

https://dspace.org/

The discovery of kidney injury-related protein biomarkers associated with first-line anteretroviral treatment in South Africa using microflow SWATH mass spectrometryre

Govender, Ireshyn Selvan

2019
https://hdl.handle.net/10539/27922

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

The Discovery of Kidney Injury-Related Protein Biomarkers Associated with First-Line Antiretroviral Treatment in South Africa using Microflow SWATH® Mass Spectrometry

by

Ireshyn Selvan Govender

Supervisors:

Dr. Stoyan Stoychev

Dr. Dalu Mancama

Prof. Demetra Mavri-Damelin



A thesis submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg in fulfilment of the requirements for the degree of Doctor of Philosophy

January 2019

Declaration

I, Ireshyn Selvan Govender (549848), am a student registered for the degree of Doctor of Philosophy in the academic year 2018.

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 work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at this or any other University.
- The information used in the Thesis has been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- 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.

This study complies with the declaration of Helsinki as all patients provided informed oral and written consent. The study was approved by local ethical committees, reference: #58/2013 (CSIR-REC) and #120612 (WITS-HREC).

The research described in this study was carried out at the Council for Scientific and Industrial Research, Biosciences, Pretoria, South Africa, under the supervision of Dr. Stoyan Stoychev and Dr. Dalu Mancama.



Ireshyn S. Govender 30th day of January 2019

Abstract

Background: South African patients account for ~10% of the global HIV/AIDS burden. In response, the country has implemented the largest antiretroviral therapy (ART) program in the world. However, not all patients respond positively, and it has been observed that up to 8-10% of patients on first-line ART experience acute kidney injury (AKI). Current tests for AKI are unreliable and only reflect positive following significant kidney damage. Hence there is a dire need to identify biomarkers for early detection of AKI. The urinary proteome provides a large and diverse source of information pertaining to a patient's health status. To date, there is no - published - standard operating protocol for reproducible collection and processing of urine for mass spectrometry-based clinical proteomics.

Methods: Extensive workflow development for urinary proteome analysis was done including multiple protein extraction, liquid chromatography and mass spectrometry acquisition methods. Acetone precipitation, of urinary protein, followed by FASP was compared to two novel sample preparation techniques using MagReSyn® HILIC and Norgen Biotek Corporation Urine Kits. The kit-based method was applied to a clinical cohort of 74 HIV positive patients. Their urinary proteomes were compared with the aim of detecting markers associated with AKI due to first-line ART using SWATH-MS.

Results and Discussion: Using MagReSyn® HILIC was the easiest, fastest and cheapest method whilst generating the most comprehensive data. The kit-based method was the most easily implemented at the point-of-care clinic and generated data that was comparable to the MagReSyn® HILIC method. Using the kit-based method, this study identified many proteins ($\geq$ 4-fold change, q value $\leq$ 0.01, $\geq$ 2 unique peptides), previously reported in literature, that have an association with kidney dysfunction. Furthermore, novel, putative markers for AKI have been identified including mimecan which has not been reported as an AKI marker before. Additionally, proteins showing significant changes in urinary abundance allude to the process of renal hepatisation since many of these proteins are usually associated with the liver. Some of these proteins have not been reported in relation to AKI and have not been reported in any South African populations before.

Conclusion: The methods developed in this study can serve as the gold-standard for future urinary proteomics experiments. Additionally, following assessment in a blinded verification cohort, previously reported and novel proteins have the potential to be used as early markers for AKI.

Acknowledgements

Many individuals and organizations have contributed significantly to the completion of this thesis and it is my sincere hope that their efforts in making this study a success have not been in vain.

Firstly, I owe my deepest gratitude to Dr. Stoyan Stoychev for being more than just a supervisor. His mentorship, guidance, understanding and tolerance allowed for the completion of this thesis whilst maintaining my sanity. I wish to thank him for sharing his knowledge with me and in turn deepening my scientific grounding. For supporting my ideas and allowing me to make mistakes and learn at my own pace. I appreciate everything that he has taught me professionally and personally that has made me a better scientist and a better person.

Thank you to Dr. Dalu Mancama, for affording me the opportunity to work on this project and develop relationships with a team that extends beyond CSIR, for always providing sound scientific advice and for imparting some of his vast knowledge onto me.

Thank you to Prof. Demetra Mavri-Damelin, for agreeing to take on the role of supervisor and for trusting that the project would proceed according to plan. I appreciate all the extra efforts that she had to put in towards the end of this project.

Thank you to my office and lab colleagues, Dr. Previn Naicker and Dr. Sindisiwe Buthelezi, for the friendship and support, for sharing the burdens and heartaches of rejected proposals and celebrating the successes of accepted abstracts.

I owe my thanks to the CSIR, for affording me the opportunity to integrate into the community of CSIR and for broadening my scientific horizons through the many collaborations and partners that I have interacted with.

To the PHRU, this study would not have been possible without the work done on site. I wish to thank the team for allowing me to implement my methods and workflows at the clinic, and for patient recruitment and management. Additionally, I am grateful to all the patients that agreed to participate in the study.

To my parents, family and friends – thank you for always believing in me and for the constant support and encouragement throughout the years.

Finally, to my partner - Renay - my moral compass, biggest critic and biggest fan. This thesis is as much hers as it is mine. Her unwavering support, understanding and encouragement, especially when I felt like I wanted to give up over the past few months, is greatly appreciated. The numerous setbacks I faced throughout this study were mere hiccups thanks to the motivation that she gave me to ‘know better and do better’. Her assistance with decoding patient samples is a highlight of my lab work thus far, thanks to the ‘glove-socks’. Above all, I appreciate the constant love and happiness that she has showered me with over the years.

List of Abbreviations

1D RP	One-dimensional reverse phase
2D RPRP	Two-dimensional reverse phase-reverse phase
2DGE	Two-dimensional gel electrophoresis
ABC	ATP-binding cassette
ABC-HSR	Abacavir-hypersensitivity reaction
ACN	Acetonitrile
AIDS	Acquired immune deficiency syndrome
AKI	Acute kidney injury
AmBic	Ammonium bicarbonate
ART	Antiretroviral therapy
ARV	Antiretroviral
AUC	Area under the curve
BCA	Bicinchoninic acid assay
BUN	Blood urea nitrogen
C18	Octadecyl bonded stationary phase
CE	Collision energy
CFRs	Circulating recombinant forms
CHAPS	3-[(3 cholamidopropyl) dimethylammonio]-1-propanesulfonate
CID	Collision-induced dissociation
CKD	Chronic kidney disease
CV	Co-efficient of variation
CW	Constant windows
Da	Dalton
DC	Direct current
DDA	Data-dependent acquisition
DIA	Data independent analysis
DN	Diabetic nephropathy
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded deoxyribonucleic acid

DTT	Dithiothreitol
EB	Elution buffer
eGFR	Estimated glomerular filtration rate
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionisation
FA	Formic acid
FASP	Filter-aided sample preparation
FDA	Food and Drug Administration
FDC	Fixed dose combination
FDR	False discovery rate
FWER	Family wise error rate
FWHM	Full width at half maximum
g	Gravitational force (RCF)
GFR	Glomerular filtration rate
GO	Gene ontology
GRAVY	Grand average of hydrophobicity
HAART	Highly active antiretroviral therapy
HFIP	Hexafluoroisopropanol
HILIC	Hydrophilic interaction chromatography
HIV	Human immunodeficiency virus
hOAT	Human organic anion transporter
HRG	Histidine-rich glycoprotein
IAA	Iodoacetamide
ID	Internal diameter
IEX	Ion-exchange chromatography
IGF	Insulin-like Growth Factor
iTRAQ	Isobaric tags for relative and absolute quantitation
KDIGO	Kidney Disease Improving Global Outcomes
LC	Liquid chromatography
LCMS	Liquid chromatography mass spectrometry
LFQ	Label-free quantification

MALDI	Matrix assisted laser desorption ionisation
MAR	Missing at random
MCAR	Missing completely at random
MDRD	Modification in Renal Disease
MNAR	Missing not at random
mRNA	Messenger ribonucleic acid
MRP-2	Multi drug resistant protein type 2
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MW	Molecular weight
m/z	Mass-to-charge
NH_4OAc	Ammonium acetate
NNRTI	Non-nucleoside reverse transcriptase inhibitors
NPC	Normal phase chromatography
NRTI	Nucleotide reverse transcriptase inhibitors
PAGE	Polyacrylamide gel electrophoresis
PCER	Per comparison error rate
PEDF	Pigment epithelium-derived factor
p*I*	Isoelectric point
PMF	Peptide mass fingerprinting
POC	Point of care
PSM	Peptide-spectrum match
PTM	Post-translational modifications
QqQ	Triple quadrupole
QTOF	Quadrupole-time-of-flight
RNA	Ribonucleic acid
RB	Rehydration buffer
RF	Radio-frequency
RmT	Room temperature
RPC	Reverse-phase chromatography
RP-RP	Reverse-phase-reverse-phase

RT	Retention time
sCr	Serum creatinine
SD	Standard deviations
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SGMT	Sequencing grade modified trypsin
SILAC	Stable isotope labelling of amino acids in cell culture
SIV	Simian immunodeficiency virus
SPE	Solid phase extraction
ssRNA	Single stranded ribonucleic acid
SRM	Selected reaction monitoring
START	Strategic timing of antiretroviral therapy
SWATH	Sequential Window Acquisition of all Theoretical Mass Spectra
TAF	Tenofovir alafenamide
TCA	Trichloroacetic acid
TDF	Tenofovir disoproxil fumarate
TMT	Tandem mass tags
TOF	Time-of-Flight
UNAIDS	United Nations Programme on HIV/AIDS
USS	Urea Sample Solution
VW	Variable windows

List of Figures

Figure 1.1	HIV family tree.	5
Figure 1.2	Schematic overview of the HIV-1 replication cycle.	7
Figure 1.3	A general overview of the time course of untreated HIV infection.	8
Figure 1.4	Timeline of antiretroviral drugs.	11
Figure 1.5	Activation and mechanism of action of tenofovir disoproxil fumarate.	13
Figure 1.6	Publications (NCBI PubMed) associated with “tenofovir” and “renal” complications between January 2001 and April 2018.	14
Figure 1.7	Schematic representation of a kidney proximal tubule cell.	15
Figure 1.8	Schematic representation of the central dogma and its relationship to the proteome.	21
Figure 1.9	Overview of acquisition schemes for tandem mass spectrometry.	26
Figure 1.10	Schematic overview of top-down and bottom-up proteomics.	28
Figure 1.11	Schematic overview of in-gel (left) and in-solution (right) workflows used in bottom-up proteomics.	30
Figure 1.12	A typical mass spectrometry-based proteomics pipeline.	31
Figure 1.13	Schematic overview of HPLC instrumentation.	36
Figure 1.14	Common reverse phase HPLC stationary phases.	39
Figure 1.15	Schematic overview of the electrospray ionisation A) source and B) process.	41
Figure 1.16	A typical mass spectrum.	42
Figure 1.17	A schematic representation of a quadrupole mass analyser.	43
Figure 1.18	An overview of a triple quadrupole mass spectrometer showing three distinct quadrupoles.	44
Figure 1.19	Schematic overview of different tandem mass spectrometry scans performed on a triple quadrupole mass spectrometer.	45
Figure 1.20	A hybrid QTOF mass analyser.	46
Figure 1.21	Schematic overview of data-dependent acquisition.	47
Figure 1.22	Schematic overview of SWATH™-MS.	49
Figure 1.23	Variable Q1 window generation using SWATH™ Variable Window Calculator tool v1.0.	50

Figure 1.24	Overview of peptide mass fingerprinting.	53
Figure 1.25	A typical tandem mass spectrometry experimental workflow.	54
Figure 1.26	Overview of quantification strategies for relative (A) and absolute (B) protein quantification.	60
Figure 1.27	Schematic overview of the two LFQ methodologies.	61
Figure 2.1	Schematic overview of the parameters optimised in the development of a robust and reproducible urinary proteome profiling workflow.	69
Figure 2.2	Overview of the comparison between the Norgen Kit, HILIC and (m)FASP methods for urinary proteome profiling.	80
Figure 2.3	A generic gradient elution profile.	83
Figure 2.4	A schematic overview depicting the concatenation strategy.	89
Figure 2.5	Overview of the complete workflow using 2D RPRP.	90
Figure 2.6	Summary of the optimised workflow developed and employed in this study for urinary proteome profiling of clinical samples.	97
Figure 2.7	Volcano plot of candidate markers after comparing all cases to all controls.	99
Figure 3.1	Summary of protein yields from various methods.	103
Figure 3.2	Summary of protein quantification methods.	105
Figure 3.3	Comparison of protein digestion using the SMART™ Digest kit and SGMT.	106
Figure 3.4	Summary of data derived after C18 clean-up.	107
Figure 3.5	Summary of data following online C18 clean-up.	108
Figure 3.6	Protein and peptide properties for Norgen, HILIC and mFASP methods assessed using ExPASy pI/MW and protein GRAVY calculators.	110
Figure 3.7	Gradient lengths of 60 and 90 minutes were assessed.	111
Figure 3.8	Summary of gradient slope experiments.	112
Figure 3.9	Summary of peptide loads assessed.	113
Figure 3.10	Peptide and protein identifications at 1% FDR after varying the number of MS/MS scans and accumulation time.	114
Figure 3.11	Peptide and protein identifications at 1% FDR using default and crowdsourced CE parameters.	115

Figure 3.12	Total ion chromatogram of a typical urinary protein sample.	116
Figure 3.13	Summary of VW vs CW comparison.	119
Figure 3.14	Graph of library matches showing peptides and proteins matched in different libraries at 1% FDR.	122
Figure 3.15	Coefficient of variation when analysing three technical replicates using the complete workflow.	123
Figure 3.16	Overlaid total ion chromatographs of a single injection from each technical workflow replicate.	124
Figure 3.17	Panther gene ontology pathway analysis of all proteins identified in the clinical sample cohort.	126
Figure 3.18	Summary of clinical characteristics used in the diagnosis of AKI.	128
Figure 3.19	Over-representation pathway map derived from Reactome v65.	138
Figure 3.20	Proteins exhibiting increasing differential abundance as kidney damage progresses.	139

List of Tables

Table 2.1	Commercially available clean-up kits that were assessed.	78
Table 2.2	Collision energy parameters.	86
Table 2.3	Additional parameters optimised for DDA analysis.	87
Table 2.4	Power analysis using ClinCalc software.	95
Table 2.5	Randomised block design for patient urinary protein analysis.	96
Table 3.1	Summary of urinary proteome profiling data according to isolation method tested.	109
Table 3.2	Summary of data derived from acquisition parameters used to optimise methodology for data dependent analysis.	117
Table 3.3	Total peptides and proteins derived from each fraction following 2DRPRP MS analysis.	120
Table 3.4	Summary of library matches showing average peptides and proteins matched using different reference libraries.	121
Table 3.5	Demographic details and relevant clinical characteristics of the patient cohort used in the assessment of kidney function.	127
Table 3.6	Putative biomarkers arranged in ascending q value (decreasing confidence) order.	129
Table 3.7	Putative biomarkers identified at 1% FDR.	132
Table 3.8	GO pathway analysis of putative biomarkers.	135

Research Outputs

Oral presentations

Ireshyn Govender, Stoyan Stoychev, Demetra Mavri-Damelin, Neil Martinson and Dalu Mancama**.** A novel and robust method for urinary proteome profiling applied in an exploratory case-control study. Human Proteome Organisation (HUPO) World Congress 2018, Orlando – Florida, USA, October 2018**.**

Ireshyn Govender, Stoyan Stoychev, Demetra Mavri-Damelin, Neil Martinson and Dalu Mancama. Evaluation of methods for urinary proteome profiling and application to acute kidney injury associated with HIV anti-retroviral therapy. South African Proteomics Symposium 2018. Johannesburg – South Africa, August 2018.

Ireshyn Govender, Stoyan Stoychev, Demetra Mavri-Damelin, Neil Martinson and Dalu Mancama. A novel and robust method for urinary proteome profiling applied in an exploratory case-control study. CSIR Emerging Researchers Symposium. Pretoria – South Africa, July 2018

Ireshyn Govender, Stoyan Stoychev, Demetra Mavri-Damelin, Neil Martinson and Dalu Mancama. The discovery of kidney-injury related protein biomarkers associated with first-line antiretroviral treatment in South Africa by SWATH-MS. WITS Molecular Biosciences Research Thrust (MBRT) Research Day. Johannesburg – South Africa, November 2017.

Ireshyn Govender, Stoyan Stoychev, Demetra Mavri-Damelin, Neil Martinson and Dalu Mancama. The discovery of kidney-injury related protein biomarkers associated with first-line antiretroviral treatment in South Africa by SWATH-MS. 8th Cross Faculty Postgraduate Symposium. Johannesburg – South Africa, October 2017.

Ireshyn Govender, Stoyan Stoychev, Demetra Mavri-Damelin, Neil Martinson and Dalu Mancama. The discovery of kidney-injury related protein biomarkers associated with first-line antiretroviral treatment in South Africa by SWATH-MS. ACGT Proteomics Mini-Symposium 2017. Pretoria – South Africa, October 2017

Ireshyn Govender, Stoyan Stoychev, Demetra Mavri-Damelin, Neil Martinson and Dalu Mancama. The discovery of kidney injury-related protein biomarkers associated with antiretroviral treatment in South Africa using mass spectrometry. 2016 African Centre for Gene Technologies (ACGT) Proteomics Symposium and Skyline Workshop. Irene – South Africa, December 2016.

Ireshyn Govender, Stoyan Stoychev, Demetra Mavri-Damelin, Neil Martinson and Dalu Mancama. The discovery of kidney injury biomarkers associated with antiretroviral treatment. SAMRC Soweto Matlosana Collaborative Centre for HIV/AIDS and TB (SoMCHAT) Young Researcher Conference. Soweto – South Africa, November 2016.

Posters

Ireshyn Govender, Stoyan Stoychev, Demetra Mavri-Damelin, Motshidisi Khumalo, Faheem Seedat, Ebrahim Variava, Neil Martinson and Dalu Mancama. The Discovery of kidney injury-related protein biomarkers associated with first-line antiretroviral treatment in South Africa using microflow SWATH® mass spectrometry. HUPO World Congress 2017. Dublin – Ireland, September 2017.

Ireshyn Govender, Stoyan Stoychev, Demetra Mavri-Damelin, Neil Martinson and Dalu Mancama. Expanding spectral library depth using 2D-LCMSMS for the discovery of ARV induced kidney injury-related protein biomarkers in urine. MaxQuant Summer School 2016. Oxford – United Kingdom, July 2016.

Ireshyn Govender, Stoyan Stoychev, Demetra Mavri-Damelin, Neil Martinson and Dalu Mancama. The discovery and validation of kidney injury-related protein biomarkers associated with antiretroviral treatment in South Africa using mass spectrometry. 7th WITS Postgraduate Cross Faculty Symposium. Johannesburg – South Africa, March 2016.

Ireshyn Govender, Stoyan Stoychev, Demetra Mavri-Damelin, Neil Martinson and Dalu Mancama. The discovery and validation of kidney injury-related protein biomarkers associated with antiretroviral treatment in South Africa using mass spectrometry. WITS MBRT Research Day. Johannesburg – South Africa, December 2015.

Ireshyn Govender, Stoyan Stoychev, Demetra Mavri-Damelin, Neil Martinson and Dalu Mancama. The discovery and validation of kidney injury-related protein biomarkers associated with antiretroviral treatment in South Africa using mass spectrometry. CSIR 5th International Conference - Ideas that Work. Pretoria - South Africa, October 2015.

Publications

Ireshyn Govender, Demetra Mavri-Damelin, Faheem Seedat, Ebrahim Variava, Neil Martinson, Stoyan Stoychev and Dalu Mancama. A novel and robust method for urinary proteome profiling applied in an exploratory case-control study. *Clinical Proteomics***. (**CLIP-D-18-00046-R2) under review.

Table of Contents

Declaration i

Abstract ii

Acknowledgements iii

List of Abbreviations iv

List of Figures viii

List of Tables xi

Research Outputs xii

Chapter 1 – Literature review **2**

1.1. Project Overview 2

1.2. HIV/AIDS: Origins, current status and future outlook 4

1.2.1. Overview of HIV/AIDS 4

1.2.2. Evolution of HIV 4

1.2.3. Life cycle, stages of infection and disease progression 6

1.2.4. Antiretroviral therapy 9

1.2.5. Treatment of HIV in South Africa 11

1.2.6. Tenofovir metabolism, activity and adverse reactions 12

1.2.7. Tenofovir associated nephropathy and kidney function testing 15

1.2.8. New nucleoside analogues on the horizon 17

1.2.9. Personalised medicine in HIV management 18

1.3. What are biomarkers? 19

1.4. The Proteomic Biomarker Discovery-Personalised Medicine Link 19

1.5. Proteomics based biomarker discovery 20

1.5.1. History of proteomics 20

1.5.2. Urinary proteomics 22

1.5.2.1. Urine as a biomarker source – the advantages 22

1.5.2.2. Urine as a biomarker source – the disadvantages 23

1.5.2.3. Urine in clinical proteomics applications 23

1.6. Mass spectrometry-based proteomics 25

1.6.1. Why mass spectrometry? 25

1.6.2. Discovery vs targeted proteomics 26
1.6.3. Top-down vs bottom-up proteomics 27
1.6.4. Bottom-up proteomics for global proteome profiling 28
1.7. Typical mass spectrometry-based proteomics pipeline 31
1.7.1. Study design 31
1.7.1.1. Sources of variation and bias 31
1.7.2. Sample collection and sample preparation 34
1.7.2.1. Sample collection 34
1.7.2.2. Sample preparation 34
1.7.2.2.1. History of liquid chromatography 35
1.7.2.2.2. Effect of flow rate on liquid chromatography 37
1.7.2.2.3. Chromatographic chemistries 38
1.7.3. Data acquisition 40
1.7.3.1. What is a mass spectrometer? 40
1.7.3.2. Ionisation techniques 40
1.7.3.3. Types of mass spectrometers 41
1.7.3.3.1. Quadrupole mass analysers 43
1.7.3.3.2. Time-of-flight mass analyser 45
1.7.3.4. Data acquisition methods 46
1.7.3.4.1. Data-dependent acquisition 46
1.7.3.4.2. Selected (multiple) reaction monitoring 48
1.7.3.4.3. Sequential Window of All Theoretical Mass Spectra (SWATH™-MS) or Data- Independent acquisition (DIA) 48
1.8. Data processing and analysis 52
1.8.1. Peptide mass fingerprinting (PMF) 52
1.8.2. Peptide sequencing by tandem mass spectrometry 53
1.8.3. False discovery rate analysis in database searching 55
1.9. Data analysis in proteomics 56
1.9.1. Protein identification overview 56
1.9.2. Sequence database 57

1.9.3. Mass tolerance 57
1.9.4. Enzyme selection and partial digestion 57
1.9.5. Modifications 58
1.10. Protein quantification strategies 58
1.10.1. Label-based quantification 59
1.10.2. Label-free quantification 60
1.10.3. Quantification considerations 62
1.11. Limitations of mass spectrometry-based proteomics 63
1.12. Aims and objectives 65
Chapter 2 - Materials and Methods 66
Part A - Method development and optimisation for urinary proteome profiling in clinical proteomics 66
2.1. Database search parameters 66
2.2. Isolation of urinary proteins 69
2.2.1. Methods employed to isolate urinary proteins 69
2.2.2. Methanol precipitation 69
2.2.3. Acetonitrile precipitation 69
2.2.4. Acetone and acetone-trichloroacetic acid precipitation 70
2.2.5. Kit-based urinary protein methods 70
2.2.5.1. Collection, concentration and preservation 70
2.2.5.2. Protein isolation following preservation 71
2.3. Quantification methods 71
2.3.1. Bicinchoninic assay 71
2.3.2. Bradford assay 72
2.3.3. Qubit Assay 72
2.3.4. 2-D GE Quantification Kit 73
2.4. Reduction, alkylation and digestion 74
2.4.1. Digestion using sequencing grade modified trypsin 74
2.4.2. Digestion using the SMART Digest™ kit 75
2.5. Peptide sample clean-up optimisation 75
2.5.1. Nest group 96-well MiniSpin clean up protocol 76
2.5.2. SOLAμ™ SPE Plates protocol 76

2.5.3. C18 Stage Tip clean up protocol 77
2.5.4. Online C18 clean-up 78
2.6. Assessment of novel methods developed in the study vs current methods for urinary proteome sample preparation and profiling 79
2.6.1. Novel, one-step clean-up and digestion, microsphere-based method for sample preparation 81
2.6.2. Sample preparation using FASP 81
2.6.3. Peptide and protein properties comparison 82
2.7. LCMS/MS data acquisition method development 82
2.7.1. Gradient elution 82
2.7.1.1. Gradient length 83
2.7.1.2. Gradient slope 84
2.7.1.3. Column load (µg) per injection 84
2.8. MS data acquisition parameters for data dependent analysis 84
2.8.1. MS/MS and accumulation time 84
2.8.2. Default Analyst® vs crowdsourced optimised rolling collision energy parameters 85
2.9. Additional parameters for data dependent analysis 86
2.10. MS data acquisition and data processing parameters for data independent (SWATH™) analysis 87
2.10.1. Constant vs variable windows 87
2.10.2. Two-dimensional reverse phase-reverse phase (2D RPRP) LCMS data-dependent analysis for SWATH™ library generation 87
2.10.3. Two-dimensional reverse phase-reverse phase methodology 88
2.11. Spectral library generation settings 91
2.12. Complete workflow technical variation assessment 91
2.13. Panther gene ontology (GO) analysis 92
2.14. Summary of data dependent analysis methodology after individual component optimisation. 92
Part B - Urinary protein biomarker discovery: An exploratory case-control study 93
2.15. Cohort details 93
2.16. Demographic and clinical data 94
2.17. Power analysis 95

2.18. Randomised block design for sample analysis 95
2.19. Summary of methods used in sample preparation and mass spectrometry data acquisition 97
2.20. Processing of patient data using Spectronaut™ 11 98
2.20.1. Selection of candidate biomarkers 99
2.20. Gene ontology pathway analysis of putative biomarkers 100
Chapter 3 **101**
Part A - Method development and optimisation for urinary proteome profiling in clinical proteomics **101**
3.1. Urinary protein isolation methods 101
3.2. Quantification methods 104
3.3. Digestion methods 106
3.4. Commercial, offline C18 clean-up kits 107
3.5. Online clean-up vs no clean-up 108
3.6. Norgen kit, HILIC and (m)FASP comparisons 109
3.7. LCMS/MS Data acquisition method development 111
3.7.1. Gradient length 111
3.7.2. Gradient slope 112
3.7.3. Column load per injection 113
3.8. MS data acquisition parameters for data dependent analysis 114
3.8.1. MS/MS and accumulation time 114
3.8.2. Rolling collision energy (CE) parameters 115
3.9. Typical total ion chromatogram derived from the optimised method. 116
3.10. MS data acquisition and processing parameters for data independent analysis 118
3.10.1. Constant vs variable SWATH™ windows 118
3.10.2. Two dimensional RPRP 120
3.10.3. Varying reference spectral library 121
3.11. Technical variation assessment 123
3.12. Gene ontology analysis 125
Part B - Urinary protein biomarker discovery: An exploratory case-control study **127**
3.13. Demographic details and clinical characteristics 127
3.14. Putative biomarkers 129

3.15. Reactome pathway gene ontology analysis 135
3.16. Markers showing increasing urinary abundance as AKI severity progresses 139
Chapter 4 – Discussion **140**
4.1. Method development and optimisation 140
4.2. Urinary protein biomarker discovery: an exploratory case-control study 142
Chapter 5 - Conclusion and future perspectives **151**
Chapter 6 - References **154**
Chapter 7 - Appendix **190**

Chapter 1

Literature Review

1.1 Project Overview

The HIV/AIDS pandemic is a health problem that receives high importance worldwide. In 2016, the number of individuals living with HIV worldwide was estimated to be 36.7 million, which included 1.8 million new infections for that year (*UNAIDS*, 2018) . In South Africa, people living with HIV account for 7.06 million of the population (~12.5 %) (Statistics South Africa, 2017), which translates to ~19 % of the global HIV/AIDS burden (*UNAIDS*| *South Africa*, 2018). For adults aged 15-49 years, ~18.9 % are HIV positive (Statistics South Africa, 2017). These alarming statistics affect the country severely as the country's workforce is diminished, increasing costs for public healthcare and a concomitant life expectancy decline.

Over the past few decades, the global scientific community has taken giant leaps towards finding and developing suitable HIV/AIDS prevention, treatment and management regimens. This is largely due to rapid advances in molecular and cellular biology. As a result, the nature of the Human Immunodeficiency Virus is now more clearly understood and thus development of target driven treatment approaches is the gold-standard.

Antiretroviral (ARV) therapy for the treatment of HIV has resulted in a shift from HIV being a fatal to a potentially chronic disease. Early monotherapy ARV treatments were flawed as they were required in high-doses and often did not achieve high efficacy. Since the mid-1990s, highly active antiretroviral therapy (HAART) - where 3 or more ARV drugs are used in combination - has become the norm (Pavlos and Phillips, 2012). Of those infected in South Africa, more than 3.5 million receive free ARV treatment and this figure is set to rise in the coming years (*UNAIDS*, 2018).

Despite ARV treatment being largely successful – there are many limitations such as cost, patient adherence and short and long-term side effects (Volberding and Deeks, 2010). Strict adherence to the treatment regimen is vital for the suppression of viral replication but even with strict compliance, a patient's health is never fully restored. Coupled to this are adverse side-effects associated with long-term ARV use including non-AIDS related complications such as

cardiovascular (Dau and Holodniy, 2008; Aberg, 2009; Palella and Phair, 2011), neurocognitive (Giunta *et al.*, 2011; Robertson, Liner and Meeker, 2012; Winston *et al.*, 2013), liver (Puoti *et al.*, 2009) and kidney injury (Choi *et al.*, 2009; Estrella, Fine and Atta, 2010; Scherzer *et al.*, 2012).

While ARV treatment is highly effective in some individuals, it has adverse effects in others. These and other non-AIDS related side effects have prompted a shift towards personalised medicine in the treatment of HIV. Personalised medicine aims to avoid these adverse effects by developing diagnostic tests based on genetic and/or molecular alterations to predict patient response to a particular therapy even before the drug or treatment is administered (Hamburg and Collins, 2010). In so doing, clinicians can tailor a treatment programme for individual patients e.g. alter the dose or prescribe different drugs completely.

In an attempt to identify kidney injury at an early stage this study focuses on changes that can be identified in the urinary proteome indicating acute kidney injury (AKI) related to first line ARV treatment. Clinicians have reported that, in certain population sub-groups in South Africa, approximately 10% of new patients enrolled onto first line ARV programs experience acute, adverse kidney related side effects (Dr Ebrahim Variava, PHRU, Tshepong Hospital, personal communication, 31 January 2014). Thus, the current study employs mass spectrometry (MS) based proteomics to compare the urinary proteomes of patients on first line ARV therapy who experience renal injury related adverse side effects (cases) and those that do not experience such effects (controls). The identification of differentially expressed proteins can potentially be used to develop a point of care diagnostic that can detect the early onset of kidney damage, or predict the progression of this damage, in patients who have initiated first line ARV treatment. This will allow for corrective treatment and monitoring prior to or limiting further ARV associated kidney complications.

1.2 HIV/AIDS: Origins, current status and future outlooks

1.2.1 Overview of HIV/AIDS

Acquired immune deficiency syndrome (AIDS) was first described as a disease in the early 1980's when a sharp increase in opportunistic infection and malignancies was documented in homosexual men. The causative agent(s) at the time was unknown. In 1984, the National Cancer Institute and Pasteur Institute jointly announced that Human T-Lymphotropic Virus Type III and Lymphadenopathy Associated Virus were the causative agents for AIDS (Marx, 1984). When it was discovered that these viruses are actually the same, The International Committee on the Taxonomy of Viruses declared that the virus responsible for the development of AIDS will be called human immunodeficiency virus (HIV) (Case, 1986). This led to intensive research into a disease syndrome, that has since become the most devastating and fear instilling communicable disease of our modern era (Sharp and Hahn, 2011; Maartens, Celum and Lewin, 2014).

HIV/AIDS is primarily classified as a sexually transmitted disease as nearly 80% of all adult cases worldwide are as a result of exposure to mucosal surfaces during sexual intercourse (Fettig *et al.*, 2014). Other methods of transmission include percutaneous (Beltrami *et al.*, 2000) and perinatal transmission (Shubber *et al.*, 2014). Since its emergence a mere 35 years ago, HIV/AIDS has plagued more than 75 million people worldwide with a mortality rate nearing 50%. Currently, UNAIDS estimates that there are 35 million people living with HIV/AIDS. The greatest HIV/AIDS burden is concentrated in developing countries with sub-Saharan Africa experiencing the highest prevalence rates (Maartens, Celum and Lewin, 2014). Statistics South Africa estimates that 7 million people are infected with HIV as of August 2017 (Statistics South Africa, 2017; *UNAIDS| South Africa*, 2018).

1.2.2 Evolution of HIV

Since its discovery in the early 1980's, there has been widespread research into the origins and evolution of HIV. In 1999, researchers demonstrated that a strain of simian immunodeficiency virus (SIV) is very similar to HIV (Gao *et al.*, 1999). It was then postulated and now widely accepted that HIV resulted from a zoonosis in middle and eastern Africa in the early 1900's (Apetrei, Robertson and Marx, 2004; Marx, Apetrei and Drucker, 2004; Sharp and Hahn, 2011; Maartens, Celum and Lewin, 2014).

The most common theory that accounts for this zoonosis is the "hunter theory". Briefly, SIV was passed onto humans as a result of chimps and other smaller primates being killed and eaten

(Volberding and Deeks, 2010). Normally the SIV would not have affected humans adversely as their immune systems would fight off the virus. However, this virus adapted and evolved resistance mechanisms into what is called HIV today. There are two distinct types of HIV virus that can infect man *viz.* HIV-1 and HIV-2 (Jaffar *et al.*, 2004; Nyamweya *et al.*, 2013). The lesser known and uncommon form, HIV-2, is concentrated in West Africa. Whereas, HIV-1 is the most predominant virus that accounts for the vast majority of cases worldwide (~95%) and generally when one refers to HIV, HIV-1 is assumed to be the reference point.

HIV-1 is a retrovirus belonging to the family of lentiviruses. Lentiviruses are characterised by a chronic course of disease, long clinical latency periods and continual viral replication (Hoffmann, Rockstroh and Rockstroh, 2015). Strains of HIV-1 can be classified into four main groups (M, N, O and P). Group M is the predominant group that accounts for most of the HIV epidemic worldwide. Further subdivisions of HIV-1 Group M into clades (subtypes) A-K have been identified, with each having genetically distinct features. Additionally, in individuals infected with more than one subtype, hybrid forms or "quasispecies" have been identified which contain genetic material from different clades – referred to as circulating recombinant forms (CFRs) (Hemelaar, 2012; Lau and Wong, 2013).

In the Americas, Europe and Australasia, subtype-B is the most dominant. While subtype-C dominates the high prevalence regions of sub-Saharan Africa and India and accounts for nearly 50% of all HIV cases worldwide (*UNAIDS*, 2018) (Figure 1.1).

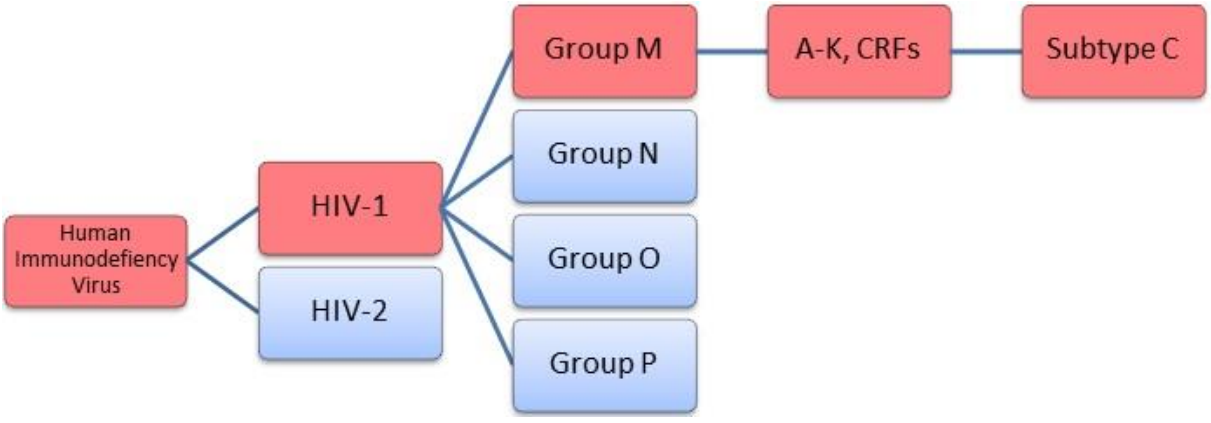


Figure 1.1: HIV family tree. HIV-1 Group M is the strain of HIV that is responsible for the HIV/AIDS pandemic. Subtype C is the most prevalent in South Africa.

1.2.3 Life cycle, stages of infection and disease progression

HIV is an adaptable virus which infects the host's immune cells and uses these host cells to its advantage by harnessing the cellular machinery to reproduce. HIV expresses gp120 proteins on its surface which facilitate binding to CD4 receptors on T cells (Arts and Hazuda, 2012). Activated CD4 T lymphocytes are the primary targets for HIV, however other cells that express chemokine co-receptors CCR5 and CXCR4 such as resting CD4 cells, dendritic cells, macrophages and monocytes can also be infected (Maartens, Celum and Lewin, 2014). Co-receptor CCR5 and/or CXCR4 recruitment is required for complete binding and subsequent viral fusion to occur. Fusion is the process in which the viral lipid layer and host cellular membrane become contiguous thereby allowing full entry of viral genetic information into the cell (Doms and Moore, 2000; Blumenthal, Durell and Viard, 2012). Once fusion with the host cell is complete, the viral capsid [containing genetic information - viral single stranded ribonucleic acid (ssRNA) and enzymes - reverse transcriptase, integrase) is uncoated and its contents are injected into the host cytoplasm (Maartens, Celum and Lewin, 2014).

Reverse transcriptase is an essential component to HIV. It converts viral ssRNA into viral deoxyribonucleic acid (DNA). It is important to note that the process of reverse transcription is highly error prone and thus multiple mutations can occur resulting in viral drug resistance (Roberts, Bebenek and Kunkel, 1988). The newly generated viral DNA is then integrated into the host DNA via the integrase enzyme. During normal cellular function, RNA polymerase converts the integrated HIV DNA into messenger ribonucleic acid (mRNA) which encodes for viral proteins (Maartens, Celum and Lewin, 2014). After synthesis at the ribosomes, these proteins are transported to the cell surface in a polyprotein chain together with a single ssRNA, where assembly, packaging and budding of immature virions takes place. At this stage the immature virions are non-infective and require further processing by a protease enzyme. The protease enzyme cuts the polyprotein chain into smaller proteins which are essential for the structure and functioning of a mature, infective virion. Finally, the newly formed infective virion is mature and able to infect other CD4 cells (Maartens, Celum and Lewin, 2014). This process continues indefinitely as long as the cell is active and thus thousands of viruses are produced in an untreated patient (Figure 1.2).

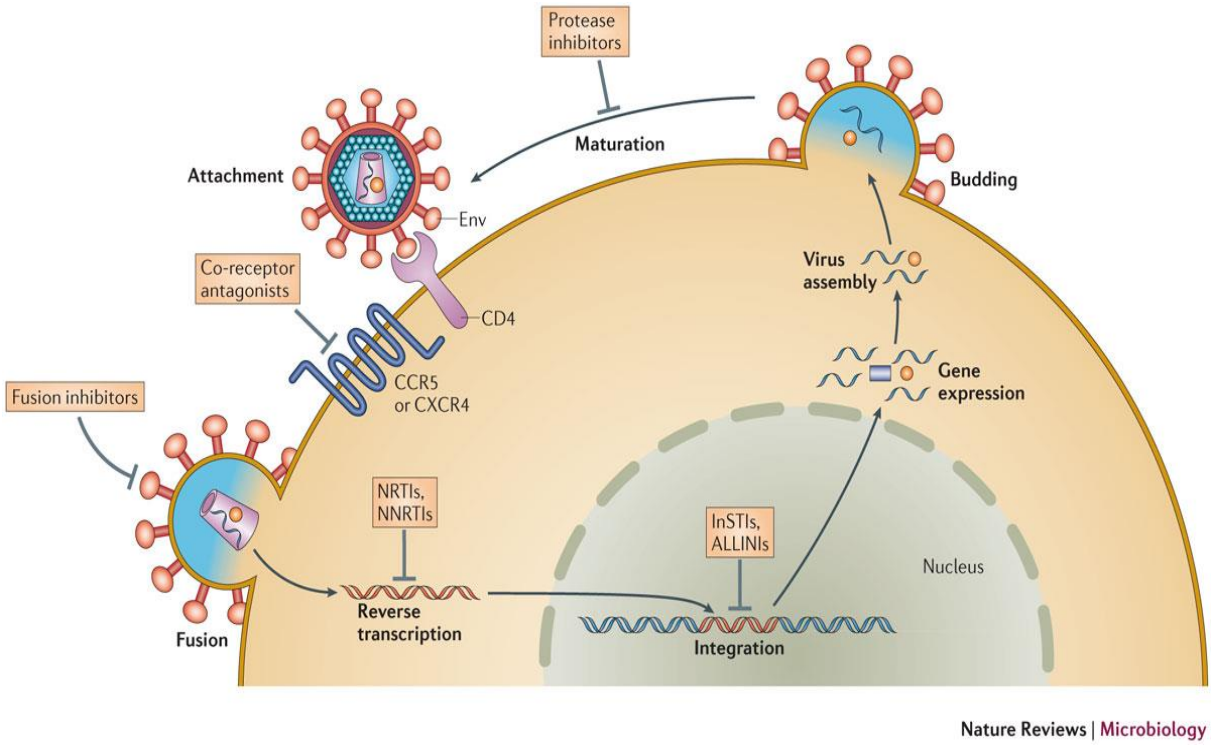


Figure 1.2: Schematic overview of the HIV-1 replication cycle. Various stages that are targeted by antiretroviral drugs are listed in each box (Laskey and Siliciano, 2014**).**

During the first few weeks after infection, the process of seroconversion takes place rapidly within the host termed "acute infection". This results in the production of antibodies by the host's immune system against HIV. Patients may feel symptoms similar to those experienced during colds and flus i.e. sore throat, headaches, muscle and joint pain and malaise therefore patients generally do not suspect HIV infection. During this stage the virus rapidly reproduces within the body with concomitant decreases in the CD4 lymphocyte pool (Volberding and Deeks, 2010).

Clinical latency or "asymptomatic infection" then ensues, often for many years, where the patient's immune system manages to seemingly keep the virus suppressed. Patients seem to be in overall good health however their immune systems are gradually being attacked by the virus as it actively replicates at low level (Langford, Ananworanich and Cooper, 2007). The virus stimulates the immune system throughout the infection and thus patients are often in a state of chronic inflammation which further aids in the progression of the infection (Pavlos and Phillips, 2012) (Figure 1.3).

The hallmark of untreated HIV is the gradual depletion and increased destruction of the CD4 T cell pool (Maartens, Celum and Lewin, 2014). The untreated infection progresses gradually and ultimately leads to AIDS – a state where a patient's immune system is completely compromised rendering them vulnerable to opportunistic infections. Once an immunocompromised patient contracts an opportunistic infection their life expectancy is reduced to approximately 1 year (Maartens, Celum and Lewin, 2014). Thus, it is critical for patients to be tested early and enrolled into antiretroviral therapy (ART) programmes to ensure that their risk for progression to AIDS is diminished.

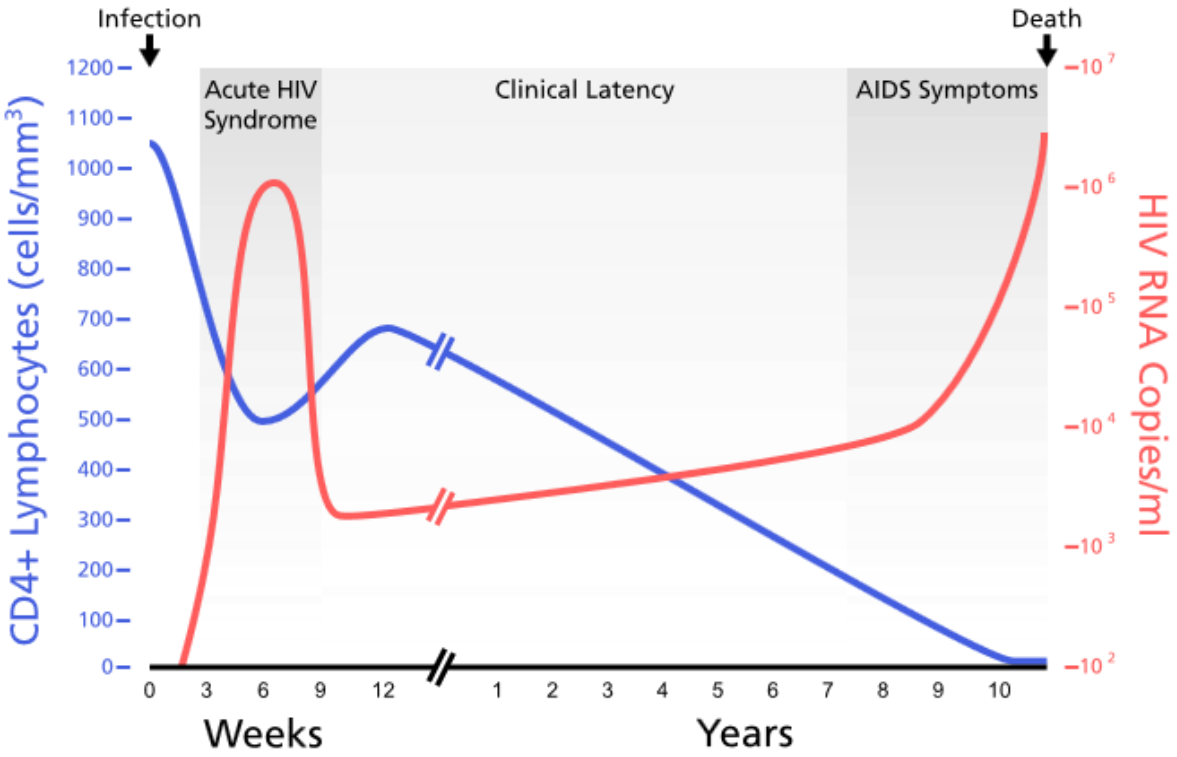


Figure 1.3: A general overview of the time course of untreated HIV infection. The relationship between CD4 counts and viral load (HIV copies) is shown for the average patient [adapted from (Pantaleo, Graziosi and Fauci, 1993)].

1.2.4. Antiretroviral therapy

The use of ART has significantly decreased the number of AIDS-related deaths, from 2.3 million in 2005 to 1 million in 2016, worldwide (Maartens, Celum and Lewin, 2014; *UNAIDS*, 2018). These drugs are currently unable to eradicate the disease completely, therefore the primary goal for the foreseeable future is to change HIV from a slow-progressing, fatal disease into a lifelong, chronic disease (Palmisano and Vella, 2011). There are multiple types of ARVs that are classed according to the stage at which they act within the HIV life cycle (Volberding and Deeks, 2010; Arts and Hazuda, 2012).

The first step in the HIV replication cycle is a logical target for antiretroviral agents. Ideally these block the ability of the virus to bind the CD4/CCR5/CXCR4 receptors and thus prevent entry into the cell. Fusion inhibitors are similar in mechanism of action to entry inhibitors as they prevent the virus from merging with the cell wall and thus prevent entry into the cell (Arts and Hazuda, 2012; Maartens, Celum and Lewin, 2014). This class of ARV is collectively referred to as entry inhibitors e.g. maraviroc and enfuvirtide.

Integrase inhibitors (integrase strand transfer inhibitors – InSTIs) block the enzyme integrase from catalysing the binding of viral DNA to host DNA e.g. raltegravir (Lau and Wong, 2013; Maartens, Celum and Lewin, 2014).

Protease inhibitors are an important class of ARVs that inhibit the viral protease enzyme thus preventing maturation of the virion. Hence newly formed virions are left in an immature and non-infective state e.g. ritonavir (Hemelaar, 2012; Lau and Wong, 2013).

Reverse transcriptase inhibitors work by blocking the activity of reverse transcriptase. Thus, conversion of viral ssRNA into viral dsDNA is prevented. There are two subclasses of reverse transcriptase inhibitors and each has a different mechanism of action:

Non-nucleoside reverse transcriptase inhibitors (NNRTIs) bind to reverse transcriptase directly in a non-competitive manner at a region near the active site of the enzyme. This complex blocks the active site and reduces ability to efficiently bind to ssRNA and convert it into dsDNA e.g. efavirenz (Volberding and Deeks, 2010; Hemelaar, 2012).

Nucleoside or nucleotide reverse transcriptase inhibitors (NRTIs) where the first class of drugs to gain United States Food and Drug Administration (FDA) approval in the late 1980's (Young,

1988). They are administered as prodrugs and require phosphorylation by endogenous kinases into their active form which carry out the antiviral effects (Arts and Hazuda, 2012). These drugs exert their effects by terminating the growing viral DNA chain by incorporating themselves into the viral DNA during replication by reverse transcriptase. NRTIs lack a 3'-hydroxyl group at its sugar moiety thereby preventing 3'-5' phosphodiester bond formation between the NRTIs and normal (endogenous) nucleotides resulting in chain termination (Arts and Hazuda, 2012; Pavlos and Phillips, 2012; Günthard *et al.*, 2016); examples include abacavir and zidovudine.

The combination of drugs from at least two different classes is now recommended by all healthcare providers to reduce the risk of HIV drug resistance. This approach is called HAART. Most HAART regimens contain two NRTIs and a single drug from one other class (Günthard *et al.*, 2016).

Zidovudine (AZT, Retrovir®), a NRTI, was the first ARV to receive FDA approval in 1987 (Hoffmann and Rockstroh 2015). AZT showed great promise as a monotherapy however its benefits as an effective ARV were marred by it being required in high doses multiple times a day. Monotherapy was the gold standard between 1988 and 1994 as more drugs such as didanosine and stavudine were developed as ARVs. In 1995/6 results from the European-Australian DELTA Study (King, 1995) and the American ACTG 175 Study (Hammer *et al.*, 1996) showed that NRTIs in combination were more effective than monotherapy. This coincided with the first protease inhibitor receiving FDA approval. This was a critical point in the HIV/AIDS treatment pipeline – thus "highly active antiretroviral therapy" was born. To date greater than 20 FDA approved ARV drugs are available to patients and the number is likely to increase (Figure 1.4).

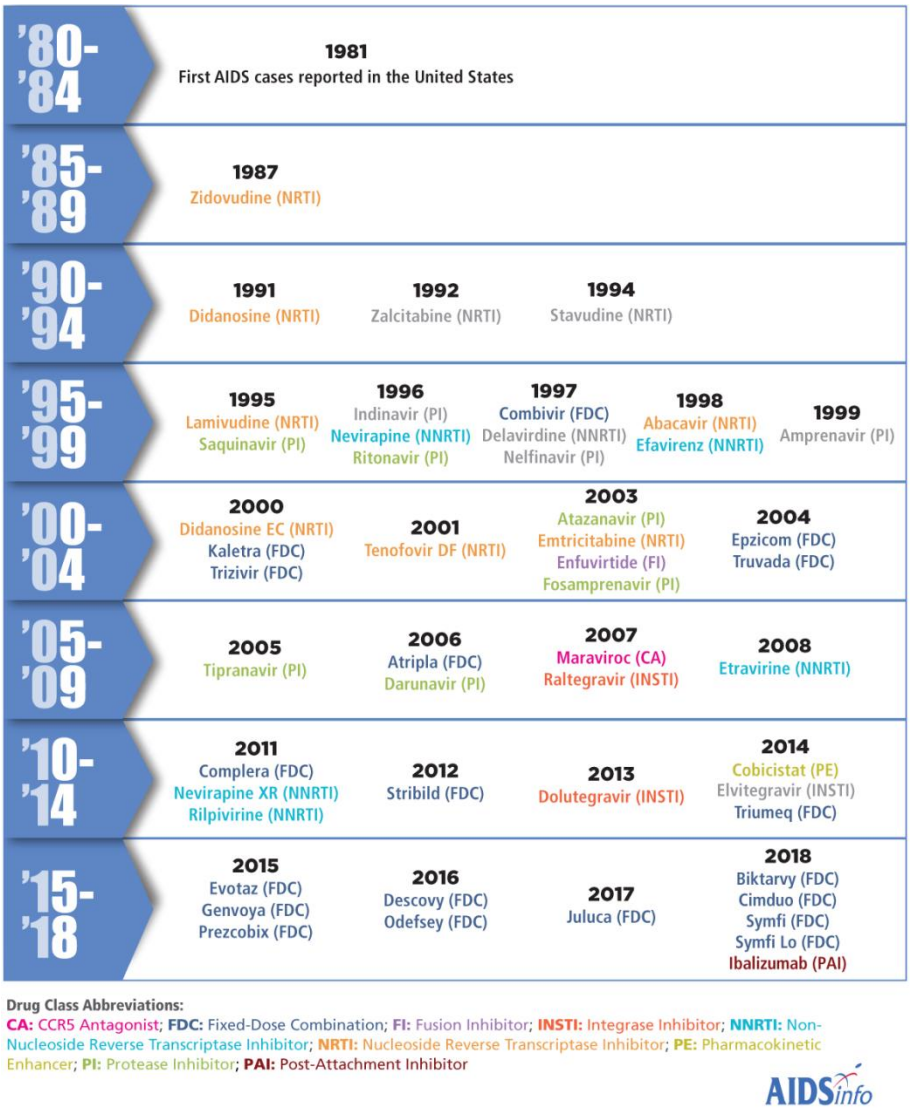


Figure 1.4: Timeline of antiretroviral drugs. The ARV drug development pipeline is continually evolving, there are a number of new drug formulations that are currently being rolled out year on year (*AIDSinfo*, 2018).

1.2.5. Treatment of HIV in South Africa

South Africa boasts the largest ARV programme in the world where treatment is made freely available to more than 3.5 million patients in South Africa (Statistics South Africa, 2017). The first fixed dose combination (FDC) drug approved by the FDA, on 12 July 2006, was only made available in South Africa in April 2013. This was a breakthrough in HIV/AIDS treatment as this was the first "one pill" ARV made available to South African HIV patients. Marketed under the trade name Atripla®, this once-daily, one-pill drug combined three drugs i.e. the nucleoside reverse transcriptase inhibitors tenofovir (Viread®) and emtricitabine (Emtriva®), and the non-nucleoside reverse transcriptase inhibitor efavirenz (Sustiva®). Due to its ease of

use, cost and - most importantly - efficacy, this treatment became the gold-standard, "blanket" treatment for all patients eligible for HAART (Meintjes, 2014; Meintjes *et al.*, 2017).

Prior to 10 May 2016 only patients with a CD4 count less than or equal to 350 cells/µl whole blood were eligible for free, state provided ART. Since then new legislation was passed which stated that all newly diagnosed patients are eligible for immediate FDC treatment irrespective of their CD4 cell count (Meintjes *et al.*, 2017). These new treatment guidelines were passed after the completion of two seminal studies in which patients were initiated on ART irrespective of the CD4 counts. These two studies: Strategic timing of antiretroviral therapy (START) and Early antiretroviral treatment and/or early isoniazid prophylaxis against tuberculosis in HIV-infected adults (TEMPRANO ANRS 12136) both showed that immediate initiation of ART lead to significant individual clinical benefit as opposed to initiation after CD4 count ≤ 350 cells/µl (Lundgren and Group, 2015; TEMPRANO ANRS 12136 Study Group *et al.*, 2015).

1.2.6. Tenofovir metabolism, activity and adverse reactions

In South Africa, all new HIV patients receive Atripla® as the gold standard FDC ARV (Meintjes *et al.*, 2017) and in the rest of the world approximately 50% of all ARV regimens contain tenofovir as one of its components (Prabhu, 2016; Watkins *et al.*, 2017).

Tenofovir is administered in an ester prodrug form as tenofovir disoproxil fumarate (TDF) and requires intracellular hydrolysation and subsequent phosphorylation to its active form tenofovir diphosphate which exerts the antiviral effect (Masho *et al.*, 2007; De Clercq, 2011). Tenofovir diphosphate is a structural analogue of deoxyadenosine triphosphate and competitively inhibits its incorporation into the newly forming viral DNA (Fung *et al.*, 2002; Masho *et al.*, 2007). This results in chain termination as tenofovir diphosphate does not contain a hydroxyl moiety that allows for attachment of the next nucleic acid by DNA polymerase (Fung *et al.*, 2002) (Figure 1.5).

Tenofovir is favoured over other NRTIs as it contains a phosphonate group attached to an adenine base, hence it requires only two further phosphorylation events into its active form as opposed to other NRTIs which require three (Fung *et al.*, 2002). This allows for much more rapid conversion of TDF into its pharmacologically active metabolite – tenofovir diphosphate.

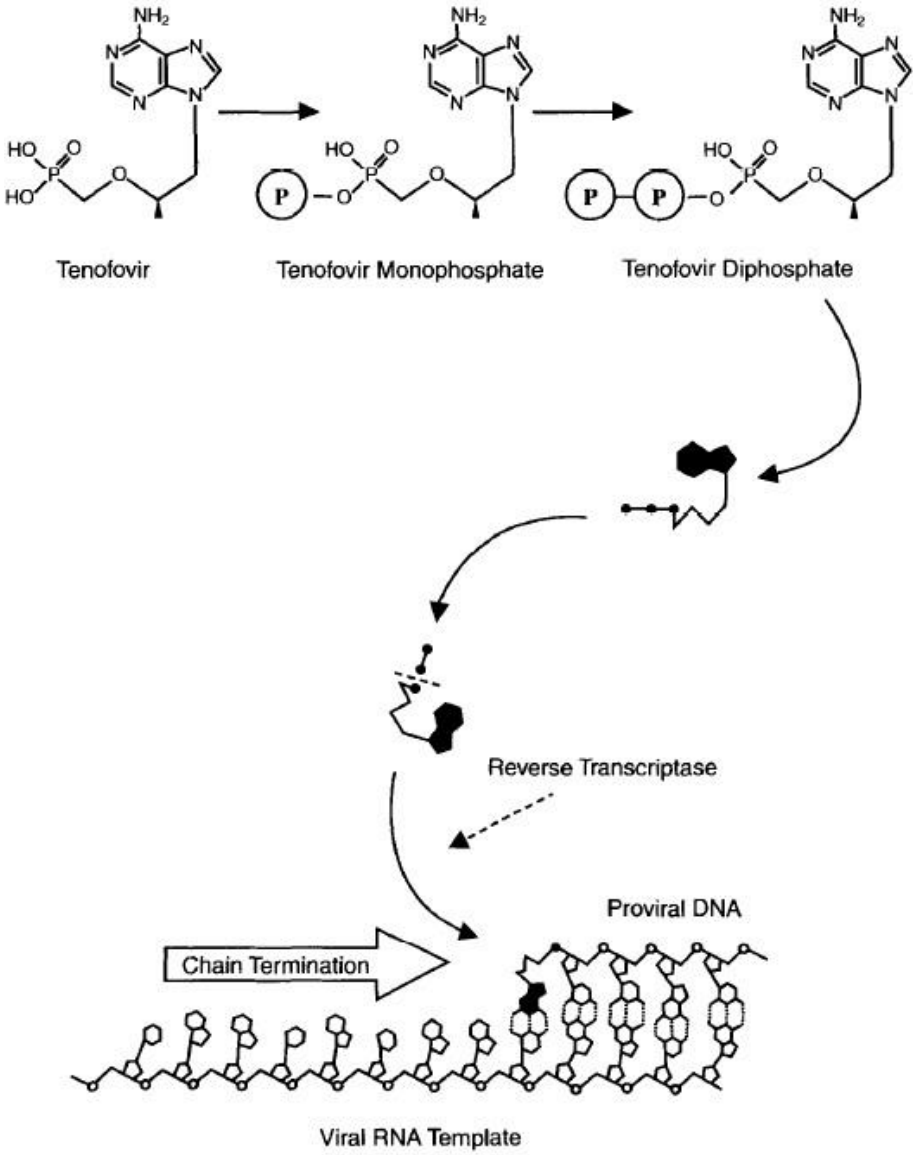


Figure 1.5: Activation and mechanism of action of tenofovir disoproxil fumarate. Tenofovir is administered in an ester prodrug and requires intracellular hydrolysation and subsequent phosphorylation to its active form tenofovir diphosphate which exerts the antiviral effect [illustration taken from (Fung *et al.*, 2002)]**.**

Despite being a highly successful combination treatment, it is known to cause kidney associated complications in approximately 8% of patients on first line ART in the Tshepong area of South Africa (Dr Ebrahim Variava, personal communication, 31 January 2014). Since the FDA approval of tenofovir, in 2001, its association with renal complications has become a subject of intense research (Likanonsakul *et al.*, 2016). A NCBI PubMed search combining "tenofovir" and "renal" generated 728 hits, as at 1 April 2018, showing increasing publications year on year (Figure 1.6).

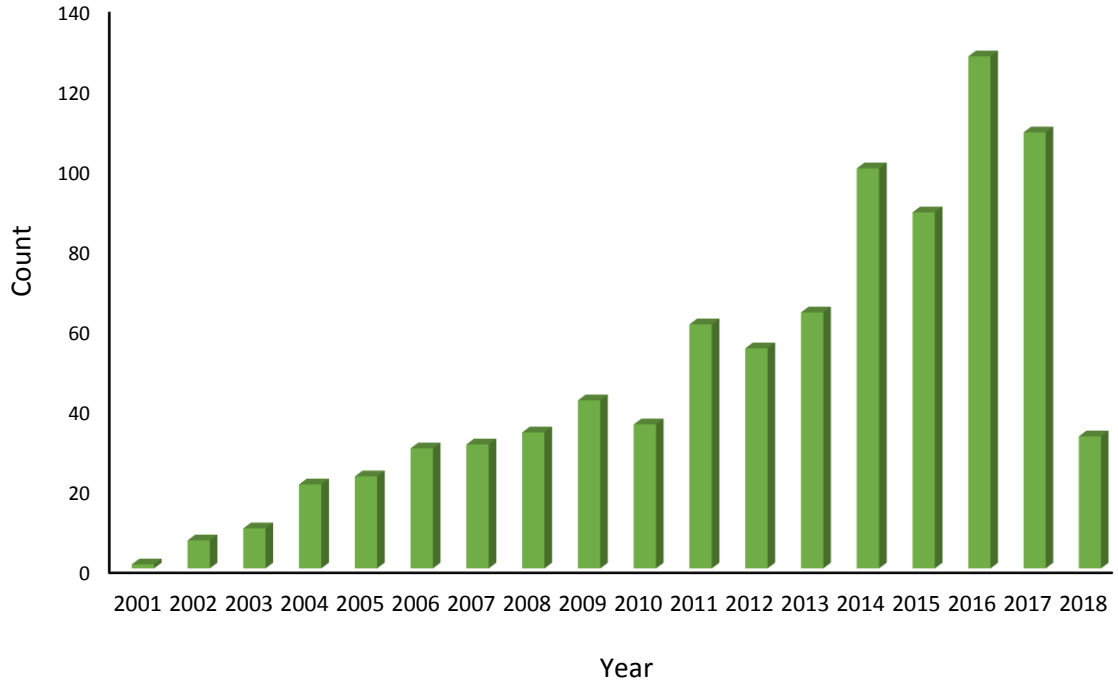


Figure 1.6: Publications (NCBI PubMed) associated with "tenofovir" and "renal" complications between January 2001 and April 2018. A steady increase in publications associated with tenofovir and renal complications has been observed since the first use of tenofovir as an antiretroviral drug.

Studies on the effects of tenofovir on kidney function in South African populations are limited (Seedat *et al.*, 2017), however strong correlation between tenofovir use and kidney dysfunction has been shown in Malaysian (Koh and Suresh, 2016), Japanese (Nishijima *et al.*, 2011), Danish (Rasch *et al.*, 2012) and Taiwanese (Huang *et al.*, 2015) patients, among others. Interestingly a long term, cross-sectional study on Ugandan patients showed no significant association between tenofovir use and kidney dysfunction over a period of 10 years (Salome *et al.*, 2016).

Unbound tenofovir diphosphate is carried via the bloodstream to the kidneys (Masho *et al.*, 2007). There it is eliminated by a combination of glomerular filtration and active tubular pumping into the proximal kidney tubules (Ray *et al.*, 2006; Estrella, Fine and Atta, 2010) (Figure 1.7). Specific uptake and efflux transporters control this process and their proper functioning is critical to this process. Tenofovir has shown a high affinity for human organic anion transporter 1 (hOAT1) and low affinity for hOAT3 (Ray *et al.*, 2006). The efflux of tenofovir into the tubular lumen is mediated by ATP-binding cassette (ABC) transporter

subfamily members multi drug resistant protein type 2 (MRP-2) and MRP-4 (Ray *et al.*, 2006; Estrella, Fine and Atta, 2010).

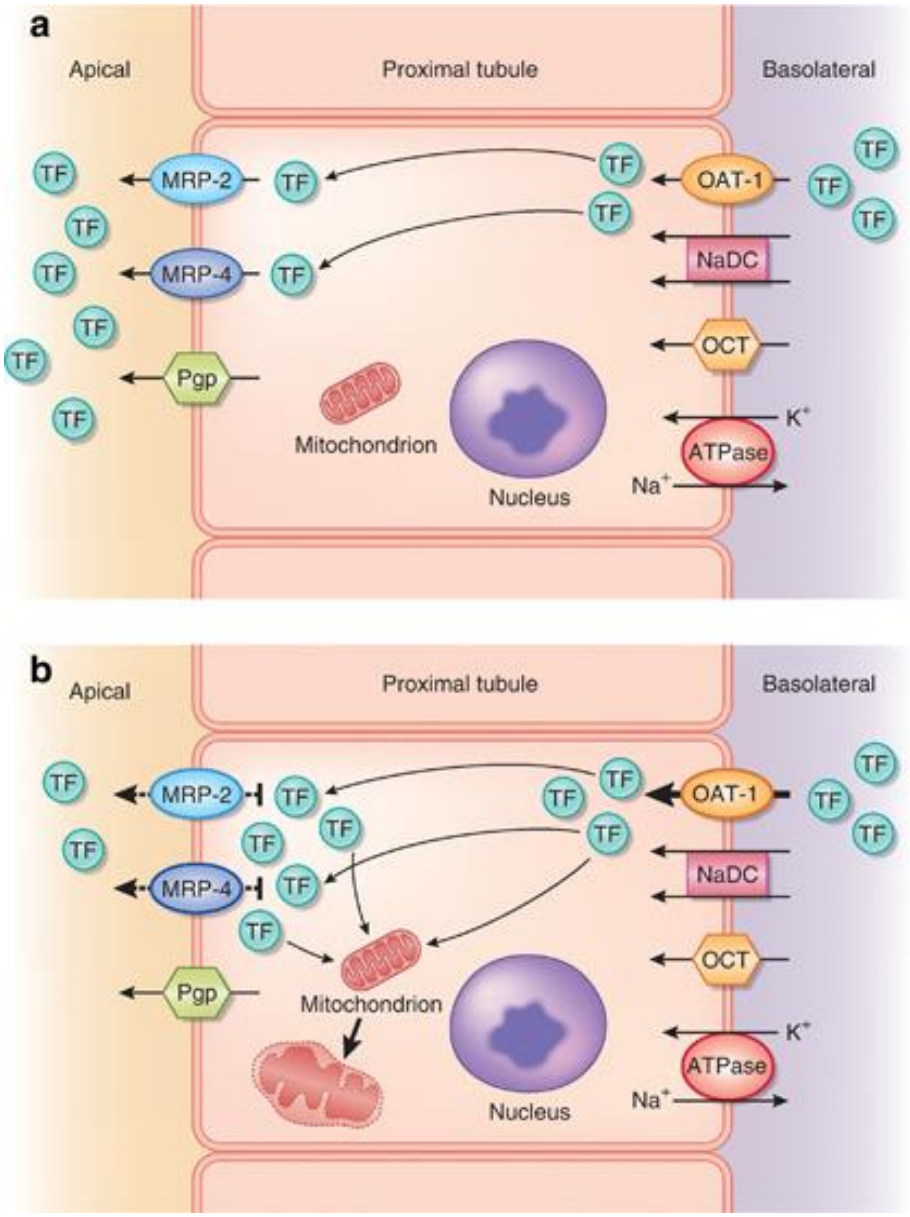


Figure 1.7: Schematic representation of a kidney proximal tubule cell. Normal tenofovir clearance and abnormal tenofovir retention (b) are shown (Perazella, 2010).

1.2.7. Tenofovir associated nephropathy and kidney function testing

The exact mechanism by which tenofovir causes nephrotoxicity is unknown, however, the interaction with its transporters and/or defective transporters due to gene and subsequently protein expression may play a key role. Recent studies have shown that when tenofovir is retained in the proximal tubules, it prevents mitochondrial DNA replication by competing with

deoxyadenosine 5′-triphosphate thereby inhibiting DNA polymerase-γ. Subsequently, these cells lose mitochondrial function and the electron transport chain is disturbed leading to oxidative stress induced cell damage (Kohler *et al.*, 2009; Lebrecht *et al.*, 2009; Perazella, 2010). Kohler *et al.* (2009) and Lebrecht *et al.* (2009) both showed that renal tubular mitochondrial toxicity resulted from tenofovir exposure in mouse and rat models, respectively. It is plausible that the same mechanism results in tenofovir induced nephropathy in humans. Furthermore, increased renal tubular retention of tenofovir could occur as energy supply required (from mitochondria in form of ATP) for active pumping of tenofovir out of the proximal tubule cell (by MRP-2/4) will be diminished.

It is likely that patients on first line ART who experience declining renal function are a minority of patients that are unable to clear tenofovir adequately. In such patients, there is a growing need to be able to identify - at a very early stage - risk factors for the development of renal complications. Accomplishing this goal will lead to increased patient adherence to ARVs, decreased patient morbidity and decreased costs arising from hospitalisation of patients with renal failure. Understanding these risk factors may also be important for improving treatment outcome with other drugs. Urinary protein biomarkers show potential as molecular markers for kidney damage and can readily be implemented as a point of care (POC) diagnostic tool. Ideally, patients exhibiting early renal injury will be changed to other ARV regimens thereby creating a personalised program based on their individual needs.

The most widely used renal function tests are based on increased serum creatinine (sCr) levels and a decline in glomerular filtration rate (GFR). Glomerular filtration rate is defined as "the volume of plasma filtered through the glomeruli per unit time" and is measured in mL/min/1.73m^2 (Traynor *et al.*, 2006). A decrease in GFR leads to a decrease in urinary excretion of creatinine leading to increased sCr. Glomerular filtration rate is generally accepted as an indicator of renal function however it is not easy to perform clinically and often generates misleading results as sCr and renal filtration are in a state of continual homeostasis even when kidney function is normal (Toffaletti, 2010). Decline in GFR coupled to increased blood urea nitrogen (BUN) and sCr are hallmarks of AKI however the rate of increase in BUN and sCr are not necessarily in parallel to decreasing GFR (Molitoris, 2012). Serum creatinine levels are also affected by age (Bhandari, 2006), diet (Preiss, Godber and Gunn, 2006), muscle mass and physical activity (Baxmann *et al.*, 2008; Molitoris, 2012). Another major criticism of increased serum creatinine levels as a marker for kidney function decline is that it manifests after significant kidney damage has occurred i.e. early changes are not seen (Toffaletti, 2010).

Decline in GFR has been associated with an increase in plasma protein cystatin *c*, however its use as a clinical biomarker is limited as its levels are also affected by age, race and gender in addition to inflammation and other chronic inflammatory states such as diabetes and HIV (Traynor *et al.*, 2006; Toffaletti, 2010). Blood urea is another indicator that is touted as a potential marker of GFR however it is not reliable. This is due to its levels being affected by high protein diets and tissue breakdown (Higgins, 2016). Inulin clearance is another method that is used to measure GFR. This is a cost and labour intensive process that involves intravenous infusion of the exogenous agent - inulin - into patients (Traynor *et al.*, 2006).

To circumvent the multiple problems associated with singular markers to measure GFR, prediction formulas have been generated and employed. The Cockcroft and Gault equation is a common method used to estimate GFR. It estimates creatinine clearance based on sCr, age, gender and weight (Cockcroft and Gault, 1976). This was one of the first prediction equations to be developed and as such was prone to have limited predictive ability. Many years later the Modification in Renal Disease (MDRD) equation for GFR was developed as it better estimated GFR by considering the effects of age, race and gender, serum urea, creatinine and albumin (Levey *et al.*, 1999). In 2009, Levey *et al.* developed yet another equation called the CKD-EPI creatinine equation that outperformed the previous equations by showing greater predictive ability than the MDRD equation.

It is important to note that these formulas are not applicable to every patient and every clinical setting (Filler *et al.*, 2005; Rule, 2010; Delanaye, Pottel and Botev, 2013). Thus, finding robust and novel ways to determine kidney function decline is still at the forefront of many research groups worldwide.

1.2.8. New nucleoside analogues on the horizon

A new prodrug of tenofovir, tenofovir alafenamide (TAF), has recently been developed as a treatment for HIV infection (Ray, Fordyce and Hitchcock, 2016). TAF shows similar efficacy to TDF however at significantly lower doses due to TAF being activated to tenofovir diphosphate intracellularly as opposed to TDF which is activated in the plasma. This results in lower circulating levels of tenofovir (Wang *et al.*, 2016). Numerous clinical studies have shown that 10 mg doses of TAF, compared to 300 mg doses of TDF, result is the same intracellular bioavailability and as much as 90% decrease in systemic (plasma) levels [reviewed in (Ray, Fordyce and Hitchcock, 2016)]. Furthermore, TAF has shown limited associations with renal complications thus far, possibly due to the 30-fold lower dosage required for viral suppression

compared to TDF (Sax *et al.*, 2015; Ray, Fordyce and Hitchcock, 2016; Wang *et al.*, 2016). Tenofovir alafenamide may be gradually implemented into existing combination drug treatments, however this process may take a long time to implement worldwide therefore continued research into novel biomarkers for TDF associated renal dysfunction is warranted. Such knowledge may also provide useful insights into the broader molecular mechanisms involved in general kidney function decline.

1.2.9. Personalised medicine in HIV management

Modern medicine is continually moving towards the ideal of treating patients based on their individual molecular needs. The "one disease-one treatment" notion is gradually fading as scientific and medical breakthroughs allow us to gather detailed patient information on multiple levels such as: genetic, epigenetic, metabolomic, proteomic and environmental factors. It is now well understood that each of these alone (and in combination) can greatly contribute to predict the outcome of a treatment regimen.

Taking these factors into consideration, whilst utilising the latest analytical and medical techniques at a molecular level, allows for the development of personalised-precision medicine. For the treatment of HIV, two well-known instances in which a "personalised" approach is used are with the drugs maraviroc and abacavir.

Abacavir is a reverse transcriptase inhibitor that is used exclusively for HIV treatment. Abacavir-hyper sensitivity reaction (ABC-HSR) is the major adverse effect of ABC treatment which occurs in approximately 5-10% of patients (Yuen, Weller and Pakes, 2008). Numerous studies have found an association of the ABC-HSR with HLA-B*5701 (Rauch *et al.*, 2008; Saag *et al.*, 2008). Current screening tests for this polymorphism are able to identify high risk patients prior to initiation of ART and effectively reduce the occurrence of ABC-HSR (Sivasubramanian, Frempong-Manso and Macarthur, 2010; Volberding and Deeks, 2010). The second successfully implemented personalised treatment programme centres on the use the of entry inhibitor maraviroc. This drug is effective against CCR-5 tropic HIV only – thus pre-screening for HIV type is necessary before a patient is put onto a maraviroc containing treatment regimen (Perry, 2010). The benefits of successfully implementing personalised medical treatment regimens are quite significant and highly effective.

For such personalised regimens, development of highly specific and targeted diagnostic and/or prognostic tools is required. This is accomplished by the identification of suitable biomarkers like those described above i.e. HLA-B*5701 and CCR-5 screening.

1.3. What are biomarkers?

By definition, a biomarker is "a characteristic that is objectively measured and evaluated as an indicator of a normal biological process, pathogenic process, or pharmacological response to a therapeutic intervention" (Atkinson *et al.*, 2001). Pathophysiological states lead to changes in gene and protein expression that are disease specific. These changes are measured clinically thus providing insight into the disease state. The use of protein biomarkers as a guide for clinicians has already been implemented in 1st world countries for various pathological states including cancer, cardiovascular disease, diabetes and HIV, among others (Phillips and Pavlos, 2011; Pavlos and Phillips, 2012).

1.4. The Proteomic Biomarker Discovery-Personalised Medicine Link

The one disease-one treatment paradigm has dominated all treatment regimens ever since the inception of modern medicine. However over the past few decades it has become increasingly apparent that a patient's response to treatment is greatly influenced by age, health status, environmental factors, genetic and epigenetic factors, among others (Vogenberg, Isaacson Barash and Pursel, 2010; Duarte and Spencer, 2016). The complete sequencing of the human genome in 2003 gave rise to a new field in which a patient's genes directed treatment – termed genomic medicine. Despite the cost of sequencing an entire human genome decreasing significantly in the past few years, the information derived from such data can only provide a portion of the information needed for complete precision medicine.

Precision (personalised) medicine has its roots in genomic medicine, however it considers the full dynamic complexity of the proteome as well. This includes protein-protein interactions, post-translational modifications and differential protein expression during diseased and healthy states.

Advances in MS technology and software have led to the generation of a draft human proteome (Kim *et al.*, 2014; Wilhelm *et al.*, 2014). This was a multicentre study in which tissues and fluids were subjected to proteomic analysis generating a set of proteins that are expressed in normal healthy tissue. This milestone represented a giant stride forward in development of the personalised proteomics field. This publicly available dataset has the potential to allow scientists and clinicians to apply these findings to patient's proteomes and thus direct treatment in a more precise, cost effective, timely and, most importantly, efficacious manner (Duarte and Spencer, 2016).

Most of the current methods to detect and quantitate biomarkers at clinical level suffer from the same shortfalls i.e. a low number of analytes and a lack of sensitivity as well as a lack of sufficient samples during the discovery phases (Duarte and Spencer, 2016; Geyer *et al.*, 2016). As more research is performed on disease states, it becomes increasingly clear that many diseases are multifactorial and, as such, the monitoring of singular biomarkers does not suffice. This is a problem that MS-based biomarker assessment can address readily due to the large number of analytes that can analysed in a short time (single analysis). Furthermore, MS based biomarker discovery has the ability to detected proteoforms (versions of a single protein based on its post-translational modifications) that are associated with a specific disease (and patients) thereby adding another layer of depth in analysis and potential source of clinical biomarkers (Smith *et al.*, 2013; Nedelkov, 2017a).

Collectively, these unique traits allow for MS based-precision medicine to be the primary tool that links a patient's proteomic profile with a specific disease or outcome. Ultimately this can lead to a more suitable treatment which can be implemented with higher success rates. The implementation of such strategies requires a holistic approach that can be translated into clinical settings i.e. being able to detect small changes in a patient's proteome due to advances in MS does not necessarily mean it will have an impact clinically. Stringent biomarker verification, validation and preclinical evaluation in proteomics studies is necessary for the dream of precision medicine to be realised.

1.5. Proteomics based biomarker discovery

1.5.1. History of proteomics

Today, the term "proteomics" is used broadly as a field of study in which proteins are studied on a large scale. The term "proteome" was coined in 1994 at the 1st Siena Meeting by Marc Wilkins (Dunn, 2014; Dunn and Kraus, 2016). It was initially proposed to describe the protein complement to any given genome however, it has since evolved into the high-throughput identification and analysis of the set of observable proteins expressed by any given organism, organ, tissue, cell or organelle at a particular time point under a defined set of conditions.

The promise of proteomics has been to generate novel, global and integrated insights into new and previously misunderstood disease states. Examining alterations in disease and healthy states would enable scientists to gain a deeper understanding of the cellular and molecular mechanisms of disease thereby generating diagnostic and prognostic markers for therapeutic intervention (Dunn, 2014).

The complexity of any given proteome has historically been the Achilles-heel to in-depth and comprehensive proteome analysis. In comparison to the static genome and semi-static transcriptome, the proteome "suffers" from a far greater complexity due to post-translational modifications (e.g. glycosylation, phosphorylation, methylation and acetylation) (Chandramouli and Qian, 2009), alternative RNA splicing and chemical modifications (e.g. oxidation) (Figure 1.8). It has been estimated that all human genes (~22 000) give rise to nearly 400 000 functional protein entities (Anderson and Anderson, 1998).

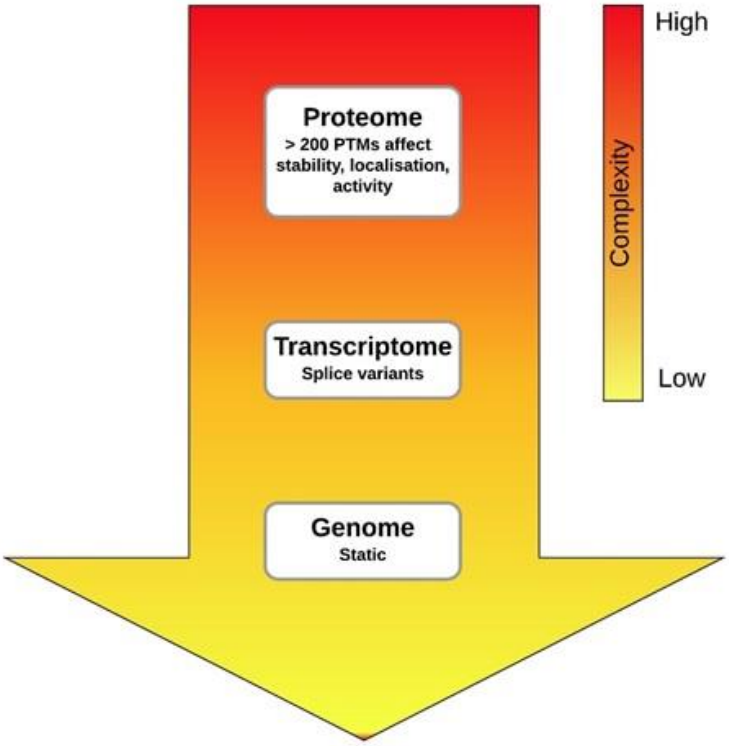


Figure 1.8: Schematic representation of the central dogma and its relationship to the proteome. With current technologies it is not possible to map an entire proteome including all PTMs in one experiment. Multiple, complementary approaches need to be employed to extract meaningful biological data from a sample of interest.

1.5.2. Urinary proteomics

Studying the human urinary proteome has fast become a method of choice in clinical proteomics studies (biomarker discovery) (Decramer *et al.*, 2008). Large sample volumes are readily available with minimal invasiveness and, importantly, urine serves as a source for a large number of diverse proteins that are relatively stable (Prunotto *et al.*, 2011; Schley *et al.*, 2015). Urine from normal, healthy individuals without any known health issues, contains many proteins and peptides that originate from the urinary tract and all other organs and tissues that are perfused with blood plasma. Due to blood being filtered at the kidneys, it is possible to identify soluble proteins and peptides, from the entire organism, in the urine thereby reflecting the health state of the individual (Decramer *et al.*, 2008; Prunotto *et al.*, 2011).

An early urinary proteome mapping experiment using MS was carried out by Spahr *et al.* (2001) where they identified 1450 peptides and 124 proteins. This study clearly showed that urine was a rich source of information that needed to be explored further. A few years on researchers were able to identify 1400 distinct protein spots using two dimensional gel electrophoresis (Pieper *et al.*, 2004) and later, 1543 proteins using a gel-based MS approach (Adachi *et al.*, 2006).

These studies laid the groundwork for the explosion of the field of urinary proteomics as they further showed that urine is an information rich, non-invasive source of (bio) markers that can be used in any health state for the benefit of the patient. A recent attempt at mapping the entire urinary proteome (urinome) was done by Santucci *et al.* (2015) where they employed numerous fractionation and sub-fractionation techniques during sample preparation followed by complementary analytical techniques. Collectively, the study generated the most comprehensive urinary proteome map containing 3 429 proteins further reinforcing the idea that urine is an excellent source for biomarkers that can reflect the systemic health state of an individual.

1.5.2.1. Urine as a biomarker source – the advantages

Urine has a few notable characteristics that make it a suitable choice for biomarker discovery and subsequent routine analysis. Firstly, the large volume of urine which can be collected with relative ease makes it easy for patient monitoring over extended time periods without adding discomfort (Kalantari *et al.*, 2015). Furthermore, the large volume affords researchers the opportunity to work with an abundance of sample thus allowing full optimisation of sample preparation protocols and methods. Secondly, soluble peptides and small proteins can be

analysed directly, without further processing (Decramer *et al.*, 2008). Thirdly, urinary proteins and peptides are generally stable when compared to plasma (Prunotto *et al.*, 2011). This is due to urine being stored in the bladder for extended periods where endogenous proteases degrade proteins therefore it is unlikely that further degradation occurs during/post voiding in contrast to plasma where protease inhibitors need to be added to limit endogenous protease activity (Decramer *et al.*, 2008). Interestingly, urinary proteins are stable for many years when stored correctly, immediately after collection, at -20 °C (Tencer *et al.*, 1997; Remer, Montenegro-Bethancourt and Shi, 2014). The urinary proteome is also less complex and has a smaller dynamic range than plasma making it easier to analyse in depth (Kalantari *et al.*, 2015). Lastly, urinary proteins have the potential to reflect the health state of the entire organism due to glomerular filtration of plasma (Decramer *et al.*, 2008; Kalantari *et al.*, 2015; Beasley-Green, 2016).

1.5.2.2. Urine as a biomarker source – the disadvantages

The most well documented disadvantage of urine as a source for protein biomarkers is the wide variation that is associated with its protein content. This is largely due to differences in diet and fluid intake. Both intra and inter patient urinary proteome variation are not easily compensated for, however recent strategies to account for this variation have been shown (Schiffer, Mischak and Novak, 2006). These include, but are not limited to, correction of protein concentration based on components of normal urine such as creatinine or collagen peptides and collecting the second morning urine to limit the variability of diet and exercise (Albalat, Mischak and Mullen, 2011). An additional downfall is that urinary proteome content is affected by the circadian rhythm, muscle metabolism and various metabolic processes. These physiological changes hamper the reproducibility of the analytical methods applied even if the method shows very low technical variation. However, these variations are limited to a small number of urinary proteins (Weissinger *et al.*, 2004) and can be compensated for by using large sample sizes. Additionally, despite the urinary proteome having a smaller dynamic range than plasma, it is still in the region of 5 – 7 orders of magnitude (Yu and Pieper, 2015; Thomas *et al.*, 2016) and poses a challenge to analyse reproducibly.

1.5.2.3. Urine in clinical proteomics applications

The rapidly advancing urinary proteomics field is moving from discovery type studies into validation and application phases. The majority of which focus on the kidney and urinary tract diseases and/or injuries due to urine expressed proteins that are mostly of urinary tract origin (70%) versus systemic origin (30%) (Thongboonkerd and Malasit, 2005) (See appendix: Table

A2). There are however multiple studies that have shown the utility of the urinary proteome in disease states emanating from regions other than the kidney.

Renal transplants suffer lowered success rates resulting from rejection of donor kidney allografts. Acute rejection occurs in many cases and many proteomic biomarker discovery studies aims to identify early markers post-surgery. Protein biomarkers associated with kidney transplant rejection have been identified in numerous independent studies (Schaub *et al.*, 2004; Wittke *et al.*, 2005; Gwinner, 2007). Despite these studies showing different marker panels that predicted renal transplant rejection, each panel was able to classify patients at risk with high sensitivity and specificity. This highlights the need for a defined and thoroughly developed standard protocol for such studies to enable reliable, reproducible and comparable results between laboratories ultimately leading to better patient management.

Diabetic nephropathy (DN) is another area in which proteomics has been applied to identify prognostic markers. The current method for DN diagnosis is based on a kidney biopsy. This is a highly invasive and costly procedure which is not suitable for every suspected case of DN. Therefore, the use of urinary proteomics is an attractive approach. Soldatos and Cooper (2008) showed a protein signature (with a high sensitivity and specificity) in type 2 diabetes patients compared to healthy controls. Urinary fetuin-A was shown to be a novel biomarker for DN in a Japanese cohort (Inoue *et al.*, 2013). Ma *et al.* (2015) identified a four-urinary protein DN-biomarker panel in streptozotocin-induced diabetic rats.

Chronic kidney disease (CKD) is closely linked to diabetes incidence and thus early detection of renal function decline is imperative to allow for better treatment and management of those patients more prone to kidney failure. This, in turn, will reduce the number of patients requiring kidney transplants. Currently, CKD is diagnosed only after significant kidney damage has occurred over a prolonged period, in part, due to a lack of biomarkers with high sensitivity and specificity. Therefore, numerous research groups have focussed their work on utilising the urinary proteome as a proxy for early detection of (chronic) kidney disease (Zimmermann *et al.*, 2006; Rampoldi *et al.*, 2011; Jang *et al.*, 2012; Musial, Bargenda and Zwolinska, 2015).

Acute kidney injury is a clinical state characterised by a rapid decrease in urine output with concomitant increase in serum creatinine levels. Despite recent advances in diagnostics and therapeutics, morbidity and mortality associated with AKI is a major problem for health institutions worldwide. It has become critical to identify new ways to detect kidney injury early thus limiting further damage. Proteomics, and in particular - urinary proteomics, has emerged

as the preferred method to address this issue. There is an increasing body of information where urinary proteomics was applied to identify novel protein markers that potentially have utility as clinical biomarkers for AKI.

Urinary proteomics has been applied to malignant diseases of the urogenital tract such as renal cell carcinoma (Rogers *et al.*, 2003), bladder cancer (Celis *et al.*, 1996; Saito *et al.*, 2005) and prostate cancer (Theodorescu *et al.*, 2005, 2008; Wu *et al.*, 2017). Interestingly, the past few years have seen an increase in the study of the urinary proteome in diseases and disease states of non-urogenital origin, such as cardiac diseases (Dwivedi *et al.*, 2016; Htun *et al.*, 2017; Röthlisberger and Pedroza-Diaz, 2017), breast cancer (Beretov *et al.*, 2014, 2015; Gajbhiye *et al.*, 2016), lung cancer (Zhang *et al.*, 2015), appendicitis (Kentsis *et al.*, 2010), rheumatoid arthritis (Kang *et al.*, 2014) and in hematopoietic stem cell transplantation (Kaiser *et al.*, 2004; Weissinger *et al.*, 2004; Weissinger, Mullen and Albalat, 2014), among others. The promising results generated from these studies as well as the overall increased study of the urinary proteome clearly illustrate that urinary proteomics has the potential to be developed into a routinely used clinical method for diagnosis, tracking prognosis and therapeutic interventions in various diseased states with minimal invasiveness.

1.6. Mass spectrometry-based proteomics

Proteomics is a technology heavy field of biological sciences which is often associated, solely, with MS. However, there are numerous ways to explore a given proteome, each with its unique advantages and disadvantages. The technique of choice is based on the type of information required i.e. the biological question. The most popular techniques include: western blotting, immunohistochemistry, enzyme assays and MS.

1.6.1. Why mass spectrometry?

Mass spectrometry has unique advantages that has led to it becoming an indispensable analytical technique in modern science. It is a highly flexible technique that can be applied to analyse entire proteomes (Magdeldin *et al.*, 2012; Kim *et al.*, 2014; Tyanova, Albrechtsen, *et al.*, 2016), protein isoforms [reviewed in (Tipton *et al.*, 2011)], protein-protein interactions (Paul, Hosp and Selbach, 2011; Smits and Vermeulen, 2016) and protein localisation (Kuckova *et al.*, 2013; Itzhak *et al.*, 2017). Furthermore, and perhaps most importantly, MS is a technique that is characterised by high sensitivity with the ability to detect small changes in protein abundance and high resolution with the ability to differentiate between closely related protein species (Aebersold and Mann, 2003).

1.6.2. Discovery vs targeted proteomics

Traditionally, proteomics is broken down into two distinct phases: discovery and verification (Figure 1.9). The discovery phase is classified as a high throughput, hypothesis free, low coverage stage in which the aim is to identify as many analytes (proteins) as possible. In this setup, the mass spectrometer selects the most abundant species for analysis resulting in lower reproducibility and dynamic range, as all peptide signals are analysed simultaneously, and lower intensity signals can be overlooked. In contrast, the verification phase, is classified as a low throughput, hypothesis driven, high coverage stage in which predefined sets of proteins are measured, in a targeted manner, with high reproducibility, dynamic range and lower limits of detection (Parker and Borchers, 2014). This is due to the mass spectrometer being set up to focus only on a defined set of peptides and peptide fragments, therefore signals from other peptides are ignored (Vidova and Spacil, 2017). Over the past few years, there has been increased usage in a new type of phase, in which discovery and targeted phases are combined, called data independent analysis (DIA) which will be discussed in detail later.

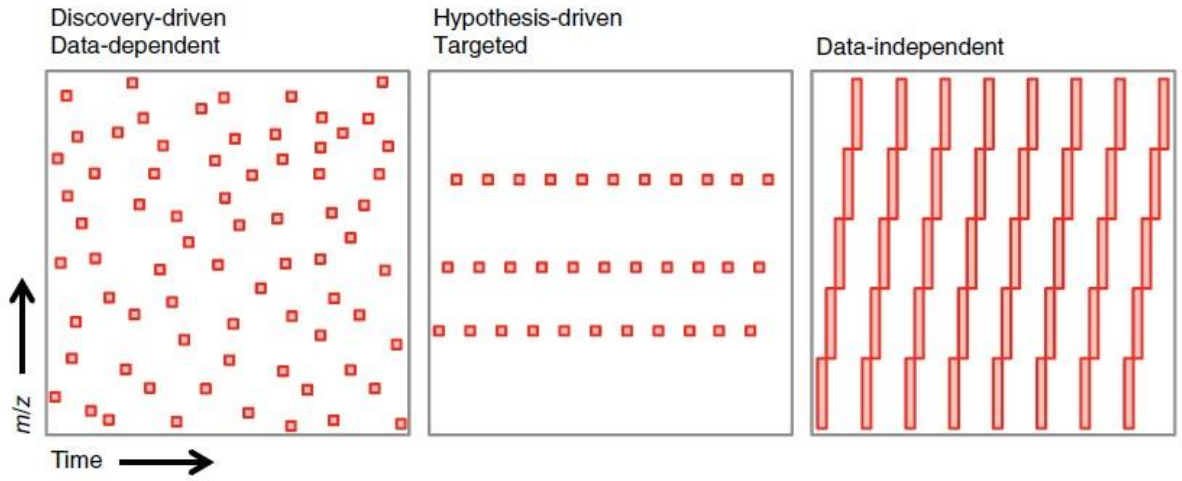


Figure 1.9: Overview of acquisition schemes for tandem mass spectrometry. The untargeted, discovery (shotgun, data-dependent acquisition) approach is easy to implement and widely used, however, it is prone to missing data points due to stochastic nature of the method. Targeted approaches require prior knowledge of the sample and the resulting data is more sensitive and reproducible, however, only a relatively small number of analytes can be analysed at one time. Data-independent approaches acquire data on all detectable precursor and fragment ions in a consistent and reproducible manner resulting in complex raw data that requires sophisticated algorithms and high-powered computing to deconvolute [adapted from (Schubert *et al.*, 2017)].

1.6.3. Top-down vs bottom-up proteomics

Proteomics experiments are further divided into bottom-up and top-down proteomics (Figure 1.10).

Top-down proteomics allows for the analysis of intact proteins, thereby decreasing sample complexity (Yates and Kelleher, 2013; Toby, Fornelli and Kelleher, 2016; Chen *et al.*, 2018). Furthermore, proteins native characteristics are preserved such as post-translational modifications, truncations and splicing events (Gregorich, Chang and Ge, 2014). Top-down proteomics suffers from the inability of liquid chromatography (LC) to efficiently separate intact proteins thus proteome-wide top-down proteomics is currently challenging (Gregorich, Chang and Ge, 2014). Furthermore, ionisation and fragmentation of intact proteins is challenging and requires ultra-high resolution mass spectrometers that are equipped with specialised fragmentation sources making this technique difficult to access and implement widely (Kelleher, 2004; Catherman, Skinner and Kelleher, 2014) .

Bottom-up proteomics is the more popular of the two techniques and is employed in ~90% of proteomics experiments (Lovric, 2011) . In bottom-up proteomics, proteins are first digested into peptides and then analysed by the mass spectrometer (Y. Zhang *et al.*, 2013; Gillet, Leitner and Aebersold, 2016). Typically, due to the proteome complexity as well as the wide dynamic range, cell/tissue lysates, plasma/serum or urinary proteins are either subjected to in-solution digestion or separated on a gel and subsequently subjected to in-gel digestion, using enzymes such as trypsin (Gregorich, Chang and Ge, 2014). The usage of peptides as the analyte in bottom-up proteomics, in contrast to intact proteins top-down proteomics, allows for easier separation by LC and, perhaps more importantly, peptides ionise more readily and fragment more easily that intact proteins (Wysocki *et al.*, 2005; Y. Zhang *et al.*, 2013). However, bottom-up proteomics increases the complexity of the input matrix since a single protein is digested into multiple peptides, thus a proteome of 1 000 proteins can easily result in greater than 10 000 tryptic peptides. Additionally, data analysis is further complicated since one peptide can match to multiple proteins or protein isoforms – this is termed "the protein inference problem" (Nesvizhskii and Aebersold, 2005). This is an important aspect considering that higher organisms can make use of protein isoforms in certain states to exert different biological functions and understanding of these isoforms can lead to deeper understanding of the biology of the organism (Yates and Kelleher, 2013). Despite the challenges with protein inference, there are many different ways to account for this [reviewed in (Li and Radivojac, 2012)].

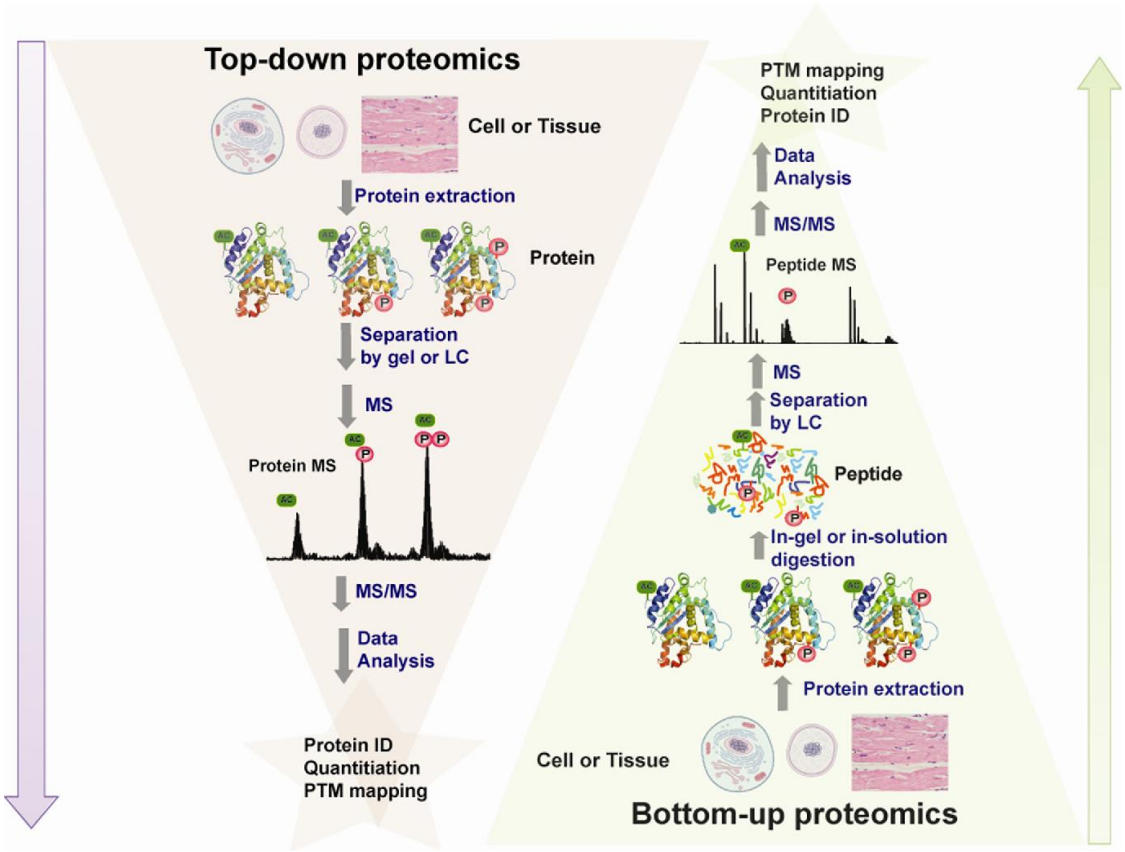


Figure 1.10: Schematic overview of top-down and bottom-up proteomics. In top-down proteomics, intact proteins are analysed directly by mass spectrometry, whereas in bottom-up proteomics, proteins are first digested into peptides and analysed by mass spectrometry (Gregorich, Chang and Ge, 2014).

1.6.4. Bottom-up proteomics for global proteome profiling

Global proteome profiling via bottom-up proteomics can be performed using either gel-free (in-solution) or gel-based approaches (Figure 1.11). In a gel-based approach, proteins are extracted and then separated on a polyacrylamide gel. Proteins may be separated using one-dimensional or two-dimensional gels.

Sodium dodecyl polyacrylamide gel electrophoresis (SDS-PAGE) is a technique that is based on the separation of proteins based on their molecular weight. Solubilised proteins are mixed with a buffer containing β-mercaptoethanol which reduces disulphide bridges and SDS which binds ratiometrically to and denatures proteins. Thus proteins are linearised and have a uniform net negative charge (Yang and Mahmood, 2012).

Protein samples are loaded into wells in a polyacrylamide gel. After a direct current electric field is applied to the apparatus, the proteins move down the gel in defined lanes creating

protein “bands”. Due to the sieve-like properties of the gel, the constituent proteins in the sample are separated according to their linear size. The entire gel is stained, and the intensity of the band is indicative of the relative abundance of the protein/s in the band. Due to co-migration of proteins this is not an absolute quantification as a single band may be made up of multiple proteins.

In an attempt to deal with this inherent problem with SDS-PAGE, two-dimensional gel electrophoresis (2DGE) was developed by O’Farrell in 1975, where he reported a method that was able to resolve up to 5000 proteins (O’Farrell, 1975). This method relies on two characteristics of a protein viz. its isoelectric point and molecular mass. Proteins are separated in two distinct stages; in the first dimension by isoelectric point via isoelectric focussing where proteins migrate to their unique isoelectric point in a pH gradient and second in an orthogonal SDS-PAGE manner (Niwa, 2008; Magdeldin *et al.*, 2014).

In the bottom-up gel-based proteomics workflow, proteins bands (1D) or spots (2D) are excised and digested using trypsin. The protein digests (peptides) are then further separated by LC before being analysed by the mass spectrometer.

Despite being relatively cheap and easy to perform, gel-based methods are time consuming, laborious, non-automatable and prone to high levels of contamination and, perhaps most importantly, is relatively irreproducible. Furthermore, gel to gel variation, especially using 2DGE, is high - making these methods error prone. Generally, gel-based approaches are difficult to apply to membrane proteins, low abundance proteins, alkaline proteins and high molecular weight proteins [reviewed in (Chevalier, 2010)]

In gel-free (in-solution) approaches, isolated proteins are digested into peptides and subjected to LC separation before being analysed by the mass spectrometer. This method is the simpler and faster method of the two (Hustoft *et al.*, 2012). It is also applicable to a wide range of samples and can be performed in a single tube thereby reducing sample contamination and losses (Feist and Hummon, 2015). The amount of starting material required is also significantly lower using an in-solution approach. Furthermore, protein concentration can be controlled in in-solution whereas protein concentration needs to be estimated using densitometry in in-gel approaches.

Whilst gel-based methods have their place in a proteomics workflow, they are gradually being replaced by in-solution (gel-free) methods due to the numerous advantages (above) and higher throughput that is often required.

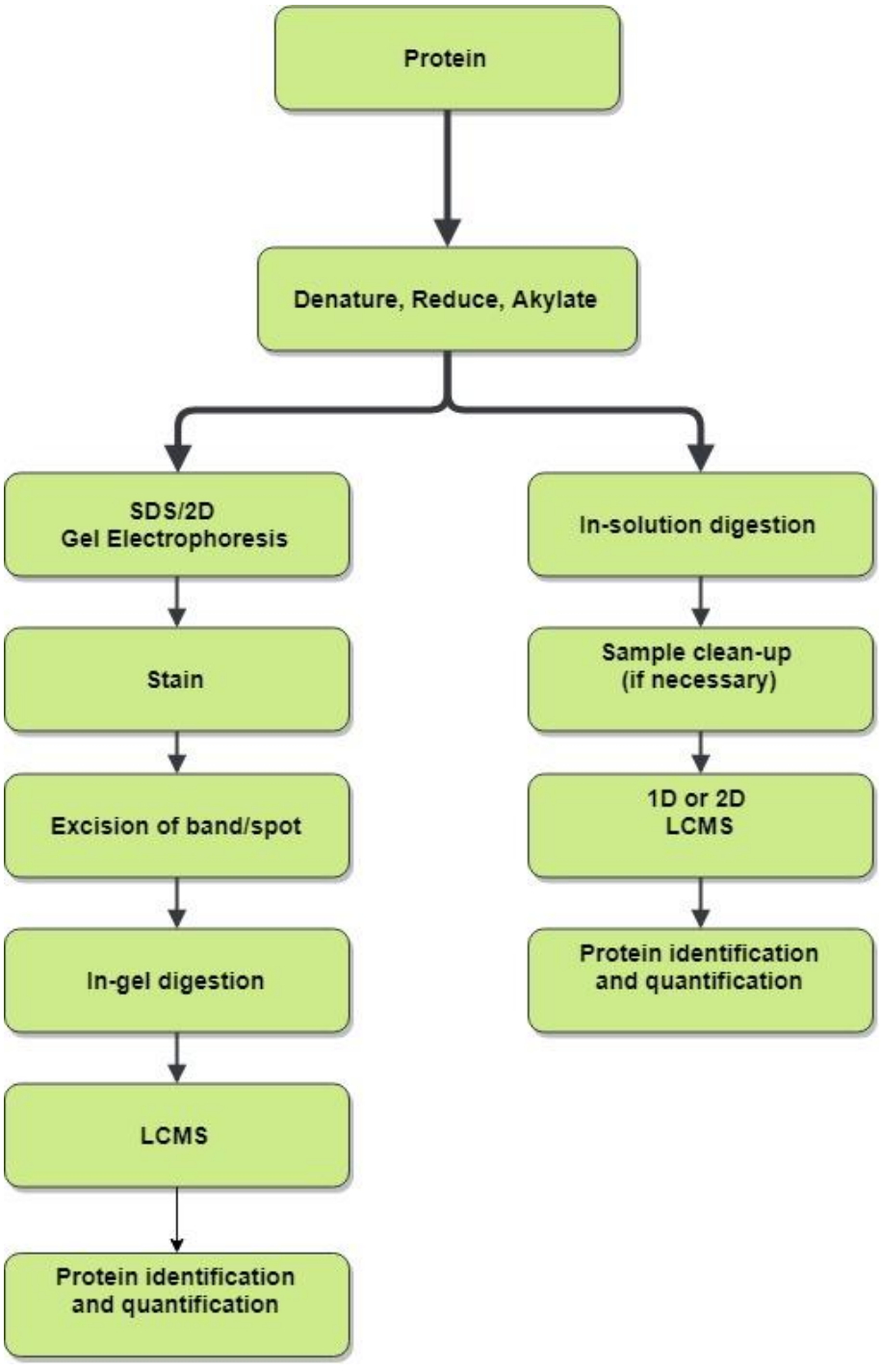


Figure 1.11: Schematic overview of in-gel (left) and in-solution (right) workflows used in bottom-up proteomics. The in-gel approach is simpler and faster to perform and introduces less variation than gel-based approaches [adapted from (Hustoft *et al.*, 2012)].

1.7. Typical mass spectrometry-based proteomics pipeline

A typical proteomics pipeline is illustrated below showing the key steps involved (Figure 1.12). Each step requires detailed planning and refinement to ensure that the experiment achieves the desired aims. The sections that follow describe the various steps and considerations that are needed for an efficient and successful proteomics experiment/study.

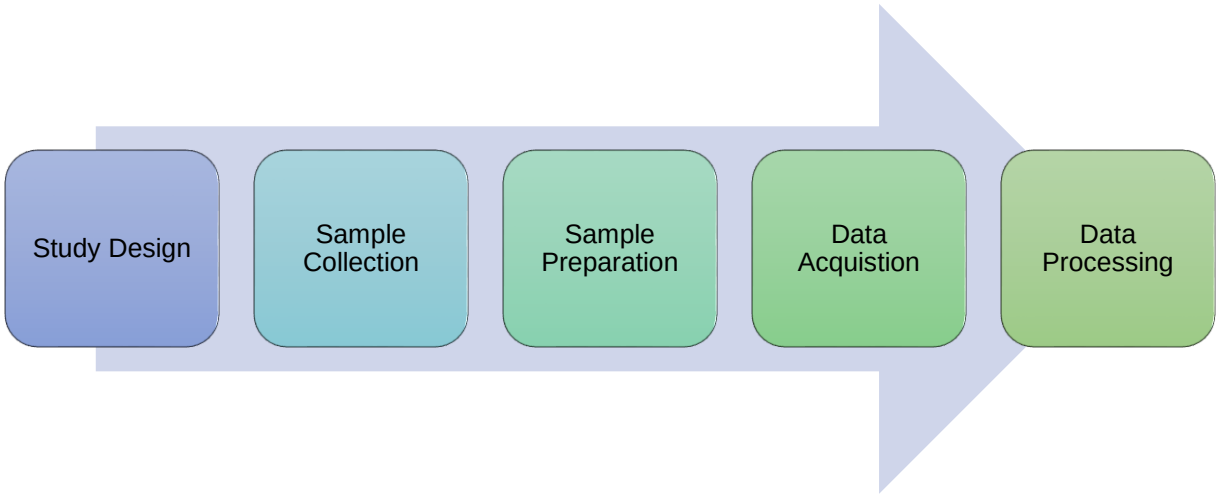


Figure 1.12: A typical mass spectrometry-based proteomics pipeline. Each step in the proteomics pipeline is critical to a successful proteomics study.

1.7.1. Study design

Study design is the most critical part of any proteomics experiment. The main components of study design include, hypothesis formulation, defining a population of interest, selection of participants that are an unbiased and accurate representation of the general population to which the conclusions apply, sample size and replicates, selecting appropriate methods to test the hypothesis and lastly, performing a pilot study to assess efficiency and variability of the chosen workflow. The goal of a study design is to ensure that bias and technical variation is minimised so that true biological changes can be determined confidently and in a statistically significant manner (Krzywinski and Altman, 2014).

1.7.1.1. Sources of variation and bias

There are many sources of variation and bias that must be accounted for during the planning of a proteomics experiment. It is critical to minimise measurements of noise so that true biological changes can be determined. Variability is inherent to all experimental workflows however, carefully controlling the experimental design by including replication, randomisation and blocking, biologically meaningful data can be generated (Altman and Krzywinski, 2015).

Replication allows the validation of data by isolating sources of variation and thus limiting their effects on the data. There are two distinct classes of replication viz. biological replicates and technical replicates.

Biological replicates are multiple measurements of distinct samples that are necessary to identify the biological changes present naturally or due to the study itself. As a rule of thumb: a) small variation allows for fewer replicates to be assessed in order to identify true differences and b) random and large variation can be accounted for by increasing the number of replicates (Oberg and Vitek, 2009). It is thus important to assess variation by conducting small-scale pilot studies to determine the levels of noise and variation in the representative population. This data can then be fed into power analysis equations to determine the sample size required to sufficiently power the study such that true biological conclusions can be drawn without analysing an unnecessarily large number of replicates (Skates *et al.*, 2013; Forshed, 2017).

Technical replicates are multiple measurements of the same sample to identify variation associated with the workflow, instrumentation and operators (Blainey, Krzywinski and Altman, 2014). Generally, the more workflow steps involved, the greater the technical variation introduced. This variation needs to be smaller the biological change being measured in order extract valid conclusions. If there is large technical variation, each step in the workflow needs to be assessed individually to determine the source of the variation and take necessary action to minimize the variation (Piehowski *et al.*, 2013; Klont *et al.*, 2018). Technical variation can also be introduced by changes in reagents, instrument calibrations and laboratory personnel (Altman and Krzywinski, 2015). However, if these are known, appropriate steps such as sample blocking can be done to limit the effect of this variation on the data.

It is common for technical variation to be small and much smaller than biological variation (Blainey, Krzywinski and Altman, 2014). Biological variation, especially in proteomics, is high due to numerous and diverse factors [reviewed in (Molloy *et al.*, 2003)]. Briefly, the dynamic nature of the proteome contributes the most to the biological variability that exists. Besides inter-individual variability, intra-individual variability greatly contributes to the biological variation observed. Protein biosynthesis and degradation occur continuously and affect the protein abundances in cells and tissues (Harper and Bennett, 2016), which subsequently affects protein concentrations in proximal fluids such as plasma and urine.

Intra-individual variation can result from changes in the proteome due to diurnal/circadian rhythm (Mauvoisin *et al.*, 2015), dietary changes (Walsh *et al.*, 2006; Nagarajan *et al.*, 2017),

fluid intake and exercise (particularly on urinary proteome) (Hussey *et al.*, 2013; Kohler, Schänzer and Thevis, 2015), to name a few. Proteome variations as a result of gender can (and should) be controlled by the experimenter (Altman and Krzywinski, 2015). It is critical to take these variations into account in designing a study to minimise their effects on the biological conclusions. Biological replicates are more valuable than technical replicates since they allow for greater inference biology of the biological population of interest (Blainey, Krzywinski and Altman, 2014).

In clinical proteomics studies, biological replicates (within financial reason) are the most important type of replication that needs to be taken into account in order to extract the most meaningful biological data from the study (Oberg and Vitek, 2009; Blainey, Krzywinski and Altman, 2014). The equation below shows a theoretical calculation for biological variation:

$$\boldsymbol{Variation}_{\boldsymbol{(Disease-Healthy)}} = \mathbf{2}\left(\frac{\sigma^2 \boldsymbol{Individual}}{\boldsymbol{I}} + \frac{\sigma^2 \boldsymbol{Sample\ Preparation}}{\boldsymbol{IJ}} + \frac{\sigma^2 \boldsymbol{Measurement\ Error}}{\boldsymbol{IJK}}\right)$$

Variation in signal is due to: biological variation (*I*), sample preparation (*J*) and instrument measurements (*K*). By increasing the number of instrument measurements (*K*), the total measurement error (red term) will decrease however it does not affect the variation introduced by sample preparation (blue term) and biological variation (green term). Similarly, performing multiple sample preparations does not decrease the total variation. However, by increasing the number of biological replicates, the total variation can be decreased significantly leading to better statistical inference (Oberg and Vitek, 2009)

The second aspect of experimental design deals with minimising systematic biases and this is done by randomising and blocking. Collectively, these two techniques can minimize experimental artefacts such as instrumental drift (Gibbons *et al.*, 2015), changes in reagent batches, changes in column batches etc. Randomising during sample selection, preparation (biological randomisation) and data acquisition (at MS level – technical randomisation) ensures that systematic biases are averaged across all the samples (Oberg and Vitek, 2009; Karp, 2010).

Blocking on a noise variable or known source of variability (such as instrumental drift in MS), allows for the data to be compared across blocks without the effect/s of the variable on the data (Altman and Krzywinski, 2015). Blocking imposes a restriction on the randomisation such that it allows randomisation whilst taking into account the known sources of variation (Oberg and Vitek, 2009). Samples analysed in a single block should be processed individually and not compared across blocks as this will introduce block-to-block variation which is difficult to

account for. Biological changes can then be assessed, with greater confidence, on average across all samples following block data processing and analysis (Oberg and Vitek, 2009; Qin *et al.*, 2014).

1.7.2. Sample collection and sample preparation

1.7.2.1. Sample collection

Sample collection needs to be done in a standardised and reproducible manner to ensure that variation is kept low. This can be achieved by setting up detailed standard operating protocols for the collection and storage of samples, particularly in clinical proteomics studies where samples are often not collected at the MS laboratory level. Samples are generally collected at point-of-care by nurses and clinicians and therefore it is critical to train these personnel on the way samples should be collected and stored. Additionally, sample collection should be done at the same time during the day to minimise the effect of the diurnal/circadian rhythm on the proteome. Finally, the collection and processing procedure should be simple to implement and perform and have minimal instrument use such that the protocol can be reproducible.

1.7.2.2. Sample preparation

Sample preparation can be broken down into 3 major steps viz. protein extraction and digestion, sample clean-up and decreasing sample complexity.

Protein extraction is the process that aims to extract the greatest number of proteins in a soluble form with minimal degradation and modifications. Extracted proteins are often soluble in buffered, detergent-based solutions at neutral pH, however this is not compatible with downstream MS and needs to be removed prior to analysis. This is due to most peptides only ionising under acidic conditions. Additionally, tryptic digestion may also be affected by certain buffers and needs to be considered and accounted for when designing an experiment.

Given these constraints, extracted proteins need to be subjected to some form of clean-up to remove interfering substances and buffer exchanged so that the resulting sample is more compatible with downstream analysis. This is a step that is necessary to avoid clogging LC flow paths and contamination of the mass spectrometer, however it is prone to sample losses and increased variability (Piehowski *et al.*, 2013). Methods to achieve this include, precipitation, solid phase extraction (SPE), and ultracentrifugation, to name a few.

One of the most common methods for sample clean-up that is almost universally applicable is based on precipitation. Precipitation can be varied based on the nature of the contaminants

and/or detergents that need to be removed, the nature of the proteins the sample, and downstream analysis methodologies. Of all precipitation methods, acetone precipitation is the most widely used technique and is highly efficient for water-soluble proteins (Thongboonkerd *et al.*, 2002; Feist and Hummon, 2015; Santa, Anjo and Manadas, 2016). Acetone precipitation is relatively simple to perform and doesn't require any specialised equipment. The general process involves adding ice-cold (-20 °C) acetone in excess volume to a sample containing solubilised proteins. This then effectively precipitates the proteins out of solution and can be collected by centrifugation. There are many other forms of organic solvent-based precipitation including, acetonitrile (Kay *et al.*, 2008), methanol (Afkarian *et al.*, 2010), trichloroacetic acid (Beretov *et al.*, 2014), and methanol-chloroform (Jiang, He and Fountoulakis, 2004; Fic *et al.*, 2010), among others. Despite precipitation being widely used and easy to perform, protein pellets collected after centrifugation are difficult to resuspend. However numerous methods have been used to aid in this process, including: vigorous vortex mixing and sonication [reviewed in (Feist and Hummon, 2015)].

The most popular method for SPE, filter-aided sample preparation (FASP), was introduced in 2009 by Wisniewski *et al.* This protocol incorporates sample clean-up and digestion. It makes use of a high-molecular weight cut-off filter which retains proteins and allows salts and smaller molecular weight compounds through the filter. These are then discarded, and the remaining proteins are subjected to alkylation and digested on the filter. The filter serves as reaction medium for the digestion process and once complete, peptides can be washed off the filter and collected. The single tube nature of the protocol means it is easy perform however it has many steps and is a tedious process to complete for many samples. Furthermore, sample losses have been well documented for this protocol (Erde, Loo and Loo, 2014). Despite this, FASP remains a widely used protocol that is amenable to a wide range of samples and workflows [reviewed in (Feist and Hummon, 2015)].

Lastly, complex peptide mixtures, derived from enzymatic digestion of proteins, needs to be simplified at the protein or peptide level typically using LC based separation (explained in detail below).

1.7.2.2.1. History of liquid chromatography

The first research published on LC appeared in 1905 by a Russian scientist, M. S. Tswett, where he described separation of plant pigments by column chromatography (Ettre and Sakodynskii, 1993; Jandera *et al.*, 2011). It was only 55 years later that Porath and Flodin

developed size exclusion chromatography (Porath and Flodin, 1987), followed by the development of a method to determine molecular mass distribution of synthetic polymers by Waters (Milford, Massachusetts, USA). These two key developments utilised a pump delivering a solvent at high pressure through a column containing small packed particles whilst monitoring the column effluent at a constant rate using a detector (Jandera *et al.*, 2011). This marked the birth of modern high pressure (performance) liquid chromatography (HPLC) as we know it today. There are 5 major components in a HPLC system and each has a distinct function (described below) to enable to accurate and reproducible separation of biomolecules (Figure 1.13).

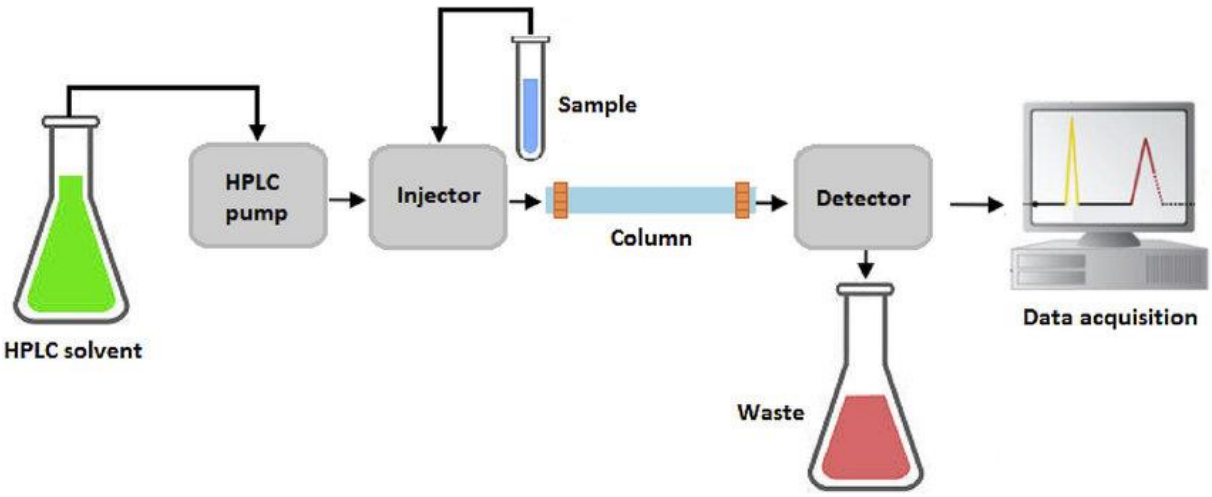


Figure 1.13: Schematic overview of HPLC instrumentation. A typical liquid chromatography setup involves the use of a pump, injector, separation column, detector and a data acquisition computer (Afsarimanesh, Mukhopadhyay and Kruger, 2018).

Pump: the purpose of the pump is to force the mobile phase through the entire system at a constant flow rate. Flow rate stability is critical to the reproducible separation of analyte molecules and should ideally be less than 1%. Modern HPLC systems can be equipped with variable flow selector modules which enable differing flowrates based on the application.

Injector: the injector is responsible for the introduction of the sample into the mobile phase at high pressure which in turn carries it into the chromatographic bed. Sample volumes can range between 0.1 - 20 µL.

Column: the most critical component of HPLC that enables the separation of the molecules of interest is the chromatographic column. This is a tube that contains the stationary phase and

comes in various sizes and chromatographic chemistries based on its application. Columns used for proteomics are often packed with small diameter e.g. 2 µm, (semi)porous, silica gel particles. The particles are coupled to chemically specific phases which are directly responsible for the separation of molecules of interest. The most popular type of column chemistry phase used in proteomics is C18. Columns are built with frits which allow the mobile phase to pass through whilst retaining the packed particles.

Detector: the detector is responsible for analysing, visualising and digitally recording the analytes that are separated by the column. This is done using ultraviolet light which passes through the sample which generate absorption data which is specific to and characteristic of the molecule of interest. When LC is coupled to MS, the mass spectrometer functions as the detector.

Data system: the data system is responsible for controlling the entire HPLC process electronically. Furthermore, it digitally records all information being sent to and received from the HPLC system. The data derived from the system provides information on the retention time and qualitative and quantitative characteristics of the molecule of interest.

In a perfect LC experimental setup, there exists a delicate balance between the chromatographic stationary phase and the mobile phase carrying samples at any given time and point along the chromatographic column (Jandera *et al.*, 2011). To simplify, as the mobile phase moves through the column, the equilibrium is in a state of constant flux as new mobile phase meets the stationary phase which temporarily adsorbs the sample of interest. The sample will remain adsorbed to the stationary phase until such time that it is more soluble in the changing mobile phase – at this point the sample is carried by the mobile phase out of the column. In this manner, molecules are separated based on their unique chemical properties, which confers different affinities to the stationary phase, and are further analysed by the detector which, in most cases, is a mass spectrometer.

1.7.2.2.2. Effect of flow rate on liquid chromatography

The flow rate of the mobile phase in LC needs to be adjusted according to the internal diameter (ID) size of the column in the system. It has been well documented that decreasing the column ID – and thus flow rate – results in decreased sample and solvent requirement whilst increasing sensitivity (Needham, 2017). Column IDs of 2.1 mm and corresponding flow rates between 250 – 600 µL/min is called standard flow LC (González Fernández-Niño *et al.*, 2015) and conversely column IDs of < 100 µm and corresponding flow rates of < 1 µL/min is called

nanoflow (Vissers, Claessens and Cramers, 1997). The benefit of standard flow is that it is robust however it requires large volumes of solvents and more importantly requires large quantities of sample. Nanoflow benefits from increased sensitivity and significantly lower solvent and sample consumption at the expense of being less robust and requiring constant specialised maintenance (Moreno-González *et al.*, 2017). Microflow LC is a method that uses flow rates between 5 – 100 µL/min and columns with IDs between 0.2 – 1 mm (Christianson, Johnson and Needham, 2013; Needham, Williams and Bell, 2016). This method boasts the benefits of nanoflow LC in terms of increased MS signal (compared to standard flow) with a slightly decreased sensitivity (compared to nanoflow) whilst maintain the robust nature of standard flow LC without the need for excessive amounts of solvents (Christianson, Johnson and Needham, 2013; Needham, Williams and Bell, 2016; Needham, 2017).

1.7.2.2.3. Chromatographic chemistries

A large and diverse range of samples can be analysed using HPLC. These include, but are not limited to, samples ranging from strongly polar to non-polar and can be synthetic polymers or highly complex biological compounds (Westermeier, Naven and Höpker, 2008). The nature of the molecule of interest dictates the type of chromatographic chemistry used in the column and the composition of the mobile phase.

Some of the commonly encountered chemistries include, normal phase chromatography (NPC), hydrophilic interaction liquid chromatography (HILIC), ion-exchange chromatography (IEX) and size exclusion chromatography (SEC), to name a few. However, the most commonly used chromatographic chemistry, used in proteomics experiments, is reverse-phase chromatography (RPC) (Mitulovic and Mechtler, 2006).

In RPC, the stationary phase (column) is non-polar or hydrophobic and the mobile phase is a polar mixture of water and organic solvent (acetonitrile) or hydrophilic. The non-polar stationary phase is hydrophobic and made up of alkyl hydrocarbon chains, of differing length (C4, C8 or C18), bonded to silica particles (Dorsey and Dill, 1989) (Figure 1.14). Reverse-phase chromatography in which the hydrocarbon chain is made up of C18 or octadecylsilane is most commonly used and is the gold-standard for bottom-up proteomics analyses as it can separate peptides with high accuracy and precision (Kondeti, Mulpuri and Meruga, 2014).

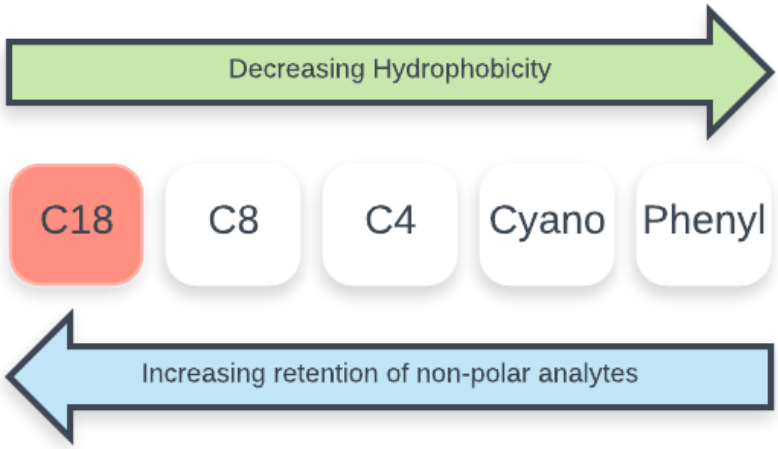


Figure 1.14: Common reverse phase HPLC stationary phases. Arranged in increasing retention of non-polar analytes and decreasing hydrophobicity. C18, the most commonly used chemistry for proteomics research, is shown in red.

In proteomics research, trypsin digested peptides are analysed using nano- and micro-columns. Peptides adsorb to the non-polar C18 material and allow polar molecules to pass through with limited interaction with the stationary phase. As the mobile phase is gradually changed to become more non-polar, peptides begin to elute from the column with increasing hydrophobicity i.e. hydrophilic peptides elute first, followed by hydrophobic peptides. In RPC the experiment can be tightly regulated by controlling variables such as pH, organic solvent type and concentration and addition of modifiers.

It has become increasingly common to perform two-dimensional column chromatography as this allows for increased analytical data to be extracted from the sample. This is achieved by using two orthogonal forms of column chemistry whereby peptides are either enriched, pre-fractionated, trapped and/or desalted on the first column (1^{st} dimension) and then separated using RPC in the second column (2^{nd} dimension) (Issaq *et al.*, 2005).

A common 2D technique used in proteomics employs RPC in both the first and second dimensions i.e. RP-RP chromatography where peptides are separate based on their hydrophobicity in both dimensions. This allows for greater resolution of the sample of interest when analysed by MS (Dowell, Vander Heyden and Li, 2006; McQueen and Krokhin, 2012).

In the 1^{st} dimension, peptides are separated at an alkaline pH and in the 2^{nd} dimension peptides are eluted at acidic pH. This methodology cleverly exploits the change in amino acid charge at

different pH. Basic amino acids are neutral and acidic amino acids are negative at alkaline pH in the 1st dimension, and in the 2nd dimension basic amino acids are positively charged, and acidic amino acids are neutral. Whilst this process is time consuming and uses large amounts of sample, the net simplification of the sample proteome of interest allows for greater depth of analytical information to be extracted (Fournier *et al.*, 2007)

1.7.3. Data acquisition

1.7.3.1. What is a mass spectrometer?

Mass spectrometry has become the most widely used and indispensable analytical technique in the biological, life and health sciences (Aebersold and Mann, 2003; Finehout and Lee, 2004; Han *et al.*, 2008; Yates, Ruse and Nakorchevsky, 2009; Parker and Borchers, 2014). The first mass spectrometer was developed in the early 1900's and has since progressed from analysing small inorganic molecules to large biological macromolecules (El-Aneed, Cohen and Banoub, 2009). In any mass spectrometer, analysis of proteins/peptides occurs stepwise in three steps: a) generation of gas-phase ions, b) separation of ions based on mass-to-charge ratio (m/z) and c) quantitative detection of separated ions (Chandramouli and Qian, 2009)

1.7.3.2. Ionisation techniques

The explosion of MS-based proteomics began after the invention of soft ionising techniques: electrospray ionisation (ESI) and matrix assisted laser desorption ionisation (MALDI) both of which led to their developers being awarded with Nobel prizes (reviewed in Aebersold and Mann, 2003). MALDI uses high powered laser pulses to ionise an analyte that is mixed with a carrier matrix. The analyte is then carried in a vacuum to a detector where its mass-to-charge ratio (m/z) is recorded. Electrospray ionisation ionises analytes by applying a high voltage to a narrow emitter carrying analytes in solution with low stability without causing significant decomposition of the ion (Garbis, Lubec and Fountoulakis, 2005). Whilst both methods create gas-phase peptides, a requirement for MS analysis, ESI is readily coupled to LC systems and is thus the most commonly used method for ionisation in proteomics experiments (Wilm, 2011) (Figure 1.15).

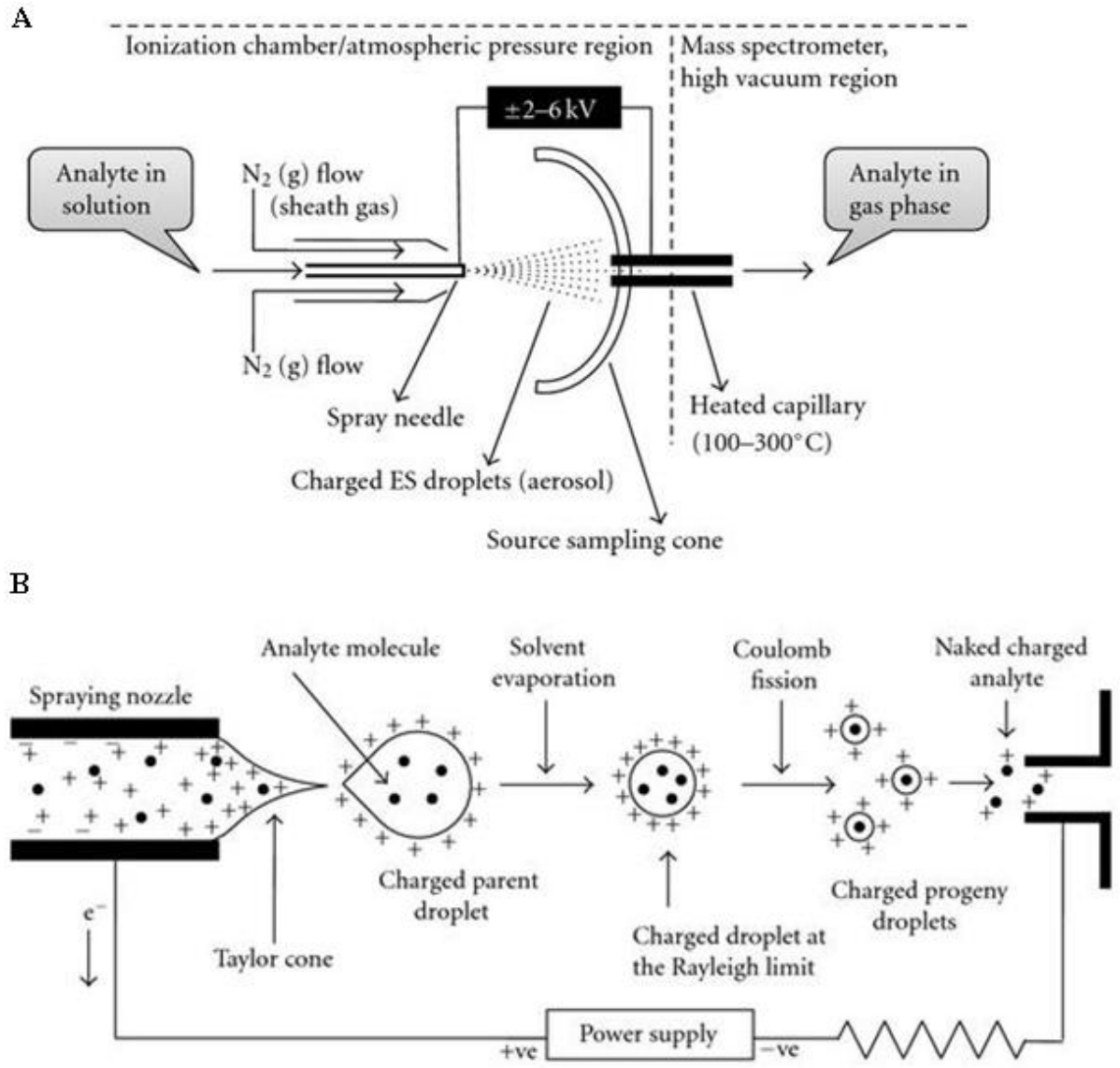


Figure 1.15: Schematic overview of the electrospray ionisation A) source and B) process. Solute surrounded analytes are passed through the electrospray ion source held at high electric potential (2-6 kV). The solute is converted into charged electrospray droplets that gradually decrease in size aided by a drying gas. Once the droplet can no longer undergo desolvation, at the Rayleigh limit, the droplet bursts (due to Coulombic repulsion) releasing charged analytes into the mass spectrometer [adapted from (Banerjee and Mazumdar, 2012)].

1.7.3.3. Types of mass spectrometers

There are many different types and forms of mass spectrometers that have their unique strengths and weaknesses. Vendors continuously push the boundaries of MS by improving on current technologies to allow greater resolution, sensitivity and speed of mass spectrometers. Mass spectrometers are made up of 3 main components: a source, a mass analyser and a detector. The source generates ions, the mass analyser resolves the ions according to their

unique m/z and the detector digitally records the m/z and intensity data for each ion (Haag, 2016). A mass spectrum is generated by the detector after converting the electrical signal received from an ion into a visual signal. The mass spectrum is a plot of the relative abundance (intensity) of the ion against its m/z (El-Aneed, Cohen and Banoub, 2009; Haag, 2016) (Figure 1.16).

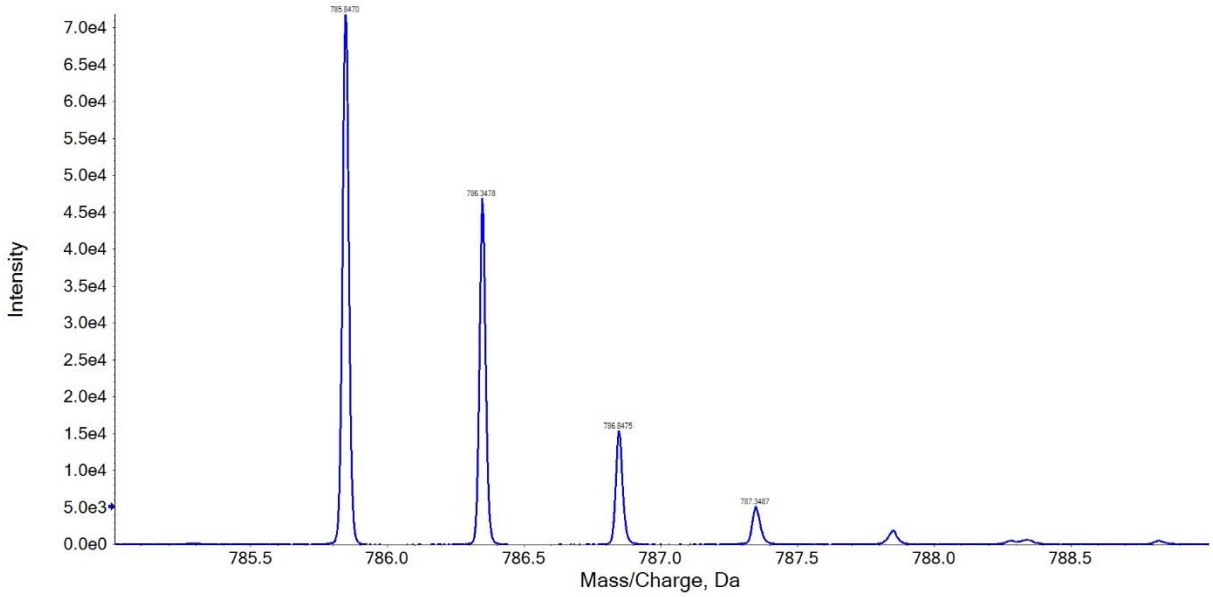


Figure 1.16: A typical mass spectrum. Derived from a [Glu[1]]-Fibrinopeptide B standard sample. Intensity of the parent ion is shown on the y-axis and m/z is shown on the x-axis.

The mass analyser is the most critical component in a mass spectrometer that is available in many different configurations as well as hybrid configurations. The major analysers include, quadrupole, Time-of-Flight (TOF), ion trap and Fourier transform analysers and an array of hybrid analysers that contain at least 2 of the aforementioned – known as tandem mass spectrometers (Haag, 2016). The choice of analyser is dependent on, but not limited to: a) m/z range, b) mass of the analyte, c) resolving power of the analyser and d) limit of detection (Glish and Vachet, 2003; El-Aneed, Cohen and Banoub, 2009; Haag, 2016). Despite there being many combinations and configurations of mass analysers available, there is no single, "ideal" analyser that can perform all functions adequately. The quadrupole and TOF analysers will be discussed below as these technologies were employed, as components of SCIEX (Framingham, Massachusetts, USA) TripleTOF® 6600, used in this study.

1.7.3.3.1. Quadrupole mass analysers

The quadrupole analyser was first described in the 1950s (El-Aneed, Cohen and Banoub, 2009) and is one of the most popular mass analysers used today due to relatively low cost, small design and robustness. They are most often found in combination with each other as a triple quadrupole (QqQ) or in tandem with a TOF (qTOF) (Haag, 2016). Quadrupole mass analysers are highly efficient mass filters that are able to accurately and reproducibly filter ions of different m/z.

This analyser is made up of four parallel rods to which direct current (DC) (2 rods) and radio-frequency (RF) potential (2-rods) are applied to opposite rods (Dawson and York, 1976). This combination of DC and RF cause ions passing through them to oscillate and only ions of a particular m/z will have stable trajectories. Unstable ions will be filtered out as they cannot be controlled and will likely collide with the surroundings. The tight control of the DC and RF potentials allow for ions of different m/z to be filtered for analysis whilst excluding those that are not of interest (Figure 1.17). The major advantage of quadrupole mass analysers is their low cost and compact size. This makes them ideal for small laboratories and they can be used on the benchtop.

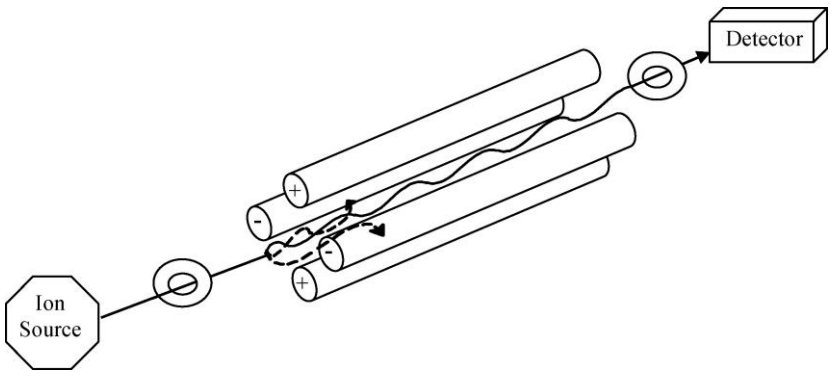


Figure 1.17: A schematic representation of a quadrupole mass analyser. This analyser is made up of four parallel rods to which direct current (DC) (2 rods) and radio-frequency (RF) potential (2-rods) are applied to opposite rods allowing filtering for specific m/z ions [Taken from (El-Aneed, Cohen and Banoub, 2009)**].**

In QqQ mass spectrometers, three quadrupoles are arranged in series. The first analyser is called Q1 and is responsible for scanning and allowing through a defined range of m/z or selected m/z ions (parent or precursor ions). These then pass into the second quadrupole, called

Q2, which acts as a collision cell and operates in RF mode. Here the precursor ions can be fragmented, in a process called collision-induced dissociation (CID), into fragment or product ions. These product ions are then passed into the last quadrupole called Q3 where they can all be detected or Q3 can be configured to detect a particular fragment ion of interest (Figure 1.18). The major types of scans that can be done using this approach include precursor ion scan, product ion scan and selected/multiple reaction monitoring (S/MRM).

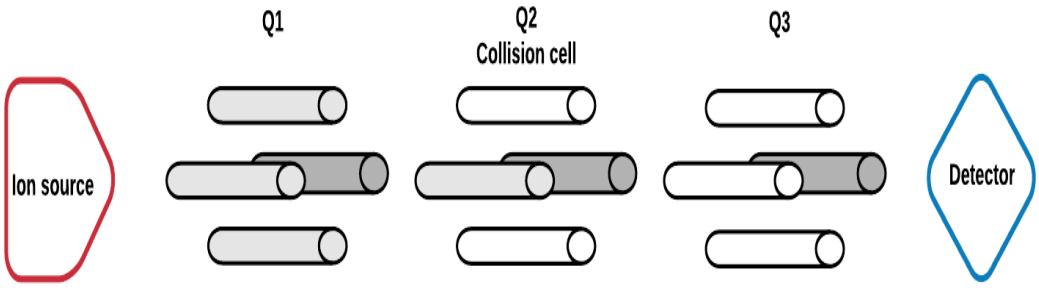


Figure 1.18: An overview of a triple quadrupole mass spectrometer showing three distinct quadrupoles. Q1 acts as a filter for parent ions, Q2 is responsible for CID and Q3 filters product ions.

In product ion scanning, Q1 is fixed on a set m/z allowing only the selected ion to pass into Q2 where is undergoes CID. The fragment ions are then analysed in Q3. The process is then repeated with Q1 fixing to another m/z. In precursor ion scanning, Q1 allows a defined range of m/z ions into Q2. There they all undergo CID and in Q3 only selected product ions are monitored. The chromatogram generated provides information on the intensity of the ions in Q3 that originated from the precursor ion in Q1. Lastly, in selected reaction monitoring, Q1 only allows a predefined m/z ion into the next quadrupole where it is fragmented. All fragments are sent to Q3 and Q3 further allows only a predefined product ion into the detector region. This combination of Q1 and Q3 fixed scanning is for a unique precursor and the selected product ion is called a transition (March and Todd, 2016; Vidova and Spacil, 2017) (Figure 1.19).

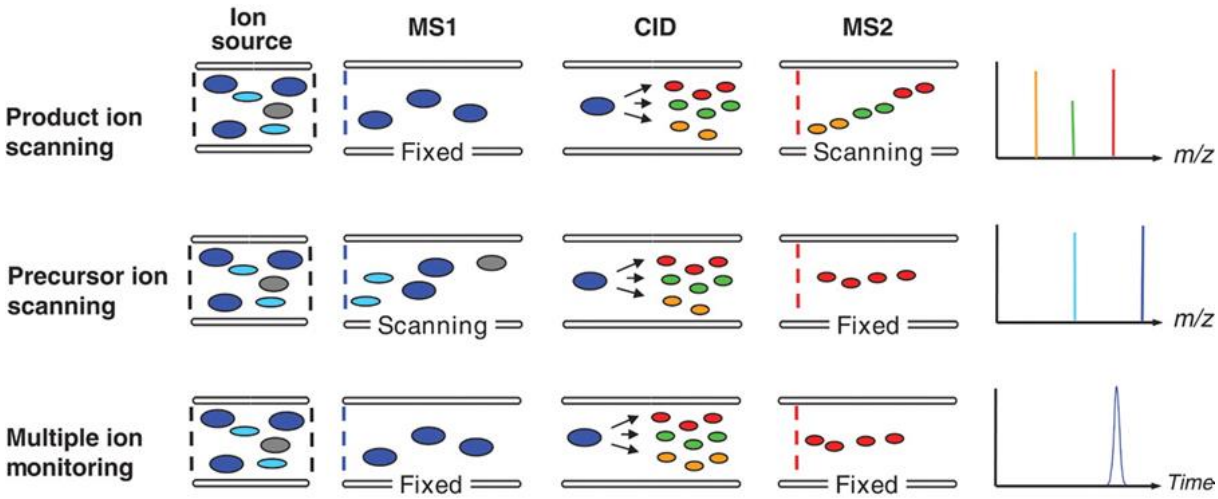


Figure 1.19: Schematic overview of different tandem mass spectrometry scans performed on a triple quadrupole mass spectrometer. Product ion scanning, precursor ion scanning and multiple reaction monitoring are shown [adapted from (Domon and Aebersold, 2006)**]**

1.7.3.3.2. Time-of-flight mass analyser

Time-of-flight analysers consist of a flight tube under vacuum and an acceleration grid that serves to move ions from the source and the detector (Aebersold and Mann, 2003). In simple terms, "if two ions of different m/z are accelerated from the ion source with the same kinetic energy and allowed to drift through the field-free region of the flight tube, their arrival time at the detector will be different" (Haag, 2016). The principle of this technology is based on the following equation:

$$\boldsymbol{m = \frac{2E_k t^2}{d^2}}$$

Since the initial kinetic energy (E_k) of the ions and the flight tube length (d) remains constant, mass (m) is directly proportional to time (t) (El-Aneed, Cohen and Banoub, 2009; Haag, 2016). Modern TOF analysers employ an ion mirror (reflector) to help improve resolution. This is achieved as ions with higher velocity and small size will penetrate deeper into the reflector than identical m/z ions with lower velocity – thus even though the m/z values are the same, these ions are easily separated from each other (Gale and Vestal, 2016).

Due to the nature of a TOF analyser, it is most often found paired to a MALDI source. However, it can be implemented into hybrid configurations that enable tandem MS. In 1996 Morris *et al.* described the quadrupole-time-of-flight (QTOF) hybrid analyser (Morris *et al.*,

1996). This configuration is now the most common hybrid TOF configuration and allows for higher mass accuracy and resolution and greater speed when compared to QqQ and TOF alone (El-Aneed, Cohen and Banoub, 2009; Haag, 2016) (Figure 1.20).

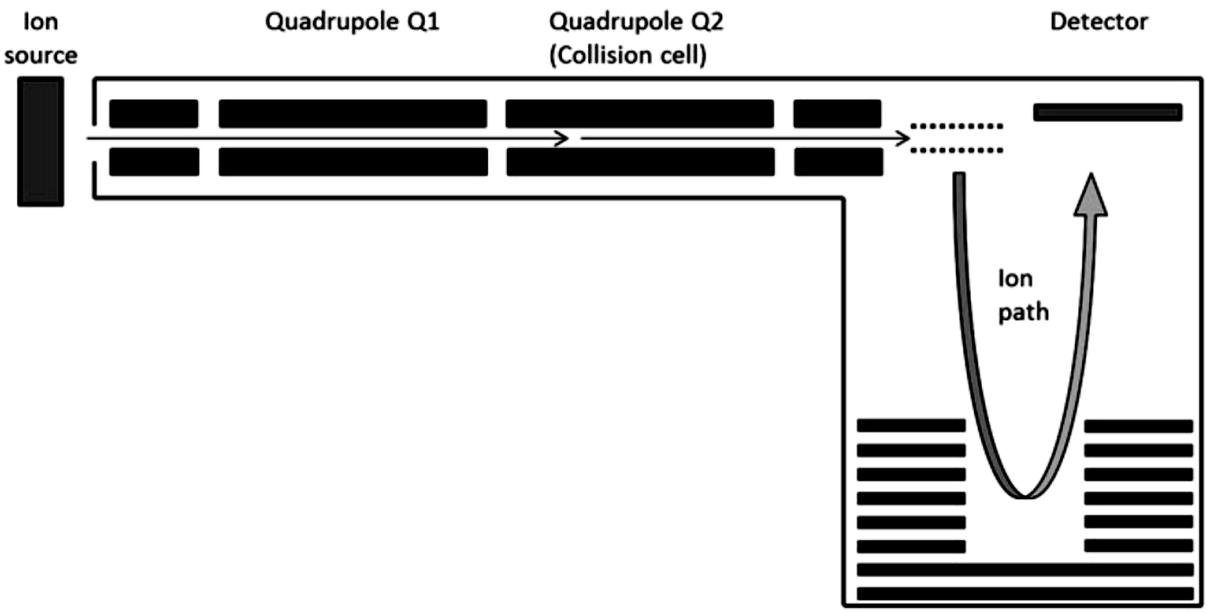


Figure 1.20: A hybrid QTOF mass analyser. The format of a triple quadrupole mass analyser is used, however the last quadrupole (Q3) is replaced by a time-of-flight analyser Adapted from (Haag, 2016).

1.7.3.4. Data acquisition methods

1.7.3.4.1. Data-dependent acquisition

Data-dependent acquisition (DDA) is defined as "Mode of data collection in tandem MS in which a fixed number of precursor ions whose m/z values were recorded in a survey scan are selected using predetermined rules and are subjected to a second stage of mass selection in an MS/MS analysis" (Mann, Hendrickson and Pandey, 2001; Murray *et al.*, 2013).

The standard "predetermined rules" is known as the TopN algorithm (Goldfarb, Wang and Major, 2016). The mass spectrometer is setup to perform a short MS1 scan in which it records intensity and m/z and identifies potential ions for analysis based on the most intense precursor signals (TopN). Thereafter the TopN ions are selected and subjected to MS/MS scans in which they are fragmented into their constituent fragment (product) ions and detected (see Figure 1.21). Precursor ions are fragmented from highest to lowest intensity. To ensure that the same precursor ions is not re-analysed in the next cycle, a dynamic exclusion window is set. During that period (usually 20 – 30 seconds) all precursor peptides that have already been subjected to

MS/MS analysis will be overlooked by the mass spectrometer allowing precursor ions of lower intensity to be analysed.

Due to the stochastic nature of this method it is irreproducible as peptides do not always ionise in the same manner, hence running a single sample multiple times may yield different results due to "missing values" (Zhang, Käll and Zubarev, 2016). Often the most abundant and intense peptide species will dominate the dataset derived from the sample and the species of lower abundance and intensity are left under-sampled or not sampled at all. Despite these apparent shortcomings, DDA still remains the most widely used shotgun tandem MS method.

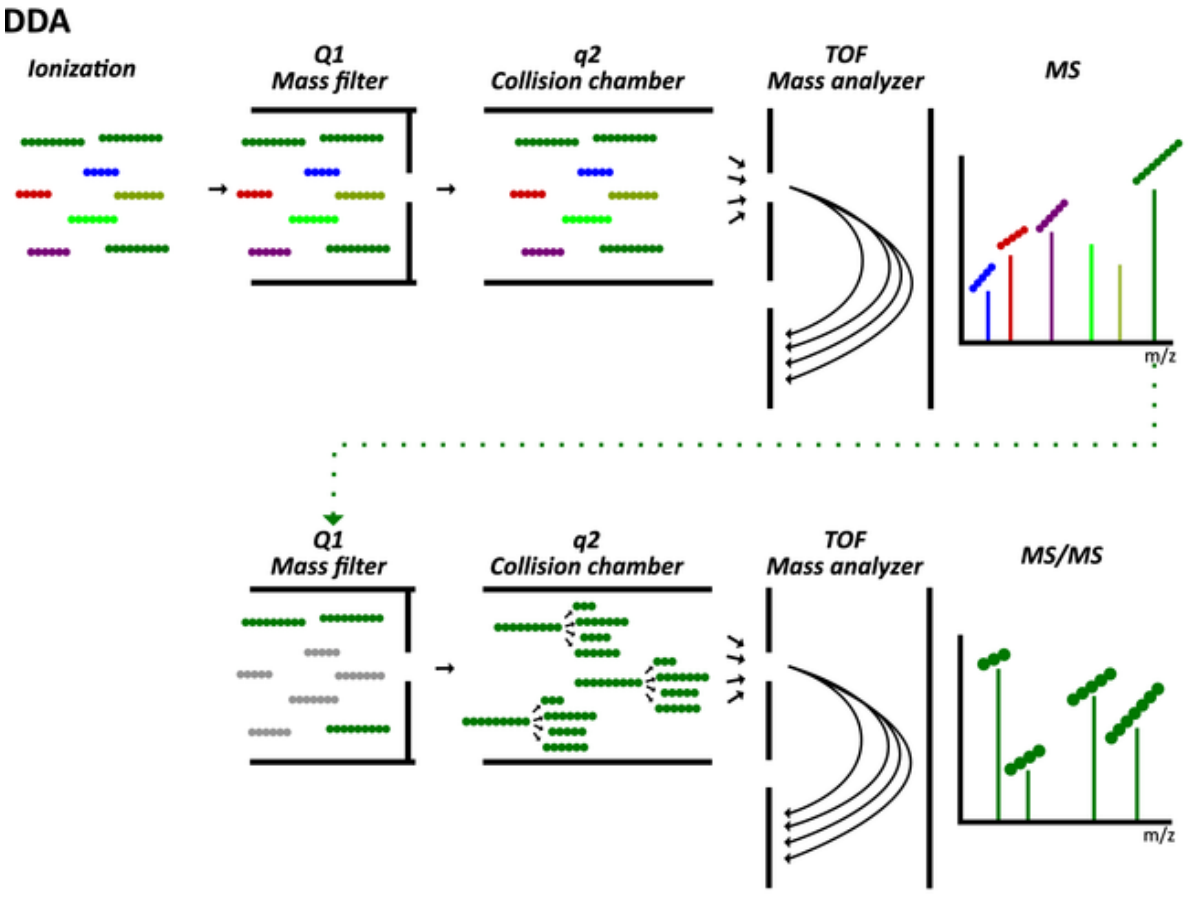


Figure 1.21: Schematic overview of data-dependent acquisition. A fixed number of precursors, identified in the first scan, are subjected to a second round of mass selection via MS/MS (Anjo, Santa and Manadas, 2017)**.**

1.7.3.4.2. Selected (multiple) reaction monitoring

Targeted MS is popular due to its inherent high sensitivity, accuracy and reproducibility. In contrast to DDA where 1000s of proteins can be monitored, selected reaction monitoring (SRM) is used to monitor a subset of proteins that have been predefined by the user (Borràs and Sabidó, 2017).

Generally, SRM is performed using triple quadrupole mass spectrometers in which Q1 is set to select a defined precursor, followed by fragmentation of the precursor in Q2 via CID and one of the fragments is monitored by Q3 which was predefined to select that m/z ion (Borràs and Sabidó, 2017; Vidova and Spacil, 2017)

The selection of a transition implies a specific precursor and its unique fragment ion.. This technique is characterised by high sensitivity and specificity since the chances of multiple transitions being identical is low (Borràs and Sabidó, 2017)

The pitfalls of this methodology lie in the experimenter needing to have prior knowledge of the sample before analysis. Furthermore, each individual transition needs to be set up and fully optimised before implementation in a workflow. This is a time consuming and tedious process. Lastly, the data is not re-mineable i.e. if a new protein is to be monitored, the entire sample needs to be analysed again after development of a transition specific to the new protein.

1.7.3.4.3. *S*equential *W*indow of *A*ll *Th*eoretical Mass Spectra (SWATH™-MS) or Data- Independent acquisition (DIA)

Since 2012, there has been an increase in the usage of DIA as the preferred acquisition method (Tsou *et al.*, 2016; Collins *et al.*, 2017). This method combines the unique strengths of both DDA and SRM acquisition methods into a hybrid methodology. In simple terms, DIA allows for unbiased data acquisition where all detectable ions are monitored resulting in a wealth of information that needs to be subjected to targeted data extraction to gain biological meaning.

The need for high quality, comprehensive, reproducible and easily implementable quantitative proteomics workflows has been at the forefront of proteomics research for the past few years. To address this issue, DIA MS has been developed. Recently, DIA has seen an upsurge in use worldwide due to its ability to collect data that is more comprehensive than traditional DDA methods. This is due to technological advances in both MS hardware and data processing software (Schilling, Gibson and Hunter, 2017). Additionally, the data generated can be mined retrospectively to answer questions that may not be apparent at the time of acquisition.

The DIA method when employed on a SCIEX (Framingham, Massachusetts, USA) TripleTOF® mass spectrometer is termed SWATH™ (**S**equential **W**indowed **A**cquisition of All **TH**eoretical Fragment Ion Mass Spectra) (Collins *et al.*, 2013; Rosenberger *et al.*, 2014, 2017) and when employed on Orbitrap-type mass spectrometers is simply referred to as DIA (Egertson *et al.*, 2013). DIA/SWATH™ generates MS/MS spectra that are more comprehensive and complex than MS/MS spectra derived from traditional DDA methods (Schilling, Gibson and Hunter, 2017).

For SWATH™ acquisition, small Q1 isolation windows are defined along the entire m/z range therefore, many precursors within that mass range are simultaneously fragmented. This leads to fewer missing values compared to DDA where only the most intense precursors are sampled for MS/MS. The fragments are then detected as complex MS coupled to MS/MS spectra (Figure 1.22), which needs targeted extraction where the software assigns fragments to peptides and then to proteins. There are two distinct methods that govern the size of the isolation windows used during the SWATH™ acquisition. One may select constant isolation windows of user-determined size by defining the window width during method development e.g. isolation window width Δ25 m/z with 1 m/z overlap between windows across 400 – 1000 m/z mass range, generates a data acquisition method containing 24 constant isolation windows.

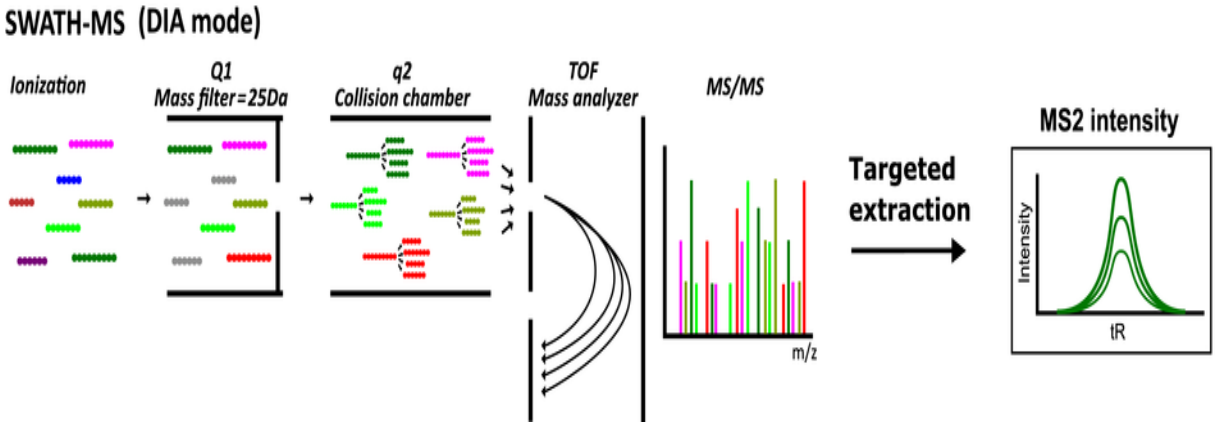


Figure 1.22: Schematic overview of SWATH™-MS. Q1 selects precursors within a defined m/z range that are fragmented in Q2. Thereafter the fragments are analysed by the TOF region. The Q1 selection in this diagram shows constant isolation windows of Δ25 m/z (Anjo, Santa and Manadas, 2017).

Alternatively, one can use software supplied by SCIEX (Framingham, Massachusetts, USA) to create a method that uses Q1 isolation windows of variable size based on the complexity of the sample. Here, the software creates smaller (narrow) windows in regions that have densely populated precursors of similar size thereby allowing fewer peptide precursors through for further analysis (Figure 1.23). This in turn allows for fewer missing values, increased specificity and a more comprehensive data acquisition ultimately leading to improved data quality (SCIEX, 2017).

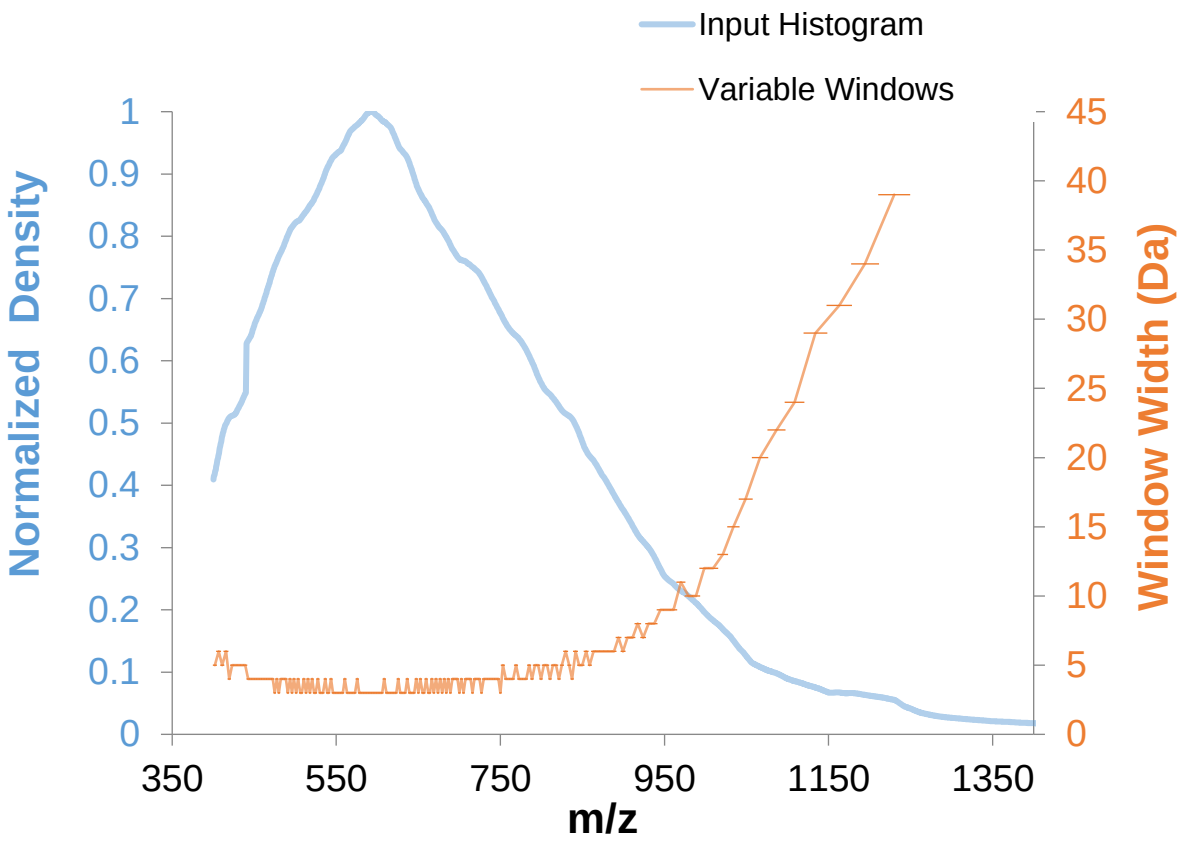


Figure 1.23: Variable Q1 window generation using SWATH™ Variable Window Calculator tool v1.0. The m/z density histogram generated from DDA TOF-MS data for a specific proteome (blue) is used to develop a SWATH™ method with variable isolation window widths (orange).

DIA/SWATH™ data can be processed using many different software packages that have been programmed specifically for SWATH™ data. The most common method involves generating a spectral library (a spectral map of precursors, product ions and their relative intensities as well as their position along a defined LC gradient) to which the SWATH™ data is searched against.

Generation of a comprehensive spectral library is critical for extracting in depth data from the SWATH™ experiments i.e. if a protein is not present in the spectral library, peptides belonging to that protein will not be assigned to it and therefore will not be identified and/or quantified in the sample. Spectral libraries are generated from DDA data, however DDA is stochastic in nature and therefore is prone to having many missing values (León *et al.*, 2013). To account for this during spectral library generation and thus ensure a comprehensive, in-depth spectral library, extensive fractionation of samples and at least three DDA technical repeats per sample are needed. Interestingly, the spectral library need not be generated from DDA sample runs that exactly mimic the LC setup and identical MS instrumentation and conditions that will be used during SWATH™ acquisition.

Chromatographic separation of peptide mixtures, to reduce the complexity of the sample and to ensure that the detector (mass spectrometer) is not overwhelmed, results in additional information being recorded for each peptide as a retention time (RT) (Escher *et al.*, 2012). The RT is unique to the particular peptide under a specific set of conditions and is influenced by the chromatographic mobile and stationary phases (solvent gradient, column length, dead volume) (Khurana and George, 2008). Additionally, RT is affected by variation in the LC system such as sample concentration, pump pressure and column aging (Escher *et al.*, 2012). Therefore, comparing samples across instruments and laboratories as well as intra-instrument over time has been a challenge. When applied to spectral library generation and SWATH™ data analysis, this poses a significant challenge since inevitable changes in peptide RTs would lead to the generation highly variable spectral libraries and subsequent targeted data extraction would be almost impossible.

To account for this, indexed retention time (iRT) – "an empirically derived dimensionless peptide-specific value that allows for highly accurate RT prediction" (Escher *et al.*, 2012), has been developed. The iRT of a peptide is a value that is relative to a standard set of iRT-peptides such as those developed by Biognosys (Schlieren, Switzerland). These are 11 non-naturally occurring synthetic peptides that have different chemical characteristics and thus elute at different times spanning the length of the LC gradient. By spiking these peptides into samples prior to separation by LC, all samples can be recalibrated (in the RT dimension) against these peptides enabling the samples to be compared more easily irrespective of the chromatographic gradient length, slope and chemistry. When applied to SWATH™, iRT recalibration enables the use of external, publicly available spectral libraries against which study specific data extraction can be conducted (Liu *et al.*, 2013; Anjo, Santa and Manadas, 2017).

Software such as SWATH™ 2.0, Skyline (MacLean, Daniela M Tomazela, *et al.*, 2010), Spectronaut™ (Khurana and George, 2008) and OpenSWATH (Röst *et al.*, 2014) make use of this strategy (Spectral library and iRT recalibration) for data extraction. DIA/SWATH™ has increased in popularity over the past two years and recently Bruderer *et al.* (2017) reported that DIA was able to generate 2-fold greater data points per protein in a single run compared to DDA.

DIA/SWATH™ has the ability to generate more comprehensive proteomic data than DDA (Rosenberger *et al.*, 2014), reproducibility that is similar to SRM, with fewer missing values, and the unique ability to provide a data-rich, proteomic map that can be re-mined at a later stage without need for new mass spectrometric analysis of the sample of interest (Pernikářová and Bouchal, 2015).

1.8. Data processing and analysis

1.8.1. Peptide mass fingerprinting (PMF)

This method is usually applied to relatively pure proteins and was one of the first methods routinely used to identify proteins in early proteomics experiments (typically 2DGE). The unknown protein is proteolytically cleaved into its characteristic constituent peptides, of specific mass and length, using an enzyme which has specific cleavage sites such as trypsin (Saraswathy and Ramalingam, 2011; Szabo and Janaky, 2015). The peptides are then analysed by the mass spectrometer (acquiring m/z and intensity) resulting in a precursor ion scan containing peptide accurate masses in a peak list (Saraswathy and Ramalingam, 2011). Data analysis is then performed on the peak list where the accurate masses of the peptides are compared to an in-silico digestion of an input protein database and the best protein match is calculated by the software (Baldwin, 2004; Saraswathy and Ramalingam, 2011) (Figure 1.24). Since the method is not suitable for complex samples, it is not widely applicable to modern biomarker discovery studies (Henzel, Watanabe and Stults, 2003; Domon and Aebersold, 2006).

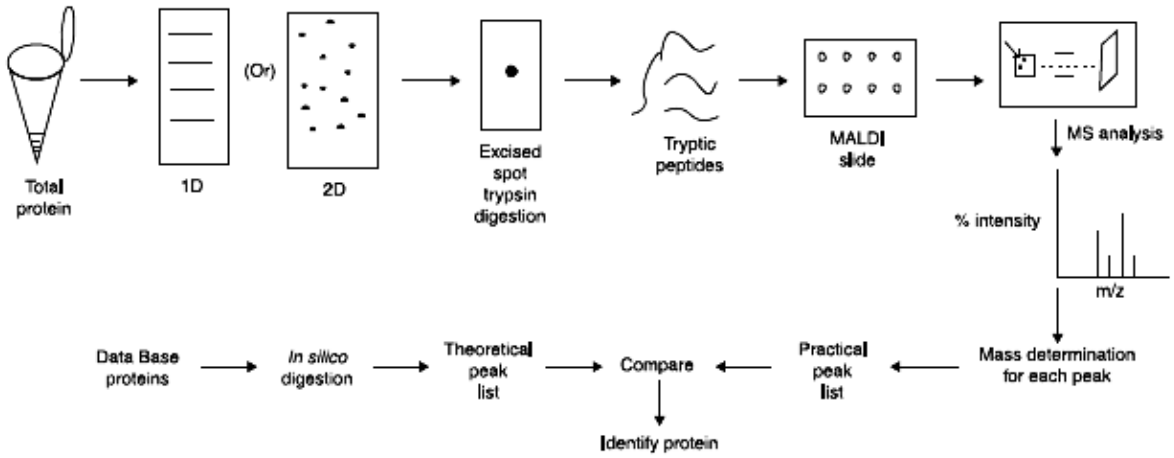


Figure 1.24: Overview of peptide mass fingerprinting. A protein sample is separated by 1D or 2DGE. The protein band/spot is excised and digested (in-gel) with trypsin into peptides. Mass spectrometric analysis yields a peak list which is then compared to an in-silico digested protein database. The software then computes the matching hits and identifies the protein (Saraswathy and Ramalingam, 2011).

1.8.2. Peptide sequencing by tandem mass spectrometry

In tandem mass spectrometry (MS/MS), as the name suggests, two stages of MS analysis are performed. Trypsin digested peptides that have been separated by LC are ionised by ESI and enter the hybrid mass analyser, such as QqQ or QTOF. There they are filtered and then fragmented into product ions which are detected.

The output data is a set of fragment spectra for each precursor ion that has been analysed. These fragment spectra need to be matched to a database to infer the protein identity. This is done using software packages that in-silico digest an input database into characteristic peptides and theoretical MS/MS spectra for each peptide. Thereafter, the measured MS/MS spectra from the acquired data are matched to the in-silico fragment MS/MS data (Figure 1.25). Finally, peptides are matched to their respective proteins. The method is costlier than PMF however the wealth of information acquired is worth the trade-off. Furthermore, this method is easily coupled to liquid chromatography mass spectrometry (LCMS) setups allowing for complex mixtures to be analysed.

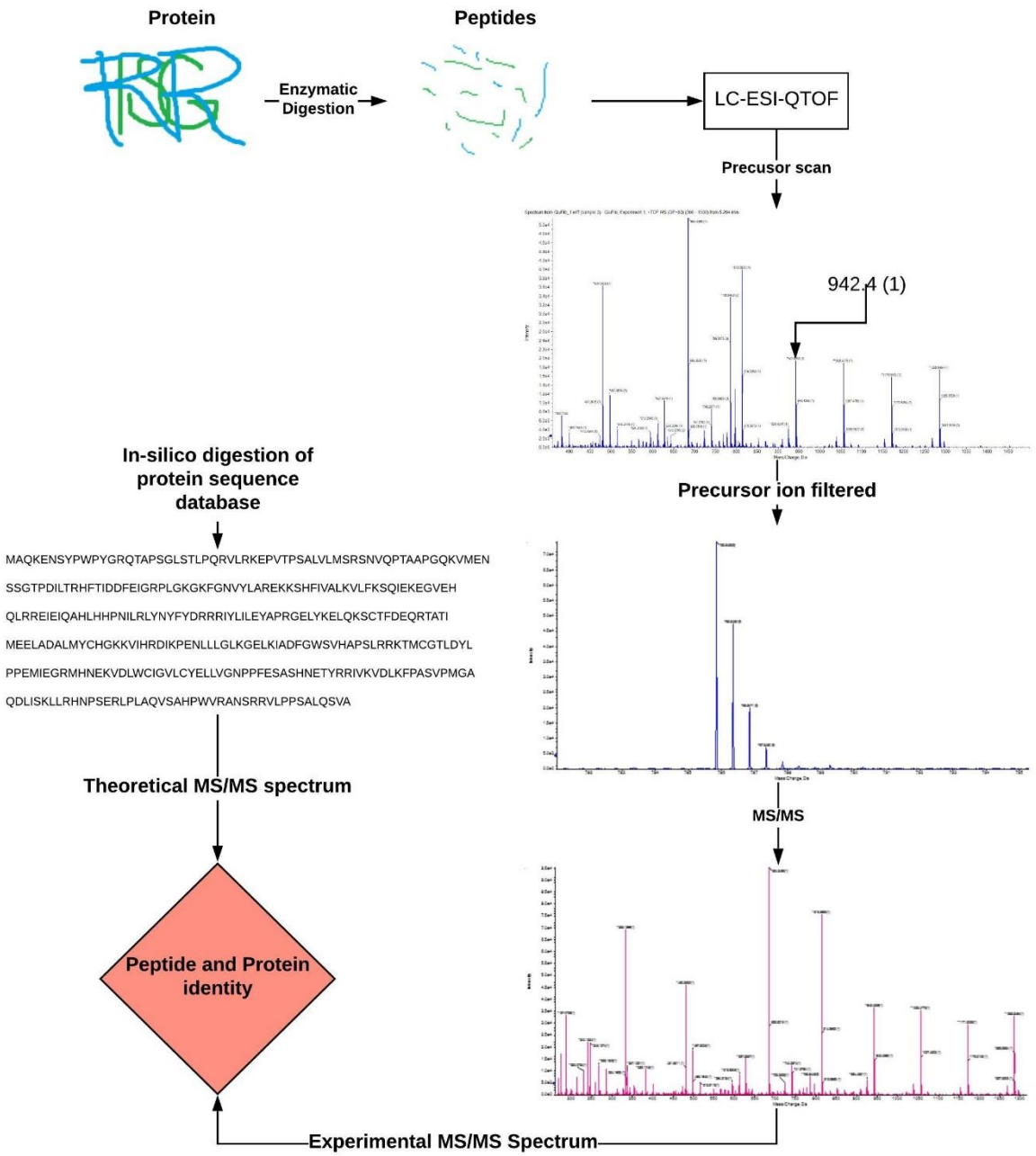


Figure 1.25: A typical tandem mass spectrometry experimental workflow. Proteins are digested into peptides, separated by liquid chromatography and ionised by electrospray ionisation. MS spectrum is scanned, and a precursor ion of 942.4 m/z is selected for tandem mass analysis. The selected precursor is then subjected to MS/MS scanning producing characteristic product ion fragments. These fragments are then used in search algorithms to identify the peptide and protein of interest by comparing the experimentally derived MS/MS spectra to theoretical MS/MS spectra derived from in-silico digestion of a protein sequence database.

1.8.3. False discovery rate analysis in database searching

The false discovery rate (FDR) describes the portion of positive tests that are incorrect i.e. the observation has occurred by random chance (Colquhoun, 2014).

In proteomics, the software used in analysis generates a value based on a statistical model that measures the quality of a match between a MS/MS event and a theoretical spectrum. This is called a peptide-spectrum match (PSM). Every MS/MS event is matched to the database and the highest scoring match is calculated (Aggarwal and Yadav, 2016). The larger the database, the higher the score needs to be, to be considered a true match. This is a single measurement that is specific to a single peptide only and is not applicable to the global or population error rate.

Hundreds to thousands of hypotheses are tested for significance with each generating an associated p-value – this is termed the "multiple testing problem" (Benjamini and Yekutieli, 2001). Bonferroni correction is, perhaps, the most well-known and widely used multiple comparison adjustment, however it is very stringent and results in low false positives for large sample sizes/tests but omits many true positives (Aggarwal and Yadav, 2016). In proteomics, the multiple testing correction of choice is the FDR approach developed by Benjamini and Hochberg (1995). This is defined as the "measure of the incorrect PSMs among all PSMs" and is less stringent than the Bonferroni correction (Aggarwal and Yadav, 2016).

In proteomics database searching, the most common method for FDR assessment is the target-decoy approach proposed by Elias and Gygi (2007). In this approach, the search databased is either: reversed, scrambled or randomised, and concatenated to the original database. Thereafter, the concatenated database is used in search engines and matches to decoy sequences are known false positives. This in turn allows the analyst to determine the number of incorrect matches in the dataset as well as set filtering criteria that can accurately and precisely differentiate between true and false positives (Elias and Gygi, 2010). The FDR approach can be applied at the PSM, peptide and protein levels (See 1.7.3 for details on protein level FDR) (The, Tasnim and Käll, 2016).

It is important to control the number of false positives when reporting on proteomics data. Two key concepts in in this regard include: global and local FDR. The global FDR is the most commonly reported value and provides information on the number of false positives that are expected out of all identifications. It is generally used to compare methods and conditions as a

global measure of the better approach (Hunter, 2015). The local FDR provides information on a specific protein of interest and answers the question: "how confident am I of this specific protein?" by providing the estimated number of false positive for a specific protein (Tang, Shilov and Seymour, 2008; Hunter, 2015).

1.9. Data analysis in proteomics

1.9.1. Protein identification overview

Bioinformatics tools are essential components of a proteomics experimental pipeline (Megger *et al.*, 2013). A plethora of commercial and open-source search engines are available to search raw data generated from MS experiments. Some of the most popular include SEQUEST (Washburn, 2015), MASCOT (Shilov *et al.*, 2007), X! Tandem (Craig and Beavis, 2004) and Comet (Eng, Jahan and Hoopmann, 2013). The Paragon™ algorithm (Shilov *et al.*, 2007) incorporated into Protein Pilot™ (SCIEX) is primarily used in the analysis of DDA data generated from SCIEX (Framingham, Massachusetts, USA) instruments. Andromeda (Cox *et al.*, 2011) is an open-source search algorithm that is incorporated into MaxQuant software (Cox and Mann, 2008) that can analyse data generated from any mass spectrometer.

Search engines compare measured m/z to theoretical masses of amino acid sequences in defined databases using MS/MS data (Mann, Hendrickson and Pandey, 2001; Kertesz-Farkas *et al.*, 2012), thereby the parent ion can be determined and can then can be mapped to the protein from which it was derived. Protein database sequences can be sourced from online repositories such as UniProt (Bateman *et al.*, 2017) and Ensembl (Zerbino *et al.*, 2018), to name a few. To minimise the FDR, decoy databases are included in the search database. Therefore, all matches to the decoy database are known false identifications and a score threshold that corresponds to a predetermined FDR can be set to increase validity of the results by minimising these false identifications (Megger *et al.*, 2013).

It is important to carefully define the search parameters before starting any search. These include: the sequence database, mass tolerance, digestion enzyme and number of missed cleavages, and variable modifications. These settings ultimately influence the search space and thus computing time and perhaps more importantly, affects the number of false positives that will be observed.

1.9.2. Sequence database

Using a study/sample specific database will ensure that false positive hits are reduced i.e. a human protein database should be used when analysing human cells, tissues, fluids etc. In the case of animal and plant studies where no protein database is available for the animal or plant model, the database of the closest related species should be used. It has been shown that using a complete, unfiltered, non-curated database such as the complete Ensembl database results in more peptide identifications (Fei *et al.*, 2011). This is due to the complete database containing isoforms of all proteins as compared to the filtered, curated, non-redundant database from the same organism e.g. Swiss-Prot (Cottrell, 2011). The search space using a complete database significantly adds to the time and computing power required for sample analysis and can lead to peptides being shared by many proteins – protein inference problem (Kertesz-Farkas *et al.*, 2012). If this is not accounted for, particularly in biomarker discovery studies, errors in differential protein abundance candidates can ensue (Fei *et al.*, 2011). Furthermore, by using a non-redundant database, protein isoforms that may be important biomarkers of disease will be missed since only one canonical sequence that represents the protein family will be present in the database (Fei *et al.*, 2011; Kertesz-Farkas *et al.*, 2012).

1.9.3. Mass tolerance

Setting mass tolerances is critical to limit the search space that is potentially matched to every MS/MS spectrum. Mass tolerance may be set at the MS (precursor) and MS/MS (fragment) levels and is dependent on the accuracy of the mass spectrometer (Eng *et al.*, 2011). Peptide mass tolerance limits the search to the theoretical peptides that have masses within a defined range. Setting this too narrow could result in the correct peptide not being used in the search and conversely, setting this too wide will increase the search space since more candidates will be assessed as possible matches for each experimental spectrum (Cottrell, 2011; Eng *et al.*, 2011).

1.9.4. Enzyme selection and partial digestion

The correct enzyme needs to be defined so that in-silico digestion can be done accurately and according to the theoretical digestion pattern of the sample (Eng *et al.*, 2011). Furthermore, it is important to set the maximum number of missed cleavages allowed since it is exceedingly likely that the enzyme of choice will not cleave at every possible cleavage site (Kertesz-Farkas *et al.*, 2012). An increase in missed cleavages allowance causes an increase in the search space and computational time. It is common to set the cleavage pattern as "specific at one end/semi-

specific" with up to two missed cleavages on the other end of the peptide (Cottrell, 2011; Eng *et al.*, 2011).

1.9.5. Modifications

Protein modifications are important variables that need to be considered during data processing. Post-translational modifications (PTMs) can occur during sample handling and processing and as a biological modification that results in a change in the protein structure/function. During a database search, modifications can broadly fall into two categories: fixed/static or variable modifications (Cottrell, 2011).

The most common static modification is the alkylation of cysteine which occurs during sample preparation and applies to all cysteine residues (Eng *et al.*, 2011). Therefore, this type of modification does not significantly add to the time or computational power required for the search.

In contrast, variable modifications are those that do not occur to all amino acids e.g. oxidation and phosphorylation. In these instances, the search space is significantly increased since the software must search all the modified and unmodified version of the same peptide. It is critical to add modifications to the search settings in a measured and calculated manner since it is easy to overwhelm the search by adding too many modifications i.e. combinatorial explosion can occur (Cottrell, 2011).

1.10. Protein quantification strategies

Mass spectrometry can simultaneously generate qualitative and quantitative data. The quantitative data can be subdivided into relative (Figure 1.26A) and absolute (Figure 1.26B) quantification. Relative quantification expresses the results of comparing multiple samples as fold changes of one compared to the other or a control, whereas absolute quantification measures the exact amount or concentration of a particular protein (Wasinger, Zeng and Yau, 2013).

Multiple methods for relative quantification have been developed, the most popular of which are shown in Figure 1.26A, including chemical tagging, metabolic labelling and label-free workflows. Absolute quantification methods require the use of internal standards (either peptides or proteins) of known concentration (Hawkridge, 2014).

The most popular methods used in quantitative proteomics are: isobaric tags for relative and absolute quantification (iTRAQ) (Ross *et al.*, 2004) or tandem mass tags (TMT) (Thompson *et*

al., 2003), stable isotope labelling of amino acids in cell culture (SILAC) (Ong *et al.*, 2002), MRM, and label-free quantification (LFQ), each with its own pros and cons.

1.10.1. Label-based quantification

Isotope labelling methods can be broken down further into chemical and metabolic labelling. In metabolic labelling, stable isotopes are incorporated into cell culture media; thus, all cells grown in that media are exposed and can incorporate the labels as part of the normal growth cycle (Wasinger, Zeng and Yau, 2013). Samples can be pooled together early (at the protein level) in the sample preparation workflow and all further steps can be performed on the combined sample. This reduces technical variation since all steps post-pooling are conducted on "one" sample. SILAC is one such method that employs metabolic labelling of heavy and light arginine and/or lysine amino acids in the culture media (Wasinger, Zeng and Yau, 2013). SILAC has also been applied to mouse models (Zanivan, Krueger and Mann, 2011; Collins *et al.*, 2013), however this is exceedingly expensive. Due to the small technical variation, small differences in protein abundance can be detected by comparing light and heavy peptides. This method is not applicable to samples such as plasma, serum and urine.

In chemical labelling, labels are chemically added to proteins and peptides during sample preparation. The most popular methods are iTRAQ and TMT. Briefly, these methods label at the peptide level, and allow for up to eight (iTRAQ) and 10 (TMT) samples pooled together into one super sample that is analysed via LCMS (Latosinska *et al.*, 2015; Moulder *et al.*, 2017; Zhang and Elias, 2017). The physicochemical properties of the peptides are not altered by the tags and thus identical peptides (from different samples) elute of the column at the same time. These are then analysed by the mass spectrometer and upon fragmentation, the tags are released, and their signature signals are measured. The various tag signals can then be correlated to the respective samples pre-pooling and relative quantification can be achieved. This method is limited by digestion efficiency and label incorporation efficiency, but it still favourable due to its ability to be multiplexed and it its amenable to many biological samples including plasma, serum and urine (Wasinger, Zeng and Yau, 2013). Additionally, since peptides are pooled after labelling, all downstream technical variation is negated.

Absolute quantification can be achieved using MRM and isotopically labelled synthetic peptides that have the same physicochemical properties of the precursor peptide in the transition (Wasinger, Zeng and Yau, 2013). Thus, when the peptides (sample and synthetic) are selected for fragmentation, their product ions can be used to determine the original absolute

peptide amount based on the signal of the isotopic synthetic peptide (Wasinger, Zeng and Yau, 2013; Vidova and Spacil, 2017). Importantly, this only provides quantification for the individual peptide therefore, multiple peptides from the sample protein are needed to extrapolate the absolute protein quantity. Despite this apparent shortcoming, MRM based absolute quantification has been used in many plasma biomarker studies [reviewed in (Wasinger, Zeng and Yau, 2013)].

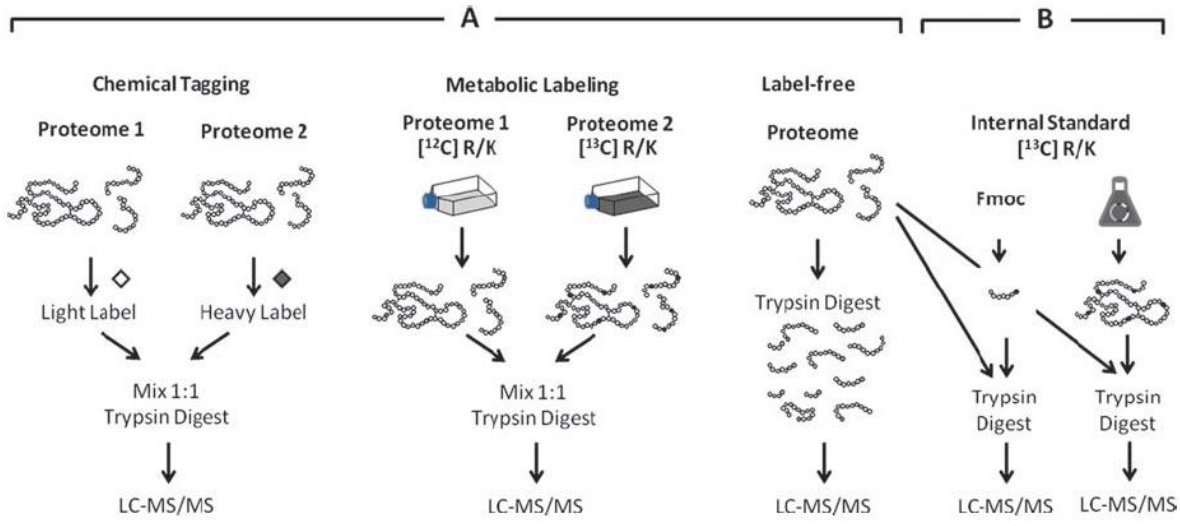


Figure 1.26: Overview of quantification strategies for relative (A) and absolute (B) protein quantification (Hawkridge, 2014)**.**

1.10.2. Label-free quantification

Label-free quantification requires a robust and highly reproducible LC system (Wasinger, Zeng and Yau, 2013). Two distinct methods exist for LFQ viz. spectral counting and area under the curve (AUC) (Figure 1.27).

Spectral counting involves counting the number of fragment ion spectra for peptides of a given protein i.e. the greater the number of fragment ions, the greater the amount of peptides and therefore the greater the abundance of the protein (Megger *et al.*, 2013). This method is not widely used, due to its inherent weaknesses of not measuring physical data and relying solely on counts for quantification and the ability to only detect large differences in abundance (Bantscheff *et al.*, 2007).

Area under the curve measurement has been shown to be linearly proportional to peptide concentration for peptides 10 fmol to 100 pmol (Neilson *et al.*, 2011)**.** Ion abundances are

measured at specific retention times as ionised peptides elute from the column into the mass spectrometer. For differential analysis, the same ions are measured at a specific m/z and retention time across samples and their AUC are calculated and compared (Figure 1.26). This method is fast, relatively simple to perform and is easy to analyse and inexpensive (Wasinger, Zeng and Yau, 2013), however there are a few key elements that need to be accounted for to ensure reproducible and accurate results. Co-eluting peptides, shifts in retention time and gradual decrease in signal over time, pose significant hurdles to accurate AUC analysis. There are ways to address these issues including increasing the LC gradient, retention time realignment and normalisation of the spectral data, respectively (Neilson *et al.*, 2011). Label free proteomics is generally applied to biomarker discovery studies and numerous studies have made use of this approach [reviewed in (Megger *et al.*, 2013, 2014)].

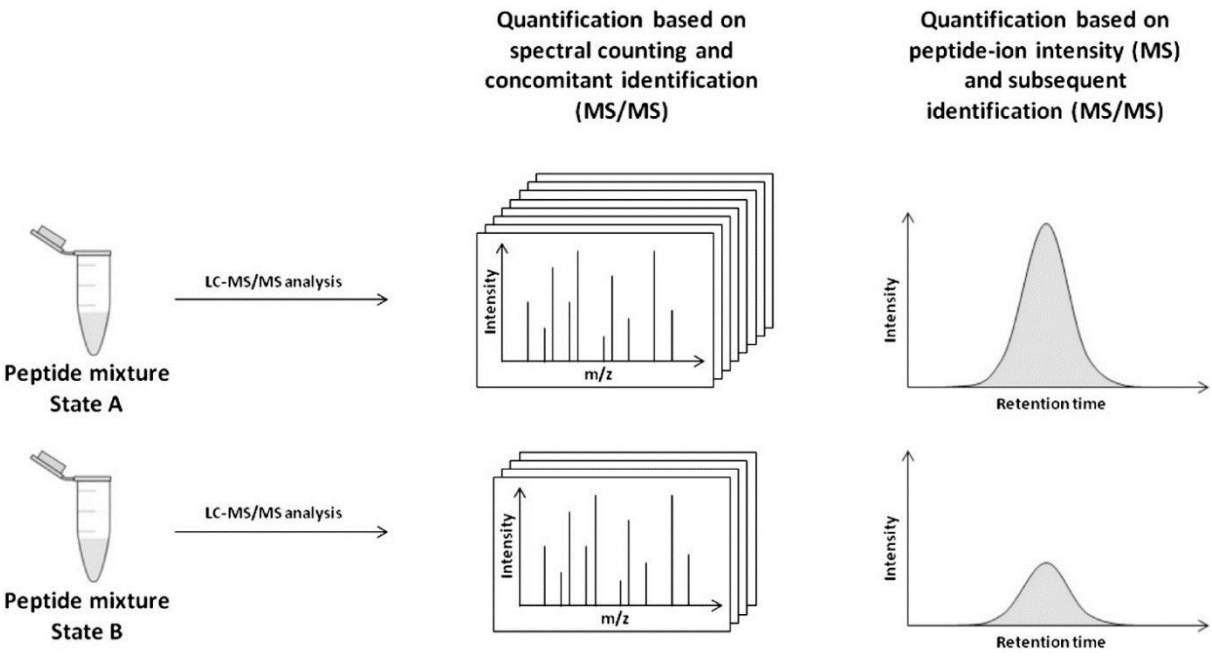


Figure 1.27: Schematic overview of the two LFQ methodologies. Spectral counting quantification is estimated based on the number of peptide spectrum matches for a particular peptide in each sample. In ion intensity based quantification, changes in peptide abundance is determined by comparing chromatographic peak areas of the same peptide in each sample (Megger *et al.*, 2013).

1.10.3. Quantification considerations

There are numerous considerations that need constant thought when analysing proteomics data. These include: outliers, missing values, normalisation, statistical considerations and controlling for multiple testing using the FDR approach.

Outliers: are defined as data points that fall outside the normal range of expected data. These can be accounted for by using defined mathematical rules e.g. Dixon's Q test or if there is a consistent and clear reason to discard the data point/s. However, use of the latter may introduce bias based on the user.

Missing values: due to the nature of LCMS-based proteomics, and DDA proteomics in particular, missing values are inevitable and represent a significant challenge during data analysis (Webb-Robertson *et al.*, 2015; Lazar *et al.*, 2016). In proteomics, missing values can be as a result of many diverse known and unknown effects such as, variation in trypsin efficiency, ion suppression and ionisation competition, dynamic range and type of software used in analysis to name a few (Lazar *et al.*, 2016). Additionally, missing values at the peptide level can be as a result of proteins of interest being absent from the sample entirely or below the detection limit (Webb-Robertson *et al.*, 2015).

Three types of missing values have been defined: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). Options to deal with missing values include, ignoring missing values and dealing with available data only, use of imputation methods and the use of statistical tests that are tolerant to missing values. Imputation methods are employed on a peptide level in bottom-up proteomics experiments. There are multiple imputation methods that can be employed based on the type of missing values each with its advantages and disadvantages [reviewed in (Webb-Robertson *et al.*, 2015; Lazar *et al.*, 2016)]

Normalisation: is the process of dividing all data by a known variable (source of variation) such that the data is free from bias associated with that variable. This allows the data to be more comparable to each other (Karp, 2010; Välikangas, Suomi and Elo, 2018). It is critical to ensure that the normalisation approach does not introduce bias. Furthermore, normalisation assumes that most of the features do not change between groups, however if this is not true then careful attention needs to be paid when choosing how and when to normalise the data.

In proteomics many thousands of peptides are measured, and their abundances can be because of true biological changes and can also be introduced by systematic errors in sample handling, preparation etc. and noisy measurements in during acquisition. Collectively, these increase the

variability in the measurement of the feature and therefore needs to be accounted for to ensure that conclusions derived from the data are statistically and biologically sound (Callister *et al.*, 2006). Numerous normalisation strategies exist and many software employ unique normalisation approaches [for details see (Callister *et al.*, 2006; Webb-Robertson *et al.*, 2011; Välikangas, Suomi and Elo, 2018)], however, the most common method, particularly in label-free proteomics, is based on total ion current normalisation (Karpievitch *et al.*, 2009; Mizuno *et al.*, 2016; Wulff and Mitchell, 2018).

Multiple testing correction: in a biomarker discovery study looking at just one protein, the mean of the protein abundance is compared between two groups. The null hypothesis is that there is no difference in abundance between the two groups. In this scenario, if the protein does show a difference between the two groups, the p value computed will be below a defined threshold (usually 0.01). This is an independent test and computes the "per comparison error rate" (PCER), where there is a 1% chance of the conclusion being incorrect. However, in a typical proteomics biomarker discovery experiment, 1000s of proteins are compared and 1000s of similar statistical tests are conducted, therefore at a 0.01 p value cut-off, 1% of all differentially abundant proteins will be false positives (Karp, 2010) – this is known as the multiple testing problem (Benjamini and Yekutieli, 2001). To account for this, control the "family wise error rate" (FWER) is required. The FWER controls the "probability of one or more false rejections out of all tests conducted" (Karp, 2010). The most common FWER control is the Bonferroni-correction, which divides the PCER by the total number of tests conducted (Armstrong, 2014). This is a very stringent approach leading to low powered studies since many true biological changes may be excluded since they will not meet the Bonferroni corrected cut-off (Karp, 2010). This type of correction is far too stringent for initial phase biomarker discovery studies where multiple subsequent verification and validation studies will be performed (Karp, 2010). To this end, controlling the FDR has emerged as the optimal approach for these types of studies as proposed by Benjamini and Hochberg in 1995 (Benjamini and Hochberg, 1995). The FDR FWER correction simply means that for a FDR of 1%, on average 1% of all proteins identified as significant will be erroneous. In 2002, Storey developed an extension of the FDR approach where a q-value is calculated for each identified feature (Storey, 2002). The q-value is the proportion of false positives expected when a defined "feature and all other significant species have a significant change in expression" (Karp, 2010).

1.11. Limitations of mass spectrometry-based proteomics

Despite the numerous advances that have seen MS grow into the leading analytical technique, MS proteomics-based biomarker discovery has its limitations which are detailed here. The promises of MS-based proteomics are yet to be fulfilled, these include: comprehensive analysis of complex biological samples, ability to detect low abundance species reproducibly, quantification over large dynamic range, and analysis of protein complexes and high throughput analysis (Chandramouli and Qian, 2009).

There are two categories of limitations in proteomic analysis: a) limitations regarding the proteome composition of the proteome of interest relative to the expression levels of proteins and b) limitations in the analytical methods and technologies.

The dynamic and diverse nature of the proteome poses a significant obstacle to biomarker discovery-based proteomics. Sample preparation needs to be well thought out and planned in the analysis of proteomes. This is due to proteins being found in various sizes, subcellular localizations, charge states, PTMs and protein complexes (Garbis, Lubec and Fountoulakis, 2005). Therefore, it is critical to isolate and prepare the correct (sub)sample to answer the research question that is being asked. Proteins are expressed in different amounts in all proteomes and this needs to be taken into consideration during proteome analysis (Graves and Haystead, 2002). Biomarker discovery using biofluids such as plasma, serum and urine, is challenging due to the wide dynamic range in protein concentration that characterises these commonly used samples (Garbis, Lubec and Fountoulakis, 2005; Hortin and Sviridov, 2010). There are various methods available to enrich low or deplete high abundance protein species and each needs to be assessed carefully when designing a proteomics experiment. Importantly, the depleted sample is not necessarily an accurate representation of the original proteome (Geyer *et al.*, 2017), it may be a useful approach for qualitative assessment of a proteome. The issues regarding sample preparation are compounded when one considers that there is no standard operating procedure for the preparation of various types of samples (Chandramouli and Qian, 2009).

The drawbacks described above when applied to biomarker discovery are compounded significantly. This is largely due to pre-analytical variation in sample collection, processing and storage. These need to be carefully and thoroughly assessed during the study design phase to minimise their effects on the final biological data.

Drawbacks of MS relate to ionisation efficiency, instrument performance and difficulty in measuring certain protein classes (Garbis, Lubec and Fountoulakis, 2005). Electrospray ionisation is never completely efficient due to changes in the physicochemical properties of the peptides being eluted of the column. The ionisation of identical m/z species in two runs in series may show significantly different ionisation efficiencies. Furthermore, during LC separation of peptides, peptides with similar characteristics but derived from different proteins will elute from the column at the same time – co-elution of peptides. This could lead to the mass spectrometer not being able to analyse each independently as they may compete during ionisation and subsequent analysis in the mass spectrometer. There are methods to limit these types of interactions from affecting the analysis, however, the mass spectrometer needs to be set up to address this issue by a skilled operator who is aware of this limitation. Furthermore, MS has a limited dynamic range of detection which is not sufficient to map proteomes in their entirety without extensive fractionation.

Lastly, operation of a mass spectrometer requires a highly skilled, diligent and consistent operator with high attention to small details that can affect mass spectrometric analysis. Instruments need to be tuned and calibrated regularly, the laboratory must be maintained at a constant temperature and humidity and the laboratory must be a dust-free environment (Garbis, Lubec and Fountoulakis, 2005). These components, although small, can influence MS in multiple, and often, detrimental ways.

Even though there are several limitations to proteomics, the rapid advancement of the hardware, technology and bioinformatics bodes well for the future of the field. When used in conjunction with other complementary technologies, it can generate data that has the potential to change the way we approach biological and health sciences.

1.12. Aims and objectives

The aim of this project was to develop a simple and robust analytical method that can be applied to clinical urine samples to comprehensively profile the urinary proteome to aid in the search for novel, sensitive and specific biomarkers for early AKI diagnosis. More specifically the objectives were to:

1) Develop and optimise a biomarker discovery pipeline using SWATH™ MS and apply the pipeline to a clinical cohort.

2) Collate the proteome analysis data, demographic data and clinical data to determine whether any correlation exists between patient proteome composition and key clinical outcomes related to first line ART induced AKI.

Chapter 2
Materials and Methods

Part A

Method development and optimisation for urinary proteome profiling in clinical proteomics

The development of a robust biomarker discovery pipeline is essential to achieve a method that is reliable and reproducible. For such a pipeline to be developed, each step of the process needs to be thoroughly tested and critically evaluated in an iterative manner. In doing so, technical variation can be limited, allowing for true biological differences to be determined. This chapter focuses on the development of such a pipeline. It describes and evaluates, in detail, the various steps taken to arrive at the final workflow used to analyse the urinary proteomes of clinical samples obtained from the patient cohort described in the next chapter.

For each of the sub-experiments carried out below, the raw data files were searched as described next (see Section 2.1) and/or processed using Spectronaut™ 11 software (Biognosys Inc, Beverly, Massachusetts, USA) (see Section 2.17). For data dependent analysis (DDA), peptide and protein identifications at 1% FDR and chromatographic peak widths (where appropriate) were used as a measure of the best configuration/settings for each sub-experiment. For data independent analysis (DIA) peptide and protein matches at 1% FDR, data points per peak, data completeness and co-efficient of variation were used to determine the optimal conditions. The experiments proceeded in the order described with the best configuration/setting/method (showing highest protein/peptide identifications) being carried forward and kept constant whilst manipulating the next variable. This process was performed in an iterative manner and the final method, to be used on clinical samples, was derived from the data after all the sub-experiments were concluded.

2.1. Database search parameters

Raw data (.wiff files) was searched against the human UNIPROT sequence database (reviewed entries, downloaded on 2 June 2017) using MaxQuant (v1.5.8.3) (Cox and Mann, 2008; Cox *et al.*, 2014). Common contaminants, such as keratins, were added to the search automatically by the software and were thus not added to the downloaded sequence database. The search settings applied were as follows: trypsin as the proteolytic enzyme with up to two missed cleavages allowed, mass tolerance of 20 ppm for the first search and 6 ppm for the main search. Oxidation of M and N-terminal acetylation were set as dynamic modifications and carbamidomethylation of C as static modification. False discovery rate was set at 1% on the peptide and protein level. The "match between runs" feature was selected for replicate injections. All other search parameters were set to default.

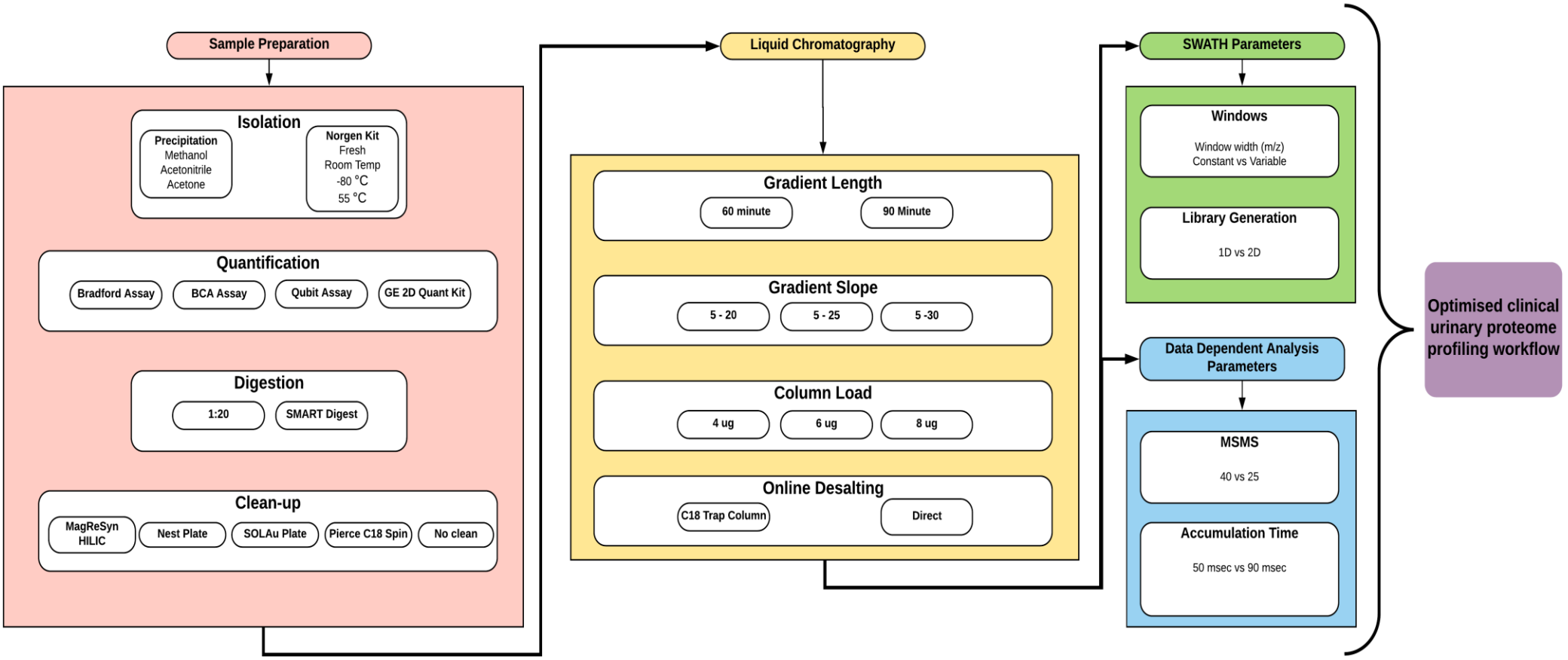


Figure 2.1: Schematic overview of the parameters optimised in the development of a robust and reproducible urinary proteome profiling workflow. Each individual component is explored and explained in detail in this chapter.

2.2. Isolation of urinary proteins

2.2.1. Methods employed to isolate urinary proteins

A range of methodologies were applied to isolate urinary proteins. These included kit-based approaches and commonly used precipitation methods. All methods were assessed in technical duplicate. Precipitation was achieved using either methanol, acetonitrile (ACN), acetone, acetone-trichloroacetic acid, ethanol and ammonium acetate. Following precipitation, the pellets were resuspended in either ammonium bicarbonate (AmBic), Hexafluoroisopropanol (HFIP), Ammonium acetate (NH_4OAc), rehydration buffer (RB) or sodium dodecyl sulfate (SDS).

The kit-based approach used the commercially available - urine protein concentration, preservation and isolation kit (Catalogue # 38200) developed by Norgen Biotek Corporation (Thorold, Ontario, Canada). The methodology is explained in detail in Section 2.2.5.

Only methodologies that were successful in isolating protein from urine are highlighted in more detail below. These include acetone, acetone-trichloroacetic acid and methanol precipitation and the kit-based approach.

2.2.2. Methanol precipitation

Protein extraction via methanol precipitation was performed following the procedure described by Wessel and Flügge (1984) with minor modifications. Crude urine (100 µL) was used as the starting material to which 400 µL methanol was added and vortex mixed well for 1 minute. Thereafter, 300 µL MS-H_2O was added and vortex mixed well for 1 minute. The sample was centrifuged at 14 000 x *g* for 1 minute followed by the addition of 400 µL methanol. A final centrifugation for 2 minutes at 14 000 x *g* was performed to pellet the proteins. Supernatant was removed by aspiration and the pellet was resuspended in 20 µL rehydration buffer (RB) [7 M urea, 2 M thiourea, 40 mM Tris-base, 1% 3-[(3 cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), 50 mM Dithiothreitol (DTT)]. Finally, a short centrifugation for 30 seconds at 14 000 x *g* was performed to pellet insoluble materials and the resuspended protein was transferred to a fresh tube and stored at -80 °C until further use.

2.2.3. Acetonitrile precipitation

Protein extraction via ACN was done according to the method of Kay *et al.* (2008). Briefly, 40 µL crude urine was used as the starting material and diluted with 80 µL MS-H_2O. Thereafter 180 µL ACN was added and sonicated in a water bath for 10 minutes, followed by brief vortex mixing and a further 10-minute sonication. The precipitated proteins were then pelleted by centrifugation at 14 000 x *g* for 10 minutes. Supernatant was removed by aspiration. The pellets were subsequently vacuum dried and resuspended in 20 µL RB. Finally, a short centrifugation for 30 seconds at 14 000 x *g* was done to pellet insoluble materials and the resuspended protein was transferred to a fresh tube and stored at -80 °C until further use.

2.2.4. Acetone and acetone-trichloroacetic acid precipitation

Protein extraction via acetone and acetone-trichloroacetic acid precipitation followed the method reported by Beretov *et al.* (2015) with slight modifications Acetone precipitation was achieved by adding 8 parts of ice-cold acetone to 1 part urine and stored at -20 °C for 1 hour. Thereafter the samples were centrifuged at 11 000 x *g* at 4 °C for 30 minutes. Acetone-Trichloroacetic acid (TCA) precipitation was achieved by the acetone precipitation procedure as detailed above followed by addition of freshly prepared 100% TCA solution to the pellet such that TCA solution was in a ratio of 1:4 with the original urine volume. This was incubated at 4 °C for 1 hour followed by centrifugation at 11 000 x *g* at 4 °C for 30 minutes. The precipitated protein pellet from each method was resuspended in 200 µL RB and/or 2% SDS. Finally, a short centrifugation for 30 seconds at 14 000 x *g* was done to pellet insoluble materials and the resuspended protein was transferred to a fresh tube and stored at -80 °C until further use.

2.2.5. Kit-based urinary protein methods

2.2.5.1. Collection, concentration and preservation

The urine protein concentration, preservation and isolation kit developed by Norgen Biotek Corporation was used according to the manufacturer's instructions. Briefly, donor/patient urine was added to urine collection tubes containing Norgen's proprietary resin (30 mL per condition/patient). The tubes were then gently invert mixed to allow the resin to interact with (bind) protein in the crude urine. Once fully resuspended, the tubes were allowed to stand at room temperature (RmT) for 15 minutes or until the resin settled completely. The protein depleted urine was then gently decanted leaving the resin (containing bound protein) within

the collection tube. Norgen's proprietary preservation and antimicrobial buffer was then added to the resin and vortex mixed gently for 15 seconds. The preserved samples were then stored according to the manufacturer's instructions as follows: a) no storage with immediate protein isolation, b) room temperature (20-26 °C), c) incubator at 50 °C and d) freezer at -80 °C. Each condition (b-d) was stored for 1 week at the specified temperature before protein isolation described below.

2.2.5.2. Protein isolation following preservation

The preserved urine samples were left to equilibrate to RmT. To the preserved urine sample 750 µL of wash solution (supplied in kit) was added and vortex mixed for 15 seconds. Thereafter the slurry was transferred, 400 µL at a time, to a mini-filter spin column and centrifuged for 1 minute at 14 000 x *g* at RmT. Wash solution (400 µL) was then added to the spin column and centrifuged at 14 000 x *g* at RmT. This step was repeated twice followed by a final 3-minute centrifugation at 14 000 x *g* at RmT to remove residual wash solution. All unbound material was discarded up to this point. The spin column was then placed in an elution tube and 200 µL of elution buffer (EB) (supplied in kit) was applied to the resin. The column was then centrifuged for 2 minutes at 2 000 x *g* at RmT followed by 2 minutes at 14 000 x *g* at RmT. Finally, 18 µL of protein neutralizer (supplied in kit) was added to the eluted protein. The isolated protein was then stored at -80 °C until further use.

2.3. Quantification methods

The quantification of crude protein isolates is a critical step in a proteomics workflow as all downstream analysis is reliant on an accurate and reproducible quantification of crude protein extracts. The next set of experiments aimed to determine which of the common total protein quantification methods displayed greatest accuracy and precision with the lowest variation.

Two standard solutions of bovine serum albumin (1 µg/uL and 2 µg/uL) were prepared and assessed on 3 consecutive days using the methods outlined below.

2.3.1. Bicinchoninic assay

The bicinchoninic assay (BCA) assay was performed according the method of Smith *et al.* (1985). The assay relies on the formation of a Cu^{2+}-protein complex under alkaline conditions. This is followed by reduction of the Cu^{2+} to Cu^{1+} where the amount of reduction is directly proportional to the protein concentration. It is an easy assay that is moderately tolerant to

detergents with a very stable product that can be measured easily using a spectrophotometer in a 96-well plate format.

The BCA Protein Assay Kit (Sigma-Aldrich, St Louis, Missouri, USA) was used according to the manufacturer's instructions.

2.3.2. Bradford assay

The Bradford assay was performed according to the method of Bradford (1976). The principle of the assay is based on the binding of Coomassie Brilliant Blue G250 dye to proteins. Under acidic conditions the brown-red form of the dye is converted into a blue form upon binding to proteins. This blue colorimetric change is then measured using a spectrophotometer at 595 nm. The amount of protein is directly proportional to the degree of colour formation. The assay is also moderately resistant to salts and chaotrophic agents however it is affected by many common sample buffer detergents such as CHAPS and SDS.

The commercially available Quick Start™ Bradford Assay kit (Bio-Rad, Hercules, California, USA) was used according to manufacturer's instructions. Briefly, protein samples and standards were added to Bradford reagent (at RmT) at a 1:50 ratio. These were vortex mixed briefly and allowed to incubate for 5 minutes, thereafter transferred (200 µL/well) to a 96-well plate and absorbance was read at 595 nm using a spectrophotometer. A standard curve was generated in Microsoft® Excel® (v1708) and used to determine sample concentration.

2.3.3. Qubit assay

The fluorescent Qubit® Protein Assay Kit developed by ThermoFisher Scientific (Waltham, Massachusetts, USA) was used, according to the manufacturer's instructions, as an alternative method to quantify proteins. This assay is highly selective and sensitive for proteins. It is moderately tolerant to common contaminants and reducing agents such as DTT however it too does not perform well in the presence of detergents such as SDS and CHAPS.

Briefly, Qubit® Protein reagent was mixed with Qubit® Protein buffer in a 1:200 ratio to create a Qubit® working solution. Qubit® standards or protein samples (10 µL) were added to Qubit® assay tubes. Thereafter, 190 µL Qubit® working solution was added to each tube and vortex mixed for 3 seconds. All tubes were then incubated at room temperature for 15 minutes. A standard curve was generated from the standards using the Qubit® 2.0 Fluorometer (ThermoFisher Scientific, Waltham, Massachusetts, USA). Finally, all sample tubes were read, and the quantification data was generated by comparing the absorbance measured in each

sample to the standard curve. This was done automatically by the Qubit® 2.0 Fluorometer internal software.

2.3.4. 2-D GE Quantification Kit

The 2-D Quant Kit (GE Healthcare, Chicago, Illinois, USA) was also assessed as a method for protein quantification. This assay was designed to quantify proteins before two-dimensional electrophoresis – a process that requires sample proteins to be resuspended in a buffer that contains many potential contaminants in MS analysis. The common buffer is made up as follows: 6-9 M urea, 0.5 – 4% detergents (CHAPS, SDS), low molar reducing agents and 0.2 – 2% ampholytes. The method precipitates proteins out of solution and leaves behind all other contaminants. This is a distinct advantage over other colorimetric methods such as BCA and Bradford assays that are not able to quantify proteins in these types of buffers due to the many components that interfere with the reduction of cupric ions to cuprous ions.

The precipitated proteins are resuspended in a copper solution and the proteins can react with a colorimetric reagent. This assay is different to the BCA and Bradford assay in that the absorbance is inversely proportional to protein concentration.

Briefly, working reagent was prepared by mixing 100 parts of colour reagent A with 1 part colour reagent B. Thereafter, 10 µL of each sample or bovine serum albumin standard (provided in the kit) was added to protein Lo-Bind tubes. Precipitant solution (500 µL) was added, vortex mixed and incubated on ice for 15 minutes. Thereafter, 500 µL co-precipitant was added and vortex mixed. The tubes were then centrifuged at 10 000 x *g* for 5 minutes at room temperature to sediment precipitated proteins. The supernatant was completely removed, and a further centrifugation was done (12 000 x *g*, 1 minute, room temperature), followed by removal of any residual supernatant without disturbing the pellet. 100 µL of copper solution and 400 µL deionised water was added to each pellet and vortex mixed. Lastly, 1 mL of working solution was added to each tube. The tubes were incubated at room temperature for 20 minutes before being read at 480 nm using a spectrophotometer. A standard curve was generated in Microsoft® Excel® (v1708) and used to determine sample concentrations.

2.4. Reduction, alkylation and digestion

Reduction and alkylation of proteins is an essential step in increasing the efficiency of a bottom-up proteomics workflow. Reduction is done using a thiol-containing reagent, such as DTT, which results in the reduction of disulfide bonds and thus protein unfolding (Konigsberg, 1972). This is followed by alkylation of free cysteines by reagents such as iodoacetamide (IAA). Collectively this results in a protein that is unfolded and less prone to refolding and/cross-linking making it easy for digestion to occur using digestion enzymes.

The digestion of proteins into peptides is a critical step in bottom-up proteomics experiments. Numerous enzymes can be used such as trypsin, pepsin and endoproteinase Lys-C, among others. The most widely used digestion enzyme is trypsin. This enzyme cleaves proteins at the carboxyl (C) terminus of lysine and arginine (Olsen, Ong and Mann, 2004; Hustoft *et al.*, 2012). The enzyme has become the preferred enzyme in modern proteomics due to its ability to produce peptides of smaller average length and more importantly tryptic peptides are more suited to the vast majority of bottom-up proteomic experiments conducted (Olsen, Ong and Mann, 2004; Hustoft *et al.*, 2012; Tsiatsiani and Heck, 2015). Moreover, because trypsin cleaves proteins at arginine and/or lysine amino acid residues, it allows at least one positive charge per peptide. This helps the ionisation process (Hustoft *et al.*, 2012).

The ratio of trypsin to protein is crucial to ensure that proteins are efficiently digested i.e. to ensure that many peptides are produced with minimal missed cleavages and autolysis of trypsin. Thus, it is critical to quantify sample protein accurately and precisely prior to digestion.

For this experiment varying trypsin to protein ratios and a novel kit-based digestion method was assessed as follows:

2.4.1. Digestion using sequencing grade modified trypsin

Protein samples, quantified using the Bradford assay, were subjected to in-solution digestion using trypsin to protein ratio 1:20. Proteins were first reduced using 10mM DTT at 40 °C for 30 minutes and then alkylated using 30mM IAA at room temperature for 30 minutes in the dark. Finally, trypsin was added to the reaction and incubated at 37 °C for 16 hours.

2.4.2. Digestion using the SMART Digest™ kit

The SMART Digest™ kit developed by ThermoFisher Scientific (Waltham, Massachusetts, United States) is a kit-based digestion method that is fast and simple with claimed higher reproducibility than traditional digestion methods due to the ability to be automated (Barattini and Humphryes, 2015).

Briefly, 50 µL of sample was added to 150 µL SMART digest buffer. The final reaction volume of 200 µL was then added to a microtube (supplied in kit). The tube was then placed in a heating block with shaking at 70 °C, 1400 rpm for 4 hours (depending on the complexity of the sample). Following digestion, the sample was cleaned up using a commercially available kit i.e. SOLAµ™ SPE Plates (see Section 2.5.2).

Peptides from the various methods were analysed using a generic LCMS method since no LCMS method optimisation had been performed at this stage of the study. Data analysis was conducted on the total number of peptides and protein groups identified as well as missed cleavages for each enzyme to protein ratio and for the kit-based method to determine the optimum method for all further experiments.

2.5. Peptide sample clean-up optimisation

Most peptide samples require some form of concentration and clean-up post digestion and prior to analysis by LCMS. This is due to the samples often containing sufficient quantities of salts and/or detergents to cause interference with analyte ionisation and may lead to increased noise and/or retention time shifts in the mass spectrum as well as column and instrument contamination (Gundry *et al.*, 2009; Tubaon, Haddad and Quirino, 2017). There are numerous ways to achieve sample clean-up either online (via LC) or offline using commercially available kits. The most widely used offline clean up protocols employ reverse-phase (C18), SPE in various formats such as tips, 96-well plates and spin columns. In general, the chemistry is the same regardless of the format type and brand of reverse-phased clean-up method used. In this experiment various commercially available kits were used as a post digestion clean-up method. These are listed (see Table 2.1) and a general methodology is explained below followed by kit specific methodologies thereafter.

The cartridge/well/tip is first washed with 80 - 100% ACN followed by a wash with 0.1% formic acid (FA). The acidified sample (acidified to 0.1% FA) is then added to the column and allowed to bind to the C18 material. The flow-through (containing unbound peptide) may be

re-applied to the material to facilitate higher peptide recoveries. The column is then washed with 0.1% FA to thoroughly remove any residual salts and ionic detergents. Finally, the bound peptides are eluted from the material using a high concentration ACN (>70%) in 0.1% FA. The peptide solution is then vacuum dried and stored at -80 °C until further analysis by LCMS.

2.5.1. Nest group 96-well MiniSpin clean up protocol

The plate was gently tapped to settle the C18 material at the bottom of the wells. Protective foil was removed from required number of wells. The entire plate was placed into a collection plate and 200 µL 100% ACN was added per well and centrifuged at 110 x *g* for 1 minute at RmT. This was repeated with 100% LCMS water to equilibrate the wells. Thereafter the peptide samples were added to the centre of the well and centrifuged at 110 x *g* for 1 minute and flow-throughs were discarded. The plate was then transferred to a new collection plate and 100 µL 80% ACN in 0.1% FA was added and centrifuged at 110 x *g* for 1 minute. The resulting eluate containing cleaned peptides was then vacuum dried and stored at -80 °C until further use.

2.5.2. SOLAµ™ SPE Plates protocol

The SOLAµ™ SPE Plates uses a 96-well plate vacuum manifold (PlatePrep 96-well vacuum manifold, Sigma-Aldrich). The vacuum was set between 750 – 800 bar for all steps in the protocol. Briefly, the required number of wells were conditioned using 200 µL 100% ACN and thereafter equilibrated with 200 µL 0.1% FA. The sample was then added to the wells and allowed to bind to the C18 material by applying the vacuum on lower strength (650 – 700 bar). Wells were then washed twice with 0.1% FA. Lastly, the peptides were eluted from the wells using 70% ACN. The resulting eluate containing cleaned peptides was then vacuum dried and stored at -80 °C until further use.

2.5.3. C18 Stage Tip clean up protocol

The stage tip clean-up protocol uses a unique pipettor for the application. The pipettor was set to 10 µL. The sample was adjusted to 0.1% FA. Thereafter, 10 µL of 50% ACN was twice aspirated into the tip to wet the C18 material. Thereafter the tip was equilibrated twice using 10 µL 0.1% FA. Thereafter the sample was aspirated, 10 µL at a time, into the C18 tip. The tip was then washed with 10 µL 5% ACN in 0.1% FA. Finally, the bound peptides were eluted into a clean tube using 10 µL 95% ACN in 0.1% FA. The resulting eluate containing cleaned peptides was then vacuum dried and stored at -80 °C until further use.

Table 2.1: Commercially available clean-up kits that were assessed. These were used as potential clean-up methods post digestion. The supplier and product code are also shown.

Product description	Supplier (Catalogue Number)
96-Well MiniSpin 300Å C18, 10-100 µl	Nest Group (Southborough, MA, USA) (SNS SS18V)
SOLAµ™ SPE Plates	ThermoFisher Scientific (60209-001)
C18 Stage Tips	ThermoFisher Scientific (89870)

The Norgen urine protein concentration, preservation and isolation kit was also able to remove salts and other contaminants contained in crude urine during the elution procedure. Hence crude isolates (from the Norgen kit) were also assessed directly on the mass spectrometer. As an additional attempt to remove any trace salt and other contaminants not removed by the kit, acetone precipitation was done post elution as previously described (see Section 2.2.4). Thereafter the precipitated proteins were resuspended in 50 mM AmBic (pH 8.0) and reduced, alkylated and digested as previously described earlier (see Section 2.3).

Peptides were analysed using a generic LCMS method since no LCMS method development had been performed yet. The total number of peptides and proteins identified using each method was then assessed as a measure of the optimal sample preparation method to be used in further experiments. Further, visual assessment of the total ion current chromatogram was done to assess the total peptide recovery using each clean up method. Additionally, monitoring back pressure in the LC system was done continually as a measure of potential contamination as well as column damage

2.5.4. Online C18 clean-up

An additional clean-up method was assessed via online desalting using a C18 trapping column. The trapping column is installed such that it is placed before the analytical column in the flow path of the LC. The main roles of the trapping column are to concentrate and desalt samples and thereby allow for better chromatography and data generation (Mitulovic *et al.* 2003). Here, peptides were derived as described (as follows: Sections 2.2.5. → 2.3.2. → 2.4.1.) but not desalted using any offline method (above). The sample was, instead, analysed directly via LCMS using a) Acclaim PepMap® 100 (100 µm x 2 cm, C18, 5 µm, 100 Å) trapping column (ThermoFisher™ Scientific, Waltham, Massachusetts, USA) and b) directly without C18 trapping.

2.6. Assessment of novel methods developed in the study vs current methods for urinary proteome sample preparation and profiling

A novel, one-step clean-up and digestion, microsphere-based method for sample preparation was developed as an alternative to the Norgen kit-based approach as described in Section 2.5.1 below. Thereafter, these 2 methods (Norgen kit and microsphere) were compared to the current methods for urinary proteome isolation and profiling which employ acetone precipitation followed by filter assisted sample preparation (FASP) according to the method of Wisniewski *et al.* (2009) (see Figure 2.2 for overview). The aim of this experiment was to assess the ease of use, cost and, most importantly, the method that generated the most comprehensive data.

Donor urine was obtained following informed consent [Ethics reference: #58/2013 (CSIR-REC) and #120612 (WITS-HREC)] and subjected to sample preparation using the Norgen kit as described previously (see Section 2.1.5), microsphere-based method (Section 2.5.1. below) and FASP (Section 2.5.2 below). All samples were analysed using the optimised LCMS settings described Section 2.14. Peptide and protein identifications at 1% FDR were used to determine the optimal method to use on clinical samples.

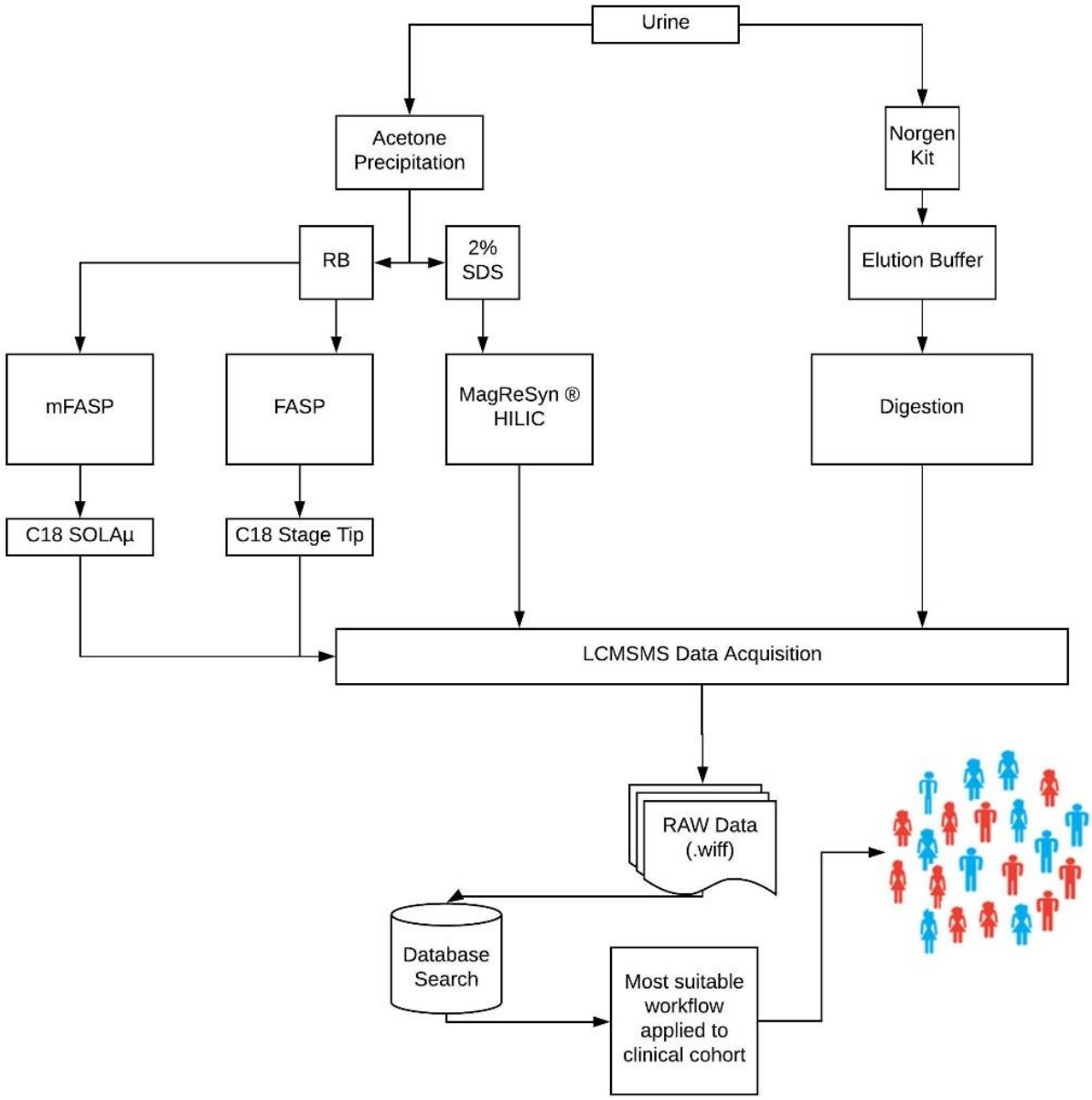


Figure 2.2: Overview of the comparison between the Norgen Kit, HILIC and (m)FASP methods for urinary proteome profiling. Commonly used methods for urinary proteome sample preparation were compared to the Norgen kit-based method and a novel HILIC method using a DDA workflow.

2.6.1. Novel, one-step clean-up and digestion, microsphere-based method for sample preparation

For samples not collected using the Norgen Kit, a novel one-step clean-up and digestion microsphere-based method for sample preparation was developed using MagReSyn® HILIC (hereafter referred to as HILIC) manufactured by ReSyn Biosciences (Edenvale, Gauteng, South Africa).

Protein resuspended in 2% SDS after acetone precipitation (as described in Section 2.1.1.3) was subjected to on-bead clean-up and digestion using HILIC beads. Briefly, 20 µg protein, quantified using the BCA assay, was reduced and alkylated (as described in Section 2.3). Thereafter, the protein was mixed with an equal volume of HILIC binding buffer (30% ACN/200 mM Ammonium acetate). HILIC beads (200 µg) were prepared according to the manufacturer's instructions. Briefly, beads were washed twice with 15% ACN/100 mM ammonium acetate (pH 4.5) and supernatant decanted using a magnetic stand. The reduced and alkylated proteins (as per Section 2.4) were added to the beads and allowed to incubate at RmT for 30 min on an Intelli-Mixer RM-2 (ELMI, Latvia) with mixing (to ensure beads did not settle); the supernatant was then removed. The beads were washed using 95% ACN and the supernatant removed. Finally, on-bead protein digestion was done using sequencing grade modified trypsin (Promega, Madison, Wisconsin, USA) in 20 mM AmBic (pH 8.0) at an enzyme:protein ratio of 1:10 for 4 h at 37 °C on an Intelli-Mixer RM-2 as described above. The protein digest (supernatant) was separated on a magnetic stand twice to remove beads. The peptides were acidified to 0.2% FA (v/v) and subjected to LCMS analysis using the optimised LCMS settings as described in Section 2.14.

2.6.2. Sample preparation using FASP

Protein isolated using precipitation and resuspended in RB was subjected to FASP as described by Wisniewski *et al.* (2009) with minor modifications. The standard FASP kit which uses a 30kDa MWCO filter was used in all experiments. Briefly, a volume of sample equal to 20 µg of protein was added to Urea Sample Solution (USS) as opposed to 30 µL as described by the manufacturer. The ratio of sample protein to USS was kept constant according to the manufacturer's instructions. The protocol was amended to use a trypsin:protein ratio of 1:20. All other steps were followed as per the manufacturer's instructions. Eluted peptides where then subjected to C18 Stage Tip (ThermoFisher Scientific, Waltham, Massachusetts, USA), clean up as per the manufacturer's instructions, vacuum dried and resuspended in 2% ACN,

0.2% FA and analysed by the optimised LCMS settings described in Section 3.9 in the next chapter.

A modified FASP (mFASP) method utilised 60 µg of protein (in RB) directly with the FASP kit without diluting in USS. The trypsin to protein ratio was kept at 1:20. All other steps were performed according to the manufacturer's instructions. Eluted peptides where then subjected to C18 SOLAµ™ (ThermoFisher Scientific, Waltham, Massachusetts, USA) clean up as per the manufacturer's instructions, vacuum dried, resuspended in 2% ACN, 0.2% FA and analysed by the optimised LCMS settings described in Section 2.14.

2.6.3. Peptide and protein properties comparison

Peptides and proteins from the Norgen, mFASP and HILIC methods were compared using open-source online-based software packages ExPASy (Gasteiger *et al.*, 2003) for isoelectric point (p*I*) and molecular weight analyses and protein GRAVY calculator (Stothard, 2000) for GRAVY score analyses. These analyses aimed to determine if any bias was introduced by the preparation methods. The data was then exported into Microsoft® Excel® (v1708) and plotted as normalised (normalised to total peptide/proteins identified in each method, respectively) frequency distributions.

2.7. LCMS/MS data acquisition method development

Microflow LCMS has gained traction as over the past few years as an alternative to nanoflow LCMS. Increasing the flow rate allows for a significant increase in ease of use as compared to nanoflow since flow path connections and lines are easily installed and maintained. Furthermore, leaks are easily detected and fixed leading to increased uptime of the LCMS system. Additionally, increasing the flow rate allows for faster loading times and greater throughput without sacrificing sensitivity at the expense of using more solvents (Christianson, Johnson and Needham, 2013; Needham, Williams and Bell, 2016). Collectively, these advantages allowed for a more robust LC setup which in turn allowed for the no-clean (desalting), no-trap workflow used in this study.

2.7.1. Gradient elution

Gradient elution is a technique that is employed to separate complex mixtures of analytes. The schematic below of a typical gradient elution shows the various parameters that can be optimised to ensure that the best possible separation of analytes is achieved.

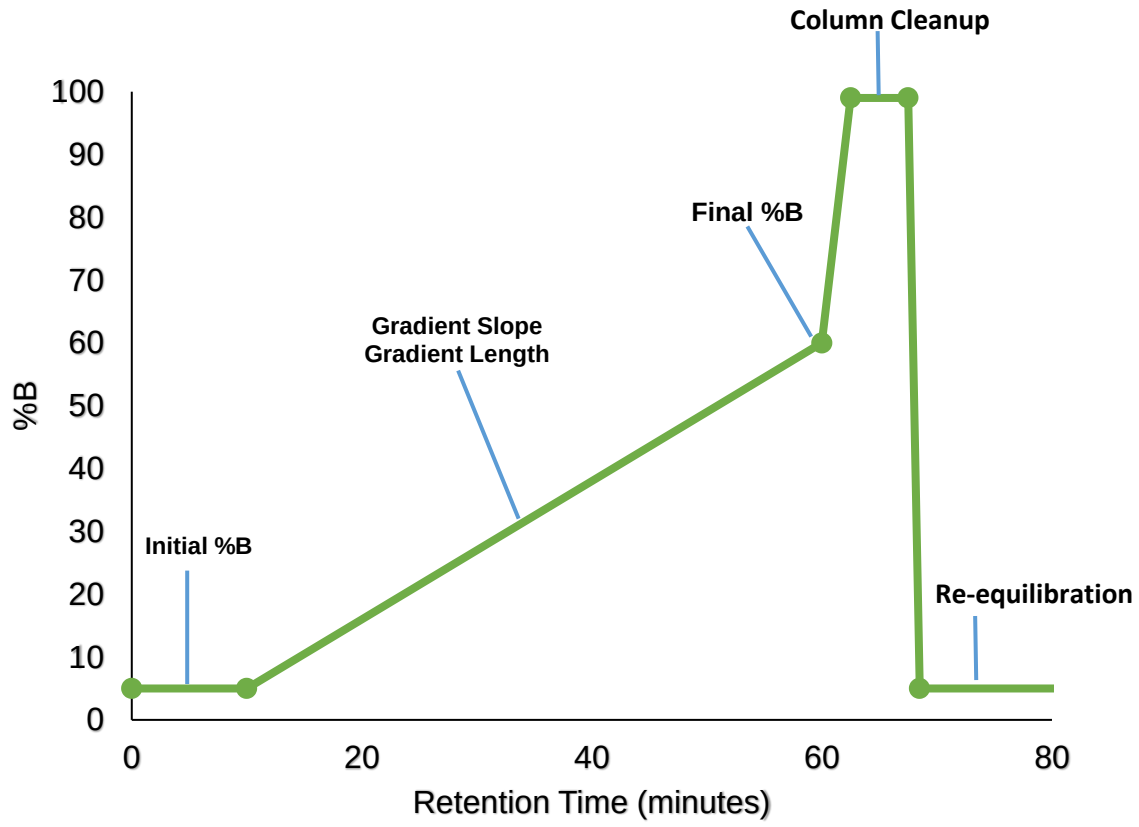


Figure 2.3: A generic gradient elution profile. Initial and final %B, gradient length and gradient slope are shown. Each of these parameters can be tweaked to achieve optimal resolution leading to greater proteome coverage whilst balancing throughput.

2.7.1.1. Gradient length

The optimal gradient length for efficient separation of peptides must delicately balance total ion signal, peak shape and total run time. This balancing act is necessary to achieve the highest quality data in the least time to enable efficient use of LC time and consumables as well as MS time. In this study, two gradient lengths (in time) were assessed whilst keeping the slope and injection load constant. Protein and peptide identifications at 1% FDR as well as peak shape properties i.e. full width at half maximum (FWHM) were used to determine the optimum gradient length. A Mann-Whitney test was conducted in GraphPad Prism 7.0 (GraphPad Software, La Jolla, California, USA) on the FWHM data to determine if there was a statistical difference between median peak shapes.

2.7.1.2. Gradient slope

Building on the data from Section 2.7.1.1 (above), the gradient length that generated the greatest number of peptides and proteins at 1% FDR was selected for further optimisation. Varying the slope of the gradient allows for better chromatography i.e. separation of peptides on the column thereby leading to higher quality data. In this experiment, the slope of the 60-minute gradient, i.e. the rate of ACN delivery to the column per min, was assessed. Gradient slopes were labelled A: 5-30 %B, B: 5-25 %B and C: 5-20 %B.

2.7.1.3. Column load (µg) per injection

Having optimised the gradient length and slope, the next experiment assessed the peptide amount loaded onto the column per injection. The amount of protein (peptides) loaded per injection onto the column can greatly affect the quality of data that is derived. There exists an optimal point where just enough protein is loaded to generate high quality chromatographic peak shapes, thereby leading to greater quality data, without over/under loading the column and skewing the data. In this experiment, peptide samples were prepared as described thus far and four, six and eight micrograms of digested protein (peptides) was loaded per injection. The data was assessed by comparing peptides and proteins at 1% FDR.

2.8. MS data acquisition parameters for data dependent analysis

Numerous studies have focussed on the development of optimised MS acquisition parameters to enable thorough and reproducible data collection (Kalli *et al.*, 2013; Bruderer *et al.*, 2017). Once the complete LC method had been optimised, the MS acquisition parameters were assessed and optimised to ensure that the data derived from the sample was comprehensive, reproducible and without bias. The source parameters were optimised regularly using a standard peptide mix and maintained by performing automated instrument calibrations between samples. The main MS parameters that were optimised were accumulation time, number of MS/MS spectra acquired and collision energy (CE) parameters.

2.8.1. MS/MS and accumulation time

The number of data dependent MS/MS scans can be user defined and plays a definitive role in the quality of the data that is extracted, particularly in proteomics experiments (Andrews *et al.*, 2011; Kalli and Hess, 2012). The number of data-dependent MS/MS scans can be set by the user and, generally, an increased allows more precursor ions to be sampled for MS/MS fragmentation. The actual number of MS/MS scans performed can be optimised depending on

the cycle time and the number of candidate peptides (Kalli *et al.*, 2013). Cycle time is calculated as MS accumulation time + MS/MS accumulation time multiplied by number of MS/MS scans (Chernushevich, Loboda and Thomson, 2001). Chromatographic peak widths play an important role in the determination of cycle time such that sufficient data points per peak are achieved which in turn affects data quality. Kalli *et al.* (2013) showed that 12-15 MS/MS scans improved proteome coverage compared to, their previously used, 5-10 MS/MS scans on a LTQ-Orbitrap instrument.

Accumulation time refers to the amount of time spent performing MS/MS scans on a single precursor that had passed the minimum threshold intensity required for MS/MS activation (Hudson *et al.*, 2014). Once a precursor ion triggers a MS/MS event, the time spent acquiring MS/MS scans of that precursor can be adjusted to extract the most meaningful data from the ion.

In this experiment the MS/MS scans (40 vs 25) and accumulation time (50 vs 90 msec) was varied within the acquisition program parameters. Data was acquired from triplicate samples, and protein and peptides identified was assessed.

2.8.2. Default Analyst® vs crowdsourced optimised rolling collision energy (CE) parameters

Optimal CE parameters are essential to acquiring and subsequent analysis of MS/MS data. The fragmentation patterns that are observed are determined by the CE that is set at an instrument level. Recent advancements have allowed for the development of linear equations (based on mass and charge) to predict the CE required to efficiently fragment peptides (MacLean, Daniela M. Tomazela, *et al.*, 2010). The equation [|CE| = (slope) x (m/z) + (intercept)] was used for rolling CE. The table below (Table 2.2) shows the factory (default) and the crowdsourced CE parameters that were assessed in this experiment. The crowdsourced parameters were derived by SCIEX (Framingham, Massachusetts, USA) from multiple users of SCIEX mass spectrometers.

Table 2.2: Collision energy parameters. The default and crowdsourced CE values used in the equation for rolling CE are shown

	Factory Default		Crowdsourced	
Charge State	**Slope**	**Intercept**	**Slope**	**Intercept**
Unknown	0.0625	-3	0.049	-1
1	0.0575	9	0.5	5
2	0.0625	-3	0.049	-1
3	0.0625	-5	0.048	-2
4	0.0625	-6	0.05	-2
5	0.0625	-6	0.05	-2

The two sets of conditions above were assessed using a single sample in duplicate injection per condition. The data was imported into GraphPad Prism 7.0 (GraphPad Software, La Jolla, California, USA) and total proteins and peptides identified at 1% FDR was plotted.

2.9. Additional parameters for data dependent analysis

Other parameters used in DDA analyses for MS and MS/MS that are essential to achieving comprehensive data are discussed briefly below (See Table 2.3).

For proteomics, it is important to filter out particular charge states of ions and ions with low intensity. Typically, singly charged species and unassigned charge states are excluded from analysis (Kalli *et al.*, 2013). Following tryptic digestion, electrospray ionisation results in multiply charged species i.e. large peptide species (Cole, 2000; Ho *et al.*, 2003; Banerjee and Mazumdar, 2012). During database searching the threshold for a peptide sequence to be identified confidently is often set to 7 amino acids and thus small peptides are normally excluded. Furthermore, many contaminants are singly charged and thus exclusion of singly charged species also reduces contaminants from being sampled and further analysis. By excluding all these ions from further analysis, it gives the mass spectrometer more time to focus on sampling truly informative ions (Kalli *et al.*, 2013).

Another parameter that has become increasingly important in DDA proteomics is dynamic exclusion. This is a parameter that enables a mass spectrometer to avoid repeated measurements of the same ions for a defined time period thereby allowing new precursors that would have otherwise been missed, to undergo further analysis by MS/MS.

Table 2.3: Additional parameters optimised for DDA analysis.

Parameter	Value
Mass range	360 – 1500 Daltons
Minimum intensity	200
Charge states	2 - 5
Dynamic exclusion	25 seconds

2.10. MS data acquisition and data processing parameters for data independent (SWATH™) analysis

2.10.1. Constant vs variable windows

As previously explained in Chapter 1, constant or variable windows (C/VW) can be used in SWATH™ data acquisition. In this experiment, one urinary protein sample was prepared using each of the optimised sample preparation methods and was analysed multiple times using differing SWATH™ acquisition parameters. This experiment aimed to determine the optimal conditions for SWATH™ data acquisition in terms of constant or variable windows. Acquisition methods were setup using Analyst® TF 1.7.1 software for constant windows (32 windows of 27 Da each with 1 Da overlap across 360 – 1100 mass range) and using SWATH™ Variable window calculator V1.0 and Analyst® TF 1.7.1 for variable windows (32, 80 and 100 variable windows across 360 – 1100 mass range).

Data acquired from each method was analysed in parallel using Spectronaut™ 11 (Biognosys Inc, Beverly, Massachusetts, USA) software (See Chapter 2B for details) using a library generated from all data acquired during DDA acquisition optimisation.

2.10.2. Two-dimensional reverse phase-reverse phase (2D RPRP) LCMS data dependent analysis for SWATH™ library generation

Studies have shown that building large, project specific spectral libraries leads to greater depth of proteome coverage in SWATH-MS experiments (Zi *et al.*, 2014; Dong *et al.*, 2016; Govaert *et al.*, 2017). Doing so requires running multiple samples with a broad biological variation to capture a wide range of detectable proteins. The urinary proteome spans a large dynamic range of protein concentrations, ~5 orders of magnitude (Yu *et al.*, 2014; Thomas *et al.*, 2016), and variability in the highly abundant protein uromodulin (Rampoldi *et al.*, 2011). Therefore, fractionation of the proteome was required to ensure deeper proteome coverage. Depletion of uromodulin was not performed since numerous studies have reported uromodulin as a potential

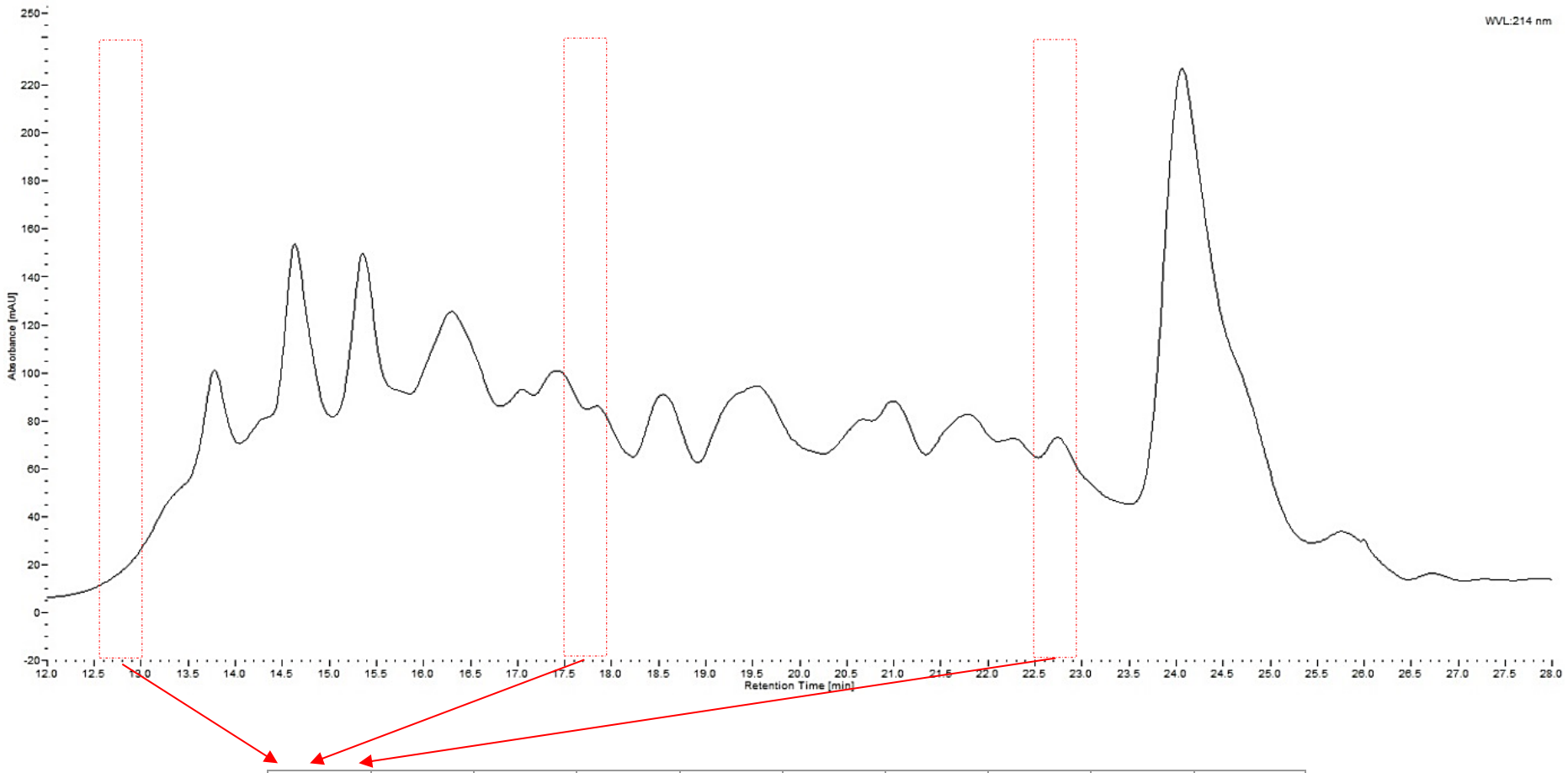
biomarker for kidney dysfunction (Wai-Hoe *et al.*, 2009; Iorember and Vehaskari, 2014; Devuyst and Bochud, 2015).

In this study, libraries generated from 1D RP DDA runs were compared to libraries generated from 2D RPRP DDA. Using 2D RPRP allows for simplification of the proteome thereby leading to greater protein identifications at the sacrifice of MS time. A single sample was run in triplicate in DDA mode and thereafter the same sample was subjected to high-pH reverse phase chromatography with fraction concatenation as described below.

2.10.3. Two-dimensional reverse phase-reverse phase methodology

High-pH reverse phase chromatography with fraction concatenation, using a Dionex UltiMate™ 3000 RSLC, was adapted from Wang *et al.*, (2011), Yang *et al.*, (2012) and Kulak *et al.*, (2017) with minor modifications. An Acclaim™ PA II column (1.0 mm x 15 cm, C18, 3 µm, 120 Å) [(ThermoFisher Scientific, Waltham, Massachusetts, USA)], was used as the stationary phase. The solvents consisted of 20 mM ammonium hydroxide (pH 9.6) as mobile phase A and 20 mM ammonium hydroxide (pH 9.6) in 80% ACN as mobile phase B. The separation was performed at an elution rate of 50 µL/min using a linear gradient as follows: 4% B for 5 min, from 4% B to 10% B in 0.1 min, from 10% B to 60% B in 35 min, to 90% B in 5 min, and thereafter held at 4% B for 15 min. Fractions were collected by time from 12.5 min to 27.5 min every 30 s. Thirty fractions were collected in total and concatenation was performed as per the example shown in Figure 2.4.

Peptide fractions were vacuum dried and then resuspended in 0.2% FA. Thereafter each fraction was analysed on a Dionex UltiMate™ 3000 RSLC (ThermoFisher Scientific, Waltham, Massachusetts, USA) in microflow configuration (Figure 2.5). An Acclaim™ PA II column (300 µm x 15 cm, C18, 3 µm, 120 Å) (ThermoFisher Scientific, Waltham, Massachusetts, USA), was used as the stationary phase. The solvents consisted of 0.1% FA as mobile phase A and 80% ACN in 0.1% FA as mobile phase B. The separation was done at a mobile phase flow rate of 8 µL/min using a linear gradient as follows: 5% B for 3.5 min, from 5% B to 30% B in 60 min, to 60% B in 5 min, and thereafter held at 99% B for 5 min.



1	2	3	4	5	6	7	8	9	10
1,11,21	2,12,22	3,13,23	4,14,24	5,15,25	6,16,26	7,17,27	8,18,28	9,19,29	10,20,30

Figure 2.4: A schematic overview depicting the concatenation strategy. For the first dimension, high-pH reverse phase chromatography, 30 fractions were collected by time, along the LC separation and were concatenated into 10 fractions by combining three sub-fractions as follows: 1, 11, 21; 2, 12, 22; etc.

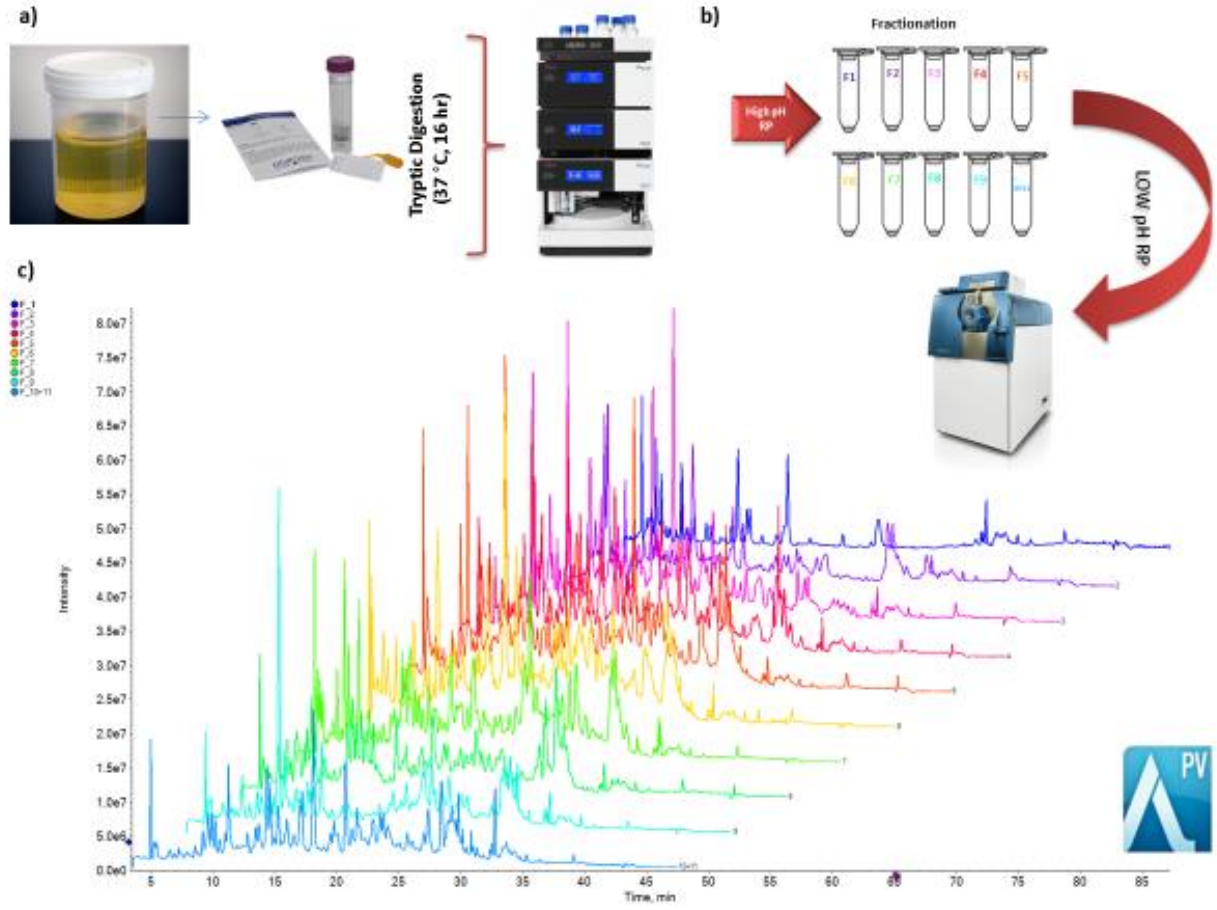


Figure 2.5: Overview of the complete workflow using 2D RPRP. A) Collection and processing of urine followed by digestion, B) 2D RPRP LCMS/MS analysis and C) total ion chromatograms of LCMS data from each concatenated, high-pH separated fraction.

2.11. Spectral library generation settings

Optimisation of the spectral library generation was done to ensure that the highest quality data was extracted from each SWATH™ run. To achieve this, libraries were generated from 1D RP DDA runs of varying length (by time) and 2D RPRP DDA runs - thereby optimising MS time. The use of a project specific library (in-house generated) and a publicly available library was also assessed. Libraries were generated from DDA data using the built-in library generation function in Spectronaut™ 11.

The final libraries used in the optimisation were generated as follows:

1. 30 minute gradient (from 3 x 30 minute DDA runs)
2. 60 minute gradient (from 3 x 60 minute DDA runs)
3. 60 minute gradient (from all patient 60 minute 1D RP DDA runs)
4. 60 minute gradient (from all patient 2D RPRP DDA)
5. Biognosys external urine library
6. Project specific in-house library (Library 3 + 4)
7. Library 5 + 6

Thereafter, the MaxQuant processed data (searched as previously described in Section 2.1) was imported directly into Spectronaut™ 11 (Biognosys Inc, Beverly, Massachusetts, USA) for new library generation.

Three sets of duplicate SWATH™ data files were then assessed using each of the libraries generated (above). The number of peptides and proteins identified at 1% FDR were used as a proxy that enabled the selection of the best library. Raw data (.wiff) was first converted into Spectronaut™ (.HTRMS) file format to enable faster processing. The analysis in Spectronaut™ 11 (Biognosys Inc, Beverly, Massachusetts, USA) proceeded with the default settings. Briefly, retention time prediction was set to dynamic iRT and correction factor for window was set at 1. Local mass calibration, MS1 and MS2 level interference correction was enabled. Cross run global normalisation, based on the algorithm by Callister *et al.* (2006), was enabled. The FDR was set at 1% on both the peptide precursor and protein level.

2.12. Complete workflow technical variation assessment

The complete, optimised workflow (from urine collection to MS analysis) was assessed on three consecutive days to determine the technical variation introduced by the all methods

utilised. A technical CV of $\leq 20\%$ is the current gold-standard for a proteomics workflow to be accepted and utilised without further optimisation (Geyer *et al.*, 2016; Fu *et al.*, 2018).

2.13. Panther gene ontology (GO) analysis

Protein classification was performed using the Panther classification system (version 13.1) (Mi *et al.*, 2017) to determine if the methodology introduced any bias by enriching specific protein classes and pathways. Molecular function, biological process, cellular component, protein class and related pathway were all assessed. Results for protein class analysis and pathway analysis results for all proteins identified using the Norgen Kit method is presented in the next chapter. Significance level was set at $p < 0.05$.

2.14. Summary of data dependent analysis methodology after individual component optimisation.

Peptides were analysed on a Dionex UltiMate™ 3000 RSLC (ThermoFisher Scientific, Waltham, Massachusetts, USA) in microflow configuration. An Acclaim™ PA II column (300 µm x 15 cm, C18, 3 µm, 120 Å) (ThermoFisher Scientific, Waltham, Massachusetts, USA), was used as the stationary phase. The solvents consisted of 0.1% FA as mobile phase A and 80% ACN in 0.1% formic acid as mobile phase B. The separation was done at a mobile phase flow rate of 8 µL/min using a linear gradient as follows: 5% B for 3.5 min, from 5% B to 30% B in 60 min, to 60% B in 5 min, and thereafter held at 99% B for 5 min.

MS/MS spectra were obtained using a TripleTOF® 6600 mass spectrometer (SCIEX, Framingham, Massachusetts, USA) in DDA mode. Eluted peptides were delivered into the mass spectrometer via a DuoSpray® ion source equipped with a 25 µm emitter electrode (SCIEX, Framingham, Massachusetts, USA). Source settings were as follows: Curtain gas - 20, Gas 1 - 16, Gas 2 - 0, temperature - 0 and ion spray voltage – 5500. Rolling collision energy was used for CID fragmentation for MS/MS spectra. Each cycle consisted of a TOFMS spectrum acquisition for 500 milliseconds (m/z 360–1500), followed by 40 MS/MS spectra of 50 milliseconds each (m/z 100–1800) for a total cycle time of 2.5 seconds. MS/MS selection was based on the following parameters: UNIT resolution, mass intensity threshold 200 cps and charge state 2–5. Dynamic exclusion was set to 25 seconds. The mass spectrometer was auto-recalibrated after every injection using a $[Glu]^1$-Fibrinopeptide B standard (Sigma-Aldrich, St. Louis, Missouri, USA).

Chapter 2
Materials and Methods

Part B

Urinary protein biomarker discovery:

An exploratory case-control study

First-line ART is freely available to 90% of HIV positive patients in South Africa, which accounts for ~10% of the global HIV/AIDS burden (Statistics South Africa, 2017). Whilst most patients benefit from ART, it can cause adverse side effects (including AKI) in others. Current tests for AKI base diagnosis on increased serum creatinine levels (Makris and Spanou, 2016; Ostermann and Joannidis, 2016), although this remains an unreliable marker of AKI in early or mild cases where creatinine levels may remain normal despite kidney damage (Waikar, Betensky and Bonventre, 2009). More effective biomarkers are thus required to either identify early renal dysfunction or to identify those with high risk of renal injury following exposure to ART and other drugs. This would allow clinicians to make early interventions, such as a change in ART regimens and/or closer monitoring of patients displaying biomarkers indicative of AKI, to limit further ART-related nephrotoxicity. Given this need, this study aimed to identify putative biomarkers for AKI, specifically in patients undergoing first-line ART.

2.15. Cohort details

Ethical approval was received for consenting patient recruitment and collection of urine specimens for this study [Ethics references: #58/2013 (CSIR-REC) and #120612 (WITS-HREC)].

This exploratory case-control study recruited patients from the Klerksdorp/Tshepong Hospital Complex based in the North West Province of South Africa. HIV positive participants were enrolled from patients admitted to the adult medicine wards, of which a portion was diagnosed with TDF-ART induced AKI (cases). The diagnosis of AKI was based on the Kidney Disease Improving Global Outcomes (KDIGO) 2014 Clinical Practice AKI Guidelines (Machado, Nakazone and Maia, 2014). Adult HIV positive participants were enrolled from patients

undergoing first line TDF-ART that did not suffer from AKI – these participants were enrolled as controls. Patients with pre-existing kidney dysfunction, those unwilling to participate in follow-up visits and those with inadequate patient records to determine prior ART regimens or prior incidents of renal injury/kidney dysfunction were excluded from the cohort. All patients were classified as TDF exposed if they had been on TDF-ART at the time of admission or defaulted for less than one week before admission.

Based on these requirements, 175 participants met the inclusion criteria for the study. However, based on power analysis (see Section 2.15 below), the exploratory case-control cohort comprised of 74 participants. Thirty-seven (37) of these patients were diagnosed with TDF induced AKI and served as the cases for the study. Control participants were gender and age matched (within three years) to the cases. Based on the power analysis, 74 participants, would provide a sufficiently powered study on which further studies could be built. Case symptomology was further classified as either mild-moderate (increase in baseline serum creatinine of 1.5 – 2.9 times greater than baseline at time of admission compared to initiation of ART; n = 12) or severe (increase in baseline serum creatinine $\geq$ 3.0 times greater than baseline at time of admission compared to initiation of ART or patient started dialysis or an absolute creatinine level greater than 353.5 µmol/L; n = 25) (Khwaja, 2012).

2.16. Demographic and clinical data

Detailed demographic and clinical data was collected for each patient in the cohort, by the clinicians and nursing staff, as per the normal requirements during admission to the hospital (See appendix A1). The data included common clinical parameter measurements, such as gender, age, race and serum creatinine level, that are used in the diagnosis and stratification of AKI based on KDIGO (Machado, Nakazone and Maia, 2014). Student's t-tests and Wilcoxon signed-rank tests were performed, where appropriate, on each parameter comparison between case and control groups using GraphPad Prism 7.0 (GraphPad Software, La Jolla, California, USA). A significance level of $p < 0.05$ was used to establish statistically significant differences between groups.

2.17. Power analysis

Power analysis was conducted using ClinCalc software (Kane, 2018), to estimate the patient sample size required for the study, at a power of 0.9 and an alpha of 0.05, when the minimum Log_2 fold-change (effect size) for differentially abundant proteins was set to 2, given an equal distribution of age and gender between matched case-control pairs. Based on the computed results, samples representing a total of 74 patients comprising 37 cases and 37 controls were selected for further analysis.

Table 2.4: Power analysis using ClinCalc software

Study Parameters	
Mean protein abundance, group 1	4
N	37
Mean protein abundance group 2	1
N	37
Alpha	0.05
Beta	0.1
Power	0.9

2.18. Randomised block design for sample analysis

Each case-control matched pair was randomly allocated an experiment number (1 – 37). Therefore, 37 individual experiments were conducted, and the data collated to extract biologically meaningful information. Further randomisation (acquisition randomisation) was performed whilst acquiring data on the mass spectrometer where experimental pairs were analysed in a random order. The clinical cohort was blocked and randomised using Microsoft® Excel® (v1708) to limit technical variation as shown in Table 2.5 below. Blocks (of case-control pairs) were analysed individually i.e. block 1 followed by block 2, block 3 and lastly, block 4. This was done to minimise mass spectrometer induced variation as it is common for a mass spectrometer's performance to change over time. If all samples were run in series, it would result in data that is biased since a sample analysed first would have greater quality compared to those analysed towards the end of the batch. Therefore, between each block, the mass spectrometer was cleaned, retuned and recalibrated. This ensured that the mass spectrometer was preforming optimally before the next block of clinical samples were analysed.

Table 2.5: Randomised block design for patient urinary protein analysis. The identification codes that were used by the clinical team to anonymise each case or control study participant are listed per experiment

Age-Race-Gender Matched Pairs				Age-Race-Gender Matched Pairs			
Block	**Experiment**	**Case**	**Control**	**Block**	**Experiment**	**Case**	**Control**
Block 1	**1**	13	186	**Block 2**	**29**	118	164
	2	91	139		**30**	60	171
	3	113	163		**31**	114	220
	4	121	161		**32**	70	223
	5	7	131		**33**	73	144
	6	58	3		**34**	109	213
	7	6	153		**35**	103	203
	8	108	141		**36**	116	190
					37	98	165

Age-Race-Gender Matched Pairs				Age-Race-Gender Matched Pairs			
Block	**Experiment**	**Case**	**Control**	**Block**	**Experiment**	**Case**	**Control**
Block 3	**19**	66	192	**Block 4**	**9**	120	194
	20	59	157		**10**	15	201
	21	119	137		**11**	1	170
	22	87	167		**12**	11	133
	23	53	221		**13**	111	215
	24	102	195		**14**	56	184
	25	115	147		**15**	83	155
	26	78	129		**16**	67	168
	27	74	4		**17**	110	183
	28	92	166		**18**	72	130

2.19. Summary of methods used in sample preparation and mass spectrometry data acquisition

Each individual patient's urinary proteome data was acquired using the optimal settings developed in this study (shown in Figure 2.6 below). Briefly, proteins were isolated using the Norgen Kit, quantified using the Bradford assay and subjected to in-solution digestion using a 1:20 enzyme to protein ratio. Thereafter, 6 µg peptides per patient (per injection) were subjected to direct DDA LCMS/MS analysis using a 60-minute gradient with 40 MS/MS and 50 millisecond accumulation time for spectral library building. To generate a more comprehensive spectral library, cases and controls were pooled, separately within each block, and subjected to 2D RPRP analysis. Two additional injections (per patient) were done using a SWATH™-MS approach (32 VW and 90 millisecond accumulation time). All samples were spiked with iRT standards before analysis by mass spectrometry.

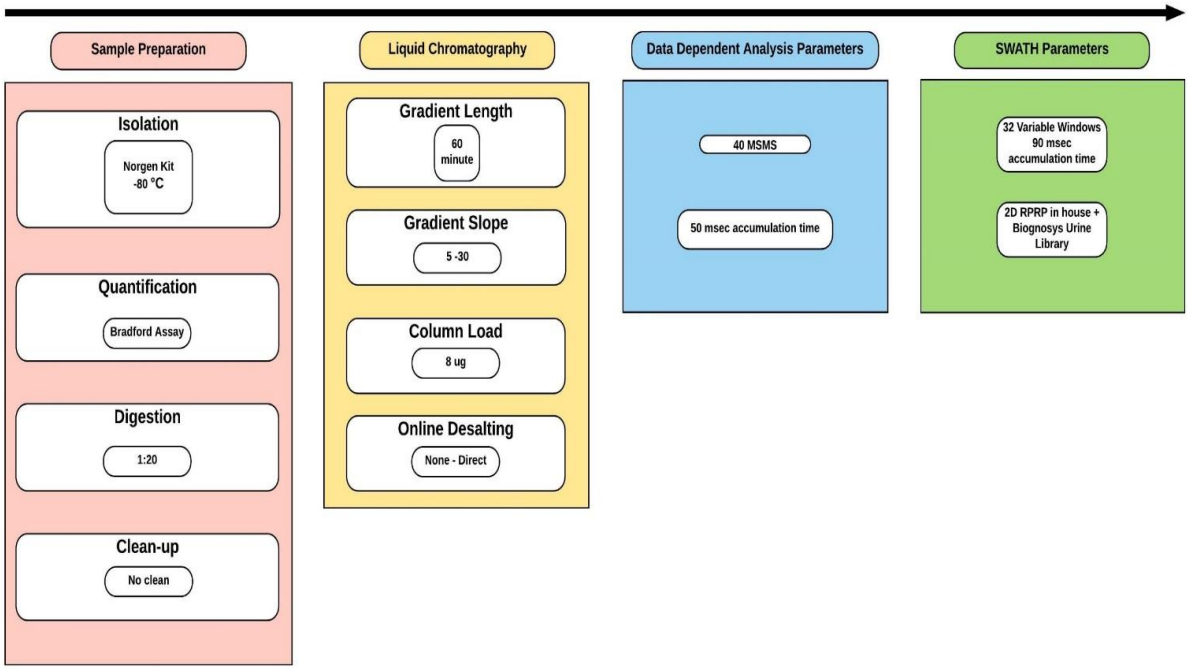


Figure 2.6: Summary of the optimised workflow developed and employed in this study for urinary proteome profiling of clinical samples.

2.20.Processing of patient data using Spectronaut™ 11

Raw SWATH™ (.wiff) data files were converted into Spectronaut™ HTRMS format and then analysed using Spectronaut™ 11. The default settings that were used for targeted analysis are described as follows: dynamic iRT retention time prediction was selected with correction factor for window 1. Mass calibration was set to local. Decoy method was set as scrambled. The FDR, based on the mProphet approach (Reiter *et al.*, 2011), was set at 1% on the peptide level. Peptide quantities were calculated as summed intensities of corresponding fragments of the peptide after interference correction (Bruderer *et al.*, 2017). Protein inference was set to Spectronaut™ default which is based on the ID picker algorithm (Zhang, Chambers and Tabb, 2007).

The study specific, in-house generated (peptides - 8947, proteins – 1391) and the publicly available Biognosys urine (peptides - 11166, proteins - 1765) spectral libraries were used as a reference for targeted data extraction.

Default settings in Spectronaut™ 11 were used for comparison analysis, on protein level, using a t-test (null hypothesis H_0 that no change in protein concentration was observed between the two groups). The t-test was performed on the $\log_2$ ratio of peptide intensities that corresponded to individual proteins. The p-values were corrected for multiple testing using the q-value approach to control FDR (Storey, 2002). The data generated from each experiment was plotted as a volcano plot ($\log_2$ fold change vs q-value), as shown in Figure 2.7, and then only candidate proteins were exported as a table into Perseus v1.6.1.1 (Tyanova, Temu, *et al.*, 2016) for filtering as described in Section 2.20.1 below.

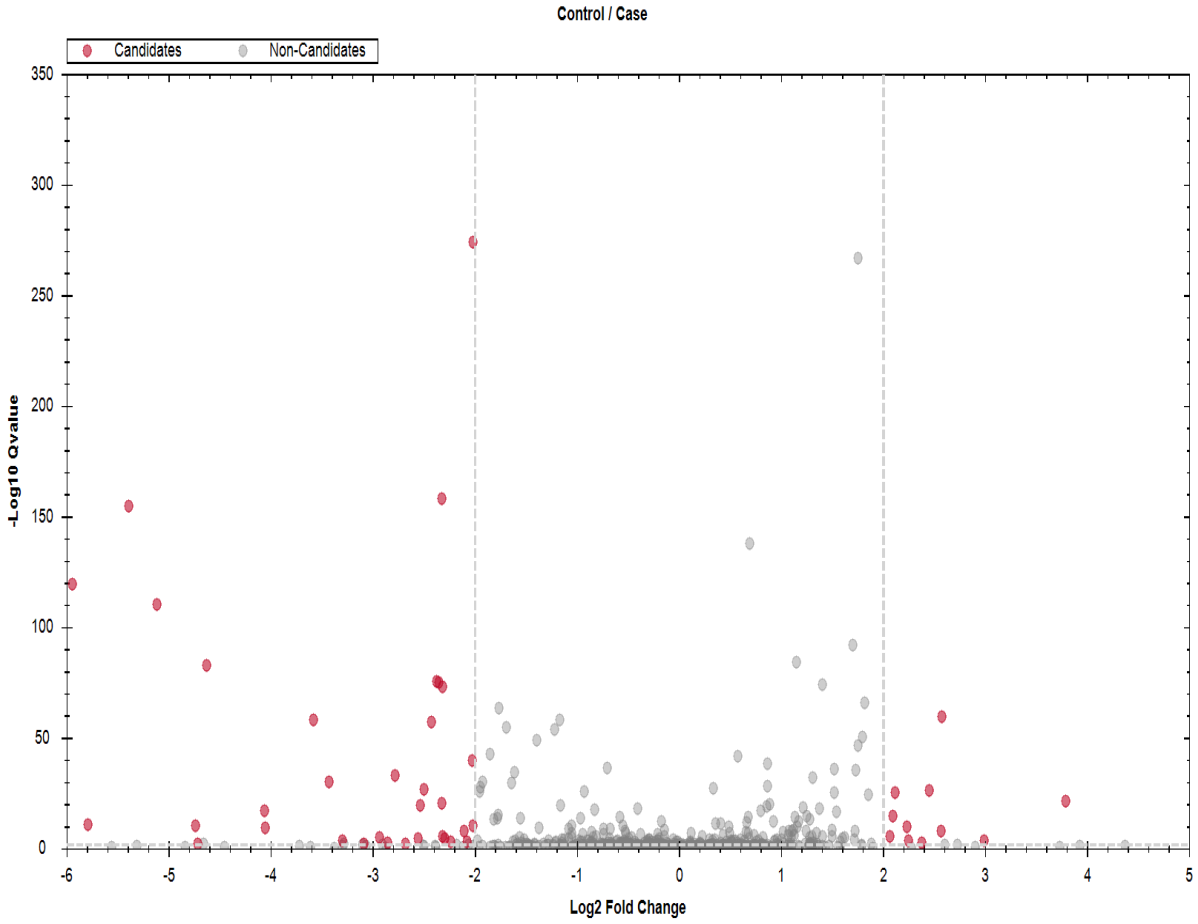


Figure 2.7: Volcano plot of candidate markers after comparing all cases to all controls. Each red point represents an identified candidate marker based on q value ≤ 0.01, Log_2 fold-change ≥ 2 and unique peptides ≥ 2.

2.20.1. Selection of candidate biomarkers

A protein was only considered a candidate biomarker if the following stringent criteria was met: Log_2 fold change greater than or equal to 2 (4-fold difference), q-value less than or equal to 0.01 (1% FDR) and greater than or equal to 2 unique peptides per protein.

Five hundred and twenty three (523) proteins (See appendix A2) out of 1391 proteins in the library met the initial criteria and were further filtered based on the statistical models develop by Skates *et al.* (2013). The model considers: discovery cohort sample size; standard deviations (SDs) of candidates separating cases and controls; number of experiments expressing the candidate protein marker and the number of experiments that will be assessed in the verification phase i.e. where the candidate markers from the discovery study are assessed in a blinded cohort. The limitation of the model is that it was developed on plasma and proximal fluids other

than urine, however, in the absence of other appropriate tools, this served as a useful method to filter the data in this study.

As the number of biological replicates increases, the proportion of cases expressing the biomarker and SDs separating the groups can decrease whilst retaining the same power (Skates *et al.*, 2013). Therefore, in our dataset, only putative protein biomarkers that were present in at least 30%, (12 out of 37), of the experiments were accepted as putative biomarkers. These putative biomarkers will be assessed beyond this study in the next phase of this research project (verification).

2.21. Gene ontology pathway analysis of putative biomarkers

Protein classification according to biological pathway function was performed using Reactome v65 (Fabregat *et al.*, 2018). Pathway analysis was done by comparing the distribution of the set of 52 putative protein biomarkers (listed in Table 3.6 and 3.7) against the entire set of gene ontologies contained within the human database. Significantly enriched pathways (FDR ≤ 0.01) containing an overrepresentation of putative biomarkers from the study were elucidated. The results are presented in in the next chapter.

Chapter 3
Results

Part A

Method development and optimisation for urinary proteome profiling in clinical proteomics

3.1. Urinary protein isolation methods

All precipitation methods were able to extract components out of urine samples as could be seen by a turbid solution after the addition of the organic solvent. These included proteins and other components such as salts. However, protein resuspension was difficult to achieve using AmBic, HFIP and NH_4OAc. Thus, urea buffers and strong detergent-based buffers, such as SDS, were used to achieve resuspension.

The methanol and ACN precipitated samples returned very low protein yields due to low protein recovery after the resuspension of pellets (~0.01 µg/µL). Therefore, these methods were excluded from all further analysis. Acetone and acetone-TCA precipitation showed the greatest yield per mL crude urine of 41.65 µg/mL and 19.0 µg/mL in RB, respectively and 46.6 µg/mL in 2% SDS, respectively. However, due to a small starting volume of crude urine being used, these yields translated to 83.3 µg, 38.0 µg and 92.8 µg of total protein being isolated in acetone (RB), acetone-TCA (RB) and acetone (2% SDS) samples, respectively (Figure 3.1).

Despite precipitation methods being relatively cheap and easy to perform, they all required pre-processing of the urine samples before being subjected to precipitation. At the clinic level (where patient samples are collected), this posed a potential point for increased technical variation. Furthermore, the very low yields and inability to scale up due to cost and capacity make these methods unsuitable for high-throughput biomarker studies. Additionally, the precipitated proteins need to be resuspended in buffers that are incompatible with downstream sample preparation (digestion) and subsequent MS analysis. These require some form of additional clean up thus substantially increasing time, cost and complexity of the sample preparation workflow.

In contrast, the kit-based collection and preservation method was easily implemented at the clinic and enabled a robust and reproducible protocol for patient urine collection. There was no need to pre-process the urine before being used with the kit. Importantly, once the urine was added to the kit, the bound proteins remained stable even after being incubated at 50 °C for one week. This is particularly important in a South African setting where clinics are often under equipped and therefore maintaining the cold-chain can be challenging. Furthermore, due to the remote location of most clinics, adequate and constant electricity supply is not guaranteed further adding to the inability to maintain a cold-chain. The downstream isolation was carried out in a reproducible manner and did not require specialised equipment. Furthermore, it was rapid and multiple samples could be processed in a day (up to 20). The protein yield was the greatest in the sample that was stored at room temperature (25 °C) for one week (17.73 µg/mL) and the lowest was from the sample stored at 50 °C for one week (9.660 µg/mL). However, all the kit-based methods returned greater than 250 µg total protein, irrespective of storage/incubation condition (Figure 3.1). The cost was calculated to be approximately R350.00, for collection and extraction, per sample. While this was the most expensive option, it was a worthwhile investment considering the ease of use, robustness of the kit and large total protein isolation per sample.

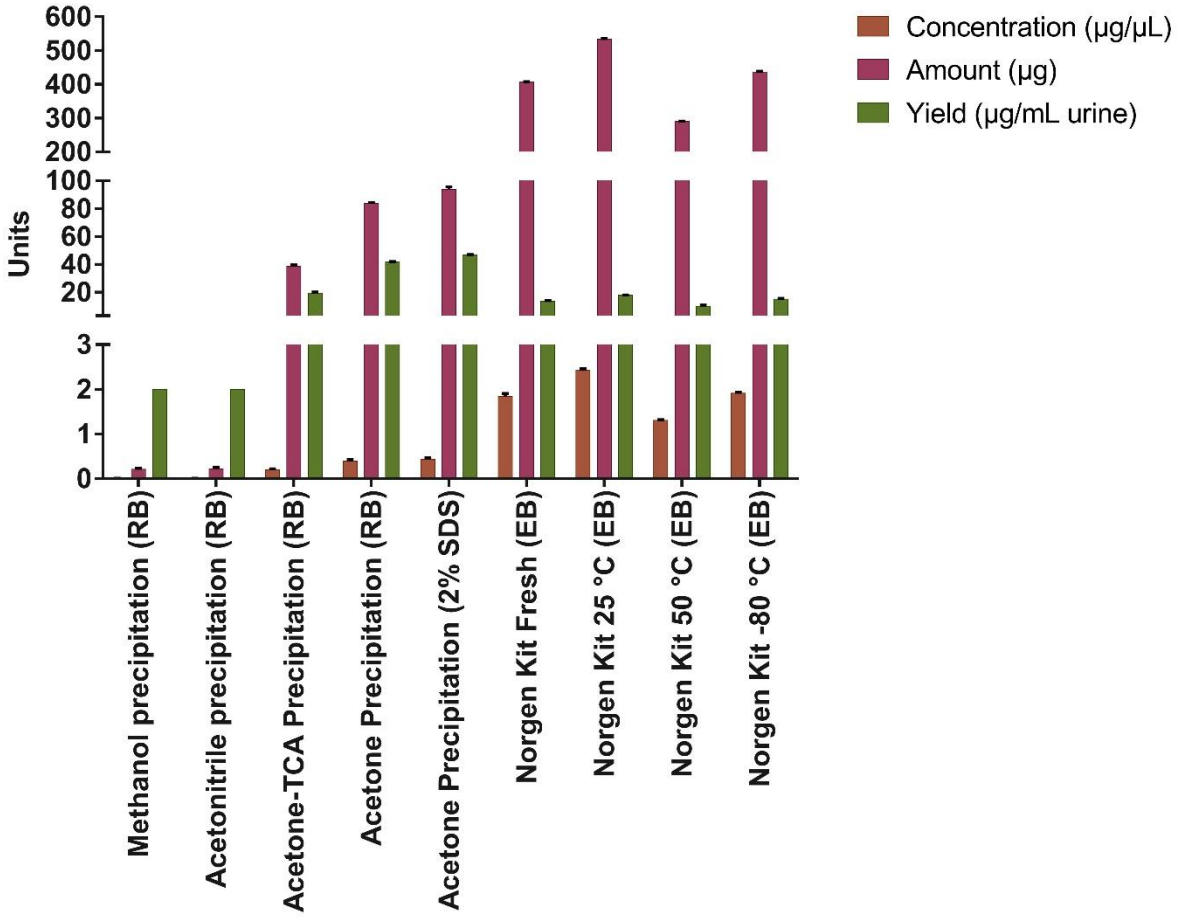


Figure 3.1: Summary of protein yields from various methods. The highest protein concentrations and amounts were obtained when urinary protein was isolated using the Norgen Kit. Furthermore, the kit preserved urinary protein even under extreme storage conditions such as at 50 °C for one week without losing the entire sample. Acetone precipitation showed the second highest yield and methanol and acetonitrile showed the lowest protein yields. RB – rehydration buffer, 2% SDS – 2% sodium dodecyl sulfate, EB – elution buffer.

3.2. Quantification methods

Every assay was able to detect protein in the standard test samples. The results below are expressed as mean estimated protein concentration ± standard deviation over three days.

The Bradford assay estimated mean protein concentrations of 1.03 ± 0.06 µg/µL and 2.05 ± 0.06 µg/µL for the 1 and 2 µg/µL standards, respectively. The assay showed the highest sensitivity and specificity over the 3-day period and was the most accurate method as it detected the 1 and 2 µg/µL BSA standards with the lowest error (Figure 3.2).

The BCA assay also reliably quantified the standard samples however it slightly over estimated the concentration in each sample. The assay overestimated the mean concentrations of both standards at 1.32 ± 0.19 µg/µL and 2.29 ± 0.16 µg/µL, respectively (Figure 3.2).

The GE-2D kit estimated the 2 µg/µL standard sample at 2.04 ± 0.08 µg/µL and overestimated the 1 µg/µL standard at 1.21 ± 0.10 µg/µL (Figure 3.2). The kit performed similarly to the Bradford assay, and accurately and precisely quantified the standards. This method is more labour intensive than the others and serves as a good confirmatory method. It is also a method that is more suited to samples that contain detergents and other interfering compounds that are not compatible with the Bradford or BCA assay.

The Qubit™ quantification method performed the worst and was unable to detect the protein concentration of the standards within a reasonable range. Furthermore, the inter-day variation between measurements was the largest for this method showing CVs of 47.63% and 36.19% for the 1 and 2 µg/µL standards, respectively (Figure 3.2).

Collectively, the results suggest that the Bradford assay is the most suitable due to its ease of use and robustness. Furthermore, the clinical urinary protein samples from this study are eluted in a low molarity salt solution and does not interfere with Bradford assay analysis. As such, the Bradford assay was used to quantify all further samples derived from the Norgen Kit.

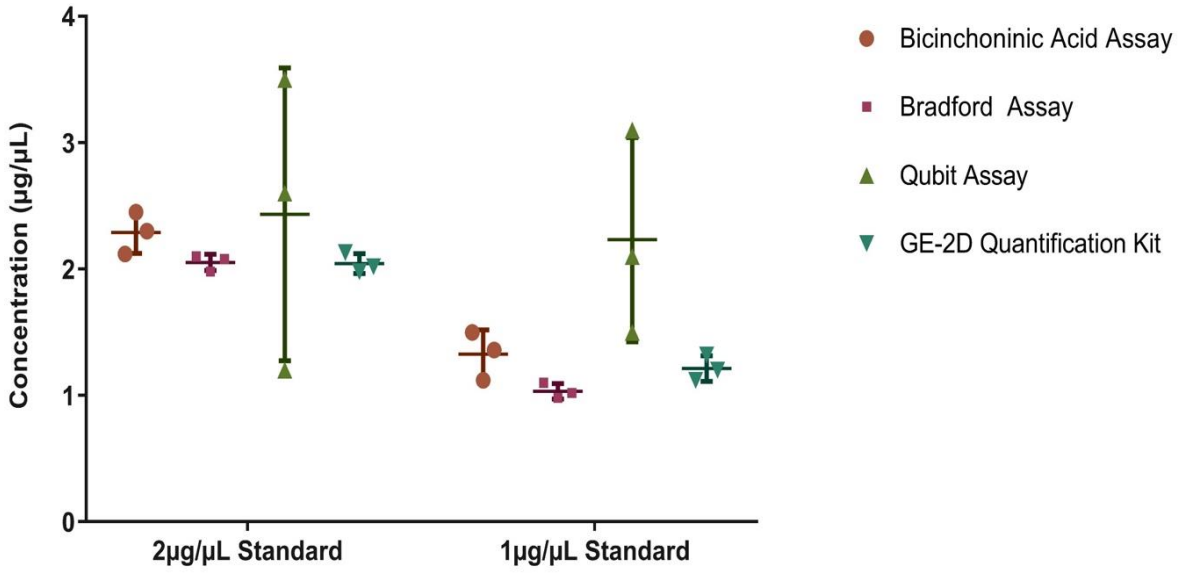


Figure 3.2: Summary of protein quantification methods. The graph above shows the mean of three data points for each quantification method performed on three consecutive days. Two standard solutions of bovine serum albumin, i.e. 1 µg/µL and 2 µg/µL, were used with each method. The Quick Start™ Bradford Protein assay displayed the lowest variability with the highest sensitivity and precision.

3.3. Digestion methods

Digesting urinary proteins using SGMT, at a 1:20 trypsin to protein ratio, outperformed the SMART™ Digest method by 3.4-fold for protein identifications at 1% FDR. Furthermore, the peptide identifications where lower in the SMART™ Digest method by greater than 10-fold. The SMART™ Digest method was easier and faster to perform requiring fewer handling steps (4 hours vs 16 hours for SGMT). Although the SGMT digestion method takes 4-fold longer to complete, it generated more comprehensive data (577 ± 8 proteins; 6007 ± 119 peptides), thus all further experiments used SGMT, at a 1:20 trypsin to protein ratio, for digestion. The SGMT and SMART™ Digest method showed a similar percentage of missed cleavages of 10.8% and 11.5%, respectively (Figure 3.3).

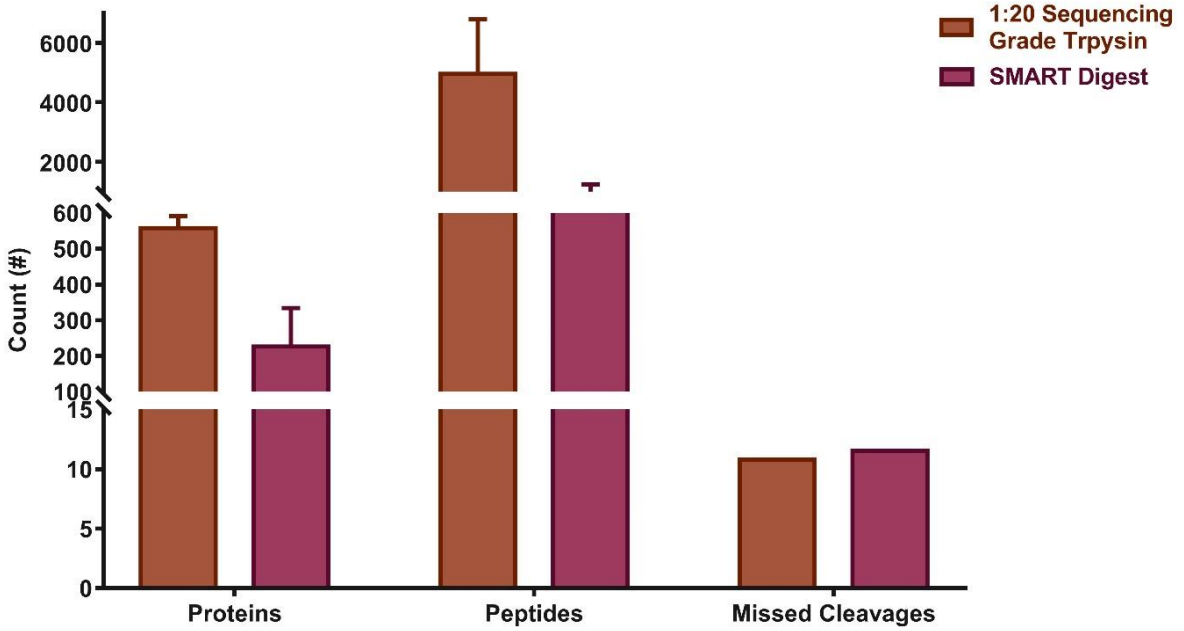


Figure 3.3: Comparison of protein digestion using the SMART™ Digest kit and SGMT. The kit-based method displayed lower protein (3.4-fold) and peptide (>10-fold) identifications (n = 3).

3.4. Commercial, offline C18 clean-up kits

The various kit-based clean-up methods all achieved the goal of removing salts and contaminants from the digested samples, however they all resulted in losses of peptides which translated into lower protein group identifications. Furthermore, the SOLAμ™ kits (ThermoFisher Scientific) displayed high intra-kit and inter-kit variation in peptide identification. Lastly, after using the 96-Well MiniSpin Plate (Nest Group) and the Pierce™ C18 Spin Columns (ThermoFisher Scientific) to clean up samples, the kits leached C18 material into the final eluted sample, which led to blockages in the LC system. The Norgen kit claims to clean up samples during the elution procedure that are directly compatible with LCMS analyses, it was shown that using this approach did not affect the LC system detrimentally, thereby supporting their claims. Furthermore, using the Norgen Kit without further offline clean-up resulted in the most comprehensive data (Figure 3.4).

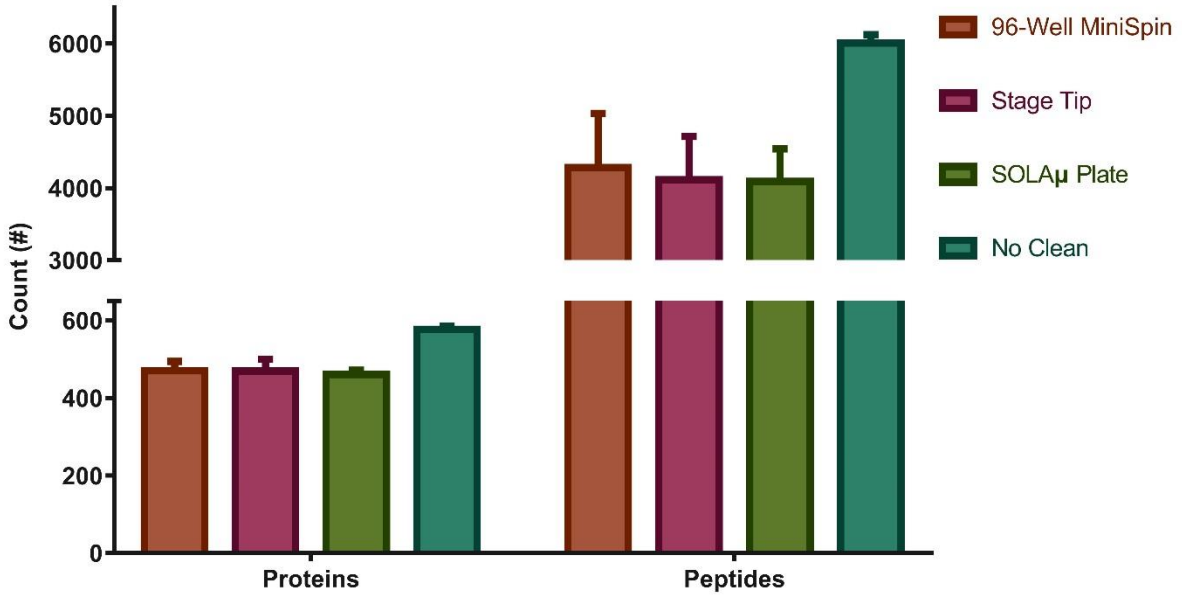


Figure 3.4: Summary of data derived after C18 clean-up. The sample that was not subjected to any clean-up post digestion returned the most comprehensive data (peptides: 6007 ± 119, proteins: 577 ± 8) (n = 3).

3.5. Online clean-up vs no clean-up

Despite trapping columns showing increased signal, decreased contamination and better overall data in other studies (Mitulovic *et al.* 2003; Mitulovic and Mechtler 2006), the data presented here shows decreases in both peptide (by 30%) and protein (by 13%) identifications when a trapping column is used (Figure 3.5). Taking these results into consideration, imparted confidence that the Norgen Kit samples were free from interfering components and therefore this approach was used in all further experiments i.e. no offline clean up post digestion was performed, and samples were analysed directly via LCMS without C18 trapping.

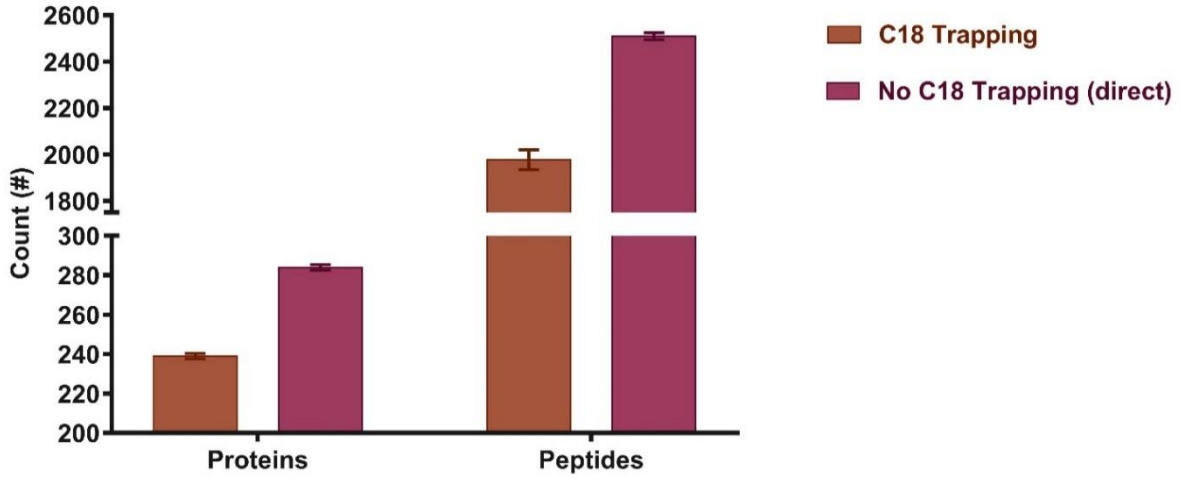


Figure 3.5: Summary of data following online C18 clean-up. The sample that was analysed directly on the analytical column, i.e. without C18 trapping, showed greater peptide and, by extension, protein recoveries (n = 3).

3.6. Norgen kit, HILIC and (m)FASP comparisons

Total peptides and proteins identified from each method are tabulated below. The widely reported acetone precipitation method followed by FASP clean-up and digestion resulted in lower peptide and protein identifications, by 8-fold and 5-fold, respectively, as compared to acetone precipitation followed by HILIC clean-up and digestion. When a modified FASP (mFASP) protocol was used, the number of peptide and protein identifications increased by 4-fold. However, the HILIC method still resulted in the highest data quality (peptides – 2567; proteins – 471). The kit-based isolation methods showed similar peptide (2590) and protein (407) identifications compared to HILIC (Table 3.1).

The acetone precipitation-based methods are less expensive to perform than the Norgen kit-based methods. However, the hands-on time for the Norgen kit-based method is significantly lower and is easier to use than the acetone precipitation-based methods (Table 3.1). Acetone precipitation followed by 2% SDS resuspension with MagReSyn® HILIC on-bead clean up and digestion is the most rapid and inexpensive method. Furthermore, the method allows for analysis to be completed within one day including the on-bead digestion which was optimized at 4 hours.

Table 3.1: Summary of urinary proteome profiling data according to isolation method tested. Peptide and protein identifications at 1% FDR after 1D LCMS/MS, cost and handling time are shown.

Isolation Method	Resuspension buffer	Clean-up method	Peptides (1% FDR)	Proteins (1% FDR)	Total time in hours (Hands on)	Cost in USD ($)
Precipitation (Acetone)	RB	FASP-Stage Tip	308	88	22 (6)	19
	RB	mFASP-SOLAμ	1377	419		
Precipitation (Acetone)	2% SDS	HILIC	2567	471	9 (4)	8
Norgen Kit	Elution Buffer	None	2590	407	18 (2)	25

The peptide molecular weight distribution and GRAVY score distributions are similar between all three methods thus showing little to no bias (Figure 3.6A and 3.6C). The mFASP methodology showed reduced selectivity for peptides with alkaline pKa as compared to HILIC and Norgen (Figure 3.6B). The molecular weight distributions were similar, with the exception of mFASP showing decreased selectivity for proteins under 30 kDa which was attributed to the molecular weight cut-off of the filter used (Figure 3.6D).

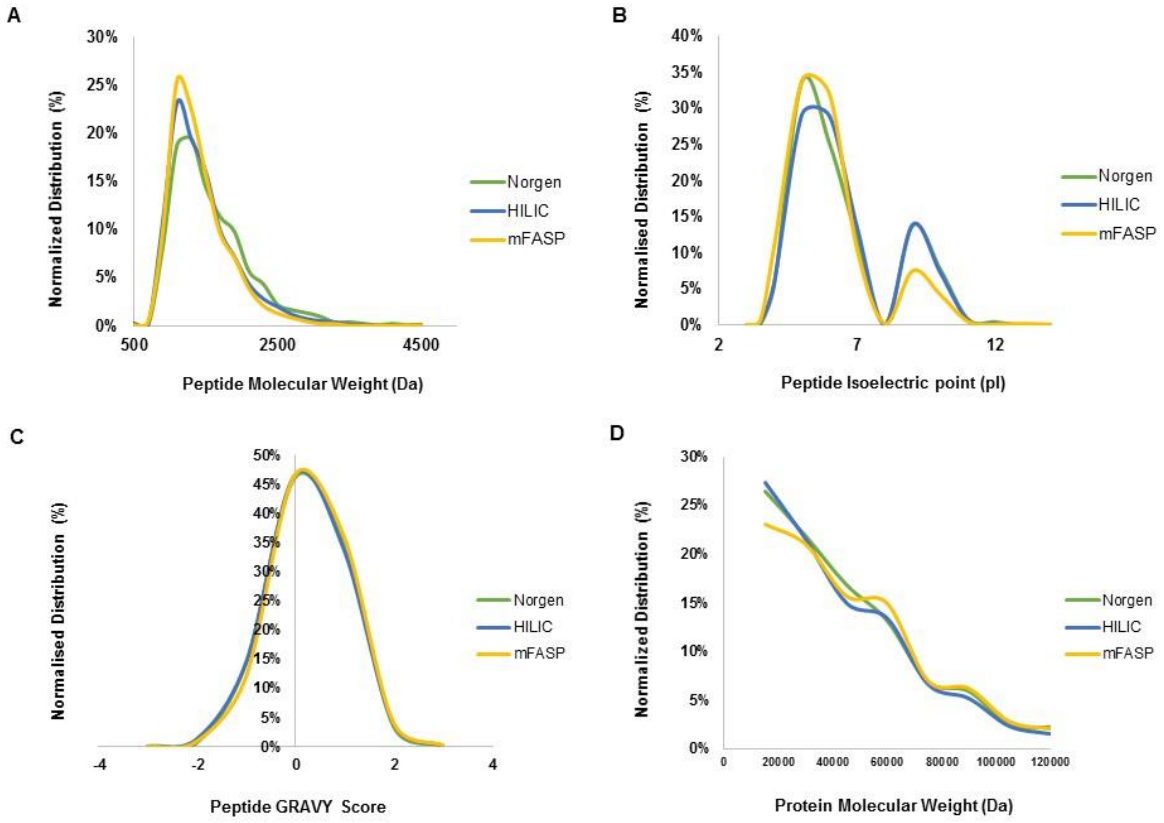


Figure 3.6: Protein and peptide properties for Norgen, HILIC and mFASP methods assessed using ExPASy p*I*/MW and protein GRAVY calculators. Peptides displayed no difference in molecular weight (A) and GRAVY score (C) distributions. The mFASP method retained fewer peptides with alkaline pKa (B) and proteins with molecular weight less than 30 kDa (D).

3.7. LCMS/MS Data acquisition method development

3.7.1. Gradient length

The 60-minute gradient returned 3803 ± 825 peptides and 319 ± 34 proteins and the 90-minute gradient returned 3414 ± 490 peptides and 290 ± 39 proteins at 1% FDR, respectively (Figure 3.7A). Furthermore, the FWHM data indicated that using a 90-minute gradient generated broader chromatographic peaks which resulted in loss of signal (Figure 3.7B). This contributed to the decrease in protein and peptide identifications observed. Therefore the 60-minute gradient was chosen as the optimal gradient length for all future experiments.

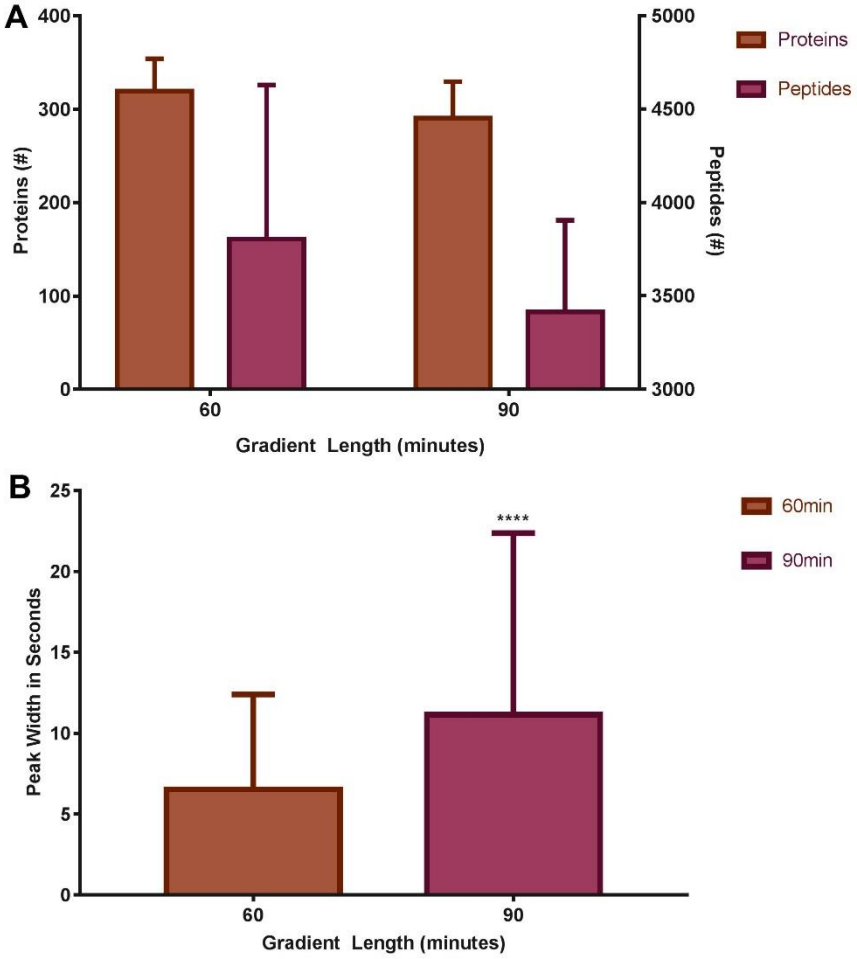


Figure 3.7: Gradient lengths of 60 and 90 minutes were assessed. The 60-minute gradient resulted A) in greater peptide and protein identifications at 1% FDR and B) narrower FWHM (Mann-Whitney test ****$p < 0.0001$) (n = 3).

3.7.2. Gradient slope

Gradient slope A, returned 3549 ± 1435 peptides and 273 ± 54 proteins, gradient slope B returned 2597 ± 424 peptides and 226 ± 15 proteins and gradient slope C, returned 2442 ± 314 peptides and 242 ± 38 proteins at 1% FDR, respectively (Figure 3.8). Therefore, gradient slope A (5-30 %B) was chosen as the optimal gradient slope for all future experiments.

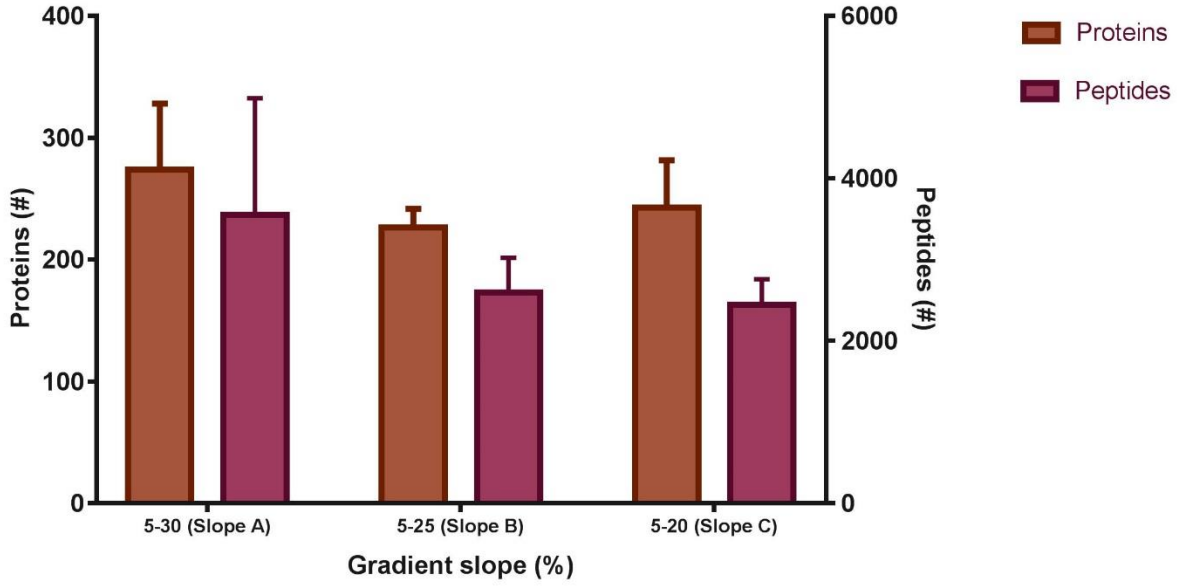


Figure 3.8: Summary of gradient slope experiments. The slope of the 60-minute gradient was varied between 5 - 20, 25 and 30 %B, respectively. Gradient A generated the greatest number of peptides and proteins identifications at 1% FDR (n = 3).

3.7.3. Column load per injection

The 4-µg load returned the lowest peptide and protein identifications. The 8-µg showed the greatest protein identifications (343 ± 19) however the peptides identified (3765 ± 219) were lower than the 6-µg load. Conversely, the 6-µg had the greatest peptide identifications (3834 ± 1031), but lower protein identifications (294 ± 24) (Figure 3.9). Based on these results, the 8-µg injection load was chosen as the best load for optimal peptide and protein identifications and was used in all future experiments.

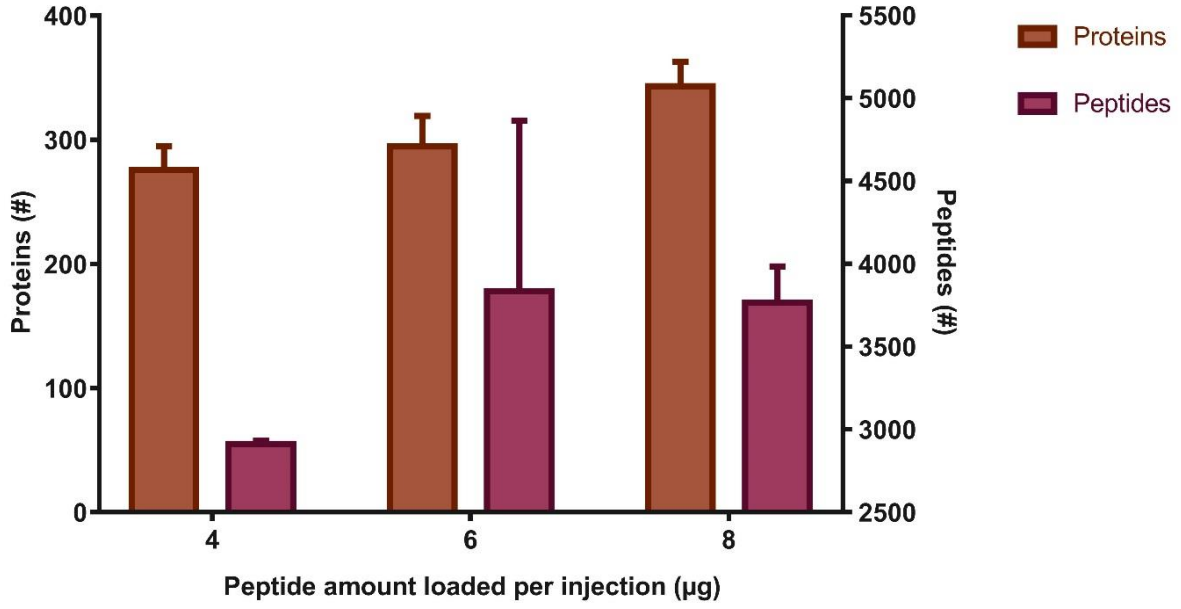


Figure 3.9: Summary of peptide loads assessed. A standard LCMS gradient was used to assess the effect of different peptide loads on protein and peptide identifications. The 8-µg load resulted in the greatest number of protein identifications at 1% FDR. (n = 3).

3.8. MS data acquisition parameters for data dependent analysis

3.8.1. MS/MS and accumulation time

The acquisition method using 40 MS/MS scans of 50 milliseconds each resulted in greater peptide (3439 ± 498) and protein (307 ± 43) identifications compared to the method using 25 MS/MS and 90 milliseconds (peptides – 2748 ± 194; proteins – 238 ± 10) (Figure 3.10).

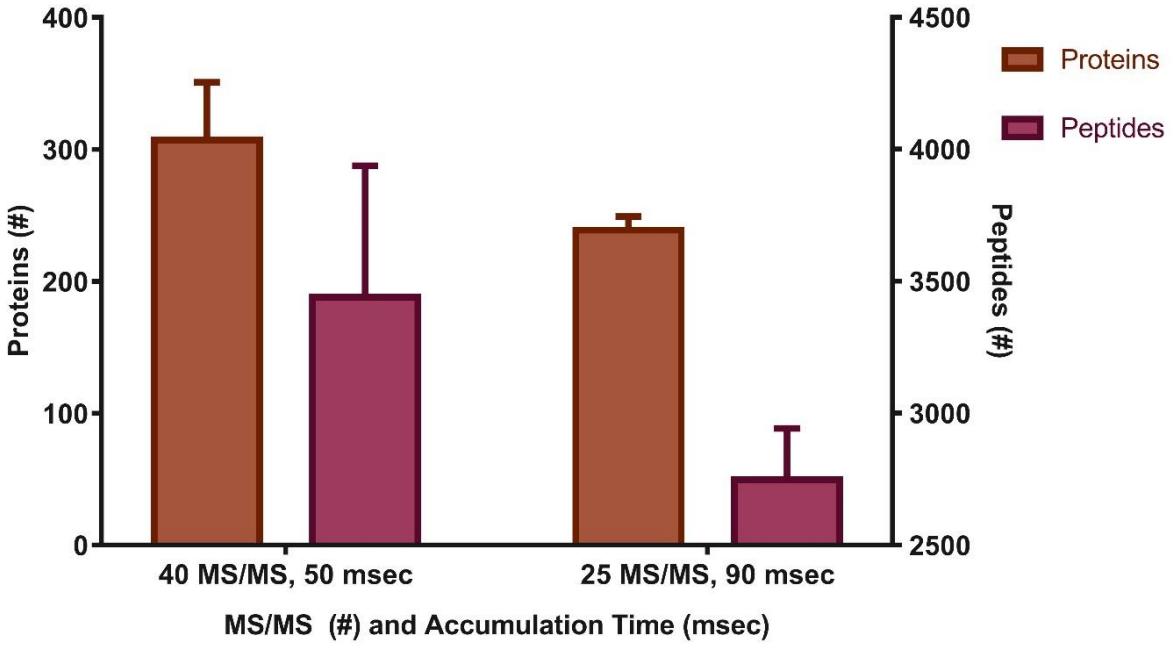


Figure 3.10: Peptide and protein identifications at 1% FDR after varying the number of MS/MS scans and accumulation time. The 50-millisecond accumulation time with 40 MS/MS scans generated the greatest number of protein and peptide identifications (n = 3).

3.8.2. Rolling collision energy (CE) parameters

The crowdsourced CE parameters returned data that showed greater peptide (770 ± 2) and protein identifications (184 ± 2) than the default settings (peptides – 681 ± 20; proteins 171 ± 2). There was an increase in peptide and protein identifications by 13% and 5%, respectively (Figure 3.11). These increases correlate to the expected increases in data quality by 5-10% as observed by SCIEX (SCIEX, 2015). Therefore, the crowdsourced CE parameters were used in all further analyses.

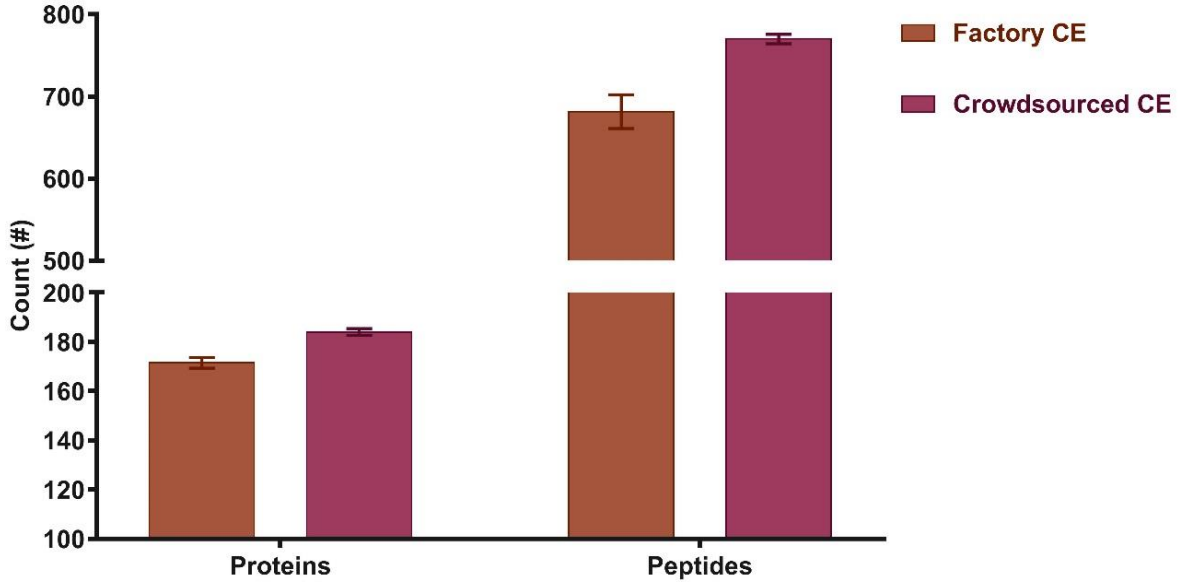


Figure 3.11: Peptide and protein identifications at 1% FDR using default and crowdsourced CE parameters. The crowdsourced CE parameters returned data that showed greatest peptide and protein identifications (n = 3).

3.9. Typical total ion chromatogram derived from the optimised method.

A total ion chromatogram derived from a typical urinary proteome analysis using the optimised settings/parameters is shown below.

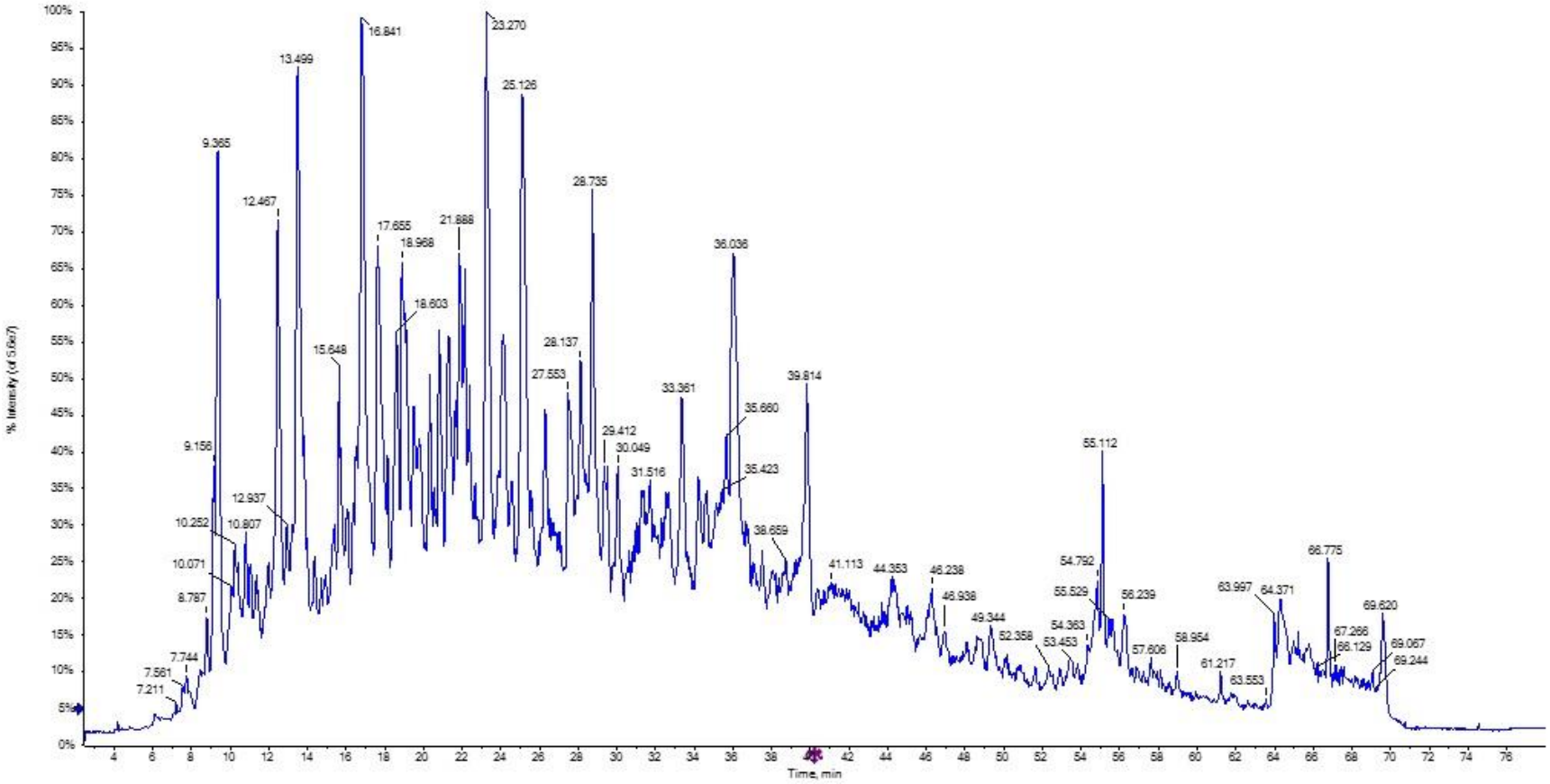


Figure 3.12: Total ion chromatogram of a typical urinary protein sample. The sample was processed using the Norgen kit, digested at 1:20 enzyme to protein ratio, non-cleaned post digestion and analysed directly (no trap) over a 60-minute gradient with 40 MS/MS and 50 millisecond accumulation time.

Table 3.2: Summary of data derived from acquisition parameters used to optimise methodology for data dependent analysis. The settings highlighted in red were, after comprehensive and iterative analysis, determined as the optimal combined settings to be used in the analysis of clinical urinary samples. Single injection, acetone precipitated (as a clean-up) samples using the optimal settings showed losses in protein and peptide recovery.

Sample	Mass injected (µg)	Gradient length (min)	Gradient slope (% ACN Start-End)	MS/MS (#)	Accumulation time (ms)	Proteins identified at 1% FDR	Peptides identified at 1% FDR
RG085 (#Crude)	4	60	5 - 30	40	50	289	2926
	4	90	5 - 30	40	50	262	2898
	6	60	5 - 30	40	50	312	4564
			5 - 25	25	90	237	2897
			5 - 20	25	90	270	2665
	6	90	5 - 30	40	50	275	3472
			5 - 25	25	90	250	2819
			5 - 20	25	90	229	2528
	8	60	5 - 30	40	50	357	3921
	8	90	5 - 30	40	50	335	3874
RG085 (*1-hour Acetone Precipitation)	8	60	5 - 30	40	50	329	3610
RG085 (*16-hour Acetone Precipitation)	8	60	5 - 30	40	50	303	3465

*precipitation was performed post extraction as a clean-up prior to digestion with trypsin. #crude implies that the extracted protein was digested and analysed directly without further clean-up. FDR – False discovery rate, ms- milliseconds.

3.10. MS data acquisition and processing parameters for data independent analysis

3.10.1. Constant vs variable SWATH™ windows

The 32CW and 32VW acquisition methods displayed similar results and both returned more comprehensive data than the 80VW and 100VW acquisition methods (Figure 3.13 A and B). Other studies have reported that increasing the number of windows over the same mass range results in higher quality and more comprehensive data (SCIEX, 2014; Schilling, Gibson and Hunter, 2017), however this did not occur in the current experiment. The 80VW and 100VW methods showed greater data points per peak (mean of 10 and 11, respectively) compared to the 32CW and VW methods (mean of 8). Studies conducted by Biognosys (developers of Spectronaut™) have shown that the optimal number of data points per peak that returns the best data with low CV is 8 (Bruderer *et al.*, 2017), which validates the 32VW method which showed a mean data points per peak of 8. Furthermore, the data completeness, which is a measure of missing values in the method, and library recovery were both lower in the 80VW and 100VW methods compared to the 32VW and 32CW methods. The greater the percentage of data completeness, the lower the missing values in the method. Therefore, these methods (80VW and 100VW) had a greater number of missing values which lead to decreased library recoveries.

The direct comparison between the 32VW and 32CW displayed similar results for data completeness and data points per peak, however, the 32VW narrowly performed better overall. The 32VW method displayed the greatest: library recovery (21%), number of peptides (2288 ± 93) and proteins (568 ± 5), and importantly, had the lower CV (5%) between the two methods. Based on these results, the 32VW method was used for data acquisition in all further SWATH™ experiments.

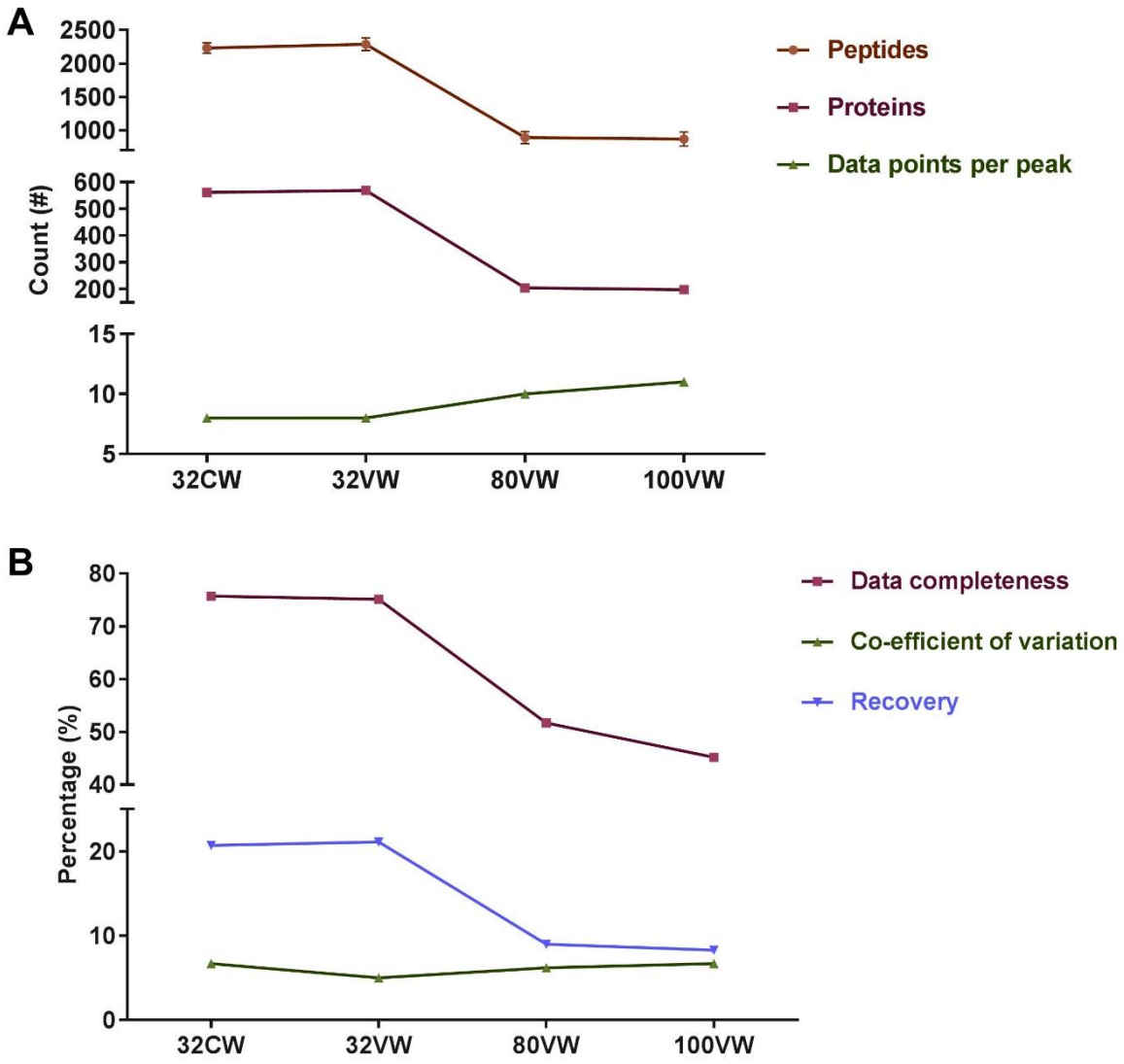


Figure 3.13. Summary of VW vs CW comparison. A) Shows peptide and protein identifications at 1% FDR and data points per peak and B) shows data completeness, CV and library recovery, for each method. Overall the 32VW data returned the most comprehensive data and was chosen as the optimal method for the clinical cohort.

3.10.2. Two dimensional RPRP

Using a 2D RPRP approach generated a 2-fold larger spectral library compared to generating a spectral library from 1D RP of the same sample. The trade-off between time and proteome depth (spectral library depth) is critical to minimise MS time whilst still achieving a large spectral library. Using the 1D RP approach required just three 60-minute runs to complete whilst the 2D RPRP approach required 60 minutes for the 1st dimension followed by 60 minutes vacuum drying and 45 minutes for each fraction in the 2nd dimension. Thus, using the 2D RPRP approach is more time consuming, however it achieves a significant increase in protein identifications by 2-fold (Table 3.2 – 1D RP = 357 proteins vs Table 3.3 - 2D RPRP = 705 proteins). Based on these results, the 2D RPRP approach was used to generate spectral libraries in all further experiments.

Table 3.3: Total peptides and proteins derived from each fraction following 2D RPRP MS analysis.

Fraction	Proteins identified	Peptides identified
1	14	23
2	183	286
3	243	390
4	251	442
5	74	103
6	334	595
7	278	515
8	186	291
9	36	52
10	33	50
Combined Fractions	705	2249

3.10.3. Varying reference spectral library

The data generated from the various libraries clearly show that having a larger library enables deeper proteome coverage. However, this is not a generic rule for any library i.e. whilst it is possible to generate data from libraries that are not generated in-house (e.g. Biognosys Urine Library), a large, project specific, in-house generated library returns higher quality data than the external library alone, no matter how large it may be. Further assessment revealed that combining libraries from in-house generated data and external data allows for the deepest proteome coverage (peptides = 4083.3 ± 1269.5, proteins 600.3 ± 155) (Table 3.4 and Figure 3.14). Therefore, all further samples were analysed using a combined library (Library 7) as a reference.

Table 3.4: Summary of library matches showing average peptides and proteins matched using different reference libraries. The data represented below shows protein and peptide matches at 1% FDR when analysed using the default settings in Spectronaut™ 11. Library 7, highlighted in red, displayed the greatest peptide and protein identifications.

Library	Peptides (± SD)	Proteins (± SD)
1	406.0 (67.5)	212.3 (36.20)
2	436.3 (71.5)	215.3 (39.70)
3	3345 (434.9)	501.6 (63.00)
4	3103 (901.0)	555.6 (147.5)
5	2477 (1057)	459.6 (171.1)
6	4053 (1114)	600.3 (137.4)
7	4083 (1269)	600.3 (155.2)

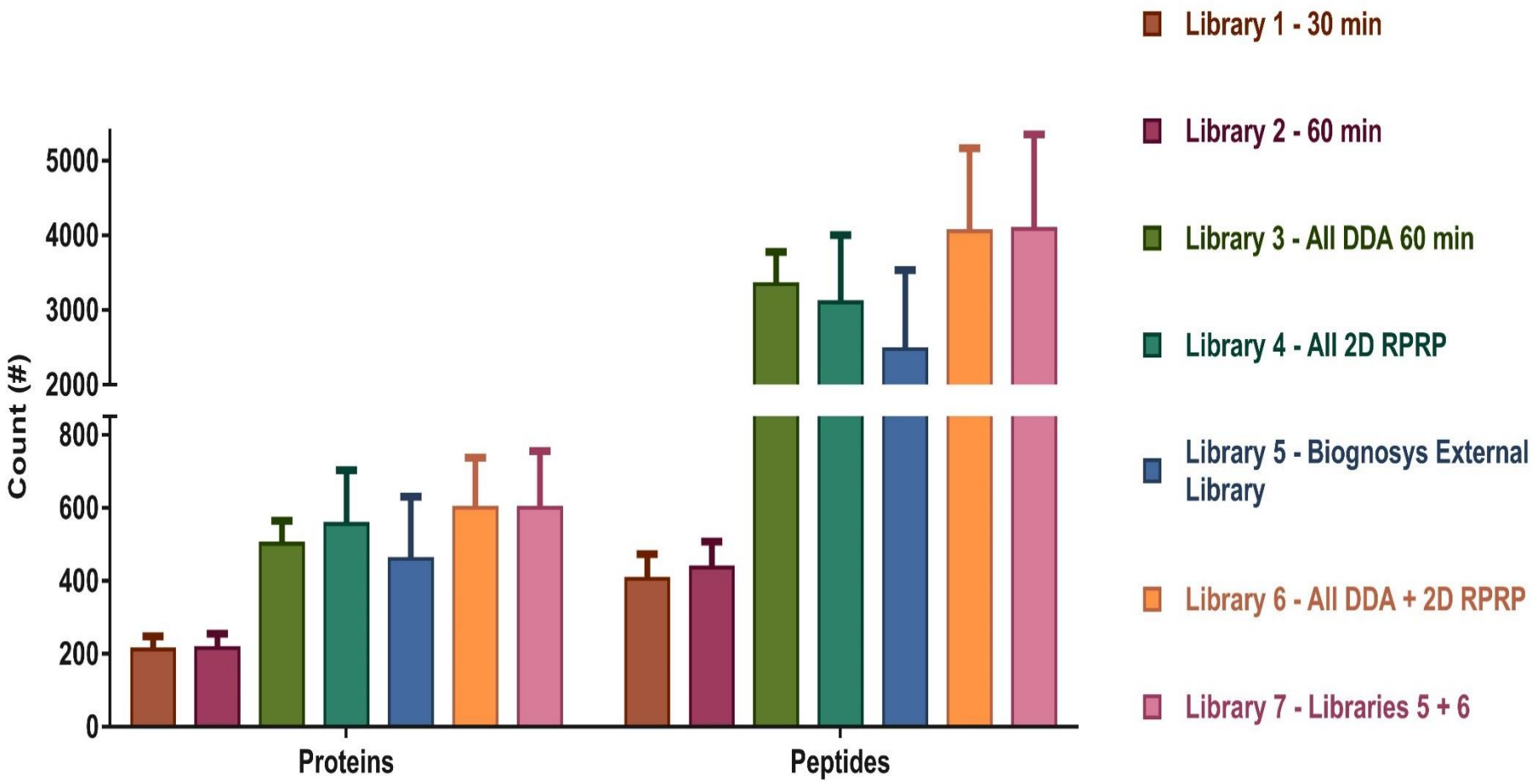


Figure 3.14: Graph of library matches showing peptides and proteins matched in different libraries at 1% FDR. The larger the library the greater the peptide and protein identifications detected. Library 7 shows the greatest number of matched peptides (4083 ± 1269) and shares the greatest number of matched proteins with library 6 i.e. 600 ± 137 and 600 ± 155 (for libraries 6 and 7, respectively).

3.11. Technical variation assessment

The complete workflow, from urine collection, processing, mass spectrometry data acquisition and analysis (using the final optimised method, see Figure 2.6), displayed a low technical variation with a median CV of 7.32; (25th-75th percentile = 3.56 - 14.4) (Figure 3.15). Total ion chromatograms displayed a high degree of similarity between technical replicates (Figure 3.16). Thus, the developed workflow was accepted, as currently optimised, for use on the clinical urine samples.

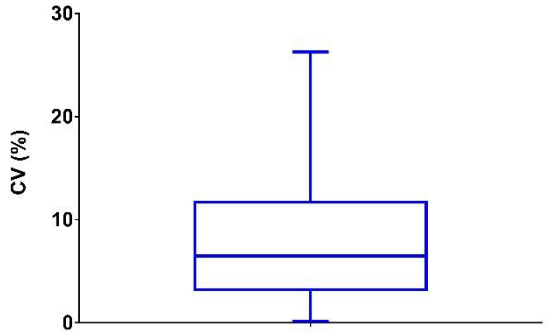


Figure 3.15: Coefficient of variation when analysing three technical replicates using the complete workflow. The complete workflow displayed a mean CV of less than 10% (n = 3).

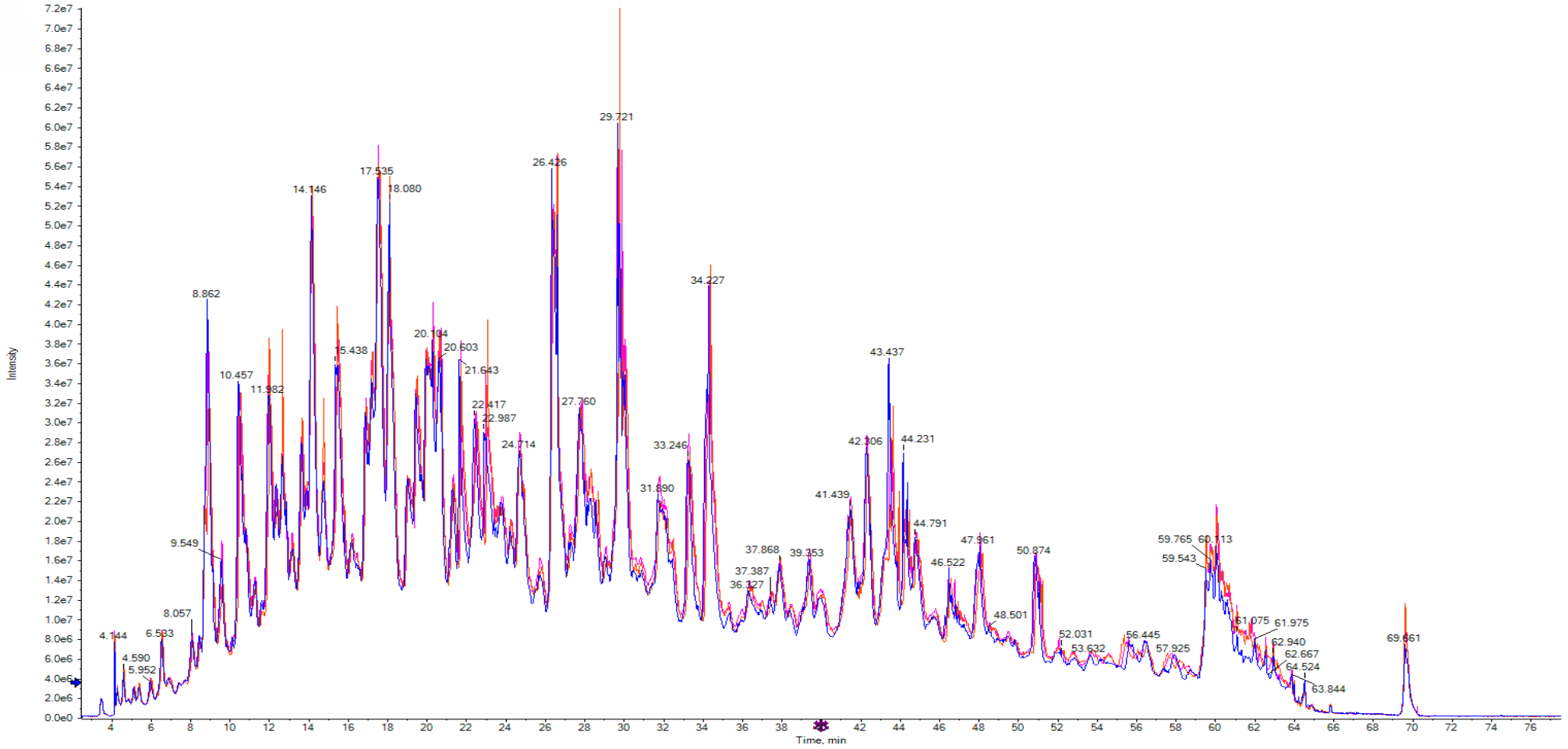


Figure 3.16: Overlaid total ion chromatographs of a single injection from each technical workflow replicate. The technical workflow replicate data showed a great consistency across different days indicating a low technical variation (n = 3).

3.12. Gene ontology analysis

The GO protein class analysis of all proteins identified using the Norgen kit method illustrates that proteins from diverse classes are detected (refer to Figure 3.17). By extension, the GO pathway analysis revealed that proteins detected are involved in multiple unrelated pathways (Figure 3.17). The developed methodology using the Norgen kit does not favour any protein class or pathway. Therefore, using this complete workflow for biomarker discovery does not have to be targeted to a specific subset of protein classes or pathways and can be done with confidence given that proteins from various pathways and classes are represented. This provides a greater chance of finding true biomarkers without introducing potential effects associated with sample preparation bias.

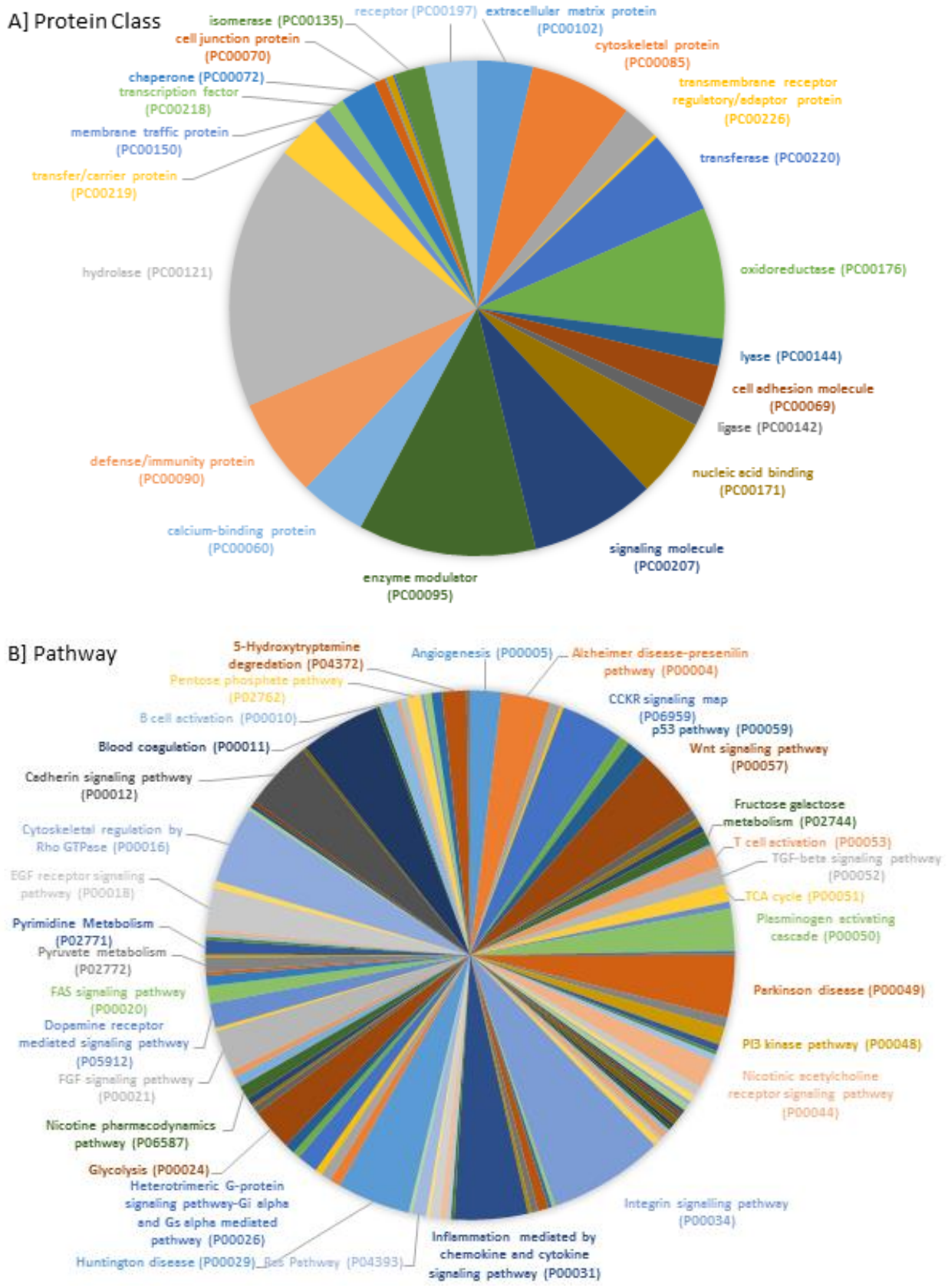


Figure 3.17: Panther gene ontology pathway analysis of all proteins identified in the clinical sample cohort. A) GO protein class analysis and B) GO pathway analysis. For pathway analysis, only pathways with 4 or more gene matches are labelled.

Chapter 3
Results

Part B
Urinary protein biomarker discovery:
An exploratory case-control study

3.13. Demographic details and clinical characteristics

No significant differences were observed in: mean age of study participants, viral load, and urinary protein:creatinine ratios between cases and controls. Creatinine and urea levels were 7-fold and 5-fold higher, respectively in cases compared to controls. Albumin was 2-fold lower and eGFR was 9-fold lower in cases compared to controls (Table 3.5). The characteristics commonly used in diagnosis of AKI are illustrated below (Figure 3.18).

Table 3.5: Demographic details and relevant clinical characteristics of the patient cohort used in the assessment of kidney function. The figures in the table represent the mean of each clinical characteristic with standard deviation shown in brackets.

Characteristic (units)		Cases	Controls	Significance (p value)
Gender	Female	23	23	
	Male	14	14	
Mean age (years)		44.7 (12.7)	43.5 (10.8)	ns
Admission Creatinine (µmol/L)		454.3 (368.5)	64.18 (24.35)	0.0001
Urea (mmol/L)		23.83 (16.05)	4.813 (3.08)	0.0001
eGFR (mL/min/1.73m^2)		21.36 (14.56)	191.3 (398.20)	0.0134
Albumin (g/L)		19.25 (6.24)	39.95 (7.21)	0.0001
Urinary protein:creatinine ratio (g/mmol creatinine)		0.20 (0.14)	0.19 (0.59)	ns

ns - not significant ($p \geq 0.05$)

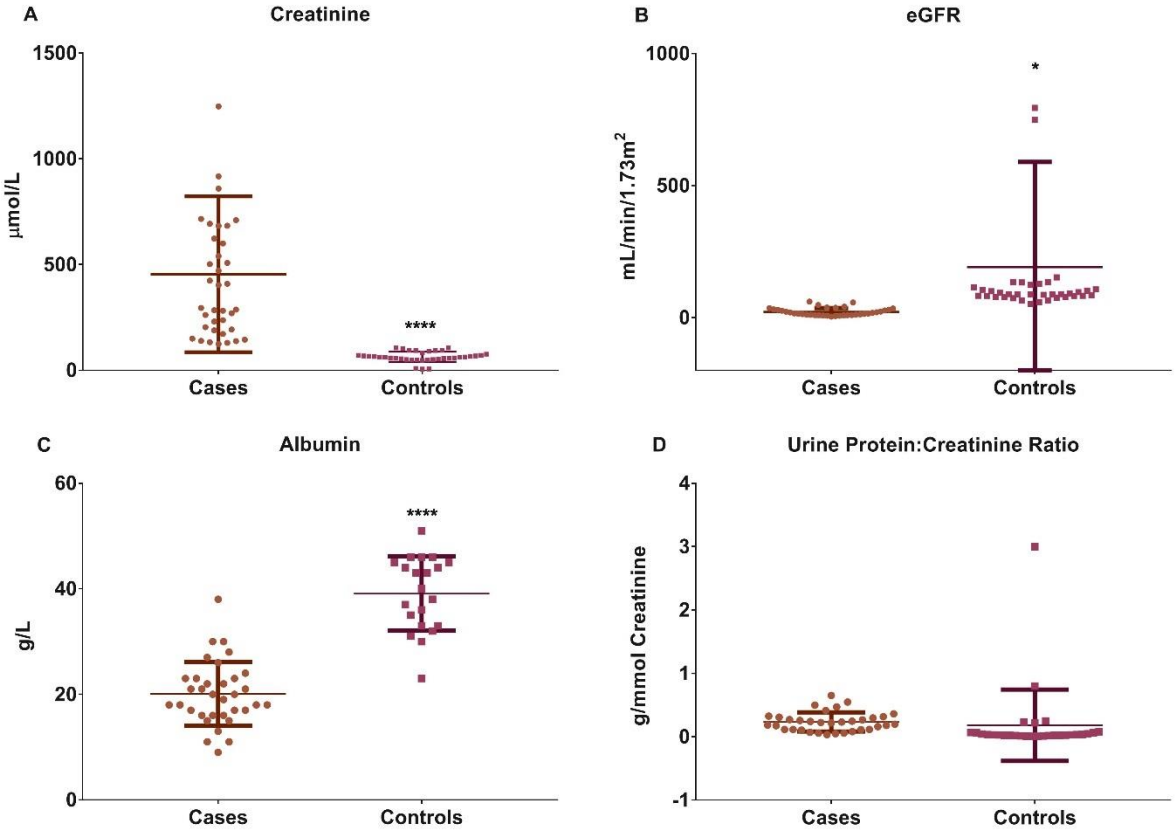


Figure 3.18: Summary of clinical characteristics used in the diagnosis of AKI. A) Serum creatinine level is a key indicator of declining kidney function. The means are significantly different (Wilcoxon signed rank test **** $p < 0.0001$) from each other where cases display higher levels of serum creatinine compared to the controls by 7-fold. B – D) Show distributions of parameters commonly used at clinical level to aid in the diagnosis of kidney injury. The distributions show large degrees of overlap and outliers which could potentially lead to misdiagnosis and further complications.

3.14. Putative biomarkers

Following sample preparation and analysis as described in section 2.19, lists of proteins showing differential abundance were generated from each experiment using Spectronaut™ 11. The data was then collated using Microsoft® Excel® (v1708) and imported into Perseus v1.6.1.1 (Tyanova, Temu, *et al.*, 2016) where the data was finally filtered using the 30% cut-off, as explained above. A combined list of 52 differentially abundant proteins common to ≥ 30% of the experiments was generated and are tabulated in Table 3.6 and Table 3.7 (Chapter 3). Table 3.6 shows candidate biomarkers arranged from most confident to least confident. The protein q-value was calculated by Spectronaut™ 11 by considering the peptides, and by extension proteins, that were accurately identified and quantified across the 37 experiments for the individual protein. The greater the number of peptides identified across subsequent samples, the greater the confidence in the peptide and, by extension, the protein's identification and quantity.

Table 3.6: Putative biomarkers arranged in ascending q value (decreasing confidence) order.

UniProt Accession Number	Protein Description	Protein Name	Experiments Expressing marker	q Value	*Mean Log_2 Fold-change
P02787	Serotransferrin	TRFE_HUMAN	18	8.22E-11	1.15
P01009	Alpha-1-antitrypsin	A1AT_HUMAN	19	2.83E-10	-2.59
P02768	Serum albumin	ALBU_HUMAN	21	1.35E-07	-0.90
P07911	Uromodulin	UROM_HUMAN	19	4.62E-07	3.74
P01011	Alpha-1-antichymotrypsin	AACT_HUMAN	17	1.11E-05	-0.63
P02765	Alpha-2-HS-glycoprotein	FETUA_HUMAN	16	2.06E-05	2.61
P00915	Carbonic anhydrase 1	CAH1_HUMAN	16	2.58E-05	-4.09
P60709	Actin, cytoplasmic 1	ACTB_HUMAN	13	4.93E-05	-0.16
P08571	Monocyte differentiation antigen CD14	CD14_HUMAN	16	5.36E-05	2.39

P0DOX7	Immunoglobulin kappa light chain	IGK_HUMAN	16	5.66E-05	1.69
P02750	Leucine-rich alpha-2-glycoprotein	A2GL_HUMAN	14	8.07E-05	1.45
P01034	Cystatin-C	CYTC_HUMAN	24	8.20E-05	-5.15
P05109	Protein S100-A8	S10A8_HUMAN	15	9.18E-05	2.48
Q9UBX5	Fibulin-5	FBLN5_HUMAN	12	0.00011	-3.14
P02647	Apolipoprotein A-I	APOA1_HUMAN	21	0.00013	-3.64
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	PGBM_HUMAN	14	0.00014	1.30
P06702	Protein S100-A9	S10A9_HUMAN	24	0.00016	3.18
P05090	Apolipoprotein D	APOD_HUMAN	17	0.00017	2.61
P04217	Alpha-1B-glycoprotein	A1BG_HUMAN	15	0.00018	2.37
P36955	Pigment epithelium-derived factor	PEDF_HUMAN	24	0.00018	-6.30
P61769	Beta-2-microglobulin	B2MG_HUMAN	26	0.0002	-4.96
P01876	Immunoglobulin heavy constant alpha 1	IGHA1_HUMAN	15	0.00028	1.90
P07737	Profilin-1	PROF1_HUMAN	16	0.0003	-3.36
P0C0L5	Complement C4-B	CO4B_HUMAN	19	0.00031	-2.93
P68871	Haemoglobin subunit beta	HBB_HUMAN	18	0.00036	-3.19
P02753	Retinol-binding protein 4	RET4_HUMAN	31	0.00041	-6.45
Q92520	Protein FAM3C	FAM3C_HUMAN	13	0.00041	-2.67
P05154	Plasma serine protease inhibitor	IPSP_HUMAN	13	0.00041	2.16
P00751	Complement factor B	CFAB_HUMAN	15	0.00043	-3.06

P27797	Calreticulin	CALR_HUMAN	16	0.00046	-3.07
P02760	Protein AMBP	AMBP_HUMAN	15	0.00054	-0.85
P01019	Angiotensinogen	ANGT_HUMAN	17	0.00056	-3.56
P01008	Antithrombin-III	ANT3_HUMAN	12	0.00058	-3.28
P04083	Annexin A1	ANXA1_HUMAN	15	0.00059	1.95
P06727	Apolipoprotein A-IV	APOA4_HUMAN	21	0.00063	-5.33
P01024	Complement C3	CO3_HUMAN	23	0.00064	-3.77
P02748	Complement component C9	CO9_HUMAN	17	0.00065	-3.66
P04004	Vitronectin	VTNC_HUMAN	13	0.00065	-2.73
P17900	Ganglioside GM2 activator	SAP3_HUMAN	18	0.00071	1.11
P02675	Fibrinogen beta chain	FIBB_HUMAN	18	0.00074	-2.99
P62937	Peptidyl-prolyl cis-trans isomerase A	PPIA_HUMAN	13	0.00079	-3.28
P69905	Haemoglobin subunit alpha	HBA_HUMAN	16	0.00088	-3.21
P01625	Ig kappa chain V-IV region Len	KV402_HUMAN	12	0.00089	1.37
P04196	Histidine-rich glycoprotein	HRG_HUMAN	15	0.00099	-3.25
P01834	Immunoglobulin kappa constant	IGKC_HUMAN	13	0.00105	2.24
Q14118	Dystroglycan	DAG1_HUMAN	12	0.00108	-2.97
P01833	Polymeric immunoglobulin receptor	PIGR_HUMAN	12	0.00123	1.80
P20774	Mimecan	MIME_HUMAN	17	0.00152	-4.08
P02679	Fibrinogen gamma chain	FIBG_HUMAN	20	0.00169	-4.14
P22352	Glutathione peroxidase 3	GPX3_HUMAN	14	0.00185	-3.12
O95425	Supervillin	SVIL_HUMAN	16	0.00359	3.42

*negative mean Log_2 fold change indicates a higher abundance in cases and vice versa

Table 3.7: Putative biomarkers identified at 1% FDR. The candidate biomarker list is arranged in descending order in terms of the number of experiments in which a protein met the criteria for inclusion as a biomarker.

UniProt Accession Number	Protein Description	Protein Name	Experiment expressing marker		q Value	*Mean Log_2 Fold-change
			Number $\frac{(x)}{37}$	Percentage		
P02753	Retinol-binding protein 4	RET4_HUMAN	31	84	0.00041	-6.45
P61769	Beta-2-microglobulin	B2MG_HUMAN	26	70	0.0002	-4.96
P01034	Cystatin-C	CYTC_HUMAN	24	65	8.20E-05	-5.15
P06702	Protein S100-A9	S10A9_HUMAN	24	65	0.00016	3.18
P36955	Pigment epithelium-derived factor	PEDF_HUMAN	24	65	0.00018	-6.30
P01024	Complement C3	CO3_HUMAN	23	62	0.00064	-3.77
P02768	Serum albumin	ALBU_HUMAN	21	57	1.35E-07	-0.90
P02647	Apolipoprotein A-I	APOA1_HUMAN	21	57	0.00013	-3.64
P06727	Apolipoprotein A-IV	APOA4_HUMAN	21	57	0.00063	-5.33
P02679	Fibrinogen gamma chain	FIBG_HUMAN	20	54	0.00169	-4.14
P01009	Alpha-1-antitrypsin	A1AT_HUMAN	19	51	2.83E-10	-2.59
P07911	Uromodulin	UROM_HUMAN	19	51	4.62E-07	3.74
P0C0L5	Complement C4-B	CO4B_HUMAN	19	51	0.00031	-2.93
P02787	Serotransferrin	TRFE_HUMAN	18	49	8.22E-11	1.15
P68871	Haemoglobin subunit beta	HBB_HUMAN	18	49	0.00036	-3.19
P17900	Ganglioside GM2 activator	SAP3_HUMAN	18	49	0.00071	1.11
P02675	Fibrinogen beta chain	FIBB_HUMAN	18	49	0.00074	-2.99
P01011	Alpha-1-antichymotrypsin	AACT_HUMAN	17	46	1.11E-05	-0.63
P05090	Apolipoprotein D	APOD_HUMAN	17	46	0.00017	2.61

P01019	Angiotensinogen	ANGT_HUMAN	17	46	0.00056	-3.56
P02748	Complement component C9	CO9_HUMAN	17	46	0.00065	-3.66
P20774	Mimecan	MIME_HUMAN	17	46	0.00152	-4.08
P02765	Alpha-2-HS-glycoprotein	FETUA_HUMAN	16	43	2.06E-05	2.61
P00915	Carbonic anhydrase 1	CAH1_HUMAN	16	43	2.58E-05	-4.09
P08571	Monocyte differentiation antigen CD14	CD14_HUMAN	16	43	5.36E-05	2.39
P0DOX7	Immunoglobulin kappa light chain	IGK_HUMAN	16	43	5.66E-05	1.69
P07737	Profilin-1	PROF1_HUMAN	16	43	0.0003	-3.36
P27797	Calreticulin	CALR_HUMAN	16	43	0.00046	-3.07
P69905	Haemoglobin subunit alpha	HBA_HUMAN	16	43	0.00088	-3.21
O95425	Supervillin	SVIL_HUMAN	16	43	0.00359	3.42
P05109	Protein S100-A8	S10A8_HUMAN	15	41	9.18E-05	2.48
P04217	Alpha-1B-glycoprotein	A1BG_HUMAN	15	41	0.00018	2.37
P01876	Immunoglobulin heavy constant alpha 1	IGHA1_HUMAN	15	41	0.00028	1.90
P00751	Complement factor B	CFAB_HUMAN	15	41	0.00043	-3.06
P02760	Protein AMBP	AMBP_HUMAN	15	41	0.00054	-0.85
P04083	Annexin A1	ANXA1_HUMAN	15	41	0.00059	1.95
P04196	Histidine-rich glycoprotein	HRG_HUMAN	15	41	0.00099	-3.25
P02750	Leucine-rich alpha-2-glycoprotein	A2GL_HUMAN	14	38	8.07E-05	1.45

P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	PGBM_HUMAN	14	38	0.00014	1.30
P22352	Glutathione peroxidase 3	GPX3_HUMAN	14	38	0.00185	-3.12
P60709	Actin, cytoplasmic 1	ACTB_HUMAN	13	35	4.93E-05	-0.16
Q92520	Protein FAM3C	FAM3C_HUMAN	13	35	0.00041	-2.67
P05154	Plasma serine protease inhibitor	IPSP_HUMAN	13	35	0.00041	2.16
P04004	Vitronectin	VTNC_HUMAN	13	35	0.00065	-2.73
P62937	Peptidyl-prolyl cis-trans isomerase A	PPIA_HUMAN	13	35	0.00079	-3.28
P01834	Immunoglobulin kappa constant	IGKC_HUMAN	13	35	0.00105	2.24
Q9UBX5	Fibulin-5	FBLN5_HUMAN	12	32	0.00011	-3.14
P01008	Antithrombin-III	ANT3_HUMAN	12	32	0.00058	-3.28
P01625	Ig kappa chain V-IV region Len	KV402_HUMAN	12	32	0.00089	1.37
Q14118	Dystroglycan	DAG1_HUMAN	12	32	0.00108	-2.97
P01833	Polymeric immunoglobulin receptor	PIGR_HUMAN	12	32	0.00123	1.80

*negative mean Log_2 fold change indicates a higher abundance in cases and vice versa

3.15. Reactome pathway gene ontology analysis

Reactome pathway analysis of all the putative markers (listed in Table 3.6 or 3.7) revealed that multiple biological pathways appear to be enriched (Table 3.8 and Figure 3.19). Pathways relating to platelets i.e. platelet degranulation and platelet activation, signalling and aggregation display highly significant enrichment as well as pathways relating to the immune system, the complement cascade and haemostasis. All pathways tabulated below are based on a FDR ≤ 1%.

Table 3.8: GO pathway analysis of putative biomarkers. A number of well-defined pathways were found to be enriched and may allude to mechanisms by which ART induced AKI occurs and/or progresses. Only significantly enriched pathways (FDR ≤ 0.01) are shown.

Pathway name	Candidates found in pathway	Total proteins in pathway	p-value	FDR
Platelet degranulation	13	137	3.77E-15	1.07E-12
Response to elevated platelet cytosolic Ca2+	13	144	7.11E-15	1.07E-12
Innate Immune System	25	1302	7.69E-13	7.77E-11
Haemostasis	20	806	3.95E-12	2.96E-10
Neutrophil degranulation	16	480	1.34E-11	8.02E-10
Platelet activation, signalling and aggregation	13	288	3.97E-11	1.99E-09
Scavenging of haem from plasma	9	106	2.43E-10	1.05E-08
Post-translational protein phosphorylation	9	109	3.1E-10	1.15E-08
Binding and Uptake of Ligands by Scavenger Receptors	10	167	6.39E-10	2.11E-08

Regulation of Insulin-like Growth Factor (IGF) transport and uptake	9	127	1.17E-09	3.5E-08
Immune System	28	2638	2.19E-08	5.92E-07
Regulation of TLR by endogenous ligand	5	31	1.33E-07	3.32E-06
Regulation of Complement cascade	7	139	9.15E-07	2.11E-05
Complement cascade	7	156	1.96E-06	4.11E-05
Activation of C3 and C5	3	7	2.79E-06	5.58E-05
Vesicle-mediated transport	13	818	7.48E-06	0.000135
Common Pathway of Fibrin Clot Formation	4	34	8.9E-06	0.000151
Amyloid fiber formation	5	78	1.17E-05	0.000188
Erythrocytes take up oxygen and release carbon dioxide	3	16	3.25E-05	0.000488
Formation of Fibrin Clot (Clotting Cascade)	4	52	4.63E-05	0.000695
Initial triggering of complement	5	120	8.98E-05	0.001233
Erythrocytes take up carbon dioxide and release oxygen	3	23	9.48E-05	0.001233
O2/CO2 exchange in erythrocytes	3	23	9.48E-05	0.001233
Retinoid metabolism and transport	4	79	0.000229	0.002691
Alternative complement activation	2	6	0.000245	0.002691

Integrin cell surface interactions	4	86	0.000316	0.003472
Metabolism of fat-soluble vitamins	4	94	0.000441	0.004407
Toll-Like Receptors Cascades	5	183	0.000621	0.006207
Signalling by high-kinase activity BRAF mutants	3	46	0.000714	0.006837
MAP2K and MAPK activation	3	47	0.00076	0.006837
Signalling by moderate kinase activity BRAF mutants	3	48	0.000807	0.007264
Paradoxical activation of RAF signalling by kinase inactive BRAF	3	49	0.000856	0.007707
Metal sequestration by antimicrobial proteins	2	13	0.001129	0.009034
Plasma lipoprotein remodelling	3	54	0.001131	0.009049

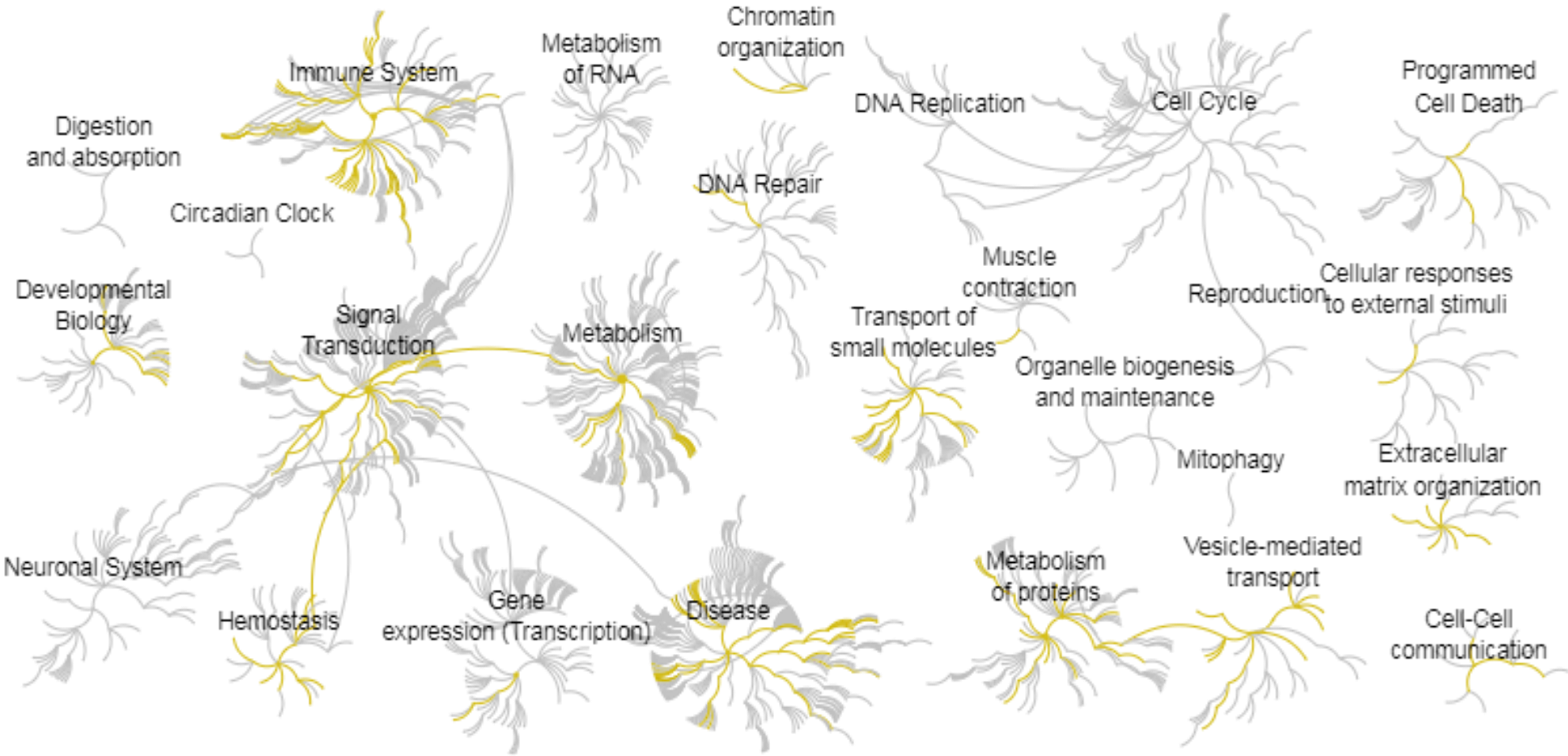


Figure 3.19. Over-representation pathway map derived from Reactome v65 (Fabregat *et al.*, 2018)**.** Pathways that contain biomarkers associated with impaired kidney function that are significantly overrepresented are highlighted in yellow.

3.16. Markers showing increasing urinary abundance as AKI severity progresses

The urinary abundance levels of beta-2-microglobulin, retinol-binding protein 4 and alpha-1-antitrypsin were observed to correlate with AKI severity. All three proteins exhibited an increase in urinary abundance as kidney damage progressed from mild-moderate to severe, however these apparent increases in abundance were not statistically significant between the two groups (Figure 3.21). No other biomarkers were observed to correlate in abundance with AKI severity.

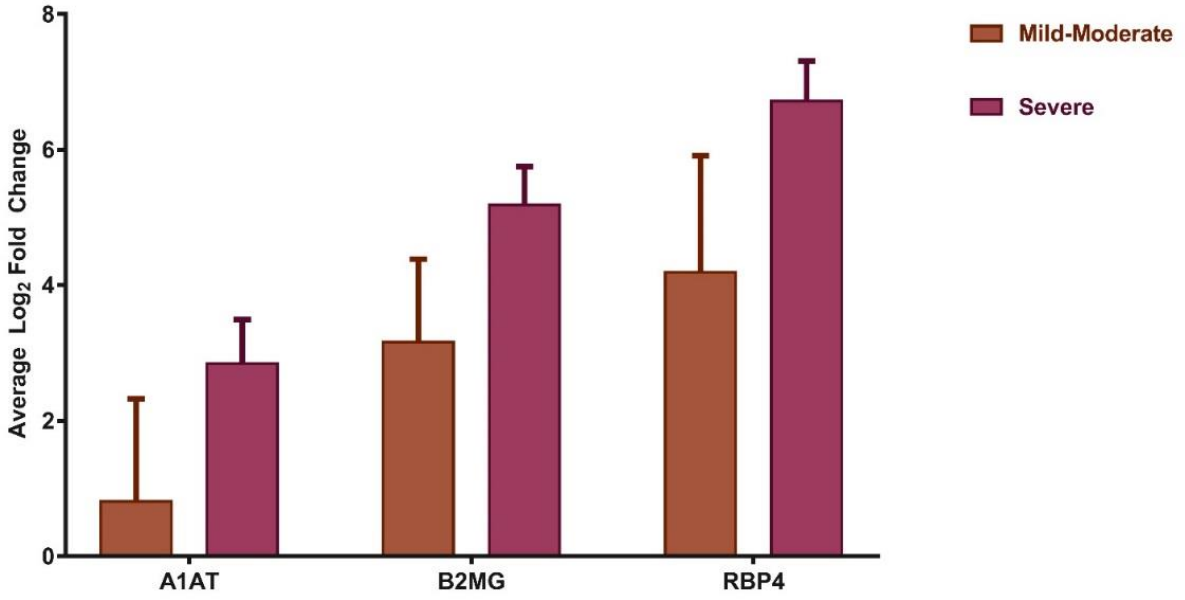


Figure 3.20: Proteins exhibiting increasing differential abundance as kidney damage progresses. Alpha-1-antitrypsin (A1AT), beta-2-microglobulin (B2MG) and retinol-binding protein-4 (RBP4) displayed an increase in urine as kidney damage progresses from mild-moderate to severe.

Chapter 4

Discussion

4.1. Method development and optimisation

Successful biomarker studies require workflows to be easily implemented and to also have intra- and inter-laboratory reproducibility. Particularly in urinary proteome studies, storage and post collection processing of urine needs to be done quickly and in a reproducible manner to limit technical variation that will inevitably skew data. Post voiding, urine and its components are not stable for long periods and furthermore, urine is susceptible to bacterial growth (Asscher *et al.*, 1966; Chambers *et al.*, 1999; Aurelius, 2009). This necessitates rapid post-collection processing to enable valuable data extraction downstream. Following precipitation of urinary proteins, protein resuspension can be difficult to achieve, and often requires the use of strong detergents and/or salts which are not compatible with downstream MS analysis (Yeung *et al.*, 2008).

It is common for urinary proteomics studies to employ organic solvent precipitation followed by FASP as a preferred method for isolation, clean-up and digestion of urinary proteins (Lacroix *et al.*, 2014; Yu and Pieper, 2015; Potriquet *et al.*, 2017), and whilst this is a widely used and a relatively simple procedure to follow, it is a labour-intensive process and is prone to sample losses. This is due to numerous handling steps which also have the potential to introduce sample contamination. Following FASP, samples need further processing, such as desalting and drying, before being analysed, substantially adding to cost and time, and perhaps more importantly, a decrease in sample recovery. Additionally, the use of molecular weight cut-off membranes in the initial steps of the FASP protocol can limit the depth of proteome coverage since small peptides and proteins can be lost. These may be important proteins in the given disease state and all information on these proteins are subsequently lost in the process. However, the use of a 30kDa membrane pore size does not mean all proteins smaller than 30kDa will be lost since the hydrodynamic volume of small proteins (when unfolded) may allow them to be retained. Lastly, the composition of the resuspension buffer can limit the method of protein quantification used for protein estimation (Friedenauer and Berlet, 1989;

Stutzenberger, 1992). These shortcomings make urinary proteome analysis using organic solvent precipitation a specialised, cumbersome and tedious process, lacking in reproducibility.

In contrast, the collection, storage and isolation procedure - of large volumes - of urine using the Norgen Kit is simple, reproducible (showing low technical variation between replicate samples) and robust, as this study demonstrated that the approach can still preserve bound proteins under extreme temperature conditions for long periods. This is particularly important in settings where clinics are not sufficiently well equipped to handle and store clinical samples for a long time. Generally, these clinics are located far from main city centres thereby posing a challenge for immediate sample transfer to analytical labs. In these instances, it is often difficult to maintain the cold-chain efficiently and therefore using traditional collection and storage methods is not advisable. Taking this into consideration, the current study details a simple and convenient method for collection of clinical samples without a need for pre-processing followed by simple storage, even up to 50 °C for relatively prolonged periods, while still being able to extract data from the precious clinical sample. Ideally, proteins should be isolated immediately since some protein and peptide losses were observed after extreme storage conditions (compared to fresh). Once protein is isolated, total protein quantification can be achieved using any of the popular quantification techniques such as Bradford and BCA assays without interference by the elution buffer used to recover the protein. Additionally, no MWCO filters are used thus smaller proteins are not lost during processing.

A novel processing method for samples that are not collected using Norgen Kits was developed in this study based on acetone precipitation of urinary proteins followed by resuspension in 2% SDS buffer with clean-up and on-bead digestion, using MagReSyn® HILIC microspheres. This method outperformed routinely used precipitation methods followed by FASP and performed marginally better than the Norgen kit-based method. Furthermore, the MagReSyn® HILIC method can be automated using magnetic handling robots, improving throughput and potentially improving reproducibility (Tape *et al.*, 2014). This method represents a viable alternative to kit-based sample processing; however, it requires immediate sample pre-processing and is more difficult to implement at point-of-care clinics. This can be addressed with formal training of clinical staff as well as the installation of the necessary equipment for processing should there be funding available.

The GO analysis from unique proteins identified in the HILIC and mFASP methods indicated that there was no significant enrichment bias in protein classes and pathways. Additionally, the

GO protein class and GO pathway analysis, from all proteins derived using the Norgen kit, clearly illustrated that there is no bias towards any particular protein class or pathway i.e. the kit-based urinary protein isolation method does not preferentially favour or enrich any particular protein groups or pathways. Therefore, the Norgen kit-based method could be used on clinical samples and would generate clinical proteomic data free from bias introduced through the processing and profiling workflow.

When compared to previously published data where multidimensional, nanoflow LC was used with long separation gradients (Nagaraj and Mann, 2011; Matafora *et al.*, 2014; Pang *et al.*, 2018), the methods detailed in this research (Norgen and HILIC) perform similarly in terms of protein groups and peptide identifications despite using shorter, microflow gradients. Thus, the methods achieve similar depth of proteome coverage whilst using less LC and mass spectrometry time which is a significant advantage. Furthermore, microflow allows for a more robust LC setup with greater uptime than nanoflow (Needham, Williams and Bell, 2016; Vowinckel *et al.*, 2018), potentially allowing for higher throughput. Additionally, using a SWATH-MS approach allows for a digital, proteomic signature to be captured which can be repeatedly mined without need for a fresh sample preparation and data acquisition.

In summary, the methods described here have been thoroughly optimised for urinary proteome profiling using a commercially available kit-based and a novel microsphere-based approach. The entire workflow has been assessed using various parameters and settings to derive a method that displays comprehensive proteome coverage, no apparent enrichment bias and low coefficient of variation between technical replicates. The method developed using the Norgen Kit exhibited good clinical utility and consequently it was applied to comprehensively analyse the clinical cohort of interest in this study.

4.2. Urinary protein biomarker discovery: an exploratory case-control study

In this exploratory study, putative protein biomarkers were identified as highlighted by the differences observed between cases and controls, as presented in Tables 3.6 and 3.7. A proportion (up to 80%) of the 52 putative biomarkers identified in this study have previously been reported in literature as putative biomarkers for various forms of kidney injuries and/or dysfunctions. However, such a study has never been undertaken in a South African population and not all the identified biomarkers have been reported as urinary proteins in this or other populations. Biomarker discovery-based clinical proteomics experiments often attempt to

identify novel proteins that can be applied uniquely to a diseased state, however in almost all studies most proteins identified as differentially abundant have some known association with the disease of interest (Geyer *et al.*, 2017). The application of big data machine learning algorithms can substantially increase the amount of information that can be derived from the dataset despite the identification of non-novel identifiers. It is unlikely that a single protein will provide sufficient sensitivity and specificity to distinguish patients at risk for AKI. The use of a biosignature, which includes a combination of proteins as well as routinely measured clinical characteristics, may be a more sensitive predictor for AKI risk. Vectra DA is an example of a protein-based, clinically applicable test is which is used in rheumatoid arthritis disease monitoring (Segurado and Sasso, 2014).

The most widely represented proteins across the 37 experiments and/or unique proteins belonging to the most significantly enriched pathways are discussed below, as these are the most likely to be true, biologically meaningful proteins to aid in AKI diagnosis. Proteins that displayed a net absolute Log_2 fold change less than two are omitted from thorough discussion as they are less likely to contribute significantly towards development of a specific and sensitive biomarker panel.

Well documented biomarkers of renal failure including: alpha-1-antitrypsin (A1AT), beta-2-microglobulin (B2M), cystatin c (CYTC), retinol-binding protein 4 (RBP4), pigment epithelium-derived factor (PEDF) and haemoglobin subunit beta (HBB) displayed an increase in urinary abundance in cases in this cohort. These findings correlate strongly with various studies pertaining to kidney diseases and dysfunctions. A study on Malaysian patients with renal calculi showed elevated urinary A1AT (Wai-Hoe *et al.*, 2009). Likewise, B2M (Schaub *et al.*, 2005; Du *et al.*, 2011; Heise *et al.*, 2011), CYTC (Ma *et al.*, 2010; Endre *et al.*, 2011; Heise *et al.*, 2011) and RBP4 (Che *et al.*, 2010), all displayed increased urinary levels in acute renal failure patients in the respective cohorts. A similar observation was made in kidney transplant patients who suffered rejection or post-operative renal complications where HBB (Cantley *et al.*, 2016), and PEDF (Moskowitz-Kassai *et al.*, 2012), were increased in urine post-surgery. In general, the sizes of the cohorts assessed in these studies ranged between 30 – 529 patients, mean 164. Collectively, these findings support the methods used in this study as well as validate these markers (as markers of kidney injury) since they were identified in patients of diverse genetic backgrounds.

Interestingly, the proteins B2M, RBP4 and A1AT, all exhibited an increased urinary abundance as kidney damage progressed from mild-moderate to severe. However, this observation was inconclusive following statistical testing, possibly due to the distributions between the two groups being small and unequal i.e. using a larger and equal sample size may yield statistically significant results to validate these findings. Nevertheless, the overall increase in abundance in these proteins represents promising targets for further development and assessment.

Serpins are a superfamily of serine protease inhibitors that are highly conserved and ubiquitously expressed in nature (Silverman *et al.*, 2001). Alpha-1-antitrypsin is a well-documented and abundantly expressed, hepatic stress, serine protease inhibitor that functions to protect cells and tissues against endogenous proteases (El-Akawi, Al-Hindawi and Bashir, 2008; Pérez-Holanda *et al.*, 2014). Recent evidence suggests that kidney injury can manifest changes that resemble hepatotoxic phenotypes in a term coined "renal hepatization" (Zager, Johnson and Frostad, 2014). Therefore, Zager *et al.* (2014) assessed whether the A1AT gene is significantly upregulated in a mouse model following induced kidney injury. Their experiments showed that A1AT mRNA was rapidly and significantly increased in proximal tubular cells and this paralleled a subsequent increase in urinary levels of A1AT. They further concluded that A1AT was upregulated in an attempt to minimise renal injury by blocking the effects of endogenous neutrophil elastase (Zager, Johnson and Frostad, 2014). Therefore, it is plausible that this mechanism occurs in humans as well and can account for the increase in urinary A1AT abundance in cases in this cohort. Recently, Mandili *et al.* (2018) reported A1AT as a urinary marker of renal cancer in von Hippel–Lindau patients. Other proteins with serine protease inhibitor-like properties including PEDF and plasma serine protease inhibitor, which displayed increases of > 20-fold and 4-fold, respectively, in urinary protein abundance in cases compared to controls. Vitronectin (VTNC) is another liver associated protein that interacts with serine protease inhibitors and has been implicated in a mouse model representing AKI (Gupta *et al.*, 2015), and in glomerulonephritis (Mesnard *et al.*, 2011). These observations can potentially be used clinically as indicators of renal injury due to TDF treatment and may manifest earlier than current markers of renal injury such as creatinine levels.

Beta-2-microglublin represents an interesting and potentially important protein that can be used in the early diagnosis of AKI in this study. The protein is a component of the major histocompatibility complex and plays a role is presenting antigens during immune response and has been implicated as a biomarker across a range of kidney associated dysfunctions [reviewed in (Argyropoulos *et al.*, 2017)]. The protein is normally filtered by the glomerulus

and 99.9% is reabsorbed by the kidney proximal tubules where it is catabolised (Wibell, 1978; Vaidya, Ferguson and Bonventre, 2008; Zeng *et al.*, 2014). Increased urinary B2M has been observed in tubular injury resulting from cardiac surgery and renal transplantation, which precedes a rise in serum creatinine levels [reviewed in (Vaidya, Ferguson and Bonventre, 2008; Endre *et al.*, 2011). Furthermore, increased serum B2M displayed the ability to predict AKI in transcatheter transplant patients (Załęska-Kocięcka *et al.*, 2017), however this is attributed to glomerular damage as opposed to tubule damage since increased serum levels are associated with decreased glomerular filtration (Argyropoulos *et al.*, 2017). In this study the increase in urinary B2M abundance (in cases) can be linked to the proposed mechanism by which TDF induces nephrotoxicity i.e. retention of TDF in proximal tubules leading to mitochondrial damage and cell death (Perazella, 2010). If TDF does cause proximal tubule cell death by this mechanism, then it is likely that there will be increased B2M in the urine due to decreased reabsorption. This can be confirmed by measuring serum/plasma B2M levels and should there be a normal level of B2M abundance, it can be concluded that adequate filtration is taking place however decreased reabsorption is leading to increased urinary levels.

Retinol binding protein-4 has also been widely reported as a biomarker for various forms of kidney injury (Che *et al.*, 2010; Domingos *et al.*, 2016; Gonzalez-Calero *et al.*, 2016). It is a transport protein that is responsible for carrying retinol within the body (Kotnik, Fischer-Posovszky and Wabitsch, 2011). Similar to B2M, RBP4 is filtered by the glomerulus and reabsorbed by the proximal tubules (Vaidya, Ferguson and Bonventre, 2008; Norden, Lapsley and Unwin, 2014). Therefore, the rise in urinary RBP4 observed in cases in this cohort would suggest decreased reabsorption by the proximal tubules due to tubular damage by TDF. Thus, further supporting the proposed mechanism of TDF induced nephrotoxicity.

The elevation of these proteins - B2M, RBP4, CYTC and A1AT - is known to be an indicator of tubular damage, and they have been shown to be significantly increased in a mixed population of American patients following cardiac surgery who develop kidney problems (Arthur *et al.*, 2014). Since the increased abundance of these four proteins show strong similarity to previously published data with respect to renal injury, this study supports the proposed mechanism by which TDF induces AKI described by Perazella (2010). This mechanism should be explored further as to the reason these potential markers for AKI are elevated.

Like A1AT and VTNC, another liver associated protein, alpha-2-HS-glycoprotein (Fetuin-A), displayed a 6-fold decrease in urinary abundance in cases. Fetuin-A is a multifactorial protein that has a wide range of functions including, but not limited to, endocytosis, mineralisation inhibition, protease inhibition and keratinocyte migration (Mori, Emoto and Inaba, 2011; Herrmann, Kinkeldey and Jahnen-Dechent, 2012). Studies have correlated urinary abundance of this protein to DN (Inoue *et al.*, 2013), kidney stones (Mehrsai *et al.*, 2017) and CKD (Vivekanandan-Giri *et al.*, 2011). The results presented here suggest that this protein is important in various diseases relating to the kidney and thus represents an important protein to monitor further.

Carbonic anhydrase (CAH1) has a well-established association with most segments of the nephron. It is particularly important in the proximal tubules where it is a key regulator of bicarbonate reabsorption (Purkerson and Schwartz, 2007). Cases in this cohort showed a 17-fold increase in urinary abundance of CAH1. Numerous urinary proteome analysis studies showed increases in CAH1 in diverse diseased states including: appendicitis (Kentsis *et al.*, 2010), breast cancer (Beretov *et al.*, 2015) and diabetes mellitus (Bellei *et al.*, 2008). Given its close association with proximal tubule cells and observed increase in urinary abundance due to TDF induced nephrotoxicity, CAH1 can represent a useful marker for AKI.

Several blood related proteins have shown significant differential urinary abundance in cases compared to controls in this cohort. These include: increases in haemoglobin subunit alpha and beta (HBA/B) by 9-fold, increases in fibrinogen beta and gamma chain (FIBB/G) by > 9-fold, antithrombin-III (ANT3) by 9-fold and a decrease in histidine-rich glycoprotein (HRG) by 9.5-fold. The proteins FIBB, FIBG and ANT3 play important roles in the blood coagulation cascade. Urinary fibrinogen has been reported as a sensitive and early diagnostic marker for AKI (Hoffmann *et al.*, 2012), and as a marker for CKD (Wang *et al.*, 2017). Haemoglobin is the key protein involved in gaseous exchange. It carries oxygen from the lungs to all parts of the body and returns carbon dioxide for excretion. The protein is made up of two alpha and two beta chains in a tetrahedral conformation (Khalighi *et al.*, 2014). Haemoglobin in the urine – haemoglobinuria – is indicative of acute nephrotoxicity. Deuel *et al.* (2016) suggested that haemoglobinuria may contribute to oxidative stress induced proximal tubule cell death contributing to AKI in mice. This suggests that the presence of excess HBA and HBB in the urine of cases could further contribute to TDF associated AKI by inducing oxidative stress mediated cell death.

Interestingly, and perhaps in response to oxidative damage, glutathione peroxidase 3 (GPX3) displayed an increase in urinary abundance by 8-fold in cases compared to controls. This protein is produced almost exclusively in the proximal tubule cells of the kidneys where it functions as a reactive oxygen species scavenger (Cutillas and Timms, 2010; Burk *et al.*, 2011). Acute kidney injury is synonymous with increased reactive oxygen species generation [reviewed in (Dennis and Witting, 2017)]. In the current population, mitochondrial damage leading to oxidative stress mediated cell death and further oxidative damage due to the presence of haemoglobinuria may account for TDF induced AKI. Increases in reactive oxygen species triggers excess production of GPX3 to scavenge the free radicals. This could account for the increase in urinary abundance observed and represents a unique and potentially useful finding to monitor further.

In the current study, cases displayed lower levels of the S100 calcium-binding protein A8 and A9 (S100A8/9) by 5.5-fold for S100A8 and 9-fold for S100A9 as compared to the controls. These proteins are calcium binding proteins that have been shown, in mice models, to be involved in macrophage mediated renal repair following injury (Dessing *et al.*, 2015). These two proteins have previously been shown to correlate with better outcome after kidney transplant rejection (Rekers *et al.*, 2016). The findings in this study suggest that monitoring these proteins for a decrease in concentration represents a potentially unique way to assess kidney damage.

Uromodulin, is the most abundantly expressed protein in healthy patient urine (Iorember and Vehaskari, 2014; Youhanna *et al.*, 2014; Devuyst and Bochud, 2015). It is produced exclusively in the kidney by the thick ascending loop of Henle (Devuyst and Bochud, 2015). The roles of uromodulin are not well defined however, it is believed to be important in maintaining osmotic balance and in the innate immune system in the kidney (Rampoldi *et al.*, 2011). *In vitro* studies show a range of functions such as activity against urinary tract infections, prevention of kidney stone formation and maintaining sodium-potassium pumps (Devuyst and Bochud, 2015). The protein was observed to be significantly decreased in patients in the current cohort possibly due, as reported elsewhere, to tubular damage leading to decreased excretion into the tubular lumen (Prajczer *et al.*, 2010). Quintana *et al.* (2009) reported similar observations where patients experiencing kidney damage expressed lower levels of uromodulin in their urine. These findings mirror the proposed mechanism by which TDF causes kidney injury i.e. by inducing aberrant functioning of the proximal tubules. It is possible that the damage caused by TDF on the proximal tubules leads to damage at the sites at which

uromodulin is produced and active thereby accounting for the decrease in production and excretion observed.

The proteins of the complement cascade, which is a component of the innate immune system, have been implicated in diseases and dysfunctions of the kidneys and in CKD in particular [see reviews (Fearn and Sheerin, 2015; Thurman, 2015)]. The patients classified as cases in this study showed elevated levels of a range of complement proteins as compared to the controls, including complement C3 (CO3), complement C4-B (CO4B), complement factor B (CFAB) and complement component C9 (CO9). Furthermore, GO analysis, using Reactome, showed a significant enrichment in the pathways related to the complement cascade and innate immune system. The proteins common to these pathways can potentially be targeted to develop a biomarker panel for the early detection of AKI. An elevation in complement proteins in the urine could indicate apoptotic cell death of the proximal tubules (Fishelson, Attali and Mevorach, 2001; Nauta *et al.*, 2003; Trouw, Blom and Gasque, 2008), which could result from mitochondrial mediated apoptotic pathways as a result of TDF induced mitochondrial injury (Perazella, 2010). Complement proteins and/or its derivatives can be routinely monitored after the initiation of ART and should a spike in concentration in the urine be observed, patients can be treated accordingly (TDF withdrawn) to minimise further ART-induced kidney complications.

Apolipoproteins were also significantly differentially abundant in this cohort. One of their major functions is the transport of lipids to various tissues and organs in the body and as co-factors for enzymes involved in lipid metabolism [reviewed in (Mahley *et al.*, 1984)]. Apolipoprotein D (ApoD) showed decreased levels, whilst apolipoproteins A-I (ApoA1) and A-IV (ApoA4) showed increased levels in the urine of cases compared to controls. Elevated serum levels of ApoA1 and ApoA4 have previously been shown to be associated with CKD in Austrian Caucasian patients (Boes *et al.*, 2006) and more recently in a multi-ethnic American population (Goek *et al.*, 2012). This study demonstrated the presence of these proteins in urine as potential markers for AKI in a South African population. Apolipoprotein A-1 is generally filtered through the glomerulus and reabsorbed by the proximal tubules (Kozyraki *et al.*, 2001; Dugué-Pujol *et al.*, 2007; Krikken *et al.*, 2010). A recent study by Madsen Svarrer *et al.* (2016), identified apolipoprotein M as a potential early urinary marker for AKI in children after heart surgery however this protein was not identified as a differentially abundant protein in the adult population in this study. Interestingly, Kasembeli *et al.* (2015) recently reported that genetic variants in the apolipoprotein L1 (APOL1) were associated with higher rates of kidney disease

in a HIV-positive South African cohort. Similar observations were reported by Friedman and Pollak (2016) after a study on an American population of African ancestry. Collectively, the available data suggests that apolipoproteins are important protein classes that can be exploited as biomarker targets for various diseases related to the kidney.

Albuminuria is a condition that is synonymous with kidney injury and has been studied in tenofovir induced kidney injury (Kelly *et al.*, 2013; Yombi *et al.*, 2015; Casado *et al.*, 2016), however there are mixed reports regarding its utility as a predictor for AKI. Zhang *et al.* (2013) and Meng *et al.* (2016) reported microalbuminuria as a good predictor of AKI, whereas Yim *et al.* (2015) reported that microalbuminuria is a poor marker for AKI. In this study, albuminuria detected by MS displayed aberrant abundance amongst cases and controls i.e. the change in abundance was not always in the same direction. Therefore, urinary albumin levels may not be an ideal marker for AKI in this cohort.

Similarly, alpha-1-microglobulin/bikunin precursor, alpha-1-antichymotrypsin, actin cytoplasmic 1, ganglioside GM2 activator, serotransferrin, basement membrane-specific heparan sulfate proteoglycan core protein, Ig kappa chain V-IV region, leucine-rich alpha-2-glycoprotein, immunoglobulin kappa light chain, polymeric immunoglobulin receptor, immunoglobulin heavy constant alpha 1 and annexin A1 all displayed a net absolute Log_2 fold change less than 2. This indicates that although their abundance levels met the criteria for inclusion as putative makers in individual experiments, the direction of the change in abundance was not uniform. Therefore, it is unlikely that these proteins would serve as a biomarker with great sensitivity and specificity. Even if these proteins were employed in a biomarker signature (panel) their individual contributions to the panel would be difficult to compensate for due to the non-uniform direction of change that was observed for each of them.

The proteins profilin-1 (PROF1), calreticulin (CALR) and fibulin-5 (FBLN) displayed an increase in urinary abundance by 8-fold. These proteins have been reported as potential urinary markers for bladder cancer (Kageyama *et al.*, 2009; Zoidakis *et al.*, 2012; Lu *et al.*, 2014; Frantzi and Vlahou, 2017). Supervillin (SVIL) displayed a decrease in urinary abundance by 9-fold in cases compared to controls in this cohort. The protein is ubiquitously expressed in almost all tissues in the body and functions as a structural support member by binding to actin (Khurana and George, 2008). The mechanisms by which these proteins (PROF1, CALR, FBLN and SVIL) change in urinary abundance in relation to AKI is yet to be determined. However,

this still represents valuable observations that can be exploited in the development of a biomarker panel for AKI diagnosis.

The ubiquitously expressed osteoglycin (mimecan) was also significantly elevated, by 8-fold, amongst cases in this cohort. The roles of mimecan are diverse due to it being implicated in both normal and diseased states in various organs and tissues [reviewed in (Deckx, Heymans and Papageorgiou, 2016)]. Baek *et al.* (2017) recently reported an association with elevated serum levels of mimecan and cardio/cerebrovascular events in patients with advanced CKD. To date, there has been no report of urinary mimecan as a putative marker for AKI. This finding represents a potentially unique marker for ARV induced AKI.

In summary, this exploratory study has identified several proteins that bear the potential to be used as urinary biomarkers for AKI diagnosis, in particular, during TDF use or during the use of drugs that elicit renal injury through similar common mechanisms. These proteins, however, require verification and validation before they can be adapted (either alone or as a biomarker signature) as potential diagnostic indicators of AKI in a clinical setting.

Chapter 5

Conclusion and future perspectives

This study has achieved the aim of developing novel, simple and robust methods for isolating human urinary proteins for whole-proteome clinical studies. Furthermore, one of the methods was applied to comprehensively identify, for the first time on the African continent, changes in urinary protein abundance that underlie acute kidney injury in HIV positive patients undergoing first-line antiretroviral therapy. The results presented here serve as the foundation for proteomics to be applied, clinically, in a precision medicine-based manner.

The current study presents data on two novel urinary proteome profiling methods which demonstrate great promise for use in future urinary proteomics research. The MagReSyn® HILIC microsphere-based method displayed suitability for small scale studies in which urine can be processed immediately. Furthermore, it is an automatable protocol, which could potentially make it suitable for large scale studies due to higher throughput and lower variation introduced by multiple user handling steps. The Norgen kit-based method allows for simple, robust and reproducible sample collection, storage and isolation of human urinary proteins, which is of utmost importance in clinical proteomics research. The ease of use and robust nature of this method made it easy to implement and highly efficient at the point-of-care clinic where patients were recruited.

The kit-based approach was applied to a cohort of HIV positive, South African patients undergoing first line TDF-based therapy, and represented the first attempt at urinary proteome profiling in a South African population using high-throughput liquid-chromatography MS. The approach demonstrated the unbiased detection of the range of proteins ordinarily present in urine samples with high confidence. Moreover, by applying the method in a large, clinically relevant cohort for a study of this nature, important changes in the urinary proteome that are associated with AKI due to TDF use were detected. These changes (putative biomarkers) will be further assessed in an independent, blinded verification study cohort before moving into full scale validation studies.

Despite the numerous putative protein biomarkers identified using stringent filtering criteria, this study presented a few notable limitations. Firstly, studies focussed on protein isoforms have shown associations of particular isoforms and disease, disease progression and disease severity, particularly in cancer (Ahmed *et al.*, 2005; Pal, Gupta and Davuluri, 2012; Wei, Zaika and Zaika, 2012; Zhang and Chen, 2016). Evaluation of proteomic isoforms was not undertaken in this study as this would greatly increase the size of the search space and would require significantly more computational power than was available for this study. However, since the SWATH™ approach generates a complete and "re-mineable" digital proteomic map for each sample (Guo *et al.*, 2015; Pernikářová and Bouchal, 2015; Dwivedi *et al.*, 2016), isoform analysis can be computed from the collected data in the future. The data analysis could potentially lead to the identification of specific isoforms of proteins that are associated with ARV-induced AKI. A similar limitation applies in the case of post-translational modifications, where potentially relevant functional changes associated with protein modifications were not assessed.

Secondly, to derive definitive conclusions, orthogonal validation of identified markers using an enzyme-linked immunosorbent assay (ELISA) for example is required, and this was not done in this study. This study was an exploratory study which presents data on which further studies can be based. Thirdly, the putative markers identified in this study need thorough independent verification and subsequent validation before the remaining putative markers can be developed further and assessed for their relevance for diagnostic use at clinical level (point-of-care). Fourthly, there exists a great degree of overlap between the proteins identified in this cohort as potential urinary biomarkers for AKI with proteins reported in other studies as serum/plasma and/or urinary biomarkers for CKD. This can potentially be addressed by using large ($n > 1000$), independent cohorts with well-defined AKI diagnosis, thereby creating an "AKI urinary proteome" which can be compared to the large volumes of data already available for CKD. Should there be any protein or sets of proteins that are significantly different between or unique to each disease state, then these would be most suitable for the development of targeted tests to distinguish between each disease state independently. Lastly, it is unlikely that all putative proteins identified in this study will be suitable for development into routinely used clinical tests. This is due to MS being more diverse than most other technologies, such as ELISAs, which are commonly used in these types of diagnostic assays. A protein (or protein panel) will need to be sufficiently abundant in urine and have suitable antibodies either currently available or easily synthesised/expressed. Mass spectrometry based diagnosis is

currently not yet ready for mainstream clinical adoption (Nedelkov, 2017b), especially in developing nations.

Despite these limitations, the current study has developed new methods in the urinary proteomics field as well as identified proteins associated with AKI of which some are novel in relation to this disease state. Furthermore, a range of proteins have been confidently identified that strongly correlate to the proposed mechanism by which TDF induces nephrotoxicity. Together, these findings contribute to the advancement of the human clinical proteomics field at large. Ultimately, the findings presented here: a) aim to become mainstream urinary proteome profiling method/s of choice, and, b) have laid the groundwork for the development of a biomarker signature for AKI diagnosis, with high sensitivity and specificity, particularly in HIV positive patients undergoing first-line antiretroviral treatment in South Africa.

Chapter 6

References

Aberg, J. A. (2009) 'Cardiovascular Complications in HIV Management: Past, Present, and Future', *J. Acquir. Immune Defic. Syndr.*, 50(1), pp. 54–64.

Adachi, J., Kumar, C., Zhang, Y., Olsen, J. V, & Mann, M. (2006) 'The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins', *Genome Biol.*, 7(9), p. R80.

Aebersold, R. & Mann, M. (2003) 'Mass spectrometry-based proteomics', *Nature*, 422(6928), pp. 198–207.

Afkarian, M., Bhasin, M., Dillon, S. T., Guerrero, M. C., Nelson, R. G., Knowler, W. C., Thadhani, R., & Libermann, T. A. (2010) 'Optimizing a proteomics platform for urine biomarker discovery', *Mol. Cell. Proteomics*, 9(10), pp. 2195–2204.

Afsarimanesh, N., Mukhopadhyay, S. C., & Kruger, M. (2018) 'Sensing technologies for monitoring of bone-health: A review', *Sensors Actuators A Phys.*, 274, pp. 165–178.

Aggarwal, S. & Yadav, A. K. (2016) 'False Discovery Rate Estimation in Proteomics', in *Methods Mol. Biol.*, pp. 119–128.

Ahmed, N., Oliva, K. T., Barker, G., Hoffmann, P., Reeve, S., Smith, I. A., Quinn, M. A., & Rice, G. E. (2005) 'Proteomic tracking of serum protein isoforms as screening biomarkers of ovarian cancer', *Proteomics*, 5(17), pp. 4625–4636.

Albalat, A., Mischak, H., & Mullen, W. (2011) 'Clinical application of urinary proteomics/peptidomics', *Expert Rev. Proteomics*, 8(5), pp. 615–629.

Altman, N. & Krzywinski, M. (2015) 'Sources of variation', *Nat. Methods*, 12(1), pp. 5–6.

Anderson, N. L. & Anderson, N. G. (1998) 'Proteome and proteomics: New technologies, new concepts, and new words', *Electrophoresis*, 19(11), pp. 1853–1861.

Andrews, G. L., Dean, R. A., Hawkridge, A. M., & Muddiman, D. C. (2011) 'Improving Proteome Coverage on a LTQ-Orbitrap Using Design of Experiments', *J. Am. Soc. Mass Spectrom.*, 22(4), pp. 773–783.

Anjo, S. I., Santa, C., & Manadas, B. (2017) 'SWATH-MS as a tool for biomarker discovery: From basic research to clinical applications', *Proteomics*, 17(3–4), p. 1600278.

Apetrei, C., Robertson, D. L., & Marx, P. A. (2004) 'The history of SIVS and AIDS: epidemiology, phylogeny and biology of isolates from naturally SIV infected non-human primates (NHP) in Africa', *Front. Biosci.* 2004/02/10, 9, pp. 225–254.

Argyropoulos, C. P., Chen, S. S., Ng, Y.-H., Roumelioti, M.-E., Shaffi, K., Singh, P. P., & Tzamaloukas, A. H. (2017) 'Rediscovering Beta-2 Microglobulin As a Biomarker across the Spectrum of Kidney Diseases.', *Front. Med.*, 4, p. 73.

Armstrong, R. A. (2014) ‘When to use the Bonferroni correction’, *Ophthalmic Physiol. Opt.*, 34(5), pp. 502–508.

Arthur, J. M., Hill, E. G., Alge, J. L., Lewis, E. C., Neely, B. A., Janech, M. G., Tumlin, J. A., Chawla, L. S., & Shaw, A. D. (2014) ‘Evaluation of 32 urine biomarkers to predict the progression of acute kidney injury after cardiac surgery’, *Kidney Int.*, 85(2), pp. 431–438.

Arts, E. J. & Hazuda, D. J. (2012) ‘HIV-1 Antiretroviral Drug Therapy’, *Cold Spring Harb. Perspect. Med.*, 2(4), p. a007161.

Asada, T., Isshiki, R., Hayase, N., Sumida, M., Inokuchi, R., Noiri, E., Nangaku, M., Yahagi, N., & Doi, K. (2016) ‘Impact of clinical context on acute kidney injury biomarker performances: differences between neutrophil gelatinase-associated lipocalin and L-type fatty acid-binding protein.’, *Sci. Rep.*, 6, p. 33077.

Asscher, A. W., Sussman, M., Waters, W. E., Davis, R. H., & Chick, S. (1966) ‘Urine as a medium for bacterial growth.’, *Lancet (London, England)*, 2(7472), pp. 1037–41.

Atkinson, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., Oates, J. A., Peck, C. C., Schooley, R. T., Spilker, B. A., Woodcock, J., & Zeger, S. L. (2001) ‘Biomarkers and surrogate endpoints: preferred definitions and conceptual framework’, *Clin. Pharmacol. Ther.* 2001/03/10, 69(3), pp. 89–95.

Aurelius, G. (2009) ‘BACTERIAL GROWTH IN URINE’, *Acta Pathol. Microbiol. Scand.*, 55(2), pp. 203–208.

Baek, S. H., Cha, R., Kang, S. W., Park, C. W., Cha, D. R., Kim, S. S. G. S., Yoon, S. A., Kim, S. S. G. S., Han, S. Y., Park, J. H., Chang, J. H., Lim, C. S., Kim, Y. S., & Na, K. Y. (2017) ‘Higher Serum Levels of Osteoglycin Are Associated with All-Cause Mortality and Cardiovascular and Cerebrovascular Events in Patients with Advanced Chronic Kidney Disease’, *Tohoku J. Exp. Med.*, 242(4), pp. 281–290.

Baldwin, M. A. (2004) ‘Protein Identification by Mass Spectrometry’, *Mol. Cell. Proteomics*, 3(1), pp. 1–9.

Banerjee, S. & Mazumdar, S. (2012) ‘Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte’, *Int. J. Anal. Chem.*, 2012, pp. 1–40.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., & Kuster, B. (2007) ‘Quantitative mass spectrometry in proteomics: a critical review’, *Anal. Bioanal. Chem.*, 389(4), pp. 1017–1031.

Barattini, V. & Humphryes, P. (2015) *Enhanced Digestion for Improved Biomarker Identification*.

Bateman, A., Martin, M. J., O’Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Castro, L. G., Figueira, L., Garmiri, P., Georghiou, G., Gonzalez, D., Hatton-Ellis, E., Li, W., Liu, W., Lopez, R., Luo, J., Lussi, Y., MacDougall, A., Nightingale, A., Palka, B., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Speretta, E., Turner, E., Tyagi, N., Volynkin, V., Wardell, T., Warner, K., Watkins, X., Zaru, R., Zellner, H., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., ArgoudPuy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M. C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., De Castro, E., Coudert, E., Cuche, B., Doche, M., Dornevil,

D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nouspikel, N., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A. L., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C., Chen, Y., Garavelli, J. S., Huang, H., Laiho, K., McGarvey, P., Natale, D. A., Ross, K., Vinayaka, C. R., Wang, Q., Wang, Y., Yeh, L. S., & Zhang, J. (2017) 'UniProt: The universal protein knowledgebase', *Nucleic Acids Res.*, 45(D1), pp. D158–D169.

Bautista, D. S., Denstedt, J., Chambers, A. F., & Harris, J. F. (1996) 'Low-molecular-weight variants of osteopontin generated by serine proteinases in urine of patients with kidney stones.', *J. Cell. Biochem.*, 61(3), pp. 402–409.

Baxmann, A. C., Ahmed, M. S., Marques, N. C., Menon, V. B., Pereira, A. B., Kirsztajn, G. M., & Heilberg, I. P. (2008) 'Influence of Muscle Mass and Physical Activity on Serum and Urinary Creatinine and Serum Cystatin C', *Clin. J. Am. Soc. Nephrol.*, 3(2), pp. 348–354.

Beasley-Green, A. (2016) 'Urine Proteomics in the Era of Mass Spectrometry.', *Int. Neurourol. J.*, 20(Suppl 2), pp. S70-75.

Bellei, E., Rossi, E., Lucchi, L., Uggeri, S., Albertazzi, A., Tomasi, A., & Iannone, A. (2008) 'Proteomic analysis of early urinary biomarkers of renal changes in type 2 diabetic patients', *Proteomics Clin. Appl.* 2008/04/01, 2(4), pp. 478–491.

Beltrami, E. M., Williams, I. T., Shapiro, C. N., & Chamberland, M. E. (2000) 'Risk and management of blood-borne infections in health care workers.', *Clin. Microbiol. Rev.*, 13(3), pp. 385–407.

Benjamini, Y. & Hochberg, Y. (1995) 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *J. R. Stat. Soc. Ser. B*, 57(1), pp. 289–300.

Benjamini, Y. & Yekutieli, D. (2001) 'The Control of the False Discovery Rate in Multiple Testing under Dependency', *Ann. Stat.*, pp. 1165–1188.

Beretov, J., Wasinger, V. C., Millar, E. K. A., Schwartz, P., Graham, P. H., & Li, Y. (2015) 'Proteomic Analysis of Urine to Identify Breast Cancer Biomarker Candidates Using a Label-Free LC-MS/MS Approach', *PLoS One*. Edited by A. Aboussekhra, 10(11), p. e0141876.

Beretov, J., Wasinger, V. C., Schwartz, P., Graham, P. H., & Li, Y. (2014) 'A Standardized and Reproducible Urine Preparation Protocol for Cancer Biomarkers Discovery', *Biomark. Cancer*, 6, p. BIC.S17991.

Bhandari, S. (2006) 'How to measure renal function in clinical practice: age affects estimated glomerular filtration rate', *Br. Med. J.*, 333(7574), p. 918.

Blainey, P., Krzywinski, M., & Altman, N. (2014) 'Replication', *Nat. Methods*, 11(9), pp. 879–880.

Blumenthal, R., Durell, S., & Viard, M. (2012) 'HIV entry and envelope glycoprotein-mediated fusion.', *J. Biol. Chem.*, 287(49), pp. 40841–9.

Blydt-Hansen, T. D., Gibson, I. W., Gao, A., Dufault, B., & Ho, J. (2015) 'Elevated urinary CXCL10-to-creatinine ratio is associated with subclinical and clinical rejection in pediatric renal transplantation.', *Transplantation*, 99(4), pp. 797–804.

Boes, E., Fliser, D., Ritz, E., König, P., Lhotta, K., Mann, J. F. E., Müller, G. A., Neyer, U., Riegel, W., Riegler, P., Kronenberg, F., & Group, for the M. S. (2006) 'Apolipoprotein A-IV predicts progression of chronic kidney disease: the mild to moderate kidney disease study.', *J. Am. Soc. Nephrol.*, 17(2), pp. 528–36.

Borràs, E. & Sabidó, E. (2017) 'What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry', *Proteomics*, 17(17–18), p. 1700180.

Bradford, M. M. (1976) 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Anal. Biochem.*, 72(1–2), pp. 248–254.

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*.

Burk, R. F., Olson, G. E., Winfrey, V. P., Hill, K. E., & Yin, D. (2011) 'Glutathione peroxidase-3 produced by the kidney binds to a population of basement membranes in the gastrointestinal tract and in other tissues.', *Am. J. Physiol. Gastrointest. Liver Physiol.*, 301(1), pp. G32-8.

Callister, S. J., Barry, R. C., Adkins, J. N., Johnson, E. T., Qian, W., Webb-Robertson, B.-J. M., Smith, R. D., & Lipton, M. S. (2006) 'Normalization Approaches for Removing Systematic Biases Associated with Mass Spectrometry and Label-Free Proteomics', *J. Proteome Res.*, 5(2), pp. 277–286.

Cantley, L. G., Colangelo, C. M., Stone, K. L., Chung, L., Belcher, J., Abbott, T., Cantley, J. L., Williams, K. R., & Parikh, C. R. (2016) 'Development of a Targeted Urine Proteome Assay for kidney diseases.', *Proteomics. Clin. Appl.*, 10(1), pp. 58–74.

Casado, J. L., Bañón, S., Santiuste, C., Serna, J., Guzman, P., Tenorio, M., Liaño, F., & Rey, J. M. del (2016) 'Prevalence and significance of proximal renal tubular abnormalities in HIV-infected patients receiving tenofovir', *AIDS*, 30(2).

Case, K. (1986) 'Nomenclature: human immunodeficiency virus', *Ann. Intern. Med.*, 105(1), p. 133.

Caterino, M., Zacchia, M., Costanzo, M., Bruno, G., Arcaniolo, D., Trepiccione, F., Siciliano, R. A., Mazzeo, M. F., Ruoppolo, M., & Capasso, G. (2018) 'Urine proteomics revealed a significant correlation between urine-Fibronectin abundance and estimated-GFR decline in patients with bardet-Biedl syndrome', *Kidney Blood Press. Res.*, 43(2), pp. 389–405.

Catherman, A. D., Skinner, O. S., & Kelleher, N. L. (2014) 'Top Down proteomics: Facts and perspectives', *Biochem. Biophys. Res. Commun.*, 445(4), pp. 683–693.

Celis, J. E., Rasmussen, H. H., Vorum, H., Madsen, P., Honoré, B., Wolf, H., & Orntoft, T. F. (1996) 'Bladder squamous cell carcinomas express psoriasin and externalize it to the urine.', *J. Urol.*, 155(6), pp. 2105–12.

Chambers, S. T., Peddie, B. A., Randall, K., & Lever, M. (1999) 'Inhibitors of bacterial growth in urine: what is the role of betaines?', *Int. J. Antimicrob. Agents*, 11(3–4), pp. 293–6.

Chandramouli, K. & Qian, P.-Y. (2009) 'Proteomics: challenges, techniques and possibilities to overcome biological sample complexity.', *Hum. Genomics Proteomics*, 2009.

Chaykovska, L., Heunisch, F., von Einem, G., Alter, M. L., Hocher, C.-F., Tsuprykov, O., Dschietzig, T., Kretschmer, A., & Hocher, B. (2016) 'Urinary Vitamin D Binding Protein and KIM-1 Are Potent New Biomarkers of Major Adverse Renal Events in Patients Undergoing Coronary Angiography.', *PLoS One*, 11(1), p. e0145723.

Che, M., Xie, B., Xue, S., Dai, H., Qian, J., Ni, Z., Axelsson, J., & Yan, Y. (2010) 'Clinical usefulness of novel biomarkers for the detection of acute kidney injury following elective cardiac surgery.', *Nephron. Clin. Pract.*, 115(1), pp. c66-72.

Chen, B., Brown, K. A., Lin, Z., & Ge, Y. (2018) 'Top-Down Proteomics: Ready for Prime Time?', *Anal. Chem.*, 90(1), pp. 110–127.

Chernushevich, I. V., Loboda, A. V., & Thomson, B. A. (2001) 'An introduction to quadrupole-time-of-flight mass spectrometry', *J. Mass Spectrom.*, 36(8), pp. 849–865.

Chevalier, F. (2010) 'Highlights on the capacities of "Gel-based" proteomics', *Proteome Sci.*, 8, p. 23.

du Cheyron, D., Daubin, C., Poggioli, J., Ramakers, M., Houillier, P., Charbonneau, P., & Paillard, M. (2003) 'Urinary measurement of Na+/H+ exchanger isoform 3 (NHE3) protein as new marker of tubule injury in critically ill patients with ARF.', *Am. J. Kidney Dis.*, 42(3), pp. 497–506.

Cho, Y., Johnson, D. W., Vesey, D. A., Hawley, C. M., Clarke, M., & Topley, N. (2015) 'Utility of Urinary Biomarkers in Predicting Loss of Residual Renal Function: The balANZ Trial.', *Perit. Dial. Int.*, 35(2), pp. 159–171.

Choi, A. I., Shlipak, M. G., Hunt, P. W., Martin, J. N., & Deeks, S. G. (2009) 'HIV-infected persons continue to lose kidney function despite successful antiretroviral therapy', *AIDS*, 23(16), pp. 2143–2149.

Christianson, C. C., Johnson, C. J., & Needham, S. R. (2013) 'The advantages of microflow LC–MS/MS compared with conventional HPLC–MS/MS for the analysis of methotrexate from human plasma', *Bioanalysis*, 5(11), pp. 1387–1396.

Ciszek, M., Paczek, L., Bartlomiejczyk, I., & Mucha, K. (2006) 'Urine cytokines profile in renal transplant patients with asymptomatic bacteriuria.', *Transplantation*, 81(12), pp. 1653–1657.

De Clercq, E. (2011) 'Discovery and Development of Tenofovir Disoproxil Fumarate', in *Antivir. Drugs*, pp. 85–101.

Cockcroft, D. W. & Gault, M. H. (1976) 'Prediction of Creatinine Clearance from Serum Creatinine', *Nephron*, 16(1), pp. 31–41.

Cole, R. B. (2000) 'Some tenets pertaining to electrospray ionization mass spectrometry.', *J. Mass Spectrom.*, 35(7), pp. 763–72.

Collins, B. C., Gillet, L. C., Rosenberger, G., Röst, H. L., Vichalkovski, A., Gstaiger, M., & Aebersold, R. (2013) 'Quantifying protein interaction dynamics by SWATH mass spectrometry: Application to the 14-3-3 system', *Nat. Methods*, 10(12), pp. 1246–1253.

Collins, B. C., Hunter, C. L., Liu, Y., Schilling, B., Rosenberger, G., Bader, S. L., Chan, D. W., Gibson, B. W., Gingras, A. C., Held, J. M., Hirayama-Kurogi, M., Hou, G., Krisp, C., Larsen, B., Lin, L., Liu, S., Molloy, M. P., Moritz, R. L., Ohtsuki, S., Schlapbach, R., Selevsek, N., Thomas, S. N., Tzeng, S. C., Zhang, H., & Aebersold, R. (2017) 'Multi-

laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry', *Nat. Commun.*, 8(1).

Colquhoun, D. (2014) 'An investigation of the false discovery rate and the misinterpretation of p-values', *R. Soc. Open Sci.*, 1(3), pp. 140216–140216.

Cottrell, J. S. (2011) 'Protein identification using MS/MS data', *J. Proteomics*, 74(10), pp. 1842–1851.

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., & Mann, M. (2014) 'Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ.', *Mol. Cell. Proteomics*, 13(9), pp. 2513–26.

Cox, J. & Mann, M. (2008) 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nat. Biotechnol.*, 26(12), pp. 1367–1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., & Mann, M. (2011) 'Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment', *J. Proteome Res.*, 10(4), pp. 1794–1805.

Craig, R. & Beavis, R. C. (2004) 'TANDEM: matching proteins with tandem mass spectra', *Bioinformatics*, 20(9), pp. 1466–1467.

Cutillas, P. R. & Timms, J. F. (2010) 'Approaches and Applications of Quantitative LC-MS for Proteomics and Activitomics', in *Mol. Cell. Biol.*, pp. 3–17.

Daggulli, M., Utangac, M. M., Dede, O., Bodakci, M. N., Hatipoglu, N. K., Penbegul, N., Sancaktutar, A. A., Bozkurt, Y., & Soylemez, H. (2016) 'Potential biomarkers for the early detection of acute kidney injury after percutaneous nephrolithotripsy.', *Ren. Fail.*, 38(1), pp. 151–156.

Dau, B. & Holodniy, M. (2008) 'The Relationship Between HIV Infection and Cardiovascular Disease', *Curr. Cardiol. Rev.*, 4(3), pp. 203–218.

Dawson, P. H. & York, A. (1976) *Quadrupole Mass Spectrometry and its Applications*.

Deckx, S., Heymans, S., & Papageorgiou, A.-P. (2016) 'The diverse functions of osteoglycin: a deceitful dwarf, or a master regulator of disease?', *FASEB J.*, 30(8), pp. 2651–2661.

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(10), pp. 1850–1862.

Delanaye, P., Pottel, H., & Botev, R. (2013) 'Con: Should we abandon the use of the MDRD equation in favour of the CKD-EPI equation?', *Nephrol. Dial. Transplant.*, 28(6), pp. 1396–1403.

Dennen, P., Altmann, C., Kaufman, J., Klein, C. L., Andres-Hernando, A., Ahuja, N. H., Edelstein, C. L., Cadnapaphornchai, M. A., Keniston, A., & Faubel, S. (2010) 'Urine interleukin-6 is an early biomarker of acute kidney injury in children undergoing cardiac surgery.', *Crit. Care*, 14(5), p. R181.

Dennis, J. M. & Witting, P. K. (2017) 'Protective Role for Antioxidants in Acute Kidney Disease', *Nutrients*, 9(7), p. 718.

Dessing, M. C., Tammaro, A., Pulskens, W. P., Teske, G. J., Butter, L. M., Claessen, N., van

Eijk, M., van der Poll, T., Vogl, T., Roth, J., Florquin, S., & Leemans, J. C. (2015) 'The calcium-binding protein complex S100A8/A9 has a crucial role in controlling macrophage-mediated renal repair following ischemia/reperfusion.', *Kidney Int.*, 87(1), pp. 85–94.

Deuel, J. W., Schaer, C. A., Boretti, F. S., Opitz, L., Garcia-Rubio, I., Baek, J. H., Spahn, D. R., Buehler, P. W., & Schaer, D. J. (2016) 'Hemoglobinuria-related acute kidney injury is driven by intrarenal oxidative reactions triggering a heme toxicity response.', *Cell Death Dis.*, 7(1), p. e2064.

Devarajan, P., Krawczeski, C. D., Nguyen, M. T., Kathman, T., Wang, Z., & Parikh, C. R. (2010) 'Proteomic identification of early biomarkers of acute kidney injury after cardiac surgery in children', *Am J Kidney Dis Off J Natl Kidney Found*, 56(4), pp. 632–642.

Devuyst, O. & Bochud, M. (2015) 'Uromodulin, kidney function, cardiovascular disease, and mortality.', *Kidney Int.*, 88(5), pp. 944–6.

Doi, K., Katagiri, D., Negishi, K., Hasegawa, S., Hamasaki, Y., Fujita, T., Matsubara, T., Ishii, T., Yahagi, N., Sugaya, T., & Noiri, E. (2012) 'Mild elevation of urinary biomarkers in prerenal acute kidney injury.', *Kidney Int.*, 82(10), pp. 1114–1120.

Doi, K., Noiri, E., Nangaku, M., Yahagi, N., Jayakumar, C., & Ramesh, G. (2014) 'Repulsive guidance cue semaphorin 3A in urine predicts the progression of acute kidney injury in adult patients from a mixed intensive care unit.', *Nephrol. Dial. Transplant*, 29(1), pp. 73–80.

Domingos, M. A. M., Moreira, S. R., Gomez, L., Goulart, A., Lotufo, P. A., Benseñor, I., & Titan, S. (2016) 'Urinary Retinol-Binding Protein: Relationship to Renal Function and Cardiovascular Risk Factors in Chronic Kidney Disease', *PLoS One*. Edited by H. Mischak, 11(9), p. e0162782.

Domon, B. & Aebersold, R. (2006) 'Mass Spectrometry and Protein Analysis', *Science (80-.).*, 312(5771), pp. 212–217.

Doms, R. W. & Moore, J. P. (2000) 'HIV-1 membrane fusion: targets of opportunity.', *J. Cell Biol.*, 151(2), pp. F9-14.

Dong, Q., Yan, X., Liang, Y., & Stein, S. E. (2016) 'In-Depth Characterization and Spectral Library Building of Glycopeptides in the Tryptic Digest of a Monoclonal Antibody Using 1D and 2D LC–MS/MS', *J. Proteome Res.*, 15(5), pp. 1472–1486.

Dorsey, J. G. & Dill, K. A. (1989) 'The molecular mechanism of retention in reversed-phase liquid chromatography', *Chem. Rev.*, 89(2), pp. 331–346.

Dowell, J. A., Vander Heyden, W., & Li, L. (2006) 'Rat Neuropeptidomics by LC-MS/MS and MALDI-FTMS: Enhanced Dissection and Extraction Techniques Coupled with 2D RP-RP HPLC', *J. Proteome Res.*, 5(12), pp. 3368–3375.

Du, Y., Zappitelli, M., Mian, A., Bennett, M., Ma, Q., Devarajan, P., Mehta, R., & Goldstein, S. L. (2011) 'Urinary biomarkers to detect acute kidney injury in the pediatric emergency center.', *Pediatr. Nephrol.*, 26(2), pp. 267–274.

Duarte, T. T. & Spencer, C. T. (2016) 'Personalized Proteomics: The Future of Precision Medicine', *Proteomes*, 4(4).

Dugué-Pujol, S., Rousset, X., Château, D., Pastier, D., Klein, C., Demeurie, J., Cywiner-Golenzer, C., Chabert, M., Verroust, P., Chambaz, J., Châtelet, F.-P., & Kalopissis, A.-D. (2007) 'Apolipoprotein A-II is catabolized in the kidney as a function of its plasma

concentration', *J. Lipid Res.*, 48(10), pp. 2151–2161.

Dunn, M. J. (2014) 'Clinical Proteomics Advances', *PROTEOMICS – Clin. Appl.*, 8(1–2), pp. 1–2.

Dunn, M. J. & Kraus, H. J. (2016) 'PROTEOMICS: The first 15 years', *Proteomics Clin. Appl.*, 10(1), pp. 1–3.

Dwivedi, R. C., Navarrete, M., Choi, N., Spicer, V., Rigatto, C., Arora, R. C., Krokhin, O., Ho, J., & Wilkins, J. A. (2016) 'A proteomic evaluation of urinary changes associated with cardiopulmonary bypass', *Clin. Proteomics*, 13(1), p. 17.

Egertson, J. D., Kuehn, A., Merrihew, G. E., Bateman, N. W., MacLean, B. X., Ting, Y. S., Canterbury, J. D., Marsh, D. M., Kellmann, M., Zabrouskov, V., Wu, C. C., & MacCoss, M. J. (2013) 'Multiplexed MS/MS for improved data-independent acquisition.', *Nat. Methods*, 10(8), pp. 744–6.

El-Akawi, Z. J., Al-Hindawi, F. K., & Bashir, N. A. (2008) 'Alpha-1 antitrypsin (alpha1-AT) plasma levels in lung, prostate and breast cancer patients.', *Neuro Endocrinol. Lett.*, 29(4), pp. 482–4.

El-Aneed, A., Cohen, A., & Banoub, J. (2009) 'Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers', *Appl. Spectrosc. Rev.*, 44(3), pp. 210–230.

Elias, J. E. & Gygi, S. P. (2007) 'Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry', *Nat. Methods*, 4(3), pp. 207–214.

Elias, J. E. & Gygi, S. P. (2010) 'Target-decoy search strategy for mass spectrometry-based proteomics.', *Methods Mol. Biol.*, 604, pp. 55–71.

Endre, Z. H., Pickering, J. W., Walker, R. J., Devarajan, P., Edelstein, C. L., Bonventre, J. V, Frampton, C. M., Bennett, M. R., Ma, Q., Sabbisetti, V. S., Vaidya, V. S., Walcher, A. M., Shaw, G. M., Henderson, S. J., Nejat, M., Schollum, J. B. W., & George, P. M. (2011) 'Improved performance of urinary biomarkers of acute kidney injury in the critically ill by stratification for injury duration and baseline renal function.', *Kidney Int.*, 79(10), pp. 1119–1130.

Eng, J. K., Jahan, T. A., & Hoopmann, M. R. (2013) 'Comet: An open-source MS/MS sequence database search tool', *Proteomics*, 13(1), pp. 22–24.

Eng, J. K., Searle, B. C., Clauser, K. R., & Tabb, D. L. (2011) 'A face in the crowd: recognizing peptides through database search.', *Mol. Cell. Proteomics*, 10(11), p. R111.009522.

Erde, J., Loo, R. R. O., & Loo, J. A. (2014) 'Enhanced FASP (eFASP) to Increase Proteome Coverage and Sample Recovery for Quantitative Proteomic Experiments', *J. Proteome Res.*, 13(4), pp. 1885–1895.

Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., MacCoss, M. J., & Rinner, O. (2012) 'Using iRT, a normalized retention time for more targeted measurement of peptides', *Proteomics*, 12(8), pp. 1111–1121.

Estrella, M. M., Fine, D. M., & Atta, M. G. (2010) 'Recent developments in HIV-related kidney disease', *HIV Ther.*, 4(5), pp. 589–603.

Ettre, L. S. & Sakodynskii, K. I. (1993) ‘M. S. Tswett and the discovery of chromatography I: Early work (1899–1903)’, *Chromatographia*, 35(3–4), pp. 223–231.

von Eynatten, M., Baumann, M., Heemann, U., Zdunek, D., Hess, G., Nawroth, P. P., Bierhaus, A., & Humpert, P. M. (2010) ‘Urinary L-FABP and anaemia: distinct roles of urinary markers in type 2 diabetes.’, *Eur. J. Clin. Invest.*, 40(2), pp. 95–102.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C. D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H., & D’Eustachio, P. (2018) ‘The Reactome Pathway Knowledgebase’, *Nucleic Acids Res.*, 46(D1), pp. D649–D655.

FDA Approval of HIV Medicines | AIDSinfo (2018).

Fearn, A. & Sheerin, N. S. (2015) ‘Complement activation in progressive renal disease.’, *World J. Nephrol.*, 4(1), pp. 31–40.

Fei, S. S., Wilmarth, P. A., Hitzemann, R. J., McWeeney, S. K., Belknap, J. K., & David, L. L. (2011) ‘Protein database and quantitative analysis considerations when integrating genetics and proteomics to compare mouse strains.’, *J. Proteome Res.*, 10(7), pp. 2905–12.

Feist, P. & Hummon, A. (2015) ‘Proteomic Challenges: Sample Preparation Techniques for Microgram-Quantity Protein Analysis from Biological Samples’, *Int. J. Mol. Sci.*, 16(12), pp. 3537–3563.

Fettig, J., Swaminathan, M., Murrill, C. S., & Kaplan, J. E. (2014) ‘Global Epidemiology of HIV’, *Infect. Dis. Clin. North Am.*, 28(3), pp. 323–337.

Fic, E., Kedracka-Krok, S., Jankowska, U., Pirog, A., & Dziedzicka-Wasylewska, M. (2010) ‘Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis’, *Electrophoresis*, 31(21), pp. 3573–3579.

Filler, G., Foster, J., Acker, A. M. Y., Lepage, N., Akbari, A., & Ehrich, J. H. H. (2005) ‘The Cockcroft-Gault formula should not be used in children’, *Kidney Int.*, 67(6), pp. 2321–2324.

Finehout, E. J. & Lee, K. H. (2004) ‘An introduction to mass spectrometry applications in biological research’, *Biochem. Mol. Biol. Educ.*, 32(2), pp. 93–100.

Fishelson, Z., Attali, G., & Mevorach, D. (2001) ‘Complement and apoptosis.’, *Mol. Immunol.*, 38(2–3), pp. 207–19.

Forshed, J. (2017) ‘Experimental Design in Clinical ‘Omics Biomarker Discovery’, *J. Proteome Res.*, 16(11), pp. 3954–3960.

Fournier, M. L., Gilmore, J. M., Martin-Brown, S. A., & Washburn, M. P. (2007) ‘Multidimensional Separations-Based Shotgun Proteomics’, *Chem. Rev.*, 107(8), pp. 3654–3686.

Frantzi, M. & Vlahou, A. (2017) ‘Ten years of proteomics in bladder cancer: Progress and future directions’, *Bl. Cancer*.

Friedenauer, S. & Berlet, H. H. (1989) ‘Sensitivity and variability of the Bradford protein assay in the presence of detergents.’, *Anal. Biochem.*, 178(2), pp. 263–8.

Friedman, D. J. & Pollak, M. R. (2016) ‘Apolipoprotein L1 and Kidney Disease in African Americans’, *Trends Endocrinol. Metab.*, 27(4), pp. 204–215.

Fu, C.-P., Lee, I.-T., Sheu, W. H.-H., Lee, W.-J., Liang, K.-W., Lee, W.-L., & Lin, S.-Y. (2012) 'The levels of circulating and urinary monocyte chemoattractant protein-1 are associated with chronic renal injury in obese men.', *Clin. Chim. Acta.*, 413(19–20), pp. 1647–1651.

Fu, Q., Kowalski, M. P., Mastali, M., Parker, S. J., Sobhani, K., Van Den Broek, I., Hunter, C. L., & Van Eyk, J. E. (2018) 'Highly Reproducible Automated Proteomics Sample Preparation Workflow for Quantitative Mass Spectrometry', *J. Proteome Res.*, 17(1), pp. 420–428.

Fufaa, G. D., Weil, E. J., Nelson, R. G., Hanson, R. L., Bonventre, J. V, Sabbisetti, V., Waikar, S. S., Mifflin, T. E., Zhang, X., Xie, D., Hsu, C.-Y., Feldman, H. I., Coresh, J., Vasan, R. S., Kimmel, P. L., & Liu, K. D. (2015) 'Association of urinary KIM-1, L-FABP, NAG and NGAL with incident end-stage renal disease and mortality in American Indians with type 2 diabetes mellitus.', *Diabetologia*, 58(1), pp. 188–198.

Fung, H. B., Stone Ea Fau - Piacenti, F. J., Piacenti, F. J., Clin, T., Stone, E. A., & Piacenti, F. J. (2002) 'Tenofovir disoproxil fumarate: a nucleotide reverse transcriptase inhibitor for the treatment of HIV infection', *Clin. Ther.*, 24(10), pp. 1515–1548.

Gajbhiye, A., Dabhi, R., Taunk, K., Vannuruswamy, G., RoyChoudhury, S., Adhav, R., Seal, S., Mane, A., Bayatigeri, S., Santra, M. K., Chaudhury, K., & Rapole, S. (2016) 'Urinary proteome alterations in HER2 enriched breast cancer revealed by multipronged quantitative proteomics', *Proteomics*, 16(17), pp. 2403–2418.

Gale, P. J. & Vestal, M. L. (2016) 'The Development of Time-of-Flight Mass Spectrometry', in *Encycl. Mass Spectrom.*, pp. 34–42.

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*. 1999/02/16, 397(6718), pp. 436–441.

Garbis, S., Lubec, G., & Fountoulakis, M. (2005) 'Limitations of current proteomics technologies.', *J. Chromatogr. A*, 1077(1), pp. 1–18.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003) 'ExPASy: The proteomics server for in-depth protein knowledge and analysis', *Nucleic Acids Res.*, 31(13), pp. 3784–3788.

de Geus, H. R. H., Fortrie, G., Betjes, M. G. H., van Schaik, R. H. N., & Groeneveld, A. B. J. (2013) 'Time of injury affects urinary biomarker predictive values for acute kidney injury in critically ill, non-septic patients.', *BMC Nephrol.*, 14, p. 273.

Geyer, P. E., Holdt, L. M., Teupser, D., & Mann, M. (2017) 'Revisiting biomarker discovery by plasma proteomics.', *Mol. Syst. Biol.*, 13(9), p. 942.

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', *Cell Syst.*, 2(3), pp. 185–195.

Gibbons, B. C., Chambers, M. C., Monroe, M. E., Tabb, D. L., & Payne, S. H. (2015) 'Correcting systematic bias and instrument measurement drift with mzRefinery.', *Bioinformatics*, 31(23), pp. 3838–40.

Gillet, L. C., Leitner, A., & Aebersold, R. (2016) 'Mass Spectrometry Applied to Bottom-Up

Proteomics: Entering the High-Throughput Era for Hypothesis Testing', *Annu. Rev. Anal. Chem.*, 9(1), pp. 449–472.

Giunta, B., Ehrhart, J., Obregon, D. F., Lam, L., Le, L., Jin, J., Fernandez, F., Tan, J., & Shytle, R. D. (2011) 'Antiretroviral medications disrupt microglial phagocytosis of amyloid and increase its production by neurons: Implications for HIV-associated neurocognitive disorders', *Mol Brain*, 4, p. 23.

Glish, G. L. & Vachet, R. W. (2003) *The basics of mass spectrometry in the twenty-first century, Nat. Rev. Drug Discov.*

Goek, O.-N., Köttgen, A., Hoogeveen, R. C., Ballantyne, C. M., Coresh, J., & Astor, B. C. (2012) 'Association of apolipoprotein A1 and B with kidney function and chronic kidney disease in two multiethnic population samples', *Nephrol. Dial. Transplant.*, 27(7), pp. 2839–2847.

Goknar, N., Oktem, F., Ozgen, I. T., Torun, E., Kucukkoc, M., Demir, A. D., & Cesur, Y. (2015) 'Determination of early urinary renal injury markers in obese children.', *Pediatr. Nephrol.*, 30(1), pp. 139–144.

Goldfarb, D., Wang, W., & Major, M. B. (2016) 'MSAcquisitionSimulator: data-dependent acquisition simulator for LC-MS shotgun proteomics.', *Bioinformatics*, 32(8), pp. 1269–71.

Gonzalez-Calero, L., Martin-Lorenzo, M., Ramos-Barron, A., Ruiz-Criado, J., Maroto, A. S., Ortiz, A., Gomez-Alamillo, C., Arias, M., Vivanco, F., & Alvarez-Llamas, G. (2016) 'Urinary Kininogen-1 and Retinol binding protein-4 respond to Acute Kidney Injury: predictors of patient prognosis?', *Sci. Rep.*, 6(1), p. 19667.

González Fernández-Niño, S. M., Smith-Moritz, A. M., Chan, L. J. G., Adams, P. D., Heazlewood, J. L., & Petzold, C. J. (2015) 'Standard flow liquid chromatography for shotgun proteomics in bioenergy research.', *Front. Bioeng. Biotechnol.*, 3, p. 44.

Govaert, E., Van Steendam, K., Willems, S., Vossaert, L., Dhaenens, M., & Deforce, D. (2017) 'Comparison of fractionation proteomics for local SWATH library building', *Proteomics*, 17(15–16), p. 1700052.

Graves, P. R. & Haystead, T. A. J. (2002) 'Molecular Biologist's Guide to Proteomics', *Microbiol. Mol. Biol. Rev.*, 66(1), pp. 39–63.

Gregorich, Z. R., Chang, Y.-H., & Ge, Y. (2014) 'Proteomics in heart failure: top-down or bottom-up?', *Pflugers Arch.*, 466(6), pp. 1199–209.

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.*, p. Unit10.25.

Günthard, H. F., Saag, M. S., Benson, C. A., & et al. (2016) 'Antiretroviral drugs for treatment and prevention of hiv infection in adults: 2016 recommendations of the international antiviral society–usa panel', *JAMA*, 316(2), pp. 191–210.

Guo, T., Kouvonen, P., Koh, C. C., Gillet, L. C., Wolski, W. E., Röst, H. L., Rosenberger, G., Collins, B. C., Blum, L. C., Gillessen, S., Joerger, M., Jochum, W., Aebersold, R., & Rost, H. L. (2015) 'Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps', *Nat Med*, 21(4), pp. 407–413.

Gupta, K. K., Donahue, D. L., Sandoval-Cooper, M. J., Castellino, F. J., & Ploplis, V. A.

(2015) ‘Abrogation of Plasminogen Activator Inhibitor-1-Vitronectin Interaction Ameliorates Acute Kidney Injury in Murine Endotoxemia’, *PLoS One*. Edited by J. Fu, 10(3), p. e0120728.

Gwinner, W. (2007) ‘Renal transplant rejection markers’, *World J. Urol.*, 25(5), pp. 445–455.

Haag, A. M. (2016) ‘Mass Analyzers and Mass Spectrometers’, in, pp. 157–169.

Hamburg, M. A. & Collins, F. S. (2010) ‘The Path to Personalized Medicine’, *N. Engl. J. Med.*, 363(4), pp. 301–304.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., Hirsch, M. S., & Merigan, T. C. (1996) ‘A Trial Comparing Nucleoside Monotherapy with Combination Therapy in HIV-Infected Adults with CD4 Cell Counts from 200 to 500 per Cubic Millimeter’, *N. Engl. J. Med.*, 335(15), pp. 1081–1090.

Han, X., Aslanian, A., Yates, J. R., & III (2008) ‘Mass spectrometry for proteomics.’, *Curr. Opin. Chem. Biol.*, 12(5), pp. 483–90.

Harper, J. W. & Bennett, E. J. (2016) ‘Proteome complexity and the forces that drive proteome imbalance’, *Nature*, pp. 328–338.

Hawkridge, A. M. (2014) ‘Practical Considerations and Current Limitations in Quantitative Mass Spectrometry-based Proteomics’, in *Quant. Proteomics*, pp. 1–25.

Heise, D., Rentsch, K., Braeuer, A., Friedrich, M., & Quintel, M. (2011) ‘Comparison of urinary neutrophil glucosaminidase-associated lipocalin, cystatin C, and alpha1-microglobulin for early detection of acute renal injury after cardiac surgery.’, *Eur. J. Cardiothorac. Surg.*, 39(1), pp. 38–43.

Hemelaar, J. (2012) ‘The origin and diversity of the HIV-1 pandemic’, *Trends Mol. Med.*, 18(3), pp. 182–192.

Henzel, W. J., Watanabe, C., & Stults, J. T. (2003) ‘Protein identification: the origins of peptide mass fingerprinting’, *J. Am. Soc. Mass Spectrom.*, 14(9), pp. 931–942.

Herrmann, M., Kinkeldey, A., & Jahnen-Dechent, W. (2012) ‘Fetuin-A Function in Systemic Mineral Metabolism’, *Trends Cardiovasc. Med.*, 22(8), pp. 197–201.

Higgins, C. (2016) ‘Urea and creatinine concentration, the urea: creatinine ratio’, *Acutecaretesting.Org*, (October), pp. 1–8.

Ho, C. S., Lam, C. W. K., Chan, M. H. M., Cheung, R. C. K., Law, L. K., Lit, L. C. W., Ng, K. F., Suen, M. W. M., & Tai, H. L. (2003) ‘Electrospray ionisation mass spectrometry: principles and clinical applications.’, *Clin. Biochem. Rev.*, 24(1), pp. 3–12.

Ho, J., Lucy, M., Krokhin, O., Hayglass, K., Pascoe, E., & Darroch, G. (2009) ‘Mass spectrometry-based proteomic analysis of urine in acute kidney injury following cardiopulmonary bypass: a nested case-control study’, *Am J Kidney Dis Off J Natl Kidney Found*, 53.

Ho, J., Wiebe, C., Rush, D. N., Rigatto, C., Storsley, L., Karpinski, M., Gao, A., Gibson, I. W., & Nickerson, P. W. (2013) ‘Increased urinary CCL2: Cr ratio at 6 months is associated with late renal allograft loss.’, *Transplantation*, 95(4), pp. 595–602.

Hoffmann, C., Rockstroh, J. K., & Rockstroh, J. K. (2015) ‘The Basics’, *HIV 2015/16*. 23rd

edn, pp. 22–23.

Hoffmann, D., Bijol, V., Krishnamoorthy, A., Gonzalez, V. R., Frendl, G., Zhang, Q., Goering, P. L., Brown, R. P., Waikar, S. S., & Vaidya, V. S. (2012) 'Fibrinogen excretion in the urine and immunoreactivity in the kidney serves as a translational biomarker for acute kidney injury', *Am. J. Pathol.*, 181(3), pp. 818–828.

Horstrup, J. H., Gehrmann, M., Schneider, B., Ploger, A., Froese, P., Schirop, T., Kampf, D., Frei, U., Neumann, R., & Eckardt, K.-U. (2002) 'Elevation of serum and urine levels of TIMP-1 and tenascin in patients with renal disease.', *Nephrol. Dial. Transplant*, 17(6), pp. 1005–1013.

Hortin, G. L. & Sviridov, D. (2010) 'The dynamic range problem in the analysis of the plasma proteome', *J. Proteomics*, 73(3), pp. 629–636.

Htun, N. M., Magliano, D. J., Zhang, Z.-Y., Lyons, J., Petit, T., Nkuipou-Kenfack, E., Ramirez-Torres, A., von zur Muhlen, C., Maahs, D., Schanstra, J. P., Pontillo, C., Pejchinovski, M., Snell-Bergeon, J. K., Delles, C., Mischak, H., Staessen, J. A., Shaw, J. E., Koeck, T., & Peter, K. (2017) 'Prediction of acute coronary syndromes by urinary proteome analysis', *PLoS One*, 12(3), p. e0172036.

Hu, H., Aizenstein, B. D., Puchalski, A., Burmania, J. A., Hamawy, M. M., & Knechtle, S. J. (2004) 'Elevation of CXCR3-binding chemokines in urine indicates acute renal-allograft dysfunction.', *Am. J. Transplant*, 4(3), pp. 432–437.

Huang, Y.-S., Chan, C.-K., Tsai, M.-S., Lee, K.-Y., Lin, S.-W., Chang, S.-Y., Hung, C.-C., & Chang, S.-C. (2015) 'Kidney dysfunction associated with tenofovir exposure in human immunodeficiency virus-1-infected Taiwanese patients', *J. Microbiol. Immunol. Infect.*

Hudson, A., Hunter, C., Schilling, B., Seymour, S., Johansen, E., & Gibson, B. (2014) *Exploring Impact of Dynamic Accumulation for Improving MS/MS Quality of QqTOF Data*.

Hunter, C. (2015) *What is the Difference between the Global and the Local FDR Rates?*, *SCIEX*.

Hussey, S. E., Sharoff, C. G., Garnham, A., Yi, Z., Bowen, B. P., Mandarino, L. J., & Hargreaves, M. (2013) 'Effect of exercise on the skeletal muscle proteome in patients with type 2 diabetes.', *Med. Sci. Sports Exerc.*, 45(6), pp. 1069–76.

Hustoft, H. K., Malerod, H., Wilson, S. R., Reubsaet, L., Lundanes, E., & Greibrokk, T. (2012) 'A critical review of trypsin digestion for LC-MS based proteomics', in *Integr. Proteomics*.

Inoue, K., Wada, J., Eguchi, J., Nakatsuka, A., Teshigawara, S., Murakami, K., Ogawa, D., Terami, T., Katayama, A., Tone, A., Iseda, I., Hida, K., Yamada, M., Ogawa, T., & Makino, H. (2013) 'Urinary Fetuin-A Is a Novel Marker for Diabetic Nephropathy in Type 2 Diabetes Identified by Lectin Microarray', *PLoS One*, 8(10), p. e77118.

Iorember, F. M. & Vehaskari, V. M. (2014) 'Uromodulin: old friend with new roles in health and disease', *Pediatr. Nephrol.*, 29(7), pp. 1151–1158.

Issaq, H. J., Chan, K. C., Janini, G. M., Conrads, T. P., & Veenstra, T. D. (2005) *Multidimensional separation of peptides for effective proteomic analysis*.

Itzhak, D. N., Davies, C., Tyanova, S., Mishra, A., Williamson, J., Antrobus, R., Cox, J., Weekes, M. P., & Borner, G. H. H. (2017) 'A Mass Spectrometry-Based Approach for

Mapping Protein Subcellular Localization Reveals the Spatial Proteome of Mouse Primary Neurons.', *Cell Rep.*, 20(11), pp. 2706–2718.

Jaffar, S., Grant, A. D., Whitworth, J., Smith, P. G., & Whittle, H. (2004) 'The natural history of HIV-1 and HIV-2 infections in adults in Africa: a literature review', *Bull. World Health Organ.* 2004/09/11, 82(6), pp. 462–469.

Jandera, P., Henze, G., Jandera, P., & Henze, G. (2011) 'Liquid Chromatography, 1. Fundamentals, History, Instrumentation, Materials', in *Ullmann's Encycl. Ind. Chem.*

Jang, H. R., Kim, S. M., Lee, Y.-J., Lee, J. E., Huh, W., Kim, D. J., Oh, H. Y., & Kim, Y.-G. (2012) 'The origin and the clinical significance of urinary angiotensinogen in proteinuric IgA nephropathy patients.', *Ann. Med.*, 44(5), pp. 448–457.

Jayakumar, C., Ranganathan, P., Devarajan, P., Krawczeski, C. D., Looney, S., & Ramesh, G. (2013) 'Semaphorin 3A is a new early diagnostic biomarker of experimental and pediatric acute kidney injury.', *PLoS One*, 8(3), p. e58446.

Jiang, L., He, L., & Fountoulakis, M. (2004) 'Comparison of protein precipitation methods for sample preparation prior to proteomic analysis', *J. Chromatogr. A*, 1023(2), pp. 317–320.

Kageyama, S., Isono, T., Matsuda, S., Ushio, Y., Satomura, S., Terai, A., Arai, Y., Kawakita, M., Okada, Y., & Yoshiki, T. (2009) 'Urinary calreticulin in the diagnosis of bladder urothelial carcinoma', *Int. J. Urol.*, 16(5), pp. 481–486.

Kaiser, T., Kamal, H., Rank, A., Kolb, H.-J., Holler, E., Ganser, A., Hertenstein, B., Mischak, H., & Weissinger, E. M. (2004) 'Proteomics applied to the clinical follow-up of patients after allogeneic hematopoietic stem cell transplantation', *Blood*, 104(2), pp. 340–349.

Kalantari, S., Jafari, A., Moradpoor, R., Ghasemi, E., & Khalkhal, E. (2015) 'Human Urine Proteomics: Analytical Techniques and Clinical Applications in Renal Diseases', *Int J Proteomics*. 2015/12/23, 2015, p. 782798.

Kalli, A. & Hess, S. (2012) 'Effect of mass spectrometric parameters on peptide and protein identification rates for shotgun proteomic experiments on an LTQ-orbitrap mass analyzer', *Proteomics*, 12(1), pp. 21–31.

Kalli, A., Smith, G. T., Sweredoski, M. J., & Hess, S. (2013) 'Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: Focus on LTQ-orbitrap mass analyzers', *J. Proteome Res.*, 12(7), pp. 3071–3086.

Kane, S. (2018) *Sample Size Calculator*.

Kang, M. J., Park, Y.-J., You, S., Yoo, S.-A., Choi, S., Kim, D.-H., Cho, C.-S., Yi, E. C., Hwang, D., & Kim, W.-U. (2014) 'Urinary Proteome Profile Predictive of Disease Activity in Rheumatoid Arthritis', *J. Proteome Res.*, 13(11), pp. 5206–5217.

Karp, N. A. (2010) 'Mass Spectrometry for Microbial Proteomics: Issues in Data Analysis with Electrophoretic or Mass Spectrometric Expression Proteomic Data', in *Mass Spectrom. Microb. Proteomics*, pp. 423–440.

Karpievitch, Y. V., Taverner, T., Adkins, J. N., Callister, S. J., Anderson, G. A., Smith, R. D., & Dabney, A. R. (2009) 'Normalization of peak intensities in bottom-up MS-based proteomics using singular value decomposition', *Bioinformatics*, 25(19), pp. 2573–2580.

Kasembeli, A. N., Duarte, R., Ramsay, M., Mosiane, P., Dickens, C., Dix-Peek, T., Limou,

S., Sezgin, E., Nelson, G. W., Fogo, A. B., Goetsch, S., Kopp, J. B., Winkler, C. A., & Naicker, S. (2015) 'APOL1 Risk Variants Are Strongly Associated with HIV-Associated Nephropathy in Black South Africans', *J. Am. Soc. Nephrol.*

Katagiri, D., Doi, K., Honda, K., Negishi, K., Fujita, T., Hisagi, M., Ono, M., Matsubara, T., Yahagi, N., Iwagami, M., Ohtake, T., Kobayashi, S., Sugaya, T., & Noiri, E. (2012) 'Combination of two urinary biomarkers predicts acute kidney injury after adult cardiac surgery.', *Ann. Thorac. Surg.*, 93(2), pp. 577–583.

Kay, R., Barton, C., Ratcliffe, L., Matharoo-Ball, B., Brown, P., Roberts, J., Teale, P., & Creaser, C. (2008) 'Enrichment of low molecular weight serum proteins using acetonitrile precipitation for mass spectrometry based proteomic analysis', *Rapid Commun. Mass Spectrom.*, 22(20), pp. 3255–3260.

Kelleher, N. L. (2004) 'Top-down proteomics.', *Anal. Chem.*, 76(11), p. 197A–203A.

Kelly, M. D., Gibson, A., Bartlett, H., Rowling, D., & Patten, J. (2013) 'Tenofovir-associated proteinuria', *AIDS*, 27(3), pp. 479–481.

Kentsis, A., Lin, Y. Y., Kurek, K., Calicchio, M., Wang, Y. Y., Monigatti, F., Campagne, F., Lee, R., Horwitz, B., Steen, H., & Bachur, R. (2010) 'Discovery and Validation of Urine Markers of Acute Pediatric Appendicitis Using High-Accuracy Mass Spectrometry', *Ann. Emerg. Med.*, 55(1), p. 62–70.e4.

Kertesz-Farkas, A., Reiz, B., P. Myers, M., & Pongor, S. (2012) 'Database Searching in Mass Spectrometry Based Proteomics', *Curr. Bioinform.*, 7(2), pp. 221–230.

Khalighi, M. A., Henriksen, K. J., Chang, A., & Meehan, S. M. (2014) 'March hemoglobinuria-associated acute tubular injury', *Clin. Kidney J.*, 7(5), pp. 488–489.

Khurana, S. & George, S. P. (2008) 'Regulation of cell structure and function by actin-binding proteins: Villin's perspective', *FEBS Lett.*, pp. 2128–2139.

Khwaja, A. (2012) 'KDIGO Clinical Practice Guidelines for Acute Kidney Injury', *Nephron Clin. Pract.*, 120(4), pp. c179–c184.

Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., Thomas, J. K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabuddhe, N. A., Balakrishnan, L., Advani, J., George, B., Renuse, S., Selvan, L. D. N., Patil, A. H., Nanjappa, V., Radhakrishnan, A., Prasad, S., Subbannayya, T., Raju, R., Kumar, M., Sreenivasamurthy, S. K., Marimuthu, A., Sathe, G. J., Chavan, S., Datta, K. K., Subbannayya, Y., Sahu, A., Yelamanchi, S. D., Jayaram, S., Rajagopalan, P., Sharma, J., Murthy, K. R., Syed, N., Goel, R., Khan, A. A., Ahmad, S., Dey, G., Mudgal, K., Chatterjee, A., Huang, T.-C., Zhong, J., Wu, X., Shaw, P. G., Freed, D., Zahari, M. S., Mukherjee, K. K., Shankar, S., Mahadevan, A., Lam, H., Mitchell, C. J., Shankar, S. K., Satishchandra, P., Schroeder, J. T., Sirdeshmukh, R., Maitra, A., Leach, S. D., Drake, C. G., Halushka, M. K., Prasad, T. S. K., Hruban, R. H., Kerr, C. L., Bader, G. D., Iacobuzio-Donahue, C. A., Gowda, H., & Pandey, A. (2014) 'A draft map of the human proteome', *Nature*, 509(7502), pp. 575–581.

King, E. (1995) 'AZT delta study: major European/Australian study finds combinations better than AZT alone', *AIDS Treat News*. 1995/10/06, (no 232), pp. 3–4.

Klont, F., Bras, L., Wolters, J. C., Ongay, S., Bischoff, R., Halmos, G. B., & Horvatovich, P. P. (2018) 'Assessment of Sample Preparation Bias in Mass Spectrometry-Based Proteomics',

Anal. Chem.

Koh, H. M. & Suresh, K. (2016) 'Tenofovir-induced nephrotoxicity: A retrospective cohort study', *Med. J. Malaysia*, 71(6), pp. 308–312.

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. Investig.*, 89(5), pp. 513–519.

Kohler, M., Schänzer, W., & Thevis, M. (2015) 'Effects of Exercise on the Urinary Proteome', in *Adv. Exp. Med. Biol.*, pp. 121–131.

Kondeti, R. R., Mulpuri, K. S., & Meruga, B. (2014) 'Advancements in column chromatography : A review', *World J. Pharm. Sci.*

Konigsberg, W. (1972) '[13] Reduction of disulfide bonds in proteins with dithiothreitol', *Methods Enzymol.*, 25, pp. 185–188.

Kotnik, P., Fischer-Posovszky, P., & Wabitsch, M. (2011) 'RBP4: A controversial adipokine', *Eur. J. Endocrinol.*, pp. 703–711.

Kovacevic, L., Lu, H., Goldfarb, D. S., Lakshmanan, Y., & Caruso, J. A. (2015) 'Urine proteomic analysis in cystinuric children with renal stones.', *J. Pediatr. Urol.*, 11(4), p. 217.e1-6.

Koyner, J. L., Coca, S. G., Thiessen-Philbrook, H., Patel, U. D., Shlipak, M. G., Garg, A. X., & Parikh, C. R. (2015) 'Urine Biomarkers and Perioperative Acute Kidney Injury: The Impact of Preoperative Estimated GFR.', *Am. J. Kidney Dis.*, 66(6), pp. 1006–1014.

Kozyraki, R., Fyfe, J., Verroust, P. J., Jacobsen, C., Dautry-Varsat, A., Gburek, J., Willnow, T. E., Christensen, E. I., & Moestrup, S. K. (2001) 'Megalin-dependent cubilin-mediated endocytosis is a major pathway for the apical uptake of transferrin in polarized epithelia', *Proc. Natl. Acad. Sci.*, 98(22), pp. 12491–12496.

Krikken, J. A., Gansevoort, R. T., Dullaart, R. P. F., & PREVEND Study Group, on behalf of the P. S. (2010) 'Lower HDL-C and apolipoprotein A-I are related to higher glomerular filtration rate in subjects without kidney disease.', *J. Lipid Res.*, 51(7), pp. 1982–90.

Krzywinski, M. & Altman, N. (2014) 'Designing comparative experiments', *Nat. Methods*, 11(6), pp. 597–598.

Kuckova, S., Sandu, I. C. A., Crhova, M., Hynek, R., Fogas, I., & Schafer, S. (2013) 'Protein identification and localization using mass spectrometry and staining tests in cross-sections of polychrome samples', *J. Cult. Herit.*, 14(1), pp. 31–37.

Kulak, N. A., Geyer, P. E., & Mann, M. (2017) 'Loss-less Nano-fractionator for High Sensitivity, High Coverage Proteomics', *Mol. Cell. Proteomics*, 16(4), pp. 694–705.

Lacroix, C., Caubet, C., Gonzalez-de-Peredo, A., Breuil, B., Bouyssié, D., Stella, A., Garrigues, L., Le Gall, C., Raevel, A., Massoubre, A., Klein, J., Decramer, S., Sabourdy, F., Bandin, F., Burlet-Schiltz, O., Monsarrat, B., Schanstra, J.-P., & Bascands, J.-L. (2014) 'Label-free quantitative urinary proteomics identifies the arginase pathway as a new player in congenital obstructive nephropathy.', *Mol. Cell. Proteomics*, 13(12), pp. 3421–34.

Lahoud, Y., Hussein, O., Shalabi, A., Nativ, O., Awad, H., Khamaisi, M., Matar, I., Nativ, O., & Abassi, Z. (2015) 'Effects of phosphodiesterase-5 inhibitor on ischemic kidney injury

during nephron sparing surgery: quantitative assessment by NGAL and KIM-1.', *World J. Urol.*, 33(12), pp. 2053–2062.

Langford, S. E., Ananworanich, J., & Cooper, D. A. (2007) 'Predictors of disease progression in HIV infection: a review', *AIDS Res. Ther.*, 4(1), p. 11.

Laskey, S. B. & Siliciano, R. F. (2014) 'A mechanistic theory to explain the efficacy of antiretroviral therapy', *Nat Rev Micro*, 12(11), pp. 772–780.

Latosinska, A., Vougas, K., Makridakis, M., Klein, J., Mullen, W., Abbas, M., Stravodimos, K., Katafigiotis, I., Merseburger, A. S., Zoidakis, J., Mischak, H., Vlahou, A., & Jankowski, V. (2015) 'Comparative Analysis of Label-Free and 8-Plex iTRAQ Approach for Quantitative Tissue Proteomic Analysis', *PLoS One*. Edited by L. Martens, 10(9), p. e0137048.

Lau, K. A. & Wong, J. J. L. (2013) 'Current trends of HIV recombination worldwide', *Infect. Dis. Rep.*, 5(1S), p. 4.

Lazar, C., Gatto, L., Ferro, M., Bruley, C., & Burger, T. (2016) 'Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies', *J. Proteome Res.*

Lebrecht, D., Venhoff, A. C., Kirschner, J., Wiech, T., Venhoff, N., & Walker, U. A. (2009) 'Mitochondrial tubulopathy in tenofovir disoproxil fumarate-treated rats', *J. Acquir. Immune Defic. Syndr.* 2009/07/08, 51(3), pp. 258–263.

León, I. R., Schwämmle, V., Jensen, O. N., & Sprenger, R. R. (2013) 'Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis.', *Mol. Cell. Proteomics*, 12(10), pp. 2992–3005.

Levey, A. S., Bosch, J. P., Lewis, J., & et al. (1999) 'A more accurate method to estimate glomerular filtration rate from serum creatinine: A new prediction equation', *Ann. Intern. Med.*, 130(6), pp. 461–470.

Levey, A. S., Stevens, L. A., Schmid, C. H., Zhang, Y., Castro, A. F., Feldman, H. I., Kusek, J. W., Eggers, P., Van Lente, F., Greene, T., & Coresh, J. (2009) 'A New Equation to Estimate Glomerular Filtration Rate', *Ann. Intern. Med.*, 150(9), pp. 604–612.

Li, Y. F. & Radivojac, P. (2012) 'Computational approaches to protein inference in shotgun proteomics.', *BMC Bioinformatics*, (Suppl 16), p. S4.

Li, Y., Fu, C., Zhou, X., Xiao, Z., Zhu, X., Jin, M., Li, X., & Feng, X. (2012) 'Urine interleukin-18 and cystatin-C as biomarkers of acute kidney injury in critically ill neonates.', *Pediatr. Nephrol.*, 27(5), pp. 851–860.

Likanonsakul, S., Suntisuklappon, B., Nitiyanontakij, R., Prasithsirikul, W., Nakayama, E. E., Shioda, T., & Sangsajja, C. (2016) 'A Single-Nucleotide Polymorphism in ABCC4 Is Associated with Tenofovir-Related Beta2-Microglobulinuria in Thai Patients with HIV-1 Infection', *PLoS One*. 2016/01/26, 11(1), p. e0147724.

Liu, B.-C., Zhang, L., Lv, L.-L., Wang, Y., Liu, D., & Zhang, X. (2006) 'Application of antibody array technology in the analysis of urinary cytokine profiles in patients with chronic kidney disease.', *Am. J. Nephrol.*, 26(5), pp. 483–490.

Liu, Y., Hüttenhain, R., Collins, B., & Aebersold, R. (2013) 'Mass spectrometric protein maps for biomarker discovery and clinical research', *Expert Rev. Mol. Diagn.*, 13(8), pp.

811–825.

Loor, J., Decruyenaere, J., Demeyere, K., Nuytinck, L., Hoste, E. A. J., Meyer, E., 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, p. 38.

Lovric, J. (2011) *Introducing proteomics : from concepts to sample separation, mass spectrometry and data analysis*.

Lu, C.-M., Lin, J.-J., Huang, H.-H., Ko, Y.-C., Hsu, J.-L., Chen, J.-C., Din, Z.-H., & Wu, Y.-J. (2014) 'A panel of tumor markers, calreticulin, annexin A2, and annexin A3 in upper tract urothelial carcinoma identified by proteomic and immunological analysis', *BMC Cancer*, 14(1), p. 363.

Lundgren, J. & Group, T. I. S. S. (2015) 'Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection', *N. Engl. J. Med.*, 373(9), pp. 795–807.

Luo, Q., Zhou, F., Dong, H., Wu, L., Chai, L., Lan, K., & Wu, M. (2013) 'Implication of combined urinary biomarkers in early diagnosis of acute kidney injury following percutaneous coronary intervention.', *Clin. Nephrol.*, 79(2), pp. 85–92.

Ma, J., Chen, X., Li, J.-S., Peng, L., Wei, S.-Y., Zhao, S.-L., Li, T., Zhu, D., He, Y.-X., Wei, Q.-J., & Li, B. (2015) 'Upregulation of podocyte-secreted angiopoietin-like-4 in diabetic nephropathy', *Endocrine*, 49(2), pp. 373–384.

Ma, Y., Li, Q., Wang, J., Xu, Z., Song, C., Zhuang, R., Yang, K., Yang, A., & Jin, B. (2010) 'Cystatin C, a novel urinary biomarker for sensitive detection of acute kidney injury during haemorrhagic fever with renal syndrome.', *Biomarkers Biochem. Indic. Expo. response, susceptibility to Chem.*, 15(5), pp. 410–417.

Maartens, G., Celum, C., & Lewin, S. R. (2014) 'HIV infection: epidemiology, pathogenesis, treatment, and prevention', *Lancet*, 384(9939), pp. 258–271.

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(3), pp. 299–307.

MacLean, B., Tomazela, D. M., Abbatiello, S. E., Zhang, S., Whiteaker, J. R., Paulovich, A. G., Carr, S. A., & Maccoss, M. J. (2010) 'Effect of Collision Energy Optimization on the Measurement of Peptides by Selected Reaction Monitoring (SRM) Mass Spectrometry', *Anal. Chem.*, 82(24), pp. 10116–10124.

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., & MacCoss, M. J. (2010) 'Skyline: An open source document editor for creating and analyzing targeted proteomics experiments', *Bioinformatics*, 26(7), pp. 966–968.

Madsen Svarrer, E. M., Andersen, H. Ø., Helvind, M., Slagman, M. C., Navis, G., Dullaart, R. P., Dahlbäck, B., & Nielsen, L. B. (2016) 'Urinary apolipoprotein M as a biomarker of acute kidney injury in children undergoing heart surgery', *Biomark. Med.*, 10(1), pp. 81–93.

Magdeldin, S., Enany, S., Yoshida, Y., Xu, B., Zhang, Y., Zureena, Z., Lokamani, I., Yaoita, E., & Yamamoto, T. (2014) 'Basics and recent advances of two dimensional- polyacrylamide gel electrophoresis.', *Clin. Proteomics*, 11(1), p. 16.

Magdeldin, S., Yoshida, Y., Li, H., Maeda, Y., Yokoyama, M., Enany, S., Zhang, Y., Xu, B., Fujinaka, H., Yaoita, E., Sasaki, S., & Yamamoto, T. (2012) 'Murine colon proteome and characterization of the protein pathways', *BioData Min.*, 5(1), p. 11.

Mahley, R. W., Innerarity, T. L., Rall, S. C., & Weisgraber, K. H. (1984) 'Plasma lipoproteins: apolipoprotein structure and function.', *J. Lipid Res.*, 25(12), pp. 1277–94.

Makris, K. & Spanou, L. (2016) 'Acute Kidney Injury: Diagnostic Approaches and Controversies.', *Clin. Biochem. Rev.*, 37(4), pp. 153–175.

Malyszko, J., Lukaszyk, E., Glowinska, I., & Durlik, M. (2015) 'Biomarkers of delayed graft function as a form of acute kidney injury in kidney transplantation.', *Sci. Rep.*, 5, p. 11684.

Mandili, G., Notarpietro, A., Khadjavi, A., Allasia, M., Battaglia, A., Lucatello, B., Frea, B., Turrini, F., Novelli, F., Giribaldi, G., & Destefanis, P. (2018) 'Beta-2-glycoprotein-1 and alpha-1-antitrypsin as urinary markers of renal cancer in von Hippel–Lindau patients', *Biomarkers*, 23(2), pp. 123–130.

Mann, M., Hendrickson, R. C., & Pandey, A. (2001) 'Analysis of Proteins and Proteomes by Mass Spectrometry', *Annu. Rev. Biochem.*, 70(1), pp. 437–473.

March, R. E. & Todd, J. F. J. (2016) 'The Development of the Quadrupole Mass Filter and Quadrupole Ion Trap', in *Encycl. Mass Spectrom.*, pp. 43–60.

Martensson, J., Glassford, N. J., Jones, S., Eastwood, G. M., Young, H., Peck, L., Ostland, V., Westerman, M., Venge, P., & Bellomo, R. (2015) 'Urinary neutrophil gelatinase-associated lipocalin to hepcidin ratio as a biomarker of acute kidney injury in intensive care unit patients.', *Minerva Anestesiol.*, 81(11), pp. 1192–1200.

Marx, J. L. (1984) 'Strong new candidate for AIDS agent', *Science (80-.).* 1984/05/04, 224(4648), pp. 475–477.

Marx, P. A., Apetrei, C., & Drucker, E. (2004) 'AIDS as a zoonosis? Confusion over the origin of the virus and the origin of the epidemics', *J. Med. Primatol.*, 33(5–6), pp. 220–226.

Masho, S. W., Wang Cl Fau - Nixon, D. E., Nixon, D. E., & Ther Clin Risk, M. (2007) 'Review of tenofovir-emtricitabine', *Ther. Clin. Risk Manag.*, 3(6), pp. 1097–1104.

Matafora, V., Zagato, L., Ferrandi, M., Molinari, I., Zerbini, G., Casamassima, N., Lanzani, C., Delli Carpini, S., Trepiccione, F., Manunta, P., Bachi, A., & Capasso, G. (2014) 'Quantitative proteomics reveals novel therapeutic and diagnostic markers in hypertension', *BBA Clin.*, 2, pp. 79–87.

Matsui, K., Kamijo-Ikemori, A., Hara, M., Sugaya, T., Kodama, T., Fujitani, S., Taira, Y., Yasuda, T., & Kimura, K. (2011) 'Clinical significance of tubular and podocyte biomarkers in acute kidney injury.', *Clin. Exp. Nephrol.*, 15(2), pp. 220–225.

Matsui, K., Kamijo-Ikemori, A., Sugaya, T., Yasuda, T., & Kimura, K. (2012) 'Usefulness of urinary biomarkers in early detection of acute kidney injury after cardiac surgery in adults.', *Circ. J.*, 76(1), pp. 213–220.

Mauvoisin, D., Dayon, L., Gachon, F., & Kussmann, M. (2015) 'Proteomics and circadian rhythms: It's all about signaling!', *Proteomics*, pp. 310–317.

McQueen, P. & Krokhin, O. (2012) 'Optimal selection of 2D reversed-phase–reversed-phase HPLC separation techniques in bottom-up proteomics', *Expert Rev. Proteomics*, 9(2), pp.

125–128.

Megger, D. A., Bracht, T., Meyer, H. E., & Sitek, B. (2013) ‘Label-free quantification in clinical proteomics’, *Biochim. Biophys. Acta - Bioenerg.*, 1834(8), pp. 1581–1590.

Megger, D. A., Pott, L. L., Ahrens, M., Padden, J., Bracht, T., Kuhlmann, K., Eisenacher, M., Meyer, H. E., & Sitek, B. (2014) ‘Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines’, *Biochim. Biophys. Acta - Proteins Proteomics*, 1844(5), pp. 967–976.

Mehrsai, A., Guitynavard, F., Nikoobakht, M. R., Gooran, S., & Ahmadi, A. (2017) ‘The relationship between serum and urinary Fetuin-A levels and kidney stone formation among kidney stone patients.’, *Cent. Eur. J. Urol.*, 70(4), pp. 394–399.

Meijer, E., Boertien, W. E., Nauta, F. L., Bakker, S. J. L., van Oeveren, W., Rook, M., van der Jagt, E. J., van Goor, H., Peters, D. J. M., Navis, G., de Jong, P. E., & Gansevoort, R. T. (2010) ‘Association of urinary biomarkers with disease severity in patients with autosomal dominant polycystic kidney disease: a cross-sectional analysis.’, *Am. J. Kidney Dis.*, 56(5), pp. 883–895.

Meintjes, G. (2014) *Adult antiretroviral therapy guidelines 2014, 2014*.

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. (2017) ‘Adult antiretroviral therapy guidelines 2017’, *South. African J. HIV Med. Vol 18, No 1 (2017)DO - 10.4102/sajhivmed.v18i1.776* .

Meng, H., Wu, P., Zhao, Y., Xu, Z., Wang, Z.-M., Li, C., Wang, L., & Yang, Z. (2016) ‘Microalbuminuria in patients with preserved renal function as a risk factor for contrast-Induced acute kidney injury following invasive coronary angiography’, *Eur. J. Radiol.*, 85(6), pp. 1063–1067.

Mesnard, L., Rafat, C., Vandermeersch, S., Hertig, A., Cathelin, D., Xu-Dubois, Y.-C., Jouanneau, C., Keller, A. C., Ribeil, J.-A., Leite-de-Moraes, M. C., & Rondeau, E. (2011) ‘Vitronectin dictates intraglomerular fibrinolysis in immune-mediated glomerulonephritis’, *FASEB J.*, 25(10), pp. 3543–3553.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., & Thomas, P. D. (2017) ‘PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements’, *Nucleic Acids Res.*, 45(D1), pp. D183–D189.

Mitulovic, G. & Mechtler, K. (2006) ‘HPLC techniques for proteomics analysis--a short overview of latest developments’, *Briefings Funct. Genomics Proteomics*, 5(4), pp. 249–260.

Mitulovic, G., Smoluch, M., Chervet, J.-P., Steinmacher, I., Kungl, A., & Mechtler, K. (2003) ‘An improved method for tracking and reducing the void volume in nano HPLC?MS with micro trapping columns’, *Anal. Bioanal. Chem.*, 376(7), pp. 946–951.

Mizuno, H., Ueda, K., Kobayashi, Y., Tsuyama, N., Todoroki, K., Zhe Min, J., & Toyo, T. (2016) ‘The great importance of normalization of LC-MS data for highly-accurate non-targeted metabolomics’.

Molitoris, B. A. (2012) ‘Measuring glomerular filtration rate in acute kidney injury: yes, but not yet’, *Crit. Care*. 2012/09/28, 16(5), p. 158.

Molloy, M. P., Brzezinski, E. E., Hang, J., McDowell, M. T., & VanBogelen, R. A. (2003) 'Overcoming technical variation and biological variation in quantitative proteomics', *Proteomics*, 3(10), pp. 1912–1919.

Moreno-González, D., Pérez-Ortega, P., Gilbert-López, B., Molina-Díaz, A., García-Reyes, J. F., & Fernández-Alba, A. R. (2017) 'Evaluation of nanoflow liquid chromatography high resolution mass spectrometry for pesticide residue analysis in food', *J. Chromatogr. A*, 1512, pp. 78–87.

Mori, K., Emoto, M., & Inaba, M. (2011) 'Fetuin-A: a multifunctional protein.', *Recent Pat. Endocr. Metab. Immune Drug Discov.*, 5(2), pp. 124–46.

Morris, H. R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Bordoli, R. S., Hoyes, J., & Bateman, R. H. (1996) 'High Sensitivity Collisionally-activated Decomposition Tandem Mass Spectrometry on a Novel Quadrupole/Orthogonal-acceleration Time-of-flight Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, 10(8), pp. 889–896.

Moskowitz-Kassai, E., MacKelaite, L., Chen, J., Patel, K., Dadhania, D. M., Gross, S. S., Chander, P., Delaney, V., Deng, L., Chen, L., Cui, X., Suthanthiran, M., & Goligorsky, M. S. (2012) 'Excretion of anti-angiogenic proteins in patients with chronic allograft dysfunction.', *Nephrol. Dial. Transplant*, 27(2), pp. 494–497.

Moulder, R., Bhosale, S. D., Goodlett, D. R., & Lahesmaa, R. (2017) 'Analysis of the plasma proteome using iTRAQ and TMT-based Isobaric labeling', *Mass Spectrom. Rev.*, 37(5), pp. 583–606.

Murray, K. K., Boyd, R. K., Eberlin, M. N., Langley, G. J., Li, L., & Naito, Y. (2013) 'Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013)', *Pure Appl. Chem.*, 85(7), pp. 1515–1609.

Musial, K., Bargenda, A., & Zwolinska, D. (2015) 'Urine matrix metalloproteinases and their extracellular inducer EMMPRIN in children with chronic kidney disease.', *Ren. Fail.*, 37(6), pp. 980–984.

Nagaraj, N. & Mann, M. (2011) 'Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome', *J Proteome Res*, 10(2), pp. 637–645.

Nagarajan, S. R., Brandon, A. E., McKenna, J. A., Shtein, H. C., Nguyen, T. Q., Suryana, E., Poronnik, P., Cooney, G. J., Saunders, D. N., & Hoy, A. J. (2017) 'Insulin and diet-induced changes in the ubiquitin-modified proteome of rat liver', *PLoS One*. Edited by M. Vinciguerra, 12(3), p. e0174431.

Nauta, A. J., Daha, M. R., van Kooten, C., & Roos, A. (2003) 'Recognition and clearance of apoptotic cells: a role for complement and pentraxins.', *Trends Immunol.*, 24(3), pp. 148–54.

Nedelkov, D. (2017a) 'Human proteoforms as new targets for clinical mass spectrometry protein tests', *Expert Rev. Proteomics*, 14(8), pp. 691–699.

Nedelkov, D. (2017b) 'Mass spectrometry protein tests: ready for prime time (or not)', *Expert Rev. Proteomics*, 14(1), pp. 1–7.

Needham, S. R. (2017) 'Microspray and microflow liquid chromatography: the way forward for LC–MS bioanalysis', *Bioanalysis*, 9(24), pp. 1935–1937.

Needham, S. R., Williams, M. D., & Bell, D. S. (2016) 'Microflow LC-MS-MS: The past, the present, and the path forward', *LC-GC North Am.*, 34(7), pp. 464–471.

Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., Van Sluyter, S. C., & Haynes, P. A. (2011) 'Less label, more free: Approaches in label-free quantitative mass spectrometry', *Proteomics*.

Nesvizhskii, A. I. & Aebersold, R. (2005) 'Interpretation of Shotgun Proteomic Data', *Mol. Cell. Proteomics*, 4(10), pp. 1419–1440.

Nickolas, T. L., Schmidt-Ott, K. M., Canetta, P., Forster, C., Singer, E., Sise, M., Elger, A., Maarouf, O., Sola-Del Valle, D. A., O'Rourke, M., Sherman, E., Lee, P., Geara, A., Imus, P., Guddati, A., Polland, A., Rahman, W., Elitok, S., Malik, N., Giglio, J., El-Sayegh, S., Devarajan, P., Hebbar, S., Saggi, S. J., Hahn, B., Kettritz, R., Luft, F. C., & Barasch, J. (2012) 'Diagnostic and prognostic stratification in the emergency department using urinary biomarkers of nephron damage: a multicenter prospective cohort study.', *J. Am. Coll. Cardiol.*, 59(3), pp. 246–255.

Nishijima, T., Komatsu, H., Gatanaga, H., Aoki, T., Watanabe, K., Kinai, E., Honda, H., Tanuma, J., Yazaki, H., Tsukada, K., Honda, M., Teruya, K., Kikuchi, Y., & Oka, S. (2011) 'Impact of small body weight on tenofovir-associated renal dysfunction in HIV-infected patients: a retrospective cohort study of Japanese patients', *PLoS One*. 2011/07/30, 6(7), p. e22661.

Niwa, T. (2008) 'Biomarker discovery for kidney diseases by mass spectrometry', *J. Chromatogr. B*, 870(2), pp. 148–153.

Norden, A. G. W., Lapsley, M., & Unwin, R. J. (2014) 'Urine retinol-binding protein 4: a functional biomarker of the proximal renal tubule.', *Adv. Clin. Chem.*, 63, pp. 85–122.

Nyamweya, S., Hegedus, A., Jaye, A., Rowland-Jones, S., Flanagan, K. L., & Macallan, D. C. (2013) 'Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis', *Rev. Med. Virol.* 2013/02/28, 23(4), pp. 221–240.

O'Farrell, P. H. (1975) 'High resolution two-dimensional electrophoresis of proteins.', *J. Biol. Chem.*, 250(10), pp. 4007–21.

O'Riordan, E., Orlova, T. N., Podust, V. N., Chander, P. N., Yanagi, S., Nakazato, M., Hu, R., Butt, K., Delaney, V., & Goligorsky, M. S. (2007) 'Characterization of urinary peptide biomarkers of acute rejection in renal allografts.', *Am. J. Transplant*, 7(4), pp. 930–940.

O'Seaghdha, C. M., Hwang, S.-J., Bhavsar, N. A., Kottgen, A., Coresh, J., Astor, B. C., & Fox, C. S. (2011) 'Lower urinary connective tissue growth factor levels and incident CKD stage 3 in the general population.', *Am. J. Kidney Dis.*, 57(6), pp. 841–849.

Oberg, A. L. & Vitek, O. (2009) 'Statistical design of quantitative mass spectrometry-based proteomic experiments', *J. Proteome Res.*, 8(5), pp. 2144–2156.

Olsen, J. V, Ong, S.-E., & Mann, M. (2004) 'Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues', *Mol. Cell. Proteomics*, 3(6), pp. 608–614.

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & Mann, M. (2002) 'Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.', *Mol. Cell. Proteomics*, 1(5), pp. 376–86.

Ostermann, M. & Joannidis, M. (2016) 'Acute kidney injury 2016: diagnosis and diagnostic workup.', *Crit. Care*, 20(1), p. 299.

Pal, S., Gupta, R., & Davuluri, R. V (2012) 'Alternative transcription and alternative splicing

in cancer.', *Pharmacol. Ther.*, 136(3), pp. 283–94.

Palella, F. J. & Phair, J. P. (2011) 'Cardiovascular Disease in HIV Infection', *Curr Opin HIV AIDS*, 6(4), pp. 266–271.

Palmisano, L. & Vella, S. (2011) 'A brief history of antiretroviral therapy of HIV infection: success and challenges', *Ann. Ist. Super. Sanita*. 2011/03/25, 47(1), pp. 44–48.

Pang, L., Li, Q., Li, Y., Liu, Y., Duan, N., & Li, H. (2018) 'Urine proteomics of primary membranous nephropathy using nanoscale liquid chromatography tandem mass spectrometry analysis', *Clin. Proteomics*, 15(1), p. 5.

Pantaleo, G., Graziosi, C., & Fauci, A. S. (1993) 'New concepts in the immunopathogenesis of human immunodeficiency virus infection', *N. Engl. J. Med.* 1993/02/04, 328(5), pp. 327–335.

Parikh, C. R., Abraham, E., Ancukiewicz, M., & Edelstein, C. L. (2005) 'Urine IL-18 is an early diagnostic marker for acute kidney injury and predicts mortality in the intensive care unit.', *J. Am. Soc. Nephrol.*, 16(10), pp. 3046–3052.

Parikh, C. R., Thiessen-Philbrook, H., Garg, A. X., Kadiyala, D., Shlipak, M. G., Koyner, J. L., Edelstein, C. L., Devarajan, P., Patel, U. D., Zappitelli, M., Krawczeski, C. D., Passik, C. S., & Coca, S. G. (2013) 'Performance of kidney injury molecule-1 and liver fatty acid-binding protein and combined biomarkers of AKI after cardiac surgery.', *Clin. J. Am. Soc. Nephrol.*, 8(7), pp. 1079–1088.

Parker, C. E. & Borchers, C. H. (2014) 'Mass spectrometry based biomarker discovery, verification, and validation – Quality assurance and control of protein biomarker assays', *Mol. Oncol.*, 8(4), pp. 840–858.

Paul, F. E., Hosp, F., & Selbach, M. (2011) 'Analyzing protein–protein interactions by quantitative mass spectrometry', *Methods*, 54(4), pp. 387–395.

Pavlos, R. & Phillips, E. J. (2012) 'Individualization of antiretroviral therapy', *Pharmgenomics. Pers. Med.*, 5, pp. 1–17.

Peco-Antic, A., Ivanisevic, I., Vulicevic, I., Kotur-Stevuljevic, J., Ilic, S., Ivanisevic, J., Miljkovic, M., & Kocev, N. (2013) 'Biomarkers of acute kidney injury in pediatric cardiac surgery.', *Clin. Biochem.*, 46(13–14), pp. 1244–1251.

Perazella, M. A. (2010) 'Tenofovir-induced kidney disease: an acquired renal tubular mitochondriopathy', *Kidney Int.* 2010/11/16, 78(11), pp. 1060–1063.

Pérez-Holanda, S., Blanco, I., Menéndez, M., & Rodrigo, L. (2014) 'Serum concentration of alpha-1 antitrypsin is significantly higher in colorectal cancer patients than in healthy controls.', *BMC Cancer*, 14, p. 355.

Pernikářová, V. & Bouchal, P. (2015) 'Targeted proteomics of solid cancers: from quantification of known biomarkers towards reading the digital proteome maps', *Expert Rev. Proteomics*, 12(6), pp. 651–667.

Perry, C. M. (2010) 'Maraviroc: A Review of its Use in the Management of CCR5-Tropic HIV-1 Infection', *Drugs*, 70(9), pp. 1189–1213.

Phillips, E. & Pavlos (2011) 'Individualization of antiretroviral therapy', *Pharmgenomics. Pers. Med.*, p. 1.

Piehowski, P. D., Petyuk, V. A., Orton, D. J., Xie, F., Moore, R. J., Ramirez-Restrepo, M., Engel, A., Lieberman, A. P., Albin, R. L., Camp, D. G., Smith, R. D., & Myers, A. J. (2013) 'Sources of Technical Variability in Quantitative LC–MS Proteomics: Human Brain Tissue Sample Analysis', *J. Proteome Res.*, 12(5), pp. 2128–2137.

Pieper, R., Gatlin, C. L., McGrath, A. M., Makusky, A. J., Mondal, M., Seonarain, M., Field, E., Schatz, C. R., Estock, M. A., Ahmed, N., Anderson, N. G., & Steiner, S. (2004) 'Characterization of the human urinary proteome: A method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots', *Proteomics*, 4(4), pp. 1159–1174.

Porath, J. & Flodin, P. (1987) 'Gel filtration: A method for desalting and group separation', *Chromatographia*, 23(5), pp. 365–369.

Potriquet, J., Laohaviroj, M., Bethony, J. M., & Mulvenna, J. (2017) 'A modified FASP protocol for high-throughput preparation of protein samples for mass spectrometry', *PLoS One*. Edited by J. M. Jacobs, 12(7), p. e0175967.

Prabhu, V. (2016) *ARV MARKET REPORT: The state of the antiretroviral drug market in low-and middle-income countries.*

Prajczer, S., Heidenreich, U., Pfaller, W., Kotanko, P., Lhotta, K., & Jennings, P. (2010) 'Evidence for a role of uromodulin in chronic kidney disease progression', *Nephrol. Dial. Transplant.*, 25(6), pp. 1896–1903.

Preiss, D. J., Godber, I. M., & Gunn, I. R. (2006) 'How to measure renal function in clinical practice: Eating cooked meat alters serum creatinine concentration and eGFR significantly', *BMJ Br. Med. J.*, 333(7577), p. 1072.

Prischl, F., Gremmel, F., Schwabe, M., Schindler, J., Balcke, P., Kopsa, H., Pinter, G., Schwarzmeier, J., & Zazgornik, J. (1989) 'Beta-2-microglobulin for differentiation between ciclosporin A nephrotoxicity and graft rejection in renal transplant recipients.', *Nephron*, 51(3), pp. 330–337.

Prowle, J. R., Ostland, V., Calzavacca, P., Licari, E., Ligabo, E. V., Echeverri, J. E., Bagshaw, S. M., Haase-Fielitz, A., Haase, M., Westerman, M., & Bellomo, R. (2012) 'Greater increase in urinary hepcidin predicts protection from acute kidney injury after cardiopulmonary bypass.', *Nephrol. Dial. Transplant*, 27(2), pp. 595–602.

Prunotto, M., Ghiggeri, G. M., Candiano, G., Lescuyer, P., Hochstrasser, D., & Moll, S. (2011) 'Urinary proteomics and drug discovery in chronic kidney disease: a new perspective', *J. Proteome Res.* 2010/08/20, 10(1), pp. 126–132.

Puoti, M., Nasta, P., Gatti, F., Matti, A., Prestini, K., Biasi, L., & Carosi, G. (2009) 'HIV-Related Liver Disease: ARV Drugs, Coinfection, and Other Risk Factors', *J. Int. Assoc. Physicians AIDS Care*, 8(1), pp. 30–42.

Purkerson, J. M. & Schwartz, G. J. (2007) 'The role of carbonic anhydrases in renal physiology', *Kidney Int.*, 71(2), pp. 103–115.

Qin, L.-X., Zhou, Q., Bogomolniy, F., Villafania, L., Olvera, N., Cavatore, M., Satagopan, J. M., Begg, C. B., & Levine, D. A. (2014) 'Blocking and randomization to improve molecular biomarker discovery.', *Clin. Cancer Res.*, 20(13), pp. 3371–8.

Quintana, L. F., Campistol, J. M., Alcolea, M. P., Banon-Maneus, E., Sol-Gonzalez, A., & Cutillas, P. R. (2009) 'Application of label-free quantitative peptidomics for the identification

of urinary biomarkers of kidney chronic allograft dysfunction.', *Mol. Cell. Proteomics*, 8(7), pp. 1658–1673.

Ramesh, G., Kwon, O., & Ahn, K. (2010) 'Netrin-1: a novel universal biomarker of human kidney injury.', *Transplant. Proc.*, 42(5), pp. 1519–1522.

Rampoldi, L., Scolari, F., Amoroso, A., Ghiggeri, G., & Devuyst, O. (2011) 'The rediscovery of uromodulin (Tamm-Horsfall protein): from tubulointerstitial nephropathy to chronic kidney disease', *Kidney Int.* 2011/06/10, 80(4), pp. 338–347.

Rasch, M. G., Engsig, F. N., Feldt-Rasmussen, B., Kirk, O., Kronborg, G., Pedersen, C., Gerstoft, J., & Obel, N. (2012) 'Renal function and incidence of chronic kidney disease in HIV patients: a Danish cohort study', *Scand. J. Infect. Dis.* 2012/06/12, 44(9), pp. 689–696.

Rauch, A., Nolan, D., Thurnheer, C., Fux, C. A., Cavassini, M., Chave, J. P., Opravil, M., Phillips, E., Mallal, S., & Furrer, H. (2008) 'Refining abacavir hypersensitivity diagnoses using a structured clinical assessment and genetic testing in the Swiss HIV Cohort Study', *Antivir. Ther.*

Ray, A. S., Cihlar, T., Robinson, K. L., Tong, L., Vela, J. E., Fuller, M. D., Wieman, L. M., Eisenberg, E. J., & Rhodes, G. R. (2006) 'Mechanism of Active Renal Tubular Efflux of Tenofovir', *Antimicrob. Agents Chemother.*, 50(10), pp. 3297–3304.

Ray, A. S., Fordyce, M. W., & Hitchcock, M. J. M. (2016) 'Tenofovir alafenamide: A novel prodrug of tenofovir for the treatment of Human Immunodeficiency Virus', *Antiviral Res.*, 125, pp. 63–70.

Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., Beck, M., Brusniak, M.-Y., Hengartner, M. O., & Aebersold, R. (2011) 'mProphet: automated data processing and statistical validation for large-scale SRM experiments', *Nat. Methods*, 8(5), pp. 430–435.

Rekers, N. V., Bajema, I. M., Mallat, M. J. K., Petersen, B., Anholts, J. D. H., Swings, G. M. J. S., van Miert, P. P. M. C., Kerkhoff, C., Roth, J., Popp, D., van Groningen, M. C., Baeten, D., Goemaere, N., Kraaij, M. D., Zandbergen, M., Heidt, S., van Kooten, C., de Fijter, J. W., Claas, F. H. J., & Eikmans, M. (2016) 'Beneficial Immune Effects of Myeloid-Related Proteins in Kidney Transplant Rejection', *Am. J. Transplant.*, 16(5), pp. 1441–1455.

Remer, T., Montenegro-Bethancourt, G., & Shi, L. (2014) 'Long-term urine biobanking: Storage stability of clinical chemical parameters under moderate freezing conditions without use of preservatives', *Clin. Biochem.*, 47(18), pp. 307–311.

Roberti, I. & Reisman, L. (2001) 'Serial evaluation of cell surface markers for immune activation after acute renal allograft rejection by urine flow cytometry--correlation with clinical outcome.', *Transplantation*, 71(9), pp. 1317–1320.

Roberts, J. D., Bebenek, K., & Kunkel, T. A. (1988) 'The accuracy of reverse transcriptase from HIV-1.', *Science*, 242(4882), pp. 1171–3.

Robertson, K., Liner, J., & Meeker, R. B. (2012) 'Antiretroviral neurotoxicity', *J. Neurovirol.*, 18(5), pp. 388–399.

Rocha, P. N., Macedo, M. N., Kobayashi, C. D., Moreno, L., Guimaraes, L. H. S., Machado, P. R. L., Badaro, R., Carvalho, E. M., & Glesby, M. J. (2015) 'Role of urine neutrophil gelatinase-associated lipocalin in the early diagnosis of amphotericin B-induced acute kidney injury.', *Antimicrob. Agents Chemother.*, 59(11), pp. 6913–6921.

Rogers, M. A., Clarke, P., Noble, J., Munro, N. P., Paul, A., Selby, P. J., & Banks, R. E. (2003) 'Proteomic profiling of urinary proteins in renal cancer by surface enhanced laser desorption ionization and neural-network analysis: identification of key issues affecting potential clinical utility.', *Cancer Res.*, 63(20), pp. 6971–83.

Rosenberger, G., Bludau, I., Schmitt, U., Heusel, M., Hunter, C. L., Liu, Y., Maccoss, M. J., Maclean, B. X., Nesvizhskii, A. I., Pedrioli, P. G. A., Reiter, L., Röst, H. L., Tate, S., Ting, Y. S., Collins, B. C., & Aebersold, R. (2017) 'Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses', *Nat. Methods*, 14(9), pp. 921–927.

Rosenberger, G., Koh, C. C., Guo, T., Rost, H. L., Kouvonen, P., & Collins, B. C. (2014) 'A repository of assays to quantify 10,000 human proteins by SWATH-MS', *Sci Data*, 1.

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., & Pappin, D. J. (2004) 'Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.', *Mol. Cell. Proteomics*, 3(12), pp. 1154–69.

Röst, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinović, S. M., Schubert, O. T., Wolski, W., Collins, B. C., Malmström, J., Malmström, L., & Aebersold, R. (2014) 'OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data', *Nat. Biotechnol.*, 32(3), pp. 219–223.

Röthlisberger, S. & Pedroza-Diaz, J. (2017) 'Urine protein biomarkers for detection of cardiovascular disease and their use for the clinic', *Expert Rev. Proteomics*, 14(12), pp. 1091–1103.

Rule, A. D. (2010) 'The CKD-EPI equation for estimating GFR from serum creatinine: real improvement or more of the same?', *Clin. J. Am. Soc. Nephrol.*, 5(6), pp. 951–3.

Saag, M., Balu, R., Phillips, E., Brachman, P., Martorell, C., Burman, W., Stancil, B., Mosteller, M., Brothers, C., Wannamaker, P., Hughes, A., Sutherland-Phillips, D., Mallal, S., Shaefer, M., Abacavir, S. of H. to, & Team, P. E. S. (2008) 'High Sensitivity of Human Leukocyte Antigen-B*5701 as a Marker for Immunologically Confirmed Abacavir Hypersensitivity in White and Black Patients', *Clin. Infect. Dis.*, 46(7), pp. 1111–1118.

Saito, M., Kimoto, M., Araki, T., Shimada, Y., Fujii, R., Oofusa, K., Hide, M., Usui, T., & Yoshizato, K. (2005) 'Proteome analysis of gelatin-bound urinary proteins from patients with bladder cancers', *Eur. Urol.*, 48(5), pp. 865–871.

Salome, T., Kasamba, I., Mayanja, B. N., Kazooba, P., Were, J., Kaleebu, P., & Munderi, P. (2016) 'The effect of Tenofovir on renal function among Ugandan adults on long-term antiretroviral therapy: a cross-sectional enrolment analysis', *AIDS Res. Ther.*, 13(1), p. 28.

Santa, C., Anjo, S. I., & Manadas, B. (2016) 'Protein precipitation of diluted samples in SDS-containing buffer with acetone leads to higher protein recovery and reproducibility in comparison with TCA/acetone approach', *Proteomics*, 16(13), pp. 1847–1851.

Santucci, L., Candiano, G., Petretto, A., Bruschi, M., Lavarello, C., Inglese, E., Righetti, P. G., & Ghiggeri, G. M. (2015) 'From hundreds to thousands: Widening the normal human Urinome (1)', *J. Proteomics*, 112(Supplement C), pp. 53–62.

Saraswathy, N. & Ramalingam, P. (2011) 'Protein Identification by Peptide Mass

Fingerprinting (PMF)', in *Concepts Tech. Genomics Proteomics*, pp. 185–192.

Satirapoj, B., Witoon, R., Ruangkanchanasetr, P., Wantanasiri, P., Charoenpitakchai, M., & Choovichian, P. (2014) 'Urine periostin as a biomarker of renal injury in chronic allograft nephropathy.', *Transplant. Proc.*, 46(1), pp. 135–140.

Sax, P. E., Wohl, D., Yin, M. T., Post, F., DeJesus, E., & Saag, M. (2015) 'Tenofovir alafenamide versus tenofovir disoproxil fumarate, coformulated with elvitegravir, cobicistat, and emtricitabine, for initial treatment of HIV-1 infection: two randomised, double-blind, phase 3, non-inferiority trials', *Lancet*, 385.

Schaub, S., Rush, D., Wilkins, J., Gibson, I. W., Weiler, T., Sangster, K., Nicolle, L., Karpinski, M., Jeffery, J., & Nickerson, P. (2004) 'Proteomic-Based Detection of Urine Proteins Associated with Acute Renal Allograft Rejection', *J. Am. Soc. Nephrol.*, 15(1), pp. 219–227.

Schaub, S., Wilkins, J. A., Antonovici, M., Krokhin, O., Weiler, T., Rush, D., & Nickerson, P. (2005) 'Proteomic-based identification of cleaved urinary beta2-microglobulin as a potential marker for acute tubular injury in renal allografts', *Am. J. Transplant.* 2005/03/12, 5(4 Pt 1), pp. 729–738.

Scherzer, R., Estrella, M., Li, Y., Choi, A. I., Deeks, S. G., Grunfeld, C., & Shlipak, M. G. (2012) 'Association of Tenofovir Exposure with Kidney Disease Risk in HIV Infection', *AIDS.*, 26(7), pp. 867–875.

Schiffer, E., Mischak, H., & Novak, J. (2006) 'High resolution proteome/peptidome analysis of body fluids by capillary electrophoresis coupled with MS', *Proteomics*, 6(20), pp. 5615–5627.

Schilling, B., Gibson, B. W., & Hunter, C. L. (2017) 'Generation of High-Quality SWATH® Acquisition Data for Label-free Quantitative Proteomics Studies Using TripleTOF® Mass Spectrometers.', *Methods Mol. Biol.*, 1550, pp. 223–233.

Schley, G., Köberle, C., Manuilova, E., Rutz, S., Forster, C., Weyand, M., Formentini, I., Kientsch-Engel, R., Eckardt, K.-U., & Willam, C. (2015) 'Comparison of Plasma and Urine Biomarker Performance in Acute Kidney Injury', *PLoS One*, 10(12), p. e0145042.

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(7), pp. 1289–1294.

SCIEX (2014) *Improved Data Quality Using Variable Q1 Window Widths in SWATH ® Acquisition Data Independent Acquisition on TripleTOF ® and X-Series QTOF Systems*.

SCIEX (2015) *Crowdsourcing: Optimized Rolling Collision Energy Curves for ID and SWATH Acquisition | SCIEX Community*.

SCIEX (2017) *Improved Data Quality Using Variable Q1 Window Widths in SWATH ® Acquisition*.

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. Retroviruses*, 33(1), pp. 33–40.

Segurado, O. G. & Sasso, E. H. (2014) 'Vectra DA for the objective measurement of disease

activity in patients with rheumatoid arthritis.', *Clin. Exp. Rheumatol.*, 32(5 Suppl 85), p. S-29-34.

Sharp, P. M. & Hahn, B. H. (2011) 'Origins of HIV and the AIDS pandemic', *Cold Spring Harb. Perspect. Med.*, 1(1), p. a006841.

Shilov, I. V, Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., Hunter, C. L., Nuwaysir, L. M., & Schaeffer, D. A. (2007) 'The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra.', *Mol. Cell. Proteomics*, 6(9), pp. 1638–55.

Shubber, Z., Mishra, S., Vesga, J. F., & Boily, M.-C. (2014) 'The HIV Modes of Transmission model: a systematic review of its findings and adherence to guidelines.', *J. Int. AIDS Soc.*, 17(1), p. 18928.

Silverman, G. A., Bird, P. I., Carrell, R. W., Church, F. C., Coughlin, P. B., Gettins, P. G., Irving, J. A., Lomas, D. A., Luke, C. J., Moyer, R. W., Pemberton, P. A., Remold-O'Donnell, E., Salvesen, G. S., Travis, J., & Whisstock, J. C. (2001) 'The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature.', *J. Biol. Chem.*, 276(36), pp. 33293–6.

Sivasubramanian, G., Frempong-Manso, E., & Macarthur, R. D. (2010) 'Abacavir/lamivudine combination in the treatment of HIV: a review', *Ther. Clin. Risk Manag.* 2010/03/18, 6, pp. 83–94.

Skates, S. J., Gillette, M. A., LaBaer, J., Carr, S. A., Anderson, L., Liebler, D. C., Ransohoff, D., Rifai, N., Kondratovich, M., Težak, Ž., Mansfield, E., Oberg, A. L., Wright, I., Barnes, G., Gail, M., Mesri, M., Kinsinger, C. R., Rodriguez, H., & Boja, E. S. (2013) 'Statistical Design for Biospecimen Cohort Size in Proteomics-based Biomarker Discovery and Verification Studies', *J. Proteome Res.*, 12(12), pp. 5383–5394.

Smith, L. M., Kelleher, N. L., Linial, M., Goodlett, D., Langridge-Smith, P., Ah Goo, Y., Safford, G., Bonilla*, L., Kruppa, G., Zubarev, R., Rontree, J., Chamot-Rooke, J., Garavelli, J., Heck, A., Loo, J., Penque, D., Hornshaw, M., Hendrickson, C., Pasa-Tolic, L., Borchers, C., Chan, D., Young*, N., Agar, J., Masselon, C., Gross*, M., McLafferty, F., Tsybin, Y., Ge, Y., Sanders*, I., Langridge, J., Whitelegge*, J., Marshall, A., & Tsybin, Y. (2013) 'Proteoform: a single term describing protein complexity', *Nat. Methods*, 10(3), pp. 186–187.

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(1), pp. 76–85.

Smits, A. H. & Vermeulen, M. (2016) 'Characterizing Protein–Protein Interactions Using Mass Spectrometry: Challenges and Opportunities', *Trends Biotechnol.*, pp. 825–834.

Soldatos, G. & Cooper, M. E. (2008) 'Diabetic nephropathy: Important pathophysiologic mechanisms', *Diabetes Res. Clin. Pract.*, 82, pp. S75–S79.

Spahr, C. S., Davis, M. T., McGinley, M. D., Robinson, J. H., Bures, E. J., Beierle, J., Mort, J., Courchesne, P. L., Chen, K., Wahl, R. C., Yu, W., Luethy, R., & Patterson, S. D. (2001) 'Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry I.Profiling an unfractionated tryptic digest', *Proteomics*, 1(1), pp. 93–107.

Srivastava, M., Eidelman, O., Torosyan, Y., Jozwik, C., Mannon, R. B., & Pollard, H. B.

(2011) ‘Elevated expression levels of ANXA11, integrins beta3 and alpha3, and TNF-alpha contribute to a candidate proteomic signature in urine for kidney allograft rejection.’, *Proteomics. Clin. Appl.*, 5(5–6), pp. 311–321.

Statistics South Africa (2017) ‘Mid-year population estimates’, *Stat. Release P0302*.

Storey, J. D. (2002) ‘A direct approach to false discovery rates’, *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, 64(3), pp. 479–498.

Stothard, P. (2000) ‘The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences.’, *Biotechniques*, 28(6), pp. 1102, 1104.

Stutzenberger, F. J. (1992) ‘Interference of the detergent Tween 80 in protein assays’, *Anal. Biochem.*, 207(2), pp. 249–254.

Szabo, Z. & Janaky, T. (2015) ‘Challenges and developments in protein identification using mass spectrometry’, *TrAC Trends Anal. Chem.*, 69, pp. 76–87.

Taneja, S., Sen, S., Gupta, V. K., Aggarwal, R., & Jameel, S. (2009) ‘Plasma and urine biomarkers in acute viral hepatitis E.’, *Proteome Sci.*, 7, p. 39.

Tang, W. H. H., Shilov, I. V. V., & Seymour, S. L. L. (2008) ‘Nonlinear Fitting Method for Determining Local False Discovery Rates from Decoy Database Searches’, *J. Proteome Res.*, 7(9), pp. 3661–3667.

Tape, C. J., Worboys, J. D., Sinclair, J., Gourlay, R., Vogt, J., McMahon, K. M., Trost, M., Lauffenburger, D. A., Lamont, D. J., & Jørgensen, C. (2014) ‘Reproducible automated phosphopeptide enrichment using magnetic TiO2 and Ti-IMAC.’, *Anal. Chem.*, 86(20), pp. 10296–302.

TEMPRANO ANRS 12136 Study Group, Danel, C., Moh, R., Gabillard, D., Badje, A., Le Carrou, J., Ouassa, T., Ouattara, E., Anzian, A., Ntakpé, J.-B., Minga, A., Kouame, G. M., Bouhoussou, F., Emieme, A., Kouamé, A., Inwoley, A., Toni, T., Ahiboh, H., Kabran, M., Rabe, C., Sidibé, B., Nzunetu, G., Konan, R., Gnokoro, J., Gouesse, P., Messou, E., Dohoun, L., Kamagate, S., Yao, A., Amon, S., Kouame, A.-B., Koua, A., Kouamé, E., Ndri, Y., Ba-Gomis, O., Daligou, M., Ackoundzé, S., Hawerlander, D., Ani, A., Dembélé, F., Koné, F., Guéhi, C., Kanga, C., Koule, S., Séri, J., Oyebi, M., Mbakop, N., Makaila, O., Babatunde, C., Babatounde, N., Bleoué, G., Tchoutedjem, M., Kouadio, A.-C., Sena, G., Yededji, S.-Y., Assi, R., Bakayoko, A., Mahassadi, A., Attia, A., Oussou, A., Mobio, M., Bamba, D., Koman, M., Horo, A., Deschamps, N., Chenal, H., Sassan-Morokro, M., Konate, S., Aka, K., Aoussi, E., Journot, V., Nchot, C., Karcher, S., Chaix, M.-L., Rouzioux, C., Sow, P.-S., Perronne, C., Girard, P.-M., Menan, H., Bissagnene, E., Kadio, A., Ettiegne-Traore, V., Moh-Semdé, C., Kouame, A., Massumbuko, J.-M., Chêne, G., Dosso, M., Domoua, S. K., N’Dri-Yoman, T., Salamon, R., Eholié, S. P., & Anglaret, X. (2015) ‘A Trial of Early Antiretrovirals and Isoniazid Preventive Therapy in Africa.’, *N. Engl. J. Med.*, 373(9), pp. 808–22.

Tencer, J., Thysell, H., Andersson, K., & Grubb, A. (1997) ‘Long-term Stability of Albumin, Protein HC, Immunoglobulin G, κ- and λ-chain-immunoreactivity, Orosomucoid and α_1-antitrypsin in Urine Stored at -20°C’, *Scand. J. Urol. Nephrol.*, 31(1), pp. 67–71.

The, M., Tasnim, A., & Käll, L. (2016) ‘How to talk about protein-level false discovery rates in shotgun proteomics.’, *Proteomics*, 16(18), pp. 2461–9.

Theodorescu, D., Fliser, D., Wittke, S., Mischak, H., Krebs, R., Walden, M., Ross, M., Eltze,

E., Bettendorf, O., Wulfing, C., & Semjonow, A. (2005) 'Pilot study of capillary electrophoresis coupled to mass spectrometry as a tool to define potential prostate cancer biomarkers in urine', *Electrophoresis*. 2005/06/28, 26(14), pp. 2797–2808.

Theodorescu, D., Schiffer, E., Bauer, H. W., Douwes, F., Eichhorn, F., Polley, R., Schmidt, T., Schöfer, W., Zürbig, P., Good, D. M., Coon, J. J., & Mischak, H. (2008) 'Discovery and validation of urinary biomarkers for prostate cancer.', *Proteomics. Clin. Appl.*, 2(4), pp. 556–570.

Thomas, S., Hao, L., Ricke, W. A., & Li, L. (2016) 'Biomarker discovery in mass spectrometry-based urinary proteomics', *Proteomics Clin Appl*, 10(4), pp. 358–370.

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., & Hamon, C. (2003) 'Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS', *Anal. Chem.*

Thongboonkerd, V. & Malasit, P. (2005) 'Renal and urinary proteomics: Current applications and challenges', *Proteomics*, 5(4), pp. 1033–1042.

Thongboonkerd, V., McLeish, K. R., Arthur, J. M., & Klein, J. B. (2002) 'Proteomic analysis of normal human urinary proteins isolated by acetone precipitation or ultracentrifugation', *Kidney Int.*, 62(4), pp. 1461–1469.

Thurman, J. M. (2015) 'Complement in kidney disease: core curriculum 2015.', *Am. J. Kidney Dis.*, 65(1), pp. 156–68.

Tipton, J. D., Tran, J. C., Catherman, A. D., Ahlf, D. R., Durbin, K. R., & Kelleher, N. L. (2011) 'Analysis of intact protein isoforms by mass spectrometry.', *J. Biol. Chem.*, 286(29), pp. 25451–8.

Titan, S. M., Vieira, J. M. J., Dominguez, W. V, Moreira, S. R. S., Pereira, A. B., Barros, R. T., & Zatz, R. (2012) 'Urinary MCP-1 and RBP: independent predictors of renal outcome in macroalbuminuric diabetic nephropathy.', *J. Diabetes Complications*, 26(6), pp. 546–553.

Toby, T. K., Fornelli, L., & Kelleher, N. L. (2016) 'Progress in Top-Down Proteomics and the Analysis of Proteoforms', *Annu. Rev. Anal. Chem.*, 9(1), pp. 499–519.

Toffaletti, J. G. (2010) 'Improving the Clinical Value of Estimating Glomerular Filtration Rate by Reporting All Values: A Contrarian Viewpoint', *Nephron Clin. Pract.*, 115(3), pp. c177–c181.

Torregrosa, I., Montoliu, C., Urios, A., Andres-Costa, M. J., Gimenez-Garzo, C., Juan, I., Puchades, M. J., Blasco, M. L., Carratala, A., Sanjuan, R., & Miguel, A. (2015) 'Urinary KIM-1, NGAL and L-FABP for the diagnosis of AKI in patients with acute coronary syndrome or heart failure undergoing coronary angiography.', *Heart Vessels*, 30(6), pp. 703–711.

Traynor, J., Mactier, R., Geddes, C. C., & Fox, J. G. (2006) 'How to measure renal function in clinical practice', *BMJ Br. Med. J.*, 333(7571), pp. 733–737.

Trouw, L. A., Blom, A. M., & Gasque, P. (2008) 'Role of complement and complement regulators in the removal of apoptotic cells', *Mol. Immunol.*, 45(5), pp. 1199–1207.

Tsiatsiani, L. & Heck, A. J. R. (2015) 'Proteomics beyond trypsin', *FEBS J.*, 282(14), pp. 2612–2626.

Tsou, C. C., Tsai, C. F., Teo, G. C., Chen, Y. J., & Nesvizhskii, A. I. (2016) 'Untargeted, spectral library-free analysis of data-independent acquisition proteomics data generated using Orbitrap mass spectrometers', *Proteomics*.

Tubaon, R. M., Haddad, P. R., & Quirino, J. P. (2017) 'Sample Clean-up Strategies for ESI Mass Spectrometry Applications in Bottom-up Proteomics: Trends from 2012 to 2016', *Proteomics*, 17(20), p. 1700011--n/a.

Tyanova, S., Albrechtsen, R., Kronqvist, P., Cox, J., Mann, M., & Geiger, T. (2016) 'Proteomic maps of breast cancer subtypes', *Nat. Commun.*, 7(1), p. 10259.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., & Cox, J. (2016) 'The Perseus computational platform for comprehensive analysis of (prote)omics data', *Nat. Methods*, 13(9), pp. 731–740.

Uettwiller-Geiger, D. L., Vijayendran, R., Kellum, J. A., & Fitzgerald, R. L. (2016) 'Analytical characteristics of a biomarker-based risk assessment test for acute kidney injury (AKI).', *Clin. Chim. Acta.*, 455, pp. 93–98.

UNAIDS (2018).

UNAIDS| South Africa (2018).

Vaidya, V. S., Ferguson, M. A., & Bonventre, J. V (2008) 'Biomarkers of acute kidney injury.', *Annu. Rev. Pharmacol. Toxicol.*, 48, pp. 463–93.

Välikangas, T., Suomi, T., & Elo, L. L. (2018) 'A systematic evaluation of normalization methods in quantitative label-free proteomics', *Brief. Bioinform.*, 19(1), pp. 1–11.

Vasconcelos, M. A., Bouzada, M. C. F., Silveira, K. D., Moura, L. R., Santos, F. F., Oliveira, J. M., Carvalho, F. F., Teixeira, M. M., E Silva, A. C. S., & Oliveira, E. A. (2011) 'Urinary levels of TGF beta-1 and of cytokines in patients with prenatally detected nephrouropathies.', *Pediatr. Nephrol.*, 26(5), pp. 739–747.

Vidova, V. & Spacil, Z. (2017) 'A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition', *Anal. Chim. Acta*, 964, pp. 7–23.

Vissers, J. P. C., Claessens, H. A., & Cramers, C. A. (1997) 'Microcolumn liquid chromatography: instrumentation, detection and applications', *J. Chromatogr. A*, 779(1–2), pp. 1–28.

Vivekanandan-Giri, A., Slocum, J. L., Buller, C. L., Basrur, V., Ju, W., Pop-Busui, R., Lubman, D. M., Kretzler, M., & Pennathur, S. (2011) 'Urine glycoprotein profile reveals novel markers for chronic kidney disease.', *Int. J. Proteomics*, 2011, p. 214715.

Vogenberg, F. R., Isaacson Barash, C., & Pursel, M. (2010) 'Personalized medicine: part 1: evolution and development into theranostics.', *P T*, 35(10), pp. 560–76.

Volberding, P. A. & Deeks, S. G. (2010) 'Antiretroviral therapy and management of HIV infection', *Lancet*, 376(9734), pp. 49–62.

Vowinckel, J., Zelezniak, A., Bruderer, R., Mülleder, M., Reiter, L., & Ralser, M. (2018) 'Cost-effective generation of precise label-free quantitative proteomes in high-throughput by microLC and data-independent acquisition', *Sci. Rep.*, 8(1), p. 4346.

Wai-Hoe, L., Wing-Seng, L., Ismail, Z., & Lay-Harn, G. (2009) 'Proteomics and detection of uromodulin in first-time renal calculi patients and recurrent renal calculi patients.', *Appl.*

Biochem. Biotechnol., 159(1), pp. 221–232.

Waikar, S. S., Betensky, R. A., & Bonventre, J. V. (2009) 'Creatinine as the gold standard for kidney injury biomarker studies?', *Nephrol. Dial. Transplant.*, 24(11), pp. 3263–3265.

Walsh, M. C., Brennan, L., Malthouse, J. P. G., Roche, H. M., & Gibney, M. J. (2006) 'Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans', *Am. J. Clin. Nutr.*, 84(3), pp. 531–539.

Wang, H., Lu, X., Yang, X., & Xu, N. (2016) 'The efficacy and safety of tenofovir alafenamide versus tenofovir disoproxil fumarate in antiretroviral regimens for HIV-1 therapy: Meta-analysis.', *Medicine (Baltimore).*, 95(41), p. e5146.

Wang, H., Zheng, C., Lu, Y., Jiang, Q., Yin, R., Zhu, P., Zhou, M., & Liu, Z. (2017) 'Urinary fibrinogen as a predictor of progression of CKD', *Clin. J. Am. Soc. Nephrol.*, 12(12), pp. 1922–1929.

Wang, Y., Yang, F., Gritsenko, M. A., Wang, Y., Clauss, T., Liu, T., Shen, Y., Monroe, M. E., Lopez-Ferrer, D., Reno, T., Moore, R. J., Klemke, R. L., Camp 2nd, D. G., & Smith, R. D. (2011) 'Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells', *Proteomics*. 2011/04/19, 11(10), pp. 2019–2026.

Washburn, K. K., Zappitelli, M., Arikan, A. A., Loftis, L., Yalavarthy, R., Parikh, C. R., Edelstein, C. L., & Goldstein, S. L. (2008) 'Urinary interleukin-18 is an acute kidney injury biomarker in critically ill children.', *Nephrol. Dial. Transplant*, 23(2), pp. 566–572.

Washburn, M. P. (2015) 'The H-Index of "An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database"', *J. Am. Soc. Mass Spectrom.*, pp. 1799–1803.

Wasinger, V. C., Zeng, M., & Yau, Y. (2013) 'Current Status and Advances in Quantitative Proteomic Mass Spectrometry', *Int. J. Proteomics*, 2013, pp. 1–12.

Watkins, M. E., Wring, S., Randolph, R., Park, S., Powell, K., Lutz, L., Nowakowski, M., Ramabhadran, R., & Domanico, P. L. (2017) 'Development of a Novel Formulation That Improves Preclinical Bioavailability of Tenofovir Disoproxil Fumarate', *J. Pharm. Sci.*, 106(3), pp. 906–919.

Webb-Robertson, B.-J. M., Matzke, M. M., Jacobs, J. M., Pounds, J. G., & Waters, K. M. (2011) 'A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors.', *Proteomics*, 11(24), pp. 4736–41.

Webb-Robertson, B.-J. M., Wiberg, H. K., Matzke, M. M., Brown, J. N., Wang, J., McDermott, J. E., Smith, R. D., Rodland, K. D., Metz, T. O., Pounds, J. G., & Waters, K. M. (2015) 'Review, Evaluation, and Discussion of the Challenges of Missing Value Imputation for Mass Spectrometry-Based Label-Free Global Proteomics', *J. Proteome Res.*, 14(5), pp. 1993–2001.

Wei, J., Zaika, E., & Zaika, A. (2012) 'p53 Family: Role of Protein Isoforms in Human Cancer.', *J. Nucleic Acids*, 2012, p. 687359.

Weissinger, E. M., Mullen, W., & Albalat, A. (2014) 'Urinary Proteomics Employing Capillary Electrophoresis Coupled to Mass Spectrometry in the Monitoring of Patients After Stem Cell Transplantation', in *Methods Mol. Biol.*, pp. 293–306.

Weissinger, E. M., Wittke, S., Kaiser, T., Haller, H., Bartel, S., Krebs, R., Golovko, I., Rupprecht, H. D., Haubitz, M., Hecker, H., Mischak, H., & Fliser, D. (2004) 'Proteomic patterns established with capillary electrophoresis and mass spectrometry for diagnostic purposes', *Kidney Int.*, 65(6), pp. 2426–2434.

Wessel, D. & Flügge, U. I. (1984) 'A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids', *Anal. Biochem.*, 138(1), pp. 141–143.

Westermeier, R., Naven, T., & Höpker, H.-R. (2008) 'Liquid Chromatography Techniques', in *Proteomics Pract.*, pp. 151–213.

Westhoff, J. H., Tonshoff, B., Waldherr, S., Poschl, J., Teufel, U., Westhoff, T. H., & Fichtner, A. (2015) 'Urinary Tissue Inhibitor of Metalloproteinase-2 (TIMP-2) * Insulin-Like Growth Factor-Binding Protein 7 (IGFBP7) Predicts Adverse Outcome in Pediatric Acute Kidney Injury.', *PLoS One*, 10(11), p. e0143628.

Westhuyzen, J., Endre, Z. H., Reece, G., Reith, D. M., Saltissi, D., & Morgan, T. J. (2003) 'Measurement of tubular enzymuria facilitates early detection of acute renal impairment in the intensive care unit.', *Nephrol. Dial. Transplant*, 18(3), pp. 543–551.

Wetz, A. J., Richardt, E. M., Wand, S., Kunze, N., Schotola, H., Quintel, M., Brauer, A., & Moerer, O. (2015) 'Quantification of urinary TIMP-2 and IGFBP-7: an adequate diagnostic test to predict acute kidney injury after cardiac surgery?', *Crit. Care*, 19, p. 3.

Wibell, L. (1978) 'The serum level and urinary excretion of beta2-microglobulin in health and renal disease.', *Pathol. Biol. (Paris).*, 26(6), pp. 295–301.

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.-H., Bantscheff, M., Gerstmair, A., Faerber, F., & Kuster, B. (2014) 'Mass-spectrometry-based draft of the human proteome', *Nature*, 509(7502), pp. 582–587.

Wilm, M. (2011) 'Principles of electrospray ionization.', *Mol. Cell. Proteomics*, 10(7), p. M111.009407.

Winston, A., Arenas-Pinto, A., Stohr, W., Fisher, M., Orkin, C. M., Aderogba, K., De Burgh-Thomas, A., O'Farrell, N., Lacey, C. J., Leen, C., Dunn, D., & Paton, N. I. (2013) 'Neurocognitive Function in HIV Infected Patients on Antiretroviral Therapy', *PLoS One*, 8(4).

Wisniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009) 'Universal sample preparation method for proteome analysis', *Nat Methods*, 6(5), pp. 359–62.

Wittke, S., Haubitz, M., Walden, M., Rohde, F., Schwarz, A., Mengel, M., Mischak, H., Haller, H., & Gwinner, W. (2005) 'Detection of Acute Tubulointerstitial Rejection by Proteomic Analysis of Urinary Samples in Renal Transplant Recipients', *Am. J. Transplant.*, 5(10), pp. 2479–2488.

Wu, D., Ni, J., Beretov, J., Cozzi, P., Willcox, M., Wasinger, V., Walsh, B., Graham, P., & Li, Y. (2017) 'Urinary biomarkers in prostate cancer detection and monitoring progression', *Crit. Rev. Oncol. Hematol.*, 118, pp. 15–26.

Wu, F., Dong, X.-J., Li, Y.-Y., Zhao, Y., Xu, Q.-L., & Su, L. (2015) 'Identification of phosphorylated MYL12B as a potential plasma biomarker for septic acute kidney injury using a quantitative proteomic approach.', *Int. J. Clin. Exp. Pathol.*, 8(11), pp. 14409–14416.

Wulff, J. E. & Mitchell, M. W. (2018) 'A Comparison of Various Normalization Methods for LC/MS Metabolomics Data', *Adv. Biosci. Biotechnol.*, 09(08), pp. 339–351.

Wysocki, V. H., Resing, K. A., Zhang, Q., & Cheng, G. (2005) 'Mass spectrometry of peptides and proteins', *Methods*, 35(3), pp. 211–222.

Xu, Z., Xu, B., & Xu, C. (2015) 'Urinary angiotensinogen as a potential biomarker of intrarenal renin-angiotensin system activity in Chinese chronic kidney disease patients.', *Ir. J. Med. Sci.*, 184(2), pp. 297–304.

Yamamoto, T., Noiri, E., Ono, Y., Doi, K., Negishi, K., Kamijo, A., Kimura, K., Fujita, T., Kinukawa, T., Taniguchi, H., Nakamura, K., Goto, M., Shinozaki, N., Ohshima, S., & Sugaya, T. (2007) 'Renal L-type fatty acid--binding protein in acute ischemic injury.', *J. Am. Soc. Nephrol.*, 18(11), pp. 2894–2902.

Yang, F., Shen, Y., Camp, D. G., & Smith, R. D. (2012) 'High pH reversed-phase chromatography with fraction concatenation as an alternative to strong-cation exchange chromatography for two-dimensional proteomic analysis', *Expert Rev. Proteomics*, 9(2), pp. 129–134.

Yang, P.-C. & Mahmood, T. (2012) 'Western blot: Technique, theory, and trouble shooting', *N. Am. J. Med. Sci.*, 4(9), p. 429.

Yanishi, M., Kinoshita, H., Mishima, T., Taniguchi, H., Yoshida, K., Komai, Y., Yasuda, K., Watanabe, M., Sugi, M., & Matsuda, T. (2017) 'Urinary l-type fatty acid-binding protein is a predictor of early renal function after partial nephrectomy.', *Ren. Fail.*, 39(1), pp. 7–12.

Yates, J. R. & Kelleher, N. L. (2013) 'Top Down Proteomics', *Anal. Chem.*, 85(13), pp. 6151–6151.

Yates, J. R., Ruse, C. I., & Nakorchevsky, A. (2009) 'Proteomics by Mass Spectrometry: Approaches, Advances, and Applications', *Annu. Rev. Biomed. Eng.*, 11(1), pp. 49–79.

Yeung, Y.-G., Nieves, E., Angeletti, R. H., & Stanley, E. R. (2008) 'Removal of detergents from protein digests for mass spectrometry analysis.', *Anal. Biochem.*, 382(2), pp. 135–7.

Yim, H., Kym, D., Seo, D. K., Yoon, J., Yang, H.-T., Lee, J., Cho, Y.-S., Hur, J., Chun, W., & Han, S.-W. (2015) 'Serum cystatin C and microalbuminuria in burn patients with acute kidney injury', *Eur. J. Clin. Invest.*, 45(6), pp. 594–600.

Yombi, J. C., Jones, R., Pozniak, A., Hougardy, J. M., & Post, F. A. (2015) 'Monitoring of kidney function in HIV-positive patients', *HIV Med.*, 16(8), pp. 457–467.

Youhanna, S., Weber, J., Beaujean, V., Glaudemans, B., Sobek, J., & Devuyst, O. (2014) 'Determination of uromodulin in human urine: Influence of storage and processing', *Nephrol. Dial. Transplant.*, 29(1), pp. 136–145.

Young, F. E. (1988) 'The role of the FDA in the effort against AIDS', *Public Health Rep.* 1988/05/01, 103(3), pp. 242–245.

Young, G.-H., Huang, T.-M., Wu, C.-H., Lai, C.-F., Hou, C.-C., Peng, K.-Y., Liang, C.-J., Lin, S.-L., Chang, S.-C., Tsai, P.-R., Wu, K.-D., Wu, V.-C., & Ko, W.-J. (2014) 'Hemojuvelin modulates iron stress during acute kidney injury: improved by furin inhibitor.', *Antioxid. Redox Signal.*, 20(8), pp. 1181–1194.

Yu, Y. & Pieper, R. (2015) 'Urine Sample Preparation in 96-well Filter Plates to Characterize

Inflammatory and Infectious Diseases of the Urinary Tract', in, pp. 77–87.

Yu, Y., Suh, M.-J., Sikorski, P., Kwon, K., Nelson, K. E., & Pieper, R. (2014) 'Urine Sample Preparation in 96-Well Filter Plates for Quantitative Clinical Proteomics', *Anal. Chem.*, 86(11), pp. 5470–5477.

Yuen, G. J., Weller, S., & Pakes, G. E. (2008) 'A review of the pharmacokinetics of abacavir', *Clin. Pharmacokinet.*

Zager, R. A., Johnson, A. C. M., & Becker, K. (2012) 'Plasma and urinary heme oxygenase-1 in AKI.', *J. Am. Soc. Nephrol.*, 23(6), pp. 1048–1057.

Zager, R. A., Johnson, A. C. M., & Frostad, K. B. (2014) 'Rapid Renal Alpha-1 Antitrypsin Gene Induction in Experimental and Clinical Acute Kidney Injury', *PLoS One*, 9(5), p. e98380.

Zager, R. A., Vijayan, A., & Johnson, A. C. M. (2012) 'Proximal tubule haptoglobin gene activation is an integral component of the acute kidney injury "stress response".', *Am. J. Physiol. Renal Physiol.*, 303(1), pp. F139-48.

Załęska-Kocięcka, M., Skrobisz, A., Wojtkowska, I., Grabowski, M., Dąbrowski, M., Kuśmierski, K., Piotrowska, K., Imiela, J., & Stępińska, J. (2017) 'Serum beta-2 microglobulin levels for predicting acute kidney injury complicating aortic valve replacement', *Interact. Cardiovasc. Thorac. Surg.*, 25(4), pp. 533–540.

Zanivan, S., Krueger, M., & Mann, M. (2011) 'In Vivo Quantitative Proteomics: The SILAC Mouse', in *Methods Mol. Biol.*, pp. 435–450.

Zappitelli, M., Washburn, K. K., Arikan, A. A., Loftis, L., Ma, Q., Devarajan, P., Parikh, C. R., & Goldstein, S. L. (2007) 'Urine neutrophil gelatinase-associated lipocalin is an early marker of acute kidney injury in critically ill children: a prospective cohort study.', *Crit. Care*, 11(4), p. R84.

Zeng, X., Hossain, D., Bostwick, D. G., Herrera, G. A., & Zhang, P. L. (2014) 'Urinary β2-Microglobulin Is a Good Indicator of Proximal Tubule Injury: A Correlative Study with Renal Biopsies', *J. Biomarkers*, 2014, pp. 1–7.

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C. G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., To, J. K., Laird, M. R., Lavidas, I., Liu, Z., Loveland, J. E., Maurel, T., McLaren, W., Moore, B., Mudge, J., Murphy, D. N., Newman, V., Nuhn, M., Ogeh, D., Ong, C. K., Parker, A., Patricio, M., Riat, H. S., Schuilenburg, H., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Zadissa, A., Frankish, A., Hunt, S. E., Kostadima, M., Langridge, N., Martin, F. J., Muffato, M., Perry, E., Ruffier, M., Staines, D. M., Trevanion, S. J., Aken, B. L., Cunningham, F., Yates, A., & Flicek, P. (2018) 'Ensembl 2018', *Nucleic Acids Res.*, 46(D1), pp. D754–D761.

Zhang, B., Chambers, M. C., & Tabb, D. L. (2007) 'Proteomic Parsimony through Bipartite Graph Analysis Improves Accuracy and Transparency', *J. Proteome Res.*, 6(9), pp. 3549–3557.

Zhang, B., Käll, L., & Zubarev, R. A. (2016) 'DeMix-Q: Quantification-Centered Data Processing Workflow.', *Mol. Cell. Proteomics*, 15(4), pp. 1467–78.

Zhang, F. & Chen, J. Y. (2016) 'A method for identifying discriminative isoform-specific peptides for clinical proteomics application', *BMC Genomics*, 17(S7), p. 522.

Zhang, H., Cao, J., Li, L., Liu, Y., Zhao, H., Li, N., Li, B., Zhang, A., Huang, H., Chen, S., Dong, M., Yu, L., Zhang, J., & Chen, L. (2015) 'Identification of urine protein biomarkers with the potential for early detection of lung cancer', *Sci. Rep.*, 5(1), p. 11805.

Zhang, L. & Elias, J. E. (2017) 'Relative Protein Quantification Using Tandem Mass Tag Mass Spectrometry', in *Methods Mol. Biol.*, pp. 185–198.

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., Yates, J. R., & III (2013) 'Protein analysis by shotgun/bottom-up proteomics.', *Chem. Rev.*, 113(4), pp. 2343–94.

Zhang, Z., Lu, B., Ni, H., Sheng, X., & Jin, N. (2013) 'Microalbuminuria can predict the development of acute kidney injury in critically ill septic patients', *J. Nephrol.*, 26(4), pp. 724–730.

Zhao, S.-L., Wei, S.-Y., Wang, Y.-X., Diao, T.-T., Li, J.-S., He, Y.-X., Yu, J., Jiang, X.-Y., Cao, Y., Mao, X.-Y., Wei, Q.-J., Wang, Y., & Li, B. (2016) 'Wnt4 is a novel biomarker for the early detection of kidney tubular injury after ischemia/reperfusion injury.', *Sci. Rep.*, 6, p. 32610.

Zi, J., Zhang, S., Zhou, R., Zhou, B., Xu, S., Hou, G., Tan, F., Wen, B., Wang, Q., Lin, L., & Liu, S. (2014) 'Expansion of the Ion Library for Mining SWATH-MS Data through Fractionation Proteomics', *Anal. Chem.*, 86(15), pp. 7242–7246.

Zimmermann, A. E., Pizzoferrato, T., Bedford, J., Morris, A., Hoffman, R., & Braden, G. (2006) 'Tenofovir-associated acute and chronic kidney disease: a case of multiple drug interactions', *Clin. Infect. Dis.*, 42.

Zoidakis, J., Makridakis, M., Zerefos, P. G., Bitsika, V., Esteban, S., Frantzi, M., Stravodimos, K., Anagnou, N. P., Roubelakis, M. G., Sanchez-Carbayo, M., & Vlahou, A. (2012) 'Profilin 1 is a potential biomarker for bladder cancer aggressiveness.', *Mol. Cell. Proteomics*, 11(4), p. M111.009449.

Zwiers, A. J. M., de Wildt, S. N., van Rosmalen, J., de Rijke, Y. B., Buijs, E. A. B., Tibboel, D., & Cransberg, K. (2015) 'Urinary neutrophil gelatinase-associated lipocalin identifies critically ill young children with acute kidney injury following intensive care admission: a prospective cohort study.', *Crit. Care*, 19, p. 181.

Chapter 7

Appendix

Figure A1: Case report form used to collect demographic and routine clinical data from all participants

INFORMED CONSENT / DEMOGRAPHICS

Date Informed Consent Signed: ☐☐ ☐☐☐ 2 0 ☐☐ (DAY / MONTH / YEAR)

Study duration: 1st August 2012 – 31st July 2015

Patient demographics

Date of Birth: ☐☐☐ 1 9 ☐☐ (MONTH / YEAR)

Race: *(check one)* **Ethnicity (if known and consented to by patient)**

☐ Black _____

☐ White _____

☐ Coloured _____

☐ Indian _____

☐ Asian _____

☐ Other (state) _____

Gender

☐ Male

☐ Female

INCLUSION CRITERIA:*Please check Yes or No.*
If any of the responses below are checked "No ", the patient is NOT eligible for the study.

	Yes	No
1. Of adult age (>18 years)	❑	❑
2. Positive diagnosis for HIV	❑	❑
3. Meets eligibility criteria for ART / already on ART	❑	❑
4. Willing to have blood taken for research	❑	❑
5. Of otherwise sound health	❑	❑

EXCLUSION CRITERIA:*Please check Yes or No.*
If any of the responses below are checked "Yes ", the patient is NOT eligible for the study.

1. Medication for chronic condition/s ; suspected concomitant illness e.g. TB	❑	❑
2. Unwilling to have blood taken for research	❑	❑
3. Poor venous access	❑	❑
4. Unwilling to comply with the study protocols	❑	❑
5. Poor treatment compliance	❑	❑

PHYSICAL ATTRIBUTES / BIOLOGICAL TISSUE SAMPLE OBTAINED

Physical attributes
Please enter all information using leading zeros as applicable.

Weight (Kgs)/ BMI ☐☐☐.☐

Height (metric) ☐☐☐.☐

CD4 cell count (viral load)

Original count & date (if available)……

Count month 6 on TDF……

Count month 12 on TDF……

BIOLOGICAL TISSUE SAMPLE - Gene analysis

Venous blood sample for gene analysis ☐

Volume (MLs)

Other (describe) _____

_____

_____

Visit Date: ☐☐ ☐☐☐ 2 0 ☐☐
DAY MONTH YEAR

BIOLOGICAL TISSUE SAMPLE - Renal Function

Venous blood sample for kidney function analysis ☐

Visit Date: ☐☐ ☐☐☐ 2 0 ☐☐
DAY MONTH YEAR

Glucose level……	Parathyroid hormone……
Uric acid level……	Alkaline phosphatase……
Potassium level……	Proteinuria……
Calcium level……	Tubular phosphate reabsorption……
Phosphorus level……	Creatinine clearance……
Glycosuria……	Albuminuria/creatinine……
Total protein……	

Creatinine level before TDF ……

Creatinine level 1 month on TDF……

Creatinine level 6 months on TDF……

Consenting patients who present with ART-induced adverse events during study, or are a control (good responders) volunteer

Venous blood sample taken ☐

Volume (MLs)

State adverse event(s) and intervention taken _____

_____

Visit Date: ☐☐ ☐☐☐ 2 0 ☐☐
DAY MONTH YEAR

MEDICAL HISTORY / MEDICATION

Existing Disease Condition or Allergy
(please enter only ONE abnormal/per line)

Year of Diagnosis

LIST DISEASE UNDER STUDY, IF APPLICABLE

_____ ☐☐☐☐

_____ ☐☐☐☐

_____ ☐☐☐☐

_____ ☐☐☐☐

CONCOMITANT MEDICATIONS

Please list all medications taken during the course of the study. ARVs taken prior to study should be listed as well*. Generic and trade names are both acceptable. If medication is ongoing, _X' the Ongoing box. Complete date fields as completely as possible.

Medication & medical condition…….. ARV code (if applicable)

Start Date ☐☐ (DAY) ☐☐☐ (MONTH) 2 0 ☐☐ (YEAR) **Stop Date** ☐☐ (DAY) ☐☐☐ (MONTH) 2 0 ☐☐ (YEAR) ☐ **Ongoing** ☐

Medication & medical condition……..

Start Date ☐☐ (DAY) ☐☐☐ (MONTH) 2 0 ☐☐ (YEAR) **Stop Date** ☐☐ (DAY) ☐☐☐ (MONTH) 2 0 ☐☐ (YEAR) ☐ **Ongoing** ☐

Medication & medical condition……..

Start Date ☐☐ (DAY) ☐☐☐ (MONTH) 2 0 ☐☐ (YEAR) **Stop Date** ☐☐ (DAY) ☐☐☐ (MONTH) 2 0 ☐☐ (YEAR) ☐ **Ongoing** ☐

Medication & medical condition……..

Start Date ☐☐ (DAY) ☐☐☐ (MONTH) 2 0 ☐☐ (YEAR) **Stop Date** ☐☐ (DAY) ☐☐☐ (MONTH) 2 0 ☐☐ (YEAR) ☐ **Ongoing** ☐

*Use ARV termination code 1) 2) 3) 4) 5)

ADVERSE EVENT REPORT

Onset date ☐☐ (DAY) ☐☐☐ (MONTH) 2 0 ☐☐ (YEAR) **End date** ☐☐ (DAY) ☐☐☐ (MONTH) 2 0 ☐☐ (YEAR) **Ongoing** ☐

Severity and event type/s _____

Action taken _____

Onset date ☐☐ (DAY) ☐☐☐ (MONTH) 2 0 ☐☐ (YEAR) **End date** ☐☐ (DAY) ☐☐☐ (MONTH) 2 0 ☐☐ (YEAR) **Ongoing** ☐

Severity and event type/s _____

Action taken _____

COMPLIANCE TO TREATMENT

Compliance counselling undertaken prior to therapeutic intervention

☐ Yes ☐ No

On a scale of 0-100 (0 being non compliant, 100 being fully compliant) at what frequency do you think the patient is taking their medication as prescribed during the preceding months

Months 0- 3	**Months 3-6**	**Months 6-9**	**Months 9-12**
_____	_____	_____	_____

If non-compliant state reason(s) given, duration, and intervention taken

END OF STUDY / EARLY TERMINATION

Date of Last study treatment intake: ☐☐ ☐☐☐ ☐☐ (DAY / MONTH / YEAR)

Did the patient complete the study completely as described in the protocol? ☐ Yes ☐ No

Date of Withdrawal: ☐☐ ☐☐☐ ☐☐ (DAY / MONTH / YEAR)

If the answer to the above question was _No', indicate primary reason for early withdrawal from study:
(Check one box only if the patient did NOT fully complete the study)

Protocol Violation: *(spec)*

Adverse Event/s
Poor Compliance
Lack of Efficacy
Concomitant illness
Other e.g. withdrew consent

Table A2: Summary of putative kidney-damage protein biomarkers identified in urine. These proteins were identified as markers for various forms of non-malignant urogenital tract diseases that have been reported in literature as at 30 June 2018.

Protein Description	Disease(s)	Potential Biomarker Usage	Reference(s)
72 kDa type IV collagenase	Chronic Kidney Disease	Early Diagnosis	(Liu *et al.*, 2006)
Alkaline phosphatase	Acute kidney failure	Early Diagnosis	(Endre *et al.*, 2011)
Alpha-1-acid glycoprotein 1	Acute kidney failure	Response To Treatment	(Devarajan *et al.*, 2010)
Alpha-1-antichymotrypsin	Kidney transplant rejection	Diagnosis	(O'Riordan *et al.*, 2007)
Alpha-1-antitrypsin	Calculus of kidney	Diagnosis	(Wai-Hoe *et al.*, 2009)
Alpha-N-acetylglucosaminidase	Acute kidney failure	Early Diagnosis	(Daggulli *et al.*, 2016)
Angiotensinogen	Chronic Kidney Disease	Indicator Of Physiological Process	(Xu, Xu and Xu, 2015)
Annexin A11	Kidney transplant rejection	Diagnosis	(Srivastava *et al.*, 2011)
APOA1 (apolipoprotein-A1)	Complications of kidney transplant	Diagnosis	(Cantley *et al.*, 2016)
Basement membrane-specific heparan sulfate proteoglycan core protein	Complications of kidney transplant	Diagnosis	(Cantley *et al.*, 2016)
Beta-2-microglobulin	Acute kidney failure Autosomal Dominant Polycystic Kidney Disease Kidney transplant rejection	Diagnosis, Early Diagnosis Indicator Of Severity Prediction Of Response To Treatment	(Prischl *et al.*, 1989; Ho *et al.*, 2009; Meijer *et al.*, 2010; Du *et al.*, 2011; Heise *et al.*, 2011)
Beta-defensin 1	Kidney transplant rejection	Diagnosis	(O'Riordan *et al.*, 2007)
Bifunctional enzyme CysN/CysC	Acute kidney failure	Diagnosis	(Malyszko *et al.*, 2015)
Cadherin-5	Chronic Kidney Disease	Indicator Of Physiological Process	(Liu *et al.*, 2006)
C-C motif chemokine 2	Autosomal Dominant Polycystic Kidney Disease Chronic Kidney Disease	Diagnosis Early Diagnosis	(Liu *et al.*, 2006; Meijer *et al.*, 2010; Fu *et al.*, 2012)
C-C motif chemokine 5	Chronic Kidney Disease	Early Diagnosis	(Liu *et al.*, 2006)
CCL2	Complications of kidney transplant	Diagnosis	(Ho *et al.*, 2013)
Chitinase-3-like protein 1	Acute kidney failure	Early Diagnosis	(Loor *et al.*, 2016)

Connective tissue growth factor	Chronic Kidney Disease Complications of kidney transplant	Staging Diagnosis	(O'Seaghdha *et al.*, 2011; Blydt-Hansen *et al.*, 2015)
C-X-C motif chemokine 10	Acute kidney failure Complications of kidney transplant	Early Diagnosis, Diagnosis Prediction Of Response To Treatment	(Hu *et al.*, 2004; Ho *et al.*, 2009; Blydt-Hansen *et al.*, 2015)
C-X-C motif chemokine 9	Complications of kidney transplant Kidney transplant rejection	Diagnosis	(Hu *et al.*, 2004)
Cystatin-C	Acute kidney failure	Diagnosis Early Diagnosis Indicator Of Severity Prediction Of Response To Treatment	(Ma *et al.*, 2010; Endre *et al.*, 2011; Heise *et al.*, 2011; Li *et al.*, 2012)
E-selectin	Chronic Kidney Disease	Early Diagnosis	(Liu *et al.*, 2006)
Fatty acid-binding protein, heart	Autosomal Dominant Polycystic Kidney Disease	Diagnosis	(Meijer *et al.*, 2010)
Fatty acid-binding protein, liver	Acute kidney failure Complications of kidney transplant Injury to Kidney Type 2 diabetes mellitus with diabetic chronic kidney disease	Diagnosis Early Diagnosis Indicator Of Severity Prediction Of Response To Treatment	(Yamamoto *et al.*, 2007; von Eynatten *et al.*, 2010; Matsui *et al.*, 2011, 2012; Katagiri *et al.*, 2012; Nickolas *et al.*, 2012; Peco-Antic *et al.*, 2013; Asada *et al.*, 2016; Yanishi *et al.*, 2017)
FLJ00123 protein	Acute kidney failure	Diagnosis	(Taneja *et al.*, 2009)
Gamma-glutamyltranspeptidase 1	Acute kidney failure	Diagnosis	(Endre *et al.*, 2011)
Glutathione S-transferase	Acute kidney failure	Early Diagnosis	(Westhuyzen *et al.*, 2003)
Glutathione S-transferase P	Acute kidney failure	Early Diagnosis	(Westhuyzen *et al.*, 2003)
Haptoglobin	Acute kidney failure	Diagnosis	(Zager, Vijayan and Johnson, 2012)
Haeme oxygenase 1	Acute kidney failure	Diagnosis	(Zager, Johnson and Becker, 2012)
Haemoglobin subunit beta	Complications of kidney transplant	Diagnosis	(Cantley *et al.*, 2016)
Hemojuvelin	Acute kidney failure	Early Diagnosis	(Young *et al.*, 2014)

Hepatitis A virus cellular receptor 1	Acute kidney failure Autosomal Dominant Polycystic Kidney Disease	Diagnosis Indicator Of Severity Prediction Of Response To Treatment	(Meijer *et al.*, 2010; Nickolas *et al.*, 2012; de Geus *et al.*, 2013; Luo *et al.*, 2013; Parikh *et al.*, 2013)
Hepcidin	Acute kidney failure	Early Diagnosis Prediction Of Response To Treatment	(Ho *et al.*, 2009; Prowle *et al.*, 2012)
Ig alpha-1 chain C region Ig gamma-1 chain C region Ig gamma-2 chain C region Ig gamma-3 chain C region Ig gamma-4 chain C region Ig kappa chain C region IGHA1 protein Immunoglobulin heavy chain Immunoglobulin heavy variable 1-46 Immunoglobulin heavy variable 4-59 Immunoglobulin kappa variable 1-5	Calculus of kidney	Diagnosis	(Wai-Hoe *et al.*, 2009)
Insulin-like growth factor-binding protein 7	Acute kidney failure	Early Diagnosis	(Uettwiller-Geiger *et al.*, 2016)
Insulin-Like Growth Factor-Binding Protein 7 (IGFBP7)	Acute kidney failure	Diagnosis	(Westhoff *et al.*, 2015)
Integrin alpha-3 and beta-3	Kidney transplant rejection	Diagnosis	(Srivastava *et al.*, 2011)
Inter-alpha-trypsin inhibitor heavy chain H1	Calculus of kidney	Diagnosis	(Kovacevic *et al.*, 2015)
Intercellular adhesion molecule 1	Chronic Kidney Disease Kidney transplant rejection	Early Diagnosis Prediction Of Response To Treatment	(Roberti and Reisman, 2001; Liu *et al.*, 2006)
Interleukin 18	Acute kidney failure	Diagnosis Early Diagnosis Indicator Of Severity	(Parikh *et al.*, 2005; Washburn *et al.*, 2008; Che *et al.*, 2010; Endre *et al.*, 2011; Koyner *et al.*, 2015)

		Prediction Of Response To Treatment	
Intestinal-type alkaline phosphatase	Acute kidney failure	Diagnosis	(Zwiers *et al.*, 2015)
Kidney injury molecule-1	Acute kidney failure Kidney disease Injury to Kidney	Diagnosis Early Diagnosis Indicator Of Physiological Process	(Ciszek *et al.*, 2006; Liu *et al.*, 2006; Dennen *et al.*, 2010; Endre *et al.*, 2011; Lahoud *et al.*, 2015; Torregrosa *et al.*, 2015; Zwiers *et al.*, 2015)
L-FABP	Acute kidney failure Kidney disease	Diagnosis	(Fufaa *et al.*, 2015; Torregrosa *et al.*, 2015)
Liver-type fatty acid binding protein	Acute kidney failure	Diagnosis	(Koyner *et al.*, 2015)
Matrilysin	Chronic Kidney Disease	Diagnosis	(Musial, Bargenda and Zwolinska, 2015)
Matrix metalloproteinase-9	Calculus of kidney Chronic Kidney Disease	Diagnosis Early Diagnosis	(Liu *et al.*, 2006; Kovacevic *et al.*, 2015)
Metalloproteinase inhibitor 2	Acute kidney failure Acute kidney failure	Diagnosis Early Diagnosis	(Wetz *et al.*, 2015; Uettwiller-Geiger *et al.*, 2016)
Monocyte differentiation antigen CD14	Kidney transplant rejection	Prediction Of Response To Treatment	(Roberti and Reisman, 2001)
Myosin regulatory light chain 12B	Acute kidney failure	Diagnosis	(Wu *et al.*, 2015)
NAG	Injury to Kidney Kidney disease	Early Diagnosis Diagnosis	(Fufaa *et al.*, 2015; Goknar *et al.*, 2015)
Netrin-1	Acute kidney failure	Diagnosis	(Ramesh, Kwon and Ahn, 2010)
NGAL	Acute kidney failure Injury to Kidney Kidney disease	Diagnosis Early Diagnosis Indicator Of Physiological Process	(Zappitelli *et al.*, 2007; Ho *et al.*, 2009; Fufaa *et al.*, 2015; Rocha *et al.*, 2015; Torregrosa *et al.*, 2015; Zwiers *et al.*, 2015)
Osteopontin	Calculus of kidney	Diagnosis	(Bautista *et al.*, 1996)
Outer dense fiber protein 2	Acute kidney failure	Diagnosis	(Taneja *et al.*, 2009)
Periostin	Complications of kidney transplant	Diagnosis	(Satirapoj *et al.*, 2014)
Pi- and alpha-GST	Acute kidney failure	Diagnosis	(de Geus *et al.*, 2013)

Pigment epithelium-derived factor	Complications of kidney transplant	Diagnosis	(Moskowitz-Kassai *et al.*, 2012)
Platelet-derived growth factor subunit B	Chronic Kidney Disease	Early Diagnosis	(Liu *et al.*, 2006)
Prostaglandin D2 synthase 21kDa	Acute kidney failure	Diagnosis	(Taneja *et al.*, 2009)
Protein AMBP	Acute kidney failure Complications of kidney transplant	Early Diagnosis Indicator Of Physiological Process Prediction Of Response To Treatment Diagnosis	(Ho *et al.*, 2009; Taneja *et al.*, 2009; Devarajan *et al.*, 2010; Cantley *et al.*, 2016)
Protein O-GlcNAcase	Acute kidney failure	Diagnosis Prediction Of Response To Treatment	(Yamamoto *et al.*, 2007; Che *et al.*, 2010; Doi *et al.*, 2012)
Protein Wnt-4	Injury to Kidney	Early Diagnosis	(Zhao *et al.*, 2016)
Retinol-binding protein 4	Acute kidney failure Chronic Kidney Disease Complications of kidney transplant	Prediction Of Response To Treatment Diagnosis	(Che *et al.*, 2010; Titan *et al.*, 2012; Cantley *et al.*, 2016)
Semaphorin-3A	Acute kidney failure	Early Diagnosis Prognosis	(Jayakumar *et al.*, 2013; Doi *et al.*, 2014)
Serotransferrin	Calculus of kidney	Diagnosis	(Wai-Hoe *et al.*, 2009)
Serum albumin	Acute kidney failure Calculus of kidney	Diagnosis Prediction Of Response To Treatment	(Taneja *et al.*, 2009; Wai-Hoe *et al.*, 2009; Devarajan *et al.*, 2010)
Sodium/hydrogen exchanger 3	Acute kidney failure	Diagnosis	(du Cheyron *et al.*, 2003)
Tenascin	Chronic Kidney Disease	Diagnosis	(Horstrup *et al.*, 2002)
Tissue inhibitor of metalloproteinase 1	Chronic Kidney Disease	Early Diagnosis	(Liu *et al.*, 2006)
Tissue Inhibitor of Metalloproteinase-2 (TIMP-2)	Acute kidney failure	Diagnosis	(Westhoff *et al.*, 2015)
Titin	Acute kidney failure	Diagnosis	(Taneja *et al.*, 2009)
Transforming growth factor beta-1	Chronic Kidney Disease Kidney disease	Early Diagnosis Diagnosis	(Liu *et al.*, 2006; Vasconcelos *et al.*, 2011)
Trefoil factor 3(TFF3)	Kidney failure	Diagnosis	(Cho *et al.*, 2015)
Tumour necrosis factor	Chronic Kidney Disease Kidney transplant rejection	Early Diagnosis Diagnosis	(Liu *et al.*, 2006; Srivastava *et al.*, 2011)

Tumour necrosis factor receptor superfamily member 6	Chronic Kidney Disease	Early Diagnosis	(Liu *et al.*, 2006)
U3 small nucleolar RNA-associated protein 14 homolog C	Calculus of kidney	Diagnosis	(Wai-Hoe *et al.*, 2009)
Uncharacterized protein	Calculus of kidney	Diagnosis	(Wai-Hoe *et al.*, 2009)
Unconventional myosin-XVIIIa	Acute kidney failure	Diagnosis	(Taneja *et al.*, 2009)
Urinary angiotensinogen (AGT)	Chronic Kidney Disease	Diagnosis	(Xu, Xu and Xu, 2015)
Urinary Kidney injury molecule-1	Injury to Kidney	Prediction Of Response To Treatment	(Lahoud *et al.*, 2015)
Urinary neutrophil gelatinase associated lipocalin (uNGAL),	Acute kidney failure Injury to Kidney	Diagnosis Indicator Of Severity Prediction Of Response To Treatment	(Lahoud *et al.*, 2015; Martensson *et al.*, 2015; Rocha *et al.*, 2015)
Urinary neutrophil gelatinase-associated lipocalin	Acute kidney failure	Diagnosis	(Zwiers *et al.*, 2015)
Uromodulin	Calculus of kidney Complications of kidney transplant	Diagnosis	(Quintana *et al.*, 2009; Wai-Hoe *et al.*, 2009)
Vascular cell adhesion protein 1	Chronic Kidney Disease	Early Diagnosis	(Liu *et al.*, 2006)
Vascular endothelial growth factor A	Chronic Kidney Disease	Early Diagnosis	(Liu *et al.*, 2006)
Vitamin-D-binding protein	Injury to Kidney	Diagnosis	(Chaykovska *et al.*, 2016)
Zinc-alpha-2-glycoprotein	Acute kidney failure	Indicator Of Physiological Process	(Taneja *et al.*, 2009)

Table A3: Differentially abundant proteins derived from all 37 experiments before filtering.

Protein Description	UniProt Identifier
14 kDa phosphohistidine phosphatase	Q9NRX4
14-3-3 protein beta/alpha	P31946
14-3-3 protein epsilon	P62258
14-3-3 protein sigma	P31947
14-3-3 protein theta	P27348
14-3-3 protein zeta/delta	P63104
3-ketoacyl-CoA thiolase, mitochondrial	P42765
3-mercaptopyruvate sulfurtransferase	P25325
4-trimethylaminobutyraldehyde dehydrogenase	P49189
60 kDa heat shock protein, mitochondrial	P10809
6-phosphogluconate dehydrogenase, decarboxylating	P52209
6-phosphogluconolactonase	O95336
78 kDa glucose-regulated protein	P11021
Acetyl-CoA acetyltransferase, mitochondrial	P24752
Acid ceramidase	Q13510
Aconitate hydratase, mitochondrial	Q99798
Actin, cytoplasmic 1	P60709
Actin-related protein 2/3 complex subunit 2	O15144
Actin-related protein 3	P61158
Adenosylhomocysteinase	P23526
Adenylyl cyclase-associated protein 1	Q01518
ADP/ATP translocase 2	P05141
Alcohol dehydrogenase [NADP(+)]	P14550
Aldehyde dehydrogenase, mitochondrial	P05091
Aldo-keto reductase family 1 member B10	O60218
Aldo-keto reductase family 1 member B15	C9JRZ8
Aldose reductase	P15121
Alpha-1-acid glycoprotein 1	P02763
Alpha-1-acid glycoprotein 2	P19652
Alpha-1-antichymotrypsin	P01011
Alpha-1-antitrypsin	P01009
Alpha-1B-glycoprotein	P04217
Alpha-2-antiplasmin	P08697
Alpha-2-HS-glycoprotein	P02765
Alpha-2-macroglobulin	P01023
Alpha-2-macroglobulin-like protein 1	A8K2U0
Alpha-actinin-1	P12814

Alpha-actinin-4	O43707
Alpha-amylase 1	P04745
Alpha-amylase 2B	P19961
Alpha-enolase	P06733
Alpha-N-acetylglucosaminidase	P54802
Amiloride-sensitive amine oxidase [copper-containing]	P19801
Aminoacylase-1	Q03154
Aminopeptidase N	P15144
Angiopoietin-related protein 2	Q9UKU9
Angiotensinogen	P01019
Annexin A1	P04083
Annexin A11	P50995
Annexin A2	P07355
Annexin A3	P12429
Annexin A5	P08758
Annexin A6	P08133
Anthrax toxin receptor 1	Q9H6X2
Antithrombin-III	P01008
Apolipoprotein A-I	P02647
Apolipoprotein A-II	P02652
Apolipoprotein A-IV	P06727
Apolipoprotein C-III	P02656
Apolipoprotein D	P05090
Apolipoprotein E	P02649
Apoptosis-associated speck-like protein containing a CARD	Q9ULZ3
Argininosuccinate synthase	P00966
Aromatic-L-amino-acid decarboxylase	P20711
ATP synthase subunit alpha, mitochondrial	P25705
ATP synthase subunit beta, mitochondrial	P06576
ATP-dependent RNA helicase A	Q08211
Basal cell adhesion molecule	P50895
Basement membrane-specific heparan sulfate proteoglycan core protein	P98160
Beta-1,4-galactosyltransferase 1	P15291
Beta-2-microglobulin	P61769
Beta-defensin 1	P60022
Beta-galactosidase	P16278
Betaine--homocysteine S-methyltransferase 1	Q93088
BH3-interacting domain death agonist	P55957
BTB/POZ domain-containing protein KCTD12	Q96CX2
C4b-binding protein alpha chain	P04003

Cadherin-1	P12830
Cadherin-11	P55287
Cadherin-3	P22223
Cadherin-6	P55285
Calmodulin-1	P0DP23
Calnexin	P27824
Calreticulin	P27797
Carbonic anhydrase 1	P00915
Carbonic anhydrase 2	P00918
Carbonyl reductase [NADPH] 1	P16152
Carboxypeptidase N subunit 2	P22792
Carboxypeptidase Q	Q9Y646
Cathepsin B	P07858
Cathepsin D	P07339
Cathepsin G	P08311
Cathepsin Z	Q9UBR2
CD59 glycoprotein	P13987
Cell division control protein 42 homolog	P60953
Ceruloplasmin	P00450
Chloride intracellular channel protein 1	O00299
Chloride intracellular channel protein 4	Q9Y696
Cholesteryl ester transfer protein	P11597
Clathrin heavy chain 1	Q00610
Clusterin	P10909
Coactosin-like protein	Q14019
Coagulation factor XII	P00748
Cofilin-1	P23528
Collagen alpha-1(III) chain	P02461
Collagen alpha-1(VI) chain	P12109
Collagen alpha-1(XV) chain	P39059
Collagen alpha-1(XVIII) chain	P39060
Collagen alpha-2(IV) chain	P08572
Collagen alpha-3(VI) chain	P12111
Complement C1s subcomponent	P09871
Complement C3	P01024
Complement C4-A	P0C0L4
Complement C4-B	P0C0L5
Complement C5	P01031
Complement component C7	P10643
Complement component C8 beta chain	P07358
Complement component C8 gamma chain	P07360

Complement component C9	P02748
Complement factor B	P00751
Complement factor D	P00746
Complement factor I	P05156
Core histone macro-H2A.1	O75367
Cornifin-A	P35321
Cornifin-B	P22528
Cornulin	Q9UBG3
Corticosteroid-binding globulin	P08185
Costars family protein ABRACL	Q9P1F3
C-reactive protein	P02741
Cubilin	O60494
Cystatin-B	P04080
Cystatin-C	P01034
Cystatin-M	Q15828
Cysteine-rich secretory protein 1	P54107
Cytochrome b-c1 complex subunit 2, mitochondrial	P22695
Cytochrome c oxidase subunit 4, mitochondrial	P13073
Cytochrome c oxidase subunit 5B, mitochondrial	P10606
Cytoplasmic aconitate hydratase	P21399
Cytosolic 10-formyltetrahydrofolate dehydrogenase	O75891
Cytosolic non-specific dipeptidase	Q96KP4
Deoxyribonuclease-1	P24855
Dermatopontin	Q07507
Desmocollin-2	Q02487
Desmoglein-2	Q14126
Dipeptidyl peptidase 1	P53634
Dipeptidyl peptidase 2	Q9UHL4
Dipeptidyl peptidase 4	P27487
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	P04844
Dystroglycan	Q14118
Ecto-ADP-ribosyltransferase 3	Q13508
EGF-containing fibulin-like extracellular matrix protein 1	Q12805
EGF-containing fibulin-like extracellular matrix protein 2	O95967
EH domain-containing protein 1	Q9H4M9
Elongation factor 1-alpha 1	P68104
Endonuclease domain-containing 1 protein	O94919
Endoplasmin	P14625
Endothelial protein C receptor	Q9UNN8
Enoyl-CoA hydratase, mitochondrial	P30084

ERO1-like protein alpha	Q96HE7
Extracellular matrix protein 1	Q16610
Extracellular superoxide dismutase [Cu-Zn]	P08294
Ezrin	P15311
Fatty acid-binding protein, adipocyte	P15090
Fatty acid-binding protein, epidermal	Q01469
Fatty acid-binding protein, heart	P05413
Fibrinogen alpha chain	P02671
Fibrinogen beta chain	P02675
Fibrinogen gamma chain	P02679
Fibronectin	P02751
Fibulin-1	P23142
Fibulin-5	Q9UBX5
Ficolin-2	Q15485
Flavin reductase (NADPH)	P30043
Fructose-bisphosphate aldolase B	P05062
Galectin-3	P17931
Galectin-3-binding protein	Q08380
Galectin-7	P47929
Gamma-glutamylaminecyclotransferase	Q9BVM4
Ganglioside GM2 activator	P17900
Gelsolin	P06396
Glucose-6-phosphate isomerase	P06744
Glutamine synthetase	P15104
Glutaminyl-peptide cyclotransferase	Q16769
Glutamyl aminopeptidase	Q07075
Glutaredoxin-1	P35754
Glutathione peroxidase 3	P22352
Glutathione S-transferase P	P09211
Glyceraldehyde-3-phosphate dehydrogenase	P04406
Glyoxylate reductase/hydroxypyruvate reductase	Q9UBQ7
Golgin subfamily A member 4	Q13439
Grancalcin	P28676
Growth/differentiation factor 15	Q99988
Guanylate-binding protein 7	Q8N8V2
Haptoglobin	P00738
Heat shock 70 kDa protein 1A/1B	P08107
Heat shock cognate 71 kDa protein	P11142
Heat shock protein beta-1	P04792
Heat shock protein HSP 90-alpha	P07900
Heat shock protein HSP 90-beta	P08238

Haeme-binding protein 2	Q9Y5Z4
Haemoglobin subunit alpha	P69905
Haemoglobin subunit beta	P68871
Haemoglobin subunit delta	P02042
Haemoglobin subunit gamma-2	P69892
Hemopexin	P02790
Heparin cofactor 2	P05546
Heterogeneous nuclear ribonucleoprotein A1	P09651
Heterogeneous nuclear ribonucleoprotein K	P61978
Heterogeneous nuclear ribonucleoproteins A2/B1	P22626
Hexokinase-3	P52790
Histidine-rich glycoprotein	P04196
Histone H1.2	P16403
Histone H2A type 1-B/E;Histone H2A.Z	P04908
Histone H2A type 2-C	Q16777
Histone H2B type 1-J	P06899
Histone H2B type 1-K	O60814
Histone H2B type 1-K	O60814
Histone H3.1	P68431
Histone H3.3	P84243
Histone H4	P62805
HLA class I histocompatibility antigen	P04439
HLA class I histocompatibility antigen	P05534
HLA class I histocompatibility antigen	P10319
HLA class I histocompatibility antigen, A-30 alpha chain	P16188
HLA class I histocompatibility antigen, B-7 alpha chain	P01889
HLA class I histocompatibility antigen, Cw-12 alpha chain	P30508
HLA class I histocompatibility antigen, Cw-2 alpha chain	P30501
Hornerin	Q86YZ3
Ig kappa chain V-I region Mev	P01612
Ig kappa chain V-I region Roy	P01608
Ig kappa chain V-II region TEW	P01617
Ig kappa chain V-III region CLL	P04207
Ig kappa chain V-III region GOL	P04206
Ig kappa chain V-III region NG9 (Fragment)	P01621
Ig kappa chain V-III region SIE	P01620
Ig kappa chain V-IV region Len	P01625
Ig lambda chain V-I region WAH	P04208
Ig lambda-2 chain C regions	P0CG05
Immunoglobulin alpha-2 heavy chain	P0DOX2
Immunoglobulin delta heavy chain	P0DOX3

Immunoglobulin gamma-1 heavy chain	P0DOX5
Immunoglobulin heavy constant alpha 1	P01876
Immunoglobulin heavy constant alpha 2	P01877
Immunoglobulin heavy constant delta	P01880
Immunoglobulin heavy constant gamma 1	P01857
Immunoglobulin heavy constant gamma 2	P01859
Immunoglobulin heavy constant gamma 3	P01860
Immunoglobulin heavy constant gamma 4	P01861
Immunoglobulin heavy constant mu	P01871
Immunoglobulin heavy variable 1-69	P01742
Immunoglobulin heavy variable 3-23	P01764
Immunoglobulin heavy variable 3-49	A0A0A0MS15
Immunoglobulin heavy variable 3-7	P01780
Immunoglobulin heavy variable 4-34	P06331
Immunoglobulin heavy variable 4-39	P01824
Immunoglobulin J chain	P01591
Immunoglobulin kappa constant	P01834
Immunoglobulin kappa light chain	P0DOX7
Immunoglobulin kappa variable 1-16	P04430
Immunoglobulin kappa variable 1-17	P01599
Immunoglobulin kappa variable 1-27	A0A075B6S5
Immunoglobulin kappa variable 1-33	P01594
Immunoglobulin kappa variable 1-5	P01602
Immunoglobulin kappa variable 1D-12	P01611
Immunoglobulin kappa variable 1D-33	P01593
Immunoglobulin kappa variable 2-24	A0A0C4DH68
Immunoglobulin kappa variable 2-29	A2NJV5
Immunoglobulin kappa variable 2D-28	P01615
Immunoglobulin kappa variable 2D-29	A0A075B6S2
Immunoglobulin kappa variable 3-15	P01624
Immunoglobulin kappa variable 3-20	P01619
Immunoglobulin kappa variable 3D-11	A0A0A0MRZ8
Immunoglobulin lambda variable 1-40	P01703
Immunoglobulin lambda variable 1-44	P01699
Immunoglobulin lambda variable 1-51	P01701
Immunoglobulin lambda variable 3-19	P01714
Immunoglobulin lambda variable 3-21	P80748
Immunoglobulin lambda variable 3-25	P01717
Immunoglobulin lambda variable 4-69	A0A075B6H9
Immunoglobulin lambda variable 6-57	P01721
Immunoglobulin lambda variable 7-46	A0A075B6I9

Immunoglobulin lambda variable 9-49	A0A0B4J1Y8
Immunoglobulin lambda-1 light chain	P0DOX8
Immunoglobulin lambda-like polypeptide 1	P15814
Immunoglobulin lambda-like polypeptide 5	B9A064
Immunoglobulin superfamily containing leucine-rich repeat protein	O14498
Immunoglobulin superfamily member 8	Q969P0
Insulin-degrading enzyme	P14735
Insulin-like growth factor-binding protein 1	P08833
Insulin-like growth factor-binding protein 2	P18065
Insulin-like growth factor-binding protein 4	P22692
Insulin-like growth factor-binding protein 6	P24592
Insulin-like growth factor-binding protein 7	Q16270
Integrin alpha-M	P11215
Inter-alpha-trypsin inhibitor heavy chain H1	P19827
Inter-alpha-trypsin inhibitor heavy chain H2	P19823
Inter-alpha-trypsin inhibitor heavy chain H4	Q14624
Interferon alpha/beta receptor 2	P48551
Junction plakoglobin	P14923
Kallistatin	P29622
Keratin, type I cytoskeletal 10	P13645
Keratin, type I cytoskeletal 13	P13646
Keratin, type I cytoskeletal 14	P02533
Keratin, type I cytoskeletal 15	P19012
Keratin, type I cytoskeletal 17	Q04695
Keratin, type I cytoskeletal 19	P08727
Keratin, type I cytoskeletal 9	P35527
Keratin, type II cytoskeletal 1	P04264
Keratin, type II cytoskeletal 2 epidermal	P35908
Keratin, type II cytoskeletal 4	P19013
Keratin, type II cytoskeletal 5	P13647
Keratin, type II cytoskeletal 6A	P02538
Keratin, type II cytoskeletal 6B	P04259
Keratin, type II cytoskeletal 6C	P48668
Keratin, type II cytoskeletal 79	Q5XKE5
Keratin, type II cytoskeletal 8	P05787
Ketimine reductase mu-crystallin	Q14894
Kininogen-1	P01042
Lactotransferrin	P02788
Lambda-crystallin homolog	Q9Y2S2
Leucine-rich alpha-2-glycoprotein	P02750
Leukocyte elastase inhibitor	P30740

Leukocyte immunoglobulin-like receptor subfamily B member 4	Q8NHJ6
Lipopolysaccharide-binding protein	P18428
L-lactate dehydrogenase A chain	P00338
L-lactate dehydrogenase B chain	P07195
Low-density lipoprotein receptor-related protein 2	P98164
Lumican	P51884
Lysosomal protective protein	P10619
Lysozyme C	P61626
Malate dehydrogenase, cytoplasmic	P40925
Malectin	Q14165
Mannan-binding lectin serine protease 1	P48740
Mannan-binding lectin serine protease 2	O00187
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	P33908
Matrilysin	P09237
Matrix metalloproteinase-9	P14780
Matrix remodeling-associated protein 8	Q9BRK3
Migration and invasion enhancer 1	Q9BRT3
Mimecan	P20774
Mitogen-activated protein kinase 15	Q6ZN16
Moesin	P26038
Monocyte differentiation antigen CD14	P08571
Multimerin-2	Q9H8L6
Myeloblastin	P24158
Myeloid-derived growth factor	Q969H8
Myeloperoxidase	P05164
Myoglobin	P02144
Myosin light polypeptide 6	P60660
Myosin regulatory light chain 12B;Myosin regulatory light chain 12A	O14950;P19105
Myosin-9	P35579
N(G),N(G)-dimethylarginine dimethylaminohydrolase 1	O94760
N(G),N(G)-dimethylarginine dimethylaminohydrolase 2	O95865
Na(+)/H(+) exchange regulatory cofactor NHE-RF1	O14745
N-acetylmuramoyl-L-alanine amidase	Q96PD5
Napsin-A	O96009
Nectin-2	Q92692
Neutrophil defensin 1	P59665
Neutrophil elastase	P08246
Nidogen-1	P14543
Nuclear transport factor 2	P61970
Nucleobindin-1	Q02818

Olfactomedin-4	Q6UX06
Osteopontin	P10451
Out at first protein homolog	Q86UD1
Pancreatic alpha-amylase	P04746
Peptidase inhibitor 16	Q6UXB8
Peptidoglycan recognition protein 1	O75594
Peptidyl-glycine alpha-amidating monooxygenase	P19021
Peptidyl-prolyl cis-trans isomerase A	P62937
Peptidyl-prolyl cis-trans isomerase C	P45877
Peptidyl-prolyl cis-trans isomerase FKBP1A	P62942
Peroxiredoxin-1	Q06830
Peroxiredoxin-2	P32119
Peroxiredoxin-5, mitochondrial	P30044
Peroxisomal sarcosine oxidase	Q9P0Z9
Phosphoglycerate kinase 1	P00558
Phosphoglycerate mutase 1	P18669
Pigment epithelium-derived factor	P36955
Plasma alpha-L-fucosidase	Q9BTY2
Plasma protease C1 inhibitor	P05155
Plasma serine protease inhibitor	P05154
Plasminogen	P00747
Plastin-2	P13796
Plastin-3	P13797
Polymeric immunoglobulin receptor	P01833
Polyubiquitin-B;Polyubiquitin-C;Ubiquitin-40S ribosomal protein S27a;Ubiquitin-60S ribosomal protein L40	P0CG47;P0CG48;P62979;P62987
Probable serine carboxypeptidase CPVL	Q9H3G5
Pro-epidermal growth factor	P01133
Profilin-1	P07737
Proline-rich protein 4	Q16378
Propionyl-CoA carboxylase beta chain, mitochondrial	P05166
ProSAAS	Q9UHG2
Prosaposin	P07602
Prostaglandin-H2 D-isomerase	P41222
Prostate-specific antigen	P07288
Prostatic acid phosphatase	P15309
Proteasome activator complex subunit 1	Q06323
Proteasome activator complex subunit 2	Q9UL46
Proteasome inhibitor PI31 subunit	Q92530
Protein ABHD14B	Q96IU4
Protein AMBP	P02760

Protein disulfide-isomerase	P07237
Protein disulfide-isomerase A3	P30101
Protein disulfide-isomerase A6	Q15084
Protein DJ-1	Q99497
Protein FAM3C	Q92520
Protein S100-A12	P80511
Protein S100-A14	Q9HCY8
Protein S100-A16	Q96FQ6
Protein S100-A6	P06703
Protein S100-A8	P05109
Protein S100-A9	P06702
Prothrombin	P00734
Protocadherin-1	Q08174
Putative PDZ domain-containing protein PDZK1P1;Na(+)/H(+) exchange regulatory cofactor NHE-RF3	A8MUH7
Pyruvate kinase PKM	P14618
Quinone oxidoreductase PIG3	Q53FA7
Rab GDP dissociation inhibitor beta	P50395
Ras-related C3 botulinum toxin substrate 1	P63000
Ras-related C3 botulinum toxin substrate 2	P15153
Ras-related protein Rab-11A	P62491
Renin receptor	O75787
Resistin	Q9HD89
Reticulocalbin-1	Q15293
Retinoic acid receptor responder protein 2	Q99969
Retinol-binding protein 4	P02753
Rho GDP-dissociation inhibitor 1	P52565
Roundabout homolog 4	Q8WZ75
Salivary acidic proline-rich phosphoprotein 1/2	P02810
Secreted and transmembrane protein 1	Q8WVN6
Serotransferrin	P02787
Serpin B3	P29508
Serpin B4	P48594
Serum albumin	P02768
Serum amyloid A-1 protein	P0DJI8
SH3 domain-binding glutamic acid-rich-like protein 3	Q9H299
Sialidase-1	Q99519
Small proline-rich protein 2D	P22532
Small proline-rich protein 3	Q9UBC9
Sorbitol dehydrogenase	Q00796
Sulfhydryl oxidase 1	O00391

Superoxide dismutase [Cu-Zn]	P00441
Supervillin	O95425
Sushi domain-containing protein 2	Q9UGT4
Syntabulin	Q9NX95
Tetranectin	P05452
Thioredoxin	P10599
Thioredoxin domain-containing protein 17	Q9BRA2
Thioredoxin domain-containing protein 5	Q8NBS9
Thioredoxin-dependent peroxide reductase	P30048
Thrombospondin-1	P07996
Thymidine phosphorylase	P19971
Thyroxine-binding globulin	P05543
Tissue alpha-L-fucosidase	P04066
Titin	Q8WZ42
Transaldolase	P37837
Transgelin-2	P37802
Transitional endoplasmic reticulum ATPase	P55072
Transketolase	P29401
Transthyretin	P02766
Triosephosphate isomerase	P60174
Tripeptidyl-peptidase 1	O14773
Tropomyosin alpha-3 chain	P06753
Tsukushin	Q8WUA8
Tubulin alpha-1B chain	P68363
Tubulin alpha-4A chain	P68366
Tubulin beta-4B chain	P68371
Unconventional myosin-VI	Q9UM54
UPF0587 protein C1orf123	Q9NWV4
Urokinase-type plasminogen activator	P00749
Uromodulin	P07911
Uteroglobin	P11684
Vascular cell adhesion protein 1	P19320
Vasorin	Q6EMK4
Ventricular zone-expressed PH domain-containing protein homolog 1	Q14D04
Vesicular integral-membrane protein VIP36	Q12907
Vimentin	P08670
Vitamin D-binding protein	P02774
Vitamin K-dependent protein S	P07225
Vitamin K-dependent protein Z	P22891
Vitelline membrane outer layer protein 1 homolog	Q7Z5L0

Vitronectin	P04004
von Willebrand factor A domain-containing protein 1	Q6PCB0
V-type proton ATPase subunit S1	Q15904
WNT1-inducible-signaling pathway protein 2	O76076
Zinc-alpha-2-glycoprotein	P25311
Zymogen granule protein 16 homolog B	Q96DA0

Table A3: Expanded GO analysis showing list of putative biomarkers involved in each pathway.

Pathway name	Total proteins in pathway	Total proteins found in pathway	FDR	Proteins identified in pathway
Platelet degranulation	137	13	1.07E-12	P01011;P04217;P07737;P62937;P01009;P02768;P02647;P02679;P04196;Q92520;P02787;P02765;P02675
Response to elevated platelet cytosolic Ca2+	144	13	1.07E-12	P01011;P04217;P07737;P62937;P01009;P02768;P02647;P02679;P04196;Q92520;P02787;P02765;P02675
Innate Immune System	1302	25	7.77E-11	P04217;P05109;P62937;P0C0L5;P17900;P68871;P02748;P01834;P01833;P02765;P61769;P01034;P01011;P04004;P06702;P60709;P00738;P08571;P01625;P01009;P02679;P02675;P00751;P01024;P02750
Haemostasis	806	20	2.96E-10	P01011;P04217;P60709;P07737;P62937;P68871;P01625;P01009;P01834;P02768;P01008;P02647;P02679;P01876;P04196;Q92520;P02787;P02765;P05154;P02675
Neutrophil degranulation	480	16	8.02E-10	P01011;P06702;P04217;P05109;P62937;P00738;P17900;P08571;P68871;P01009;P01833;P02765;P61769;P01024;P02750;P01034
Platelet activation, signalling and aggregation	288	13	1.99E-09	P01011;P04217;P07737;P62937;P01009;P02768;P02647;P02679;P04196;Q92520;P02787;P02765;P02675
Scavenging of haem from plasma	106	9	1.05E-08	P02760;P00738;P68871;P01625;P01834;P02768;P02647;P01876;P69905
Post-translational protein phosphorylation	109	9	1.15E-08	P01009;P02768;P01008;P02647;P02679;P02787;P02765;P01024;P01034
Binding and Uptake of Ligands by Scavenger Receptors	167	10	2.11E-08	P02760;P00738;P68871;P01625;P01834;P02768;P02647;P01876;P69905;P27797
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by	127	9	3.50E-08	P01009;P02768;P01008;P02647;P02679;P02787;P02765;P01024;P01034

Insulin-like Growth Factor Binding Proteins (IGFBPs)				
Immune System	2638	28	5.92E-07	P04217;P05109;P62937;P0C0L5;P17900;P68871;P02748;P04083;P01834;P01833;P02765;P61769;P01034;P01011;P04004;P06702;P60709;P00738;P00915;P08571;P01625;P01009;P02679;P02675;P00751;P01024;P27797;P02750
Regulation of TLR by endogenous ligand	31	5	3.32E-06	P06702;P05109;P08571;P02679;P02675
Regulation of Complement cascade	139	7	2.11E-05	P04004;P0C0L5;P02748;P01625;P01834;P00751;P01024
Complement cascade	156	7	4.11E-05	P04004;P0C0L5;P02748;P01625;P01834;P00751;P01024
Activation of C3 and C5	7	3	5.58E-05	P0C0L5;P00751;P01024
Vesicle-mediated transport	818	13	1.35E-04	P02760;P60709;P00738;P68871;P01625;P01009;P01834;P02768;P02647;P01876;P02787;P69905;P27797
Common Pathway of Fibrin Clot Formation	34	4	1.51E-04	P01008;P02679;P05154;P02675
Amyloid fiber formation	78	5	1.88E-04	P06727;P02647;P61769;P98160;P01034
Erythrocytes take up oxygen and release carbon dioxide	16	3	4.88E-04	P00915;P68871;P69905
Formation of Fibrin Clot (Clotting Cascade)	52	4	6.95E-04	P01008;P02679;P05154;P02675
Initial triggering of complement	120	5	0.001232969	P0C0L5;P01625;P01834;P00751;P01024
Erythrocytes take up carbon dioxide and release oxygen	23	3	0.001232969	P00915;P68871;P69905
O2/CO2 exchange in erythrocytes	23	3	0.001232969	P00915;P68871;P69905
Retinoid metabolism and transport	79	4	0.002690855	P06727;P02647;P02753;P98160
Alternative complement activation	6	2	0.002690855	P00751;P01024
Integrin cell surface interactions	86	4	0.003472011	P04004;P02679;P02675;P98160
Metabolism of fat-soluble vitamins	94	4	0.004406949	P06727;P02647;P02753;P98160
Toll-Like Receptors Cascades	183	5	0.006206646	P06702;P05109;P08571;P02679;P02675

Signalling by high-kinase activity BRAF mutants	46	3	0.006836911	P60709;P02679;P02675
MAP2K and MAPK activation	47	3	0.006836911	P60709;P02679;P02675
Signalling by moderate kinase activity BRAF mutants	48	3	0.0072635	P60709;P02679;P02675
Paradoxical activation of RAF signalling by kinase inactive BRAF	49	3	0.007706666	P60709;P02679;P02675
Metal sequestration by antimicrobial proteins	13	2	0.009034046	P06702;P05109
Plasma lipoprotein remodelling	54	3	0.009048769	P06727;P02768;P02647

Table A4: Summary of urinary proteome studies showing sample preparation and LCMS data acquisition methods as well as total protein identifications.

Reference	Extraction/Clean up method	Nano/micro-flow	1D/2D LC	Gradient length (min per injection)	Protein groups (SD)
(Yu *et al.*, 2014)	Ultrafiltration-FASP	nano	1D	130	1063 (15)
	Ultrafiltration-96FASP	nano	1D	130	894 (18)
(Dwivedi *et al.*, 2016)	Ultrafiltration-FASP	nano	2D	30 (80 min x 10 fractions)	1324
		nano	1D SWATH	80	730
(Matafora *et al.*, 2014)	Ultrafiltration-FASP	nano	1D	320	812
(Caterino *et al.*, 2018)	Ultrafiltration-Precipitation-1D SDS	nano	1D	45	Male group - 73 Female group - 68
(Beretov *et al.*, 2015)	Precipitation	nano	1D	60	59 differentially abundant
(Nagaraj and Mann, 2011)	Ultrafiltration-Precipitation-StageTip	nano	1D	120	633
(Adachi *et al.*, 2006)	1D SDS – Albumin depletion	analytical	2D	54	
		nano	1D	90 (x 54 fractions)	1543
(Pang *et al.*, 2018)	Precipitation-Ultrafiltration	analytical	2D	45	
		nano	1D	150 (x 10 fractions)	509