of all people living with HIV know their status, have 95% of people who know their status to be on treatment and that 95% of those on treatment will be virally suppressed (95-95-95) (UNAIDS, 2020, 2021b; Frescura *et al*., 2022). As of 2021, South Africa had 94% of all people living with HIV aware of their status, of which 79% are on HIV treatment and 92% of those on HIV treatment have suppressed viral loads (UNAIDS, 2021c).

South Africa complies with the first-line ART recommendations made by WHO for ART initiation and as of 2019, the new preferred first-line regime: tenofovir disoproxil fumarate/lamivudine/dolutegravir was introduced into the public health system due to its high

resistance barrier, minimal drug interactions and improved side effects over the previous preferred regimen (which was prone to causing drug resistance) (Republic of South Africa National Department of Health, 2020). The preferred first-line regiment will be initiated in an ART-naïve recipient after clinical assessments conclude an absence of meningitis, tuberculosis (TB) and renal disease, while awaiting laboratory baseline assessments (Republic of South Africa National Department of Health, 2020).

Those who are re-initiating ART or have a failed regimen and/or are experiencing pre-treatment drug resistance, have pushed the South African government to ensure that second- and third-line ART were made available to the public (Moorhouse *et al.,* 2019). In 2019, approximately 4% of South Africans on ART are on second-line HIV regimens (after failure of previous preferred first-line regimen) and a national third-line ART program has been established as well, which gives clear guidance and recommendations for individual patient regimens (Moorhouse *et al.,* 2019).

Aside from the cost of freely initiating various ART regimens, the South African health system faces challenges with poor treatment adherence, in addition to operational and logistical issues as the HIV program expands to meet the 95-95-95 UNAIDS goals (Bekker *et al*., 2014). Poor treatment adherence can lead to switching to regimens that can be less tolerable and more inconvenient to maintain, as well as to drug resistance that can result in resistant HIV transmission and even treatment failure, which places strain on the public health system (Mills *et al*., 2006; Bekker *et al*., 2014; Moorhouse *et al*., 2019). Logistical and operational issues that need to be overcome include consistent drug supply, acquainting diagnosed patients with local care centers and clinics, sufficient health staff and infrastructure, diagnostic and program monitoring services (Bekker *et al*., 2014).

The recent COVID-19 pandemic also posed a challenge to the expansion of the national ART program by reducing the number of anticipated ART initiations for people living with HIV (Benade *et al*., 2022). Although access to healthcare was permitted during the lockdown periods, access to public transport was scarce, lack of or redistributed medical supplies to COVID-19 centers, reduced health staff, and fear of COVID-19 transmission, deterred routine clinic access and disrupted ART initiation and re-initiation (Benade *et al*., 2022). The difficulty in accessing HIV treatment for the untreated posed a great risk of contracting COVID-19 and being severely ill, with potential fatalities depending on the health state of the individual, thus increasing the mortality of untreated South African’s living with HIV and delaying the South African Department of Health from achieving the 95-95-95 UNAIDS goals (Benade *et al*., 2022).

1.2.4) ARV-related toxicity

Although ART has transformed HIV infection into a more manageable chronic disease, people living with HIV can still experience adverse drug reactions caused by ARV treatment (National HIV Curriculum, 2023). ARVs can cause short- and long-term adverse reactions that range from mild intolerance to life-threatening toxicities (Meintjes *et al.,* 2017). ARV-related toxicities are reported as the most common reason for treatment non-adherence, drug substitutions and treatment discontinuation (World Health Organization, 2017, 2020). These adverse reactions can either appear as symptoms in patients or only in laboratory tests (Fernandez-Montero *et al*., 2013).

ARV-related toxicities can arise due to ARV usage, the existence of underlying conditions and infections (such as tuberculosis (TB) and kidney disease) that are exacerbated by ARV usage, as well as potential drug interactions with medicines treating pre-existing conditions and infections (Meintjes *et al.,* 2017). Furthermore, if treatment was initiated long after diagnosis, unusual clinical and laboratory toxicities can manifest (Fernandez-Montero *et al*., 2013). ARV-related toxicities can occur at any stage of HIV treatment, from ART initiation to prolonged ART exposure and it can be attributed to a specific drug or class of ARV drugs (World Health Organization, 2017).

ARV-related adverse effects range from weight gain, allergic reactions (hypersensitivity and skin reactions), gastrointestinal and neuropsychiatric reactions, mitochondrial and metabolic dysfunction, as well as nephro- and hepato-toxicity, to name a few (Table 1.1) (Meintjes *et al.,* 2017). Adverse reactions like hypersensitivity, hepatotoxicity and neuropsychiatric reactions can be observed shortly after initiating ART, whereas toxicities like mitochondrial dysfunction, metabolic abnormalities and renal toxicity are usually observed after months or years of exposure (Fernandez-Montero *et al*., 2013).

It is challenging to differentiate between ARV-related toxicities and the impediments associated with HIV infection, as a result, monitoring toxicity in HIV positive patients is crucial (Boswell and Rossouw, 2017a; World Health Organization, 2017). Further, with the 'universal test and treat' policy calling for the scale-up of HIV testing and ART initiation, more people will experience prolonged ART exposure, with potential to experience ARV-related toxicities and thus indicating the need for active pharmacovigilance in the HIV-treated population (World Health Organization, 2017).

Pharmacovigilance systems consist of laboratory tests and clinical assessments that are conducted prior to and following ART initiation (World Health Organization, 2017; Department of Health and Human Services, 2022). The data generated before ART initiation are used to confirm HIV diagnosis and attain the patient's medical profile to ensure that an appropriate

HIV regimen is prescribed (World Health Organization, 2017; Department of Health and Human Services, 2022). Additionally, the data obtained post ART initiation are used to monitor potential ARV-related toxicities and the efficacy of treatment regimens, as well as informing the changes made to HIV regimens (World Health Organization, 2017; Department of Health and Human Services, 2022). The HIV patient monitoring system is used at ART initiation as well as at 3, 6 and 12 months of treatment, then annually, especially in persons at risk of experiencing ARV toxicity (Republic of South Africa National Department of Health, 2020; World Health Organization, 2017, 2021).

Monitoring toxicity in this way, allows clinicians to detect any potentially unfavourable drug responses and make the appropriate recommendations to a patient's HIV regimen before any irreversible damage can occur (Meintjes *et al.,* 2017). Regular toxicity monitoring also aids in detecting the long-term effects of ART usage (Moosa *et al.,* 2019), it assesses the quality and performance of the ART program, it can improve ART guidelines and next generation ARV drugs (World Health Organization, 2017). However, pharmacovigilance systems may be underdeveloped in resource-limited regions due to the high influx of patients at clinics, inefficient reporting systems for toxicity and the lack of the necessary monitoring tools (Venter *et al.,* 2018), thus resulting in inadequate routine toxicity monitoring.

When ARV toxicity is observed, ART is not immediately discontinued, but the drug in question is substituted in the patient's HIV regimen when the adverse effects persist (National HIV Curriculum, 2023). In the case of irreversible damage, ART can be discontinued until the adverse effects are resolved and with the appropriate monitoring, ART can be re-initiated with a different HIV regimen, if possible (Meintjes *et al.,* 2017).

Table 1.1: List of available ARV drugs in southern Africa and related adverse effects*. The available ARVs are listed by their generic name, their corresponding drug class and common adverse effects associated with it (Meintjes et al., 2017).*

Generic name of ARV	Class of ARV	Common adverse effects of ARV
Tenofovir (TDF)	NRTI	Renal failure, tubular wasting syndrome, reduced bone mineral density
Lamivudine (3TC)	NRTI	Anaemia (rare)
Emtricitabine (FTC)	NRTI	Palmar hyperpigmentation
Abacavir (ABC)	NRTI	Hypersensitivity reaction
Zidovudine (AZT)	NRTI	Anaemia, GI upset, myopathy, lipoatrophy
Efavirenz (EFV)	NNRTI	Central nervous system symptoms (dizziness, confusion, psychosis, vivid dreams), rash, hepatitis, gynaecomastia
Etravirine (ETR)	NNRTI	Rash, hepatitis (both are uncommon)
Lopinavir/ritonavir (LPV/r)	Boosted PI	GI upset, dyslipidaemia, hepatitis
Darunavir (DRV)	PI	GI upset, rash, hepatitis, dyslipidaemia, sulpha allergy
Dolutegravir (DTG)	InSTI	Headache, CNS side effects, GI upset, hepatitis, rash
Maraviroc (MVC)	CCR5 blocker	Rash, hepatitis, fever, abdominal pain, cough, dizziness

1.3) Justification for study in South African cohort

ARV-related toxicities are a concern for the treatment-exposed HIV population, and ARV-associated nephrotoxicity in particular, is of concern in the black sub-Saharan African populations, as these patients have a greater likelihood of developing renal complications and diseases than patients of other races (Assaram *et al*., 2018; Fabian *et al*., 2022). This is further exacerbated by limited accessibility to renal function screening facilities, pathology laboratories for testing and programs for early detection of kidney dysfunction in the region (Assaram *et al*., 2018; Fabian *et al*., 2022). The preferred first-line regimen is widely administered and consists of tenofovir disoproxil fumarate (TDF), which is an ARV drug known to be tolerable and safe (Boswell and Rossouw, 2017; Assaram *et al*., 2018; Republic of South Africa National Department of Health, 2020). However, more severe adverse reactions such as renal toxicity (and even Fanconi syndrome or acute tubular necrosis) and decreased bone mineral density are seen with long-term TDF usage (Venter *et al*., 2018; National HIV Curriculum, 2023).

Adverse reactions associated with ARVs can be reversed when detected prior to any severe damage, by discontinuing and substituting the afflicting ARV drug in the HIV treatment regimen (Boswell and Rossouw, 2017; Venter *et al*., 2018). The challenge with early detection of ARV-associated nephrotoxicity is that patients can be asymptomatic for prolonged periods and in the sub-Saharan African region, diagnostic tools used to determine kidney function are not validated for the population, making it difficult to accurately assess renal function of patients to make timely, well-informed decisions on patient health and treatment regimens (Boswell and Rossouw, 2017; Waal *et al*., 2017; Mtisi *et al*., 2019; Fabian *et al*., 2022).

As a result, finding methods that will improve detection of ARV-associated nephrotoxicity, that are suitable for the sub-Saharan African region is crucial, and early detection will ensure that people living with HIV in the region can adhere to their treatment or that their treatment can be changed reliably without the burden of acquiring potentially life-threatening comorbidities.

1.3.1) TDF-associated nephrotoxicity in South Africa

A third of the sub-Saharan African HIV population resides in South Africa and with TDF-based first-line regimens widely used in this resource-strapped region, the risk of TDF-associated nephrotoxicity is high and the need to investigate TDF-related consequences is growing substantially (Waal *et al*., 2017; Venter *et al*., 2018; Mtisi *et al*., 2019). A genetic predisposition for renal injury has been postulated amongst people of African ancestry (George *et al*., 2018; Fabian *et al*., 2022). Studies have shown that African Americans are four times more likely to

develop end stage renal disease than Americans of European origin, and that this occurrence is greatly increased in PLHIV (Kasembeli *et al*., 2015; Wearne, 2015). Although a higher prevalence of renal complications/disease is observed in black Africans, there's an insufficient amount of genetic information available, to confirm the presence of specific mutations or variants with respect to kidney disease in African populations (George *et al*., 2018; Fabian *et al*., 2022).

Furthermore, genetic variations associated with a high prevalence of renal injury in African Americans cannot be extrapolated to Africans living in Africa, as the African American genome has been influenced by American and European populations, as well as other socio-economic and lifestyle factors that result in an inaccurate reflection of the incidence of renal injury in African populations (Fabian *et al*., 2022). Therefore, performing more genetic association studies of kidney disease in the South African population (George *et al*., 2018; Fabian *et al*., 2022) as well as improving documentation of renal disease/complications, will provide accurate information on the prevalence of kidney disease and help determine measures to equip the public health system to better manage the increasing burden, especially in PLHIV (Wearne, 2015; World Health Organization, 2017).

Additionally, South African clinicians have reported that approximately 10% of patients newly initiated on first-line ART, experience acute adverse nephrotoxicity (Seedat *et al*., 2017) and that extensive clinical administration of TDF-based regimens can cause TDF-associated nephrotoxicity (Waal *et al*., 2017; Mtisi *et al*., 2019). TDF-associated nephrotoxicity can progress into TDF-associated acute kidney injury (AKI), with the potential to cause irreversible renal damage that threatens ART continuation (Boswell and Rossouw, 2017). Furthermore, the increase in testing and treating of PLHIV, increases the number of individuals exposed to long-term ART and this increases the rate of ARV-related adverse reactions observed (World Health Organization, 2017). This emphasizes the need to detect and monitor potential nephrotoxicity in ART-initiated PLHIV from the onset of the condition and/or determine a patient's likelihood to develop the condition, to minimize the adverse effects of TDF-associated nephrotoxicity.

Monitoring nephrotoxicity in PLHIV is a routine procedure, especially within the first year of treatment (Republic of South Africa National Department of Health, 2020) and mainly consists of assessing renal function by measuring serum creatinine levels to estimate the glomerular filtration rate (eGFR) (Fabian *et al*., 2022). The eGFR is commonly calculated with the modification of diet in renal disease (MDRD) formula to assess renal function in South African adults and adolescents (≥ 16 years) (Republic of South Africa National Department of Health, 2020). However, the MDRD and the chronic kidney disease – epidemiology collaboration

(CKD-EPI) equations have shown to overestimate the GFR in African patients (Fabian *et al*., 2022). This is due to the limited validity of these equations for African populations, even with the race-based adjustments made to the equations, as they mainly accommodate African American populations and not Africans residing in Africa (De Waal *et al*., 2017; Fabian *et al*., 2022).

The overestimation of renal function in the treated South African HIV population, suggests that asymptomatic patients as well as mild and moderate nephrotoxic events can go undetected, causing misinformation about the state of kidney function and allowing for unimpeded progression to irreversible renal damage (Boswell and Rossouw, 2017; Venter *et al*., 2018; Mtisi *et al*., 2019). This indicates that current GFR-estimating equations are not reliable for early detection of TDF-associated nephrotoxicity (De Waal *et al.*, 2017; Fabian *et al*., 2022). Developing appropriate diagnostic tools for the South African population is crucial for accurately determining renal function (Waal *et al*., 2017; Mtisi *et al*., 2019) and diagnosing the onset of renal injury to reduce TDF-induced AKI, treatment non-adherence and discontinuation in PLHIV.

1.3.2) Mechanism of TDF-associated nephrotoxicity

As previously mentioned, TDF is a widely used ARV drug in HIV treatment and, in preventative medicines like pre-exposure prophylaxis (PrEP) and post-exposure prophylaxis (PEP) (Meintjes *et al*., 2017; Venter *et al*., 2018). TDF falls within the NRTI class of ARV drugs and is an orally administered prodrug of the active compound tenofovir (TFV) (Venter *et al*., 2018). Upon consumption, prodrug TDF is rapidly absorbed in the gut and cleaved by plasma esterases to release tenofovir (TFV) (van Bömmel and Berg, 2008; Wassner *et al*., 2020). TFV is then absorbed by CD4 positive lymphocytes and undergoes 2 subsequent intracellular phosphorylation events to produce tenofovir diphosphate (TFV-DP); this has favoured the use of nucleotide NRTIs rather than nucleoside NRTIs, which require 3 phosphorylation events to yield the active metabolite/compound and complete its therapeutic action (van Bömmel and Berg, 2008; Hamlin *et al*., 2019).

TFV-DP is a structural analogue of deoxyadenosine-5’-triphosphate, the natural substrate for HIV-1 reverse transcriptase to elongate the cDNA chain (van Bömmel and Berg, 2008; Kohler *et al*., 2009; Duwal *et al*., 2012). However, TFV-DP lacks a hydroxyl group at the 3’ carbon of the deoxyribose sugar and as a result, when TFV-DP competes for the reverse transcriptase binding site, its inclusion into the growing cDNA chain hinders the addition of more nucleotides, by ensuring that the 5’ to 3’ phosphodiester bonds cannot be formed (van Bömmel and Berg, 2008; Duwal *et al*., 2012; De Clercq, 2013). This terminates cDNA elongation, competitively

inhibits reverse transcription of HIV-1 RNA and leads to reduction of HIV replication (van Bömmel and Berg, 2008; Duwal *et al*., 2012).

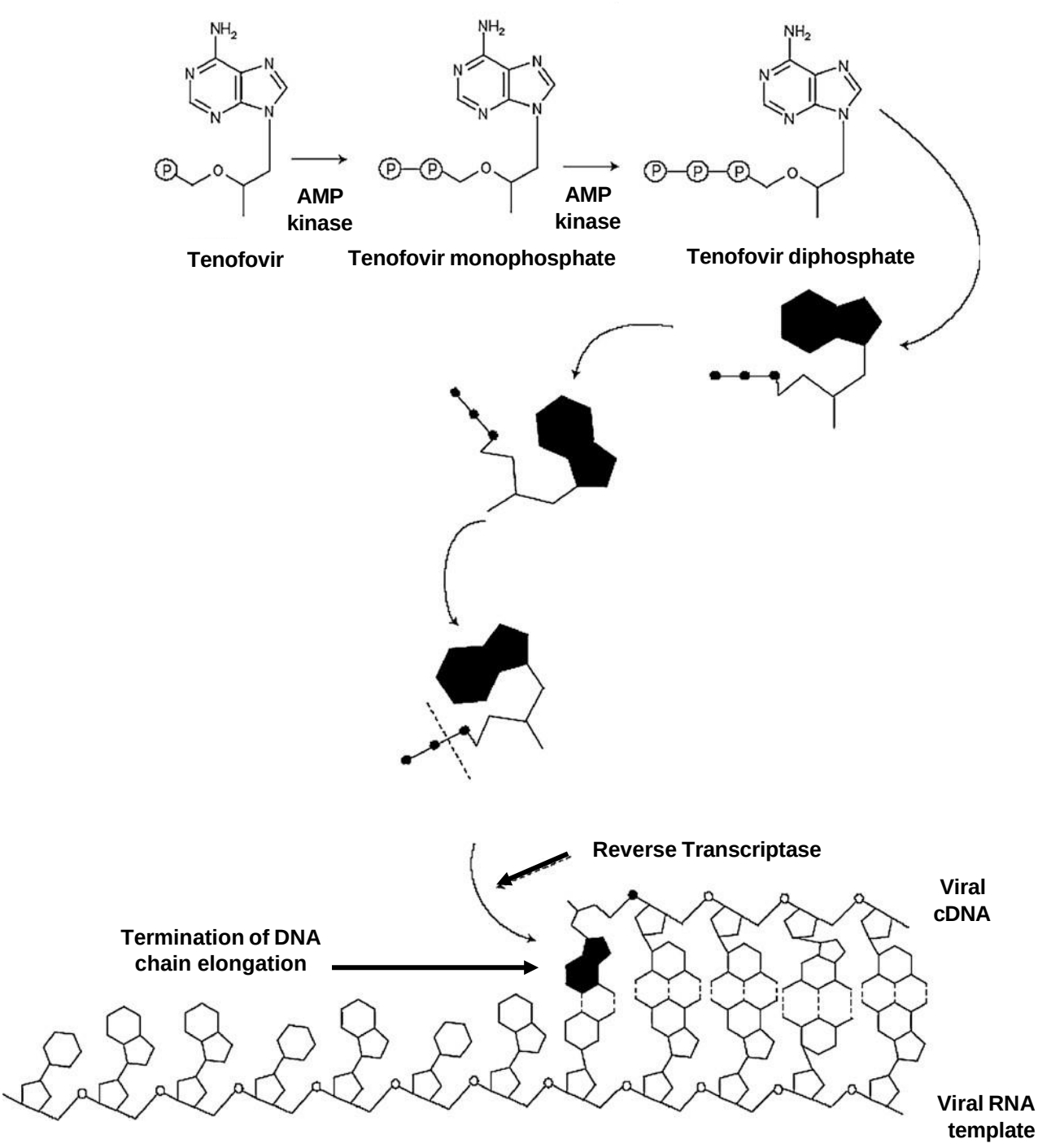


Figure 1.4: Mechanism of action of TDF against HIV reverse transcriptase*. The conversion of TDF to tenofovir monophosphate in vivo. The phosphorylation of tenofovir monophosphate results in tenofovir diphosphate and this binds to the reverse transcriptase and terminates the DNA elongation. Figure adapted from De Clercq, 2013 (License number for use of figure issued by Elsevier Books: 5765500948964).*

Subsequent to inhibiting reverse transcription, TFV-DP is mainly excreted in the urine unmodified through glomerular filtration and active tubular secretion in the proximal tubule of the kidney (Jafari *et al*., 2014; Tun-Yhong *et al*., 2017; Venter *et al*., 2018; Wassner *et al*., 2020). Up to 30% of the total TFV elimination occurs during active tubular secretion, where TFV-DP entry into the proximal tubule from the blood is facilitated by organic anion transporter (OAT) protein 1 and 3 (Jafari *et al*., 2014; Venter *et al*., 2018). Thereafter, multi-drug proteins (MRP) 2 and 4 mediate the transport of TFV-DP from the proximal tubule into the tubular lumen, where it reaches the collecting duct and is excreted from the body in the urine (Figure 1.5) (Jafari *et al*., 2014; Venter *et al*., 2018).

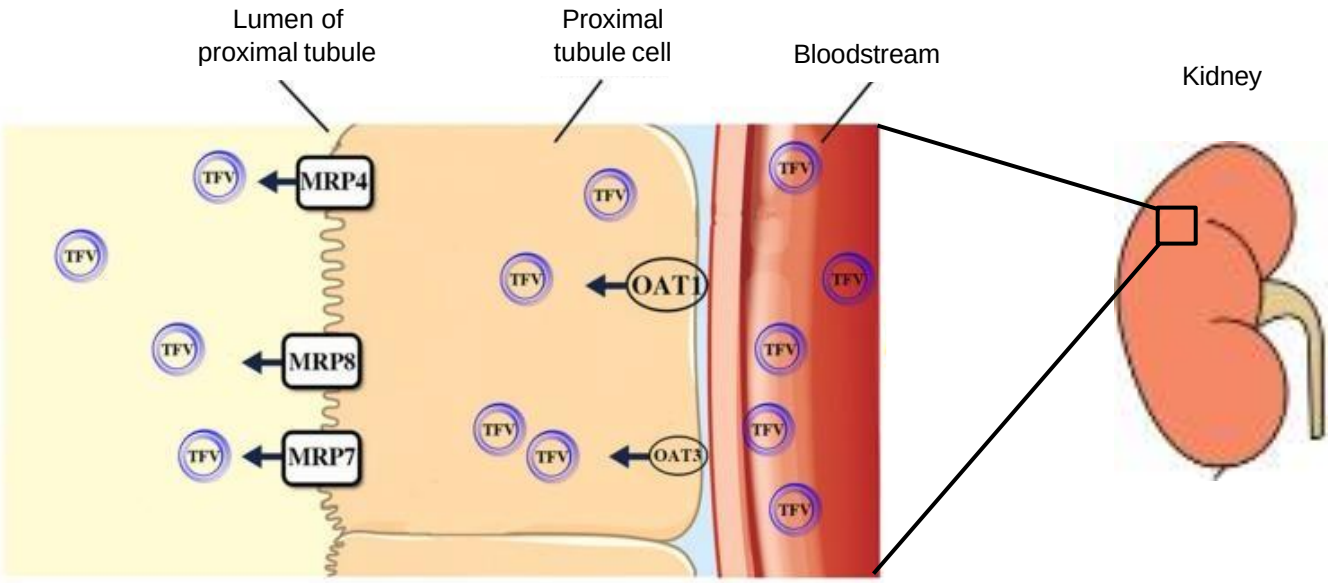


Figure 1.5: Movement of tenofovir diphosphate (TFV) into lumen of renal proximal tubule. *The movement of TFV from the blood into the proximal tubule is facilitated by OAT 1 and 3 proteins and then the subsequent movement of TFV into the lumen of the proximal tubule is mediated by MRPs. The lumen of the proximal tubule allows for TFV to move downstream and be excreted into urine. Figure is adapted from Tun-Yhong et al., 2017.*

The dysfunction of the proximal tubule is a characteristic of TDF-associated nephrotoxicity (Gallant and Deresinski, 2003; Jafari *et al*., 2014; Venter *et al*., 2018; Wassner *et al*., 2020). Damage caused to the proximal tubule results in the inadequate transport of metabolites (such as creatinine or drugs) filtered from the blood and secreted into the tubule, as well as those metabolites reabsorbed into the blood (such as calcium ions and small proteins) that can be lost to urine after tubular secretion (Gallant and Deresinski, 2003; Boswell and Rossouw, 2017; Venter *et al*., 2018; Wassner *et al*., 2020).

The weakly cytotoxic impact of TFV-DP on tubular function is usually undetectable clinically and requires laboratory testing to identify (Jafari *et al*., 2014; Boswell and Rossouw, 2017), unless exacerbated by risk factors such as pre-existing renal complications, a low $CD4^+$ cell count, aging, use of boosted PIs, co-morbidities and nephrotoxins that accompany HIV infection, to name a few (Jafari *et al*., 2014; Boswell and Rossouw, 2017; Tun-Yhong *et al*., 2017; Venter *et al*., 2018). The mechanism of action for TDF-associated nephrotoxicity has been proposed through studies of other NRTIs like adefovir and cidofovir that are known to induce nephrotoxicity (Jafari *et al*., 2014; Tun-Yhong *et al*., 2017).

It has been postulated that TDF-associated nephrotoxicity occurs due to the accumulation of TFV-DP in the proximal tubular cells, which can inhibit mitochondrial DNA (mtDNA) polymerase-γ activity and induce oxidative stress that can lead to mtDNA damage, mitochondrial injury and depletion (Kohler *et al*., 2009; Jafari *et al*., 2014; Tun-Yhong *et al*., 2017; Venter *et al*., 2018). Mitochondrial injury promotes structural defects, leakage of mitochondrial proteins into the cytoplasm and negatively impacts the molecular transport system necessary for tubular secretion and reabsorption (Venter *et al.,* 2018). It is suggested that OAT 1 and MRP 4 play a role in the regulation and accumulation of TFV-DP due to the increased influx of the compound into the proximal tubule by OAT 1 (assisted by other OAT proteins) and the decreased efflux of TFV-DP into the tubular lumen by MRP 4 (Jafari *et al*., 2014; Tun-Yhong *et al*., 2017; Venter *et al*., 2018). The damage caused to mitochondria can lead to apoptosis of the proximal tubular cells and promote development of acute kidney injury (AKI) or chronic kidney disease (CKD) (Boswell and Rossouw, 2017; Venter *et al.,* 2018).

CKD is characterized by the persistent and gradual decline of kidney function over time and usually arises due to the presence of chronic illnesses. CKD cannot be reversed but can be managed to preserve kidney function (George *et al*., 2018; Zuk and Bonventre, 2019). AKI, on the other hand, refers to a sudden episode of significant kidney damage or failure that persists over several days and can be a result of illness, structural injury or the presence of drugs; in many cases AKI can be reversed when the underlying condition is resolved (Boswell and Rossouw, 2017). Both conditions can range from mild kidney damage to complete organ

failure (Boswell and Rossouw, 2017). Both CKD and AKI result in the accumulation of salts and waste products in the blood and are diagnosed by measuring increased levels of serum creatinine, as well as measuring a decrease in the eGFR and the apparent decline in urine output (Hall *et al*., 2011; Boswell and Rossouw, 2017; De Waal *et al*., 2017).

Individuals suffering from either CKD or AKI can be asymptomatic in the early stages of damage and symptoms usually arise when severe damage has occurred to the kidneys (Boswell and Rossouw, 2017; Assaram *et al*., 2018). Furthermore, the clinical distinction between AKI and CKD is challenging at presentation, as both conditions have common biochemical abnormalities (Boswell and Rossouw, 2017; Assaram *et al*., 2018). Additionally, AKI is a risk factor for the development of CKD and vice versa (Ferenbach and Bonventre, 2016; Boswell and Rossouw, 2017; Assaram *et al*., 2018). As a result, early detection and prompt treatment of kidney dysfunction, especially AKI, is important as the abnormal levels of chemicals and waste in the blood can poorly affect other organs and pose the risk of organ damage/failure (Boswell and Rossouw, 2017; Assaram *et al*., 2018).

The timeous detection and treatment/resolution of underlying factors that can cause AKI, initiate reparative steps that lead to the reversal of the kidney injury (Basile *et al*., 2012; Ferenbach and Bonventre, 2016). The kidney responds to AKI with an adaptive repair mechanism, where damaged tubular epithelial cells are regenerated through cellular dedifferentiation and proliferation of surviving epithelial cells (Basile *et al*., 2012; Ferenbach and Bonventre, 2016; Yao and Wang, 2020). In addition, monocytes become the predominant leukocyte present at the site of cellular tubule injury and transform from a pro-inflammatory to a pro-repair phenotype to support this regenerative process; when successful, resident tubular epithelial cells and in turn kidney function, is restored to pre-injury level (Basile *et al*., 2012; Ferenbach and Bonventre, 2016). In this way, TDF-associated nephrotoxicity leading to AKI can be reversed when detected early.

There are various pathophysiological mechanisms that contribute to AKI and to understand these underlying mechanisms, reproducible animal models of AKI and CKD are studied experimentally (Harwood *et al*., 2022). The main animal models used to assess mechanisms related tubular injury in AKI, are the unilateral ureteral obstruction (UUO), cisplatin-induced AKI, and renal ischaemic-reperfusion injury (I/RI) models. The UUO model is used to cause renal fibrosis, where the primary feature of UUO is tubular injury due to obstructed urine flow (Martínez-Klimova *et al*., 2019). The cisplatin-induced AKI model is used to study nephrotoxicity; cisplatin is a potent anti-cancer drug that can cause tubular injury during its excretion from the kidneys and can induce renal cell death (McSweeney *et al*., 2021). The I/RI model, which is a commonly used model to study AKI and progression of CKD, occurs

when organ tissues receive reduced blood supply causing an ischaemic event. However, the crucial restoration of adequate blood supply for the survival of ischaemic tissues can cause reperfusion injury, resulting in further tissue damage and affecting the health of the organ (Cowled and Fitridge, 2011). These models provide information on the underlying mechanisms that induce AKI and will help in elucidating the role of identified protein markers in this AKI study.

1.4) Protein biomarkers for AKI detection

To ensure early detection of TDF-associated nephrotoxicity, it is necessary to find new methods for diagnosing kidney injury, other than measuring increased serum creatinine levels and/or decreased eGFR and urinary output to assess kidney function (De Waal *et al*., 2017; Venter *et al*., 2018; Fabian *et al*., 2022). New diagnostic tools can be developed through the discovery, verification, and validation of biomarkers (Geyer *et al*., 2017). Biomarkers are biological molecules (or their metabolites) that are found in the body serving as vital medical signs, that can be accurately measured and reproduced (Strimbu and Tavel, 2010). These molecular markers are objective indicators of normal biological, pathogenic and pharmacologic processes (Strimbu and Tavel, 2010), where patient response to treatment can be evaluated or the risk of a patient developing a medical condition can be determined (Geyer *et al.,* 2017). Biological markers can be found at a tissue, cellular or molecular level, and in bodily fluids (Strimbu and Tavel, 2010).

There are various types of biomarkers, including diagnostic biomarkers which are used in the screening, diagnosing and monitoring of disease progression (Geyer *et al.,* 2017) Ideal diagnostic biomarkers are usually proteins that can be easily measured in an accessible medium such as blood or urine and are sensitive in identifying disease at a primitive stage, as well as in distinguishing diseased patients from healthy patients (Mayeux, 2004; Sakhuja, 2023). Furthermore, these protein biomarkers would ideally identify an individual's susceptibility to a disease or condition, even in asymptomatic cases (Mayeux, 2004). Lastly, a preferrable diagnostic biomarker would have high specificity for the diseased condition without influence of the effects of co- morbidities (Mayeux, 2004; Sakhuja, 2023). Consequently, the discovery and use of AKI-related diagnostic biomarkers would aid in diagnosing and resolving TDF-associated AKI, before or soon after disease onset, and identify individuals who are at risk of TDF-associated AKI, in South Africans living with HIV.

Although serum creatinine remains the biomarker of choice for assessing kidney function, it is limited in that serum creatinine levels are influenced by multiple external factors (such as sex,

age, and medications, to name a few) (Schmid *et al*., 2015; Sakhuja, 2023), and in the case of AKI, serum creatinine levels are more indicative of functional changes of the kidney (i.e., changes in the eGFR), which occurs as a delayed response to the structural changes that already occurred at the onset of AKI (Vaidya *et al*., 2008; Schmid *et al*., 2015; Kashani *et al*., 2017; Sakhuja, 2023). As a result, over the past years, a few novel biomarkers for AKI have been evaluated in clinical trials, namely kidney injury molecule-1 (KIM-1), interleukin-18 (IL-18), insulin-like growth factor-binding protein 7 (IGFBP-7), tissue inhibitor of metalloproteinase-2 (TIMP-2), N-acetyl-β-D-glucosaminidase (NAG), retinol-binding protein (RBP) and cystatin-C, to name a few (Vaidya *et al*., 2008; Siew *et al*., 2011; van Duijl *et al*., 2019).

Known AKI biomarkers are listed in Table 1.2, and although promising, these markers are under investigation for clinical use and as a result, not many new biomarkers have been introduced for diagnosing AKI (Schmid *et al*., 2015; Kashani *et al*., 2017; Zou *et al*., 2022). This further emphasizes the need for discovering AKI biomarkers, especially within the South African context, that are sensitive and specific in identifying nephrotoxicity and resulting tubular injury in PLHIV.

***Table 1.2: List of published markers of AKI**. The protein biomarkers listed below are published in literature and/or sourced from markerDB but are under clinical investigation for patient use.*

Protein Name	Protein Description	Reference(s)
NGAL_HUMAN	Neutrophil Gelatinase-associated lipocalin	Vaidya *et al*., 2008, Siew *et al*., 2011, Schmid *et al*., 2015, Kashani *et al*., 2017, van Duijl *et al*., 2019, Wishart *et al*., 2021
HAVR1_HUMAN	Hepatitis A virus cellular receptor/ Kidney injury molecule-1	Vaidya *et al*., 2008, Siew *et al*., 2011, Schmid *et al*., 2015, Kashani *et al*., 2017, van Duijl *et al*., 2019, Wishart *et al*., 2021
IL6_HUMAN	Interleukin 6	Siew *et al*., 2011, Kashani *et al*., 2017
IL8_HUMAN	Interleukin 8	Kashani *et al*., 2017
IL18_HUMAN	Interleukin 18	Vaidya *et al*., 2008, Siew *et al*., 2011, Schmid *et al*., 2015, van Duijl *et al*., 2019, Wishart *et al*., 2021
OGA_HUMAN	N-acetyl-beta-D-glucosaminidase	Vaidya *et al*., 2008, Siew *et al*., 2011, Schmid *et al*., 2015, van Duijl *et al*., 2019, Wishart *et al*., 2021
FABPL_HUMAN	Liver fatty acid binding protein	Vaidya *et al*., 2008, Siew *et al*., 2011, Schmid *et al*., 2015, Kashani *et al*., 2017
GSTA3_HUMAN	Glutathione transferases	Siew *et al*., 2011
ALBU_HUMAN	Albumin	Vaidya *et al*., 2008
TRFE_HUMAN	Serotransferrin	Vaidya *et al*., 2008, Wishart *et al*., 2021
IGG1_HUMAN	Immunoglobulin gamma-1 heavy chain	Wishart *et al*., 2021
RET4_HUMAN	Retinol-binding protein 4	Vaidya *et al*., 2008, van Duijl *et al*., 2021
CYTC_HUMAN	Cystatin-C	Vaidya *et al*., 2008, Siew *et al*., 2011, Schmid *et al*., 2015, Wishart *et al*., 2021, van Duijl *et al*., 2021
PEDF_HUMAN	Pigment epithelium-derived factor	Vaidya *et al*., 2008, Wishart *et al*., 2021
B2MG_HUMAN	β-2 microglobulin	Vaidya *et al*., 2008, van Duijl *et al*., 2021
TIMP2_HUMAN	Tissue inhibitor of metalloproteinases-2	Kashani *et al*., 2017, van Duijl *et al*., 2019
IBP7_HUMAN	Insulin-like growth factor-binding protein-7	Kashani *et al*., 2017, van Duijl *et al*., 2019
CLUS_HUMAN	Clusterin	Vaidya *et al*., 2008
CRIP1_HUMAN	Cysteine-rich protein	Vaidya *et al*., 2008
OSTP_HUMAN	Osteopontin	Vaidya *et al*., 2008
VEGFC_HUMAN	Vascular endothelial growth factor	Wishart *et al*., 2021
HGF_HUMAN	Hepatocyte growth factor	Wishart *et al*., 2021
CXL10_HUMAN	Chemokine interferon inducible protein 10	Vaidya *et al*., 2008, Wishart *et al*., 2021, van Duijl *et al*., 2021
S10A8/9_HUMAN	Calprotectin (Protein S100-A8 and S100-A9 complex)	Kashani *et al*., 2017
ANGT_HUMAN	Angiotensinogen	Kashani *et al*., 2017
Creatinine	Chemical Biomarker	Vaidya *et al*., 2008, Wishart *et al*., 2021
Urea	Chemical Biomarker	Wishart *et al*., 2021
Urine microRNAs	miR-21, miR-101-3p, miR-127-3p, miR-210-3p, miR-126-3p	Kashani *et al*., 2017

	miR-26b-5p, miR-29a-3p, miR-146a-3p, miR-27a-3p, miR93-3p, miR-10a-5p	

1.5) Mass spectrometry-based proteomics for biomarker discovery in plasma

1.5.1) Background on proteomics

Proteomics is the study and application of tools that identify and quantify proteomes, and their functional activities within a biological system (Wilkins *et al*., 1996; Aslam *et al*., 2017). A proteome refers to the overall protein content that is or can be produced in a cell, tissue, organ, or in biofluids (Wilkins *et al*., 1996; Graves and Haystead, 2002; Aslam *et al*., 2017). Proteomes can fluctuate over time and can differ from cell to cell or organ to organ in an individual; proteomes can also change in response to external stimuli (Wilkins *et al*., 1996; Graves and Haystead, 2002; Aslam *et al*., 2017).

Proteomics is an interdisciplinary field combining biochemical and analytical chemistry techniques, as well as computational biology and bioinformatics, to name a few (Aebersold and Mann, 2003; Sinha and Mann, 2020). Proteomic experiments generally assess and gather data on protein location, abundance/turnover and post-translational modifications (Graves and Haystead, 2002; Orchard and Charles, 2022). These data can be used directly or to infer further information depending on the experimental design (Graves and Haystead, 2002; Orchard and Charles, 2022).

The profiling of the proteome and its perturbations, provides an accurate representation of the functional state of the biological system and provides comprehensive information on underlying molecular mechanisms of normal physiological and pathophysiological conditions (Cho, 2007; Gupta and Kumar, 2013; Sinha and Mann, 2020). In clinical research, proteomic profiling enables the identification and detection of diagnostic biomarkers for early diagnosis of disease, disease prognosis and monitoring disease progression, amongst other uses (such as drug development and monitoring response to treatment) (Graves and Haystead, 2002; Aslam *et al*., 2017; Sinha and Mann, 2020). Quantitative proteomic profiling can assess the relative differences in protein abundance levels between diseased and healthy patient samples, with the objective of developing a clinically viable diagnostic assay (Graves and Haystead, 2002; Aslam *et al*., 2017; Sinha and Mann, 2020).

One of the major challenges in proteomic analysis, is the complexity of proteomes. The dynamic nature of proteomes is attributed to the post-translational modifications (PTMs) on proteins, the alternative splicing of mRNA to produce different proteoforms as well as the formation of protein complexes (Kavallaris and Marshall, 2005; Hongzhan *et al*., 2007; Chandramouli and Qian, 2009). Additionally, unlike nucleic acids, proteins cannot be amplified and thus, a loss of protein in any sample preparation and analysis stages can poorly affect protein identification and quantification (Cho, 2007; Nature Methods Editorial, 2023).

Furthermore, the large dynamic range of proteins (e.g. > 10^{10} human plasma proteome) present in different sample types (e.g., cells, tissues or biofluids), increases the difficulty in evaluating the proteome in depth and can leave many low abundant proteins undetected (Kavallaris and Marshall, 2005; Chandramouli and Qian, 2009; Nature Methods Editorial, 2023).

1.5.2) Mass spectrometry-based proteomics

1.5.2.1) Mass spectrometry and related technologies

Mass spectrometry (MS) is currently the most comprehensive approach to routinely explore proteomes (Aebersold and Mann, 2003; Schubert *et al*., 2017; Sinha and Mann, 2020). MS allows for the detection, identification, and quantification of proteins within a sample based on the mass-to-charge (m/z) ratio of charged peptide/protein particles and the corresponding relative abundance of these charged peptide molecules using mass spectrometers (Patterson and Aebersold, 2003; Parker *et al*., 2010). MS is a highly sensitive technique, that reliably allows for the accurate measurement of analytes, even at attomolar [10^{-18}] ranges in some mass spectrometers (Forsgard *et al*., 2009; Parker *et al*., 2010; Schubert *et al*., 2017). Mass spectrometers can typically quantify hundreds of proteins per sample (Aslam *et al*., 2017) and have three main components: the ion source, mass analyser(s), and detector (Sinha and Mann, 2020).

Electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) are commonly used soft ionization techniques that produce gaseous peptide ions for detection (Sinha and Mann, 2020). Upon ionization, gaseous peptide ions enter the mass analyser and the ions (now referred to as precursor ions) are separated according to their m/z ratios within an electric and magnetic field (Pitt, 2009; Krisp and Molloy, 2017; Sinha and Mann, 2020). Mass analysers differ in how they separate molecular ions, and common mass analysers used in proteomic research are quadrupoles in combination with time-of-flight (TOF) or Orbitrap analysers (Sinha and Mann, 2020). The mass detector then measures the relative abundance of each precursor ion that it encounters, resulting in MS spectra (Sinha and Mann, 2020).

In tandem MS, three or more mass analysers that are coupled together; one of the mass analysers functions as a collision cell found between the other mass analysers (Mittal, 2015; Sinha and Mann, 2020). Precursor ions enter the first mass analyser and are separated accordingly. Selected precursor ions move into the collision cell, where fragmentation at the lowest energy bond (peptide bond) occurs, through collision with an inert gas like nitrogen or argon (Sinha and Mann, 2020). The fragmented ions then enter the next mass analyser, are

separated by their m/z ratios, and detected by the mass detector, resulting in MS/MS spectra (Mittal, 2015; Sinha and Mann, 2020).

Various methods can be applied to a mass spectrometer to acquire data, such as data-dependent and independent acquisition methods (DDA and DIA) in untargeted proteomics experiments (see section 1.5.2.4) (Ignjatovic *et al*., 2019; Sobsey *et al*., 2020). The selected MS acquisition method is important as it can affect factors such as reproducibility, selectivity, and the limit of detection (Sinha and Mann, 2020). The information provided on protein identity, structure, and chemical modifications indicates that MS can manage the complexities associated with proteomes (Han *et al*., 2008; Parker *et al*., 2010). Underlying molecular mechanisms can be elucidated with minimal bias, as all proteins in a sample are directly measured simultaneously (Schubert *et al*., 2017; Sinha and Mann, 2020).

1.5.2.2) Other proteomic techniques and related technologies

Proteomic techniques besides MS analysis, can also be used within a proteomics workflow. The technique of choice will largely depend on the research question and the complexity of the samples at hand, among other things (Chandramouli and Qian, 2009; Aslam *et al*., 2017; Beeton-Kempen, 2020). Techniques such as western blotting, protein microarray chips, and enzyme-linked immunosorbent assay (ELISA) rely on antibodies available to target specific proteins in a sample (Cho, 2007; Chandramouli and Qian, 2009; Aslam *et al*., 2017). Protein microarray chips are high-throughput immunoassays that can provide expression as well as functional protein data (Cho, 2007; Chandramouli and Qian, 2009; Aslam *et al*., 2017).

Immunoassay techniques are limited by the number of analytes that can be measured, as antibodies are not commercially available for every conceivable analyte, and as a result investigators may need to produce their own, which is a labour-intensive and costly process (Sakamoto *et al*., 2018), especially to the scale with which mass spectrometers can detected analytes. Recently, Olink platforms have become popular in plasma proteomics workflows. These immunoassay-based technologies use antibodies to bind to target proteins and generate a signal. This technology is advantageous for analysis of plasma proteome because it uses small volumes (1 – 2 µl) of plasma (and other biofluids), has sample preparation that is quick, is easily scalable, eliminates the need for fractionation and/or depletion processes, and can measure many samples and proteins (>5 400 proteins) simultaneously over 10 orders of magnitude in concentrations (Wang *et al*., 2023, 2022; Yesavage, 2023).

Gel-based methods such as two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) separates protein samples by charge (first dimension) and molecular weight (second dimension). Differential gel electrophoresis (2D-DIGE) separates protein samples labelled with different fluorescent dyes similarly to 2D-PAGE, and sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) separates denatured protein samples by molecular weight (Cho, 2007; Chandramouli and Qian, 2009; Aslam *et al*., 2017). Other techniques are chromatography based and separate proteins from complex mixtures by their charge, molecular weight, or reversible affinity binding between specific ligands (Cho, 2007; Chandramouli and Qian, 2009; Aslam *et al*., 2017). Currently, these gel- and chromatography-based techniques can be used to prepare protein samples for MS analysis (Chandramouli and Qian, 2009; Aslam *et al*., 2017; Beeton-Kempen, 2020).

1.5.2.3) Top-down vs bottom-up MS-based proteomic workflows

MS-based proteomic workflows are separated into two groups: top-down and bottom-up proteomic workflows (Han *et al*., 2008; Al-Amrani *et al*., 2021). The top-down proteomics approach uses MS analysis to identify and characterize intact proteins. Charged intact proteins are introduced into the mass spectrometer allowing for direct matching of peptide fragments to the protein sequence, and thus the ability to characterize the associated PTMs, protein isoforms and complexes (Wehr, 2006; Han *et al*., 2008; Catherman *et al*., 2014; Schubert *et al*., 2017; Al-Amrani *et al*., 2021). The workflow generally consists of protein separation, MS detection and data analysis (Wehr, 2006; Catherman *et al*., 2014). However, large-scale proteome-wide analysis of intact proteins is challenging, due to the lack of adequate separation techniques for on-line prefractionation of complex samples and the low efficiency of fragmentation methods used in tandem MS (Wehr, 2006; Han *et al*., 2008; Catherman *et al*., 2014; Al-Amrani *et al*., 2021). High performance mass spectrometers that support high resolution, mass accuracy and sensitivity are crucial for separation and accurate identification and is usually associated with high instrument cost (Wehr, 2006; Han *et al*., 2008; Catherman *et al*., 2014; Schubert *et al*., 2017; Al-Amrani *et al*., 2021).

Bottom-up proteomics is the more widely used proteomics workflow (Wehr, 2006; Schubert *et al*., 2017; Al-Amrani *et al*., 2021). The bottom-up proteomics approach identifies and characterizes peptides derived from enzymatic digestion using MS. The m/z ratio and the specific fragment ion pattern for each peptide ion detected are used to reliably identify peptides present in the sample to infer protein identity (Wehr, 2006; Han *et al*., 2008; Schubert *et al*., 2017; Al- Amrani *et al*., 2021). The general workflow usually consists of tryptic digestion of samples, which increases sample complexity; separation of peptide samples by liquid chromatography methods, followed by MS analysis of peptide ions and data analysis (Wehr,

2006; Han *et al*., 2008; Schubert *et al*., 2017; Al-Amrani *et al*., 2021).

Unlike with top-down proteomics, protein digestion prior to MS analysis allows for high-throughput analysis of complex samples, supported by up-front gel and chromatographic-based separation methods (Wehr, 2006; Han *et al*., 2008; Schubert *et al*., 2017; Al-Amrani *et al*., 2021). However, the digestion of intact proteins can result in a loss of valuable information because not all peptides generated will be measured, therefore a limited protein sequence is evaluated and regions containing PTMs and encoding for protein isoforms may be lost (Wehr, 2006; Han *et al*., 2008; Schubert *et al*., 2017; Al-Amrani *et al*., 2021). This also contributes to the protein inference problem, where the under-sampling of peptides in bottom-up proteomics results in identical peptide sequences being shared by multiple homologous proteins, making it challenging to infer protein identity accurately and hinders comprehensive proteome characterization (Huang *et al*., 2012; Ivanov *et al*., 2021).

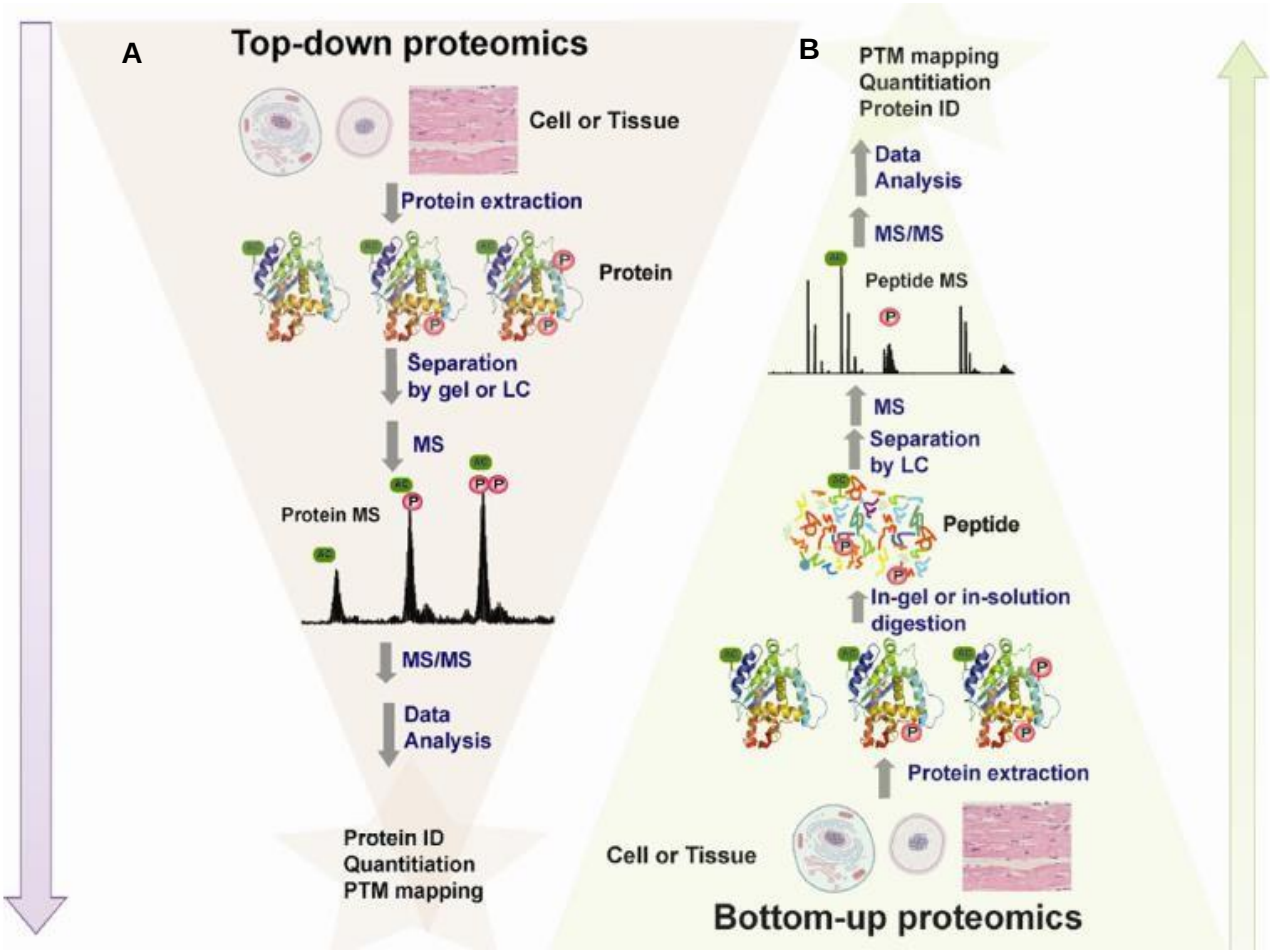


Figure 1.6: Schematic image of top-down(A) and bottom-up (B) proteomics workflows. *Proteins are extracted from cells or tissues and in A) the proteins are separated by gel or liquid chromatography (LC) and analysed by MS, whereas in B) the extracted proteins are digested in-gel or in-solution, and peptides are separated by liquid chromatography (LC), then analysed by MS. Protein identification and quantification then follows. Figure taken from Gregorich et al., 2014*.

1.5.2.4) Targeted vs untargeted MS-based proteomic workflows

MS-based proteomics is ideal for the discovery, verification, and validation of novel biomarkers in the proteome, and is separated into untargeted and targeted proteomics (Ignjatovic *et al*., 2019; Sobsey *et al*., 2020) (Figure 1.7). The untargeted approach has conventionally been used for biomarker discovery, with the intention of identifying as many proteins as possible in a hypothesis-free manner (Schubert *et al*., 2017; Ignjatovic *et al*., 2019; Sobsey *et al*., 2020). Two main data acquisition strategies are employed for untargeted workflows: data-dependent acquisition (DDA) and data-independent acquisition (DIA).

In DDA workflows, the mass spectrometer is pre-programmed according to user specifications, to select a specified number of the most abundant or intense precursor ions within a specific m/z range for detection (MS) or for fragmentation and further separation in the next mass analyser in tandem MS, where the MS/MS spectra are matched to corresponding peptide sequences using sequence database searching (Gillet *et al*., 2012; Krisp and Molloy, 2017; Sinha and Mann, 2020). There are generally more peptide ions to be analysed than what the provided analysis time can accommodate, and due to the stochastic nature of this acquisition method, some precursor ions (especially low abundant precursor ions) go undetected (MS) or unselected for fragmentation (MS/MS), generating missing values (Krisp and Molloy, 2017; Sinha and Mann, 2020). DDA method is limited in the reproducibility of quantified peptides as repeated experiments produce different results and it can also make fragmentation of low abundant peptide ions a challenge (Schubert *et al*., 2017; Ignjatovic *et al*., 2019) (Figure 1.8).

DIA workflows enable mass spectrometers to acquire data by consistently detecting all precursor ions present in relatively large, predetermined isolation windows along the m/z dimension (Gillet *et al*., 2012; Schubert *et al*., 2017). Sequential Window Acquisition of all Theoretical fragment-ion spectra or SWATH-based MS is a specific DIA method, where all precursor peptide ions of a specific m/z range are selected using smaller m/z windows across the MS data produced; this ensures deeper coverage of all detectable precursor ions, especially those of low abundance (Gillet *et al*., 2012; Ignjatovic *et al*., 2019; Sinha and Mann, 2020). SWATH-MS greatly depends on the reproducibility of the sample preparation and results in reproducible measurement of proteomes, with few missing values (Pitt, 2009) (Figure 1.8).

The targeted approach favours the verification of biomarker development and is hypothesis-driven with the aim of analysing a few pre-selected proteins of interest (Schubert *et al*., 2017; Ignjatovic *et al*., 2019; Sobsey *et al*., 2020). For targeted approaches such as selected/multiple reaction monitoring MS (SRM/MRM-MS) or parallel reaction monitoring MS (PRM-MS), mass

spectrometers are programmed to select a set of predefined peptide and peptide fragment ions for analysis to provide consistent and accurate quantification (Schubert *et al*., 2017; Ignjatovic *et al*., 2019; Sobsey *et al*., 2020) (Figure 1.7).

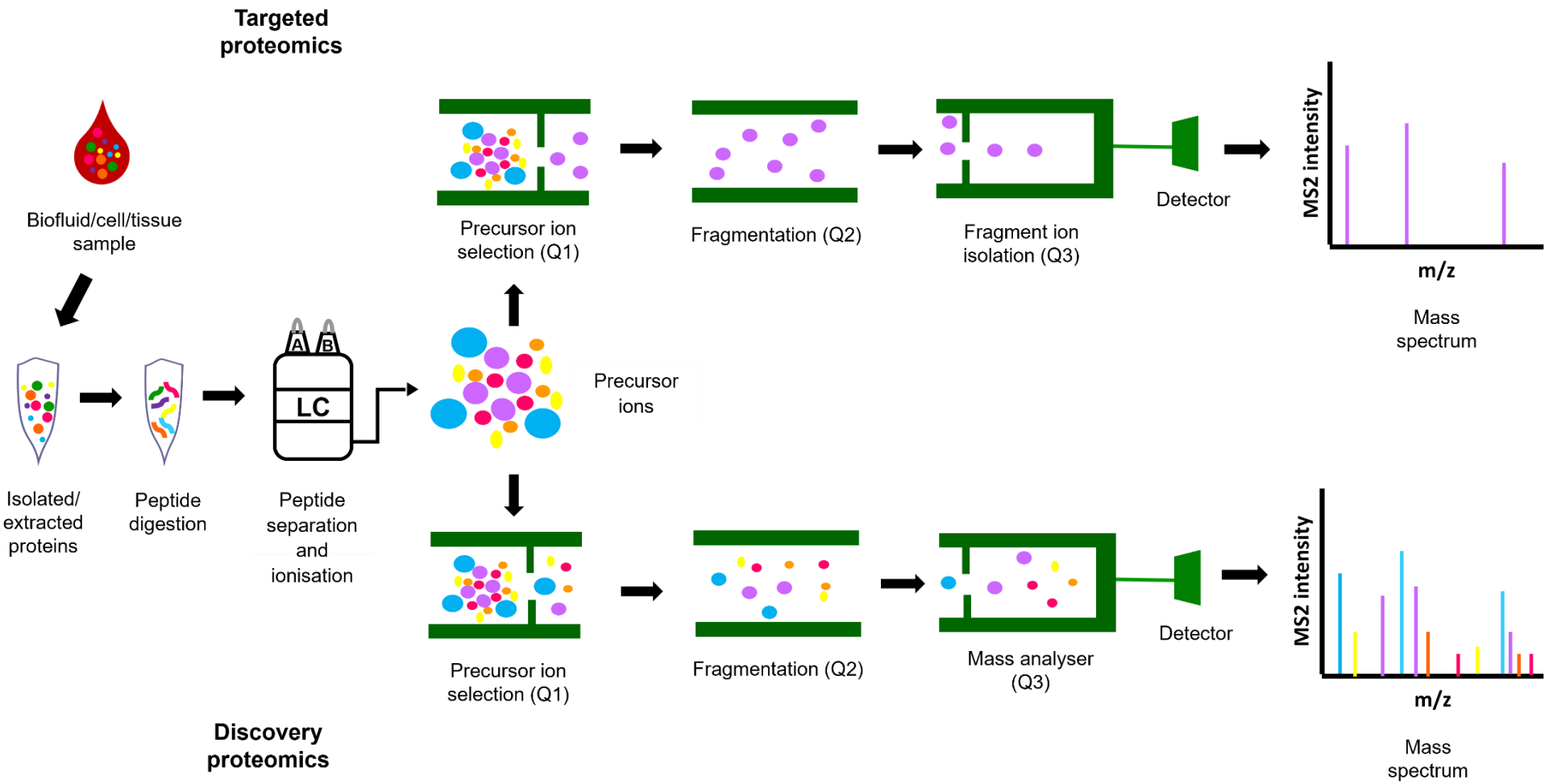


Figure 1.7: Diagram of overview of targeted and untargeted (discovery) proteomics (bottom-up) workflow in plasma.

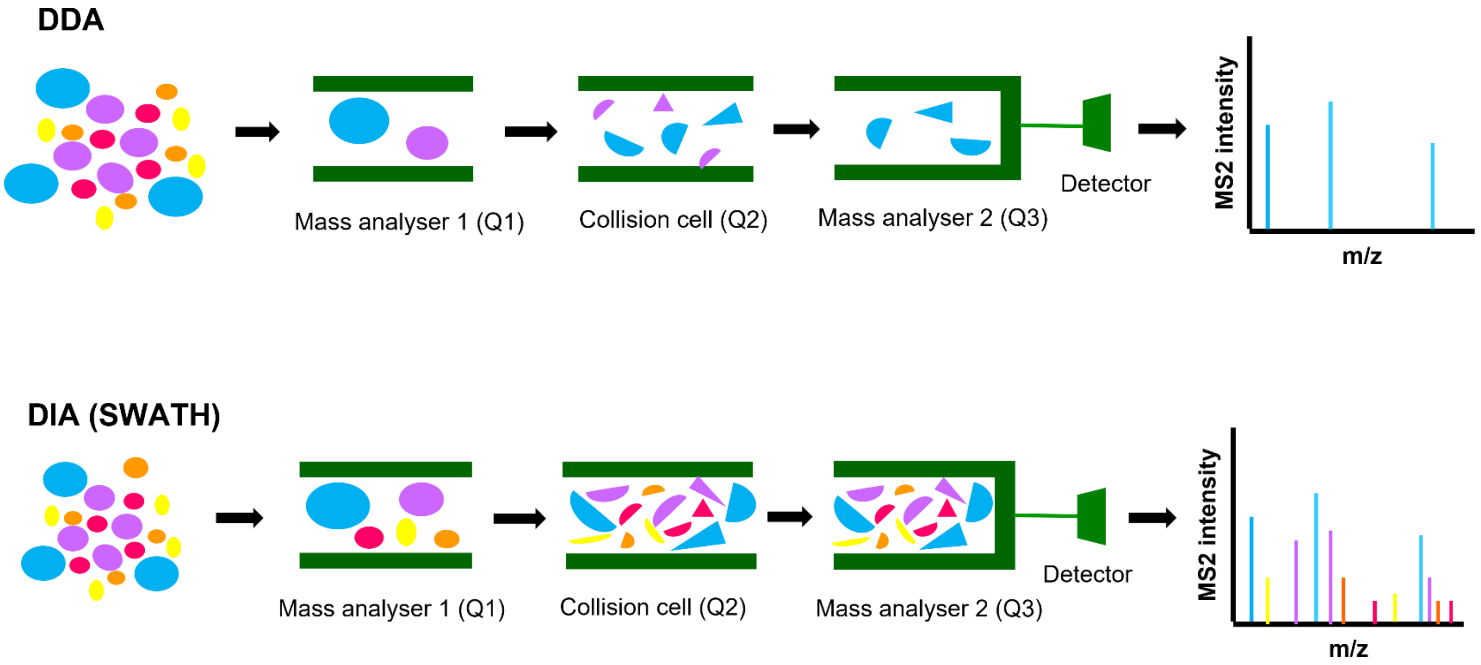


Figure 1.8: Schematic diagram of DDA and DIA (SWATH) acquisition methods during untargeted tandem mass spectrometry analysis.

1.5.3) MS-based quantification strategies

Mass spectrometry (MS) can produce qualitative and quantitative proteomics data (Zhang *et al*., 2010; Gale *et al*., 2014; Rozanova *et al*., 2021). Quantification can occur at a protein (top-down workflow) or peptide (bottom-up workflow) level and can be separated into relative and absolute quantification (Zhang *et al*., 2010; Gale *et al*., 2014; Rozanova *et al*., 2021). Relative quantification results from the comparison of hundreds or thousands of proteins between two or more samples, whereas absolute quantification results from measuring the exact concentration of a specific protein within a sample (Zhang *et al*., 2010; Gale *et al*., 2014; Rozanova *et al*., 2021). To ensure accurate quantitation, unwanted experimental variation introduced during sample preparation and analysis must be minimised and the workflow must be optimised for reproducibility (Zhang *et al*., 2010).

Relative quantification can be performed using label-based (e.g. chemical, metabolic, and enzymatic labelling) and label-free methods (Zhang *et al*., 2010; Rozanova *et al*., 2021), while absolute quantification of a target protein can be determined by using internal standards (e.g. heavy-labeled reference peptides or proteins, at a known concentration) (Zhang *et al*., 2010; Ignjatovic *et al*., 2019). The most common methods of quantitative proteomics use isobaric labelling, which consists of tandem mass tags (TMT) (Thompson *et al*., 2003) and isobaric tag for relative and absolute quantification (iTRAQ) (Ross *et al*., 2004), label-free quantification, stable isotope labelling of amino acids in cell culture (SILAC) (Ong *et al*., 2002), and multiple reaction monitoring (MRM).

1.5.3.1) Label-based quantification strategies

Label-based quantification techniques incorporate artificial labelling of proteins and peptides (Rozanova *et al*., 2021). In metabolic labelling, stable isotopes are introduced into the cell culture media (SILAC) when the sample is still metabolically active, thus labelling proteins during their normal growth cycle (Zhang *et al*., 2010; Wasinger *et al*., 2013; Rozanova *et al*., 2021). This allows samples to be pooled together before subsequent sample preparation steps to be prepared and analysed as one sample, and in this way reduces the technical variation of the workflow (Zhang *et al*., 2010; Wasinger *et al*., 2013; Rozanova *et al*., 2021). The SILAC method employs heavy and light arginine and/or lysine amino acids in the culture media, and by comparing the heavy and light peptides produced, small differences in protein abundance can be monitored because of the small technical variation associated with the method (Zhang *et al*., 2010; Wasinger *et al*., 2013; Rozanova *et al*., 2021).

In chemical labelling, heavy stable isotopes (e.g., ^{13}C, ^{15}N, ^{18}O, ^{2}H) and isobaric labels (e.g. TMT and iTRAQ) can be introduced to samples during sample preparation (Zhang *et al*., 2010;

Rozanova *et al*., 2021). TMT and iTRAQ are popular methods that label peptides and allow for multiplexing of up to 10 and 8 samples into a single pooled sample for LC-MS analysis, respectively (Zhang *et al*., 2010; Rozanova *et al*., 2021). When tagged samples are analysed by the mass spectrometer, the tags are released upon fragmentation, and their signals are measured. The different tag signals are then correlated to the respective samples prior to multiplexing and relative quantification can be achieved. Multiplexing methods are favourable as they allow for high sample throughput, enabling the analysis of larger sample cohorts and technical variation introduced downstream of peptide labelling can be considered negligible; these isobaric tags are limited by digestion efficiency and the efficiency of added labels (Zhang *et al*., 2010; Wasinger *et al*., 2013; Rozanova *et al*., 2021).

Absolute quantification commonly employs synthetic peptides labelled with heavy stable isotopes that are incorporated into the peptide sample to serve as internal standards (of known concentration) (Wasinger *et al*., 2013; Rozanova *et al*., 2021). The synthetic labelled peptides are identical in sequence and have the same physicochemical properties as the peptides naturally present in the sample and are analysed using MRM. Upon fragmentation of selected peptides (sample and synthetic), their product ions can be used to determine the absolute abundance of the peptide based on the measured signal of the synthetic peptide; this provides quantification for peptide fragment ions and as a result, multiple peptides are required from the sample protein to extrapolate the absolute protein quantity (Wasinger *et al*., 2013; Rozanova *et al*., 2021).

1.5.3.2) Label-free quantification strategies

Label-free quantification approaches rely on the comparison of proteins or peptides detected between independent LC-MS or LC-MS/MS runs and require a highly reproducible and robust LC system (Wasinger *et al*., 2013; Rozanova *et al*., 2021). Label-free quantification is categorised by two methods, which are spectral counting and measuring the area under the curve (intensity-based quantification) (Wasinger *et al*., 2013; Rozanova *et al*., 2021).

Spectral counting relies on counting the number of peptide ions or fragment ions acquired for a particular protein, such that the greater the number of peptide ions and/or fragment ions detected during analysis, the greater the abundance of the protein (Megger *et al*., 2013; Rozanova *et al*., 2021). Spectral counting is limited by its inability to measure small changes in protein abundance, thus restricting the quantification of low abundance proteins and that it does not measure physical data, relying only on counts for quantification (Bantscheff *et al*., 2007; Rozanova *et al*., 2021).

The area under the curve (AUC) or peak of each ionised peptide eluted from the LC column at a specific retention time is quantified and this measured signal intensity correlates with the ion abundance (Neilson *et al*., 2011; Rozanova *et al*., 2021). With high-resolution mass

spectrometers, this method allows for measurements that accommodate a wide dynamic range and high precision (Rozanova *et al*., 2021). When conducting differential analysis, the same ions are measured across samples at a specific retention time and m/z, allowing for their AUCs to be calculated and compared. This intensity-based quantification method is considered more accurate and sensitive than spectral counting, especially when using a high-resolution mass spectrometer (Rozanova *et al*., 2021) and is relatively easy to perform and analyse (Wasinger *et al*., 2013). However, it is important to take into consideration that shifts in the retention time, co-eluting peptides and the decrease in signal over time, can negatively impact accurate AUC analysis. To account for these potential hurdles, the retention time can be realigned, the LC gradient time can be increased, and the spectral data can be normalized (Neilson *et al*., 2011; Rozanova *et al*., 2021).

1.5.4) Considerations for MS-based proteomics biomarker discovery

There are general considerations that need to be taken into regard prior to executing any MS-based proteomics biomarker discovery study, such as the study design, sample collection and preparation, quality control measures for sample analysis, as well as the tools that will be used for data analysis.

1.5.4.1) Study design

A well-designed plasma proteomics study requires a clear research question with a hypothesis based on theoretical deductions. The appropriate tools and statistical tests selected for hypothesis testing must be noted. Additionally, study-specific factors such as the study settings, cohort sample size and replicates, cohort variables, and reference samples, are to be clearly defined (Krzywinski and Altman, 2014; Ignjatovic *et al*., 2019).

The study setting refers to the type of investigation that will be conducted on the proteome, for example, investigating the effects of a therapeutic drug on the proteome or investigating a disease(s) of interest in a case-control study (Ignjatovic *et al*., 2019; Nakayasu *et al*., 2021). The appropriate reference samples should be noted, i.e., appropriate control groups are needed for the diseased groups or treated samples (Ignjatovic *et al*., 2019).

With regards to the cohort sample size, cohorts with < 100 plasma samples were commonly assessed but in recent times, the introduction of automated technologies and the advancement of high-throughput methods has enabled larger cohorts (> 100 plasma samples) to be analysed (Ignjatovic *et al*., 2019) for example, a plasma proteome profiling study conducted with 1005 plasma samples took 15 weeks to complete the wet-lab experiments (Cominetti *et al*., 2016). The larger the sample size, the better powered the studies would be and this aids in distinguishing the

biological signal from the technical variability. Unfortunately, the larger the sample size, the more instrument time is required, and more variability is potentially introduced into the study (Zhou *et al*., 2012; Ignjatovic *et al*., 2019).

The use of replicates within a plasma proteomics study is essential in ensuring confidence in and reproducibility of the observed effects in the study. There are two types of replicates, technical and biological (Bell, 2016). Technical replicates refer to repeated measurements of the same sample to evaluate the variation associated with the protocol, instrumentation, and user (Bell, 2016). Biological replicates refer to the measurement of biologically different samples with the intention to identify biological variation in the study (Bell, 2016). To assess the variability and efficiency associated with the MS-based proteomics workflow, a preliminary study should be performed on a smaller group of patient samples (Hassan *et al*., 2006).

Additional variables (such as age, sex, and ethnicity) associated with the cohort are also to be accounted for, to ensure that (bio)markers discovered are disease-specific and not biased to a particular sex, racial group, or age range (Ignjatovic *et al*., 2019). With large cohorts, it is recommended to group biological samples into randomized and balanced blocks, this requires matching of variables (such as age, sex, and ethnicity); this allows for comparison across blocks as the bias associated with the variable is reduced (Ignjatovic *et al*., 2019; Burger *et al*., 2021).

1.5.4.2) Sample collection and preparation

1.5.4.2.1) Sample collection

Sample collection can be a source of variation within a study and should be performed in a standardized manner (Geyer *et al*., 2019; Ignjatovic *et al*., 2019). With this in mind, it is important that standard operating procedures (SOPs) are strictly followed to ensure reproducibility in sampling (Ignjatovic *et al*., 2019; Nakayasu *et al*., 2021). In the case of MS-based plasma proteomics, blood collection requires collection tubes with anti-coagulant to inhibit blood clotting for plasma generation as well as trained clinical staff to ensure blood collection is done accurately and safely (Ignjatovic *et al*., 2019; Nakayasu *et al*., 2021). Additionally, temperature changes during blood draw from 37 °C to 20 °C can potentially contribute to changes in the proteome, as well as the time of day the sample is collected (Ignjatovic *et al*., 2019). Other factors like the needle gauge size during blood collection, the number of times each patient is sampled, the inter-technician/nurse variability, and sample storage time before downstream processing need to be considered prior to blood collection (Ignjatovic *et al*., 2019; Nakayasu *et al*., 2021).

Collected blood is centrifuged twice to ensure the removal of all cellular components, leaving behind plasma (Hassis *et al*., 2015; Ignjatovic *et al*., 2019; Nakayasu *et al*., 2021). Factors like the

centrifugation speed and duration, the temperature during centrifugation, delay in preparation of plasma, storage and potential protein degradation by unstable proteases can alter the proteome composition (Zander *et al*., 2014; Hassis *et al*., 2015; Ignjatovic *et al*., 2019). Once stored (usually at - 80 °C for long-term storage) subjecting plasma samples to multiple freeze-thaw cycles during analysis can increase variation (Zander *et al*., 2014; Ignjatovic *et al*., 2019).

1.5.4.2.2) Sample Preparation

Sample preparation is a critical part of the MS-based proteomics workflow, as it is prone to variability and can significantly affect the quality and reproducibility of downstream MS analysis (Zhou *et al*., 2012; Ignjatovic *et al*., 2019). In a bottom-up LC-MS/MS plasma proteomics workflow, the typical steps include dilution of crude plasma with buffers containing detergent or precipitation with organic solvents (such as acetone) to dilute out or remove contaminants in the plasma samples, respectively (Luque- Garcia and Neubert, 2007; Ignjatovic *et al*., 2019; O'Rourke *et al*., 2023). Due to the large dynamic range of plasma proteins, depletion, enrichment, or fractionation protocols can be integrated into the workflow to remove the top abundant proteins, enrich the low abundant proteins, or reduce the sample complexity before further sample preparation, respectively (Luque-Garcia and Neubert, 2007; de Jesus *et al*., 2017; Ignjatovic *et al*., 2019). The steps that follow are reduction of disulphide bonds and alkylation of cysteine residues to ensure proteins remain unfolded for optimal enzymatic digestion downstream (Müller and Winter, 2017; Ignjatovic *et al*., 2019).

Protein samples need to have the incompatible detergents and solvents removed prior to enzymatic digestion and MS analysis, to ensure optimal protein digestion, peptide ionization, and MS analysis (Ignjatovic *et al*., 2019; O'Rourke *et al*., 2023; Rischke *et al*., 2023). Trypsin is commonly used for protein digestion, and the resulting peptides are separated using liquid chromatography (LC) to reduce sample complexity, then ionized using electrospray ionization (ESI) to produce gaseous peptide ions for tandem MS (MS/MS) analysis (Ignjatovic *et al*., 2019; O'Rourke *et al*., 2023; Rischke *et al*., 2023).

1.5.4.3) Quality control measures for LC-MS/MS analysis

As mentioned in section 1.5.4.1, LC-MS/MS experiments are subject to large variation which is introduced into the experiment at various stages from sample collection to MS/MS analysis. Variation is also expected due to sample degradation and instrument drift over time (Bittremieux *et al*., 2018). The different sources of variation within these workflows can influence each other and can generate bias in the results.

Large-scale proteomic studies require that experimental conditions be consistent to ensure that the data acquired can be accurately compared (Bittremieux *et al*., 2018; Pino, 2019). Preventative measures such as regular instrument maintenance are necessary to minimize technical variation; additionally, quality control (QC) measures can also be employed to monitor and control for existing variation (Bittremieux *et al*., 2018; Pino, 2019). There are various applications for MS and MS/MS analysis and as a result, having a single approach to standardizing QC is difficult. QC samples that can be incorporated in this MS-based proteomics workflow range from simple peptide mixtures (digest of a single protein or a few proteins) to complex cell lysates and are used to monitor the performance of the LC-MS/MS system (Bittremieux *et al*., 2018; Pino, 2019). The simple peptide mixture (QC1) is analysed frequently (a few times per day) and usually interleaved systematically between the biological samples to quickly evaluate LC- MS/MS system by looking at the peak signal, width, and retention time of the peptides (Bittremieux *et al*., 2018; Pino, 2019). QC1 is analysed using shorter gradients to minimize excessive instrument time (Bittremieux *et al*., 2018; Pino, 2019).

The complex cell lysate, e.g., Hela cell lysate (QC2) should be analysed under the same conditions the biological samples are analysed to evaluate the suitability of LC-MS/MS system for complex sample analysis. QC2 is not analysed frequently (once per day), as it would consume more valuable instrument time (Bittremieux *et al*., 2018; Pino, 2019). QC2 is usually integrated into the workflow at the beginning and end of each batch of experiments (Bittremieux *et al*., 2018; Pino, 2019).

1.5.4.4) Statistical analysis of data

With regards to identifying differential abundance of proteins between treatment/study groups, it is important that the appropriate statistical tests are performed by a software platform or manually by the user prior to bioinformatic interrogation of the processed data for hypothesis testing. Additionally, due to the hundreds and thousands of features (i.e. proteins or genes) produced by omics datasets, it is crucial that the statistical tests are corrected for the multiple hypothesis problem. The multiple hypothesis problem arises as hypothesis testing occurs for each individual protein identified in the study, and as a result numerous hypothesis tests are performed simultaneously, influencing the error rate, and allowing for increased identification of false positives, resulting in misleading data (Higdon, 2013). In order to maintain a low rate of false positives and still identify as many significant differentially abundant proteins, the false discovery rate (FDR) can be applied to control for the multiple hypothesis problem; the FDR adjusts the resulting statistical p-value for each comparison made (Pike, 2011; Aggarwal and Yadav, 2016).

When investigating statistically significant differences between proteins identified in the study

conditions, statistical tests such as the Student's t-test can be performed, however, the large sets of biological data generated in proteomics studies are challenging to analyse and interpret using classical statistical methods (Ng *et al*., 2023; Torun *et al*., 2023). As a result, machine learning (ML) platforms have also been incorporated into data analysis of proteomic studies, to provide an additional layer of information that can be extracted from the processed data (Ng *et al*., 2023; Torun *et al*., 2023). OmicLearn (v1.4) is a ready-to-use, open-access online ML platform. This is an exploratory tool specifically developed for omics datasets and allows the non-specialist user to explore different options for building preferrable machine learning pipelines for their omics related datasets for the classification of treatment/study groups and thus supports ML exploration for biomarker discovery without the need for initial coding or programming experience unlike with other available ML platforms (Torun *et al*., 2023). Various options for data pre-processing and for selecting an appropriate feature selection method as well as a machine learning model/classifier can be explored, and the performance of the different machine learning pipelines generated is evaluated based on the scores given for **accuracy**, **precision**, **recall**, **F1_score**, **ROC_AUC** and **PR_AUC** (Torun *et al*., 2023).

The accuracy metric provides a ratio of the number of correct predictions against the total number of predictions made by the model (Jiao and Du, 2016; Orozco-Arias *et al*., 2020; Torun *et al*., 2023), and specifically for binary classifications, accuracy refers to the sum of all true positives and negatives predicted against the sum of all true and false positives as well as true and false negatives (Jiao and Du, 2016; Orozco-Arias *et al*., 2020; Torun *et al*., 2023).

Precision refers to the proportion of accurately estimated positive identifications made by the classification models and is represented by the ratio of true positives against the sum of true and false positives, where a model with no false positives would have a precision of 1 (Jiao and Du, 2016; Orozco-Arias *et al*., 2020; Torun *et al*., 2023). The recall metric refers to the proportion of true positives that were accurately identified by the model and is represented by the ration of true positives against the sum of true positives and false negatives, where a model with no false negatives would have a recall of 1 (Jiao and Du, 2016; Orozco-Arias *et al*., 2020; Torun *et al*., 2023). The precision and recall metrices are indirectly proportional to one another, and along with the accuracy metric enable a more insightful evaluation of the performance of classification models explored (Jiao and Du, 2016; Orozco-Arias *et al*., 2020; Torun *et al*., 2023). Additionally, the F1_score represents a weighted average of the recall and precision metrices, where perfect recall and precision result in an F1_score of 1 and an F1_score value of 0 is associated with either a recall or precision ratio of 0 (Jiao and Du, 2016; Orozco-Arias *et al*., 2020; Torun *et al*., 2023).

The ROC_AUC metric refers to the measurement of the area under the curve (AUC) of the receiver operating characteristics (ROC) graph. The ROC curve represents the performance of a binary classification model at different classification thresholds and is used to evaluate a model's

ability to differentiate between two different classes, i.e. the AKI and non-AKI groups, by plotting the true positive rate (TPR) (y-axis) against the false positive rate (FPR) (x-axis) (Jiao and Du, 2016; Orozco-Arias *et al*., 2020; Torun *et al*., 2023). The area under the ROC curve is a measurement of the model's predictive performance in distinguishing between classes, with the aim of having the greatest TPR and the lowest FPR, such that an AUC of 1 would indicate that the model can distinguish between classes 100% of the time (Jiao and Du, 2016; Orozco-Arias *et al*., 2020; Torun *et al*., 2023).

The PR_AUC metric refers to the measurement of the AUC of the precision-recall (PR) curve. The PR curve represents the relationship between precision (or positive predictive value) (y- axis) and recall (or sensitivity) (x-axis) of a binary classification model at different classification thresholds and reflects the overall performance of the model (Jiao and Du, 2016; Orozco-Arias *et al*., 2020; Torun *et al*., 2023). Like with the ROC_AUC, the area under the PR curve measures the model's predictive performance in distinguishing between classes, such that an AUC of 1 would represent a model that successfully distinguishes the classes from each other 100% of the time (Jiao and Du, 2016; Orozco-Arias *et al*., 2020; Torun *et al*., 2023). The PR curves are often preferred when the classes are imbalanced, as their predictive performance is not impacted by the addition of features (i.e. proteins or genes), unlike the ROC curve (Jiao and Du, 2016; Orozco-Arias *et al*., 2020; Torun *et al*., 2023).

Based on the scores observed for the different ML models explored, the preferred ML model will then be applied to for classification of treatment/study groups. Additionally, a confusion matrix can be used to understand the performance metrics by showing the number of samples classified in each class, therefore revealing any misclassifications and the overall predictive accuracy of the ML model (Bhandari, 2020; Markoulidakis *et al*., 2021; Torun *et al*., 2023).

1.5.5) The utility of plasma in clinical proteomics studies

The analysis of blood is a common tool for diagnosing diseases and conditions in a clinical setting (Geyer *et al*., 2016; Ignjatovic *et al*., 2019). The liquid component of blood, i.e., plasma, is profiled using MS-based proteomics to discover potential biomarkers for the diagnosis, prognosis, and the therapeutic response to an illness (Aslam *et al*., 2017; Geyer *et al*., 2017; Ignjatovic *et al*., 2019). The plasma proteome is categorized by three classes of proteins: 1) consists of abundant proteins such as human serum albumin, apolipoproteins, globulins, and blood coagulation proteins, 2) tissue leakage proteins, such as aspartate and alanine aminotransferases (ASAT and ALAT) and tissue-specific proteins, such as cardiac troponins, and 3) signalling proteins such as protein hormones and low abundant proteins like cytokines (Geyer *et al*., 2017). Plasma is a good source of new biomarkers, as it is easily accessible requiring minimal invasiveness, and it circulates throughout the body, serving as an indicator of a patient's health status (Geyer *et al*.,

2016; Ignjatovic *et al*., 2019).

MS-based plasma proteomics is limited by the wide range of protein concentrations (> 10 orders of magnitude), making it difficult to detect low abundant proteins that are easily masked by the high abundant class 1 plasma proteins (Geyer *et al*., 2016; Ignjatovic *et al*., 2019). Additionally, many of the existing plasma biomarkers lack specificity as they could indicate the presence of illness in any organ, unless the biomarker itself is site-specific (Geyer *et al*., 2016; Ignjatovic *et al*., 2019). MS-based plasma proteomics uses methods to enable detection of low abundant proteins, such as depletion workflows to remove the high abundant proteins, enrichment workflows to enrich or concentrate proteins and fractionation workflows to reduce the sample complexity by dividing plasma samples into less complex fractions for MS analysis (Geyer *et al*., 2016; Ignjatovic *et al*., 2019). Although beneficial, the addition of these workflows to the proteomics pipeline can introduce variation and increases sample handling at each stage (Geyer *et al*., 2016; Ignjatovic *et al*., 2019).

Similarly, urine is also a rich source of novel biomarkers, and in the case of renal dysfunction, the urinary proteome directly reflects changes in the kidney and urogenital tract (Wu *et al*., 2010; Gao, 2015). As a result, evaluating the urinary and plasma proteome of the same patients/treatment groups in kidney diseases would provide great insight into the pathophysiological mechanisms at play that contribute to renal dysfunction and disease. A urinary proteome profiling study of the same study groups, i.e. AKI and non-AKI groups, was previously undertaken, and its results will be compared to the plasma proteome profiling data generated in this AKI study. The urinary proteome profiling study is briefly introduced in section 1.6.

1.6) Prior ARV-associated AKI urinary biomarker discovery study in a South African cohort

Previously, an exploratory case-control study was conducted with a clinical cohort of 74 South African PLHIV on first-line ART, with an equal number (37) of patients in the AKI and non-AKI groups. The urinary proteomes of the AKI and non-AKI groups were compared with the aim of identifying markers that are associated with AKI. The urinary AKI study identified proteins that have been previously reported in literature but were not confirmed in the South African population. Additionally, mimecan was identified as a novel candidate marker of AKI not reported prior in literature (Govender, 2019). The urinary proteome profiling data previously generated will be compared to the data produced from the plasma proteome profiling study of the same patient samples in their respective study conditions (AKI and non-AKI groups).

Proteome profiling of both urine and plasma is worthwhile as these biofluids are commonly used in clinical practice for diagnosing illnesses (Mayeux, 2004; Decramer *et al*., 2008;

Vanmassenhove *et al*., 2013; Ignjatovic *et al*., 2019). Urine and plasma are sources of novel biomarkers and are easily accessible sample types (Mayeux, 2004; Decramer *et al*., 2008; Vanmassenhove *et al*., 2013; Ignjatovic *et al*., 2019). In healthy individuals, ~ 70% of the urinary proteome is obtained from the kidneys and urinary tract and, ~ 30% is derived from glomerular filtration of plasma, therefore indicating that like the plasma proteome, the urinary proteome has the potential to represent the health state of an individual (Decramer *et al*., 2008; Vanmassenhove *et al*., 2013).

Unlike with plasma, urine can be non-invasively obtained in large volumes, allowing for repeated sampling over a period of time (Decramer *et al*., 2008; Vanmassenhove *et al*., 2013). Additionally, urinary proteins and peptides are relatively more stable than those in plasma, as protein degradation is likely to be completed by endogenous proteolysis in bladder awaiting voiding of the bladder, whereas in plasma this degradation my still be viable post collection (Decramer *et al*., 2008; Vanmassenhove *et al*., 2013). The urinary proteome has a less complex dynamic range than plasma, with protein concentration spanning ~10 orders of magnitude (Decramer *et al*., 2008; Vanmassenhove *et al*., 2013). Furthermore, the fluctuations of the urinary proteome result in wide variation of the protein content, influenced by diet, intake of fluids and exercise, to name a few (Decramer *et al*., 2008; Vanmassenhove *et al*., 2013).

1.6.1) Brief overview of results from previous urinary proteome profiling of AKI study

The urinary proteome profiling was performed using the SWATH-MS data acquisition method and a total of 217 differentially abundant proteins were identified using a selection criterion: q- value ≤ 0.05 and a $\log_2$FC ≥ 0.58 (FC ≥ 1.5). Gene ontology pathway analysis of the urinary proteins identified, indicated that pathways relating to platelets (i.e., platelet activation, signalling, aggregation, and degranulation) are highly enriched in this study, as well as pathways of the immune system, of haemostasis and of the complement cascade, at an FDR of 1% (Govender, 2019).

Well reported proteins of renal failure such as beta-2- microglobulin (B2MG_HUMAN), retinol-binding protein 4 (RET4_HUMAN) and alpha-1-antitrypsin (A1AT_HUMAN) were identified as differentially abundant proteins in the urinary protein analysis, but the differences were not considered statistically significant. These proteins indicated an increase in their urinary abundance, which correlated to progression of AKI severity from mild-moderate to severe. Other differentially abundant proteins identified in this study and discussed as proteins of interest are cystatin c, pigment epithelium-derived factor, haemoglobin subunit alpha and beta, vitronectin, alpha-2-HS-glycoprotein, carbonic anhydrase 1, fibrinogen beta and gamma chain, antithrombin-III, glutathione peroxidase 3, S100 calcium-binding protein A8 and A9, and uromodulin (Govender,

2019).

1.7) Project Aim

The aim of this biomarker discovery study is to identify and quantify plasma proteins that are differentially abundant and can distinguish between PLHIV in South Africa diagnosed with AKI and those not diagnosed with AKI, to aid in early detection of ARV-related nephrotoxicity. In addition, a comprehensive proteomic map related to ARV-associated AKI will be generated by comparing the plasma and urine proteomic profiles of the same study groups.

1.8) Project Objectives

1) To perform a pilot study to assess the appropriateness and variability of the semi-automated, undepleted plasma sample preparation and analysis workflow.
2) To use SWATH-based MS workflows to generate patient specific proteomic profiles.
3) To apply statistical and machine learning methods to identify proteins that are differentially abundant in plasma that can distinguish between AKI (case) and non-AKI (control) study groups.
4) To compare findings from the plasma proteome profiles with previously determined urinary proteome profiles of the same study groups.

Chapter 2

Materials and Methods

2.1) Materials

All chromatography solvents and buffers were prepared using LC-MS grade water, 0.1% formic acid (FA) in water, isopropanol, and pure acetonitrile (ACN), and were obtained from either Thermo Scientific™ (Massachusetts, USA), Honeywell (North Carolina, USA) or VWR Chemicals (Pennsylvania, USA). Other solvents used in sample preparation and clean-up were prepared with ultrapure (type 1) water (Milli-Q® Purification System, Merck Millipore, Massachusetts, USA). Ampules of pure trifluoroacetic acid (TFA) and formic acid (FA) used in some buffers and solvents were purchased from Thermo Scientific™ (Massachusetts, USA). The ammonium bicarbonate (ABC), ammonium acetate (NH_4Ac), sodium dodecyl sulphate (SDS), tris(hydroxymethyl)aminomethane (Tris), and iodoacetamide (IAA) were purchased from Sigma-Aldrich (Missouri, USA). Dithiothreitol (DTT) was purchased from Sigma-Aldrich (Missouri, USA). The Bicinchoninic Acid (BCA) assay kit was obtained from Sigma-Aldrich (Missouri, USA), and the Pierce Quantitative Colorimetric Peptide assay kit was obtained from Thermo Scientific™ (Massachusetts, USA). The Hela cell digest and Glu-fibrinopeptide β quality control samples were purchased from Thermo Scientific™ (Massachusetts, USA). Sequencing grade trypsin and endopeptidase LysC were purchased from Promega (Wisconsin, USA) and Thermo Scientific™ (Massachusetts, USA), respectively. The magnetic MagResyn HILIC microspheres were purchased from ReSyn Biosciences (Edenvale, South Africa). Protein Lo-Bind plastic tubes (2 ml) were acquired from Eppendorf (Hamburg, Germany). Shallow and deep 96-well KingFisher plates were purchased from Thermo Scientific™ (Massachusetts, USA). Evotips were purchased from Evosep Biosystems (EV2011; Odense, Denmark).

2.2) Study design

Research question:

Can differences in individual plasma protein abundances identified using mass spectrometry-based proteomics, between the AKI and non-AKI groups of PLHIV on first-line ART in South Africa, distinguish the study groups?

Hypothesis testing:

Hypothesis testing (H_0 vs H_1) was performed on the quantitative data acquired from experimental analysis using an unpaired Student's t-test, to determine which hypothesis was statistically significant:

- **H_0**: There is no significant statistical difference in the individual protein abundances between the AKI and non-AKI groups.
- **H_1**: There is a significant statistical difference in the individual protein abundances between the AKI and non-AKI groups.
- Multiple hypotheses testing was corrected using the false discovery rate (FDR) estimation.

Study participants:

- **AKI group:** PLHIV on first-line ART regimen, diagnosed with AKI (according to Kidney Disease Improving Guidelines Outcome or KDIGO)
- **Non-AKI group**: PLHIV on first-line ART regimen with no AKI diagnosis
- The **exclusion criteria** for this study:
    - PLHIV with pre-existing renal complications or prior renal injury
    - PLHIV unwilling to attend follow-up visits at the clinic
    - PLHIV with records lacking information on prior HIV treatment usage

Sample blocking strategy:

- 188 plasma samples (from study participants) were arranged in a randomized block design.
- The AKI and non-AKI samples were matched in each block according to:
    - gender (e.g., females to females),
    - age (ages within a 3-year range where matched),
    - race (e.g., lack participant to black participant, etc.)
    - TDF exposure (e.g., TDF exposed participant to TDF exposed participant)

- Each block had 31 – 43 matched participant samples and were analysed per day over 5 days.
- LC-MS/MS data was acquired as per the blocking strategy (Figure 2.1)

Quality control (QC) measures in block design:

- Hela cell digests and sample peptide pool replicates were analysed in the blocks to assess LC-MS/MS variation over the five days.
- A batch reference sample (pooled crude participant samples) was prepared and added in each block to assess block-to-block variation across the five blocks.
- Glu-fibrinopeptide β (GluFib) is a calibrant sample that was interleaved between participant samples to assess LC-MS/MS performance frequently during analysis.

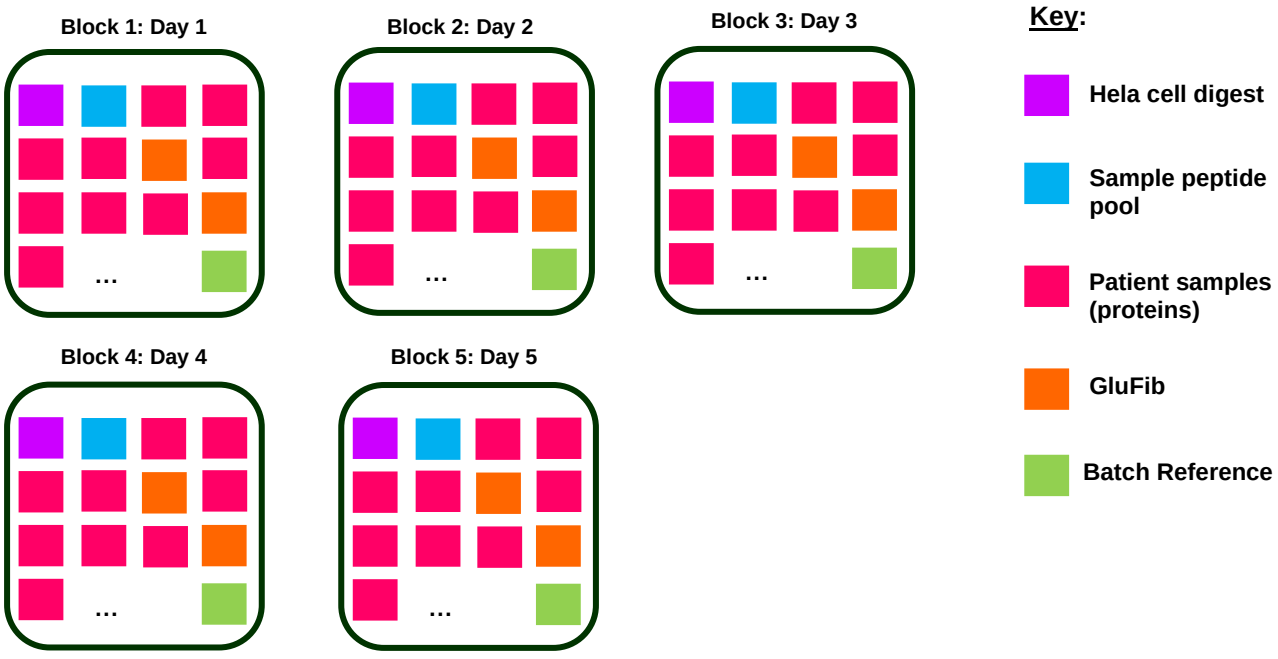


Figure 2.1: Schematic diagram of randomized block design of AKI study. *Participant samples were randomly blocked ensuring that participants in AKI and non-AKI groups were matched according to gender, age, race and TDF exposure. Quality control samples were incorporated and arranged in the five blocks.*

2.3) Overview of experimental procedures

The experimental procedures conducted for the AKI study comprise of sample collection and preparation, sample clean-up and digestion, LC-MS/MS sample analysis, followed by data processing, data and statistical analysis, as well as the exploration of machine learning models (Figure 2.2). Blood was collected from consenting patients that met the inclusion criteria for this study, and plasma was prepared. The plasma samples were subjected to dilution, reduction, and alkylation (see section 2.7).

Thereafter, sample clean-up using the hydrophilic interaction liquid chromatography (HILIC) technique with magnetic microspheres, and tryptic protein digestion took place in an automated sample handling station (KingFisher™ Flex Purification System, Thermo Scientific™, Massachusetts, USA). The resulting peptides were fractionated by 1-dimensional reverse phase liquid chromatography (1-D RPLC) using an Evosep One LC system (Evosep Biosystems, Odense, Denmark) coupled to a Sciex TripleTOF 5600 mass spectrometer (Sciex, Toronto, Canada) using a SWATH-based acquisition method for tandem MS analysis. Raw MS data was searched against a Uniprot FASTA file of reviewed human sequences for peptide identification and protein inference, allowing for the identification of differentially abundant proteins using an unpaired t-test in Spectronaut™ (v17) (Biognosys AG, Schlieren, Switzerland). Additionally, an exploratory machine learning (ML) model was applied to the plasma proteome profiling dataset using the OmicLearn (v1.4) platform (Torun *et al*., 2023), to estimate how well ML- based classification can differentiate the AKI group from the non-AKI group and which important proteins enabled the ML-based classification.

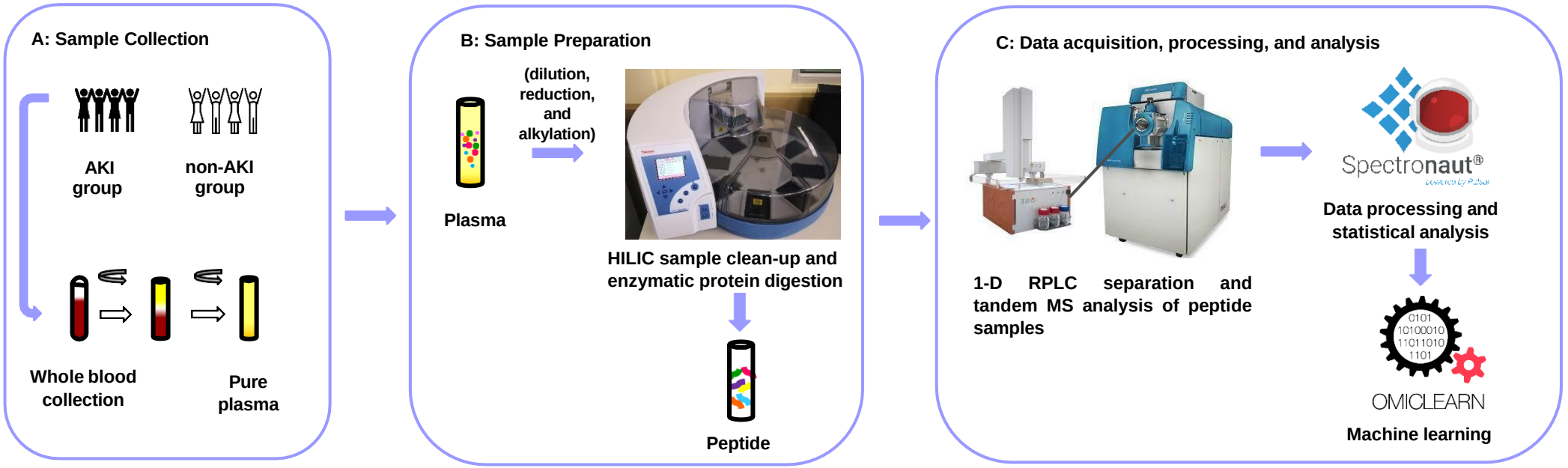


Figure 2.2: Overview of experimental procedures of the AKI study. *The schematic indicates the flow of methodology (guided by the blue arrows) from sample collection (**A**), sample preparation (**B**) as well as data acquisition, processing, and analysis (**C**). The various methods will be explained in detail in sections below.*

2.4) Patient recruitment and sample collection

2.4.1) Patient recruitment

Patient recruitment and sample collection were conducted at the Klerksdorp Tshepong Hospital, in the North West Province of South Africa. The necessary ethical clearance [Ethics references: #58/2013 (CSIR-REC) and #120612 (WITS-HREC)] was obtained to proceed with requesting patient consent and sample collection. Following written and oral consent, blood samples were collected from HIV positive patients in the adult (18 years and older) wards by trained clinical staff. These samples were collected from patients on ART regimens that included TDF and have been diagnosed with AKI, according to the 2014 Clinical Practice Kidney Disease Improving Guidelines Outcome (KDIGO) (Kellum *et al*., 2012; Machado *et al*., 2014). These patients served as the case group. Additionally, samples were collected from patients on ART regimens that included TDF but showed no indication of developing AKI, and this group served as the control group.

Demographic and common renal function clinical data, such as age, race, gender, estimated glomerular filtration rate (eGFR) and serum creatinine (srCr) levels, were collected for data and statistical analysis. In total, 200 patients met the inclusion criteria and 100 patients belonging to the case group were matched according to age (± 3 years), gender, race, and exposure to TDF, to 100 patients in the control group (see section 2.5). Patients that were excluded from this study were those with pre-existing renal complications or had prior renal injury, as well as patients who were unwilling to have follow-up visits and those with patient records lacking sufficient information for determining any prior use of HIV treatment.

Of the 200 patient samples obtained, only 188 samples were prepared and analysed (Figure 2.3). The remaining samples were illegibly labelled and could not be cross-referenced with the demographic and clinical data collected. From the 188 samples analysed in this AKI study, 135 patients were exposed to TDF-ART regimens, with 53 patients in the AKI (case) group and 82 patients in the non-AKI (control) group, and this subset of data was used for data and statistical analysis.

2.4.2) Sample collection

Approximately 9 ml of blood was collected by venipuncture from each patient in ethylenediaminetetraacetic acid (EDTA) collection tubes. EDTA is an anticoagulant and prevents blood clotting by binding calcium in the blood (Gordon Hadley and Larson, 1953; Wiseman *et al*., 1996). Immediately upon drawing blood into the collection tubes, the tubes were inverted 8 – 10 times to ensure adequate mixing of the blood with the anticoagulant and

was placed in wet ice for plasma generation within 1 hour of collection (Gordon Hadley and Larson, 1953; Wiseman *et al*., 1996). The blood samples were centrifuged at 1 000 x g for 10 minutes at 4°C to separate the cellular components of the blood from the plasma (liquid component of blood) (Gordon Hadley and Larson, 1953; Wiseman *et al*., 1996). Approximately 3 ml of plasma was collected per blood vial into a sterile tube (without disturbing the buffy white layer in the tube) and the remaining sample was used to collect red, white, and other blood cells. The plasma tubes were then further centrifuged at 2 500 x g for 10 minutes at 4°C to ensure that all cellular blood contaminants were removed from the plasma. The centrifuged plasma (~ 1 ml) was aliquoted into 1.5 ml cryovials for storage in a - 80°C freezer, with minimal disturbance to the pellets formed. Stored plasma aliquots were used to prepare the samples for downstream MS-based proteomic analysis.

The specific blood collection protocol used was as per the standard operating procedures (SOPs) from the CSIR Next Generation Health Cluster (see section 6.1).

2.5) Use of randomised block design and quality control for sample analysis

2.5.1) Randomised block design

A randomised experimental block design allows for the arrangement of patient samples into groups with equal sample sizes, referred to as blocks (Festing, 2020; Burger *et al*., 2021). This experimental design minimizes the effects of known (e.g., sex, age, or race, etc.) and unknown (e.g., a drug treatment that is unaccounted for during sample collection) confounding variables and ensures that the results obtained are comparable and interpretable across the blocks (Festing, 2020; Burger *et al*., 2021).

In MS-based proteomic experiments, variation can be introduced at various steps along the workflow from sample collection to data processing, and using a randomised block design limits the impact of technical variability that occurs (Ignjatovic *et al*., 2019; Burger *et al*., 2021), for example, during data acquisition the sensitivity of mass spectrometer can change with time due to instrument drift, and a randomised block design minimizes the effect of instrument-specific variation in the number of proteins identified and their measured abundances. In this study, a randomised block design was used, and five blocks were generated (Figure 2.1), with an equal number of cases (AKI group) and controls (non-AKI group) per block. Additionally, variables such as age, sex, and race greatly influence proteomic data, and as a result, the AKI group in each block had a similar number of male and female patients, with all age ranges and races equally represented. The same was done for the non-AKI group in each block. The blocks were analysed sequentially (i.e., block 1 followed by block 2, 3, 4 and 5), with one block analysed per day over five days.

2.5.2) Quality control measures for sample analysis

The performance of the LC-MS system was also evaluated by assessing the peptide peak-derived parameters associated with liquid chromatography (LC), and by analysing replicates of samples of known concentration. In this study, peptides refer to modified fragment ion sequences as per Spectronaut™ (v17).

The LC-derived peak parameters were the number of data points per peak, the peak capacity, and the median full width at half maximum (FWHM) of peaks. The number of data points acquired per peptide peak define the peak shape for accurate quantitation of the area under the curve (AUC) of each peak (Karpievitch *et al*., 2010; Bruderer *et al*., 2017; Hermes *et al*., 2018). Approximately 8-10 data points per peak are considered optimal for reproducible quantitation of the resolved peptide peaks, but for analysis of complex protein mixtures, 5-8 data points per peak are needed for a well-defined peak shape with reproducible quantitation

of the AUC of each peptide peak (Karpievitch *et al*., 2010; Bruderer *et al*., 2017). The peak capacity and median FWHM are important parameters for assessing peak resolution. The greater the peak capacity, the narrower the median FWHM, and this results in good peak separation enabling the observation of more peptide peaks (Karpievitch *et al*., 2010; Bruderer *et al*., 2017; Hermes *et al*., 2018).

Furthermore, replicates of a Hela cell digest, and a sample peptide pool were prepared and analysed in the blocks (one of each replicate per block) to assess LC-MS/MS variation at the beginning of each day over the five days of analysis (Figure 2.1). The Hela cell digest, and the sample peptide pool served as quality control samples. A stock of Hela cell digest was prepared and diluted with solvent A (LC-MS grade water with 0.1% FA) and 0.1 µg of the Hela cell digest was injected into the mass spectrometer for analysis. To prepare the sample peptide pool, 2.5 µl of sample peptide digest from each patient sample in block 1 was pooled together (an average of ~ 1.35 µg of sample peptide digest) into a 2 ml Lo-Bind Eppendorf tube and aliquoted to prepare replicate samples after thorough mixing. All Hela cell digest and sample peptide pool replicates were prepared on day 1 of LC-MS/MS analysis.

Other quality control samples were the batch reference replicates and the Glu-fibrinopeptide β (GluFib) calibrant samples. The batch reference stock was prepared to assess block-to-block variation by pooling together 10 µl of crude plasma from each patient sample in block 1 into a 2 ml Lo-Bind Eppendorf tube. The pooled sample was thoroughly mixed and aliquoted to prepare replicate samples, with each replicate allocated to the end of each block (Figure 2.1) for LC-MS/MS analysis along with the patient samples. The batch reference replicates were prepared with the patient samples (see section 2.5 and 2.6).

Lastly, a stock of 5 pmol/µl of GluFib was diluted with resuspension buffer (2% acetonitrile (ACN), 0.2% formic acid (FA); v/v) to 1 pmol/µl in a 2 ml Lo-Bind Eppendorf tube. GluFib is a calibrant sample that was interleaved after every 6 patient samples in each block (Figure 2.1) to frequently assess LC-MS/MS performance and calibrate the mass spectrometer.

2.6) Protein quantification using the BCA assay

The stored crude plasma samples were thawed on wet ice and inverted often to aid the thawing process. Approximately, 5 µl of crude plasma was diluted 10x with buffer (50 mM Tris-HCl, 2% SDS; pH 8) to a final volume of 50 µl and the entire diluted sample was used for protein concentration determination in duplicate, prior to sample preparation. Samples were vortexed and centrifuged to ensure adequate mixing. Protein quantification of samples was performed in the same order as the pre-determined randomised blocks at the beginning of each day of analysis (Figure 2.1).

The total protein concentration of the samples was determined using the Pierce Bicinchoninic Acid (BCA) Protein Assay kit. The BCA protein assay is a colorimetric assay, in which sample proteins reduce Cu^{2+} ions to Cu^{+} ions in an alkaline environment, to form a faint light blue colour; this is referred to as the biuret reaction (Smith *et al*., 1985; Walker, 1996). When BCA is added to the reaction and is in the presence of Cu^{+} ions, the samples develop a purple colour (Smith *et al*., 1985; Walker, 1996). A UV-Vis spectrophotometer plate reader (BioTek PowerWave HT UV-Vis Microplate Reader, BioTek Instruments, Vermont, USA) was used to measure the absorbance of the purple-coloured samples (at 562 nm) to quantify their protein concentration in reference to a standard curve of known protein concentrations (Smith *et al*., 1985; Walker, 1996). The BCA protein assay was performed in duplicates, in a 96-well plate format for increased sample throughput, and the assay kit was used as per manufacturer's instructions.

2.7) Sample preparation

Following protein quantification, ~ 20 µl of crude plasma was diluted 25x with buffer (50 mM Tris- HCl, 2% SDS; pH 8) to dilute out sample contaminants; the SDS within the buffer denatured and unfolded the plasma proteins into their primary structures. Samples were vortexed and centrifuged to ensure adequate mixing. Based on the BCA quantification assay, 20 µg of each diluted plasma sample was added to a fresh 0.5 ml Lo-Bind tube and the sample was topped-up to a final volume of ~ 20 µl with buffer (50 mM Tris-HCl, 2% SDS; pH 8).

Thereafter, the samples were reduced with 10 mM dithiothreitol (DTT) for 30 minutes at 60°C. DTT reduces accessible disulphide bonds in the denatured proteins by a two-step sequential thiol-disulphide exchange reaction, that prevents the formation of disulphide bonds between cysteine residues of a protein or peptide (Gundry *et al*., 2009; Müller and Winter, 2017). Subsequently, the reduced samples were alkylated with 40 mM iodoacetamide (IAA) for 30 minutes at room temperature, in a dark cupboard. IAA is an alkylating agent that covalently binds to the cysteine thiol groups of the reduced proteins, to further ensure that disulphide bonds do not re-form, and the unfolded protein structure can be maintained (Gundry *et al*., 2009; Müller and Winter, 2017). The alkylation process was then quenched with 10 mM DTT. After quenching, the samples were added to a 96 deep-well KingFisher plate as per Figure 2.3.

2.8) Automated HILIC sample clean-up and on-bead protein digestion

Protein digestion and sample clean-up were performed using the KingFisher Flex™, an automated magnetic sample handling station, with magnetic hydrophilic interaction liquid chromatography (HILIC) microspheres from ReSyn Biosciences (Edenvale, South Africa), referred to as MagReSyn® HILIC microspheres. For the MagReSyn® HILIC clean-up process, eight KingFisher plates with 96 deep-wells and their corresponding solvents or buffers were prepared as per Figure 2.3.

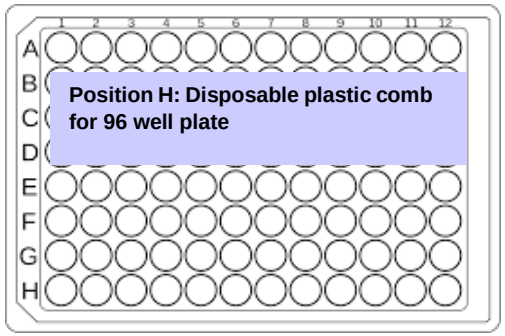
Plate 1 contained a disposable plastic comb to cover the magnetic rods in the KingFisher Flex™ for transfer of the MagReSyn® HILIC microspheres from plate 2 to subsequent plates. The MagReSyn® HILIC microspheres consist of multi-mode functional groups on the surface and make up the polar stationary phase, allowing for a high binding affinity for proteins. The mobile phase consists of acetonitrile miscible in water, which enabled separation of charged, polar and amphiphilic proteins from less polar ones by trapping them in the water-rich layer that forms on the surface of the microspheres. The retention of proteins on the surface of the microspheres may be caused by hydrogen bonding and the weak electrostatic interactions that occur (ReSyn Biosciences, 2022). The MagReSyn® HILIC microspheres were added to plate 2 in a 1:10 amount ratio of protein sample:microspheres and were transferred to plate 3 for equilibration in 100 mM ammonium acetate (NH_4Ac), 15% ACN; pH 4.5 (Figure 2.3).

Equilibrated microspheres were transferred to plate 4, which consisted of the alkylated samples in a 1:1 volume ratio with binding buffer (200 mM NH_4Ac, 30% ACN; pH 4.5), for binding of proteins to the surface of the microspheres. Thereafter, the protein-bound microspheres were transferred to plate 5 and 6 to wash off sample contaminants with 95% ACN. In this way, MagReSyn® HILIC clean-up removes salts, detergents, and buffers that are incompatible with downstream LC-MS/MS analysis (ReSyn Biosciences, 2022) (Figure 2.3).

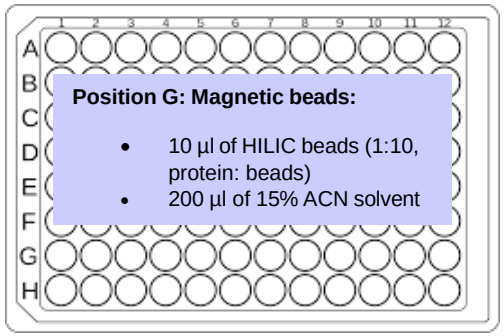
The washed protein-bound microspheres were transferred to plate 7 for on-bead protein digestion with trypsin (1:20; enzyme: protein) and LysC (1:100; enzyme: protein) in 50 mM ammonium bicarbonate (ABC) for 2 hours, at 47°C. Trypsin hydrolyses peptide bonds by cleaving the C-terminal side of lysine and arginine residues (Laskay *et al*., 2013). Other proteases such as, LysC are often paired with trypsin in protein digestion to reduce tryptic missed cleavages and improve digestion at lysine residues (Giansanti *et al*., 2016). Thereafter, the microspheres were transferred to plate 8 for elution of any microsphere-bound peptides, and a final concentration of 1% trifluoroacetic acid (TFA) was used to stop the digestion process (Figure 2.3). Volumes of solvents and buffers added to KingFisher™ plate are detailed in Figure 2.3.

After digestion, the peptides produced were transferred to clean 2 ml Lo-Bind tubes, the samples were stored at - 80°C, and the frozen peptides were then dried by solvent evaporation through sublimation using a vacuum concentrator (CentriVap® Concentrator, Labconco, Missouri, USA). Thereafter, the dried peptides were resuspended in 40 µl of solvent (2% ACN, 0.2% FA; v/v) and quantified (see section 2.9) for accurate and uniform loading of peptides for LC-MS/MS analysis.

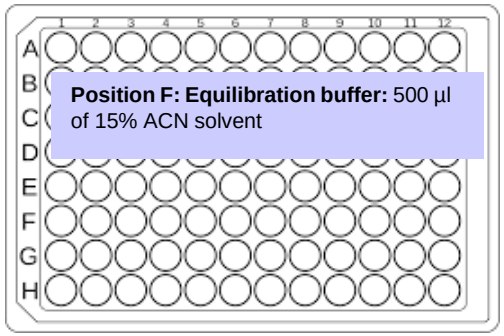
1.



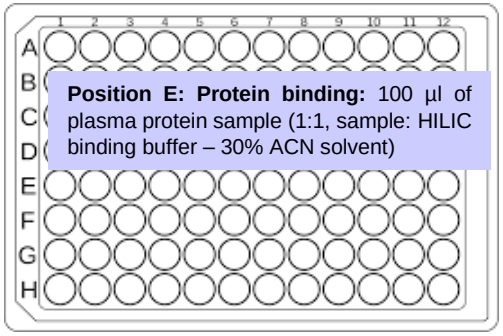
2.



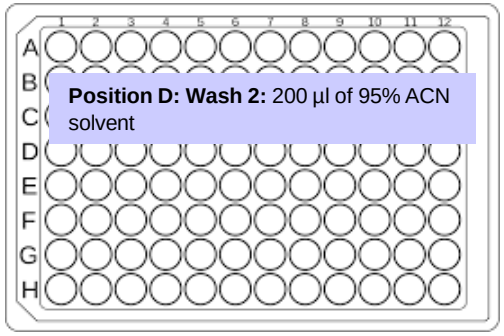
3.



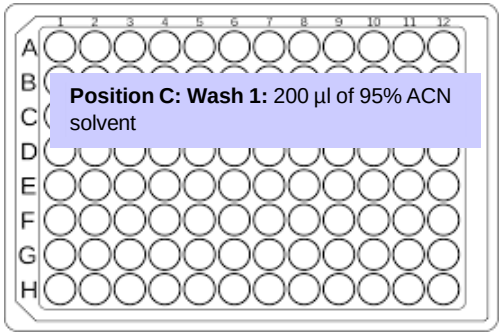
4.



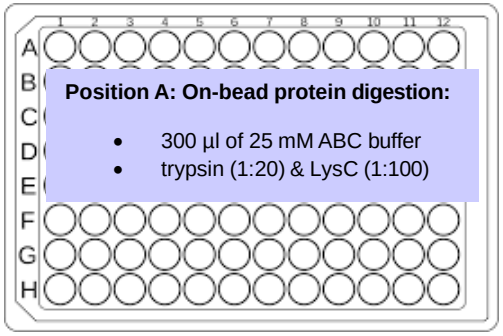
5.



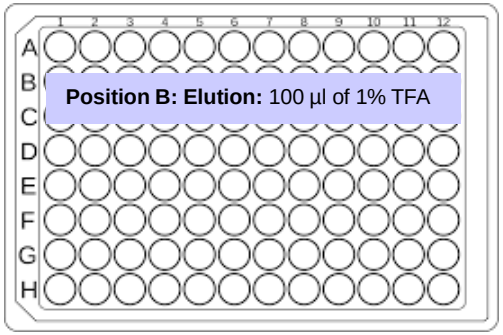
6.



7.



8.





Loading Position for KingFisher Flex™. Plates are loaded from plate 1-8 shown by green arrow.

Figure 2.3: Diagram of 96-well plate positioning in KingFisher Flex™ for sample clean-up and on-bead protein digestion. *Eight 96 deep-well plates shown are labelled 1-8, in order of plate loading sequence as per the in KingFisher Flex™ system prompting.*

2.9) Peptide quantification using Pierce Quantitative Colorimetric Peptide Assay

From each peptide sample, 8 µl of resuspended peptides were diluted with 32 µl of solvent (2% ACN, 0.2% FA (v/v)) into a clean 2 ml Lo-Bind tube and the samples were vortexed and centrifuged to ensure adequate mixing. Peptide quantification of samples was performed in the same order as the pre-determined randomised blocks (Figure 2.1).

The Pierce Colorimetric Peptide Assay kit reagents allow for the biuret reaction to occur, where the amide backbone of peptides reduces Cu^{2+} ions to Cu^{+} ions in an alkaline environment, and when the kit's proprietary chelator is added, it interacts with the Cu^{+} ions to form a bright red complex (Thermo Scientific™, Massachusetts, USA). A UV-Vis spectrophotometer plate reader (BioTek PowerWave HT UV-Vis Microplate Reader, BioTek Instruments, Vermont, USA) was used to measure the absorbance of the samples (at 480 nm) to quantify their peptide concentration in reference to a standard curve of known protein concentrations (Thermo Scientific™, Massachusetts, USA). This peptide assay was performed in single replicates, in a 96-well plate format for increased sample throughput at room temperature, and the assay kit was used as per manufacturer's instructions.

2.10) Acquisition of LC-MS/MS data for the AKI study

Plasma proteome data for the patient samples was acquired in their blocks (Figure 2.1) using an Evosep One LC system (EV-1000, Evosep Biosystems, Odense, Denmark) coupled to a Sciex TripleTOF 5600 mass spectrometer (Sciex, (Sciex, 2017)), that was set in the SWATH acquisition mode. The Evosep One is a standardized liquid chromatography platform with six methods that can analyse up to 30, 60, 100, 200, 300, 500 samples per day (SPD), with individual specific gradient lengths (min), cycle times (min), and flow rates (µl/min) programmed in a plug-and-play set-up (Evosep Biosystems, Odense, Denmark). Evosep One uses disposable Evotips (EV-2011) that resemble pipette tips with a C-18 plug at the base to serve as a trap column for sample loading and desalting, prior to partial elution into the Evosep One. The partial elution associated with the Evotips reduces cross-contamination and variation between sample injections and preserves the life of the columns used.

As per peptide quantification (see section 2.9), 0.5 µg of peptides per sample were loaded onto Evotips that had been washed, conditioned, and kept wet throughout Evotip loading as per manufacturer's instructions, to a final volume of 20 µl. The QC samples (see section 2.5) were also prepared for MS analysis using the Evotips as per the manufacturer's instructions. The eluted peptides were separated on a C-18 Evosep performance column (EV-1109: 8 cm, 150 µm x 1.5 µm) at a flow rate of 1 µl/min. The generic 60 SPD method consisted of a 21-minute gradient, with a preformed gradient of 5 - 30% ACN (solvent B) and an offset gradient with a lower percentage of ACN using solvent A (LC-MS grade water with 0.1% FA) and solvent B (LC-MS grade 100% ACN with 0.1% FA).

Following peptide separation with the Evosep One LC system, the eluted peptides were measured using a Sciex TripleTOF 5600 mass spectrometer, that was operated in a positive, SWATH acquisition mode. The SWATH acquisition method consisted of 48 variable windows that have a 1 m/z overlap between the variable windows (Figure 2.4), and measured 48 MS/MS scans as well as a single MS scan over a mass range of 400 – 1 100 m/z. The size of the variable SWATH windows across the m/z range depend on the number of precursor ions found in that m/z range; SWATH windows with narrower m/z range have a greater number of precursor ions that those with wider m/z ranges (Sciex, 2017). The gradient length was 21 minutes, the period cycle time was 1 062 milliseconds and the accumulation time for the MS and MS/MS scans was 50 and 20 milliseconds, respectively.

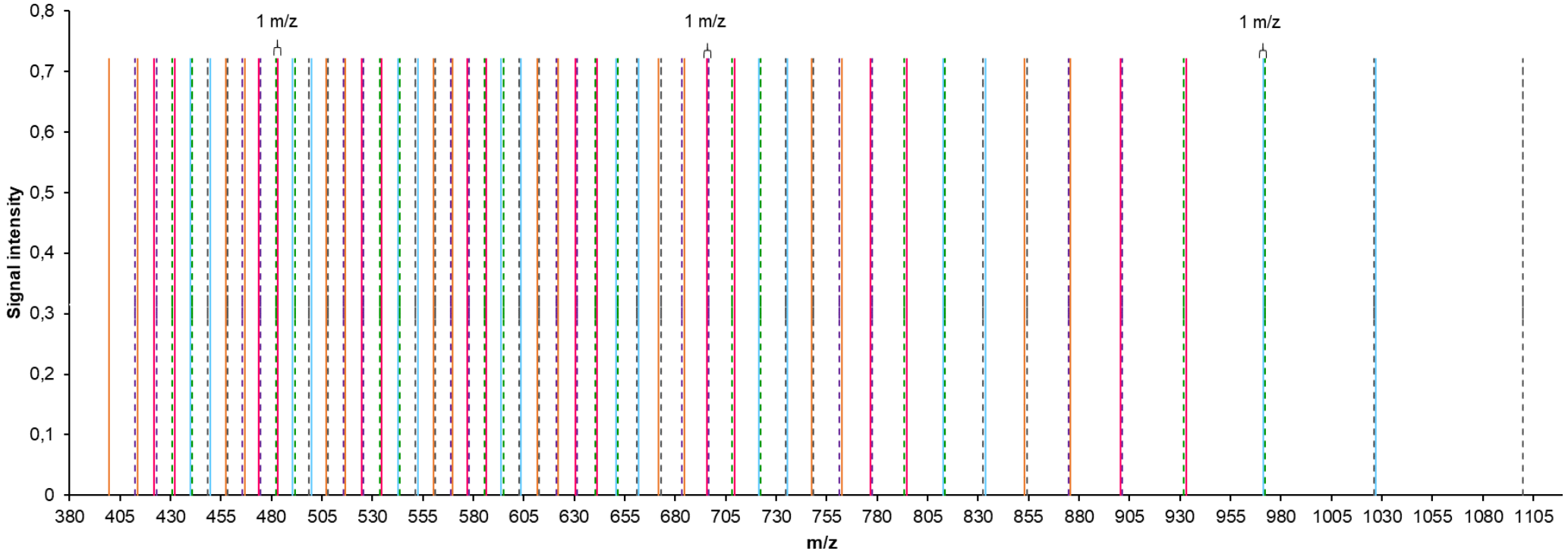


Figure 2.4: Diagram of 48 variable window (VW) schema used for acquiring data with the SWATH acquisition method and an overlap of 1 m/z between the windows over a m/z range of 399 – 1100 m/z.

2.11) Processing of LC-MS/MS patient data using Spectronaut™ 17

The patient data acquired (section 2.10) were processed and analysed by uploading the raw SWATH (.wiff) files onto Spectronaut™ 17, a software independent of mass spectrometer vendor (Biognosys AG, Schlieren, Switzerland). The data were processed using the directDIA+ (Deep) workflow, allowing for the SWATH runs to be searched in reference to a protein database instead of generating a specific spectral library for peptide identification and protein inference. The Biognosys (BGS) default factory settings for searching and extracting MS and MS/MS data were applied: specific enzyme cleavage by trypsin/P and LysC/P allowing for 2 consecutive missed cleavages sites to be missed by the proteases, the fixed and variable modifications were set for cysteine carbamidomethyl and acetyl at protein N-terminal as well as methionine oxidation, identifications were performed at a false discovery rate (FDR) of 1% at a peptide and protein group level, with a minimum peptide length of 7 amino acids, and peptides were filtered by selecting the best scoring fragment ions per precursor ion, with a minimum of 3 and maximum of 6 fragment ions per precursor ion. For quantification, no imputation strategy was applied, the protein LFQ method applied was automatic, quantity type was set to area, i.e. area under the curve (AUC), and quantification was based on MS/MS measurements. Protein inference was performed by the IDPicker inference algorithm. An appropriate database was selected (consisting of all reviewed human protein sequences in a FASTA format, including sequences of expected contaminant proteins, such as keratin polypeptides from Uniprot Knowledge Database, with 20 507 protein entries – downloaded: 6 June 2022) and the study conditions (AKI or non-AKI) were specified.

2.12) Retrospective power analysis

A retrospective power analysis of the acquired data was performed using MSstats (v3.4.1) (Choi *et al*., 2014) in RStudio (v2022.12.0 + 353) (R Core Team, 2022). The power analysis was conducted to determine a desired fold-change (or effect size) for the identified differentially abundant proteins to be considered significant for this study. The retrospective power analysis was performed at a power of 80%, with an FDR of 1%. A conservative sample size of 53 patient samples was used to perform the power analysis so as to not overestimate the desired fold-change, as the study consisted of 53 AKI and 82 non-AKI patient samples.

2.13) Gene ontology functional enrichment and network interaction analysis of candidate markers

A functional enrichment analysis of the candidate markers (i.e. significant differentially abundant proteins) was performed using Enrichr (Chen *et al*., 2013; Kuleshov *et al*., 2016). Enrichr enabled the observation of enriched gene ontology (GO) biological processes, molecular functions, and cellular compartments, as well as enriched pathways from the Reactome, KEGG, and Wikipathways databases associated with the identified candidate markers. Enrichr conducted the functional enrichment analysis by comparing the genes of candidate markers against the total set of human genes present in the various databases available. Enrichr presents the pathways and gene ontology processes and functions ranked by p-value, adjusted p-value, odds ratio, and combined scores from most to least significant (Chen *et al*., 2013; Kuleshov *et al*., 2016). A p-value cut-off of 0.05 was used and significantly enriched GO processes, functions, and pathways were reported.

Clusters of interaction networks displaying the physical subnetwork between the identified candidate markers as well as between proteins not considered to be statistically significant candidate markers was illustrated using the STRING database (STRING-db) (Szklarczyk *et al*., 2015) within Cytoscape (v3.10.1) (Shannon *et al*., 2003), at a confidence interval of 70%. Enrichment data was uploaded to the network for visualisation of enriched gene ontology (GO) biological processes, molecular functions, and cellular compartments, as well as enriched pathways within Cytoscape (v3.10.1).

2.14) Statistical and Machine Learning analysis

Statistical analysis of the patient demographic factors, such as race, sex, and the median age, as well as the clinical AKI characteristics, such as serum creatinine (srCr) levels and estimated glomerular filtration rate (eGFR), was performed using non-parametric statistical tests. The Chi-squared test of independence was used to compare sex and race respectively, to determine if there was a statistically significant difference between the AKI and non-AKI groups. Likewise, the Mann-Whitney U Test (Wilcoxon signed rank test) was performed to compare the median age, srCr levels, and the eGFR, respectively, to determine if there was a statistically significant difference between the AKI and non-AKI groups. A p-value cut-off < 0.05 indicated a statistically significant difference between the two study groups, and the statistical tests were performed manually in excel (v2308) and confirmed using an online platform called Epitools for epidemiological and clinical statistics (Sergeant, ESG, 2018).

The average intensities of the proteins identified in this AKI study were compared across runs in Spectronaut™ (v17) using default settings: differential abundance testing was set to unpaired t-test and identification was performed at a protein q-value cutoff of 5% for each run and 1% across the experiment. The average intensities were log_2 transformed (absolute log_2 was reported in this study) and an unpaired Student's t-test was performed to determine which differentially abundant proteins between the AKI and non-AKI groups were statistically significant. The p-value of the differentially abundant proteins was corrected for the multiple hypothesis testing by controlling the false discovery rate (FDR) and a q-value was then computed for the differentially abundant proteins identified.

Using the OmicLearn platform (v1.4) (Torun *et al*., 2023), machine learning (ML) classifiers and other parameters, such as feature selection methods and imputation of missing values, were explored to build appropriate ML pipelines. Within the OmicLearn platform (v1.4), the feature tables were imported using the Pandas package (v1.0.1) and manipulated using the Numpy package (v1.18.1), the ML pipelines were built using the scikit-learn package (v0.22.1) and ML was completed in Python (v3.8.8) (Geyer *et al*., 2021; Torun *et al*., 2023). Firstly, the 'no imputation' strategy was applied to the dataset, with the XGBoost classifier (the only available classifier when using data with missing values) and the performance of five feature selection methods, ExtraTrees, k - best (mutual_info_classif), k - best (f_classif), k - best (chi-squared) and none was evaluated. All feature selection methods excluding the 'None' method had a default of 20 maximum number of features, and the 'ExtraTrees' method had a default of 100 trees in the forest.

Secondly, 'ExtraTrees' was selected as the best performing feature selection method and with the 'zero imputation' strategy, pipelines consisting of the seven ML classifiers (AdaBoost, LogisticRegression, RandomForest, k-NeighboursClassifier, DecisionTree, LinearSVC, and XGBoost) were explored for the best performing ML model. Lastly, the best performing ML model was compared with a similar model, that substituted the 'zero imputation' strategy for the 'median imputation' strategy to identify the best performing imputation method for the dataset. The ML pipelines were explored at a random state of 23, with no normalization performed (performed prior in Spectronaut™, v17), and the following cross-validation settings: 'RepeatedStratifiedKFold' method with 5 cross-validation splits and 10 cross-validation repeats.

The best performing ML model for the AKI proteomics dataset was selected to generate a list of proteins identified by the ML model as important in distinguishing the AKI from the non-AKI group.

Chapter 3

Result

3.1) Patient demographic details and clinical characteristics of AKI

Statistical analysis of the patient demographic factors, such as race, sex and the median age showed no significant difference ($p > 0.05$) between the AKI and non-AKI groups. All the patients in both AKI and non-AKI conditions were exposed to an ART regimen that included TDF (Table 3.1). Furthermore, the median estimated glomerular filtration rate (eGFR), as well as the median serum creatinine (srCr) levels showed a significant statistical difference ($p \leq 0.05$) between the AKI and non-AKI groups. The AKI group displays an eGFR of 21,25 mL/min/1.73m^2, whereas the non-AKI group has an eGFR of 82,03 mL/min/1.73m^2. The srCr levels in the AKI and non-AKI groups is 276 µmol/L and 67 µmol/L, respectively. This is indicative of kidney dysfunction, and these are the main parameters used to identify the presence of AKI in the study participants (Table 3.1).

Table 3.1: Demographic details and AKI clinical characteristics of study participants in the AKI and non-AKI groups*. The demographic details consist of the sex, median age, and race/ethnicity of study participants. The median clinical characteristics (eGFR and srCr levels) of the patients are also represented in the table.*

Category	Normal ranges	AKI (cases) n = 54 Median [IQR]	non-AKI (controls) n = 82 Median [IQR]	p-value
Sex:				**0,4555**
Females, n (%)	**-**	32 (59,26%)	55 (67,07%)	
Males, n (%)	**-**	22 (40,74%)	27 (32,93%)	
Age:				**0,3774**
In years	**-**	42 [36,25 - 51,75]	40,5 [35 - 52,75]	
Race:				**0,4432**
Black, n (%)	**-**	52 (96,30%)	81 (98,78%)	
Coloured, n (%)	**-**	1 (1,85%)	1 (1,22%)	
Indian, n (%)	**-**	1 (1,85%)	0	
Serum creatinine levels (µmol/L):				**< 0,0001**
	53 - 97,2 (females)	276 [146,25 - 520]	67 [58 - 81,5]	
	61,9 - 114,9 (males)			
Estimated glomerular filtration rate (mL/min/1.73m^2):				**< 0,0001**
	≥ 90	21,25 [9,93 - 35,21]	82,03 [68,58 - 92,88]	

The Chi-square Test of Independence was performed on the sex, and race.

The Mann-Whitney U Test (Wilcoxon signed rank test) was performed on data for the median age, eGFR and srCr levels.

Data for normal ranges indicated were retrieved from Abcar *et al*., 2004 for the serum creatinine levels and from Tarwater, 2011 for the estimated glomerular filtration rate levels.

3.2) Preliminary analysis of the AKI study

The preliminary analysis was conducted on block 1 (23 AKI; 16 non-AKI), to verify the adequacy of the internal LC-MS/MS workflow to identify and quantify differentially abundant plasma proteins. Block 1 was prepared and measured independently from the rest of the cohort for this initial assessment of the LC-MS/MS workflow. Additionally, the preliminary assessment was also performed to indicate which candidate markers and their associated biological processes would be determined in the AKI study. The preliminary data were processed with directDIA search in Spectronaut™ (v17) and an average of 4 465 precursors, 3 175 modified peptides, 3 056 unmodified peptides, 358 protein groups, 409 proteins, and an average of ~ 37% peptide missed cleavages were identified in the AKI and non-AKI samples within block 1 (Table 3.2).

Table 3.2: Overview of the average number of identifications determined in the preliminary analysis. *The average number of precursors, modified and unmodified peptides, protein groups, proteins, and the percentage of missed cleavages are listed.*

Average no. of identifications	Preliminary analysis
Precursor	4 465
Modified peptides	3 175
Unmodified (stripped) peptides	3 056
Protein groups	358
Proteins	409
% of missed cleavages	37%

An unpaired Student's t-test was performed on the preliminary data to identify differentially abundant plasma proteins between the two study conditions. The volcano plot below shows that 65 differentially abundant proteins were identified using the default selection criterion in Spectronaut™ (v17): q-value ≤ 0.05 and an absolute log_2FC ≥ 0.58 (FC ≥ 1.5); these proteins are considered candidate markers. There were 42 differentially abundant proteins that showed increased abundance (pink) and 23 differentially abundant proteins that showed decreased abundance (blue) in the AKI group compared to the non-AKI group (Figure 3.1). The list of the 65 differentially abundant proteins can be found in Table 6.1.

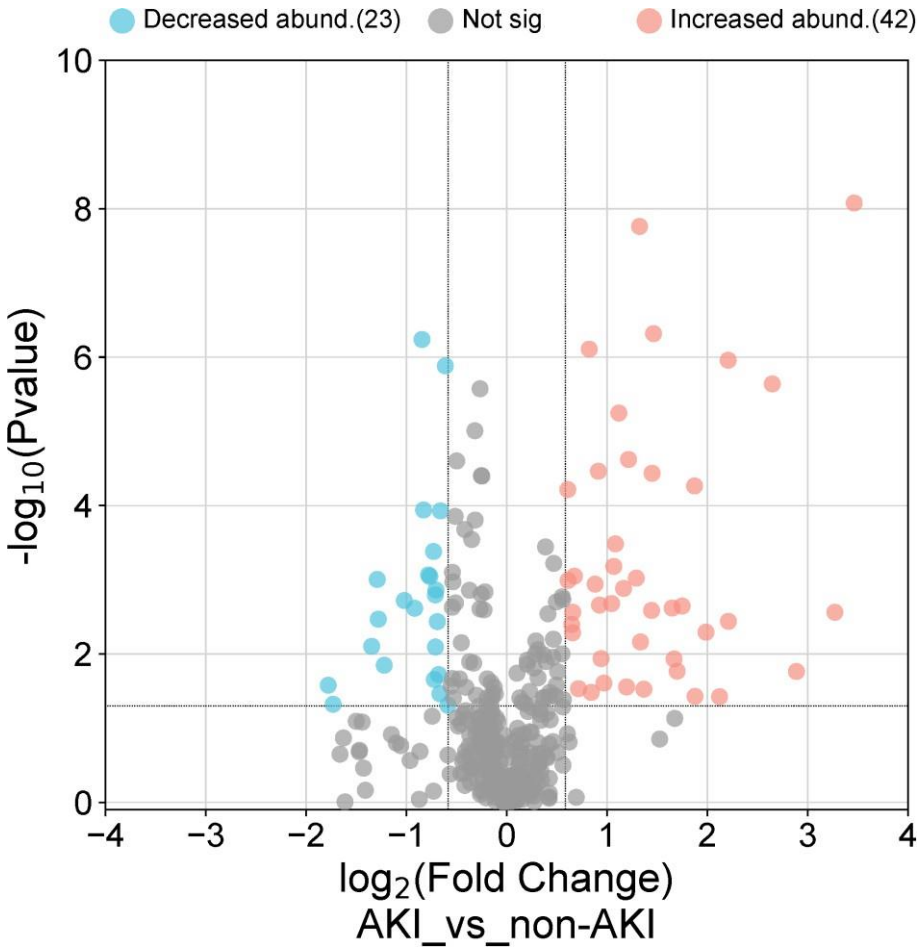


Figure 3.1: Volcano plot of differentially abundant proteins generated from the preliminary analysis*. Candidate markers in pink showed increased abundance and markers in blue showed decreased abundance, in the AKI group compared to the non-AKI group (q-value ≤ 0.05, absolute log_2 FC ≥ 0.58; FC ≥ 1.5). Proteins in grey not significantly differentially abundant. Volcano plot generated in SRplot.*

Additionally, a list of enriched gene ontology (GO) biological processes as well as pathways from the KEGG, Reactome and Wikipathways databases, associated with the genes of the identified candidate markers was generated (Table 3.3). The enriched pathways of signal transduction, the immune system, as well as in cell-cell communication and extracellular matrix organization were identified. Enriched GO biological processes and pathways are tabulated from most to least number of participating candidate markers (adjusted p-value ≤ 0.01).

Table 3.3: GO biological processes and pathway analysis of candidate markers*. Enriched GO biological processes as well as KEGG, Reactome and Wiki pathways associated with the genes of the identified candidate markers are tabulated below (adjusted p-value ≤ 0.01). Data was generated using Enrichr analysis and tabulated in descending order of the number of candidate markers found in each process or pathway name.*

	Name	No. of candidates found in process/pathway	Candidate marker gene names	Adjusted p-value
GO Biological Processes	Defence Response to Bacterium (GO:0042742)	5	*PGLYRP2; DEFA1; RNASE1; S100A9; S100A8*	1,21 E-02
	Leukocyte Aggregation (GO:0070486)	2	*S100A9; S100A8*	1,21 E-02
	Peptidyl-Cysteine Modification (GO:0018198)	2	*S100A9; S100A8*	1,38 E-02
KEGG Pathways	PI3K-Akt signalling pathway (map04151)	5	*COMP; HSP90AA1; FN1; THBS4; EIF4B*	7,50 E-03
	Lipid and atherosclerosis (map05417)	4	*HSP90AA1; APOA1; LBP; ICAM1*	7,50 E-03
	Malaria (map05144)	3	*COMP; THBS4; ICAM1*	6,76 E-03
	ECM-receptor interaction (map04512)	3	*COMP; FN1; THBS4*	7,50 E-03
	IL-17 signalling pathway (map04657)	3	*HSP90AA1; S100A9; S100A8*	7,50 E-03
	Staphylococcus aureus infection (map05150)	3	*FCGR3A; DEFA1; ICAM1*	7,50 E-03
Reactome Pathways	Immune System R-HSA-168256; Innate Immune System R-HSA-168249	12	*PIGR; FCN3; HSP90AA1; PGLYRP2; CFHR4; FN1; SAA2; LBP; RNASE1; S100A9; S100A8; ICAM1*	6,52 E-03
	Signalling By Interleukins R-HSA-449147; Interleukin-4 And Interleukin-13 Signalling R-HSA-6785807	6	*HSP90AA1; FN1; SAA2; LBP; S100A9; ICAM1*	6,86 E-03

	Binding And Uptake of Ligands by Scavenger Receptors R-HSA-2173782; Scavenging by Class B Receptors R-HSA-3000471; Scavenging of Heme from Plasma R-HSA-2168880	4	*CD163; HSP90AA1; APOA1; SAA2*	3,06 E-04
	Toll Like Receptor 4 (TLR4) Cascade R-HSA-166016; Regulation of TLR By Endogenous Ligand R-HSA-5686938	4	*SAA2; LBP; S100A9; S100A8*	6,21 E-03
	RHO GTPases Activate NADPH Oxidases R-HSA-5668599	3	*RNASE1; S100A9; S100A8*	1,10 E-03
	Integrin Cell Surface Interactions R-HSA-216083	3	*COMP; FN1; ICAM1*	7,22 E-03
	Amyloid Fiber Formation R-HSA-977225	3	*H2AC4; APOA1; SAA2*	1,03 E-02
	Antimicrobial Peptides R-HSA-6803157	3	*PGLYRP2; S100A9; S100A8*	1,21 E-02
	Sema3A PAK Dependent Axon Repulsion R-HSA-399954; Chaperone Mediated Autophagy R-HSA-9613829	2	*HSP90AA1; RNASE1*	1,03 E-02
	Metal Sequestration by Antimicrobial Proteins R-HSA-6799990; MyD88 Deficiency (TLR2/4) R-HSA-5602498; IRAK4 Deficiency (TLR2/4) R-HSA-5603041	2	*S100A9; S100A8*	2,64 E-03
Wikipathways	Complement system WP2806	4	*FCGR3A; CFHR4; APOA1; ICAM1*	2,27 E-03
	Vitamin B12 metabolism WP1533	3	*APOA1; SAA2; ICAM1*	3,91 E-03

3.3) Quality control for AKI study analysis

The variability in the performance of the LC-MS system was evaluated over the course of five days. This was done by analysing replicates of the Hela protein digest (Hela_QC) and the sample peptide pool (Peptide_Pool_QC) at the beginning of each day, with each block of patient samples (Figure 2.1). Five blocks were prepared for analysis, consisting of 31 - 43 patient plasma samples, a batch reference sample and a calibrant sample (Glufib_QC) interleaved after every 6 patient samples. Each block was analysed using the 60 samples per day (60SPD) method on the Evosep One (HPLC system).

To assess variability between the Hela_QC and Peptide_Pool_QC replicates, the number of protein groups and peptides identified across the five days was compared, and Pearson correlations of the protein groups and peptides identified between the QC replicates were performed. Hela_QC replicates analysed on days 1, 2 and 3, had a similar number of protein groups and peptides identified, with a slight decline observed on day 4 and a further decline on day 5 (Figure 3.2A). QC_4 and QC_5 have ~ 2.84% and ~ 5.52% less protein groups and peptides identified than QC_1, 2 and 3, respectively (Figure 3.2A). The number of protein groups and peptides identified in the Peptide_Pool_QC replicates was similar throughout the five days of analysis, with ~ 2.39% less protein groups identified in QC_2 and ~ 2.28% less peptides identified in QC_5, compared to the other Peptide_Pool_QC replicates (Figure 3.3B).

A strong positive correlation (≥ 0.93) of the protein groups identified, was observed between the Hela_QC (Figure 3.3A) and Peptide_Pool_QC (Figure 3.3B) replicates. Similarly, the identified peptides had a strong positive correlation (≥ 0.875) between the Hela_QC (Figure 3.3C) and Peptide_Pool_QC (Figure 3.3D) replicates.

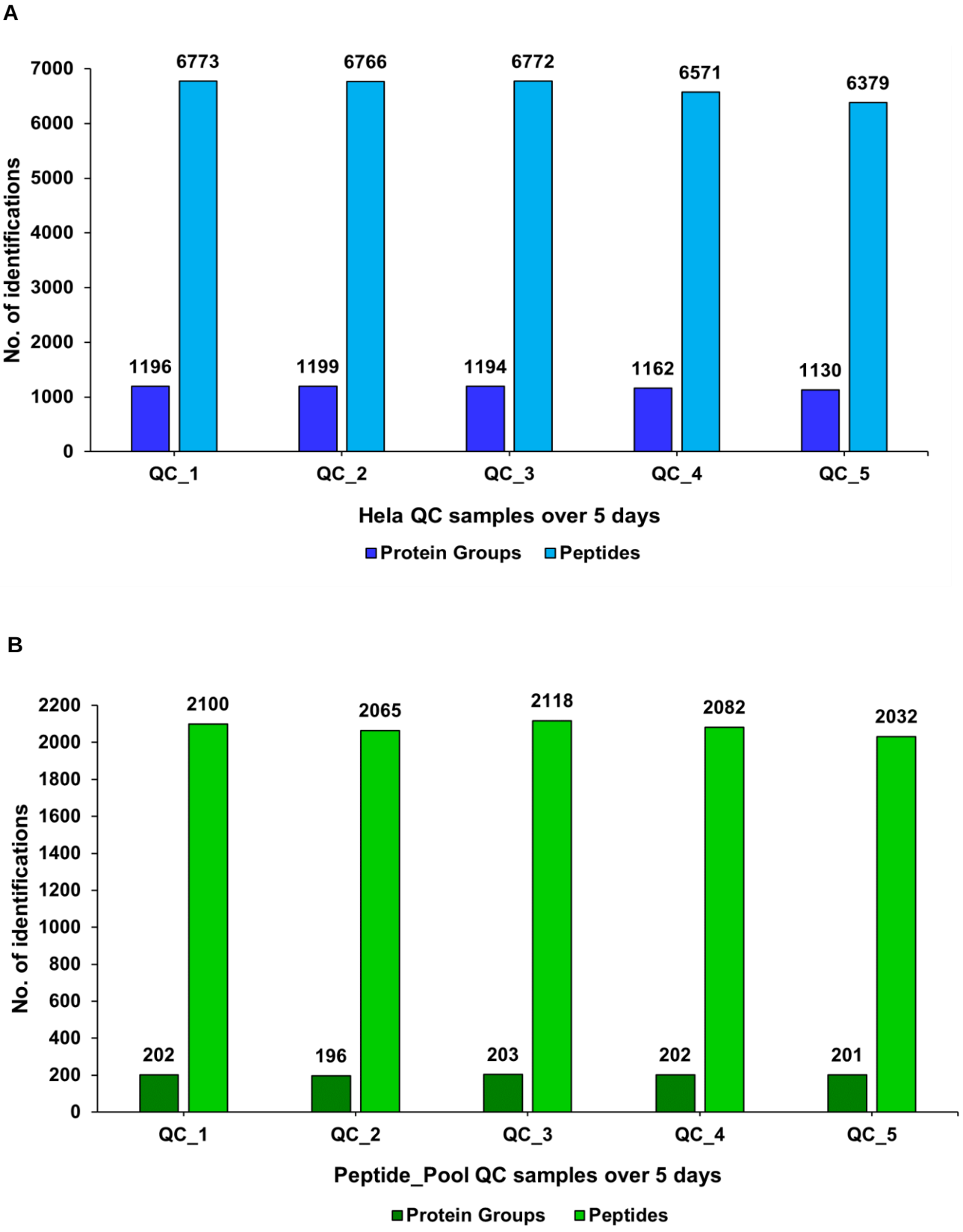


Figure 3.2: Bar graphs indicating the number of protein groups and modified peptides identified in the Hela_QC (A) and Peptide_Pool_QC (B) replicates over 5 days*. **A**) Protein groups in dark blue and modified peptides in light blue. **B**) Protein groups in dark green and modified peptides in light green.*

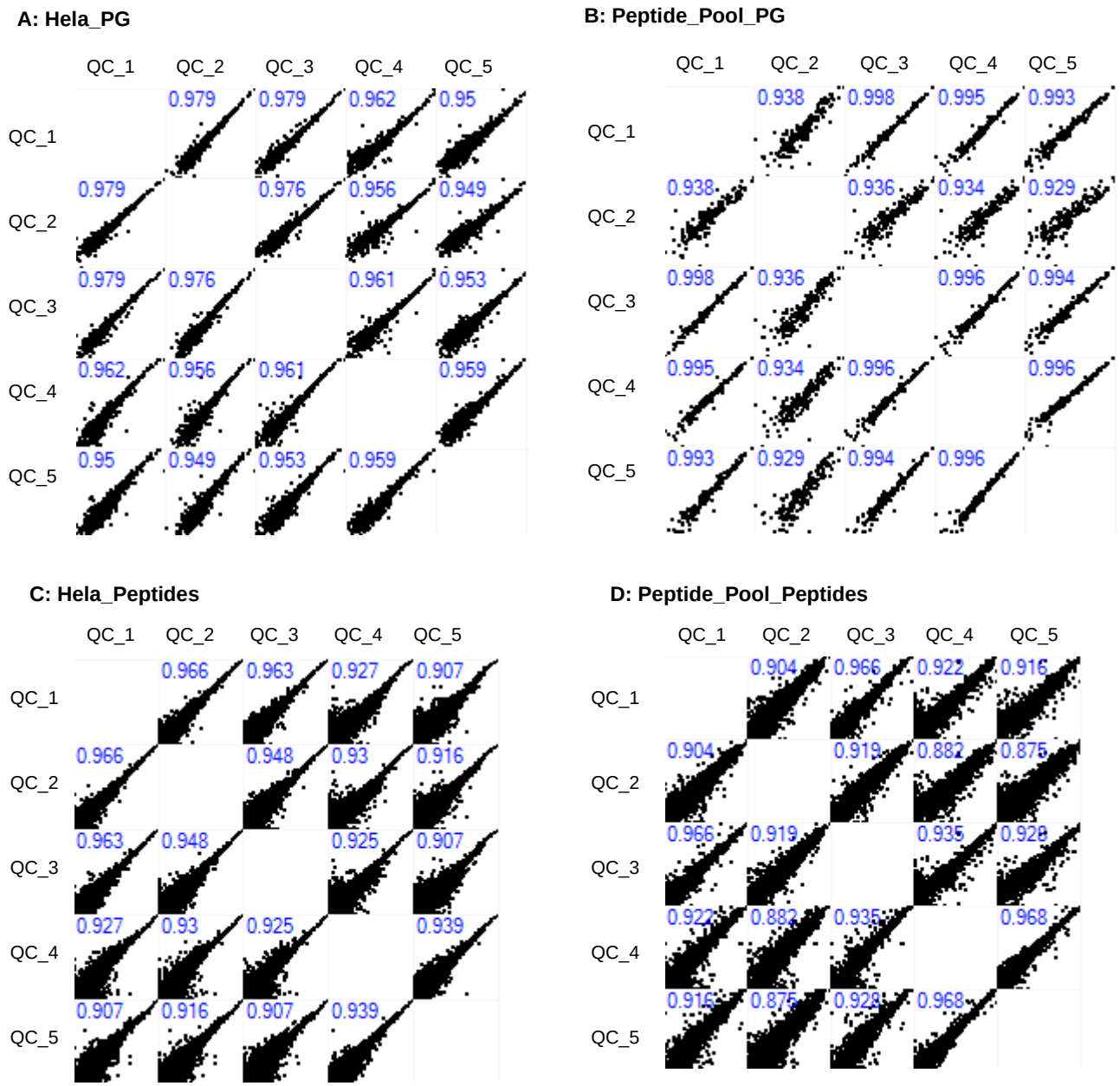


Figure 3.3: Multi-scatter plots showing Pearson correlation of identified protein groups and peptides in Hela_QC (A, C) and Peptide_Pool_QC (B, D) replicates analysed over 5 days.

In addition, the median percentage coefficient of variation (CV) was used to illustrate the inter-day variation observed for the protein groups identified in the Hela and Peptide_Pool QC replicates, respectively (Figure 3.4). The Hela and Peptide_Pool QC replicates had a median CV of 20,5% and 26,3%, respectively (Figure 3.4), indicating that there was low technical variation within the 5 days of LC-MS/MS analysis for both QC measurements.

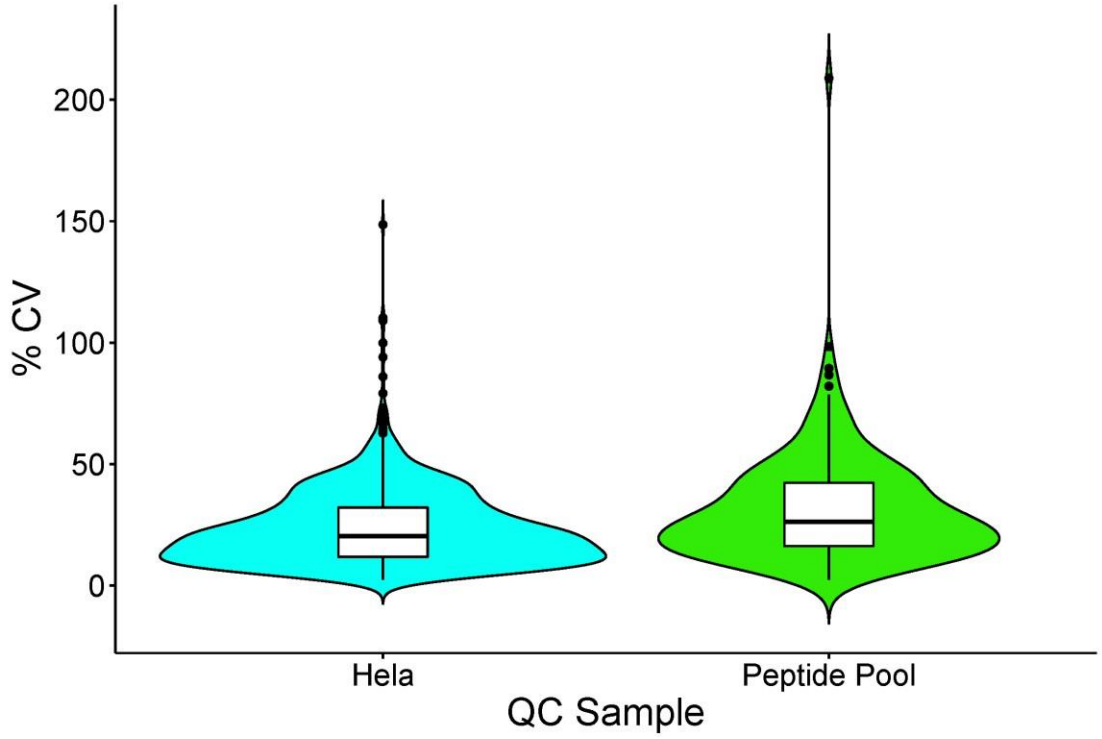


Figure 3.4: Boxplot showing the median percentage coefficient of variation (% CV) for the Hela (A) in blue and Peptide_Pool (B) in green QC replicates.

To further assess the quality and reliability of the experimental data, LC-derived peak parameters were tabulated below. A decline in the peak capacity of the Hela and Peptide_Pool QC replicates was observed between day 4 and 5, during the acquisition period. Peak capacity is a common metric of separation power in gradient elution, and the typical peak capacity for peptide separation ranges from 100 to 400. The peak capacity for all Hela and Peptide_Pool QC replicates was within the expected range for peptide separation (Table 3.4).

Furthermore, as the peak capacity decreases, the median full width half maximum (FWHM) of the peaks resolved in each replicate widens (indicated by an increase in the median FWHM), resulting in reduced peak resolution. Additionally, the average data points per peak for the Hela and Peptide_Pool QC replicates are 6.1 and 7.1, respectively (Table 3.4). The observed average data points per peak are sufficient for reproducible quantitation.

Table 3.4: Table of LC derived peak parameters*. The average data points per peak, peak capacity and the median FWHM in minutes is tabulated for each Hela_QC and Peptide_Pool_QC replicate analysed.*

	Average data points per peak	**Peak Capacity**	**Median FWHM (minutes)**
Hela_QC_1	5,9	154,85	0,0799
Hela_QC_2	5,98	152,53	0,0811
Hela_QC_3	5,92	146,12	0,0846
Hela_QC_4	6,54	107,51	0,1151
Hela_QC_5	6,18	98,33	0,1258
Average data points for Hela	6,10		
Peptide_Pool_QC_1	6,57	178,49	0,0693
Peptide_Pool_QC_2	6,88	174,35	0,0709
Peptide_Pool_QC_3	6,62	170,98	0,0723
Peptide_Pool_QC_4	7,88	125,33	0,0987
Peptide_Pool_QC_5	7,52	118,98	0,104
Average data points for Peptide_Pool	7,09		

3.4) Identified candidate markers and gene ontology enrichment analysis of AKI study

Block 1 was prepared and re-analysed together with blocks 2 - 5 and the full AKI cohort (54 AKI; 86 non-AKI) was subjected to the LC-MS/MS analysis and data processing workflow established in the preliminary assessment of the study (section 3.2). An average of 4 370 precursors, 3 212 modified peptides, 2 952 unmodified peptides, 363 protein groups and 401 proteins were identified across all the AKI and non-AKI patient samples (Table 3.5). Within the individual study conditions, an average of 3 365 modified peptides and 380 protein groups were identified in the AKI group, while an average of 3 113 modified peptides and 352 protein groups were identified in the non-AKI group (Figure 3.5). Furthermore, the retrospective power analysis performed showed that at a power of 80%, the appropriate fold-change cut-off for the candidate markers identified in the AKI study is ~ 1,68 (Figure 3.6).

Table 3.5: Overview of the average number of identifications determined in the preliminary analysis. *The average number of precursors, modified and unmodified peptides, protein groups, and proteins are listed.*

Average no. of identifications	AKI study (54 AKI, 82 non-AKI)
Precursor	4 370
Modified peptides	3 212
Unmodified (stripped) peptides	2 952
Protein groups	363
Proteins	401

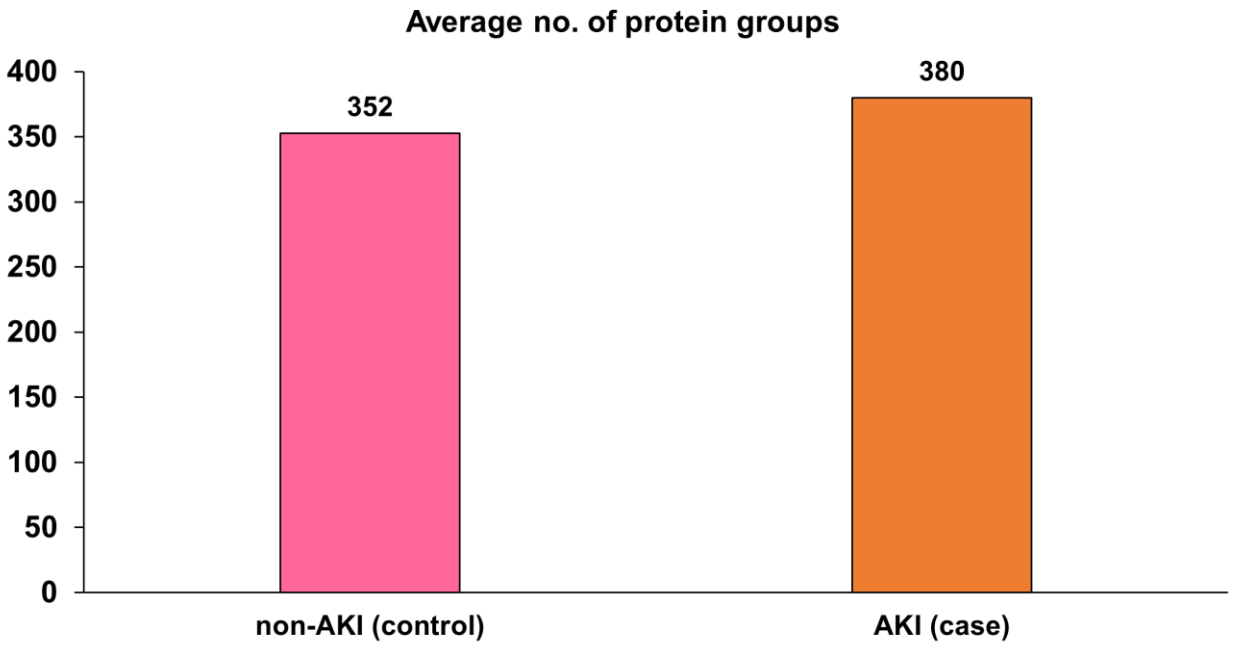


Figure 3.5A: Bar graph indicating the average number of protein groups identified between the non-AKI (pink) and AKI (orange) study conditions.

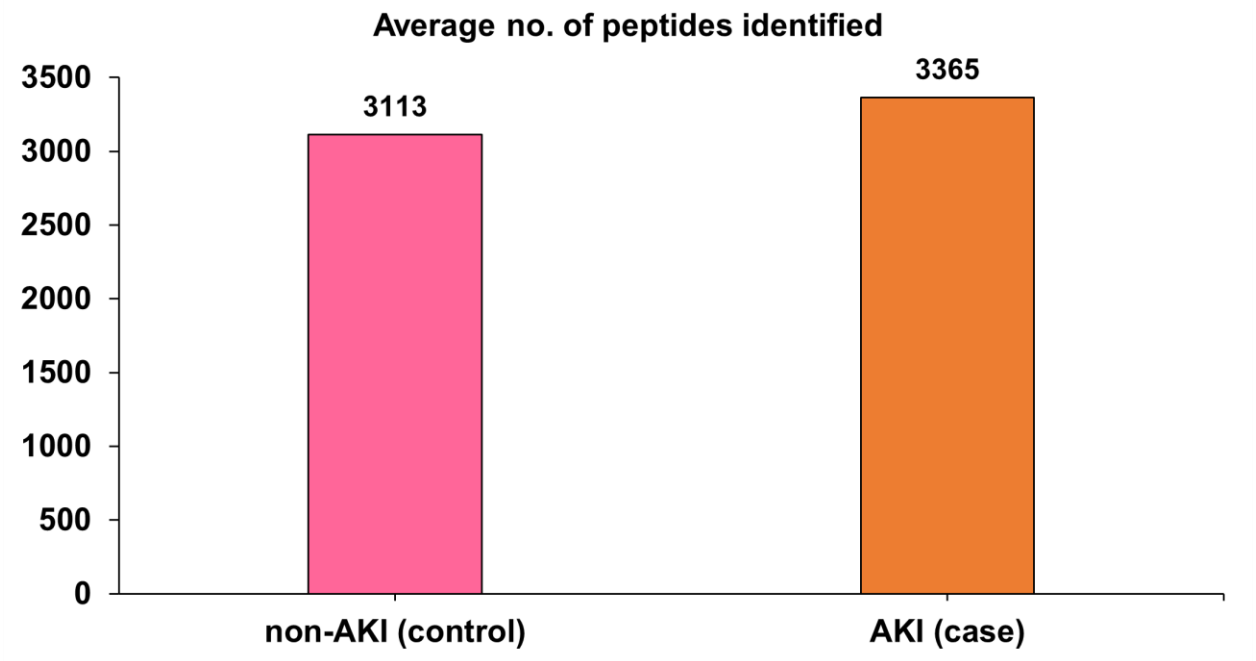


Figure 3.5B: Bar graph indicating the average number of peptides identified between the non-AKI (pink) and AKI (orange) study conditions.

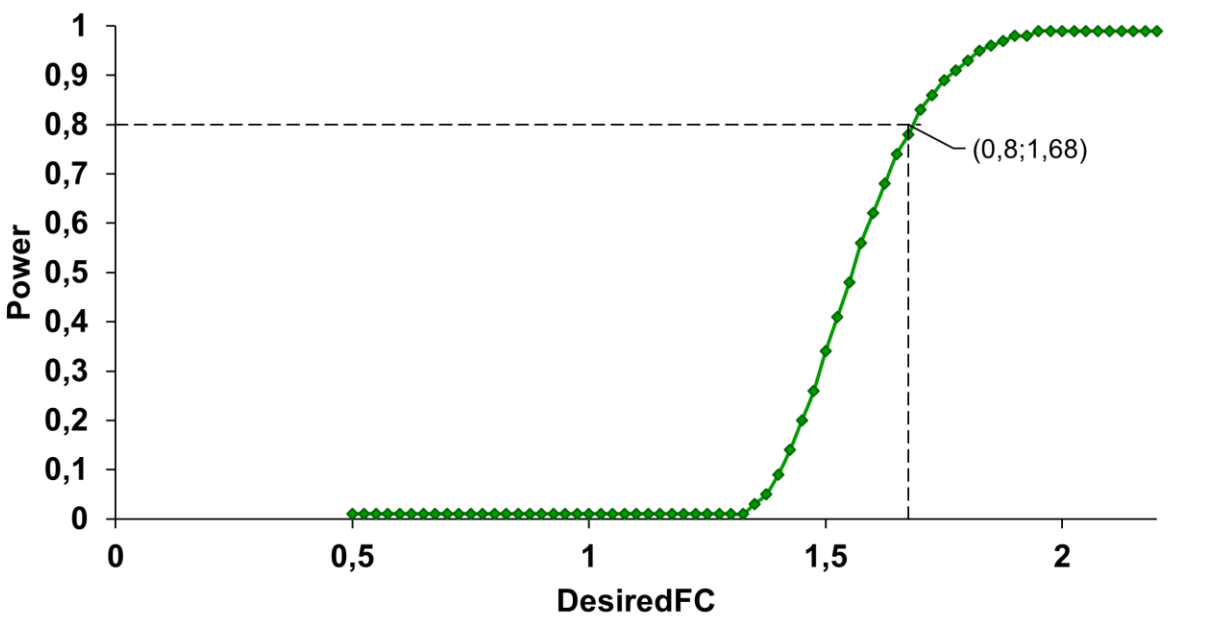


Figure 3.6: Line graph indicating the estimated power of the AKI study and the corresponding desired fold-change (FC) defining the candidate markers*. The desired FC is given at an FDR of 1%. Power analysis completed using MSstats (v3.4.1) in RStudio.*

An unpaired Student's t-test was performed, and with the desired fold-change determined by the retrospective power analysis, the volcano plot below shows that 34 candidate markers were identified using the following selection criterion: q-value ≤ 0.05, absolute $\log_2$FC ≥ 0.75 (FC ≥ 1.68). There are 3 differentially abundant proteins that showed increased abundance (pink) and 31 differentially abundant proteins that showed decreased abundance (blue) in the AKI group compared to the non-AKI group (Figure 3.7). The list of the 34 differentially abundant proteins can be found in Table 6.2.

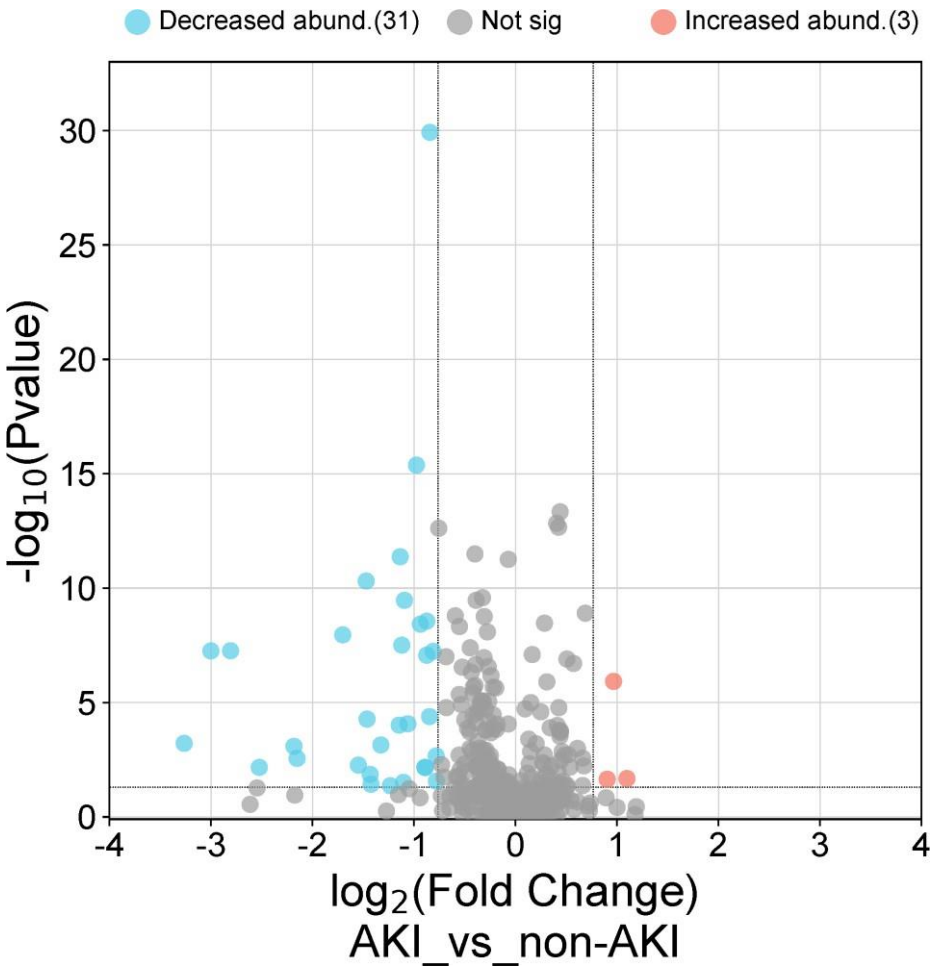


Figure 3.7: Volcano plot of differentially abundant proteins generated from the AKI study*. The candidate markers in pink have increased abundance and markers in blue have decreased abundance, in the AKI group compared to the non-AKI group (q-value ≤ 0.05, absolute $\log_2$ FC ≥ 0.75; FC ≥ 1.68). Proteins in grey show no significant differential abundance. Volcano plot generated in SRplot.*

The enrichment analysis for the GO biological processes, molecular functions, cellular compartments, and pathways that are associated with the candidate markers, indicated various enriched processes and pathways of signal transduction (e.g., intrinsic/extrinsic apoptotic signalling), as well as immune responses (e.g., leukocyte cytotoxicity), metabolism (e.g., RNA and cholesterol) and small molecule transport (plasma lipoprotein assembly/remodelling) (Table 3.6).

Table 3.6: GO biological processes, molecular functions, and pathway analysis of candidate markers*. Enriched GO biological processes, molecular functions, cellular compartments as well as KEGG, Reactome, and Wikipathways associated with the identified candidate markers is tabulated below. Data generated with Enrichr and is tabulated in descending order of the number of candidate markers found in each process, function, or pathway name (adjusted p-value ≤ 0.05).*

	Name	No. of candidates found in process/pathway	Candidate marker gene names	p-value	Adjusted p-value
GO Biological Processes	Regulation Of Very-Low-Density Lipoprotein Particle Remodelling (GO:0010901); Regulation of CoA-transferase Activity (GO:1905918); Regulation of Cholesterol Transport (GO:0032374); Regulation of Lipid Catabolic Process (GO:0050994); Regulation of Cytokine Production Involved in Immune Response (GO:0002718)	1	*APOA2*	8,72 E-03	1,32 E-01
	Positive Regulation of Cell Migration by Vascular Endothelial Growth Factor Signalling Pathway (GO:0038089); Vascular Endothelial Growth Factor Receptor Signalling Pathway (GO:0048010)	1	*HSPB1*	1,22 E-02	1,32 E-01
	Regulation Of Oxidative Stress-Induced Intrinsic Apoptotic Signalling Pathway (GO:1902175); Negative Regulation of Oxidative Stress-Induced Cell Death (GO:1903202)	1	*HSPB1*	3,10 E-02	1,32 E-01

	Regulation Of Leukocyte Mediated Cytotoxicity (GO:0001910); Negative Regulation of Extrinsic Apoptotic Signalling Pathway Via Death Domain Receptors (GO:1902042); Negative Regulation of Endothelial Cell Apoptotic Process (GO:2000352); T Cell Activation Involved in Immune Response (GO:0002286)	1	*ICAM1*	1,56 E-02	1,32 E-01
	Spindle Assembly Checkpoint Signalling (GO:0071173)	1	*KNTC1*	4,45 E-02	1,32 E-01
	RNA Phosphodiester Bond Hydrolysis (GO:0090501)	1	*PARN*	4,79 E-02	1,32 E-01
GO Molecular Function	poly(A)-specific Ribonuclease Activity (GO:0004535); Telomerase RNA Binding (GO:0070034)	1	*PARN*	2,08 E-02	2,06 E-01
	Peptidase Inhibitor Activity (GO:0030414); Endopeptidase Regulator Activity (GO:0061135); Endopeptidase Inhibitor Activity (GO:0004866)	1	*PZP*	6,28 E-02	2,09 E-01
	Serine-Type Endopeptidase Activity (GO:0004252); Serine-Type Peptidase Activity (GO:0008236); Endopeptidase Activity (GO:0004175)	1	*CTRB2*	1,97 E-01	3,59 E-01
GO Cellular Compartment	Desmosome (GO:0030057)	1	*DSG1*	2,93 E-02	1,42 E-01
KEGG Pathways	Cholesterol metabolism (map04979)	2	*ABCA1; APOA2*	3,46 E-03	1,21 E-01
	T cell receptor signalling pathway (map04660)	1	*PAK6*	1,67 E-01	2,89 E-01
	NF-kappa B signalling pathway (map04064); TNF signalling pathway (map04668); Natural killer cell mediated cytotoxicity (map04650)	1	*ICAM1*	1,67 E-01	2,89 E-01
Reactome Pathways	Regulation Of mRNA Stability by Proteins That Bind AU-rich Elements R-HSA-450531	2	*PARN; HSPB1*	1,04 E-02	3,43 E-01
	HDL Assembly R-HSA-8963896; Plasma Lipoprotein Assembly R-HSA-8963898	1	*ABCA1*	1,22 E-02	3,43 E-01
	Uptake Of Dietary Cobalamins into Enterocytes R-HSA-9758881; Cobalamin (Cbl, Vitamin B12) Transport and Metabolism R-HSA-196741; Activation of Matrix Metalloproteinases R-HSA-1592389	1	*CTRB2*	1,74 E-02	3,43 E-01
	Apoptotic Cleavage of Cell Adhesion Proteins R-HSA-351906; Apoptosis R-HSA-109581	1	*DSG1*	1,91 E-02	3,43 E-01

	KSRP (KHSRP) Binds and Destabilizes mRNA R-HSA-450604; Deadenylation Of mRNA R-HSA-429947; ATF4 Activates Genes in Response to Endoplasmic Reticulum Stress R-HSA-380994	1	*PARN*	2,93 E-02	3,43 E-01
	Interleukin-10 Signalling R-HSA-6783783; Interleukin-4 And Interleukin-13 Signalling R-HSA-6785807; Interferon Gamma Signalling R-HSA-877300	1	*ICAM1*	7,59 E-02	3,43 E-01
	MAPK6/MAPK4 Signalling R-HSA-5687128; VEGFA-VEGFR2 Pathway R-HSA-4420097; Signalling by VEGF R-HSA-194138	1	*HSPB1*	1,46 E-01	3,65 E-01
Wikipathways	Metabolic Pathway of LDL HDL And TG Including Diseases WP4522; Statin Inhibition of Cholesterol Production WP430	2	*ABCA1; APOA2*	3,98 E-04	2,63 E-02
	Vitamin B12 Metabolism WP1533; Folate Metabolism WP176; Urotensin II Mediated Signalling Pathway WP5158	2	*ABCA1; ICAM1*	7,04 E-03	7,75 E-02

Using STRING database (STRING-db) through Cytoscape (v3.10.1), clusters of interaction networks are observed between the candidate markers (p-value ≤ 0.05, absolute $log_2FC \geq 0.75$) and background protein groups. Candidate markers APOA-2- and ABCA-1- are involved in cholesterol transport and plasma lipoprotein assembly and remodelling. APOA-2, LEG-7 (*LGALS7B*), CH3L-1 and HSPB-1 are involved in regulation of cytokine production. CH3L-1 and ICAM-1 participate in pathways of signal transduction. Both DSG-1 and DESP interact with each other and a background protein (PLAK – *JUP*) associated with signal transduction pathways. DESP and HSPB-1 participate in wound healing processes. Furthermore, PZP participates in peptidase and endopeptidase activities and interacts with CTRB-2. KNTC-1, PARN, AIG-1, and PAK-6 show no participation in the enriched pathways, processes, and diseases in the AKI study (Figure 3.8).

The biological processes, pathways, and tissues enriched in the AKI study were: plasma lipoprotein assembly and remodelling (red), kidney disease (pink), peptidase and endopeptidase activity (purple), liver disease (brown), adaptive immune responses (green), complement cascade and activation (orange), cholesterol metabolism and transport, as well as chylomicron assembly and remodelling (navy), regulation of cytokine production (light pink), wound healing (light green), signal transduction (light purple) and lastly, haemostasis (blue).

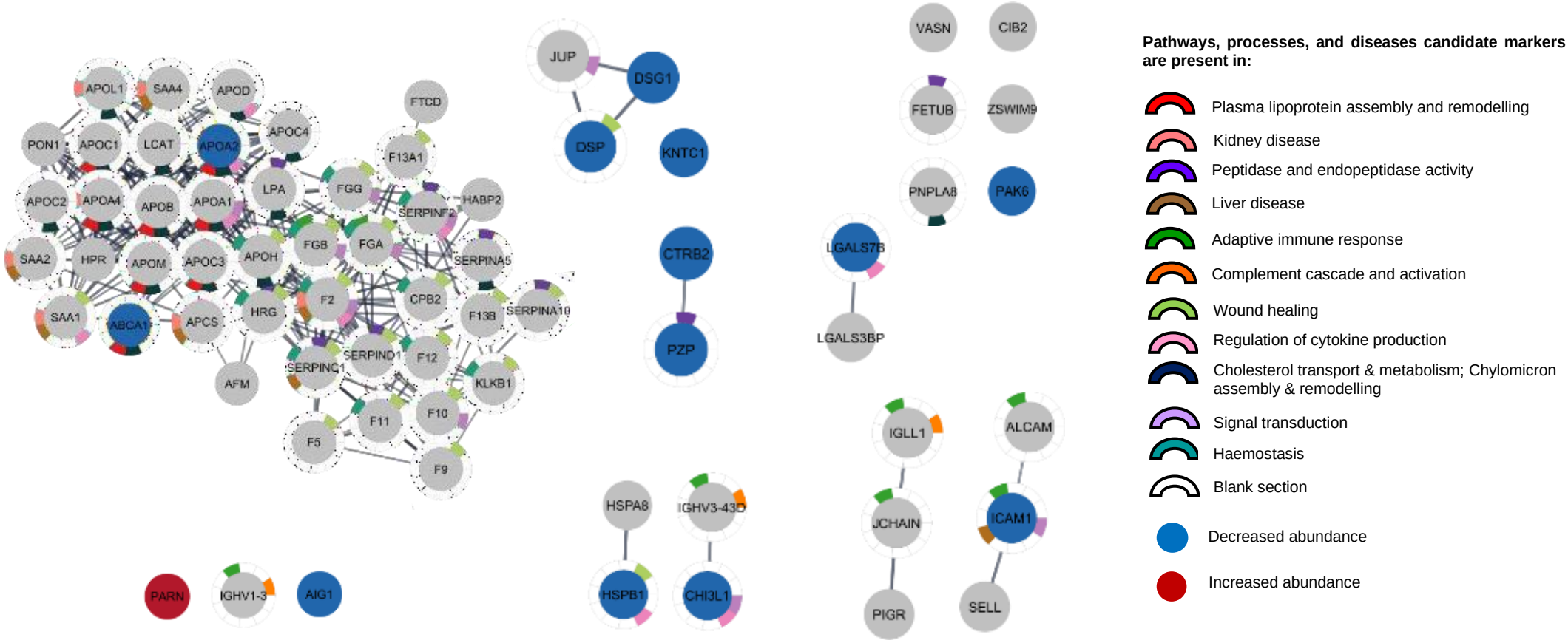


Figure 3.8: Interactions and networks of candidate markers identified with the pathways, processes, and tissues in which they are associated with. *The red and blue nodes represent candidate markers with increased and decreases abundance, respectively in the case group (AKI). The coloured sections surrounding the nodes represent the associated pathway(s) or tissue(s). Interactions were observed between candidate markers as well as background proteins (grey nodes – no statistically significant differential abundance) in the study; the interactions shown at a confidence level of 70%.*

3.5) Using machine learning to identify features that distinguish the conditions of the AKI study

In addition to the statistically significant candidate markers identified using an unpaired t-test (section 3.4), features (i.e., proteins or genes) that were able to distinguish the AKI group from the non-AKI group were also generated using the machine learning (ML) platform OmicLearn. OmicLearn was used to explore the most appropriate options of a ML pipeline(s) to generate the best performing ML model, that can use 20 of the most important proteins that contribute to differentiating between the two study conditions. There are various options available for data imputation, feature selection as well as ML classifiers to generate ML pipelines. The performance of the different ML models explored was evaluated based on the scores (0 – 1; 0: poor performance and 1: excellent performance) given for accuracy, precision, recall, F1_score, ROC_AUC and PR_AUC.

Firstly, the performance of the 5 feature selection methods was evaluated using the XGBoost classifier with a 'no imputation' method. For each feature selection method, a maximum of 20 features (default setting in OmicLearn) were identified to contribute to the training of the classifier and in this way the data input was reduced from 474 available features to the 20 most contributing features. The XGBoost classifier is the only classifier available when using a 'no imputation' method in OmicLearn. The 'None' and 'ExtraTrees' feature selection methods performed similarly, but the 'ExtraTrees' option was preferred as the 'None' option can create bias in the important protein features identified by the selected ML model(s). The performance scores for each ML model were reported in Table 3.7A.

Secondly, the performance of 7 ML classifiers was assessed using the 'ExtraTrees' feature selection method and the 'zero imputation' method, where 'zero' was imputed in the place of missing values in the dataset. The XGBoost classifier had the best performance across all classifiers evaluated. The performance scores for each ML model were reported in Table 3.7B.

Lastly, the performance of the 'no', 'zero', as well as the 'median' imputation methods was evaluated using the 'ExtraTrees' feature selection method and the XGBoost classifier. Imputation is used to assign values to missing data in the input data matrix and 26.71% (17 092) of the data matrix were missing values. Although >70% of the data matrix was available, imputation methods were evaluated to explore the impact of imputation on the performance of the ML models. The 'no' and 'median' imputation methods performed similarly and the 'no imputation' method was preferred to the 'median imputation' method, to ensure that no bias or loss of information from the features with missing values occurs. The performance scores for each ML model were reported in

Table 3.7C.

Table3.7A: Mean performance values for ML models to evaluate feature selection methods. *ML models were built using the 'no' imputation strategy and XGBoost classifier to evaluating various feature selection methods (ExtraTrees, k-best [mutual_info_classif, chi2 and f-classif] and none) in the ML pipelines for the AKI study.*

	ExtraTrees	k-best (mutual_info_classif)	k-best (f_classif)	k-best (chi2)	None	Best Score	Best Score (excl. 'none')
Accuracy	0,7363	0,7304	0,7215	0,7126	0,7578	**None**	**ExtraTrees**
Precision	0,6884	0,6839	0,6601	0,6428	0,7223	**None**	**ExtraTrees**
Recall	0,6360	0,6169	0,6185	0,6342	0,6385	**None**	**ExtraTrees**
F1 Score	0,6485	0,6375	0,6276	0,6269	0,6671	**None**	**ExtraTrees**
ROC_AUC	0,8154	0,7985	0,8082	0,7604	0,8327	**None**	**ExtraTrees**
PR_AUC	0,7360	0,7175	0,7181	0,6515	0,7515	**None**	**ExtraTrees**

Table 3.7B: Mean performance values for ML models to evaluate ML classifiers. *ML models were built using the 'zero' imputation strategy and the ExtraTrees feature selection method to explore the various ML classifiers (AdaBoost, LogisticRegression, RandomForest, k-NeighboursClassifier, DecisionTree and LinearSVC) in the ML pipelines for the AKI study.*

	AdaBoost	LogisticRegression	RandomForest	k-Neighbours Classifier	DecisionTree	LinearSVC	XGBoost	Best Score
Accuracy	0,7207	0,7119	0,7215	0,6074	0,6578	0,7074	0,7244	**XGBoost**
Precision	0,6667	0,6455	0,6766	0,0000	0,5774	0,6452	0,6732	**RandomForest**
Recall	0,6153	0,6162	0,5867	0,0000	0,5425	0,5338	0,5924	**LogisticRegression**
F1 Score	0,6263	0,6193	0,6155	0,0000	0,5505	0,5668	0,6211	**AdaBoost**
ROC_AUC	0,7859	0,7942	0,8087	0,7076	0,6371	0,7994	0,8102	**XGBoost**
PR_AUC	0,6747	0,6684	0,7105	0,6836	0,6496	0,7064	0,7183	**XGBoost**

Table 3.7C: Mean performance values for ML models to evaluate imputation methods. *ML models were built using the ExtraTrees feature selection method with the XGBoost classifier to evaluate the 'no', 'zero' and 'median' imputation strategies in the ML pipelines for the AKI study.*

	No imputation	Zero imputation	Median imputation	Best Score
Accuracy	0,7363	0,7244	0,7370	**Median imputation**
Precision	0,6884	0,6732	0,6883	**No imputation**
Recall	0,6360	0,5924	0,6329	**No imputation**
F1 Score	0,6485	0,6211	0,6507	**Median imputation**
ROC_AUC	0,8154	0,8102	0,8239	**Median imputation**
PR_AUC	0,7360	0,7183	0,7504	**Median imputation**

The most appropriate ML model explored for the AKI study consisted of the 'no' imputation method, with the 'ExtraTrees' feature selection method and XGBoost classifier. The confusion matrix generated for the selected ML model accurately predicted 64% of AKI patients to the AKI group, while the model inaccurately predicted 36% of AKI patients to the non-AKI group. Additionally, the ML model correctly assigned 80% of the non-AKI patients to the non-AKI group, while the model incorrectly assigned 20% of the non-AKI patients to the AKI group (Figure 3.9).

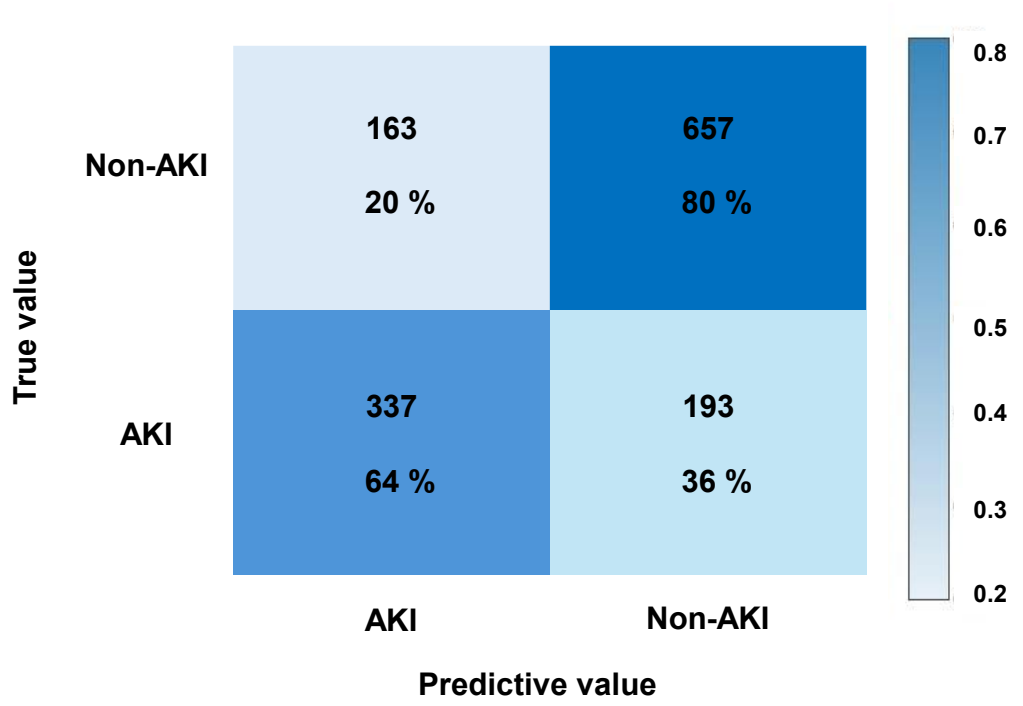


Figure 3.9: Confusion matrix of the selected machine learning (ML) model. *The percentage of patients that are accurately and inaccurately predicted and assigned to either the AKI or non-AKI groups.*

The selected ML model utilised the 20 features selected by the feature selection method used in the ML pipeline. These 20 features contributed the most to distinguishing the AKI and non-AKI groups from each other and were ranked according to the most important to the least important protein feature (Table 3.8). The feature importance value for each protein were reported in Table 6.3.

Table 3.8: List of top 20 proteins differentiating the two study conditions ranked in order of importance using machine learning (ML). *The 20 most important proteins are listed along with their protein descriptions. The ML model used: No imputation, ExtraTrees feature selection method and XGBoost classifier, using the OmicLearn platform.*

Protein Names	Protein Descriptions
STXB5	Syntaxin-binding protein 5
PROF1	Profilin-1
PLF4	Platelet factor 4
B2MG	Beta-2-microglobulin
ITA2B	Integrin alpha-IIb
1433Z	14-3-3 protein zeta/delta
RP1BL	Ras-related protein Rap-1b-like protein
TAGL2	Transgelin-2
G3P	Glyceraldehyde-3-phosphate dehydrogenase
PARN	Poly(A)-specific ribonuclease PARN
RNAS1	Ribonuclease pancreatic
APOH	Beta-2-glycoprotein 1
UBB	Polyubiquitin-B
TBB5	Tubulin beta chain
ZA2G	Zinc-alpha-2-glycoprotein
PLXA2	Plexin-A2
PZP	Pregnancy zone protein
A1BG	Alpha-1B-glycoprotein
O51B5	Olfactory receptor 51B5
DCA13	DDB1- and CUL4-associated factor 13

A principal component analysis (PCA) plot comparing the AKI group to the non-AKI group using the 20 most contributing features for the classification of the two study groups by the selected machine learning (ML) model is shown below. Figure 3.10 showed no distinct separation between the AKI and non-AKI group using the contributing features identified during ML.

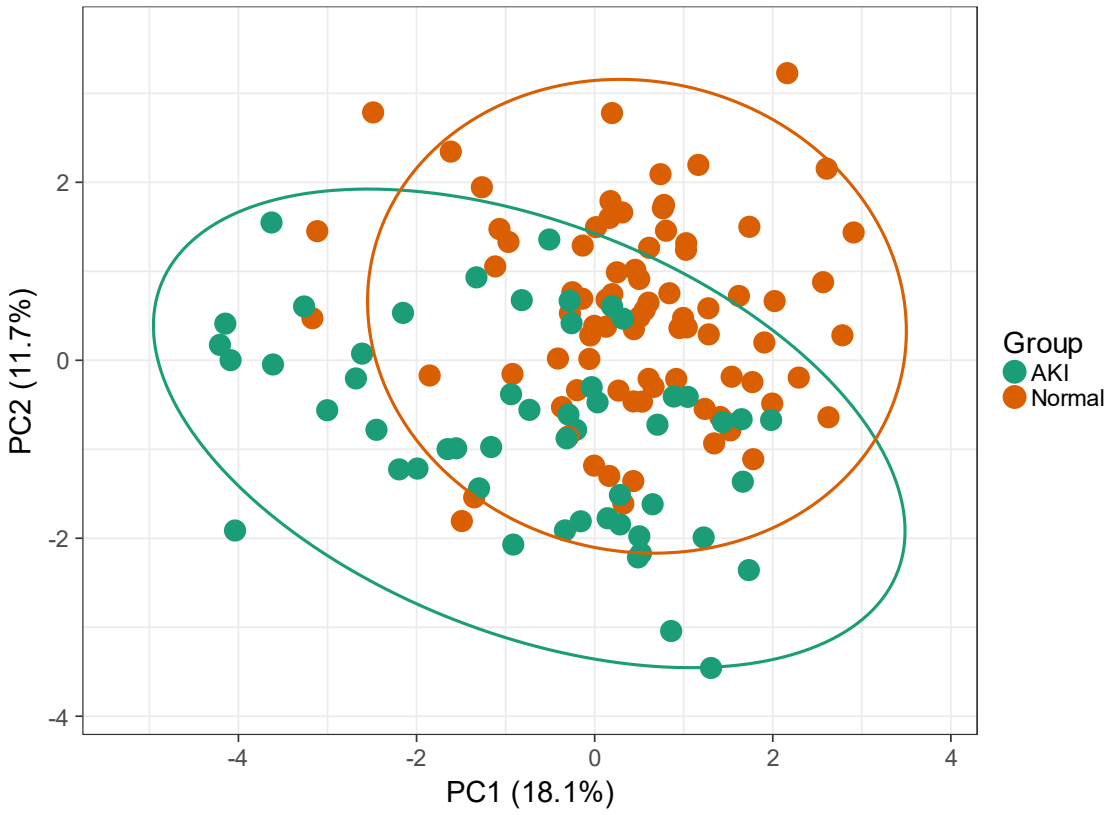


***Figure 3.10**: **Principal component analysis plot for features contributing to machine learning (ML) classification of study groups**. AKI cohort samples were compared between the AKI (green) and non-AKI (orange) groups using the 20 most contributing features identified by the ML model used.*

Furthermore, the important protein features identified with the ML model (See Table 3.8) were compared to the candidate markers identified using the unpaired t-test, in a Venn diagram. There are 3 proteins (STXB-5, PZP and PARN) that were found to overlap the candidate markers of Student's t-test, and the list of important protein features used by the ML model applied to distinguish the AKI group from the non-AKI group (Figure 3.11).

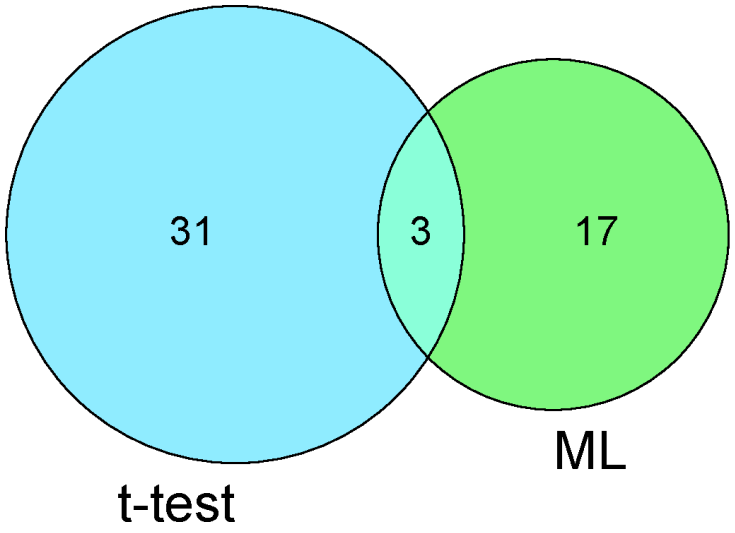


Figure 3.11: Venn diagram comparing differentially abundant proteins identified by an unpaired t-test and a machine learning (ML) model of the same study group*. The t-test-derived markers are shown in blue, and the ML-derived markers are shown in green. The Venn diagram shows 3 proteins (STXB-5, PZP, and PARN) overlapping both methods of identifying differentially abundant proteins.*

3.6) Comparing identified candidate markers from the plasma and urinary AKI cohorts

A Venn diagram was used to compare the candidate plasma markers (section 3.5) to the candidate markers (not shown) generated from the urinary proteomes of the same study groups (q-value < 0.05, absolute $\log_2$FC ≥ 0.58; FC ≥ 1.5). Five candidate markers (LEG-7, CH3L-1, IGFBP-6, HSPB-1, and KVD-33) were identified in both the urinary and plasma proteomes of the study groups (Figure 3.12). The list of candidate markers identified in the plasma and urinary cohorts were reported in Table 6.4 and 6.5, respectively.

The fold-change of LEG-7 was 3.55-times less in plasma than in urine and was greater in the non-AKI group than the AKI group. The fold-change of CH3L-1 was 1.99-times greater in urine than in plasma and the fold-change of CH3L-1 in urine was greater in the AKI group, but in plasma was greater in the non-AKI group. The fold-change of IGFBP-6 was 1.29-times greater in urine than in plasma and was greater in the AKI group than in the non-AKI group. The fold-change of HSPB- 1 was 0.83 times greater in plasma than in urine and was greater in the non-AKI group than the AKI group. The fold-change of KVD-33 was 1.46-times greater in urine than in plasma and was greater in the non-AKI group than the AKI group (Table 3.9).

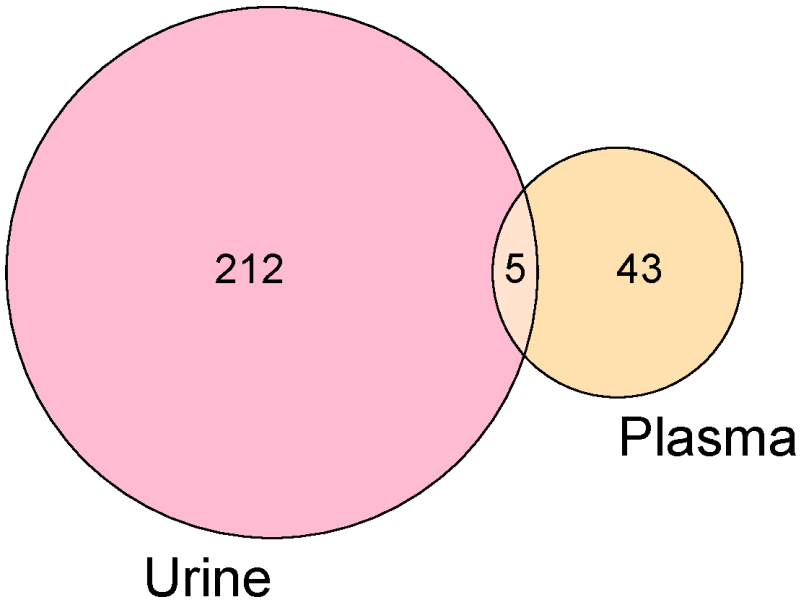


***Figure 3.12**: **Venn diagram comparing candidate markers identified in the urinary and plasma proteomes of the same study groups.** The urinary markers are shown in pink, and the plasma markers are shown in orange. The Venn diagram shows that 5 candidate markers (LEG-7, CH3L-1, IGFBP-6, HSPB-1, and KVD-33) are overlapping both studies (q-value ≤ 0.05, absolute log_2FC ≥ 0.58; FC ≥ 1.5).*

***Table 3.9: List of candidate markers overlapping the urinary and plasma proteome of the study groups.** The average log2 fold-change of LEG-7, CH3L-1, IGFBP-6, HSPB-1, and KVD-33 in urine and in plasma.*

Protein Name	Ave Log_2FC	
	Urine	Plasma
LEG-7	-0,920	-3,265
CH3L-1	1,770	-0,890
IGFBP-6	1,168	0,903
HSPB-1	-1,018	-1,233
KVD-33	-1,181	-0,811

3.7) Summary of Results

A summary of the results obtained from the AKI study is highlighted below. The patient demographic factors, such as race, sex and the median age showed no significant statistical difference ($p > 0.05$), and all patients studied were exposed to an ART-regimen including TDF. The AKI clinical parameters, eGFR and srCr levels of patients, indicated a significant statistical difference between the AKI and non-AKI groups.

The preliminary study was conducted on a subset of patient samples (39), and the resulting data showed that 65 proteins were differentially abundant using the default selection criterion in Spectronaut™ (v17): q-value ≤ 0.05 and an absolute $\log_2$FC ≥ 0.58 (FC ≥ 1.5). Gene ontology enrichment revealed that mainly biological processes and pathways of the immune system and signal transduction were enriched.

Quality control (QC) measures put in place to assess variability in the performance of the LC-MS/MS system over five days, indicated that Hela and Peptide_Pool QC replicates had low technical variation and that the peak related parameters performed adequately for reproducible quantitation.

In the AKI study, an average of 363 protein groups were identified across the study conditions and in each condition, an average of 380 and 352 protein groups were identified in the AKI and non-AKI groups, respectively. A retrospective power analysis was performed to determine the appropriate fold-change for the AKI study. There were 34 differentially abundant proteins (i.e., candidate markers) that were identified, and gene ontology enrichment indicated mainly pathways and processes of signal transduction, immune response, and small molecule transport were highly enriched. The cluster of interaction networks shows the interaction(s) of candidate markers and background proteins within their relevant enriched processes and pathways.

Additionally, the performance of different machine learning (ML) models was evaluated, to identify proteins that can distinguish the AKI group from the non-AKI group using the most appropriate model. The best performing ML model was built from a pipeline consisting of the 'no' imputation method, with the 'ExtraTrees' feature selection method and the XGBoost classifier. This ML model identified 20 proteins that distinguish the AKI and non-AKI groups from each other. There were 3 proteins (STXB5_HUMAN, PZP_HUMAN and PARN_HUMAN) found overlapping the candidate markers identified with the unpaired t-test and the proteins identified as important differentiators of the study conditions.

Lastly, the candidate markers identified in the AKI plasma and urinary studies were compared, and 5 markers (LEG-7, CH3L-1, IGFBP-6, HSPB-1, and KVD-33) were found overlapping in both studies.

Chapter 4

Discussion

South Africa accounts for approximately 20% of the global burden of HIV and has responded by establishing the largest antiretroviral therapy (ART) program globally, that freely administers ART to approximately 65% of people living with HIV (PLHIV) (Rees and Bisnauth, 2022; Statistics South Africa, 2022). Although ART has greatly assisted PLHIV, some patients experience toxic side effects associated with ARV drugs, with approximately 10% of African patients taking first-line ART developing AKI (Seedat *et al*., 2017). Monitoring ARV toxicity is necessary to detect adverse drug responses and evaluating long-term drug effects (Meintjes *et al.,* 2017). ARV-induced AKI can be reversed with early detection of the renal insult by substituting ARVs in the treatment regimen (Boswell and Rossouw, 2017; Venter *et al*., 2018). Currently, AKI is diagnosed by a reduction in estimated glomerular filtration rate (eGFR) and an increase in serum creatinine (srCr) levels; however, these parameters serve as unreliable markers of AKI, as srCr levels can remain relatively unchanged in early or mild cases of AKI despite kidney damage (Vaidya *et al*., 2008; Schmid *et al*., 2015; Kashani *et al*., 2017; Sakhuja, 2023). As a result, more biomarkers that can effectively and efficiently identify early kidney dysfunction are required for early treatment interventions, in order to limit ARV-related AKI.

The purpose of this study was to evaluate the plasma of PLHIV that were diagnosed with AKI and those that were not diagnosed with AKI, to identify differentially abundant proteins between the two study groups for the discovery of potential protein markers related to kidney damage. Plasma was used as it is a commonly analysed medium for clinical diagnoses and is an ideal source of novel biomarkers (Ignjatovic *et al*., 2019).

4.1) Statistical assessment of patient demographic and AKI clinical details

Preparation of and access to well-documented patient demographic profiles is important, as they provide insight into factors such as age, race, or gender, that can influence or cause bias in the interpretation of the outcomes of a biological study (Festing, 2020; Burger *et al*., 2021). Additionally, the clinical characteristics detailing AKI diagnosis, such as srCr levels and the eGFR of patients, are required to ensure that the patient samples included in the study clinically represent their study groups, i.e., AKI (case) or non-AKI (control). Table 3.1 showed that there was no significant statistical difference in the median age, race, and gender of the TDF-exposed patients between the AKI and non-AKI groups, indicating that the findings of downstream proteomic analysis were likely independent of the effects of the demographic factors mentioned.

Lastly, Table 3.1 also showed a significant statistical difference in the median srCr levels and eGFR of patients between the AKI and non-AKI groups, confirming that patients in the AKI group have elevated srCr levels and lower eGFRs in comparison to the patients in the non-AKI group, as supported by literature (Vaidya *et al*., 2008; Schmid *et al*., 2015; Kashani *et al*., 2017; Sakhuja, 2023). Following this statistical assessment, a preliminary analysis of the AKI study was conducted.

4.2) Preliminary results of the AKI study

Preliminary analysis was performed on the 39 patient samples in block 1 (Figure 2.1) to evaluate the adequacy of the plasma preparation and analysis workflow by observing the number of protein groups (i.e. proteins and their isoforms) identified, the differentially abundant proteins found between the AKI and non-AKI groups, and the biological pathways and processes that are associated with them.

In this preliminary study, an average of 358 ($\pm$ 23) protein groups and 3 175 ($\pm$ 181) modified peptides were identified from the undepleted patient plasma samples; this corresponded with the number of proteins identified under similar conditions in some literature (Geyer *et al*., 2016; O'Rourke *et al*., 2023). This showed that the workflow could be used adequately to analyse the full AKI cohort (135 patient samples). However, with the advancement of sample preparation techniques and the increased sensitivity of state-of-the-art LC-MS instrumentation and software, greater protein identifications ($\geq$ 1 000 proteins) have been observed in cohorts analysing undepleted plasma (Park *et al*., 2020).

Using the default significance cut-off (q-value < 0.05, absolute $log_2FC \geq 0.58$; $FC \geq 1.5$) in Spectronaut™ (v17), 65 differentially abundant proteins were identified as significant, and were considered candidate markers for the preliminary study. Of the top 10 candidate markers that showed up- regulated or increased abundance (Figure 3.1 – pink dots) in the AKI group compared to the non-AKI group, protein S100-A8 and -A9, as well as heat shock protein 90 alpha (HSP90-A) were identified in many of the enriched GO biological processes and pathways (Table 3.3) and are discussed below.

Calprotectin (S100A8/9) activates pro-inflammatory mechanisms shown to mediate renal damage and progression of renal fibrosis in obstructive nephropathy, observed in a murine model of unilateral ureteral obstruction (Tammaro *et al*., 2018). Table 3.3 further indicated that S100A8/9 participates in enriched processes and pathways relating to immune response such as signalling of inflammatory cytokines (such as interleukins), leukocyte aggregation which can cause obstruction to blood flow, antimicrobial peptide activity and the release of reactive oxygen species from immune cells for defence purposes; further emphasizing the role of S100A8/9 in renal dysfunction and its potential usefulness as a biomarker of AKI.

Under AKI conditions, HSP90 proteins (e.g. HSP90-A) expression is increased in the PTCs during the late regeneration phase where damaged cells are renewed and differentiation of new cells occurs, signifying the role HSP90 proteins in adaptive repair (Chebotareva *et al*., 2017). Furthermore, overexpression of HSP90 and 70 proteins is observed in I/RI associated with kidney transplantation (O'Neill *et al*., 2014; Nayak Rao, 2016). HSP90-A is shown (Table 3.3) to interact in enriched processes and pathways relating to immune response, such as inflammatory cytokine (such as interleukins) signalling, as well as chaperone-mediated autophagy degrading proteins (that are likely irreparably misfolded) and in the PI3K-Akt signalling pathway, known to regulate cell physiology and pathology including cell survival and proliferation (Shi *et al*., 2019), emphasising the potential role that HSP90-A plays in AKI survival.

On the other hand, tissue inhibitor of metalloproteinase 1 (TIMP-1) was also identified in the top 10 markers with increased abundance but was not found to participate in the enriched biological processes and pathways reported in Table 3.3. However, TIMP-1 has been reported as a marker of tubulointerstitial injury (Mizdrak *et al*., 2022). TIMPs regulate the activity of matrix metalloproteinases (MMPs), which are endopeptidases that regulate inflammation, cell proliferation and apoptosis, EMT, and angiogenesis, and their dysregulation is associated with renal diseases, such as AKI (Tan and Liu, 2012). Furthermore, TIMP-2 is a predictive and prognostic marker of AKI in conjunction with IGFBP- 7 (Wang *et al*., 2021). This suggests that TIMP-1 may have utility as a potential marker of AKI.

Of the top 10 candidate markers that showed down-regulated or decreased abundance (Figure 3.1 – blue dots) in the AKI group compared to the non-AKI group, fibronectin (FINC) was found to participate in multiple enriched GO biological processes and pathways (Table 3.3) and is discussed below. Fetuin-B was also identified in the top 10 markers with decreased abundance but was not found to participate in the enriched biological processes and pathways reported in Table 3.3.

Fetuins are cysteine protease inhibitors and are synthesised in the liver and secreted into the blood (Olivier *et al*., 2000; Chekol Abebe *et al*., 2022). Down-regulated mRNA levels of Fetuin-A and Fetuin-B were observed during the acute phase of inflammation that was induced experimentally in rats (Olivier *et al*., 2000), and have been implicated in bone resorption, which is an adverse effect associated with some nephrotoxic ARV drugs, and in response to inflammation among other functions (Denecke *et al*., 2003). Urinary fetuin-A has been identified as a marker of AKI and showed notable increases in rats following administration of nephrotoxic agent, cisplatin, prior to changes observed in serum creatinine and histological evidence of tubular damage (Zhou *et al*., 2006). Additionally, urinary fetuin-A was also found in high concentration in the cytoplasm of damaged proximal tubule cells with increasing severity of injury (Zhou *et al*., 2006). As fetuin- A and fetuin-B share structural and functional similarities, investigating fetuin-B in AKI can potentially serve as an indication of renal tubular injury.

Fibronectin (FINC) is a multi-domain glycoprotein and forms an important component of the ECM, playing a vital role in the repair of tissues (To and Midwood, 2011). Changes in FINC expression, degradation and organisation have been associated with disease. Tubulointerstitial fibrosis is among various risk factors associated with the pathophysiological mechanisms of AKI and is characterised by an excessive deposition of the ECM and myofibroblast activation in the renal tubules (Guan *et al*., 2020). FINC participates in enriched processes and pathways noted in literature, such as being a receptor for the ECM and integrin cell surface interactions, as well as the PI3K-Akt signalling pathway (regulates cell physiology and pathology (Shi *et al*., 2019)), and also participates in interleukin signalling and immune responses (Bonventre and Yang, 2011), as observed in Table 3.3; this emphasises that FINC can serve as a potential marker of AKI.

The identification of candidate markers that can potentially serve as markers of AKI, suggests that the workflow used for plasma proteome profiling in the preliminary study was adequate in identifying proteins that are associated with renal dysfunction and is suitable for the complete AKI study.

4.3) Quality control for AKI study

The Hela_QC and Peptide_Pool_QC replicates served as system suitability and sample-specific system suitability measures, respectively, and were incorporated into the AKI study to assess the variability in the performance of the LC-MS/MS system during the five days of analysis (Figure 2.1). The variability in LC-MS/MS performance was evaluated at a peptide and protein level. Figure 3.2 showed that a similar number of protein groups and peptides were identified between the Hela_QC replicates, with no more than 5.52% variation observed, and the Peptide_Pool_QC replicates, with no more than 2.39% variation observed. This indicated that there was low technical variation in LC-MS/MS performance between the five days.

Figure 3.3 also indicated that there was low technical variation between the individual QC replicates, as a strong positive correlation was found between the identified protein groups and peptides. The violin plots shown in Figure 3.4 further confirmed that there was low technical variation in LC-MS/MS performance between the five days. This suggests that the variation observed in the AKI study was mainly driven by biological variation.

In addition, the LC-derived peak parameters, which are peak capacity, median full width half maximum (FWHM) of the peak and the number of data points per peak were used to assess the quality and reliability of peak separation and peptide quantitation.

The peak capacity and median FWHM are indirectly proportional parameters that refer to the separation power of peptides during a gradient elution. Table 3.4 showed that the peak capacity of the Hela_QC and Peptide_Pool_QC replicates was in the typical range for peptide separation, indicating that there was adequate peak separation of the detected peptides. Furthermore, the decrease in peak capacity of the individual QC replicates on day 4 and 5, corresponds with the increase on day 4 and 5 of the median FWHM of the peaks identified and was indicative of a drift in the chromatographic performance, which results in a decline in the sensitivity of the LC-MS/MS system with time (also shown by the decline in protein groups and peptides observed on day 4 and 5 in Figure 3.2A). The anticipated decline in LC-MS/MS sensitivity was accounted for by the block randomisation strategy used to acquire LC-MS/MS data (Burger *et al*., 2021).

Lastly, the number of data points obtained per LC peak is an important metric to monitor, as more data points present per peak, result in a more defined peak shape for accurate and reproducible quantitation of the area under the peak (or curve). However, acquiring more data points per peak requires longer run times, resulting in a trade-off with the available analysis time on the mass spectrometer. This can be counteracted by increasing the scan speed of

each analysis window, but can negatively affect data quality (Bruderer *et al*., 2015; Ignjatovic *et al*., 2019). As a result, the 6 – 7 average data points observed per peak for the Hela_QC and Peptide_Pool_QC replicates was sufficient for reproducible quantitation, indicating that the intensity measurements were reliably acquired for the detected fragment/peptide ions (Table 3.4) (Bruderer *et al*., 2015).

The incorporation and evaluation of quality control measures showed relatively low technical variability with enough data points per peak for sufficient quantitation, suggesting that the data from the proteome profiling of the AKI cohort was reliably and reproducibly acquired.

4.4) Differentially abundant proteins identified in AKI study

The clinical utility of MS-based plasma proteome profiling is in its ability to detect changes at the protein level that represent the underlying pathophysiological mechanisms of disease. This information can assist in the discovery of biomarkers and the development of novel therapeutic treatments (Geyer *et al*., 2017; Karayel *et al*., 2022). Proteome profiling of patient samples can identify proteins that are uniquely expressed or proteins with differences in their abundance levels between normal and diseased patient groups (Geyer *et al*., 2017; Karayel *et al*., 2022).

In this AKI study, the proteomic profiling of the plasma patient samples (53 AKI; 82 non-AKI) identified 34 differentially abundant protein groups as significant, i.e., candidate makers (q - value < 0.05, absolute $\log_2$ FC $\geq$ 0.75; FC $\geq$ 1.68) between the AKI and non-AKI groups. Most candidate markers shown in Figure 3.7 have been reported in the literature as associated with renal damage and subsequent repair processes, however, to date, this study has not been undertaken in the South African population.

The candidate markers that were involved in the enriched GO biological processes, molecular functions, cellular compartments as well as KEGG, Reactome, and Wiki pathways (Table 3.6) are discussed below, including candidate markers shown to be associated with AKI according to the Harmonizome (v3.0) database (p-value < 0.05) (Rouillard *et al*., 2016). The Harmonizome (v3.0) database is a collection of processed datasets allowing the search of genes or proteins and their functional terms extracted and organized from over a hundred publicly available resources (Rouillard *et al*., 2016). The CTD gene-disease association dataset was used, and the association was measured by the use of manually curated literature (Rouillard *et al*., 2016).

Of the three candidate markers shown with an upregulated or increased abundance (Figure 3.7 - pink dots) in the AKI group compared to the non-AKI group, the insulin-like growth factor-binding protein 6 (IGFBP-6) was associated with AKI as per the Harmonizome (v3.0) database and the poly(A)-specific ribonuclease (PARN) participated in enriched GO biological processes, molecular functions and Reactome pathway enrichment shown in Table 3.6.

On the other hand, of the thirty-one candidate markers shown with down-regulated or decreased abundance (Figure 3.7 - blue dots) in the AKI group compared to the non-AKI group, heat shock protein β 1 (HSPB-1), apolipoprotein A-2 (APOA-2), phospholipid-transporting ATP binding cassette transporter A-1 (ABCA-1), intercellular adhesion molecule 1 (ICAM-1), serine/threonine-protein kinase PAK 6 (PAK-6), chymotrypsinogen B 2 (CTRB-2), pregnancy zone protein (PZP), desmoglein 1 (DSG-1) and Kinetochore-associated protein 1

(KNTC-1) participated in enriched GO biological processes, molecular functions and Reactome, KEGG and Wiki pathway enrichment shown in Table 3.6. Additionally, desmoplakin (DESP), galectin 7 (LEG-7), chitinase 3-like protein 1 (CH3L-1) and vinculin (VINC) were associated with AKI as per the Harmonizome (v3.0) database. These dysregulated candidate markers are discussed under the main biological processes, functions and pathways observed in Table 3.6 that are key mechanisms of AKI.

4.4.1) Immune Response

This AKI study identified candidate markers participating in enriched processes and pathways of innate and adaptive immune responses. These immune responses facilitate kidney injury and recovery from AKI, with immune cells, such as T- and B-lymphocytes, neutrophils, and monocytes, performing particular roles in the development of AKI (Bonavia and Singbartl, 2018; Singbartl *et al*., 2019). AKI can be caused by septic and aseptic (nephrotoxic or ischaemic) insults that activate the innate immune system and usually involve changes in inflammatory pathways (Bonavia and Singbartl, 2018; Singbartl *et al*., 2019). Damaged renal tubular epithelial and endothelial cells induce local inflammation, which is followed by leukocyte recruitment and infiltration to the site of renal inflammation (Figure 4.1) (Bonavia and Singbartl, 2018; Singbartl *et al*., 2019).

The necrotic renal cells release damage-associated molecular pattern (DAMPs) that bind and induce toll-like receptor (TLR) activity in the early phase of innate immune response, thus mediating pro-inflammatory pathways that involve interleukin-18 (IL-18), -6 (IL-6) and -1β (IL-1β) (Bonavia and Singbartl, 2018; Singbartl *et al*., 2019). Some infiltrating leukocytes such as neutrophils are recruited by endothelial adhesion molecules, such as ICAM-1, and produce and release cytotoxic substances (e.g. reactive oxygen species (ROS)), while adhering to and migrating across the endothelium to the site of inflammation (Figure 4.1) (Bonavia and Singbartl, 2018; Singbartl *et al*., 2019). Furthermore, ICAM-1 was also shown to participate in enriched pathways that regulate leukocyte mediated cytotoxicity (Table 3.6). In addition, T-lymphocytes are also recruited during inflammation and have been shown to contribute to the presence of nephrotoxicity in AKI (Bonavia and Singbartl, 2018; Singbartl *et al*., 2019). In this study, PAK-6 and ICAM-1 were shown to participate in enriched T-cell receptor signalling and activation pathways involved in immune responses.

Conversely, leukocytes such as T-regulatory cells exhibit reno-protective effects, while other leukocytes such as monocytes, differentiate into M1 (pro-inflammatory) or M2 (anti-inflammatory) macrophages depending on the stage of disease progression (Bonavia and Singbartl, 2018; Singbartl *et al*., 2019). During the early phase of the innate immune response,

the binding of DAMPs to surface pattern recognition molecules (e.g. TLRs) activates M1 macrophages that release pro-inflammatory cytokines and chemokines, among other molecules. The M2 macrophages are involved in immunoregulation and release anti-inflammatory molecules, such as IL-10 and transforming growth factor-β (TGF-β) to reduce the inflammatory response and promote wound healing (Bonavia and Singbartl, 2018; Singbartl *et al*., 2019). IL-10, as well as IL-4 and -13 signalling were shown to be enriched in this study, with ICAM-1 participating in these immunoregulatory pathways (Table 3.6). Additionally, CH3L-1 has also been reported to be involved in IL-13 induced inflammation (Lee *et al*., 2009). Furthermore, M2 macrophages are involved in generating the extracellular matrix (ECM), and when dysregulated, these immune cells can contribute to renal fibrosis (Bonavia and Singbartl, 2018; Singbartl *et al*., 2019) (Figure 4.1).

The renal dysfunction associated with AKI can have an added effect on the inflammatory processes, such as increasing cytokine production and reducing cytokine clearance, indicated by an increase in plasma cytokines appearing at the onset of AKI, such as tumour necrosis factor-α (TNF-α) and IL-6 and -8 (Bonavia and Singbartl, 2018; Singbartl *et al*., 2019). ICAM-1 was shown to participate in enriched TNF signalling pathways (Table 3.6). Regulation of cytokine production was also enriched in this study, with Table 3.6 indicating APOA-2 involvement in this enriched process and Figure 3.9 indicating the participation of LEG-7, CH3L-1, and HSPB-1 in cytokine production as well. PZP is associated with clearance of pro-inflammatory cytokines as well as modulating T-helper immune cells (Cater *et al*., 2019; Löb *et al*., 2022), and its downregulation in the AKI group of this study, suggests that cytokine clearance can be potentially reduced in this group. In persistent AKI, the increase in cytokine production can result in an increase systemic inflammation, which can cause injury to other organs, such as the lungs or heart, and can lead to immunosuppression (Bonavia and Singbartl, 2018; Singbartl *et al*., 2019).

Lastly, 11 immunoglobulin lambda, kappa, and heavy variables (see Table 7.2) known to be released from B-lymphocytes are responsible for antigen (Ag) recognition in adaptive immune responses and were downregulated in the AKI group in this study. IL-4, -10 and -13 mediated signalling pathways enriched in this AKI study, have been reported in literature to share immunoregulatory functions on B-lymphocytes (Bonavia and Singbartl, 2018; Chomarat and Banchereau, 1998). B-lymphocytes have been suggested to play a pathogenic role in AKI by secreting immunoglobulins and limiting wound healing (Bonavia and Singbartl, 2018; Singbartl *et al*., 2019). Therefore, the decline in B-lymphocytes can be associated with kidney repair, suggesting that the downregulation of the immunoglobulin variables in the AKI group of this study may potentially reflect recovery from AKI (Bonavia and Singbartl, 2018; Singbartl *et al*., 2019).

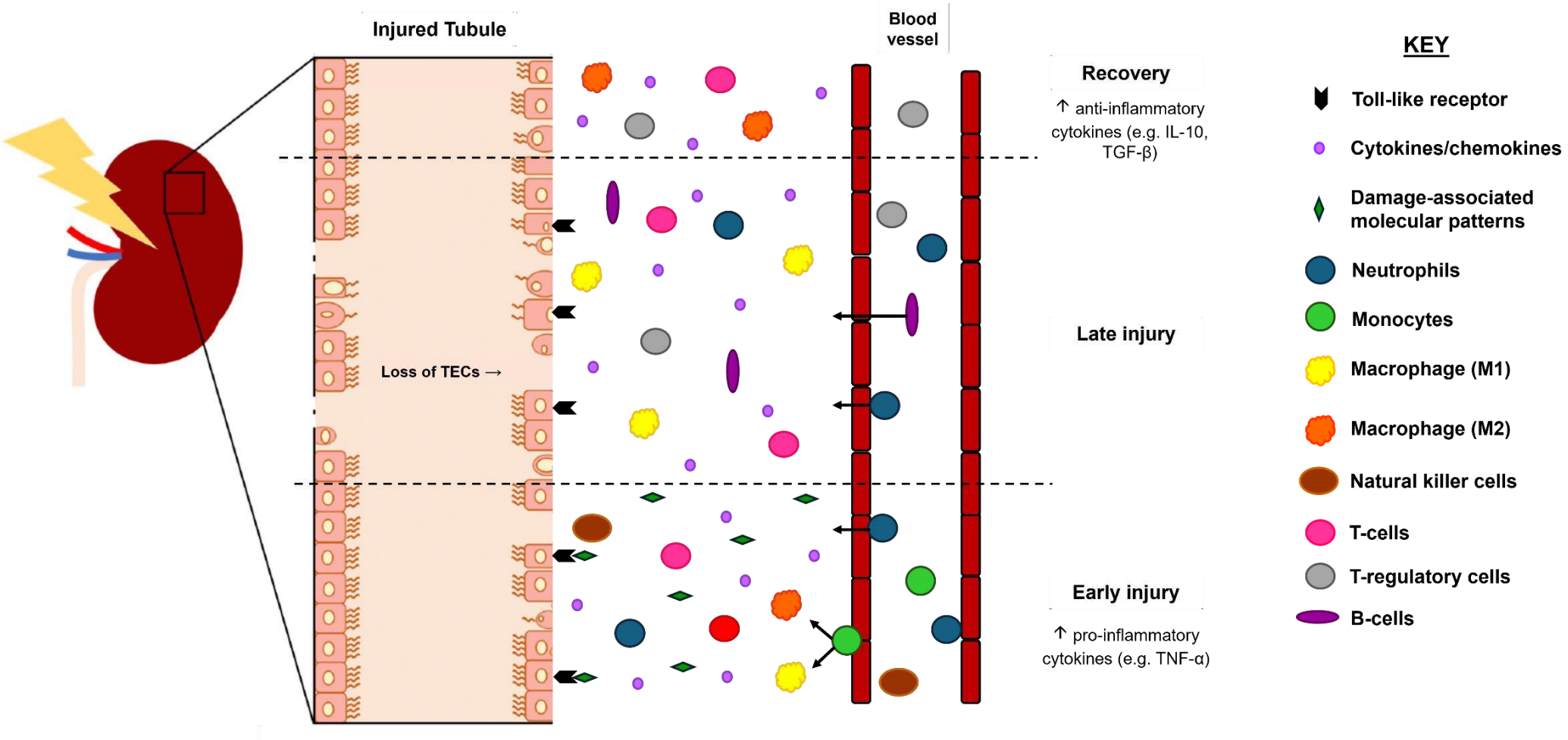


Figure 4.1: Immune responses associated with injury to the kidney proximal tubule. *The immune cells and molecules associated with early and late injury as well as the recovery phase of AKI are illustrated. Figure adapted from Bonavia and Singbartl, 2018; Black et al., 2019 (Figure 2).*

4.4.2) Programmed cell death: apoptosis and necrosis

Pathways of regulated programmed cell death (PCD) are necessary to maintain the healthy functioning of an organism, however, dysregulated cell death is key in the pathogenesis of various kidney diseases (Havasi and Borkan, 2011; Sanz *et al*., 2023). Apoptotic cell death is a characteristic feature of nephrotoxin-induced AKI, with proximal tubule cells being highly susceptible to apoptosis and this injury contributes to organ dysfunction (Vaidya *et al*., 2008; Havasi and Borkan, 2011; Sanz *et al*., 2023). Additionally, endothelial, and epithelial renal cell death also contributes to AKI (Vaidya *et al*., 2008; Havasi and Borkan, 2011; Sanz *et al*., 2023). Regulated apoptosis functions by intrinsic and extrinsic caspase-mediated pathways (Figure 4.2), where necrosis can facilitate the release of DAMPs into the extracellular space, initiating the innate and adaptive immune responses (Havasi and Borkan, 2011; Sanz *et al*., 2023).

In this study, ICAM-1 is shown to be involved in enriched pathways that negatively regulate extrinsic apoptotic signalling via death receptors (Table 3.6), which is often promoted by pro-inflammatory cytokines (e.g. TNF-α), as well as negatively regulating endothelial cell apoptotic processes (Table 3.6). Additionally, CH3L-1, a known biomarker of AKI, can limit renal tubule and inflammatory cell apoptosis (De Loor *et al*., 2016; Conroy *et al*., 2018; Zhang and Parikh, 2019). CH3L-1 down-regulation in this study may suggest persistent activation of apoptotic pathways in the AKI group. This indicated that apoptotic pathways that contribute to AKI were identified in this study.

Nephrotoxic agents can induce cell death in renal tubules using apoptotic and necrotic pathways, and the disruption and inhibition of these pathways has been shown to be reno-protective, as a result studies investigating renal cell death can generate new therapies for renal injury (Sanz *et al*., 2023). Cisplatin-induced AKI is a well-studied model for nephrotoxin-induced AKI, and the presence of cisplatin causes tubular epithelial cell toxicity, which injures the proximal tubule and decreases renal function, it also causes inflammation by inducing the release of pro-inflammatory cytokines, as well as causing vasoconstrictions of the renal vasculature (Sanz *et al*., 2023).

Cisplatin has been shown to primarily target DNA, to increase p53, thus mediating mitochondrial apoptotic pathways and to also disrupt mitochondrial ATPases, thus decreasing the membrane potential and oxidative phosphorylation of the mitochondria (Sanz *et al*., 2023). PARN is involved in the enriched pathways of RNA phosphodiester bond hydrolysis and poly(A)-specific ribonuclease activity, as well as regulation of mRNA stability of proteins along with HSPB-1 (Table 3.6). Additionally, PARN was also found to participate in an enriched pathway that activate genes in response to endoplasmic reticulum stress (Table 3.6). PARN is known to mediate p53 pathways, and the up-regulation of PARN can cause down-regulation

of p53 mRNA, therefore hindering p53-mediated apoptotic processes (Shukla *et al*., 2019; Huynh *et al*., 2023). The earliest response of the kidney to cisplatin, activates p38 mitogen-activated protein (MAP) kinase-mediated apoptotic cell death and oxidative stress (Havasi and Borkan, 2011; Sanz *et al*., 2023). HSPB-1 has been shown to participate in enriched signalling pathways that negatively regulate oxidative-stress induced cell death and specifically, the regulation of oxidative stress-induced intrinsic apoptosis, in this study (Table 3.6) (Figure 4.2).

Furthermore, significantly dysregulated candidate markers that were not participating in enriched pathways identified in this study also indicated having apoptotic functions. For example, LEG-7 contributes to pro-apoptotic signalling pathways that lead to cytochrome c release and PAK-6 contributes to anti-apoptotic signalling pathways that protect from apoptosis by phosphorylation of BAD, thus deactivating pro-apoptotic effect of BAD (Kuwabara *et al*., 2002; Zhang *et al*., 2010) (Figure 4.2). Additionally, IGFBP-6 has been suggested to have similar pro-apoptotic functions in IGFBP-3 and -7 (a known marker of AKI) by inducing apoptosis by p53-mediated IGF-independent and -dependent pathways (Oh, 2012; S. Wang *et al.*, 2021).

Lastly, DSG-1 is involved in enriched pathways of apoptosis and apoptotic cleavage of cell adhesion proteins (Table 3.6) to enable the clearance of damaged and unwanted cells by disrupting cell-cell and cell-ECM adhesion, and VINC, a cell adhesion protein, has emerged as a regulator of apoptosis. VINC deficient cells are resistant to caspase-3 mediated apoptosis under conditions that normally induce an apoptotic response (Peng *et al*., 2011). Additionally, serine-type endopeptidase activity as well as endopeptidase inhibition and regulatory activity were also enriched in this AKI study, with CTRB-2 and PZP participating in these molecular functions respectively (Table 3.6). Serpins are a family of serine protease inhibitors that are involved in coagulation, inflammation, and immune responses, and perform anti-inflammatory and -apoptotic activities, making these serine protease inhibitors appropriate candidates for monitoring AKI development (El-Hefnawy *et al*., 2022). Furthermore, CTRB-2 was also found participating in enriched pathways that activate matrix metalloproteinases, which are responsible for ECM degradation and regulation of inflammation, cell proliferation and apoptosis, EMT, and angiogenesis (Tan and Liu, 2012). This highlights the participation of serine-type endopeptidases and their inhibitors in the progression of AKI.

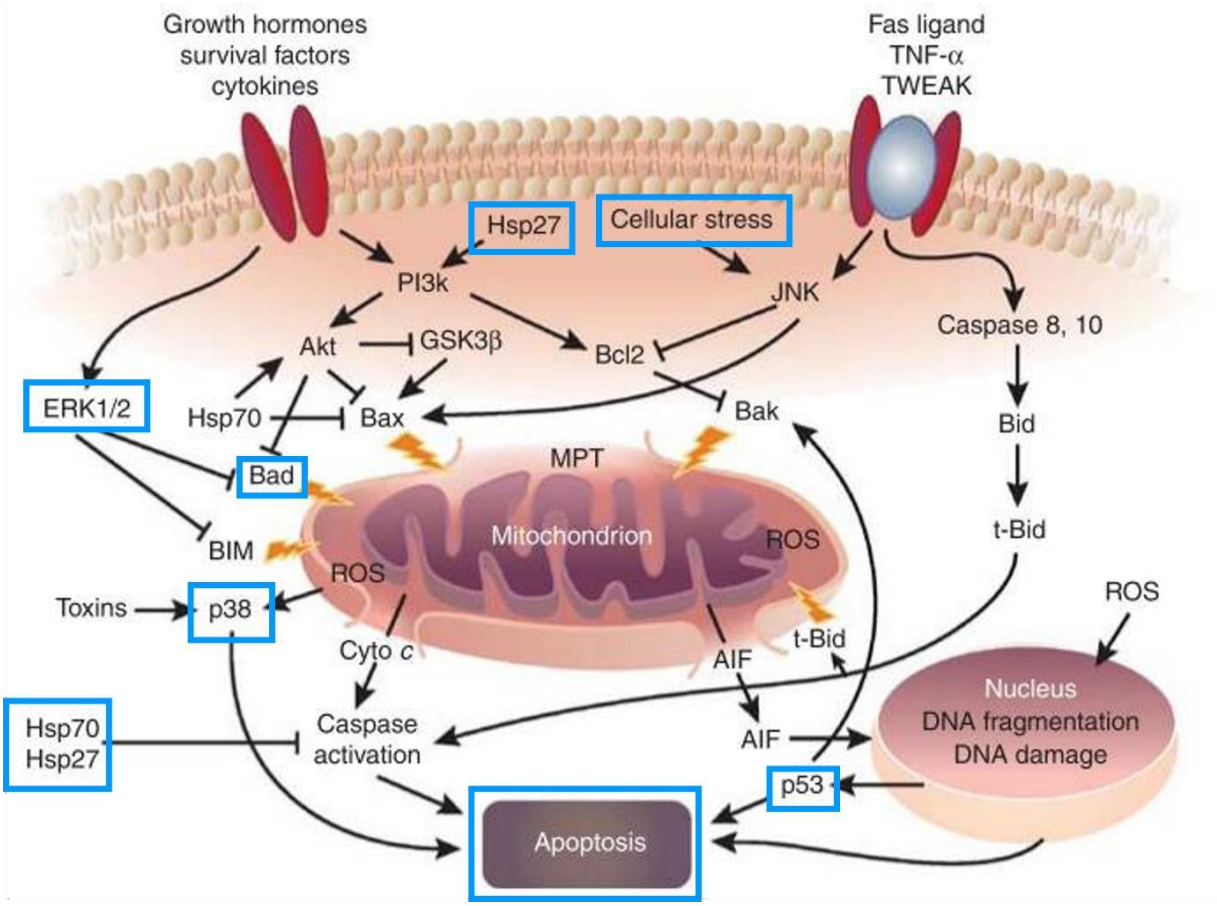


Figure 4.2: Simplified illustration of main intrinsic and extrinsic signalling events and associated proteins known to promote apoptosis in renal proximal tubule epithelial cells (PTECs) under cellular stress*. The proteins highlighted with blue boxes are those discussed above. Figure adapted from Havasi and Borkan, 2011 (Figure 1).*

4.4.3) Plasma lipoprotein assembly and cholesterol/lipid metabolism

Metabolic abnormalities such as aberrant lipid metabolism, are closely associated with the dysfunction of renal physiology, allowing for the development and progression of kidney diseases, including AKI (Lund-Katz and Phillips, 2010; Bulbul *et al*., 2018; Lin and Duann, 2020; Xu *et al*., 2023). Lipid metabolism employs multiple interdependent pathways to generate, store and transport lipids using plasma lipoprotein particles in the blood (Lund-Katz and Phillips, 2010; Bulbul *et al*., 2018; Lin and Duann, 2020; Xu *et al*., 2023). There are different types of lipoprotein particles, namely low-density (LDL), high-density (HDL), very low-density (VLDL) lipoproteins and chylomicrons, and they are composed of varying content of amphipathic lipids and apolipoproteins to stabilise their structures and fulfil their functions of mediating the movement of insoluble triglycerides and cholesterol using the lipoprotein hydrophobic core (Lund-Katz and Phillips, 2010; Bulbul *et al*., 2018; Lin and Duann, 2020; Xu *et al*., 2023). LDLs and VLDLs transport cholesterol, triglycerides, and other insoluble lipids from the liver and intestines to muscles and adipose tissues, while HDLs transport excess cholesterol from these peripheral tissues back to the liver for elimination (Lund-Katz and Phillips, 2010; Bulbul *et al*., 2018; Lin and Duann, 2020; Xu *et al*., 2023). Chylomicrons transport cholesterol, triglycerides, and other insoluble lipids into circulation to be degraded into free fatty acids and other metabolites by lipoprotein lipase; the free fatty acids generated serve as the substrate for mitochondrial metabolism via fatty acid oxidation in kidney tissue (Lund-Katz and Phillips, 2010; Bulbul *et al*., 2018; Lin and Duann, 2020; Xu *et al*., 2023).

Renal dysfunction can cause significant changes in renal lipid metabolism that can lead to dyslipidaemia (i.e. an abnormal amount of lipids in the blood), which contributes to the progression of renal dysfunction (Ogata *et al*., 1981; Bulbul *et al*., 2018; Lin and Duann, 2020). The dysregulation of renal lipid metabolism is noted by a decline in HDL particles, an increase in LDL particles and triglycerides due to variation in lipid and apolipoprotein composition, causing disruptions to cholesterol and free fatty acid metabolism in patients with AKI and chronic kidney disease (CKD) (Ogata *et al*., 1981; Bulbul *et al*., 2018; Lin and Duann, 2020). The decrease of HDL particles is likely due to alterations in the HDL proteome composition, with specific HDL associated proteins, APOA-2, APOA-1, and lecithin–cholesterol acyltransferase (LCAT), an enzyme responsible for the esterification of cholesterol in HDLs declining in patients with kidney disease (Strazzella *et al*., 2021; Pan, 2022) (Figure 4.3A-B). Decreasing HDL particles thus poorly affects reverse cholesterol transport and HDL antioxidant functions (Lin and Duann, 2020).

The downregulation of APOA-2 in the AKI group of this study coincides with the reported decline in HDL particles and its associated proteins in renal diseases. With declining HDL

particles, excess lipids are not removed from peripheral tissues leading to endoplasmic reticulum stress and mitochondrial dysfunction, contributing to cellular damage and apoptosis; this can further contribute the progression of renal injury (Lin and Duann, 2020).

Additionally, ABCA-1, an ATP-dependent transmembrane protein that regulates cholesterol and phospholipid movement from peripheral tissues (Oram, 2003; Yvan-Charvet *et al*., 2010), was also down-regulated in the AKI group of this study. ABCA-1 interacts with the amphipathic domains of lipid-poor HDL apolipoproteins, and expels excess cholesterol from peripheral cells (Oram, 2003; Yvan-Charvet *et al*., 2010) (Figure 4.3C). Down-regulation of ABCA-1 has been shown to contribute to defective cholesterol metabolism and is closely associated with the development and progression of kidney disease (Jacobo-Albavera *et al*., 2021). Furthermore, studies have indicated that the mechanisms responsible for role of the ABC transporters and HDL in reverse cholesterol transport are associated with immunosuppressive and anti-inflammatory functions of HDL (Yvan-Charvet *et al*., 2010) (Figure 4.3C); further highlighting mechanisms that can contribute to the progression AKI.

Lastly, a study investigating the dysregulated lipid metabolic pathways in kidney injury showed that patients with ischaemic and cisplatin-induced AKI had greater suppression of fatty acid oxidation compared to CKD patients, emphasising the sudden and dysfunctional impact of AKI on normal renal mitochondrial functions, and thus the physiological functions of the kidney (Xu *et al*., 2023).

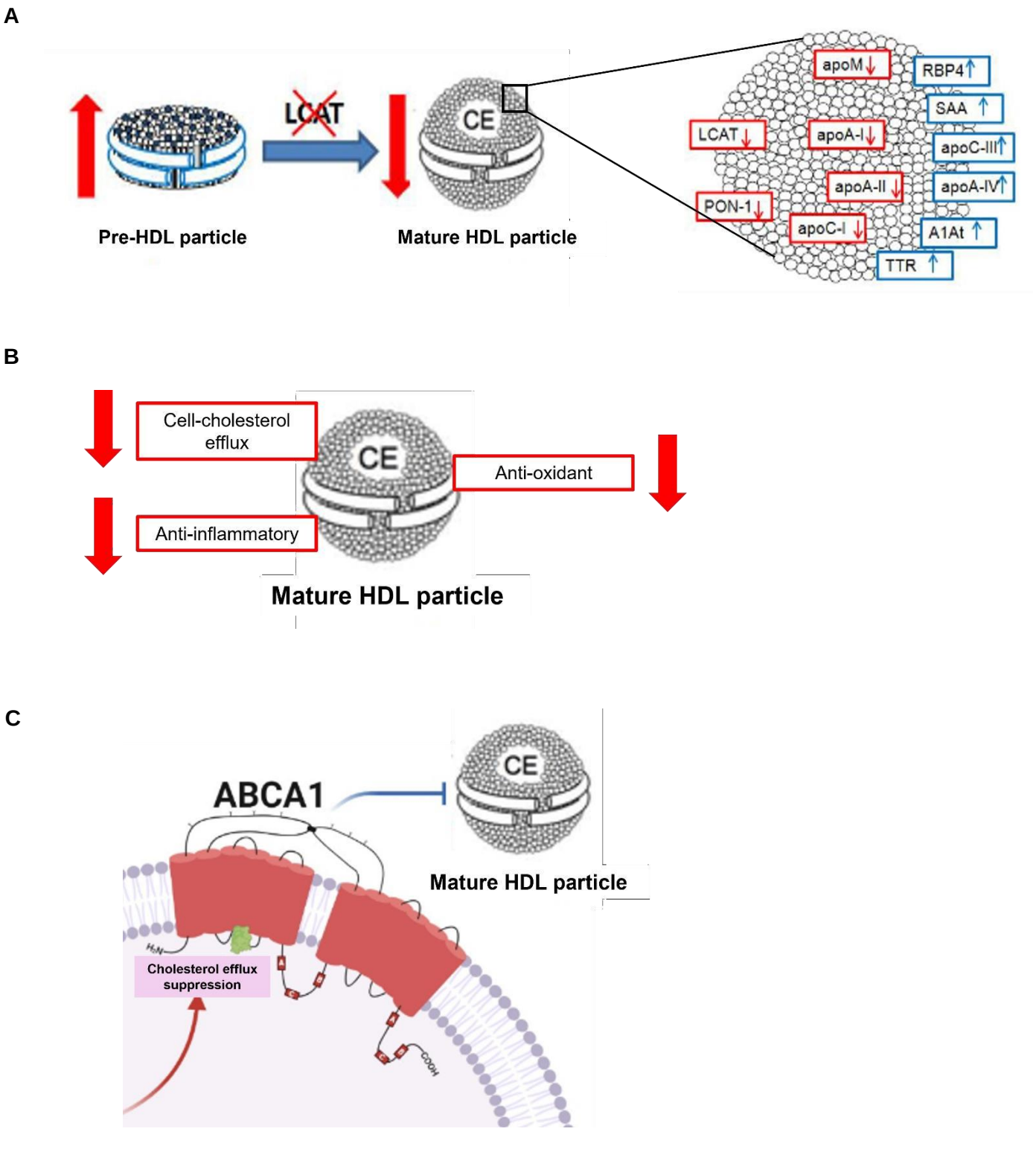


Figure 4.3: Changes in the mature HDL structural protein composition (A) and the associated functional changes of HDL (B) as well as interaction of ABCA-1 and mature HDL particle during cholesterol efflux suppression (C). *Figures adapted from Jacobo-Albavera et al., 2021 (Figure 2); Strazzella et al., 2021 (Figure 1).*

4.4.4) Cellular processes in AKI cell survival, repair, and regeneration

A variety of injuries can occur to the kidneys and generally cause damage to the major structures of the kidney, including damage to the renal tubules and can lead to necrosis of tubule cells (Bonventre and Yang, 2011; Basile *et al*., 2012; Ferenbach and Bonventre, 2016). Tubular damage can be mainly attributed to ischaemic or nephrotoxic insults and are associated with a decline in the eGFR of patients, resulting in AKI (Bonventre and Yang, 2011; Basile *et al*., 2012; Ferenbach and Bonventre, 2016). Clinically, acute injury to the renal tubules is marked by phases that reflect the cellular processes that occur during injury and recovery of tubular epithelial cells (TECs) (Bonventre and Yang, 2011; Basile *et al*., 2012).

The initiation phase consists of cellular injury, a decrease in renal blood flow that results in cellular ATP depletion causing cell dysfunction, as well as structural and functional changes in the renal tubular epithelial cells (RTECs) that hinder normal renal functions (Bonventre and Yang, 2011; Basile *et al*., 2012). This is followed by the extension phase, where cellular injury persists, and cellular damage elicits inflammatory responses, with a continued decline in the eGFR being observed (Bonventre and Yang, 2011; Basile *et al*., 2012). The maintenance phase consists of the initiation of cellular repair in damaged but viable cells, as well as migration, apoptosis and proliferation of cells replacing those lost to cell death during the initiation and extension phases to re-establish tubule integrity and improve cellular function (Bonventre and Yang, 2011; Basile *et al*., 2012). Lastly, the recovery phase sees a continuation of cell proliferation and differentiation, and the restoration of normal cellular and organ function (Bonventre and Yang, 2011; Basile *et al*., 2012) (Figure 4.4).

The maintenance and recovery phases after injury show that restoration of the structure and function of the proximal tubule requires a variety of coordinated cellular activities. Post ischaemic or nephrotoxic AKI, the basement membrane of the proximal tubule has spaces were necrotic and apoptotic proximal TECs (PTECs) have detached, indicating a loss of cell-cell and cell-extracellular matrix (ECM) adhesion, as well as disruption of the cytoskeletal organisation (Bonventre and Yang, 2011; Basile *et al*., 2012; Ferenbach and Bonventre, 2016). As a result, the remaining viable PTECs enter the cell cycle and promote cell proliferation. Dedifferentiated PTECs are produced and postulated to migrate and repopulate the proximal tubule basement membrane, where they differentiate into normal PTECs, restoring cell-cell and cell-ECM adherence to re-establish proximal tubule function (Bonventre and Yang, 2011; Basile *et al*., 2012; Ferenbach and Bonventre, 2016) (Figure 4.4).

VINC, DSG-1, DESP, and LEG-7 are proteins that participate in cell-cell and cell-ECM adhesion processes, with VINC forming focal adhesion (FA) complexes (Bays and DeMali,

2017) and regulating cadherin cell-cell adhesion (Peng *et al*., 2011), DSG-1 and DESP forming desmosome junction complexes (Lee *et al*., 2006; Jiang *et al*., 2011; Fu *et al*., 2022), and LEG-7 regulating epithelial cell proliferation, differentiation and cell-cell adhesion (Yang *et al*., 2001; Rondanino *et al*., 2011; Sewgobind *et al*., 2021). VINC, DSG-1, DESP, and LEG-7 are significantly down-regulated in the AKI group compared to the non-AKI group in this study, suggesting that the cellular events necessary for repair are disrupted in this study group.

Furthermore, Table 3.6 showed that spindle assembly checkpoint signalling (an M phase checkpoint) was an enriched process in this AKI study. KNTC-1, an important protein of the mitotic checkpoint that prevents cells from exiting mitosis prematurely (Kops *et al*., 2005), was found participating in the spindle assembly checkpoint signalling process and was down-regulated in the AKI group in this study, indicating dysregulation of cell cycle events of viable tubular cells in AKI that negatively affect cell proliferation and in turn proximal tubule repair.

Cytoprotective pathways of heat shock proteins (HSPs) such as HSPB-1, and those of stress-activated protein kinases participate in renal stress responses stimulated by injury to the renal tubules (Basile *et al*., 2012). HSPB-1 is induced by stress conditions and enables the inhibition of apoptosis via caspase-apoptotic signalling pathways, participation in signal transduction and regulation of cell development and differentiation (O'Neill *et al*., 2014; Matsumoto *et al*., 2015; Chebotareva *et al*., 2017) (Figure 4.2). HSPB-1 has also been postulated to preserve tight junction formation and reduce inflammation (Basile *et al*., 2012). HSPs, including HSPB-1 function as chaperone proteins to support cell survival by assisting with new protein assembly and the repair of damaged ones, thus hindering the aggregation of proteins under cellular stress (Basile *et al*., 2012; Matsumoto *et al*., 2015; Chebotareva *et al*., 2017).

Additionally, HSPB- 1 also plays a role in modulating reactive oxygen species (ROS) as well as polymerising and depolymerising actin filaments in the cell (Matsumoto *et al*., 2015; Chebotareva *et al*., 2017). Following renal injury, HSPs play an important role in the stabilisation and recovery of the cytoskeletal structure, and HSPB-1 has been reportedly identified in cytoskeletal protein aggregates from injured kidneys (Basile *et al*., 2012). Furthermore, HSPB-1 is a pro-angiogenic factor that influences the formation of new blood vessels from existing vessels during pathogenesis and repair processes (Evans *et al*., 2008). Vascular endothelial growth factor (VEGF) is essential for angiogenesis and cell migration and is reported to induce HSPB-1 phosphorylation using p38 mitogen-activated protein kinase (MAPK) dependent signalling pathways (Evans *et al*., 2008); this finding coincides with the participation of HSPB-1 in the enriched VEGF and MAPK signalling pathways identified in this AKI study (Table 3.6). HSPB- 1 down-regulation in the AKI group compared to the non-AKI group in this study, is indicative of inadequate cell survival and regeneration.

The stress-activated protein kinases (SAPKs) and extracellular-regulated protein kinases (ERKs), participate in MAPK signalling pathways and are activated by cellular stress as a result of oxidants, inflammatory cytokines, and DNA damage in tubular injury (Basile *et al*., 2012); further highlighting the likely involvement of MAPK signalling pathways in AKI (Figure 4.2). Significant induction of HSPB-1 in murine models of ischaemic/reperfusion injury (I/RI)- and cisplatin-induced AKI have also been reported (Basile *et al*., 2012).

Lastly, PAK-6, a significantly down-regulated protein in the AKI group of this study compared to the non-AKI group is involved in the MAPK signalling pathway, suggesting the role of PAK proteins in stress response (Kaur *et al*., 2005). In addition, PAK proteins are implicated in regulating actin cytoskeletal rearrangement, gene transcription, cell cycle progression, cell motility, and apoptosis, which are processes that are essential for cellular repair (Schrantz *et al*., 2004; Ha *et al*., 2015).

The downregulation of the proteins identified in this study that participate in renal PTEC repair after an AKI event, suggests that the persistence of pathways and processes that enable AKI progression or those that indicate maladaptive repair of the renal PTECs are at play or an indication of post-AKI recovery. Maladaptive renal repair is characterised by persistent injury to the kidneys after the recovery phase, with persistent inflammation and accumulation of ECM components observed, and can lead to the development of renal fibrosis as well as progression of CKD (Ferenbach and Bonventre, 2015, 2016; Basile *et al*., 2016). Additionally, maladaptive renal repair may occur when the PTECs are unable to fully differentiate after dedifferentiation during recovery or when proliferating PTECs become arrested in the G2/M phase of the cell cycle, affecting adequate tubular cell proliferation (Ferenbach and Bonventre, 2015, 2016; Basile *et al*., 2016). On the other hand, adaptive renal repair is characterised by sustained recovery and restoration of renal function (Ferenbach and Bonventre, 2015, 2016; Basile *et al*., 2016) (Figure 4.5).

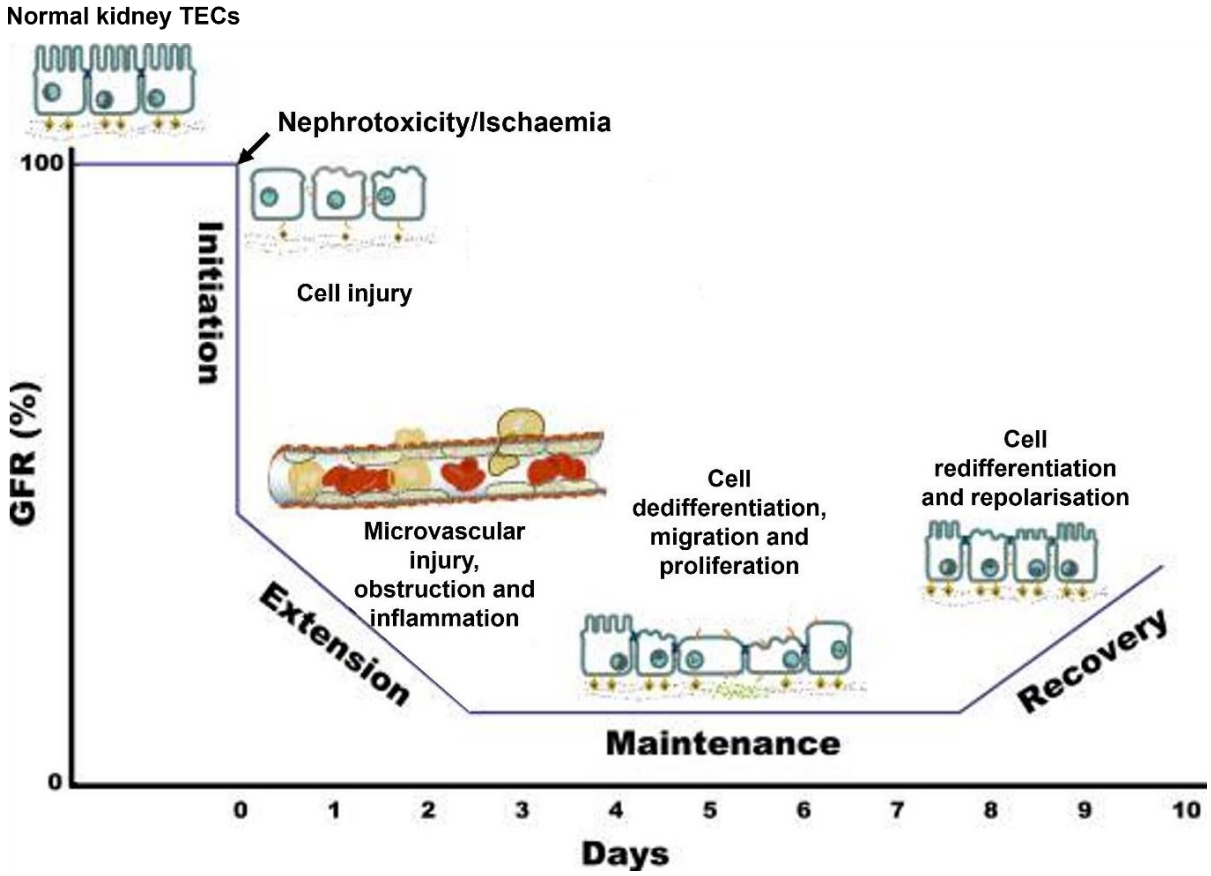


Figure 4.4: Phases of AKI progression due to nephrotoxic or ischaemic-AKI and the associated percentage glomerular filtration rate*. The state of the proximal tubule epithelial cells (PTECs) at each phase is illustrated. Figure is adapted from Basile et al., 2012 (Figure 1).*

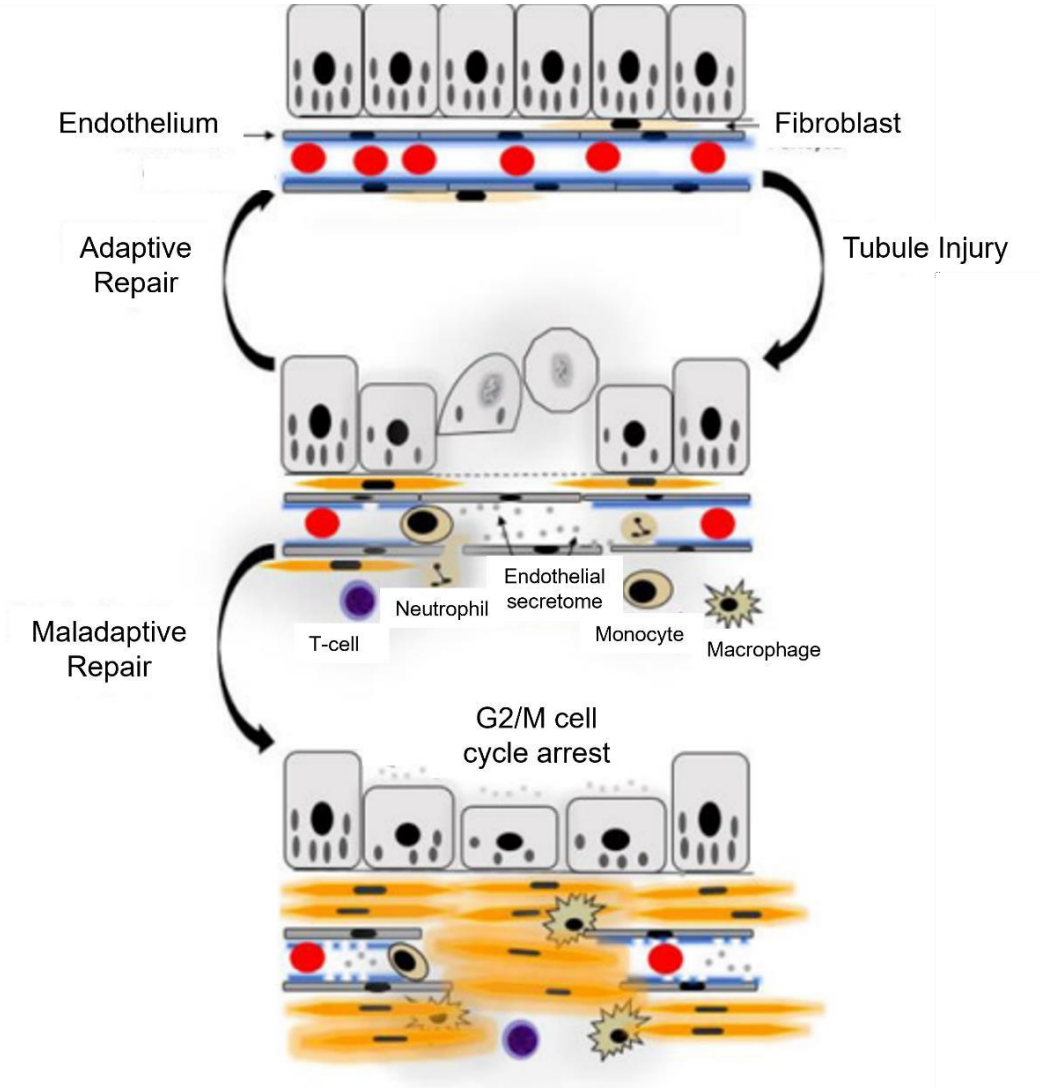


Figure 4.5: Illustration indicating the result of adaptive and maladaptive repair responses to AKI. *Adaptive repair leads to complete kidney repair. Maladaptive repair occurs as a result of severe and/or repeated kidney injury. The role of immune and ECM-related cells in adaptive and maladaptive repair are shown. Figure adapted from Zuk and Bonventre, 2019 (Figure 1).*

The dysregulated candidate markers discussed above are shown to participate in immune responses, including inflammation as well as apoptotic and necrotic renal cell death, and disrupted lipid metabolism, which are essential mechanisms in renal injury and dysfunction. These markers have the potential to function as markers of renal injury or repair and may guide the development of diagnostic tools in being able to determine the onset and progression of AKI, as well as additional mechanisms associated with AKI manifestation and persistence.

4.5) Proteins distinguishing the AKI patient group from the non-AKI group identified by machine learning

4.5.1) Exploring machine learning models for distinguishing study groups in AKI study

The application of machine learning (ML) algorithms to biomarker discovery studies, allows for the identification of robust patterns in large datasets, that can be used to classify groups or predict outcomes based on input information such as relative protein abundances (Ng *et al*., 2023; Torun *et al*., 2023); this can provide new insights into understanding the biology of diseases and drive precision medicine. Currently, MS-based proteomics is an ideal tool for biomarker discovery, allowing for the in-depth analysis of hundreds of samples (Geyer *et al*., 2016; Torun *et al*., 2023), and as with other omics technologies, generates large sets of biological data that are challenging to analyse and interpret using classical statistical methods, such the Student's t-test (Ng *et al*., 2023; Torun *et al*., 2023).

The conventional statistical methods used to analyse data are bottlenecked by large omics datasets that consist of hundreds or thousands of features (i.e. proteins or genes) that do not often share a linear relationship, which also makes it challenging for humans to visualise, recognise and interpret patterns (Ng *et al*., 2023; Torun *et al*., 2023). These methods also assume a specific distribution of the data that cannot always be applied to omics datasets (Ng *et al*., 2023; Torun *et al*., 2023). As a result, ML approaches were applied using OmicLearn in this AKI study to identify proteins that can classify the AKI group from the non-AKI group that may have not been revealed when performing differential abundance testing (using the Student's t-test) in section 3.4.

For this AKI study, data preprocessing, such as normalisation, was not performed in OmicLearn as it was performed prior in Spectronaut (v17). However, options like missing value imputation, which provides strategies for assigning values to missing data to ensure complete data matrices, as well as the various options for the feature selection methods and the ML classifiers were explored, as shown in Table 3.7A-C. Missing values are common in MS-based proteomic data and are due to the low abundance of peptides or peptides that are identified in some but not all runs or the under-sampling of peaks for fragmentation in MS/MS, or due to peptides that infer numerous proteins, among others (Karpievitch *et al*., 2010).

Different ML pipelines were built for this study, and the features selected for each pipeline were used to train ML classifiers, and evaluate the performance of the ML models generated

based on the scores (0 – 1; where 0 represents poor performance and 1 represents excellent performance) given for the performance metrices: accuracy, precision, recall, F1_score, ROC_AUC and PR_AUC, with the intention of generating a ML model that is suitable for exploring the proteomic data obtained from this cohort. The performance metrices evaluated are defined in section 1.5.4.4. The scores of the performance metrices were recorded in Table 3.7A-C to compare the ML pipelines tested and select the best performing ML model, which was later used to identify proteins that may be used to distinguish between the AKI and non-AKI groups.

Firstly, the input proteomic data was unimputed, i.e. consisted of missing values, and the 'none' imputation method was selected to explore the five feature selection methods available (ExtraTrees, k-best (mutual_info_classif), k-best (f_classif), k-best (chi2) and none) as shown in Table 3.7A. The XGBoost classifier, the only classifier available to train on data with missing values, was used in the five ML pipelines explored. Similar scores were observed for the recall and F1_score metrices for all five models, but differences were noted in the accuracy, precision, ROC_AUC and PR_AUC metrices, with the model using the 'none' feature selection method being the best scoring model followed closely by the model using the 'ExtraTrees' feature selection method.

The 'none' feature selection method was not preferred as it meant that the input dataset was not separated into a training and test datasets, and that the selected ML classifier could possibly train and test on the same data, which can result in bias in the protein features selected as important in distinguishing the AKI group from the non-AKI group by the classifier (Torun *et al*., 2023). The 'ExtraTrees' feature selection method was then used to further explore other ML classifiers and imputation strategies.

Secondly, to explore the available ML classifiers (AdaBoost, LogisticRegression, RandomForest, k-NeighboursClassifier, DecisionTree, and LinearSVC) apart from the XGBoost classifier, the 'zero' imputation strategy was employed along with the 'ExtraTrees' feature selection method, as the other ML classifiers are unable to train on and test data with missing values. The ML models using the AdaBoost, LogisticRegression, RandomForest, and XGBoost classifiers showed similar scores across the performance metrices, especially the Accuracy, F1_score and ROC_AUC scores (Table 3.7B). The model using the XGBoost classifier was the best performing, while the model using the k-NeighboursClassifier, was the worst performing with 0 for the F1_score, precision, and recall scores.

Lastly, from Table 3.7A and B the ML pipelines built using the 'ExtraTrees' feature selection and the XGBoost classifier were selected for comparison, irrespective of whether the 'none'

or ‘zero’ imputation methods were used. As a result, Table 3.7C explored the best imputation strategy by comparing ML models using the ‘none’, ‘zero’, and ‘median’ imputation methods along with the ‘ExtraTrees’ feature selection as well as the XGBoost classifier. The best performing models were those using the ‘none’ and ‘median’ imputation methods as they performed similarly across the performance metrices observed.

The ‘none’ imputation strategy was preferred over the ‘median’ as it enabled exploration of the AKI study’s proteomic data as inputted into OmicLearn, without introducing bias or loss of important information about the protein features (that have missing values) caused by data imputation. The preferred ML model explored for the AKI study used the ‘none’ imputation method, with the ‘ExtraTrees’ feature selection method and the XGBoost classifier.

Additionally, a confusion matrix (Figure 3.9) was generated by OmicLearn for the selected ML model to further evaluate the performance of the model and identify mis-classifications to help improve the predictive outcome of the model, especially in imbalanced datasets; ML models may predict the class with majority data for most instances and have a high accuracy score when the model is not predicting the class with minority data (Bhandari, 2020; Markoulidakis *et al*., 2021; Torun *et al*., 2023). The performance metrices evaluated in this study (e.g. accuracy, precision, recall and F1_score) can be calculated from the confusion matrix of the model. The trained XGBoost classifier used in the selected ML model was on average, able to accurately assign 64% of patients with AKI to the AKI group and 80% of patients without AKI to the non-AKI group. This indicated that the ML model selected is better at predicting study control group than the disease group but had the best predictive value for assigning AKI patients to the correct group than the other ML models explored using the dataset provided.

The superior performance of the XGBoost (extreme gradient boosting) classifier has been reported in various fields of study (Yu *et al*., 2020; Wang *et al*., 2021; Desaire *et al*., 2022). It is a decision tree-based boosting algorithm, which combines many weaker single tree classifiers to generate a strong classifier (Wang *et al*., 2021). XGBoost is widely used, and especially preferable in proteomics experiments as it enables training on data with missing values and promises accurate classification of said data (Desaire *et al*., 2022). The XGBoost classifier was also the preferred classifier in studies that used OmicLearn to explore the utility of ML in their proteomic studies (Geyer *et al*., 2021; Karayel *et al*., 2022). A study screening food-derived antihypertensive peptides for the treatment of hypertension compared a model built using the XGBoost classifier with models built using RandomForest, LinearSVC, and k-NearestNeighbour classifiers, and the XGBoost classifier had the greater performance and predictive ability across the performance metrices observed (Wang *et al*., 2021). Additionally,

studies reported on the strong performance of XGBoost models in the prediction of protein submitochondrial localisation (Yu *et al*., 2020) and of miRNA-disease association predictions (Chen *et al*., 2018).

Furthermore, the PCA plot in Figure 3.10 visually represented the comparison of the AKI and non-AKI group using the top 20 contributing features (Table 3.8) utilised by the ML model. The AKI (green) and non-AKI groups are shown to overlap with each other, indicating that these top 20 contributing proteins do not distinguish the study groups well. This suggests that these contributing proteins alone may not be able to differentiate the AKI from the non-AKI group.

4.5.2) Top five protein features used to distinguish the AKI study groups by selected ML model

The preferred ML model explored in OmicLearn for the AKI study, generated a list of 20 protein features that were ranked in order of importance for their ability to distinguish between the AKI and non-AKI groups (Table 3.8). The top five distinguishing protein features were syntaxin-binding protein 5 (STXB-5), profilin-1 (PROF-1), platelet factor 4 (PLF-4), beta-2-microglobulin (B2MG), and integrin alpha-IIb (ITA2B).

Syntaxin-binding proteins (STXBs) interact and bind to syntaxins (STX), a family of transmembrane proteins that are involved in the release of neurotransmitters by inducing the soluble N-ethylmaleimide-sensitive factor activating protein receptor (SNARE) protein complex formation, thus regulating syntaxin exocytotic activity at plasma membranes (Zhu *et al*., 2014; Yang *et al*., 2023). The exocytic machinery in endothelial cells is similarly found in neurons and exocytosis is triggered by injury such as inflammation, physical injury, and release of endothelial cell granules, during pathologic events that damage blood vessels (Zhu *et al*., 2014). The released granule content consists of pro-inflammatory and -thrombotic mediators, adhesion molecules (e.g. P-selectin) that enable leukocyte rolling along endothelial cells which is associated with leukocyte movement across the endothelium to the site of injury, as well as von Willebrand Factors (vWF) that initiate platelets adherence to damaged blood vessels (Zhu *et al*., 2014). STXB-5 is enriched in various tissues including kidney tissue and is associated with alteration of vWF in human blood (Zhu *et al*., 2014).

STXB-5 interacts with STX-1A in neurons and inhibits the release of neurotransmitters from neurons, and an *in vitro* study observed STXB-5 interaction with the exocytic machinery and found it to be a regulator of mammalian endothelial exocytosis as well (Zhu *et al*., 2014). STXB-5 has a SNARE-like motif at the C-terminal domain, which mediates interaction with STX-1A at SNARE domain, inhibiting SNARE complex formation and hindering leukocyte rolling on endothelial cells (Zhu *et al*., 2014). Additionally, the abnormal activation of platelets

in sepsis results in increased SNARE-mediated secretion of platelet granules implicated in pro-inflammatory responses and sepsis-related impairments; STXB-5 was observed to influence secretion of α and dense platelet granules (Xu and Sun, 2015).

Furthermore, autosomal dominant polycystic kidney disease (ADPKD), a disease that causes the formation of cysts in normal kidney tubules, is characterised by decreased cytosolic Ca^{2+} ions that serve to inhibit dysregulated cell proliferation, polarisation, and secretory functions in kidney tubules (Geng *et al*., 2008). The loss-of-function mutations associated with ADPKD encode for proteins that initiate cyst formation; the mutated proteins that would interact with STX-5 to form Ca^{2+} channels under normal conditions, inhibit channel formation and activity, resulting in increased cytosolic Ca^{2+} ions, enabling dysregulated kidney tubule cell proliferation and secretion (Geng *et al*., 2008). STX-5 functions in ER-vesicle transport and fusion and is regulated by STXB-5 (Geng *et al*., 2008). The role of STXB-5 in renal injury has not been frequently studied but the role that STXB-5 plays in regulating syntaxins associated with leukocyte rolling, pro-inflammation, and ADPKD, suggests that with further investigation STXB-5 may serve as potential marker of kidney injury.

PROF-1 is a member of the profilin family of small monomeric actin-binding proteins that consists of four isoforms identified in humans and is involved in regulating actin polymerisation in response to extracellular signals, enhancing ADP to ATP exchange on G-actin as well as cell proliferation, survival, motility, endocytosis, cell cycle and gene transcription (Ding *et al*., 2012; Zheng *et al*., 2014; Mallipattu, 2023). At high concentrations of PROF-1, actin polymerisation is prevented (Zheng *et al*., 2014). Abnormal expression or deletion of PROF-1 influences normal physiology of cells, leading to disease development (Wang *et al*., 2021).

PROF-1 has been reported as a biomarker for the diagnosis and prognosis of cancers, including renal cell carcinomas (RCC). PROF-1 expression in RCC is lower than in normal tissue in the early stage of tumours, whereas high PROF-1 expression is indicative of late stage of tumours, with a worsened prognosis (Wang *et al*., 2021; González *et al*., 2023). Additionally, in podocytes PROF-1 deficiency triggers cell cycle re-entry that results in defective cytokinesis with a loss in rRNA processing, leading to actin-related programmed cell death and a loss in podocytes, affecting kidney filtration abilities (Mallipattu, 2023); this can contribute to AKI pathophysiological mechanisms and has the potential to serve as an AKI marker.

PLF-4, also known as CXCL-4, is a small cytokine belonging to the chemokine family and is released during platelet aggregation from α-granules of activated platelets. PLF-4 has a high affinity for heparin, an anticoagulant, and by neutralising it on the endothelium of blood vessels, can promote blood coagulation (Jansen *et al*., 2017; Cai *et al*., 2020; Liu *et al*., 2023). PLF-4 has been reported to participate in wound repair and inflammation, as it is a

chemoattractant for neutrophils, fibroblasts, and monocytes; PLF-4 secretion is up-regulated when platelets are activated in response to infection or injury and is involved in innate and adaptive immune responses (Cai *et al*., 2020).

PLF-4 complexed with heparin forms an antigen on the surface of platelets and causes the production of IgG antibodies directed against the heparin-PLF-4 complex, resulting in heparin-induced thrombocytopenia (HIT), therefore enabling platelet activation and the release of more PLF-4, thus establishing a positive feedback loop of platelet activation (Liu *et al*., 2023; Thomas *et al*., 2023). HIT places patients at risk of thrombosis, and is associated with skin necrosis and organ ischaemia, and can cause a variety of pathophysiological conditions (Liu *et al*., 2023; Thomas *et al*., 2023). PLF-4 is also involved in inhibition of the haematopoietic process, cell proliferation, differentiation, and activation of immunocytes, and leukocyte activation, trafficking and recruitment at the site of vascular injury (Liu *et al*., 2023).

PLF-4 has been reported to enhance phagocytosis of bacteria and improve the outcome of sepsis in a murine model, additionally, anti-PLF-4 antibodies have been observed in patients with bacterial sepsis (Liu *et al*., 2023), as well as being associated with development of severe AKI (Thomas *et al*., 2023). Hospitalised patients that tested positive for PLF-4 exposure had a 56% higher chance of developing severe AKI, than the patients that tested negative for PLF-4 exposure (Thomas *et al*., 2023). PLF-4 has also been reported to play a role in fibrotic diseases of the skin, lungs and liver when overexpressed (Liu *et al*., 2023).

Lastly, PLF-4 and CXCL-5 were observed on platelets in the kidneys after AKI event (Jansen *et al*., 2018). Additionally, ischaemia/reperfusion injury (I/RI) can lead to activation of blood coagulation and inflammation, which are processes mediated by platelets, and can result in necrotic death of renal tubular epithelial cells (Jansen *et al*., 2017). In I/RI mice, platelet activation aggravates AKI, whereas platelet inhibition has been associated with decreased cellular necrosis and inflammation (Jansen *et al*., 2017). PLF-4 has been associated with pro-inflammatory processes, fibrotic diseases, and AKI, suggesting that it has the potential to serve as a marker of AKI.

B2MG is a component of the major histocompatibility class (MHC) 1 molecules that are expressed on the cell surface of all nucleated cells. The B2MG is the light chain of the MHC 1 molecules and dissociates from the heavy chain during cellular turnover and enters the blood circulation in its monomeric form (Vaidya *et al*., 2008). B2MG has a high sequence homology with the light chains of the immunoglobulin (Ig) family (Bernier, 1980), and functions, along with its heavy chain, in presenting peptide fragments of degraded proteins in the cytoplasm to cytotoxic T-cells, to trigger an immune response against the presented antigen, thus mediating immune responses (Bernier, 1980; Vaidya *et al*., 2008).

B2MG is a low molecular weight protein readily filtered through the glomerulus of the kidney and is mostly reabsorbed by the proximal tubule (Bernier, 1980; Vaidya *et al*., 2008). Upon tubular injury, B2MG reabsorption is hindered, resulting in an increase in urinary B2MG; urinary B2MG is a biomarker of AKI, serving as an early marker of tubular injury that precedes the elevation of serum creatinine (Bethea and Forman, 1990; Vaidya *et al*., 2008; Zeng *et al*., 2014; Jana *et al*., 2023). Upon injury to the glomerulus, serum B2MG accumulates, indicating its inability to be filtered through the kidney and serves as a marker of glomerular injury; high levels of serum B2MG have also been reported in kidney transplant rejection (Bethea and Forman, 1990; Zeng *et al*., 2014; Jana *et al*., 2023). Although serum B2MG may be specifically indicative of glomerular injury, further investigations may reveal additional roles as a marker of AKI.

ITA2B is a subunit chain that non-covalently complexes with the ITB3 chain to form integrin alpha-IIb/beta-III (α2b/β3), a glycoprotein that spans platelet membranes. ITA2B/B3 is a platelet specific integrin and is the most prominent adhesion receptor and membrane protein expressed on the surface of platelets (Payrastre *et al*., 2000; Quinn *et al*., 2003; Corbel *et al*., 2005). As a platelet cell adhesion molecule, ITA2B/B3 is crucial for platelet aggregation, which is initiated after platelet activation occurs. Platelet activation induces conformational changes in ITA2B/B3 from the inactive state, where it is unable to bind ligands (fibrinogen and vWF) to the active state, allowing ITA2B/B3 to bind fibrinogen and forms intercellular bridges between platelets, thus enabling platelet aggregation (Payrastre *et al*., 2000; Quinn *et al*., 2003; Corbel *et al*., 2005).

ITA2B/B3 is also found in platelet α-granule membranes to increase cell surface expression upon activation, increasing platelet adhesion capabilities (Quinn *et al*., 2003). Platelets function to stop bleeding upon injury to vasculature and the mechanisms involved in activating platelet functions are strictly controlled. These controls are lost in pathological conditions and can lead to excessive bleeding or excessive platelet aggregation, which can obstruct blood vessels and result in eventual organ failure (Savage *et al*., 1992; Payrastre *et al*., 2000).

ITA2B/B3 plays a role in the adhesion of platelets to neutrophils and vascular endothelial cells and mediates an inflammatory response (Maas *et al*., 2018). When activated ITA2B/B3 bind fibrinogen, fibrinogen can interact with a neutrophilic adhesion molecule MAC-1 and form neutrophil-platelet complexes that are recruited to the endothelium at the site of inflammation. This indicates that ITA2B/B3 on platelet surface plays a role in neutrophil recruitment (Maas *et al*., 2018). Additionally, activated platelets release pro-inflammatory mediators such as cytokines, chemokines, and growth factors, that support cellular apoptosis and leukocyte recruitment (Maas *et al*., 2018). Furthermore, platelets influence leukocyte recruitment by

changing the chemotactic, adhesive, and proteolytic properties of endothelial cells, and can influence the permeability of the vascular endothelium (Maas *et al*., 2018); this can facilitate the movement of leukocytes across the endothelium to the site of inflammation.

Lastly, *in vivo* studies have revealed that platelet activation is involved I/RI, and that inhibiting platelet activity reduces ischaemic injury in tissues (Hu *et al*., 2011). Clopidogrel, an anti-platelet drug used in an I/RI murine model showed decreased expression of ITA2B/B3 in the kidneys. The control I/RI mice showed widespread positive immunofluorescence for ITA2B in the renal tubules, glomerular capsules, and the small blood vessels, but the I/RI mice treated with clopidogrel showed a significant decline in fluorescence, therefore inhibiting platelet aggregation in the kidney post I/RI (Hu *et al*., 2011). This indicates that ITA2B, in conjunction with the ITB3 subunit can be used to monitor pro-inflammatory events associated with kidney injury.

ITA2B, PLF-4, B2MG and STXB-5 have been shown to participate in immune responses of the innate and/or adaptive immune system (Vaidya *et al*., 2008; Zhu *et al*., 2014; Maas *et al*., 2018; Cai *et al*., 2020; Liu *et al*., 2023), while PROF-1 is involved in regulating actin polymerisation and the cellular processes that promote cellular proliferation and repair (Zheng *et al*., 2014; Mallipattu, 2023); indicating that the top 5 proteins used to distinguish the AKI group from the non-AKI group, participate in cellular processes that occur during injury and recovery from AKI.

4.5.3) Comparing ML top contributing features with statistically derived candidate markers

A Venn diagram was used to identify overlapping protein features between the 20 most important proteins used by the ML model to distinguish the AKI from the non-AKI group (Table 3.8), and the 34 candidate markers identified using the Student's t-test (Figure 3.7). Only 3 proteins, STXB-5, PARN, and PZP were found to overlap both methods of analysis (Figure 3.11). The small overlap between the candidate markers identified by the Student's t-test and the contributing features used by the ML model suggests that the contributing features in ML analysis did not meet the filtering criteria when performing the t-test.

STXB-5 regulates exocytosis in endothelial cells and is triggered by inflammation and physical injury, as described previously (Zhu *et al*., 2014). PZP is a serine protease inhibitor found to modulate inflammatory responses, such as facilitating leukocyte infiltration, and functions as a chaperone, clearing pro-inflammatory cytokines and aggregated proteins (see section 4.4) (Cater *et al*., 2019). PARN primarily regulates mRNA turnover and can destabilise cytoplasmic mRNA with its deadenylation activity, such as p53 mRNA, which can severely affect cell survival when overexpressed (see section 4.4) (Shukla *et al*., 2019; Huynh *et al*., 2023). These

3 proteins are associated with biological processes that can contribute to renal injury and recovery, with pathways of innate and adaptive immune responses, as well as pathways contributing to apoptotic cell death being particularly enriched in both methods of data analysis; further investigation into their roles in AKI may be beneficial in identifying novel AKI biomarkers, especially within the population of PLHIV in South Africa.

The 17 remaining protein features in Table 3.8 were not identified as significantly dysregulated by the Student's t-test, suggesting the potential utility of ML models in distinguishing between conditions or treatment groups with proteins that would have otherwise been overlooked using only classical statistical methods and expanding the search space for novel AKI candidate biomarkers.

This study emphasised the importance of novel AKI candidate biomarkers for early detection of AKI in PLHIV using TDF-ART to ensure that potential renal disease doesn't result in non-adherence of HIV treatment. Patient plasma samples were acquired at a single time point for both patient groups and subjected to plasma proteome profiling. The resulting candidate markers reflect statistically significant proteome changes between the AKI and non-AKI group, providing new potential markers for improved detection of AKI. However, these candidate makers do not provide specific insights to proteome changes that occur as AKI progresses in the patients. As a result, the current experimental study design limits the study from identifying potential markers for early detection.

Future work for this research would be to analyse longitudinal plasma samples of patients with AKI to observe and interrogate the proteome changes during progression/continuation of AKI, which may provide insights related to early markers of AKI. As reported in Geyer *et al*. (2021), longitudinal investigation of serum COVID-19 patient samples allowed for the evaluation of the degree and nature of the serum proteome changes that occurred during the course of infection in patients. This study also analysed COVID-19 serum samples obtained at a single time point and noted that these samples provide insights on potentially regulated proteins, while the longitudinal data revealed fluctuating proteome changes in clusters of specific serum proteins, providing a clear set of potential biomarkers for disease progression.

4.6 Comparing plasma and urinary proteome profiles of the same study groups

4.6.1) Benefits of analysing plasma and urine proteomic profiles of the same study groups

The proteomic profiling of biofluids has become a promising approach for the identification of novel biomarkers. Plasma and urine are the most used media for diagnostic analysis of disease in clinical practice. Pathological conditions result in disturbances of normal cellular functions and pathways, and result in changes in the proteome of the biofluids that serve as

diagnostic markers (Wu *et al*., 2010; Geyer *et al*., 2016; Shama *et al*., 2023).

The use of plasma for diagnostic analysis has been the routine procedure in clinical settings, because it provides an informative view on the health state of the individual as blood connects all major organs and gathers changes in protein composition from the body (Gao, 2015; Geyer *et al*., 2016). Urine, as a filtrate of blood, also collects protein changes that are expelled from the blood and hold a biochemical record of an individual's health state, but the abundance and composition of the urinary proteome is also a direct reflection of changes in the kidney and urogenital tract (Wu *et al*., 2010; Gao, 2015). The use of plasma and urine have fast-tracked the identification and development of novel biomarkers as both media are rich sources of proteins, they are easily accessible with minimal or no invasiveness, and can be acquired repeatedly at relatively large volumes (Wu *et al*., 2010; Geyer *et al*., 2016; Shama *et al*., 2023).

The kidneys function to remove waste, drug, and endogenous metabolites from the body by filtering these products from blood into urine by glomerular filtration (1) (Figure 4.6). The waste products are passively filtered from the kidney's glomeruli, allowing for low molecular weight proteins (< 60kDa) to pass through as well. Subsequently, proximal tubular reabsorption occurs (2), actively allowing for water and other molecules, such as glucose and small proteins to be reabsorbed back into the blood from the proximal tubule (Figure 4.6). Thereafter, proximal tubular secretion occurs (3) and removes ions, such as potassium and other waste material from the blood into urine (Figure 4.6) (Ogobuiro and Tuma, 2023). When large proteins, blood cells or even an increase in molecules that should be reabsorbed are identified in the urine, it is an indication of damage to the glomeruli or proximal tubules, resulting in potential renal dysfunction (Ogobuiro and Tuma, 2023). As a result, monitoring the urine and plasma protein content can reveal changes in the normal functional capabilities of the kidney and aid in predicting and diagnosing pathological renal conditions (Ogobuiro and Tuma, 2023).

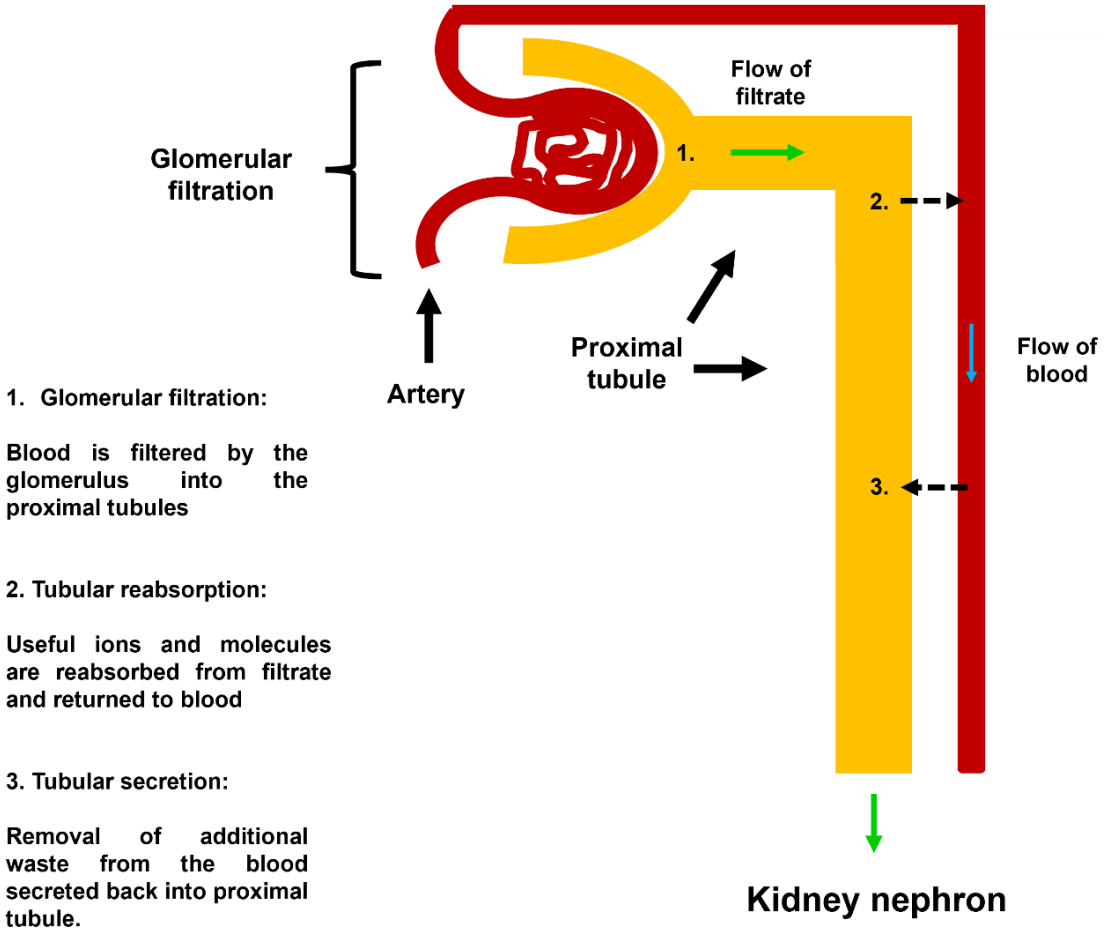


Figure 4.6: Simplified diagram indicating glomerular filtration, tubular reabsorption, and secretion in the kidney nephron.

Although few quantitative proteomic studies have been performed that simultaneously analyse plasma and urine for the same study groups (Araumi *et al*., 2021), the proteome profiling of both biofluids can provide in-depth information about the pathophysiological mechanisms at play in a condition (Qu *et al*., 2023) and can identify candidate markers that would otherwise go undetected. For example, neutrophil gelatinase-associated lipocalin (NGAL), a known marker of AKI, has been found in plasma and urine as an excellent, independent predictor of AKI (Devarajan, 2011).

4.6.2) Comparing plasma and urinary proteome-derived AKI candidate markers

The plasma proteome profiles of the study groups analysed in this AKI study were compared to the urinary proteome profiles of the same study groups. There were 217 candidate protein groups identified across the patient urinary proteome profiles (Table 6.5), while 48 candidate protein groups were identified across the plasma proteome profiles (Table 6.4), at a significance cut-off: q-value < 0.05, absolute log2FC ≥ 0.58; FC ≥ 1.5. Only 5 significantly dysregulated proteins were found overlapping the urinary (u) and plasma (p) protein groups identified (Figure 3.12). LEG-7, CH3L-1, IGFBP-6, HSPB-1, and immunoglobulin kappa variable 1D-33 (KVD-33) are the overlapping proteins reported in this AKI study.

The fold-change of pLEG-7 was lower than that of uLEG-7, and both pLEG-7 and uLEG-7 were lower in the AKI group compared to the non-AKI group (Table 3.8). The role of pLEG-7 and uLEG-7 in AKI has been infrequently studied. Most evidence for LEG-7 has been identified using various forms of microscopy and immunostaining of LEG-7 within renal tubular epithelial cells (RTECs) (Gendronneau *et al*., 2008; Rondanino *et al*., 2011) LEG-7 participates in a variety of cellular functions involved in AKI recovery phase, including cell-cell and cell-ECM adhesion, cell differentiation, and regulation of cell proliferation and apoptosis (Yang *et al*., 2001; Rondanino *et al*., 2011; Sewgobind *et al*., 2021); increased LEG-7 in RTECs is associated with increased epithelial cell motility and migration, promoting wound healing (Rondanino *et al*., 2011; Sewgobind *et al*., 2021).

LEG-7 and -3 are shown to directly interact with one another (Figure 3.8) and play a role in wound repair in simple epithelia of MD canine kidney cells (Rondanino *et al*., 2011; Sewgobind *et al*., 2021). In LEG-3 deficient mice, re-epithelializing of a corneal wound was delayed and a reduction in LEG-7 was also observed, however with the addition of recombinant LEG-7, an increase in wound healing was noted in the mice (Rondanino *et al*., 2011; Sewgobind *et al*., 2021). Plasma LEG-3 is mostly secreted by immune cells and plays a pro-inflammatory and profibrotic role in the kidney when up-regulated after AKI (Kolatsi-Joannou *et al*., 2011; Wang *et al*., 2023). In patients with CKD, increased pLEG-3 is inversely associated with a decline in the eGFR (Kolatsi-Joannou *et al*., 2011; Wang *et al*., 2023). Up-regulation of uLEG-3 is associated with severe renal fibrosis and increased risk of renal disease progression. Additionally, increased uLEG-3 is also inversely associated with decreasing eGFR and increasing serum creatinine (Ou *et al*., 2022). The interaction between LEG-7 and -3, as well as their shared functions in re-epithelialisation of injured kidney tissue suggests that

decreased pLEG-7 and uLEG-7 may be a result of recovery from AKI, and that LEG-7 can potentially serve as a maker of AKI.

In addition, pCH3L-1 had a lower fold-change than uCH3L-1, and uCH3L-1 was greater in the AKI group compared the non-AKI group, while pCH3L-1 was lower in the AKI group compared to the non-AKI group (Table 3.9). Increased uCH3L-1 in AKI has been associated with tubular injury in patients with stage 2 or more severe AKI within 24 hours of injury (Conroy *et al*., 2018; Zhao *et al*., 2020). Likewise, pCH3L-1 secretion is promoted mainly by activated macrophages and neutrophils that are triggered by pro-inflammatory cytokines (e.g. IFN-γ) in response to renal damage (Conroy *et al*., 2018; Zhao *et al*., 2020). Elevated pCH3L-1 and uCH3L-1 have been reported as good predictors of AKI associated with elective cardiac surgery (Conroy *et al*., 2018; Zhao *et al*., 2020). Down-regulation of pCH3L-1 in AKI is associated with increased AKI recovery and is suppressed by anti-inflammatory cytokines (e.g. IL-4) (Conroy *et al*., 2018; Zhao *et al*., 2020).

Furthermore, pHSPB-1, pKVD-33, and pIGFBP-6 showed a similar fold-change to uHSPB-1, uKVD-33, and uIGFBP-6, respectively (Table 3.9). Plasma and urinary HSPB-1 and KVD-33 were lower in the AKI group compared to the non-AKI group, whereas pIGFBP-6 and uIGFBP-6 were greater in the AKI group compared to the non-AKI group (Table 3.9).

Up-regulation of pHSPB-1 and uHSPB-1 has been noted in CKD patients. Plasma HSPB-1 is elevated in response to cellular and renal tissue damage triggered by increased ROS, enhanced apoptosis, and additional activation of chronic inflammation by pro-inflammatory cytokines in CKD (Matsumoto *et al*., 2015; Chebotareva *et al*., 2017); similarly in chronic allograft nephropathy, elevated pHSPB-1 was associated with progressive kidney disfunction (Djamali *et al*., 2007). Urinary HSPB-1 was suggested to be elevated in response to increased cell damage and to compensate for the increased pHSPB-1 levels in CKD (Matsumoto *et al*., 2015; Chebotareva *et al*., 2017), providing a possible explanation for the similar fold-change of pHSPB-1 and uHSPB-1 observed in the AKI group. The elevation of pHSPB-1 and uHSPB-1 in response to renal cell damage during injury, suggests that declined pHSPB-1 and uHSPB-1 levels may be reflective of adaptive repair in AKI recovery.

KVD-33 is a free immunoglobulin (Ig) light chain in plasma responsible for antigen recognition when complexed with an Ig heavy chain, thus mediating adaptive immune response during injury or infections (McHeyzer-Williams *et al*., 2012). Plasma KVD-33, among many other free light chains (FLCs), is up-regulated in response to cellular injury in AKI. FLCs are catabolised in the kidneys by endocytosis of FLCs in the RPTECs and the remaining FLCs are removed in urine (Hutchison *et al*., 2011; Desjardins *et al*., 2013; Parmar, 2022). Increased FLCs can result in excessive endocytosis in the RPTECs, inducing inflammation and triggering apoptotic

pathways, contributing to RPTEC cytotoxicity (Hutchison *et al*., 2011; Desjardins *et al*., 2013; Parmar, 2022). Under healthy conditions low concentrations of FLCs are identified in urine; increased uFLCs occur when pFLC production is increased or when the renal proximal tubule is unable to reabsorb filtered protein (Parmar, 2022). KVD-33 and KV1-39 have been identified as primary sources of FLCs associated with Fanconi's syndrome, which is characterised by renal tubular injury (Hutchison *et al.*, 2011). Renal recovery is associated with reduced pFLCs (Hutchison *et al.*, 2011), suggesting that the downregulation of pKVD-33 and uKVD-33 in the AKI group compared to the non-AKI group may be reflective of renal recovery, which is associated with anti-inflammatory mechanisms of the immune system.

IGFBP-6 has been infrequently studied in kidney diseases and related conditions, but available evidence of IGFBP family indicated functions in kidney development as well as primary and secondary kidney diseases (S. Wang *et al*., 2021). IGFBP-6, along with family members function in inducing cell-ECM interaction, cell migration and apoptosis among others (Oh, 2012; S. Wang *et al*., 2021). Plasma IGFBP-6 is significantly up-regulated in patients with CKD and end stage renal disease (ESRD) (S. Wang *et al*., 2021), and in pre-pubertal children with chronic renal failure (CRF), increased pIGFBP-6 with low glomerular filtration rates (GFR) was observed and directly correlated with elevated levels of pIGFBP-3 using an immunoradiometric assay (Powell *et al*., 1997). Before and after kidney transplantation, as well as at transplant rejection, pIGFBP-6 levels were monitored and reflected renal function status in patients with chronic renal insufficiency (S. Wang *et al*., 2021). Urinary IGFBP-6 was shown to increase with kidney development and is associated with hindered kidney development in children with CKD (Powell *et al*., 1997; S. Wang *et al*., 2021). Additionally, uIGFBP-6 levels decreased in adults with CKD, while those with nephrotic syndrome showed increased uIGFBP-6 (Oh, 2012). This suggests that up-regulated pIGFBP-6 and uIGFBP-6 are associated with progression of renal dysfunction, which is further emphasised in this AKI study by the up-regulation of pIGFBP-6 and uIGFBP-6 in the AKI group.

Plasma and urinary comparison of LEG-7, CH3L-1, HSPB-1, KVD-33, and IGFBP-6 provided evidence for the renal physiological dynamics associated with these candidate markers in renal injury and disease. Observation of these candidate markers can give insight into the progression of AKI phases of injury and recovery; however, it may prove more useful to monitor AKI biomarkers in combination and to develop a biomarker panel to improve the sensitivity and specificity of detecting and monitoring the onset and progression of AKI at a point-of-care setting (van Duijl *et al*., 2021). As a result, protein abundance fluctuations in patients can be used to improve early diagnose of AKI and improve our understanding of underlying pathophysiological mechanisms at play.

Chapter 5

Conclusion

This biomarker discovery study accomplished the aim of identifying and quantifying the differential abundance of plasma proteins that are involved in the development of acute kidney injury (AKI) as well as its recovery in people living with HIV, that were administered a TDF-based first-line antiretroviral therapy regimen. The use of bottom-up mass spectrometry (MS)-based proteomic profiling techniques enabled, for the first time, comprehensive identification of differentially abundant proteins in this cohort of patients in the South African population. The liquid-chromatography MS-based approaches used in this study allowed for high-throughput, unbiased detection of a range of proteins present in undepleted plasma samples, from a well-defined cohort of clinically relevant patients to detect key protein changes associated with TDF-related AKI in this population.

The preliminary proteomics analysis showed that the protocol used for the smaller subset of patients was suitable for further analysis of the complete AKI cohort. Additionally, the quality control measures assessed indicated that the data acquisition for the complete cohort was reliable and reproducible. Moreover, statistically significant differentially abundant markers (i.e. candidate markers) were identified between the AKI and non-AKI groups and were found to participate mainly in biological processes and pathways of immune response, apoptosis, plasma lipoprotein assembly and cholesterol/lipid metabolism, and those of cellular survival and recovery. Furthermore, the machine learning models explored identified proteins that were not found to be statistically significant using the Student's t-test but were able to distinguish between the AKI and non-AKI groups. Lastly, the plasma and urinary proteome profiles of the same study groups were compared and highlighted potential biomarkers associated with pathophysiological mechanisms that underlie AKI.

Although putative protein markers of AKI were identified using a stringent significance criterion, the study presented some limitations. The cohort design for this study measured the AKI and non-AKI samples at a single time point, making it challenging to interrogate the data for early markers of AKI, whereas obtaining longitudinal plasma samples would enable monitoring of the plasma proteome for AKI progression over time from a non-AKI state to an AKI state. This is especially because the plasma proteome is more constant within individuals over time than between individuals. Additionally, the proteomic analysis of crude plasma is challenging due to the large dynamic range of protein abundances spanning more than 10

orders of magnitude, which limits the identification of low abundant plasma proteins that may serve as putative protein biomarkers, thus requiring depletion or enrichment methods that can introduce additional variation into the study.

In addition, MS-based clinical diagnoses are not yet used at a point-of-care level, and as a result the validation of these putative markers requires an independent cohort and orthogonal measurements using techniques such as enzyme-linked immunosorbent assays (ELISA) for definitive diagnoses. This may prove challenging as MS-based proteomic profiling results in the identification of a large number of diverse proteins that may not have commercially available antibodies for protein detection and will then affect the suitability of using the candidate markers reliably for routine diagnosis, and result in designing new antibodies can be costly.

Furthermore, diagnostic tools require large enough differences in protein abundance between case and control groups to identify the presence of a disease or condition, and although large differences may be statistically significant, they may not be biologically relevant. As a result, small differences that are not statistically significant but are biologically important in distinguishing between treatment groups can go ignored in studies solely using classical statistical methods. For example, this study identified known markers of AKI, such as insulin-like growth factor-binding protein 7 (IGFBP-7), tissue inhibitor of metalloproteinase-2 (TIMP- 2), and cystatin-C, but their differential abundance between the AKI and non-AKI groups were statistically insignificant, suggesting how important markers can go unnoticed. Comparing candidate markers identified using classical statistical methods with machine learning models can aid in deriving more value from the dataset and enables better distinguishing between study groups.

Lastly, some of the markers identified for AKI overlap with CKD, which can make differentiating the two related renal diseases challenging, therefore emphasising the need for sensitive and specific biomarkers for disease diagnosis. Additionally, not all AKI-related markers identified indicate only tubular injury which is associated with drug-related AKI but also reflect other mechanisms of AKI, such as glomerular injury, thus highlighting the need to verify and validate the identified candidate markers.

Future work for this study entails improving protein-level coverage by employing new plasma preparation protocols such as extracellular vesicle (EV) profiling that allow for enrichment of EVs whilst depleting high-abundant plasma proteins. Additionally, the candidate markers identified would need to undergo further assessment, to verify the markers using an independent verification cohort and thereafter conducting validation studies, using targeted

MS-based proteomic or orthogonal techniques to attain absolute quantitation of the verified candidate markers.

Chapter 6

Appendix

6.1) Plasma separation protocol

Universal precautions must be used when working with blood. Use of personnel protective equipment is mandatory. Use of eye protection is mandatory and serum/plasma/whole blood must be handled inside a BL2 safety cabinet. Patient blood samples will be treated and processed according to stringent biohazard safety procedures and SOPs employed at CSIR Biosciences for the handling of venepuncture-sourced samples. These procedures are in line with local regulatory requirements and are outlined in detail in the protocols approved under the University of the Witwatersrand Human REC (Medical): Ethics reference number 120612 and CSIR REC: REF 58/2013.

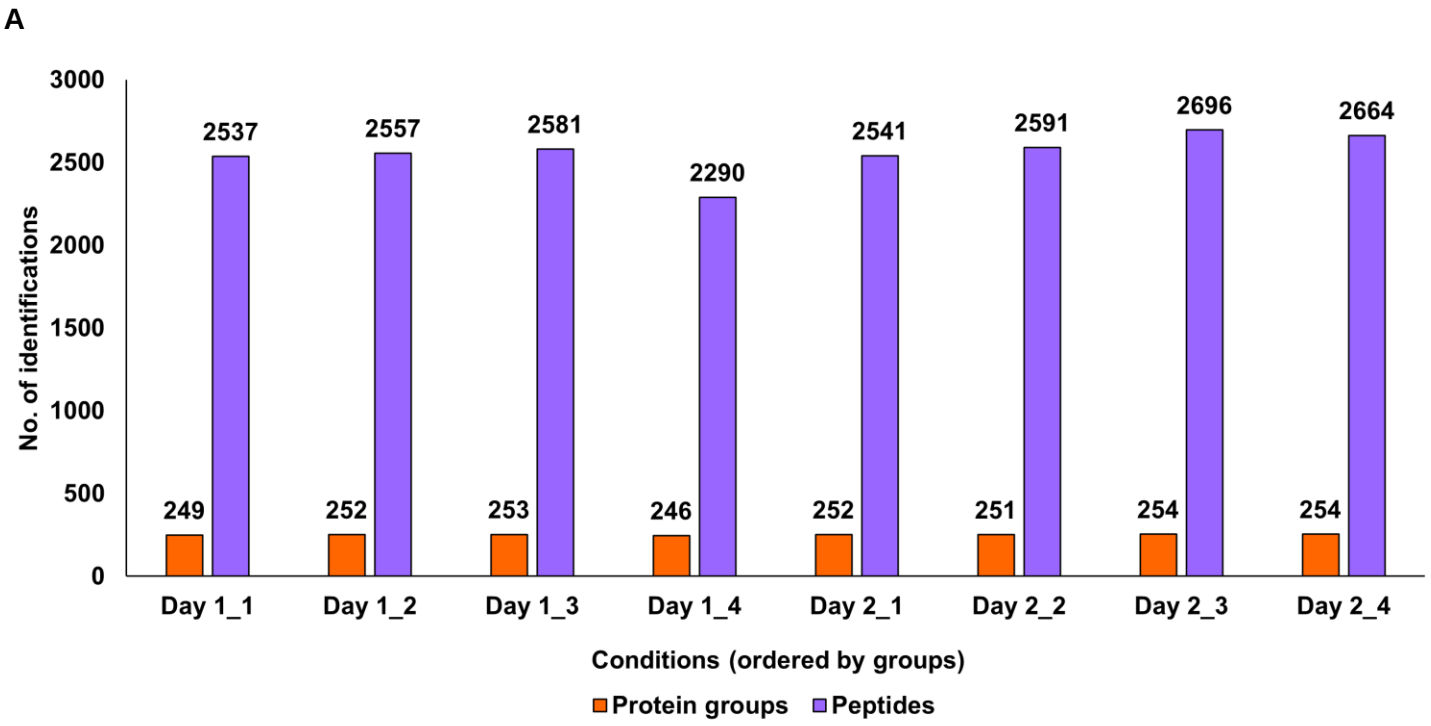
Double spinning is essential and ensures that all cellular contaminants are removed. Therefore, it is imperative that the buffy coat (1st spin) and any pellet (2nd spin) is not touched or disturbed whilst transferring plasma into new tubes.

Procedure:

1. Collect blood (~ 9 ml) in EDTA/ACD tubes and immediately invert mix 8-10 times.
2. Place on wet ice and process within one hour.
3. Spin tubes at 1 000 x g for 10 minutes at 4°C to separate red and white blood cells from plasma.
4. Using clean pipette technique - collect ~3mL plasma per vial into a sterile 15 mL tube. (Remaining sample can be used for PBMC, RBC and other white blood cells isolation).
5. Centrifuge 15 ml tube at 2 500 x g for 10 minutes at 4°C.
6. Aliquot 1 ml plasma into labelled 1.5 mL cryovials and store immediately at -80°C or snap-freeze in liquid nitrogen and transfer to a -80°C freezer.

6.2) Technical reproducibility of protocol with KingFisher™ Flex

The plasma preparation protocol was evaluated using an automated sample handler, the KingFisher™ Flex (Thermo Scientific™, Massachusetts, USA), to ensure that the method would be reproducible for medium-large sample cohorts. This sample preparation workflow had low technically variability between batches analysed on day 1 and 2 (Figure 6.1). The bar graph in Figure 6.1A, shows that the number of identified protein groups and peptides (per replicate) on day 1 and 2 are similar, with an average of 251 proteins groups and 2 557 peptides identified. The median percentage coefficient of variation (%CV) for the box and whisker plots in Figure 6.1B, indicates the technical variability between the replicates in day 1 (18.5%), day 2 (11.7%), as well as the variability between the replicates of both days (22.8%). The low median %CV for technical replicates was observed and therefore, the workflow can maintain a low technical variability when utilising the KingFisher™ Flex.



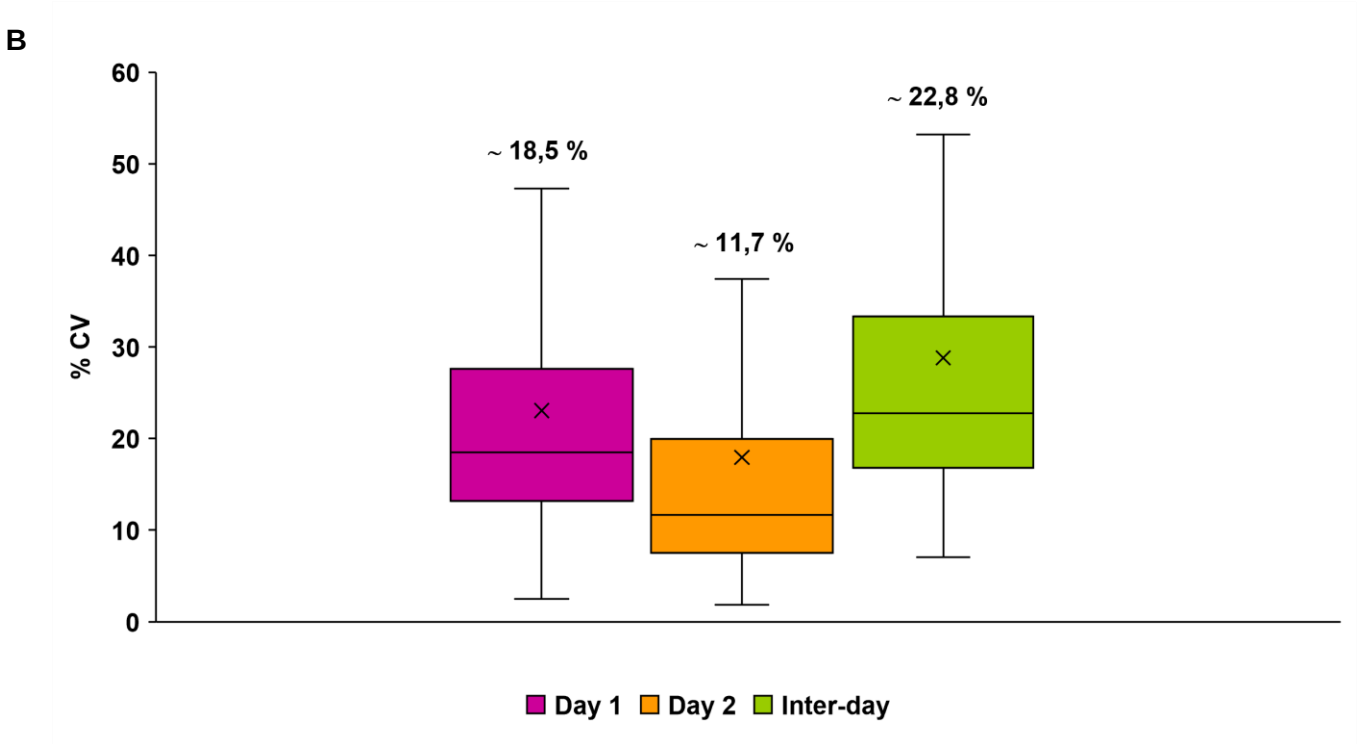


Figure 6.1: Bar graph (A) and boxplot (B) indicating technical reproducibility of the plasma sample preparation workflow between and across Day 1 and 2 using the KingFisher Flex. A *represents protein groups and peptides identified in the plasma sample replicates prepared over day 1 and 2.* ***B*** *indicates intra- and inter-sample replicates showing the median percentage coefficient of variation (% CVs) of the plasma sample replicates on and across day 1 and 2.*

6.3) List of 65 candidate markers identified in preliminary data

Table 6.1: List of differentially abundant proteins generated from the preliminary study. *Spectronaut (v17) generated 65 candidate markers from the preliminary data and are arranged in descending order of average log2 ratio (fold-change) at q-value < 0.05.*

Protein Names	Protein Descriptions	Average Log2 Ratio (fold change)
IF4B	Eukaryotic translation initiation factor 4B	3,46
H2A1-B	Histone H2A type 1-B/E	3,27
KVD-30	Immunoglobulin kappa variable 2D-30	2,89
S10-A8	Protein S100-A8	2,65
HS90-A	Heat shock protein HSP 90-alpha	2,23
T132A	Transmembrane protein 132A	2,15
HBG-1	Haemoglobin subunit gamma-1	2,06
HSPB-1	Heat shock protein beta-1	2,00
S10-A9	Protein S100-A9	1,87
FRIL	Ferritin light chain	1,72
QSOX-1	Sulfhydryl oxidase 1	1,70
TIMP-1	Metalloproteinase inhibitor 1	1,68
COMP	Cartilage oligomeric matrix protein	1,65
SH3L-3	SH3 domain-binding glutamic acid-rich-like protein 3	1,46
DEF-1	Neutrophil defensin 1	1,45
SAA-2	Serum amyloid A-2 protein	1,44
CH3L-1	Chitinase-3-like protein 1	1,37
KI-16B	Kinesin-like protein KIF16B	1,33
FCGBP	IgGFc-binding protein	1,32
RNAS-1	Ribonuclease pancreatic	1,29
FCG3-A	Low affinity immunoglobulin gamma Fc region receptor III-A	1,21
ENPL	Endoplasmin	1,19
LV-319	Immunoglobulin lambda variable 3-19	1,16
HV-158	Immunoglobulin heavy variable 1-58	1,12
C163-A	Scavenger receptor cysteine-rich type 1 protein M130	1,08
PIGR	Polymeric immunoglobulin receptor	1,07
TSP-4	Thrombospondin-4	1,05
LV-657	Immunoglobulin lambda variable 6-57	0,97
PVR	Poliovirus receptor	0,94
LV-223	Immunoglobulin lambda variable 2-23	0,92
LV-321	Immunoglobulin lambda variable 3-21	0,88
ICAM-1	Intercellular adhesion molecule 1	0,83
FCN-3	Ficolin-3	0,82
FCN-2	Ficolin-2	0,74
1433G	14-3-3 protein gamma	0,71
LBP	Lipopolysaccharide-binding protein	0,67
TAGL-2	Transgelin-2	0,65
HV-383	Probable non-functional immunoglobulin heavy variable 3-38-3	0,65
ERMAP	Erythroid membrane-associated protein	0,61
APOA-1	Apolipoprotein A-I	-0,61
PGRP-2	N-acetylmuramoyl-L-alanine amidase	-0,66
LV-545	Immunoglobulin lambda variable 5-45	-0,68

FETU-B	Fetuin-B	-0,69
KVD-40	Immunoglobulin kappa variable 2D-40	-0,71
PROZ	Vitamin K-dependent protein Z	-0,71
IGHG-2	Immunoglobulin heavy constant gamma 2	-0,71
HV-338	Probable non-functional immunoglobulin heavy variable 3-38	-0,72
IGKC	Immunoglobulin kappa constant	-0,73
FHR-4	Complement factor H-related protein 4	-0,77
FINC	Fibronectin	-0,78
HV-364D	Immunoglobulin heavy variable 3-64D	-0,83
KV-37	Probable non-functional immunoglobulin kappa variable 3-7	-0,84
PZP	Pregnancy zone protein	-0,91
MT-1E	Metallothionein-1E	-1,02
KV-621	Immunoglobulin kappa variable 6-21	-1,22
IGD	Immunoglobulin delta heavy chain	-1,28
LV-539	Immunoglobulin lambda variable 5-39	-1,29
LV-211	Immunoglobulin lambda variable 2-11	-1,35
ZSWM-9	Uncharacterized protein ZSWIM9	-1,78

6.4) Hela and Peptide_Pool QC replicate signal acquired during five days of analysis

Boxplots of the original and normalised responses of each Hela (Figure 6.2) and Peptide_Pool (Figure 6.3) QC replicates are shown below. The median normalisation of the original responses enables comparison of the data acquired. The signal of the normalised responses in both Figures 6.2 and 6.3 show a similar signal for Hela and Peptide_Pool_QC 1, 2 and 3, with an increase in the spread of the data acquired in the Hela and Peptide_Pool_QC 4 and 5.

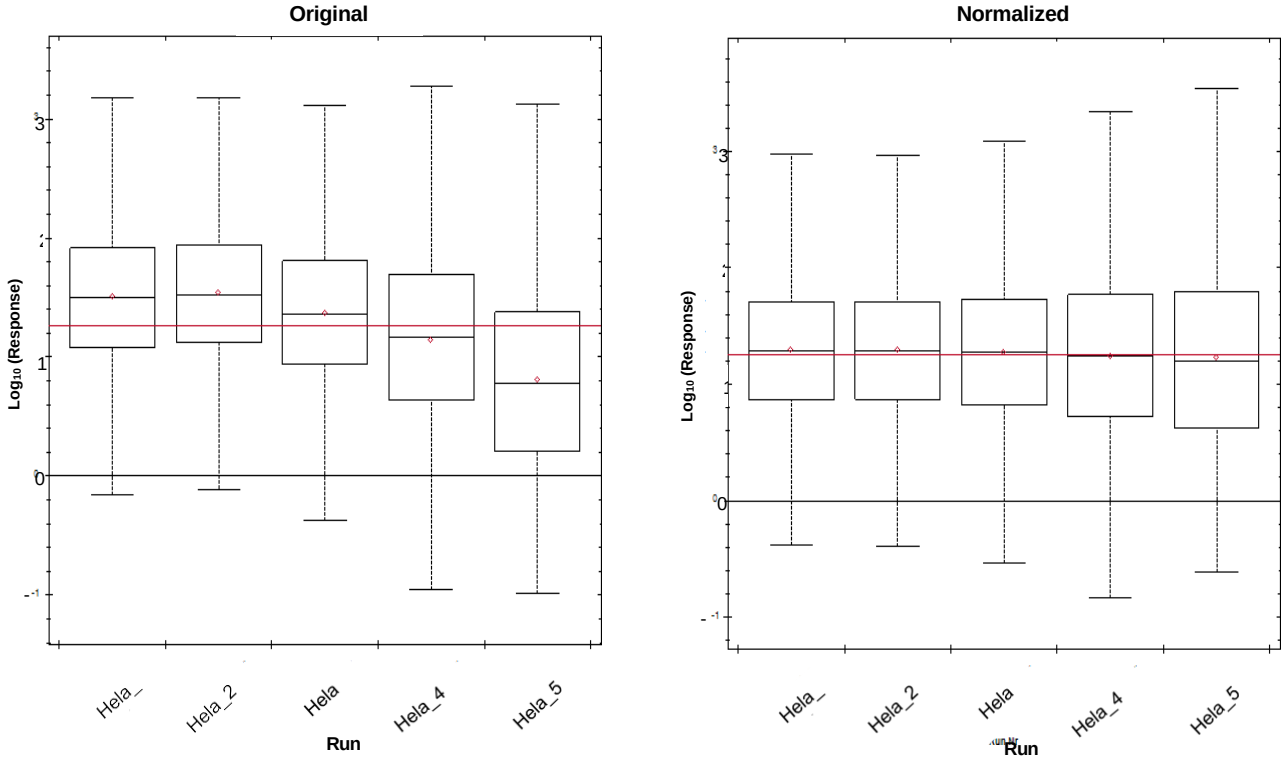


Figure 6.2: Boxplots indicating the original and normalised (log_{10}) response signal for the Hela_QC replicates analysed over 5 days. *The signal of the original responses of the 5 Hela_QC replicates was normalised around the median of the data.*

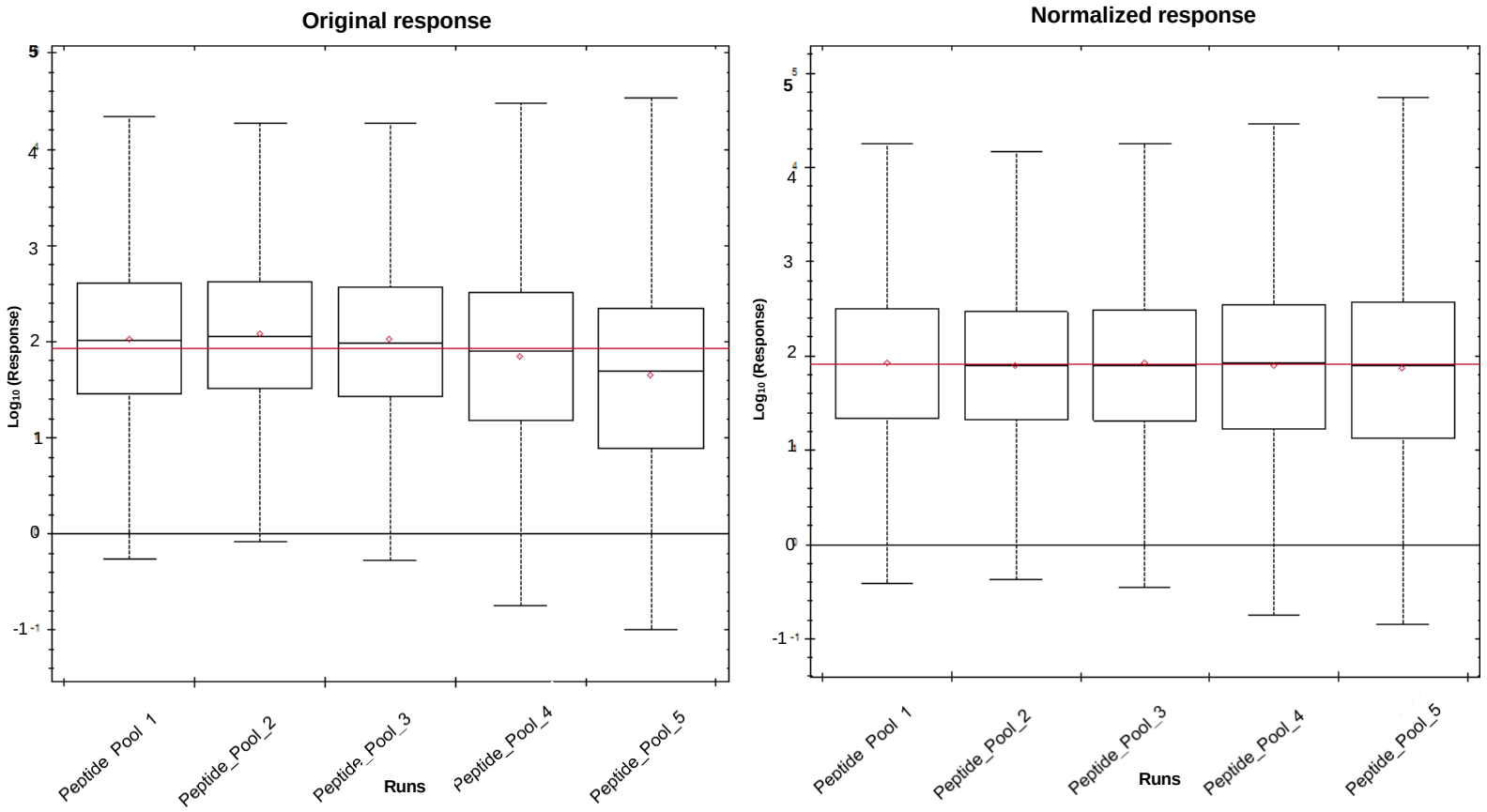


Figure 6.3: Boxplots indicating the original and normalised (log_{10}) response signal for the Peptide_Pool_QC replicates analysed over 5 days. *The signal of the original responses of the 5 Peptide_Pool QC replicates was normalised around the median of the data.*

6.5) List of 34 candidate markers identified in the AKI study

Table 6.2: List of differentially abundant proteins generated from the AKI study. *Spectronaut (v17) generated 34 candidate markers in the AKI study and are arranged in descending order of average log2 ratio (fold-change) at q-value < 0.05.*

Protein Names	Protein Descriptions	Average Log2 Ratio (fold change)
RIMS-3	Regulating synaptic membrane exocytosis protein 3	1,10
PARN	Poly(A)-specific ribonuclease PARN	0,97
IGFBP-6	Insulin-like growth factor-binding protein 6	0,90
KV-116	Immunoglobulin kappa variable 1-16	-0,78
AIG-1	Androgen-induced gene 1 protein	-0,78
KVD-33	Immunoglobulin kappa variable 1D-33	-0,81
STXB-5	Syntaxin-binding protein 5	-0,84
DYH-9	Dynein axonemal heavy chain 9	-0,85
LV-208	Immunoglobulin lambda variable 2-8	-0,87
SIG-16	Sialic acid-binding Ig-like lectin 16	-0,88
CH3L-1	Chitinase-3-like protein 1	-0,89
AL3A-2	Aldehyde dehydrogenase family 3 member A2	-0,89
PZP	Pregnancy zone protein	-0,94
HV-364	Immunoglobulin heavy variable 3-64	-0,98
TFP-11	Tuftelin-interacting protein 11	-1,06
LV-211	Immunoglobulin lambda variable 2-11	-1,09
ICAM-1	Intercellular adhesion molecule 1	-1,10
PAK-6	Serine/threonine-protein kinase PAK 6	-1,12
KNTC-1	Kinetochore-associated protein 1	-1,14
WRP-73	WD repeat-containing protein WRAP73	-1,15
HSPB-1	Heat shock protein beta-1	-1,23
HV-374	Immunoglobulin heavy variable 3-74	-1,33
LV-552	Immunoglobulin lambda variable 5-52	-1,42
DESP	Desmoplakin	-1,43
KVD-37	Probable non-functional immunoglobulinn kappa variable 1D-37	-1,46
DSG-1	Desmoglein-1	-1,47
CTRB-2	Chymotrypsinogen B2	-1,55
APOA-2	Apolipoprotein A-II	-1,70
VINC	Vinculin	-2,15
ABCA-1	Phospholipid-transporting ATPase ABCA1	-2,18
IGHG-1	Immunoglobulin heavy constant gamma 1	-2,52
LV-545	Immunoglobulin lambda variable 5-45	-2,81
LV-312	Immunoglobulin lambda variable 3-12	-3,00
LEG-7	Galectin-7	-3,27

6.6) Feature importance rank values for distinguishing AKI group from non-AKI group using selected ML model

Table 6.3: List of top 20 proteins distinguishing the two study conditions ranked in order of importance using machine learning (ML). *The 20 most important proteins are listed along with their rank value from most to least important. The ML model used: No imputation, ExtraTrees feature selection method and XGBoost classifier, using the OmicLearn platform.*

Protein Names	Feature (protein) importance rank
STXB5	0,08
PROF1	0,067
PLF4	0,054
B2MG	0,052
ITA2B	0,051
1433Z	0,047
RP1BL	0,035
TAGL2	0,031
G3P	0,03
PARN	0,03
RNAS1	0,027
APOH	0,025
UBB	0,021
TBB5	0,019
ZA2G	0,019
PLXA2	0,018
PZP	0,018
A1BG	0,017
O51B5	0,016
DCA13	0,016

6.7) List of identified protein groups in the plasma AKI cohort

Table 6.4: List of identified protein groups generated from the plasma AKI cohort. *Spectronaut (v17) identified 48 protein groups from the AKI plasma cohort and are arranged in descending order of average log2 ratio (fold-change) at a cut-off of: q-value < 0.05, log2FC ≥ 0.58.*

Protein Names	Protein Descriptions	Average Log2 Ratio (fold change)
RIMS3	Regulating synaptic membrane exocytosis protein 3	1,10
PARN	Poly(A)-specific ribonuclease PARN	0,97
IBP6	Insulin-like growth factor-binding protein 6	0,90
PLF4	Platelet factor 4	0,69
OX2G	OX-2 membrane glycoprotein	0,67
HV108	Immunoglobulin heavy variable 1-8	0,66
HV43D	Immunoglobulin heavy variable 3-43D	0,66
BIP	Endoplasmic reticulum chaperone BiP	0,61
PLPL8	Calcium-independent phospholipase A2-gamma	-0,58
KV37	Probable non-functional immunoglobulin kappa variable 3-7	-0,59
SNX33	Sorting nexin-33	-0,59
TRY2	Trypsin-2	-0,68
FTCD	Formimidoyltransferase-cyclodeaminase	-0,68
HV601	Immunoglobulin heavy variable 6-1	-0,70
KVD13	Immunoglobulin kappa variable 1D-13	-0,73
KVD40	Immunoglobulin kappa variable 2D-40	-0,76
KV116	Immunoglobulin kappa variable 1-16	-0,78
AIG1	Androgen-induced gene 1 protein	-0,78
KVD33	Immunoglobulin kappa variable 1D-33	-0,81
STXB5	Syntaxin-binding protein 5	-0,84
DYH9	Dynein axonemal heavy chain 9	-0,85
LV208	Immunoglobulin lambda variable 2-8	-0,87
SIG16	Sialic acid-binding Ig-like lectin 16	-0,88
CH3L1	Chitinase-3-like protein 1	-0,89
AL3A2	Aldehyde dehydrogenase family 3 member A2	-0,89
PZP	Pregnancy zone protein	-0,94
HV364	Immunoglobulin heavy variable 3-64	-0,98
TFP11	Tuftelin-interacting protein 11	-1,06
LV211	Immunoglobulin lambda variable 2-11	-1,09
ICAM1	Intercellular adhesion molecule 1	-1,10
PAK6	Serine/threonine-protein kinase PAK 6	-1,12
KNTC1	Kinetochore-associated protein 1	-1,14
WRP73	WD repeat-containing protein WRAP73	-1,15
HSPB1	Heat shock protein beta-1	-1,23
HV374	Immunoglobulin heavy variable 3-74	-1,33
LV552	Immunoglobulin lambda variable 5-52	-1,42
DESP	Desmoplakin	-1,43
KVD37	Probable non-functional immunoglobulinn kappa variable 1D-37	-1,46

DSG1	Desmoglein-1	-1,47
CTRB2	Chymotrypsinogen B2	-1,55
APOA2	Apolipoprotein A-II	-1,70
VINC	Vinculin	-2,15
ABCA1	Phospholipid-transporting ATPase ABCA1	-2,18
IGHG1	Immunoglobulin heavy constant gamma 1	-2,52
H2A1B	Histone H2A type 1-B/E	-2,55
LV545	Immunoglobulin lambda variable 5-45	-2,81
LV312	Immunoglobulin lambda variable 3-12	-3,00
LEG7	Galectin-7	-3,26

6.8) List of identified protein groups in the urine AKI cohort

Table 6.5: List of identified protein groups generated from the urine AKI cohort. *Spectronaut (v14) identified 217 protein groups from the AKI urine cohort and are arranged in descending order of average log2 ratio (fold-change) at a cut-off of: q-value < 0.05, log2FC ≥ 0.58.*

Protein Names	Protein Descriptions	Average Log2 Ratio (fold-change)
RET4	Retinol-binding protein 4	4,80
APOA4	Apolipoprotein A-IV	4,70
PEDF	Pigment epithelium-derived factor	4,54
CYTC	Cystatin-C	4,00
SAA2	Serum amyloid A-2 protein	3,96
CFAD	Complement factor D	3,94
HBA	Hemoglobin subunit alpha	3,73
SAA1	Serum amyloid A-1 protein	3,61
HBB	Hemoglobin subunit beta	3,55
TBB2A	Tubulin beta-2A chain	3,35
APOA1	Apolipoprotein A-I	3,23
MUC1	Mucin-1	3,09
HBG1	Hemoglobin subunit gamma-1	2,96
MYG	Myoglobin	2,88
CO3	Complement C3	2,83
CRP	C-reactive protein	2,61
FKB1A	Peptidyl-prolyl cis-trans isomerase FKBP1A	2,52
RARR2	Retinoic acid receptor responder protein 2	2,50
PP1RA	Serine/threonine-protein phosphatase 1 regulatory subunit 10	2,43
CO9	Complement component C9	2,42
TTHY	Transthyretin	2,32
B2MG	Beta-2-microglobulin	2,22
LBP	Lipopolysaccharide-binding protein	2,18
IQGA1	Ras GTPase-activating-like protein IQGAP1	2,18
APOC2	Apolipoprotein C-II	2,15
C1S	Complement C1s subcomponent	2,14
IN80D	INO80 complex subunit D	2,14
ASC	Apoptosis-associated speck-like protein containing a CARD	2,13
APOC3	Apolipoprotein C-III	2,13
GPX3	Glutathione peroxidase 3	2,09
IBP4	Insulin-like growth factor-binding protein 4	2,08
DAG1	Dystroglycan	2,08
MAGI3	Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 3	2,07
BLVRB	Flavin reductase (NADPH)	2,02
IBP2	Insulin-like growth factor-binding protein 2	1,98
HEP2	Heparin cofactor 2	1,98
PROF1	Profilin-1	1,97

MIME	Mimecan	1,96
MMP7	Matrilysin	1,91
TALDO	Transaldolase	1,88
A1AT	Alpha-1-antitrypsin	1,83
DCD	Dermcidin	1,81
HRG	Histidine-rich glycoprotein	1,81
SODM	Superoxide dismutase [Mn], mitochondrial	1,79
CH3L1	Chitinase-3-like protein 1	1,77
FIBB	Fibrinogen beta chain	1,76
CO4A	Complement C4-A	1,68
PPIA	Peptidyl-prolyl cis-trans isomerase A;	1,65
B4GT1	Beta-1,4-galactosyltransferase 1	1,59
CO8G	Complement component C8 gamma chain	1,54
OAF	Out at first protein homolog	1,51
TXND5	Thioredoxin domain-containing protein 5	1,51
CRYAB	Alpha-crystallin B chain	1,49
ABRAL	Costars family protein ABRACL	1,46
PCSK1	ProSAAS	1,45
COTL1	Coactosin-like protein	1,45
FIBG	Fibrinogen gamma chain	1,45
CALR	Calreticulin	1,44
FBLN5	Fibulin-5	1,42
CAH1	Carbonic anhydrase 1	1,37
ATS1	A disintegrin and metalloproteinase with thrombospondin motifs 1	1,33
CFAB	Complement factor B	1,32
FAM3C	Protein FAM3C	1,32
ANT3	Antithrombin-III	1,30
DHRS2	Dehydrogenase/reductase SDR family member 2, mitochondrial	1,29
BID	BH3-interacting domain death agonist	1,27
CO4B	Complement C4-B	1,21
TCTP	Translationally-controlled tumor protein	1,20
ZN217	Zinc finger protein 217	1,20
SLIT1	Slit homolog 1 protein	1,18
IBP6	Insulin-like growth factor-binding protein 6	1,17
VTNC	Vitronectin	1,12
VTDB	Vitamin D-binding protein	1,12
WNT9A	Protein Wnt-9a	1,11
FA49B	Protein FAM49B	1,11
ANGT	Angiotensinogen	1,11
TAGL2	Transgelin-2	1,10
FA12	Coagulation factor XII	1,10
MLEC	Malectin	1,10
MYDGF	Myeloid-derived growth factor	1,09
S10A4	Protein S100-A4	1,09
CO7	Complement component C7	1,06
LEG3	Galectin-3	1,05
EF1G	Elongation factor 1-gamma	1,00
MYOC	Myocilin	1,00

A2MG	Alpha-2-macroglobulin	0,99
IDHC	Isocitrate dehydrogenase [NADP] cytoplasmic	0,97
A2AP	Alpha-2-antiplasmin	0,97
PEBP1	Phosphatidylethanolamine-binding protein 1	0,97
FUCO	Tissue alpha-L-fucosidase	0,96
IGHG4	Immunoglobulin heavy constant gamma 4	0,95
NID2	Nidogen-2	0,95
GFI1	Zinc finger protein Gfi-1	0,95
LYPA1	Acyl-protein thioesterase 1	0,93
APOC1	Apolipoprotein C-I	0,93
FUCO2	Plasma alpha-L-fucosidase	0,91
H2AY	Core histone macro-H2A.1	0,85
FA9	Coagulation factor IX	0,82
NGAL	Neutrophil gelatinase-associated lipocalin	0,82
FCGBP	IgGFc-binding protein	0,79
AOC1	Amiloride-sensitive amine oxidase [copper-containing]	0,78
KAIN	Kallistatin	0,77
AK1A1	Aldo-keto reductase family 1 member A1	0,76
ASB18	Ankyrin repeat and SOCS box protein 18	0,74
APOE	Apolipoprotein E	0,74
IC1	Plasma protease C1 inhibitor	0,73
CBR1	Carbonyl reductase [NADPH] 1	0,68
ITIH1	Inter-alpha-trypsin inhibitor heavy chain H1	0,63
AACT	Alpha-1-antichymotrypsin	0,63
GDIR2	Rho GDP-dissociation inhibitor 2	0,62
DPP4	Dipeptidyl peptidase 4	0,59
K2C6A	Keratin, type II cytoskeletal 6A	-0,58
NAPSA	Napsin-A	-0,60
SODE	Extracellular superoxide dismutase [Cu-Zn]	-0,62
CD59	CD59 glycoprotein	-0,63
ANGL2	Angiopoietin-related protein 2	-0,63
CPVL	Probable serine carboxypeptidase CPVL	-0,66
S10AB	Protein S100-A11	-0,69
ANTR1	Anthrax toxin receptor 1	-0,70
HV69D	Immunoglobulin heavy variable 1-69D 1-8	-0,71
ISLR	Immunoglobulin superfamily containing leucine-rich repeat protein	-0,72
LMAN2	Vesicular integral-membrane protein VIP36	-0,76
CERU	Ceruloplasmin	-0,76
DNAS1	Deoxyribonuclease-1	-0,76
KV401	Immunoglobulin kappa variable 4-1	-0,76
FKTN	Fukutin	-0,76
CLIC1	Chloride intracellular channel protein 1	-0,77
ENOG	Gamma-enolase	-0,78
FKBP4	Peptidyl-prolyl cis-trans isomerase FKBP4	-0,79
PAPP2	Pappalysin-2	-0,79
AMPN	Aminopeptidase N	-0,81
HV323	Immunoglobulin heavy variable 3-23	-0,81
ERO1A	ERO1-like protein alpha	-0,83
COMP	Cartilage oligomeric matrix protein	-0,85

CDCA4	Cell division cycle-associated protein 4	-0,86
CBG	Corticosteroid-binding globulin	-0,86
COFA1	Collagen alpha-1(XV) chain	-0,88
IL18	Interleukin-18	-0,89
IBP7	Insulin-like growth factor-binding protein 7	-0,90
PGBM	Basement membrane-specific heparan sulfate proteoglycan core protein	-0,91
UTP20	Small subunit processome component 20 homolog	-0,91
LEG7	Galectin-7	-0,92
IGHA1	Immunoglobulin heavy constant alpha 1	-0,92
SAP3	Ganglioside GM2 activator	-0,92
QPCT	Glutaminyl-peptide cyclotransferase	-0,92
TRFL	Lactotransferrin	-0,93
VASN	Vasorin	-0,93
TETN	Tetranectin	-0,94
IGM	Immunoglobulin mu heavy chain	-0,95
COBA2	Collagen alpha-2(XI) chain	-1,01
HSPB1	Heat shock protein beta-1	-1,02
PSG3	Pregnancy-specific beta-1-glycoprotein 3	-1,04
HMCN1	Hemicentin-1	-1,04
IGKC	Immunoglobulin kappa constant	-1,05
A2GL	Leucine-rich alpha-2-glycoprotein	-1,05
KNG1	Kininogen-1	-1,07
IGK	Immunoglobulin kappa light chain	-1,08
KV320	Immunoglobulin kappa variable 3-20	-1,10
HV307	Immunoglobulin heavy variable 3-7	-1,15
PCDH1	Protocadherin-1	-1,16
KVD33	Immunoglobulin kappa variable 1D-33	-1,18
CAD11	Cadherin-11	-1,18
IGA2	Immunoglobulin alpha-2 heavy chain	-1,19
CLMP	CXADR-like membrane protein	-1,20
DJC13	DnaJ homolog subfamily C member 13	-1,20
UROK	Urokinase-type plasminogen activator	-1,21
PLST	Plastin-3	-1,21
STAT3	Signal transducer and activator of transcription 3	-1,21
AATC	Aspartate aminotransferase, cytoplasmic	-1,23
SYNP2	Synaptopodin-2	-1,23
SI1L1	Signal-induced proliferation-associated 1-like protein 1	-1,27
TRFE	Serotransferrin	-1,27
PIGR	Polymeric immunoglobulin receptor	-1,29
ANXA1	Annexin A1	-1,30
VMO1	Vitelline membrane outer layer protein 1 homolog	-1,32
SPR2B	Small proline-rich protein 2B	-1,32
WNK1	Serine/threonine-protein kinase WNK1	-1,32
ZG16B	Zymogen granule protein 16 homolog B	-1,35
CETP	Cholesteryl ester transfer protein	-1,37
1433S	14-3-3 protein sigma	-1,37
CD14	Monocyte differentiation antigen CD14	-1,38
FABP5	Fatty acid-binding protein 5	-1,42
ANXA2	Annexin A2	-1,44

PGRP2	N-acetylmuramoyl-L-alanine amidase	-1,46
KVD29	Immunoglobulin kappa variable 2D-29	-1,47
MMRN2	Multimerin-2	-1,47
APOD	Apolipoprotein D	-1,48
S10A9	Protein S100-A9	-1,51
TPP1	Tripeptidyl-peptidase 1	-1,55
THBG	Thyroxine-binding globulin	-1,60
IGJ	Immunoglobulin J chain	-1,61
CO6A1	Collagen alpha-1(VI) chain	-1,62
UROM	Uromodulin	-1,66
MXRA8	Matrix remodeling-associated protein 8	-1,73
GARE2	GRB2-associated and regulator of MAPK protein 2	-1,74
ITIH4	Inter-alpha-trypsin inhibitor heavy chain H4	-1,77
CADH1	Cadherin-1	-1,80
S10A8	Protein S100-A8	-1,82
EPCR	Endothelial protein C receptor	-1,84
A1BG	Alpha-1B-glycoprotein	-1,92
FETUA	Alpha-2-HS-glycoprotein	-1,94
EGF	Pro-epidermal growth factor	-2,00
SPR1A	Cornifin-A	-2,07
ARMC5	Armadillo repeat-containing protein 5	-2,13
PDE8B	High affinity cAMP-specific and IBMX-insensitive 3',5'-cyclic phosphodiesterase 8B	-2,16
OSTP	Osteopontin	-2,17
FBLN2	Fibulin-2	-2,25
CRNN	Cornulin	-2,26
CD248	Endosialin	-2,26
PROZ	Vitamin K-dependent protein Z	-2,48
SZT2	KICSTOR complex protein SZT2	-2,50
SPR1B	Cornifin-B	-2,66
IPSP	Plasma serine protease inhibitor	-2,71
S10AE	Protein S100-A14	-2,74
SRPX2	Sushi repeat-containing protein SRPX2	-2,85
THIL	Acetyl-CoA acetyltransferase, mitochondrial	-3,13
SPRR3	Small proline-rich protein 3	-3,16

Chapter 7

References

Abcar, A.C., Chan, L., Yeoh, H., 2004. What To Do for the Patient with Minimally Elevated Creatinine Level? Perm J **8**, 51–53.

Abram, M.E., Ferris, A.L., Shao, W., Alvord, W.G., Hughes, S.H., 2010. Nature, Position, and Frequency of Mutations Made in a Single Cycle of HIV-1 Replication. J Virol **84**, 9864.

Aebersold, R., Mann, M., 2003. Mass spectrometry-based proteomics. Nature **422**, 198–207.

Aggarwal, S., Yadav, A.K., 2016. False Discovery Rate Estimation in Proteomics. Methods Mol Biol **1362**, 119–128.

Al-Amrani, S., Al-Jabri, Z., Al-Zaabi, A., Alshekaili, J., Al-Khabori, M., 2021. Proteomics: Concepts and applications in human medicine. World J Biol Chem **12**, 57–69.

Araumi, A., Osaki, T., Ichikawa, K., Kudo, K., Suzuki, N., Watanabe, S., Watanabe, M., Konta, T., 2021. Urinary and plasma proteomics to discover biomarkers for diagnosing between diabetic nephropathy and minimal change nephrotic syndrome or membranous nephropathy. Biochem Biophys Rep **27**, 101102.

Arts, E.J., Hazuda, D.J., 2012. HIV-1 Antiretroviral Drug Therapy. Cold Spring Harb Perspect Med **2**.

Aslam, B., Basit, M., Nisar, M.A., Khurshid, M., Rasool, M.H., 2017. Proteomics: Technologies and Their Applications. J. Chromatogr. Sci. **55**, 182–196.

Assaram, S., Mashamba-Thompson, T.P., Magula, N.P., 2018. Risk factors and co-morbidities associated with changes in renal function among antiretroviral treatment-naïve adults in South Africa: A chart review. South. Afr. J. HIV Med. **19**.

Autran, B., Carcelain, G., Li, T.S., Blanc, C., Mathez, D., Tubiana, R., Katlama, C., Debré, P., Leibowitch, J., 1997. Positive effects of combined antiretroviral therapy on CD4+ T cell homeostasis and function in advanced HIV disease. Science **277**, 112–116.

Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L., 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science **220**, 868–871.

Barré-Sinoussi, F., Ross, A.L., Delfraissy, J.-F., 2013. Past, present and future: 30 years of HIV research. Nat Rev Microbiol **11**, 877–883.

Basile, D.P., Anderson, M.D., Sutton, T.A., 2012. Pathophysiology of Acute Kidney Injury. Compr Physiol **2**, 1303–1353.

Basile, D.P., Bonventre, J.V., Mehta, R., Nangaku, M., Unwin, R., Rosner, M.H., Kellum, J.A., Ronco, C., 2016. Progression after AKI: Understanding Maladaptive Repair Processes to Predict and Identify Therapeutic Treatments. J Am Soc Nephrol **27**, 687.

Bays, J.L., DeMali, K.A., 2017. Vinculin in cell–cell and cell–matrix adhesions. Cell. Mol. Life Sci. **74**, 2999–3009.

Bbosa, N., Kaleebu, P., Ssemwanga, D., 2019. HIV subtype diversity worldwide. Curr Opin HIV AIDS **14**, 153.

Beeton-Kempen, N., 2020. Proteomics: Principles, Techniques and Applications. Technology Networks. URL: https://www.technologynetworks.com/proteomics/articles/proteomics-principles-techniques-and-applications-343804 (accessed 11.18.23).

Bekker, L.-G., Venter, F., Cohen, K., Goemare, E., Van Cutsem, G., Boulle, A., Wood, R., 2014. Provision of Antiretroviral Therapy in South Africa: The Nuts and Bolts. Antivir Ther **19**, 105–116.

Bell, G., 2016. Replicates and repeats. BMC Biol. **14**.

Benade, M., Long, L., Rosen, S., Meyer-Rath, G., Tucker, J.-M., Miot, J., 2022. Reduction in initiations of HIV treatment in South Africa during the COVID pandemic. BMC Health Serv. Res. **22**, 428.

Bernier, G.M., 1980. β_2-Microglobulin: Structure, Function and Significance. Vox Sang. **38**, 323–327.

Bethea, M., Forman, D.T., 1990. Beta 2-microglobulin: its significance and clinical usefulness. Ann Clin Lab Sci **20**, 163–168.

Bhandari, A., 2020. Understanding & Interpreting Confusion Matrix in Machine Learning (Updated 2024). Analytics Vidhya. URL: https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/ (accessed 3.11.24).

Bittremieux, W., Tabb, D.L., Impens, F., Staes, A., Timmerman, E., Martens, L., Laukens, K., 2018. Quality control in mass spectrometry-based proteomics. Mass Spectrom. Rev. **37**, 697–711.

Black, L.M., Lever, J.M., Agarwal, A., 2019. Renal Inflammation and Fibrosis: A Double-edged Sword. J Histochem Cytochem. **67**, 663–681.

Bonavia, A., Singbartl, K., 2018. A review of the role of immune cells in acute kidney injury. Pediatr Nephrol **33**, 1629–1639.

Bonventre, J.V., Yang, L., 2011. Cellular pathophysiology of ischemic acute kidney injury. J Clin Invest **121**, 4210–4221.

Boswell, M.T., Rossouw, T.M., 2017. Approach to acute kidney injury in HIV-infected patients in South Africa. South. Afr. j. HIV med. **18**.

Broder, S., 2010. The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. Antiviral Res **85**, 1–18.

Broder, S., 1990. Clinical applications of 3′-azido-2′,3′- dideoxythymidine (AZT) and related dideoxynucleosides. Med. Res. Rev. **10**, 419–439.

Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinović, S.M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., Vitek, O., Rinner, O., Reiter, L., 2015. Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues*[S]. Mol Cell Proteomics **14**, 1400–1410.

Bruderer, R., Bernhardt, O.M., Gandhi, T., Xuan, Y., Sondermann, J., Schmidt, M., Gomez-Varela, D., Reiter, L., 2017. Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results*. Mol Cell Proteomics **16**, 2296–2309.

Buchbinder, S.P., Liu, A., 2011. Pre-Exposure Prophylaxis and the Promise of Combination Prevention Approaches. AIDS Behav **15**, 72–79.

Bulbul, M.C., Dagel, T., Afsar, B., Ulusu, N.N., Kuwabara, M., Covic, A., Kanbay, M., 2018. Disorders of Lipid Metabolism in Chronic Kidney Disease. Blood Purif **46**, 144–152.

Burger, B., Vaudel, M., Barsnes, H., 2021. Importance of Block Randomization When Designing Proteomics Experiments. J. Proteome Res. **20**, 122–128.

Cai, G., Zhang, X., Hong, Q., Shao, F., Shang, X., Fu, B., Feng, Z., Lin, H., Wang, J., Shi, S., Yin, Z., Chen, X., 2008. Tissue inhibitor of metalloproteinase-1 exacerbated renal interstitial fibrosis through enhancing inflammation. Nephrol Dial Transplant **23**, 1861–1875.

Cai, Z., Greene, M.I., Zhu, Z., Zhang, H., 2020. Structural Features and PF4 Functions that Occur in Heparin-Induced Thrombocytopenia (HIT) Complicated by COVID-19. Antibodies (Basel) **9**, 52.

Campbell-Yesufu, O.T., Gandhi, R.T., 2011a. Update on Human Immunodeficiency Virus (HIV)-2 Infection. Clin Infect Dis **52**, 780–787.

Cater, J.H., Kumita, J.R., Zeineddine Abdallah, R., Zhao, G., Bernardo-Gancedo, A., Henry, A., Winata, W., Chi, M., Grenyer, B.S.F., Townsend, M.L., Ranson, M., Buhimschi, C.S., Charnock-Jones, D.S., Dobson, C.M., Wilson, M.R., Buhimschi, I.A., Wyatt, A.R., 2019. Human pregnancy zone protein stabilizes misfolded proteins including preeclampsia- and Alzheimer's-associated amyloid beta peptide. Proc Natl Acad Sci **116**, 6101–6110.

Catherman, A.D., Skinner, O.S., Kelleher, N.L., 2014. Top-Down Proteomics: Facts and Perspectives. Biochem Biophys Res Commun **445**, 683–693.

CDC: HIV Prevention Division, 2023. Prevention | HIV Basics | HIV/AIDS | CDC. URL: https://www.cdc.gov/hiv/basics/prevention.html (accessed 2.22.23).

Chandramouli, K., Qian, P.-Y., 2009. Proteomics: Challenges, Techniques and Possibilities to Overcome Biological Sample Complexity. Hum Genomics Proteomics **2009**, 239204.

Chebotareva, N., Bobkova, I., Shilov, E., 2017. Heat shock proteins and kidney disease: perspectives of HSP therapy. Cell Stress Chaperones **22**, 319–343.

Chekol Abebe, E., Tilahun Muche, Z., Behaile T/Mariam, A., Mengie Ayele, T., Mekonnen Agidew, M., Teshome Azezew, M., Abebe Zewde, E., Asmamaw Dejenie, T., Asmamaw Mengstie, M., 2022. The structure, biosynthesis, and biological roles of fetuin-A: A review. Front Cell Dev Biol **10**, 945287.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., Ma'ayan, A., 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics **14**, 128.

Chen, X., Huang, L., Xie, D., Zhao, Q., 2018. EGBMMDA: Extreme Gradient Boosting Machine for MiRNA-Disease Association prediction. Cell Death Dis **9**, 1–16.

Chesney, M.A., 2000. Factors Affecting Adherence to Antiretroviral Therapy. Clin. Infect. Dis. **30**, S171–S176.

Cho, W.C.S., 2007. Proteomics Technologies and Challenges. Genom. Proteom. Bioinform. **5**, 77–85.

Choi, M., Chang, C.-Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., Vitek, O., 2014. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics **30**, 2524–2526.

Chomarat, P., Banchereau, J., 1998. Interleukin-4 and Interleukin-13: Their Similarities and Discrepancies. Int. Rev. Immunol. **17**, 1–52.

Clavel, F., Guétard, D., Brun-Vézinet, F., Chamaret, S., Rey, M.-A., Santos-Ferreira, M.O., Laurent, A.G., Dauguet, C., Katlama, C., Rouzioux, C., Klatzmann, D., Champalimaud, J.L., Montagnier, L., 1986. Isolation of a New Human Retrovirus from West African Patients with AIDS. Science **233**, 343–346.

Cohen, M.S., Chen, Y.Q., McCauley, M., Gamble, T., Hosseinipour, M.C., Kumarasamy, N., Hakim, J.G., Kumwenda, J., Grinsztejn, B., Pilotto, J.H.S., Godbole, S.V., Mehendale, S., Chariyalertsak, S., Santos, B.R., Mayer, K.H., Hoffman, I.F., Eshleman, S.H., Piwowar-Manning, E., Wang, L., Makhema, J., Mills, L.A., Bruyn, G. de, Sanne, I., Eron, J., Gallant, J., Havlir, D., Swindells, S., Ribaudo, H., Elharrar, V., Burns, D., Taha, T.E., Nielsen-Saines, K., Celentano, D., Essex, M., Fleming, T.R., 2011a. Prevention of HIV-1 Infection with Early Antiretroviral Therapy. N Engl J Med **365**, 493.

Cohen, M.S., Shaw, G.M., McMichael, A.J., Haynes, B.F., 2011b. Acute HIV-1 Infection. N Engl J Med **364**, 1943–1954.

Conroy, A.L., Hawkes, M.T., Elphinstone, R., Opoka, R.O., Namasopo, S., Miller, C., John, C.C., Kain, K.C., 2018. Chitinase-3-like 1 is a biomarker of acute kidney injury and mortality in paediatric severe malaria. Malar. J. **17**, 82.

Corbel, C., Vaigot, P., Salaun, J., 2005. alphaIIb integrin, a novel marker for hemopoietic progenitor cells. Int. J. Dev. Biol. **49**, 279–284.

Cowled, P., Fitridge, R., 2011. Pathophysiology of Reperfusion Injury, in: Fitridge, R., Thompson, M. (Eds.), Mechanisms of Vascular Disease: A Reference Book for Vascular Specialists. University of Adelaide Press, Adelaide (AU).

Craigie, R., Bushman, F.D., 2012. HIV DNA Integration. Cold Spring Harb Perspect Med **2**, a006890.

Davies, N.E., 2013. Fixed-dose combination for adults accessing antiretroviral therapy. South. Afr. j. HIV med. **14**, 41–43.

De Clercq, E., 2013. The Holý Trinity: The Acyclic Nucleoside Phosphonates, in: De Clercq, E. (Ed.), Advances in Pharmacology, Chapter 8: Antiviral Agents. Academic Press, pp. 293–316.

De Clercq, E., 2004. Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs): Past, Present, and Future. Chem. Biodiversity **1**, 44–64.

de Jesus, J.R., da Silva Fernandes, R., de Souza Pessôa, G., Raimundo, I.M., Arruda, M.A.Z., 2017. Depleting high-abundant and enriching low-abundant proteins in human serum: An evaluation of sample preparation methods using magnetic nanoparticle, chemical depletion and immunoaffinity techniques. Talanta **170**, 199–209.

De Loor, J., Decruyenaere, J., Demeyere, K., Nuytinck, L., Hoste, E.A.J., Meyer, E., 2016. Urinary chitinase 3-like protein 1 for early diagnosis of acute kidney injury: a prospective cohort study in adult critically ill patients. Crit Care **20**, 38.

De Waal, R., Cohen, K., Fox, M.P., Stinson, K., Maartens, G., Boulle, A., Igumbor, E.U., Davies, M.-A., 2017. Changes in estimated glomerular filtration rate over time in South African HIV-1-infected patients receiving tenofovir: a retrospective cohort study. J. Int. AIDS Soc. **20**.

Decramer, S., de Peredo, A.G., Breuil, B., Mischak, H., Monsarrat, B., Bascands, J.-L., Schanstra, J.P., 2008. Urine in Clinical Proteomics. Mol. Cell. Proteomics **7**, 1850–1862.

Deeks, S.G., Overbaugh, J., Phillips, A., Buchbinder, S., 2015. HIV infection. Nat Rev Dis Primers **1**, 1–22.

Denecke, B., Gräber, S., Schäfer, C., Heiss, A., Wöltje, M., Jahnen-Dechent, W., 2003. Tissue distribution and activity testing suggest a similar but not identical function of fetuin-B and fetuin-A. Biochem J **376**, 135–145.

Department of Health and Human Services, 2022. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents with HIV. URL: https://clinicalinfo.hiv.gov/en/guidelines/hiv-clinical-guidelines-adult-and-adolescent-arv/tests-initial-assessment-follow-up (accessed 3.9.23).

Desaire, H., Go, E.P., Hua, D., 2022. Advances, obstacles, and opportunities for machine learning in proteomics. Cell Rep. Phys. Sci. **3**, 101069.

Desjardins, L., Liabeuf, S., Lenglet, A., Lemke, H.-D., Vanholder, R., Choukroun, G., Massy, Z.A., 2013. Association between Free Light Chain Levels, and Disease Progression and Mortality in Chronic Kidney Disease. Toxins (Basel) **5**, 2058–2073.

Devarajan, P., 2011. Biomarkers for the Early Detection of Acute Kidney Injury. Curr Opin Pediatr **23**, 194–200.

Ding, Z., Bae, Y.H., Roy, P., 2012. Molecular insights on context-specific role of profilin-1 in cell migration. Cell Adh Migr **6**, 442–449.

Djamali, A., Sadowski, E.A., Muehrer, R.J., Reese, S., Smavatkul, C., Vidyasagar, A., Fain, S.B., Lipscomb, R.C., Hullett, D.H., Samaniego-Picota, M., Grist, T.M., Becker, B.N., 2007. BOLD-MRI assessment of intrarenal oxygenation and oxidative stress in patients with chronic kidney allograft dysfunction. Am J Physiol Renal Physiol **292**, F513–F522.

Duwal, S., Schütte, C., von Kleist, M., 2012. Pharmacokinetics and Pharmacodynamics of the Reverse Transcriptase Inhibitor Tenofovir and Prophylactic Efficacy against HIV-1 Infection. PLoS ONE **7**, e40382.

Egger, M., May, M., Chêne, G., Phillips, A.N., Ledergerber, B., Dabis, F., Costagliola, D., Monforte, A.D., Wolf, F. de, Reiss, P., Lundgren, J.D., Justice, A.C., Staszewski, S., Leport, C., Hogg, R.S., Sabin, C.A., Gill, M.J., Salzberger, B., Sterne, J.A., 2002. Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. The Lancet **360**, 119–129.

El-Hefnawy, S.M., Kasemy, Z.A., Eid, H.A., Elmadbouh, I., Mostafa, R.G., Omar, T.A., Kasem, H.E., Ghonaim, E.M., Ghonaim, M.M., Saleh, A.A., 2022. Potential impact of serpin peptidase inhibitor clade (A) member 4 SERPINA4 (rs2093266) and SERPINA5 (rs1955656) genetic variants on COVID-19 induced acute kidney injury. Hum Gene (Amst) **32**, 101023.

Evans, I.M., Britton, G., Zachary, I.C., 2008. Vascular endothelial growth factor induces heat shock protein (HSP) 27 serine 82 phosphorylation and endothelial tubulogenesis via protein kinase D and independent of p38 kinase. Cell. Signal. **20**, 1375–1384.

Evering, T.H., Markowitz, M., 2007. Raltegravir (MK-0518): an integrase inhibitor for the treatment of HIV-1. Drugs Today (Barc) **43**, 865–877.

Fabian, J., Kalyesubula, R., Mkandawire, J., Hansen, C.H., Nitsch, D., Musenge, E., Nakanga, W.P., Prynn, J.E., Dreyer, G., Snyman, T., Ssebunnya, B., Ramsay, M., Smeeth, L., Tollman, S., Naicker, S., Crampin, A., Newton, R., George, J.A., Tomlinson, L., Banda, L., Bello, S., Branson, K., Chisambo, C., Chithodwe, O., Kanyenda, C., Katundu, C., Kayange, N., Kayolo, M., Kuchipanga, V., Kyumba, D.M., Malunga, A., Mvula, B., Mweso, E., Mwiba, E., Ngwira, L., Nkhwazi, L., Thindwa, M., Adams, I., Barrow, K., Beltramo, C., Bouter, C., Candy, G., Chipungu, S., Chitagu, T., Dlamini, P., Casas, X.G.O., Gondwe, M., Gumede, P., Kabudula, C., Kagodora, B., Kahn, K., Khambule, L., Khoza, B., Khoza, D., Khoza, S., Khoza, W., Levin, J., Lesolang, D., Mabuza, M., Maher, H., Mahime, N., Malupi, W., Maphatahanyi, G., Mashaba, N., Mathebula, G., Mayindi, B., Mdaka, B., Mhembere, M., Mkansi, M., Mnisi, R., Mnisi, V., Mogane, C., Mokoena, T., Ndlovu, W., Ndlovu, Z., Ngobeni, F., Ngobeni, K., Ngobeni, N., Ngobeni, T.,

Ngwarai, N., Nkuna, D., Ntimane, M., Ntimane, T., Nxumalo, O., Ohene-Kwofie, D., Sihlangu, F., Silubane, B., Soo, C., Tibane, J., Twine, R., Ubisi, S., Vangu, M.-D.-T., Wade, A.N., Wafawanaka, F., Asiki, G., Bukenya, D., Erone, I., Tumwekase, G., Kabunga, E., Kakande, A., Kaleebu, P., Kamali, A., Kiranda, R.A., Kushemererwa, S., Kwizera, M., Nabukenya, K.J., Nakimera, T., Naluggwa, C., Nassimbwa, S., Nyirenda, M., Nabwato, R., Makanga, R., Seeley, J., Sejjemba, N.E., Seremba, G., Alumadri, V., Bird, N., 2022. Measurement of kidney function in Malawi, South Africa, and Uganda: a multicentre cohort study. Lancet Glob Health **10**, e1159–e1169.

Ferenbach, D.A., Bonventre, J.V., 2016. Acute kidney injury and chronic kidney disease: from the laboratory to the clinic. Nephrol Ther **12**, S41–S48.

Ferenbach, D.A., Bonventre, J.V., 2015. Mechanisms of maladaptive repair after AKI leading to accelerated kidney ageing and CKD. Nat Rev Nephrol **11**, 264–276.

Ferguson, T.F., Stewart, K.E., Funkhouser, E., Tolson, J., Westfall, A.O., Saag, M.S., 2002. Patient-perceived barriers to antiretroviral adherence: Associations with race. AIDS Care **14**, 607–617.

Fernandez-Montero, J.V., Eugenia, E., Barreiro, P., Labarga, P., Soriano, V., 2013. Antiretroviral drug-related toxicities – clinical spectrum, prevention, and management. Expert Opin. Drug Saf. **12**, 697–707.

Festing, M.F.W., 2020. The “completely randomised” and the “randomised block” are the only experimental designs suitable for widespread use in pre-clinical research. Sci Rep **10**, 17577.

Fischl, M.A., Richman, D.D., Grieco, M.H., Gottlieb, M.S., Volberding, P.A., Laskin, O.L., Leedom, J.M., Groopman, J.E., Mildvan, D., Schooley, R.T., 1987. The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial. N Engl J Med **317**, 185–191.

Forsgard, N., Salehpour, M., Possnert, G., 2009. Accelerator mass spectrometry in the attomolar concentration range for 14C-labeled biologically active compounds in complex matrixes. J. Anal. At. Spectrom. **25**, 74–78.

Frankel, A.D., Young, J.A.T., 1998. HIV-1: Fifteen Proteins and an RNA. Annu. Rev. Biochem. **67**, 1–25.

Frescura, L., Godfrey-Faussett, P., A, A.F., El-Sadr, W., Syarif, O., Ghys, P.D., Group, on and behalf of the 2025 testing treatment target W., 2022. Achieving the 95 95 95 targets for all: A pathway to ending AIDS. PLoS ONE **17**.

Fu, R., Jiang, X., Li, G., Zhu, Y., Zhang, H., 2022. Junctional complexes in epithelial cells: sentinels for extracellular insults and intracellular homeostasis. The FEBS Journal **289**, 7314–7333.

Gallant, J.E., Deresinski, S., 2003. Tenofovir Disoproxil Fumarate. Clin. Infect. Dis. **37**, 944–950.

Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M., Hahn, B.H., 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature **397**, 436–441.

Gao, Y., 2015. Urine Is a Better Biomarker Source Than Blood Especially for Kidney Diseases, in: Gao, Y. (Ed.), Urine Proteomics in Kidney Disease Biomarker Discovery, Advances in Experimental Medicine and Biology. Springer Netherlands, Dordrecht, pp. 3–12.

Gendronneau, G., Sidhu, S.S., Delacour, D., Dang, T., Calonne, C., Houzelstein, D., Magnaldo, T., Poirier, F., 2008. Galectin-7 in the Control of Epidermal Homeostasis after Injury. MBoC **19**, 5541–5549.

Geng, L., Boehmerle, W., Maeda, Y., Okuhara, D.Y., Tian, X., Yu, Z., Choe, C., Anyatonwu, G.I., Ehrlich, B.E., Somlo, S., 2008. Syntaxin 5 regulates the endoplasmic reticulum channel-release properties of polycystin-2. Proc Natl Acad Sci **105**, 15920–15925.

George, C., Yako, Y.Y., Okpechi, I.G., Matsha, T.E., Kaze Folefack, F.J., Kengne, A.P., 2018. An African perspective on the genetic risk of chronic kidney disease: a systematic review. BMC Med. Genet. **19**, 187.

German Advisory Committee Blood, 2016. Human Immunodeficiency Virus (HIV). Transfus Med Hemother **43**, 203.

Geyer, P.E., Arend, F.M., Doll, S., Louiset, M., Virreira Winter, S., Müller-Reif, J.B., Torun, F.M., Weigand, M., Eichhorn, P., Bruegel, M., Strauss, M.T., Holdt, L.M., Mann, M., Teupser, D., 2021. High-resolution serum proteome trajectories in COVID-19 reveal patient-specific seroconversion. EMBO Mol Med **13**, e14167.

Geyer, Philipp E, Holdt, L.M., Teupser, D., Mann, M., 2017. Revisiting biomarker discovery by plasma proteomics. Mol Syst Biol **13**.

Geyer, P.E., Kulak, N.A., Pichler, G., Holdt, L.M., Teupser, D., Mann, M., 2016. Plasma Proteome Profiling to Assess Human Health and Disease. Cells **2**, 185–195.

Giansanti, P., Tsiatsiani, L., Low, T.Y., Heck, A.J.R., 2016. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. Nat Protoc **11**, 993–1006.

Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., Aebersold, R., 2012. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics **11**, O111.016717.

González, M.A., Barrera-Chacón, R., Peña, F.J., Belinchón-Lorenzo, S., Robles, N.R., Pérez-Merino, E.M., Martín-Cano, F.E., Duque, F.J., 2023. Proteomic research on new urinary biomarkers of renal disease in canine leishmaniosis: Survival and monitoring response to treatment. Res Vet Sci **161**, 180–190.

Gordon Hadley, G., Larson, N.L., 1953. Use of Sequestrene as an Anticoagulant. Am. J. Clin. Pathol. **23**, 613–618.

Govender, I.S., 2019. The discovery of kidney injury-related protein biomarkers associated with first-line antiretroviral treatment in South Africa using microflow SWATH mass spectrometry (Thesis).

Graves, P.R., Haystead, T.A.J., 2002. Molecular Biologist's Guide to Proteomics. Microbiol Mol Biol Rev **66**, 39–63.

Gregorich, Z.R., Chang, Y.-H., Ge, Y., 2014. Proteomics in heart failure: top-down or bottom-up? Pflugers Arch - Eur J Physiol **466**, 1199–1209.

Guan, H., Peng, R., Mao, L., Fang, F., Xu, B., Chen, M., 2020. Injured tubular epithelial cells activate fibroblasts to promote kidney fibrosis through miR-150-containing exosomes. Exp. Cell Res. **392**, 112007.

Gundry, R.L., White, M.Y., Murray, C.I., Kane, L.A., Fu, Q., Stanley, B.A., Van Eyk, J.E., 2009. Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. Curr Protoc Mol Biol CHAPTER, Unit **10**.25.

Gupta, A., Kumar, A., 2013. Pros and Cons of the Proteomics. Biomedical **37**, 163.

Gupta, G.R., Parkhurst, J.O., Ogden, J.A., Aggleton, P., Mahal, A., 2008. Structural approaches to HIV prevention. The Lancet **372**, 764–775.

Gürtler, L.G., Hauser, P.H., Eberle, J., von Brunn, A., Knapp, S., Zekeng, L., Tsague, J.M., Kaptue, L., 1994. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. J Virol **68**, 1581–1585.

Ha, B.H., Morse, E.M., Turk, B.E., Boggon, T.J., 2015. Signaling, Regulation, and Specificity of the Type II p21-activated Kinases *. J. Biol. Chem **290**, 12975–12983.

Hall, A.M., Hendry, B.M., Nitsch, D., Connolly, J.O., 2011. Tenofovir-associated kidney toxicity in HIV-infected patients: a review of the evidence. Am J Kidney Dis **57**, 773–780.

Hamlin, A.N., Tillotson, J., Bumpus, N.N., 2019. Genetic variation of kinases and activation of nucleotide analog reverse transcriptase inhibitor tenofovir. Pharmacogenomics **20**, 105–111.

Han, X., Aslanian, A., Yates, J.R., 2008. Mass Spectrometry for Proteomics. Curr Opin Chem Biol **12**, 483–490.

Harwood, R., Bridge, J., Ressel, L., Scarfe, L., Sharkey, J., Czanner, G., Kalra, P.A., Odudu, A., Kenny, S., Wilm, B., Murray, P., 2022. Murine models of renal ischemia reperfusion injury: An opportunity for refinement using noninvasive monitoring methods. Physiol Rep **10**, e15211.

Hassan, Z.A., Schattner, P., Mazza, D., 2006. Doing A Pilot Study: Why Is It Essential? Malays Fam Physician **1**, 70–73.

Havasi, A., Borkan, S.C., 2011. Apoptosis and acute kidney injury. Kidney Int **80**, 29–40.

Hemelaar, J., Gouws, E., Ghys, P.D., Osmanov, S., 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS **20**, W13.

Hermes, N., Jewell, K.S., Wick, A., Ternes, T.A., 2018. Quantification of more than 150 micropollutants including transformation products in aqueous samples by liquid chromatography-tandem mass spectrometry using scheduled multiple reaction monitoring. J. Chromatogr. A **1531**, 64–73.

Higdon, R., 2013. Multiple Hypothesis Testing, in: Dubitzky, W., Wolkenhauer, O., Cho, K.-H., Yokota, H. (Eds.), Encyclopedia of Systems Biology. Springer, New York, NY, pp. 1468–1469.

Hogg, R.S., Yip, B., Kully, C., Craib, K.J., O'Shaughnessy, M.V., Schechter, M.T., Montaner, J.S., 1999. Improved survival among HIV-infected patients after initiation of triple-drug antiretroviral regimens. CMAJ **160**, 659–665.

Hongzhan, H., Shukla, H.D., Cathy, W., Satya, S., 2007. Challenges and Solutions in Proteomics. Curr Genomics **8**, 21–28.

Hu, H., Batteux, F., Chéreau, C., Kavian, N., Marut, W., Gobeaux, C., Borderie, D., Dinh-Xuan, A.T., Weill, B., Nicco, C., 2011. Clopidogrel protects from cell apoptosis and oxidative damage in a mouse model of renal ischaemia–reperfusion injury. J. Pathol. **225**, 265–275.

Hu, W.-S., Hughes, S.H., 2012. HIV-1 Reverse Transcription. Cold Spring Harb Perspect Med **2**, a006882.

Huang, T., Wang, J., Yu, W., He, Z., 2012. Protein inference: a review. Brief Bioinform **13**, 586–614.

Hutchison, C.A., Batuman, V., Behrens, J., Bridoux, F., Sirac, C., Dispenzieri, A., Herrera, G.A., Lachmann, H., Sanders, P.W., 2011. The pathogenesis and diagnosis of acute kidney injury in multiple myeloma. Nat Rev Nephrol **8**, 43–51.

Ignjatovic, V., Geyer, P.E., Palaniappan, K.K., Chaaban, J.E., Omenn, G.S., Baker, M.S., Deutsch, E.W., Schwenk, J.M., 2019. Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data. J Proteome Res **18**, 4085–4097.

Ihtiyar, E., Yaşar, N.F., Erkasap, N., Köken, T., Tosun, M., Oner, S., Erkasap, S., 2011. Effects of doxycycline on renal ischemia reperfusion injury induced by abdominal compartment syndrome. J Surg Res **167**, 113–120.

Ivanov, M.V., Solovyeva, E.M., Bubis, J.A., Gorshkov, M.V., 2021. Improving the Protein Inference from Bottom-Up Proteomic Data Using Identifications from MS1 Spectra. J. Am. Soc. Mass Spectrom. **32**, 1258–1262.

Ivers, L.C., Kendrick, D., Doucette, K., 2005. Efficacy of Antiretroviral Therapy Programs in Resource-Poor Settings: A Meta-analysis of the Published Literature. Clin. Infect. Dis. **41**, 217–224.

Ja, Levy, Ad, H., Sm, K., Ja, Landis, Jm, S., Ls, O., 1984. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. Science (New York, N.Y.) **225**.

Jacobo-Albavera, L., Domínguez-Pérez, M., Medina-Leyte, D.J., González-Garrido, A., Villarreal-Molina, T., 2021. The Role of the ATP-Binding Cassette A1 (ABCA1) in Human Disease. Int. J. Mol. Sci. **22**, 1593.

Jafari, A., Khalili, H., Dashti-Khavidaki, S., 2014. Tenofovir-induced nephrotoxicity: incidence, mechanism, risk factors, prognosis and proposed agents for prevention. Eur J Clin Pharmacol **70**, 1029–1040.

Jana, S., Mitra, P., Roy, S., 2023. Proficient Novel Biomarkers Guide Early Detection of Acute Kidney Injury: A Review. Diseases **11**, 8.

Jansen, M.P.B., Emal, D., Teske, G.J.D., Dessing, M.C., Florquin, S., Roelofs, J.J.T.H., 2017. Release of extracellular DNA influences renal ischemia reperfusion injury by platelet activation and formation of neutrophil extracellular traps. Kidney Int. **91**, 352–364.

Jansen, M.P.B., Florquin, S., Roelofs, J.J.T.H., 2018. The role of platelets in acute kidney injury. Nat. Rev. Nephrol. **14**, 457–457.

Jiang, R., Shi, Z., Johnson, J.J., Liu, Y., Stack, M.S., 2011. Kallikrein-5 Promotes Cleavage of Desmoglein-1 and Loss of Cell-Cell Cohesion in Oral Squamous Cell Carcinoma*. J. Biol. Chem. **286**, 9127–9135.

Jiao, Y., Du, P., 2016. Performance measures in evaluating machine learning based bioinformatics predictors for classifications. Quant Biol **4**, 320–330.

Karayel, O., Virreira Winter, S., Padmanabhan, S., Kuras, Y.I., Vu, D.T., Tuncali, I., Merchant, K., Wills, A.-M., Scherzer, C.R., Mann, M., 2022. Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease. Cell Rep Med **3**, 100661.

Karpievitch, Y.V., Polpitiya, A.D., Anderson, G.A., Smith, R.D., Dabney, A.R., 2010. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. Ann Appl Stat **4**, 1797–1823.

Kasembeli, A.N., Duarte, R., Ramsay, M., Naicker, S., 2015. African origins and chronic kidney disease susceptibility in the human immunodeficiency virus era. World J Nephrol **4**, 295–306.

Kashani, K., Cheungpasitporn, W., Ronco, C., 2017. Biomarkers of acute kidney injury: the pathway from discovery to clinical adoption. Clin. Chem. Lab. Med. (CCLM) **55**, 1074–1089.

Kaur, R., Liu, X., Gjoerup, O., Zhang, A., Yuan, X., Balk, S.P., Schneider, M.C., Lu, M.L., 2005. Activation of p21-activated Kinase 6 by MAP Kinase Kinase 6 and p38 MAP Kinase *. J. Biol. Chem. **280**, 3323–3330.

Kavallaris, M., Marshall, G.M., 2005. Proteomics and disease: opportunities and challenges. Med. J. Aust. 182.

Kellum, J.A., Lameire, N., Aspelin, P., Barsoum, R.S., Burdmann, E.A., Goldstein, S.L., Herzog, C.A., Joannidis, M., Kribben, A., Levey, A.S., MacLeod, A.M., Mehta, R.L., Murray, P.T., Naicker, S., Opal, S.M., Schaefer, F., Schetz, M., Uchino, S., 2012. Kidney disease: Improving global outcomes (KDIGO) acute kidney injury work group. KDIGO clinical practice guideline for acute kidney injury. Kidney Int. Suppl. **2**, 1–138.

Kohler, J.J., Hosseini, S.H., Hoying-Brandt, A., Green, E., Johnson, D.M., Russ, R., Tran, D., Raper, C.M., Santoianni, R., Lewis, W., 2009. Tenofovir renal toxicity targets mitochondria of renal proximal tubules. Lab Invest **89**, 513–519.

Kolatsi-Joannou, M., Price, K.L., Winyard, P.J., Long, D.A., 2011. Modified Citrus Pectin Reduces Galectin-3 Expression and Disease Severity in Experimental Acute Kidney Injury. PLOS ONE **6**, e18683.

Kops, G.J.P.L., Kim, Y., Weaver, B.A.A., Mao, Y., McLeod, I., Yates, J.R., Tagaya, M., Cleveland, D.W., 2005. ZW10 links mitotic checkpoint signaling to the structural kinetochore. J Cell Biol **169**, 49–60.

Krakower, D.S., Jain, S., Mayer, K.H., 2015. Antiretrovirals for Primary HIV Prevention: the Current Status of Pre- and Post-exposure Prophylaxis. Curr HIV/AIDS Rep **12**, 127–138.

Kuiken, C., Leitner, T., Foley, B., Hahn, B., Marx, P., McCutchan, F., Foundation, H.M.J., Wolinsky, S., Korber, B., Bansal, G., Abfalterer, W., Athreya, G., Fischer, W., Funkhouser, B., Gaschen, B., Hraber, P., Lo, C.-C., Macke, J., Szinger, J.J., Thurmond, J., Yoon, H., Zhang, M., 2008. HIV Seq. Compend 2008.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., McDermott, M.G., Monteiro, C.D., Gundersen, G.W., Ma'ayan, A., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res **44**, W90–W97.

Kuritzkes, D.R., 2009. HIV-1 Entry Inhbitors: An Overview. Curr Opin HIV AIDS **4**, 82–87.

Kuwabara, I., Kuwabara, Y., Yang, R.-Y., Schuler, M., Green, D.R., Zuraw, B.L., Hsu, D.K., Liu, F.-T., 2002. Galectin-7 (PIG1) exhibits pro-apoptotic function through JNK activation and mitochondrial cytochrome c release. J Biol Chem **277**, 3487–3497.

Laskay, Ü.A., Lobas, A.A., Srzentić, K., Gorshkov, M.V., Tsybin, Y.O., 2013. Proteome digestion specificity analysis for rational design of extended bottom-up and middle-down proteomics experiments. J Proteome Res **12**, 5558–5569.

Lee, C.G., Hartl, D., Lee, G.R., Koller, B., Matsuura, H., Da Silva, C.A., Sohn, M.H., Cohn, L., Homer, R.J., Kozhich, A.A., Humbles, A., Kearley, J., Coyle, A., Chupp, G., Reed, J., Flavell, R.A., Elias, J.A., 2009. Role of breast regression protein 39 (BRP-39)/chitinase 3-like-1 in Th2 and IL-13–induced tissue responses and apoptosis. J. Exp. Med. **206**, 1149–1166.

Lee, D.B.N., Huang, E., Ward, H.J., 2006. Tight junction biology and kidney dysfunction. Am J Physiol Renal Physiol **290**, F20–F34.

Lin, P.-H., Duann, P., 2020. Dyslipidemia in Kidney Disorders: Perspectives on Mitochondria Homeostasis and Therapeutic Opportunities. Front. physiol. **11**.

Liu, Z., Li, L., Zhang, H., Pang, X., Qiu, Z., Xiang, Q., Cui, Y., 2023. Platelet factor 4(PF4) and its multiple roles in diseases. Blood Rev. 101155.

Löb, S., Vattai, A., Kuhn, C., Mittelberger, J., Herbert, S.-L., Wöckel, A., Schmoeckel, E., Mahner, S., Jeschke, U., 2022. The Pregnancy Zone Protein (PZP) is significantly downregulated in the placenta of preeclampsia and HELLP syndrome patients. J. Reprod. Immunol. **153**, 103663.

Lund-Katz, S., Phillips, M.C., 2010. High Density Lipoprotein Structure–Function and Role in Reverse Cholesterol Transport. Subcell Biochem **51**, 183–227.

Luque-Garcia, J.L., Neubert, T.A., 2007. Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry. J Chromatogr A **1153**, 259–276.

Maas, S.L., Soehnlein, O., Viola, J.R., 2018. Organ-Specific Mechanisms of Transendothelial Neutrophil Migration in the Lung, Liver, Kidney, and Aorta. Front Immunol **9**, 2739.

Machado, M.N., Nakazone, M.A., Maia, L.N., 2014. Acute kidney injury based on KDIGO (Kidney Disease Improving Global Outcomes) criteria in patients with elevated baseline serum creatinine undergoing cardiac surgery. Rev Bras Cir Cardiovasc **29**, 299–307.

Mallipattu, S.K., 2023. The loss of profilin1 is catastrophic to podocytes. J Clin Invest **133**.

Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A., Doulamis, N., 2021. Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. Technologies **9**, 81.

Martínez-Klimova, E., Aparicio-Trejo, O.E., Tapia, E., Pedraza-Chaverri, J., 2019. Unilateral Ureteral Obstruction as a Model to Investigate Fibrosis-Attenuating Treatments. Biomolecules **9**, 141.

Matsumoto, T., Urushido, M., Ide, H., Ishihara, M., Hamada-Ode, K., Shimamura, Y., Ogata, K., Inoue, K., Taniguchi, Y., Taguchi, T., Horino, T., Fujimoto, S., Terada, Y., 2015. Small Heat Shock Protein Beta-1 (HSPB1) Is Upregulated and Regulates Autophagy and Apoptosis of Renal Tubular Cells in Acute Kidney Injury. PLOS ONE **10**, e0126229.

Mauclère, P., Loussert-Ajaka, I., Damond, F., Fagot, P., Souquières, S., Monny Lobe, M., Mbopi Keou, F.X., Barré-Sinoussi, F., Saragosti, S., Brun-Vézinet, F., Simon, F., 1997. Serological and virological characterization of HIV-1 group O infection in Cameroon. AIDS **11**, 445–453.

Mayeux, R., 2004. Biomarkers: Potential Uses and Limitations. NeuroRx **1**, 182–188.

McHeyzer-Williams, M., Okitsu, S., Wang, N., McHeyzer-Williams, L., 2012. Molecular programming of B cell memory. Nat Rev Immunol **12**, 24–34.

McIntyre, J., Gray, G., 2002. What can we do to reduce mother to child transmission of HIV? Br. Med. J **324**, 218.

McSweeney, K.R., Gadanec, L.K., Qaradakhi, T., Ali, B.A., Zulli, A., Apostolopoulos, V., 2021. Mechanisms of Cisplatin-Induced Acute Kidney Injury: Pathological Mechanisms, Pharmacological Interventions, and Genetic Mitigations. Cancers (Basel) **13**, 1572.

Meintjes, G., Moorhouse, M.A., Carmona, S., Davies, N., Dlamini, S., van Vuuren, C., Manzini, T., Mathe, M., Moosa, Y., Nash, J., Nel, J., Pakade, Y., Woods, J., Van Zyl, G., Conradie, F., Venter, F., 2017a. Adult antiretroviral therapy guidelines 2017. South Afr J HIV Med **18**, 776.

Meyer-Rath, G., Johnson, L.F., Pillay, Y., Blecher, M., Brennan, A.T., Long, L., Moultrie, H., Sanne, I., Fox, M.P., Rosen, S., 2017. Changing the South African national antiretroviral therapy guidelines: The role of cost modelling. PLoS One **12**.

Mills, E.J., Nachega, J.B., Bangsberg, D.R., Singh, S., Rachlis, B., Wu, P., Wilson, K., Buchan, I., Gill, C.J., Cooper, C., 2006. Adherence to HAART: A Systematic Review of Developed and Developing Nation Patient-Reported Barriers and Facilitators. PLOS Medicine **3**, e438.

Mizdrak, M., Kumrić, M., Kurir, T.T., Božić, J., 2022. Emerging Biomarkers for Early Detection of Chronic Kidney Disease. J Pers Med **12**, 548.

Moorhouse, M., Maartens, G., Venter, W.D.F., Moosa, M.-Y., Steegen, K., Jamaloodien, K., Fox, M.P., Conradie, F., 2019. Third-Line Antiretroviral Therapy Program in the South African Public Sector: Cohort Description and Virological Outcomes. J Acquir Immune Defic Syndr **80**, 73–78.

Moosa, A., Gengiah, T.N., Lewis, L., Naidoo, K., 2019. Long-term adherence to antiretroviral therapy in a South African adult patient cohort: a retrospective study. BMC Infect. Dis. **19**, 775.

Motomura, K., Chen, J., Hu, W.-S., 2008. Genetic Recombination between Human Immunodeficiency Virus Type 1 (HIV-1) and HIV-2, Two Distinct Human Lentiviruses. J. Virol. **82**, 1923–1933.

Mtisi, T.J., Ndhlovu, C.E., Maponga, C.C., Morse, G.D., 2019a. Tenofovir-associated kidney disease in Africans: a systematic review. AIDS Res Ther **16**.

Müller, T., Winter, D., 2017. Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents. Mol Cell Proteomics **16**, 1173–1187.

National HIV Curriculum, 2023. Adverse Effects of Antiretroviral Medications - Core Concepts.

Nature Methods Editorial, 2023. Single-cell proteomics: challenges and prospects. Nat. Methods. **20**, 317–318.

Nayak Rao, S., 2016. The role of heat shock proteins in kidney disease. J Transl Int Med **4**, 114–117.

Ng, S., Masarone, S., Watson, D., Barnes, M.R., 2023. The benefits and pitfalls of machine learning for biomarker discovery. Cell Tissue Res **394**, 17–31.

Novak, K.B., Le, H.D., Christison-Lagay, E.R., Nose, V., Doiron, R.J., Moses, M.A., Puder, M., 2010. Effects of metalloproteinase inhibition in a murine model of renal ischemia-reperfusion injury. Pediatr Res **67**, 257–262.

Ogata, F., Hirasawa, Y., Takahashi, S., Tajiri, M., Kanno, K., Deguchi, T., Nakamura, T., 1981. Disturbance of Serum Lipid Metabolism in Acute Renal Failure. Clin Exp Dial Apheresis **5**, 361–371.

Ogobuiro, I., Tuma, F., 2023. Physiology, Renal, in: StatPearls [Internet]. StatPearls Publishing.

Oh, Y., 2012. The insulin-like growth factor system in chronic kidney disease: Pathophysiology and therapeutic opportunities. Kidney Res Clin Pract. **31**, 26–37.

Olivier, E., Soury, E., Ruminy, P., Husson, A., Parmentier, F., Daveau, M., Salier, J.P., 2000. Fetuin-B, a second member of the fetuin family in mammals. Biochem J **350**, 589–597.

O'Neill, S., Harrison, E.M., Ross, J.A., Wigmore, S.J., Hughes, J., 2014. Heat-Shock Proteins and Acute Ischaemic Kidney Injury. Nephron Exp Nephrol **126**, 167–174.

Oram, J.F., 2003. HDL Apolipoproteins and ABCA1. Arterioscler Thromb Vasc Biol **23**, 720–727.

Orchard, S., Charles, P., 2022. Proteomics - An introduction to EMBL-EBI resources. URL: https://www.ebi.ac.uk/training/online/courses/proteomics-an-introduction (accessed 10.11.23).

O'Rourke, M.B., Januszewski, A.S., Sullivan, D.R., Lengyel, I., Stewart, A.J., Arya, S., Ma, R.C., Galande, S., Hardikar, A.A., Joglekar, M.V., Keech, A.C., Jenkins, A.J., Molloy, M.P., 2023. Optimised plasma sample preparation and LC-MS analysis to support large-scale proteomic analysis of clinical trial specimens: Application to the Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) trial. Proteom. - Clin. Appl. **17**, 2200106.

Orozco-Arias, S., Piña, J.S., Tabares-Soto, R., Castillo-Ossa, L.F., Guyot, R., Isaza, G., 2020. Measuring Performance Metrics of Machine Learning Algorithms for Detecting and Classifying Transposable Elements. Processes **8**, 638.

Ou, S.-M., Tsai, M.-T., Chen, H.-Y., Li, F.-A., Lee, K.-H., Tseng, W.-C., Chang, F.-P., Lin, Y.-P., Yang, R.-B., Tarng, D.-C., 2022. Urinary Galectin-3 as a Novel Biomarker for the Prediction of Renal Fibrosis and Kidney Disease Progression. Biomedicines **10**, 585.

Pantaleo, G., Fauci, A.S., 1996. Immunopathogenesis of HIV Infection. Annu. Rev. Microbiol. **50**, 825–854.

Park, J., Kim, H., Kim, S.Y., Kim, Y., Lee, J.-S., Dan, K., Seong, M.-W., Han, D., 2020. In-depth blood proteome profiling analysis revealed distinct functional characteristics of plasma proteins between severe and non-severe COVID-19 patients. Sci Rep **10**, 22418.

Parker, C.E., Warren, M.R., Mocanu, V., 2010. Mass Spectrometry for Proteomics, in: Alzate, O. (Ed.), Neuroproteomics, Frontiers in Neuroscience. CRC Press/Taylor & Francis, Boca Raton (FL).

Parmar, M.S., 2022. Light Chain-Associated Renal Disorders: Practice Essentials, Pathophysiology, Etiology.

Patterson, S.D., Aebersold, R.H., 2003. Proteomics: the first decade and beyond. Nat Genet **33**, 311–323.

Payrastre, B., Missy, K., Trumel, C., Bodin, S., Plantavid, M., Chap, H., 2000. The integrin αIIb/β3 in human platelet signal transduction. Biochem. Pharmacol. **60**, 1069–1074.

Peeters, M., Honoré, C., Huet, T., Bedjabaga, L., Ossari, S., Bussi, P., Cooper, R.W., Delaporte, E., 1989. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon: AIDS **3**, 625–630.

Peng, X., Nelson, E.S., Maiers, J.L., DeMali, K.A., 2011. New Insights into Vinculin Function and Regulation. Int Rev Cell Mol Biol **287**, 191–231.

Pike, N., 2011. Using false discovery rates for multiple comparisons in ecology and evolution. Methods Ecol. Evol. **2**, 278–282.

Pino, L.K., 2019. Methods for harmonizing and calibrating quantitative mass spectrometry experiments (Thesis).

Plantier, J.-C., Leoz, M., Dickerson, J.E., De Oliveira, F., Cordonnier, F., Lemée, V., Damond, F., Robertson, D.L., Simon, F., 2009. A new human immunodeficiency virus derived from gorillas. Nat Med **15**, 871–872.

Popovic, M., Sarngadharan, M.G., Read, E., Gallo, R.C., 1984. Detection, Isolation, and Continuous Production of Cytopathic Retroviruses (HTLV-III) from Patients with AIDS and Pre-AIDS. Science **224**, 497–500.

Powell, D.R., Liu, F., Baker, B.K., Hintz, R.L., Durham, S.K., Brewer, E.D., Frane, J.W., Tonshoff, B., Mehls, O., Wingen, A.-M., Watkins, S.L., Hogg, R.J., Lee, P.D.K., 1997. Insulin-Like Growth Factor-Binding Protein-6 Levels Are Elevated in Serum of Children with Chronic Renal Failure: A Report of the Southwest Pediatric Nephrology Study Group*. J Clin Endocrinol Metab. **82**, 2978–2984.

Qu, Y., Yao, Z., Xu, N., Shi, G., Su, J., Ye, S., Chang, K., Li, K., Wang, Y., Tan, S., Pei, X., Chen, Y., Qin, Z., Feng, J., Lv, J., Zhu, J., Ma, F., Tang, S., Xu, Wenhao, Tian, X., Anwaier, A., Tian, S., Xu, Wenbo, Wu, X., Zhu, S., Zhu, Y., Cao, D., Sun, M., Gan, H., Zhao, J., Zhang, H., Ye, D., Ding, C., 2023. Plasma proteomic profiling discovers molecular features associated with upper tract urothelial carcinoma. Cell Rep Med **4**, 101166.

Quinn, M.J., Byzova, T.V., Qin, J., Topol, E.J., Plow, E.F., 2003. Integrin αIIbβ3 and Its Antagonism. Arterioscler Thromb Vasc Biol **23**, 945–952.

Rees, K., Bisnauth, M., 2022. HIV treatment in South Africa: how to help people stay on ARVs when life gets in the way. The Conversation. URL: http://theconversation.com/hiv-treatment-in-south-africa-how-to-help-people-stay-on-arvs-when-life-gets-in-the-way-190684 (accessed 2.26.23).

Republic of South Africa National Department of Health, 2020. Updated 2019 ART Clinical Guidelines for the Management of HIV in Adults, Pregnancy, Adolescents, Children, Infants and Neonates.

Resyn Biosciences, E., South Africa, 2022. Preparation and HILIC SPE clean-up of SDS extracted plasma proteins for MS analysis - HILIC-PLASMA.

Riddell, J., IV, Amico, K.R., Mayer, K.H., 2018. HIV Preexposure Prophylaxis: A Review. JAMA **319**, 1261–1268.

Rischke, S., Hahnefeld, L., Burla, B., Behrens, F., Gurke, R., Garrett, T.J., 2023. Small molecule biomarker discovery: Proposed workflow for LC-MS-based clinical research projects. J. Mass Spectrom. Adv. Clin. Lab. **28**, 47–55.

Robins, T., Plattner, J., 1993. HIV protease inhibitors: their anti-HIV activity and potential role in treatment. J Acquir Immune Defic Syndr (1988) **6**, 162–170.

Rodger, A.J., Cambiano, V., Bruun, T., Vernazza, P., Collins, S., van Lunzen, J., Corbelli, G.M., Estrada, V., Geretti, A.M., Beloukas, A., Asboe, D., Viciana, P., Gutiérrez, F., Clotet, B., Pradier, C., Gerstoft, J., Weber, R., Westling, K., Wandeler, G., Prins, J.M., Rieger, A., Stoeckle, M., Kümmerle, T., Bini, T., Ammassari, A., Gilson, R., Krznaric, I., Ristola, M., Zangerle, R., Handberg, P., Antela, A., Allan, S., Phillips, A.N., Lundgren, J., for the PARTNER Study Group, 2016. Sexual Activity Without Condoms and Risk of HIV Transmission in Serodifferent Couples When the HIV-Positive Partner Is Using Suppressive Antiretroviral Therapy. JAMA **316**, 171–181.

Rondanino, C., Poland, P.A., Kinlough, C.L., Li, H., Rbaibi, Y., Myerburg, M.M., Al-bataineh, M.M., Kashlan, O.B., Pastor-Soler, N.M., Hallows, K.R., Weisz, O.A., Apodaca, G., Hughey, R.P., 2011. Galectin-7 modulates the length of the primary cilia and wound repair in polarized kidney epithelial cells. Am J Physiol Renal Physioly **301**, F622–F633.

Rouillard, A.D., Gundersen, G.W., Fernandez, N.F., Wang, Z., Monteiro, C.D., McDermott, M.G., Ma'ayan, A., 2016. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database 2016, **baw100**.

Sakamoto, S., Putalun, W., Vimolmangkang, S., Phoolcharoen, W., Shoyama, Y., Tanaka, H., Morimoto, S., 2018. Enzyme-linked immunosorbent assay for the quantitative/qualitative analysis of plant secondary metabolites. J Nat Med **72**, 32–42.

Sakhuja, A., 2023. Novel Biomarkers of Kidney Function: Introduction and Overview, Characteristics of an Ideal Marker for Kidney Disease, Biomarkers of Acute Kidney Injury.

Santiago, M.L., Range, F., Keele, B.F., Li, Y., Bailes, E., Bibollet-Ruche, F., Fruteau, C., Noë, R., Peeters, M., Brookfield, J.F.Y., Shaw, G.M., Sharp, P.M., Hahn, B.H., 2005. Simian Immunodeficiency Virus Infection in Free-Ranging Sooty Mangabeys (*Cercocebus atys atys*) from the Taï Forest, Côte d'Ivoire: Implications for the Origin of Epidemic Human Immunodeficiency Virus Type 2. J Virol **79**, 12515–12527.

Sanz, A.B., Sanchez-Niño, M.D., Ramos, A.M., Ortiz, A., 2023. Regulated cell death pathways in kidney disease. Nat Rev Nephrol **19**, 281–299.

Savage, B., Shattil, S.J., Ruggeri, Z.M., 1992. Modulation of platelet function through adhesion receptors. A dual role for glycoprotein IIb-IIIa (integrin alpha IIb beta 3) mediated by fibrinogen and glycoprotein Ib-von Willebrand factor. J. Biol. Chem. **267**, 11300–11306.

Schmid, M., Dalela, D., Tahbaz, R., Langetepe, J., Randazzo, M., Dahlem, R., Fisch, M., Trinh, Q.-D., Chun, F.K.-H., 2015. Novel biomarkers of acute kidney injury: Evaluation and evidence in urologic surgery. World J Nephrol **4**, 160–168.

Schrantz, N., Correia, J. da S., Fowler, B., Ge, Q., Sun, Z., Bokoch, G.M., 2004. Mechanism of p21-activated Kinase 6-mediated Inhibition of Androgen Receptor Signaling *. J. Biol. Chem. **279**, 1922–1931.

Schubert, O.T., Röst, H.L., Collins, B.C., Rosenberger, G., Aebersold, R., 2017. Quantitative proteomics: challenges and opportunities in basic and applied research. Nat Protoc **12**, 1289–1294.

Sciex, T., Canada, 2017. How Is the SWATH® Acquisition Variable Window Calculator Excel Sheet Used? URL https://sciex.com/support/knowledge-base-articles/how-to-use-the-swath-variable-calculator-excel-sheet_en_us (accessed 3.18.24).

Seedat, F., Martinson, N., Motlhaoleng, K., Abraham, P., Mancama, D., Naicker, S., Variava, E., 2017. Acute Kidney Injury, Risk Factors, and Prognosis in Hospitalized HIV-Infected Adults in South Africa, Compared by Tenofovir Exposure. AIDS Res. Hum. **33**, 33–40.

Sewgobind, N.V., Albers, S., Pieters, R.J., 2021. Functions and Inhibition of Galectin-7, an Emerging Target in Cellular Pathophysiology. Biomolecules **11**, 1720.

Shama, A., Soni, T., Jawanda, I.K., Upadhyay, G., Sharma, A., Prabha, V., 2023. The Latest Developments in Using Proteomic Biomarkers from Urine and Serum for Non-Invasive Disease Diagnosis and Prognosis. Biomark Insights **18**, 11772719231190218.

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 Res **13**, 2498–2504.

Sharp, P.M., Hahn, B.H., 2011. Origins of HIV and the AIDS Pandemic. Cold Spring Harb Perspect Med **1**, a006841–a006841.

Sheng, L., Zhuang, S., 2020. New Insights Into the Role and Mechanism of Partial Epithelial-Mesenchymal Transition in Kidney Fibrosis. Front. physiol. **11**.

Shi, W., Wan, T.-T., Li, H.-H., Guo, S.-B., 2023. Blockage of S100A8/A9 ameliorates septic nephropathy in mice. Front. pharmacol. **14**.

Shi, X., Wang, J., Lei, Y., Cong, C., Tan, D., Zhou, X., 2019. Research progress on the PI3K/AKT signaling pathway in gynecological cancer (Review). Mol. Med. Rep. **19**, 4529–4535.

Siew, E.D., Ware, L.B., Ikizler, T.A., 2011. Biological Markers of Acute Kidney Injury. J Am Soc Nephrol **22**, 810.

Simon, F., Mauclère, P., Roques, P., Loussert-Ajaka, I., Müller-Trutwin, M.C., Saragosti, S., Georges-Courbot, M.C., Barré-Sinoussi, F., Brun-Vézinet, F., 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat Med **4**, 1032–1037.

Singbartl, K., Formeck, C.L., Kellum, J.A., 2019. Kidney-Immune System Crosstalk in AKI. Semin Nephrol **39**, 96–106.

Sinha, A., Mann, M., 2020. A beginner's guide to mass spectrometry–based proteomics. Biochem. **42**, 64–69.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., Klenk, D.C., 1985. Measurement of protein using bicinchoninic acid. Anal Biochem **150**, 76–85.

Sobsey, C.A., Ibrahim, S., Richard, V.R., Gaspar, V., Mitsa, G., Lacasse, V., Zahedi, R.P., Batist, G., Borchers, C.H., 2020. Targeted and Untargeted Proteomics Approaches in Biomarker Development. Proteomics **20**, 1900029.

Statistics South Africa, 2022. Statistical Release (P0302): Mid-year population estimates 2022.

Strazzella, A., Ossoli, A., Calabresi, L., 2021. High-Density Lipoproteins and the Kidney. Cells **10**, 764.

Strimbu, K., Tavel, J.A., 2010a. What are Biomarkers? Curr Opin HIV AIDS 5, 463–466.

Sundquist, W.I., Kräusslich, H.-G., 2012. HIV-1 Assembly, Budding, and Maturation. Cold Spring Harb Perspect Med **2**, a006924.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J., von Mering, C., 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res **43**, D447-452.

Tammaro, A., Florquin, S., Brok, M., Claessen, N., Butter, L.M., Teske, G.J.D., de Boer, O.J., Vogl, T., Leemans, J.C., Dessing, M.C., 2018. S100A8/A9 promotes parenchymal damage and renal fibrosis in obstructive nephropathy. Clin Exp Immunol **193**, 361–375.

Tan, R.J., Liu, Y., 2012. Matrix metalloproteinases in kidney homeostasis and diseases. Am J Physiol Renal Physiol **302**, F1351–F1361.

Tarwater, K., 2011. Estimated Glomerular Filtration Rate Explained. Mo Med 108, 29–32.

Taylor, B.S., Sobieszczyk, M.E., McCutchan, F.E., Hammer, S.M., 2008. The Challenge of HIV-1 Subtype Diversity. N Engl J Med **358**, 1590–1602.

Thomas, C., Ali, R., Park, I., Kim, H., Short, S., Kaunfer, S., Durai, L., Yilmam, O.A., Shenoy, T., Battinelli, E.M., Al-Samkari, H., Leaf, D.E., 2023. Platelet Factor 4 Antibodies and Severe AKI. Kidney360 **4**, 1672.

Tilton, J.C., Doms, R.W., 2010. Entry inhibitors in the treatment of HIV-1 infection. Antiviral Res **85**, 91–100.

To, W.S., Midwood, K.S., 2011. Plasma and cellular fibronectin: distinct and independent functions during tissue repair. Fibrogenesis Tissue Repair **4**, 21.

Torun, F.M., Virreira Winter, S., Doll, S., Riese, F.M., Vorobyev, A., Mueller-Reif, J.B., Geyer, P.E., Strauss, M.T., 2023. Transparent Exploration of Machine Learning for Biomarker Discovery from Proteomics and Omics Data. J. Proteome Res. **22**, 359–367.

Tun-Yhong, W., Chinpaisal, C., Pamonsinlapatham, P., Kaewkitichai, S., 2017. Tenofovir Disoproxil Fumarate Is a New Substrate of ATP-Binding Cassette Subfamily C Member 11. Antimicrob Agents Chemother **61**, e01725-16.

UNAIDS, 2021a. HIV and AIDS Estimates, South Africa (2021), Country Factsheets. URL: https://www.unaids.org/en/regionscountries/countries/southafrica (accessed 1.21.23).

UNAIDS, 2021b. 2025 AIDS Targets.

UNAIDS, 2021c. Joint United Nations Programme on HIV/AIDS: Data 2021, Geneva.

UNAIDS, 2020. World AIDS Report: Prevailing Against Pandemics.

UNAIDS, E.E., 2022. Global HIV & AIDS statistics — Fact sheet. URL: https://www.unaids.org/en/resources/fact-sheet (accessed 1.13.23).

Vaidya, V.S., Ferguson, M.A., Bonventre, J.V., 2008. Biomarkers of Acute Kidney Injury. Annu Rev Pharmacol Toxicol **48**, 463–493.

van Bömmel, F., Berg, T., 2008. Role of tenofovir in the treatment of chronic HBV infection. Future Virol. **3**, 207–220.

van Duijl, T.T., Ruhaak, L.R., de Fijter, J.W., Cobbaert, C.M., 2019. Kidney Injury Biomarkers in an Academic Hospital Setting: Where Are We Now? Clin Biochem Rev **40**, 79–97.

van Duijl, T.T., Soonawala, D., de Fijter, J.W., Ruhaak, L.R., Cobbaert, C.M., 2021. Rational selection of a biomarker panel targeting unmet clinical needs in kidney injury. Clin. Proteom. **18**, 10.

Vanmassenhove, J., Vanholder, R., Nagler, E., Van Biesen, W., 2013. Urinary and serum biomarkers for the diagnosis of acute kidney injury: an in-depth review of the literature*. Nephrol Dial Transplant **28**, 254–273.

Venter, Willem D.F., Fabian, J., Feldman, C., 2018. An overview of tenofovir and renal disease for the HIV-treating clinician. South Afr J HIV Med **19**, 817.

Visseaux, B., Damond, F., Matheron, S., Descamps, D., Charpentier, C., 2016. HIV-2 molecular epidemiology. Infect. Genet. Evol. **46**, 233–240.

Walker, J.M., 1996. The Bicinchoninic Acid (BCA) Assay for Protein Quantitation, in: Walker, J.M. (Ed.), The Protein Protocols Handbook. Humana Press, Totowa, NJ, pp. 11–14.

Wang, F., Eliaz, A., Chertow, G., Eliaz, I., Peng, Z., 2023. The potential roles of galectin-3 in AKI and CKD. Front. Physiol. **14**.

Wang, L., Niu, D., Wang, X., Khan, J., Shen, Q., Xue, Y., 2021. A Novel Machine Learning Strategy for the Prediction of Antihypertensive Peptides Derived from Food with High Efficiency. Foods **10**, 550.

Wang, S., Chi, K., Wu, D., Hong, Q., 2021. Insulin-Like Growth Factor Binding Proteins in Kidney Disease. Front Pharmacol **12**, 807119.

Wang, Siwen, Song, R., Wang, Z., Jing, Z., Wang, Shaoxiong, Ma, J., 2018. S100A8/A9 in Inflammation. Front Immunol **9**, 1298.

Wang, Y., Wang, Yang, Wan, R., Hu, C., Lu, Y., 2021. Profilin 1 Protein and Its Implications for Cancers. Oncology Vol 35, Issue **7**, 402–402.

Wassner, C., Bradley, N., Lee, Y., 2020. A Review and Clinical Understanding of Tenofovir: Tenofovir Disoproxil Fumarate versus Tenofovir Alafenamide. J Int Assoc Provid AIDS Care **19**, 2325958220919231.

Wearne, N., 2015. Morbidity and mortality of black HIV-positive patients with end-stage kidney disease receiving chronic haemodialysis in South Africa. SAMJ: S. Afr. Med. J. **105**, 105–106.

Wehr, T., 2006. Top-Down versus Bottom-Up Approaches in Proteomics. LCGC North America, LCGC North America-09-01-2006, **24**, 1004–1010.

Weinberg, J.L., Kovarik, C.L., 2010. The WHO Clinical Staging System for HIV/AIDS. AMA J. Ethics **12**, 202–206.

Weller, S., Davis, K., 2002. Condom effectiveness in reducing heterosexual HIV transmission. Cochrane Database Syst Rev CD003255.

Wilen, C.B., Tilton, J.C., Doms, R.W., 2012. HIV: Cell Binding and Entry. Cold Spring Harb Perspect Med **2**, a006866.

Wilkins, M.R., Sanchez, J.-C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F., Williams, K.L., 1996. Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How To Do It. Biotechnol Genet Eng **13**, 19–50.

Wiseman, J.D., Dawson, J.B., Slockbower, J.M., Hill, B.M., Calam, R.R., Coyne, E., Kozlowski, C.A., Rew, T.D., Bessman, J.D., National Committee for Clinical Laboratory Standard, 1996. Evacuated Tubes and Additives for Blood Specimen Collection: Approved Standard (1996). NCCLC.

Wishart, D.S., Bartok, B., Oler, E., Liang, K.Y.H., Budinski, Z., Berjanskii, M., Guo, A., Cao, X., Wilson, M., 2021. MarkerDB: an online database of molecular biomarkers. Nucleic Acids Res. **49**, D1259–D1267.

World Health Organization, 2021. Consolidated guidelines on HIV prevention, testing, treatment, service delivery and monitoring: recommendations for a public health approach. URL: https://www.who.int/publications-detail-redirect/9789240031593 (accessed 2.23.23).

World Health Organization, 2020. Consolidated HIV strategic information guidelines - Executive summary. URL: https://www.who.int/publications-detail-redirect/consolidated-hiv-strategic-information-guidelines-executive-summary (accessed 3.9.23).

World Health Organization, 2018. Updated recommendations on first-line and second-line antiretroviral regimens and post-exposure prophylaxis and recommendations on early infant diagnosis of HIV. URL: https://www.who.int/publications-detail-redirect/WHO-CDS-HIV-18.51 (accessed 2.24.23).

World Health Organization, 2017. Consolidated guidelines on person-centred HIV patient monitoring and case surveillance. World Health Organization, Geneva.

World Health Organization, H.F.S., 2022a. HIV. URL: https://www.who.int/news-room/fact-sheets/detail/hiv-aids (accessed 1.10.23).

World Health Organization, H.F.S., 2022b. Estimated number of people (all ages) living with HIV. URL: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/estimated-number-of-people--living-with-hiv (accessed 1.13.23).

Wu, J., Chen, Y., Gu, W., 2010. Urinary proteomics as a novel tool for biomarker discovery in kidney diseases. J Zhejiang Univ Sci B **11**, 227–237.

Xu, B., Li, W., Zhang, Y., Chen, Y., Feng, J., Song, X., 2023. Untargeted and spatial-resolved metabolomics characterize serum and tissue-specific metabolic reprogramming in acute kidney injury. Heliyon **9**, e21171.

Xu, X., Sun, B., 2015. Platelet granule secretion mechanisms: Are they modified in sepsis? Thromb. Res. **136**, 845–850.

Yang, R.-Y., Hsu, D.K., Yu, L., Ni, J., Liu, F.-T., 2001. Cell Cycle Regulation by Galectin-12, a New Member of the Galectin Superfamily *. J. Biol. Chem. **276**, 20252–20260.

Yang, X., Tu, W., Gao, X., Zhang, Q., Guan, J., Zhang, J., 2023. Functional regulation of syntaxin-1: An underlying mechanism mediating exocytosis in neuroendocrine cells. Front. Endocrinol. **14**.

Yao, Y., Wang, C., 2020. Dedifferentiation: inspiration for devising engineering strategies for regenerative medicine. npj Regen Med **5**, 1–11.

Young, T., Arens, F.J., Kennedy, G.E., Laurie, J.W., Rutherford, G.W., 2007. Antiretroviral post-exposure prophylaxis (PEP) for occupational HIV exposure. Cochrane Database Syst Rev.

Yu, B., Qiu, W., Chen, C., Ma, A., Jiang, J., Zhou, H., Ma, Q., 2020. SubMito-XGBoost: predicting protein submitochondrial localization by fusing multiple feature information and eXtreme gradient boosting. Bioinformatics **36**, 1074–1081.

Yvan-Charvet, L., Wang, N., Tall, A.R., 2010. Role of HDL, ABCA1, and ABCG1 Transporters in Cholesterol Efflux and Immune Responses. ATVB **30**, 139–143.

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. J. Biomark. 2014, e492838.

Zhang, M., Siedow, M., Saia, G., Chakravarti, A., 2010. Inhibition of p21-activated Kinase 6 (PAK6) Increases Radiosensitivity of Prostate Cancer Cells. Prostate **70**, 807–816.

Zhang, W.R., Parikh, C.R., 2019. Biomarkers of Acute and Chronic Kidney Disease. Annu Rev Physiol **81**, 309–333.

Zhao, T., Su, Z., Li, Y., Zhang, X., You, Q., 2020. Chitinase-3 like-protein-1 function and its role in diseases. Sig Transduct Target Ther **5**, 1–20.

Zheng, J., Lu, Z., Kocabas, F., Böttcher, R.T., Costell, M., Kang, X., Liu, X., DeBerardinis, R.J., Wang, Q., Chen, G.-Q., Sadek, H., Zhang, C.C., 2014. Profilin 1 is essential for retention and metabolism of mouse hematopoietic stem cells in bone marrow. Blood **123**, 992–1001.

Zhou, H., Pisitkun, T., Aponte, A., Yuen, P.S.T., Hoffert, J.D., Yasuda, H., Hu, X., Chawla, L., Shen, R.-F., Knepper, M.A., Star., R.A., 2006. Exosomal Fetuin-A identified by proteomics: a novel urinary biomarker for detecting acute kidney injury. Kidney Int **70**, 1847–1857.

Zhu, Q., Yamakuchi, M., Ture, S., de la Luz Garcia-Hernandez, M., Ko, K.A., Modjeski, K.L., LoMonaco, M.B., Johnson, A.D., O'Donnell, C.J., Takai, Y., Morrell, C.N., Lowenstein, C.J., 2014. Syntaxin-binding protein STXBP5 inhibits endothelial exocytosis and promotes platelet secretion. J Clin Invest **124**, 4503–4516.

Zou, C., Wang, C., Lu, L., 2022. Advances in the study of subclinical AKI biomarkers. Front. physiol. **13**.

Zuk, A., Bonventre, J.V., 2019. Recent advances in acute kidney injury and its consequences and impact on chronic kidney disease. Curr Opin Nephrol Hypertens **28**, 397–405.