

https://dspace.org/

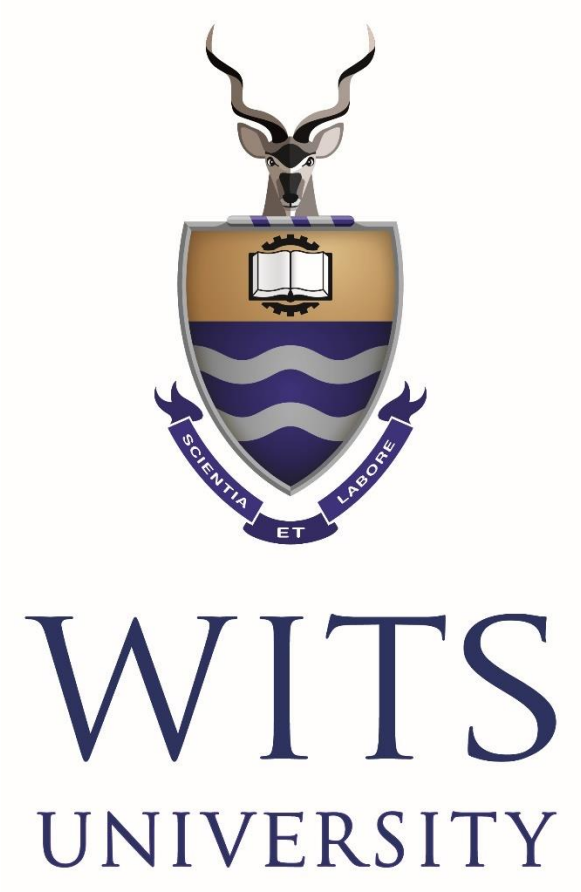
Method optimization for the identification of proteins associated with pancreatic ductal adenocarcinoma in plasma using mass spectrometr y

Dubazana, Sinegugu

2024
https://hdl.handle.net/10539/37701

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

METHOD OPTIMIZATION FOR THE IDENTIFICATION OF PROTEINS ASSOCIATED WITH PANCREATIC DUCTAL ADENOCARCINOMA IN PLASMA USING MASS SPECTROMETRY



Sinegugu Dubazana (1607284)

A thesis submitted in fulfilment of the requirement for the Degree of Master of Science

DEPARTMENT OF SURGERY, FACULTY OF HEALTH SCIENCES,
UNIVERSITY OF WITWATERSRAND, JOHANNESBURG.

MARCH 2023

DECLARATION

I, Sinegugu Dubazana declare that this thesis entitled “Method Optimization for the Identification of Proteins Associated with Pancreatic Ductal Adenocarcinoma in Plasma using Mass Spectrometry” is my unaided work. It is being submitted for the Degree of Master of Science at the University of Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other university.



_____ 22/06/2023

Sinegugu Dubazana Date

DEDICATION

Primarily, I thank God Almighty for being my source of wisdom and understanding throughout this project and for his continued blessing towards the successful completion of my work. Additionally, I would like to dedicate this work to my friends and family who have shown me immense support and encouragement.

RESEARCH OUTPUTS: CONFERENCE PRESENTATIONS

Oral presentations:

1. **Sinegugu Dubazana,** Sindisiwe Buthelezi, Previn Naicker, and Ekene Nweke. Method optimization for the verification of Pancreatic Ductal Adenocarcinoma reporter proteins in plasma. 49th Surgical Research Society of Southern Africa Meeting (Hybrid), GE Conference Centre, Melrose Estate, Johannesburg, 8th July 2022.

2. **Sinegugu Dubazana,** Sindisiwe Buthelezi, Previn Naicker and Ekene Nweke. Method optimization for the verification of Pancreatic Ductal Adenocarcinoma reporter proteins in plasma. Faculty of Health Sciences Research Day and Postgraduate Expo 2022, Department of Surgery, University of the Witwatersrand, Johannesburg, 15th September 2022.

3. **Sinegugu Dubazana,** Sindisiwe Buthelezi, Previn Naicker and Ekene Nweke. Method optimization for the verification of Pancreatic Ductal Adenocarcinoma reporter proteins in plasma. The 2022 HPBASA Congress, Wanderers, Johannesburg, 11th-13th November 2022.

ABSTRACT

Background: Pancreatic ductal adenocarcinoma (PDAC) is a leading cause of cancer-related deaths. The poor survival rate for PDAC is mostly due to the lack of effective biomarkers for early diagnosis. This study aimed to develop a standard operating procedure for a high-throughput mass spectrometry-based workflow that can be used for plasma-based biomarker discovery studies.

Methods: Total protein was extracted from a master pool consisting of 28 PDAC samples (M211158). SDS-PAGE was conducted to visualize the dynamic range of the plasma proteins. Three sample preparation methods were compared, Method 1: plasma diluted 25X with buffer (50mM Tris-HCl pH 8 with 2% SDS), Method 2: plasma diluted 5X with 50mM TEAB buffer, and Method 3: precipitation (using 80% and 100% acetone) followed by resuspension in 50 mM Tris-HCl, 2% SDS. The Hydrophilic Interaction Liquid Chromatography (HILIC) and protein aggregation capture (PAC) based protocols were compared for clean-up and digestion of the proteins. Low pH reverse phase chromatography was applied to separate the peptides and SWATH-MS analysis was conducted for label-free quantification. The methods were compared based on reproducibility, ease of automation and the number of peptides identified. A pilot study including 6 PDAC, 6 Chronic pancreatitis, and 6 healthy individuals was conducted to test the selected optimized method.

Results: The combination of the 25-fold dilution of Method 1 and the 2 hrs HILIC bead-based method (with a Trypsin and LysC combination) were effective in protein sample clean-up and digestion, yielding the highest number of peptides (mean: 2856). The combination of Trypsin and LysC enzymes contributed to the improvements in technical variation and peptide quantitation. The preliminary analysis identified 25 differentially expressed proteins (22 downregulated and 3 upregulated) between the healthy and PDAC patients. Most of these proteins are associated with tumorigenesis.

Conclusion: This study developed a standard operating procedure for processing plasma samples and can be applied in future studies for the identification of potential markers for PDAC.

ACKNOWLEDGEMENT

- I would like to express my utmost gratitude to the Council for Scientific and Industrial Research (CSIR) for hosting and allowing access to the laboratory and for financial aid. To the University of the Witwatersrand for allowing me the opportunity to further my studies, I am highly grateful.

- I would also like to thank my colleagues, Rethabile Mokoena and Pateswana Mokwena at the CSIR for their continued support and assistance during my research.

- I would like to give special thanks to my supervisors, Dr Ekene Emmanuel Nweke, Dr Sindisiwe Buthelezi and Dr Previn Naicker for their encouragement, dedication, and patient guidance. The mentorship, advice, and knowledge I have gained from them have shaped me for the future.

- Lastly, I would also like to thank Dr Nnenna Elebo and the clinical team at the Hepatopancreatobiliary unit of Chris Hani Baragwanath Hospital, Johannesburg for assistance with sample collection and patient identification, respectively.

TABLE OF CONTENTS

DECLARATION i
DEDICATION ii
RESEARCH OUTPUTS: CONFERENCE PRESENTATIONS iii
ABSTRACT iv
ACKNOWLEDGEMENT v
CHAPTER 1 1
LITERATURE REVIEW 1
1.1. The Pancreas: structure and function 1
1.2. Pancreatic ductal adenocarcinoma 1
1.2.1 Epidemiology of Pancreatic cancer 1
1.2.2. Pancreatic cancer types and PDAC development 3
1.2.3. Risk factors that predispose individuals to PDAC 4
1.2.4. Diagnosis of Pancreatic ductal adenocarcinoma 5
1.2.5. Classification of PDAC 7
1.2.6. Treatment of PDAC 7
1.3. Biomarker discovery: The role of Proteomics 8
1.4. Mass-spectrometry technologies 11
1.5. Plasma proteomics, depletion, fractionation 14
1.6. Method standardisation for biomarker discovery 15
1.7. Justification of study 16
1.8. Aim 16
1.9. Objectives 16
CHAPTER 2 18
METHODS AND MATERIALS 18
2.1. Experimental strategy 18
2.2. Solutions, Buffers and Solvents 20
2.3. Sample collection and processing 20

2.4. Sample preparation 20
2.4.1. Sodium dodecyl-sulfate polyacrylamide gel electrophoresis 20
Sodium dodecyl-sulfate polyacrylamide gel electrophoresis 20
2.4.2. Strong Anion Exchange (SAX) 21
2.4.3. Protein Depletion 22
2.4.4. Solution-based method optimization 23
2.5. Building the spectral library 25
2.6. Data acquisition for the spectral library building 25
2.7. Evotip loading and data acquisition for the pilot study 26
2.8. Pathway analysis and network interaction 27
2.9. Statistical analysis 27
CHAPTER 3 28
RESULTS 28
Section 1: Method optimization 28
3.1. Assessment of the dynamic range of study samples. 28
3.2. Spectral library building 31
3.3. Experiment 1: Evaluation of plasma protein preparation methods 32
3.4. Experiment 2: evaluation of selected protein preparation methods and sample clean-up methods. 36
3.5. Experiment 3: Evaluation of digestion conditions (time) 39
3.6. Experiment 4: Evaluation of digestion condition and enzyme efficiency 41
3.7. Final standard operating procedure 45
Section 2: Pilot study 46
3.8. Proteome profiling of plasma samples 46
CHAPTER 4 54
DISCUSSION 54
Section 1: Method optimization 54
4.1. Plasma contains highly abundant proteins 54
4.2. Spectral library building: Depletion and fractionation of abundant proteins 56
4.3. Plasma sample preparation for protein analysis 57
4.4. Sample dilution, clean-up and protein digestion 59
Section 2: Pilot study 62

4.5. Dysregulated proteins may be involved in PDAC development 62
4.6. Dysregulated proteins are enriched in key pathways involved in carcinogenesis 64
4.7. Significantly dysregulated proteins in plasma may be secreted by tumours 67
CHAPTER 5 69
Conclusion and future work 69
REFERENCES 70
APPENDIX 82

LIST OF FIGURES

Figure 1. 1. Pancreatic ductal adenocarcinoma initiation and progression. 4
Figure 1. 2. Schematic representation of the electrospray ionization process..... 12
Figure 1. 3. Schematic diagram of SWATH-MS principle. 14

Figure 2. 1. Schematic overview of experimental strategy. 19
Figure 2. 2. Schematic representation of hydrophilic interaction liquid chromatography (HILIC)-based clean-up and digestion on an automated KingFisher™ Duo magnetic handling station (Thermo Scientific). 24
Figure 2. 3. Schematic representation of hydrophilic interaction liquid chromatography (HILIC)-based clean-up and digestion on an automated KingFisher™ FLEX magnetic handling station (Thermo Scientific).....25

Figure 3. 1. Visualization of plasma dynamic range on SDS-PAGE. 29
Figure 3. 2. Depleted plasma samples analyzed via SDS-PAGE. 30
Figure 3. 3. Sample fractionation by SAX analyzed on SDS-PAGE.. 30
Figure 3. 4. Bar graphs of identified peptides and protein groups from depletion methods... 31
Figure 3. 5. Bar graph of identified peptides and protein groups from SAX fractionation. ... 32
Figure 3. 6. Normalization plot for experiment 1. 33
Figure 3. 7. Box plot indicating inter-day variability per condition over 2 days (A and B) in experiment 1. 34
Figure 3. 8. Box plot indicating variability across all replicates per condition in experiment 1. 35
Figure 3. 9. Bar graph of identified peptides and protein groups from experiment 1..... 36
Figure 3. 10. Box plot indicating inter-day variability per condition over 2 days (A and B) in experiment 2. 37
Figure 3. 11. Box plot indicating variability across all replicates per condition in experiment 2..... 38
Figure 3. 12. Bar graph of identified peptides and protein groups from experiment 2..... 39
Figure 3. 13. Box plot indicating inter-day variability per condition over 2 days (A and B) in experiment 3. 40
Figure 3. 14. Box plot indicating intra-day variability per condition in experiment 3. 40
Figure 3. 15. Bar graph of identified peptides and protein groups from experiment 3..... 41
Figure 3. 16. Box plot indicating inter-day variability per condition over 2 days (A and B) in experiment 4. 42

Figure 3. 17. Box plot indicating intra-day variability per condition in experiment 4. 43
Figure 3. 18. Bar graph of identified peptides and protein groups from experiment 4. 44
Figure 3. 19. Final standardized workflow. 45
Figure 3. 20. Volcano plots displaying dysregulated proteins between (A) PDAC and healthy patients and (B) PDAC and CP patients. 49
Figure 3. 21. Protein-protein network analysis. 52
Figure 3. 22. Venn diagram and table of candidate proteins from tissue and plasma study. .. 53

Figure 4.4. 1. Protein denaturation by SDS. 60

Figure S 1. MS1 total ion chromatograms (TICs) of 100% acetone precipitation replicates overlaid using Peak View. 85
Figure S 2. MS1 total ion chromatograms (TICs) of 5x TEAB replicates overlaid using Peak View. 85
Figure S 3. MS1 total ion chromatograms (TICs) of 80% acetone precipitation replicates overlaid using Peak View. 86
Figure S 4. MS1 total ion chromatograms (TICs) of 25x Tris-HCl replicates overlaid using Peak View. 86
Figure S 5. MS1 total ion chromatograms (TICs) of TEAB_HILIC (TEAB_H) replicates overlaid using Peak View. 87
Figure S 6. MS1 total ion chromatograms (TICs) Tris_HILIC (Tris_H) replicates overlaid using Peak View. 88
Figure S 7. MS1 total ion chromatograms (TICs) of Tris_PAC (Tris_P) replicates overlaid using Peak View. 88
Figure S 8. Normalization plot for experiment 2. 89
Figure S 9. MS1 total ion chromatograms (TICs) of 4 hrs digestion replicates overlaid using Peak View. 90
Figure S 10. MS1 total ion chromatograms (TICs) of 16 hrs digestion replicates overlaid using Peak View. 91
Figure S 11. Normalization plot for experiment 3. 91
Figure S 12. MS1 total ion chromatograms (TICs) of 4 hrs trypsin digestion replicates overlaid using Peak View. 92
Figure S 13. MS1 total ion chromatograms (TICs) of 4-hour trypsin-LysC digestion replicates overlaid using Peak View. 93
Figure S 14. MS1 total ion chromatograms (TICs) of 2 hrs trypsin-LysC digestion replicates overlaid using Peak View. 94

Figure S 15. MS1 total ion chromatograms (TICs) of 16 hrs trypsin digestion replicates overlaid using Peak View. 94
Figure S 16. MS1 total ion chromatograms (TICs) of 16 hrs trypsin-LysC digestion replicates overlaid using Peak View. 95
Figure S 17. Normalization plot for experiment 4. 96

LIST OF TABLES

Table 3. 1. Overview of parameters for the generated spectral library. 32
Table 3. 2. differentially expressed proteins measured in group comparisons (healthy patients versus patients with PDAC). 46
Table 3. 3. The enriched pathways from the dysregulated proteins between the healthy and PDAC patients. 50

Table S 1. A list of all buffer solutions and solvents with their respective components and company manufacturer. 82
Table S 2. Patient demographic characteristics and clinical stage of disease. 82
Table S 3. Power analysis for a minimum of 6 samples. 96
Table S 4. Power analysis identifying the minimum number of samples required for a power of 0.9 and a fold change of 1.5. 97

LIST OF ABBREVIATIONS

18F-FDG- 18F-flurodeoxyglucose

A2M- Immunoglobulin heavy constant alpha 2

ACN- Acetonitrile

ADM- Acinar-to-duct metaplasia

AP-Acute pancreatitis

APOC1- Apolipoprotein C1

APOC2- Apolipoprotein C2

APOE- Apolipoprotein E

AUC- Area under the curve

BMI- Body mass index

C3- Complement component 3

C4BPA- Complement component 4 binding protein

CBD- Common bile duct

CDKN2A- Cyclin-dependant kinase inhibitor 2A

CEA- Carcinoembryonic antigen

CFB- Complement factor B

CID- Collision-induced dissociation

CLEC3B- C-type lectin domain family 3 member B

CLU- Clusterin

CP- Chronic pancreatitis

CPTAC- Clinical Proteomic Tumour Analysis Consortium

CT- Computerized tomography

CV- Coefficient of variation

DIA- Data independent acquisition

DTT- Dithiothreitol

EDTA- Ethylenediaminetetraacetic acid

ESI- Electrospray ionisation

FDR- False discovery rate

FGB- Fibrinogen

GB- Gallbladder

HILIC- Hydrophilic interaction liquid chromatography

HP- Haptoglobin

HREC- Human research ethics committee

IgA- Immunoglobulin A

IgG- Immunoglobulin G

IgM- Immunoglobulin M

IAA- Iodoacetamide

IPMNs- Intraductal papillary mucinous neoplasms

KRAS- Kirsten rat sarcoma virus

LC-MS/MS- Liquid chromatography-mass spectrometry

LRG1- Leucine-rich alpha-2-glycoprotein 1

MALDI-TOF/MS- Matrix-assisted laser desorption/ionization-time of flight/mass spectrometry

MDCT- Multidetector computed tomography

MRI- Magnetic resonance imaging

MS- Mass spectrometry

NAT- Neoadjuvant therapy

NCCN- National comprehensive cancer network

NPLC-MS/MS- Normal phase liquid chromatography-mass spectrometry

NRF- National research foundation

PanINs- pancreatic intra-epithelial neoplasias

PC- Pancreatic cancer

PCS- Primary sclerosis cholangitis

PDAC- Pancreatic ductal adenocarcinoma

PET- Position emission tomography

PRSS1- Serine protease 1

PSM- Peptide-spectrum match

PTM- Post-translational modification

Q1- First quadrupole mass analyzer

Q2- Second quadrupole collision cell

Q3- Third quadrupole mass analyzer

QC- Quality control

RF- Radio frequency

RPM- Revolutions per minute

RT- Radiation therapy

SAA1- Serum amyloid

SAMRC- South African medical research council

SAX- Strong anionic exchange

SDS-PAGE- Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SID- Stable isotope dilution

SMAD4- Mothers against decapentaplegic homologue 4

SOP- Standard operating procedure

SPINK1- Serine protease inhibitor kazal-1

SWATH- Sequential window acquisition of all theoretical mass spectra

T2DM- Type-2 diabetes mellitus

T3cDM- Type-3c diabetes mellitus

TEAB- Triethylammonium bicarbonate

TICs- Total ion chromatograms

TP53- Tumour suppressor protein 53

TTR- Tranthyretin

WHO- World health organisation

CHAPTER 1

LITERATURE REVIEW

1.1. The Pancreas: structure and function

The pancreas is an endoderm-derived organ located behind the stomach that regulates carbohydrate and protein digestion and glucose homeostasis (Hezel *et al.*, 2006). It is labelled as an endocrine gland however; it conducts both endocrine and exocrine functions (Aier *et al.*, 2019). The organ comprises the endocrine region composed of small irregular patches of cells known as the islets of Langerhans (Aier *et al.*, 2019) and exocrine (acinar and ductal) cells which account for 80% of the tissue mass of the organ (Hezel *et al.*, 2006). The exocrine acinar cells secrete digestive enzymes such as proteinases, amylases, and lipases, which become secreted in pancreatic ducts before moving into the small intestines breaking down proteins, fats, and carbohydrates for absorption (Zhou and Melton, 2018). Despite making up < 5% of the overall pancreatic mass, there are a billion and above cells in the endocrine islets of Langerhans in humans. The five main types of islets cells synthesize and secrete principal hormones including glucagon (α-cells), insulin (β-cells), somatostatin (δ-cells), ghrelin (ε-cells) and polypeptide (PP cells). Through a complex intra-islets vascular network, insulin and glucagon are directly released into the bloodstream and play crucial roles in controlling blood glucose levels (Zhou and Melton, 2018).

1.2. Pancreatic ductal adenocarcinoma

1.2.1 Epidemiology of Pancreatic cancer

Pancreatic cancer (PC) is one of the most lethal cancers with a 5-year survival rate of 12% (Siegel *et al.*, 2023). PC is also the third leading cause of cancer deaths in both women and men (Siegel *et al.*, 2023). Pancreatic ductal adenocarcinoma (PDAC) makes up 90% of PC cases (Adamska *et al.*, 2017). In the year 2020, GLOBOCAN conducted a survey that revealed 495,773 new cases of PDAC, and the number of deaths secondary to PDAC was 466,003 per year worldwide (Ushio *et al.*, 2021). Multiple studies have reported that individuals of African ancestry experience 40%-90% more incidence and have the poorest prognosis (survival rate of 10%-20%) of the disease compared to other ethnic groups (Heller et al., 2020, Chang et al., 2005, Khawja et al., 2015). The poor statistics could be attributed to

genetic and social factors such as diabetes and smoking, as well as late diagnoses (Silverman *et al.*, 2003). This suggests a need for understanding the disease biology in these patients. PDAC has been observed to be more common in males than females and the incidence rate increases with age for all genders (Rawla *et al.*, 2019).

1.2.1.1. Diseases of the pancreas: Chronic pancreatitis

Various diseases occur in the pancreas, one of them being chronic pancreatitis (CP), a disease associated with the exocrine pancreas that drastically reduces life expectancy and quality of life. Presently, there is no curative measure available. This condition is highly common worldwide, but there are notable variations in how it appears and is treated in different countries. These differences can be attributed to a range of factors, such as diverse causes, individual factors, the way symptoms manifest, and disparities in healthcare accessibility. Additionally, these variations may also stem from biases present in regional studies, including biases related to participant selection, recruitment methods, or the timing of diagnosis (Desai et al., 2022). Individuals with CP will exhibit symptoms such as pancreatic insufficiency, weight loss and steatorrhea (fat presence in the stool), and weight loss. Pancreatic insufficiency occurs when more than 90% of the pancreas is damaged (Benjamin and Lappin, 2023).

CP is an inflammatory condition that gives rise to permanent loss of endocrine and exocrine pancreatic parenchyma replacement and/or atrophy with fibrotic tissue (Benjamin and Lappin, 2022, Gandhi *et al*., 2022). In certain conditions or diseases, such as CP, the normal healthy pancreatic tissue can be progressively replaced by fibrotic tissue. Fibrosis is the formation of excessive connective tissue, characterized by the deposition of collagen and other fibrous proteins. As fibrosis progresses, it can lead to the atrophy (shrinkage) of the pancreatic parenchyma, which is the functional tissue of the organ. Additionally, if the endocrine function of the pancreas is impaired, it will eventually lead to a type of diabetes called pancreatogenic diabetes or Type 3c diabetes (Benjamin and Lappin, 2023). In CP, including tropical pancreatitis (a type of CP that predominantly affects individuals living in tropical regions, particularly in developing countries) and hereditary pancreatitis (a rare form of CP that is inherited in an autosomal dominant pattern), long-lasting inflammation causes an environment in the pancreatic tissue that can lead to carcinogenesis by elevating cell turnover and stellate cell proliferation. Additionally, previous studies have observed CP to be a risk factor for pancreatic cancer (PC), a condition characterized by the formation of cancerous cells in the pancreatic tissues. (Han *et al.*, 2022).

1.2.2. Pancreatic cancer types and PDAC development

Cancer of the pancreas can occur in the endocrine and exocrine glands of the pancreas (Litman-Zawadzka *et al.,* 2019) and there are various types of exocrine pancreatic cancers including squamous cell carcinoma, adenosquamous carcinoma, colloid carcinoma, and pancreatic ductal adenocarcinoma (PDAC). PDAC is the most common among pancreatic neoplasms, accounting for more than 90% of PC diagnoses (Kamisawa *et al.,* 2016). PDAC is a mucin-producing gland-forming neoplasm that causes excessive stromal desmoplastic reactions (Orth *et al.,* 2019). Cancer cells have the ability to both destroy the existing extracellular matrix and stimulate the production of new matrix through the activation of stromal cells. This response, known as the desmoplastic reaction, involves a dynamic remodeling process that includes both the creation and breakdown of the matrix (Ohtani, 1998). The acinar cells are characterized by their high degree of plasticity and are regarded to be the drivers of pancreas regeneration as well as homeostasis (Orth *et al.,* 2019). The acinar cells undergo lineage reprogramming into epithelial (duct-like) phenotypes in a process known as acinar-to-duct metaplasia (ADM) when there are specific micro and macroenvironmental stimuli such as tissue damage (Kopp *et al.,* 2012).

ADM enables acinar cells to acquire progenitor cell-like characteristics that make them vulnerable to pro-oncogenic hits, for example, the activation of mutations in the proto-oncogene *KRAS* (Orth *et al.,* 2019). The cells are transformed into pancreatic intraepithelial neoplasias (PanINs), a transformation considered the first step in the development of PDAC before sequential progression (Kanda *et al.,* 2012) (Figure 1.1). PanINs are known to be the precursors of PDAC and progress from low to high-grade tumours carrying the faulty *KRAS* gene (Aier *et al.,* 2019). Another group of tumour lesions found within the pancreatic duct is intraductal papillary mucinous neoplasms (IPMNs) which also act as cancer precursors and contain moderate amounts of the *KRAS* genes (Schmidt *et al.,* 2007).

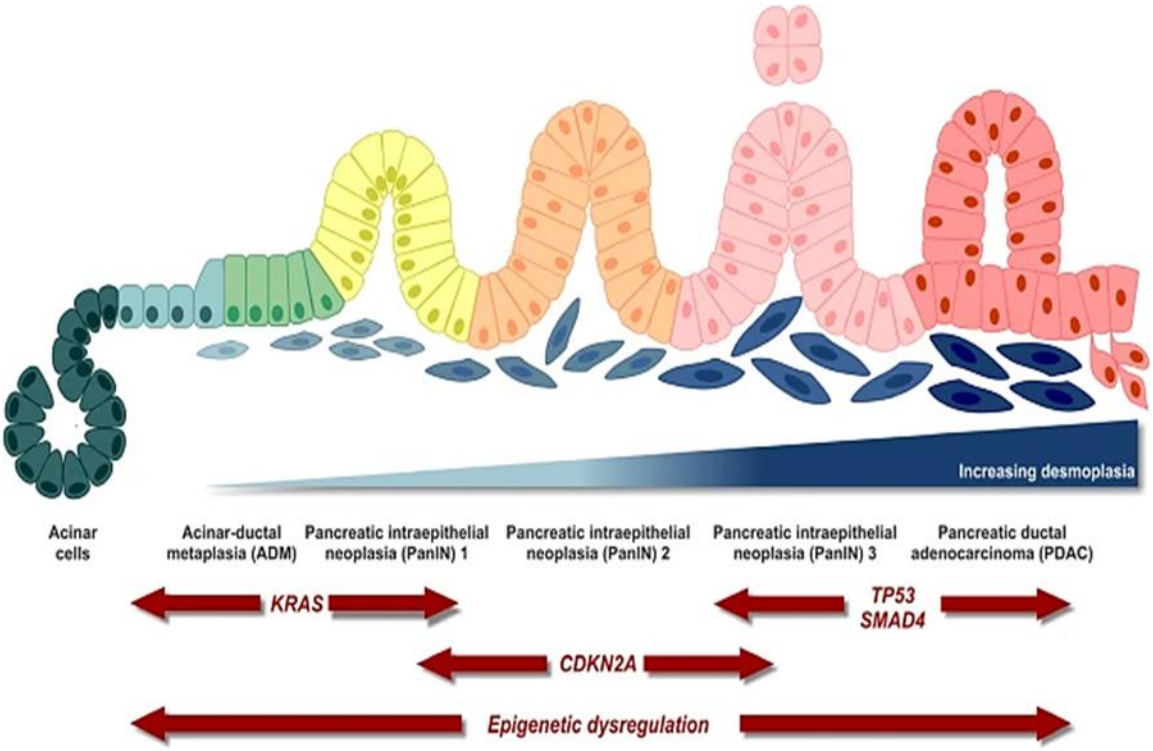


Figure 1. 1. Pancreatic ductal adenocarcinoma initiation and progression. The KRAS activity drives differentiated pancreatic cells such as acinar cells into a dedifferentiated, ductal state that progresses into pancreatic epithelial neoplasia (PanIN) and ultimately PDAC. Adapted from: (Orth et al., 2019) which is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

Studies involved in the investigation of PDAC's mutational and transcriptional landscape identified that the most frequently occurring mutations are found in *KRAS* gene(Orth *et al.*, 2019). Several other tumour suppressor genes such as mothers against decapentaplegic homologue 4 (*SMAD4*), tumour suppressor protein 53 (TP53) and cyclin-dependant kinase inhibitor 2A (*CDKN2A*) show repeated alterations in PDAC as well (Waddell *et al.*, 2015).

1.2.3. Risk factors that predispose individuals to PDAC

Several risk factors predispose individuals to PDAC. Familial PDAC occurs in approximately 10% of PDAC cases (Diaz and Lucas, 2019). Individuals with a family history of PDAC have an increased risk (6.79 times greater) of developing the disease and this risk increases in individuals younger than 50 years. (Brune *et al.*, 2010).

Genetic disorders such as hereditary pancreatitis are also risk factors for developing PDAC. Hereditary pancreatitis is linked with mutations in the serine peptidase inhibitor kazal type 1 (*SPINK1)* and serine protease 1 *(PRSS1)* genes (Ushio *et al.*, 2021). Individuals with hereditary pancreatitis have an increased risk (60-87 fold) of developing PDAC (Rebours *et*

al., 2008). In addition, Lynch syndrome, familial adenomatous polyposis, familial atypical multiple mole melanoma syndrome, and Peutz–Jeghers syndrome increase an individual's risk of developing PDAC (Ushio *et al.,* 2021).

Approximately 25-50% of PDAC patients would have already developed diabetes mellitus about 1-3 years before being diagnosed with PDAC (Becker *et al.,* 2014), with Type 3c diabetes mellitus occurring in 30% of PDAC patients (Cui and Andersen, 2012). Obesity has also been observed to promote the development of cancers through mechanisms including changes in gastrointestinal peptide secretion, autophagic processes, and gut microbiota (Park *et al.,* 2014, Allott and Hursting, 2015). Additionally, an increased waist circumference of 10 cm and body mass index (BMI) of 5 kg/m^2 increases the risk of PDAC onset by 1.11 and 1.10-fold, respectively (Ushio *et al.,* 2021).

Smoking cigarettes has also been reported to increase PDAC risk by 1.68-fold compared to non-smokers (Matsuo *et al.,* 2011). Also, large-scale studies have indicated that consuming extensive amounts of alcohol may increase the risk of developing pancreatic cancer by 60% in comparison to non-drinkers (Tramacere *et al.,* 2010, Aier *et al.,* 2019).

As mentioned above, there have been reports that CP is a risk factor for PDAC development (Ushio *et al.,* 2021), and CP patients were found to have a 13.3-fold greater risk of the PDAC disease in a study (meta-analysis) conducted by Malka and colleagues (2002) (Malka *et al.,* 2002). CP is a disease distinguished by progressive inflammation of the pancreas and scarring with parenchymal classifications.

1.2.4. Diagnosis of Pancreatic ductal adenocarcinoma

The poor statistics for the survival rate for pancreatic ductal adenocarcinoma (PDAC) are attributed to most patients being diagnosed at an advanced stage when the cancer has metastasized (Society, 2014). The early-state tumour is usually asymptomatic and there are no reliable biomarkers to aid with early detection (Elbanna *et al.,* 2020). Majority of individuals with PDAC present with symptoms such as abdominal pain, loss of appetite, jaundice, dark-coloured urine, light stool and itchy skin.

There are several methods used for the diagnosis of PDAC such as imaging methods which include ultrasound, magnetic resonance imaging (MRI), computerized tomography (CT) and positron emission tomography (PET). The advantage of using ultrasound is that it is cost-effective, widely accessible, and non-invasive (Winter *et al.,* 2019). However, it is not

recommended for PDAC detection and progress evaluation as it does not show small tumours of < 2 cm (Khorana *et al.,* 2017, Takaori et al., 2016). Additionally, it poorly visualizes the degree of local progression, the topography of changes, and tumour localization related to surrounding organs (Winter *et al.,* 2019).

Multiphase multi-detector row CT (MDCT) is considered to have the highest accuracy in determining distant metastases and resectability, the extent of the primary tumour, vascular invasion and locoregional extension (Kramer-Marek *et al.,* 2013). However, the CT scan is unable to reliably detect small lesions of < 1 cm, isoattenuating (a state in which a particular structure or tissue has a similar or nearly identical density as the surrounding tissues on imaging studies) primary tumours (Elbanna *et al.,* 2020) and differentiate between benign conditions and malignant lesions (Tummers *et al.,* 2018).

MRI has a high soft tissue contrast which makes it a preferred method for evaluating cystic lesions in the pancreas and is more precise in detecting distant metastases and enlarged lymph nodes compared to CT (Vachiranubhap *et al.,* 2009). A limitation of MRI is its inability to reliably distinguish necrotic or residual tumours from fibrosis and radiation changes after treatment (Tummers *et al.,* 2018). PET allows for whole-body imaging to enable the staging of diseases (Tummers *et al.,* 2018) and has a radiotracer which is a glucose analogue known as 18F-fluorodeoxyglucose (18F-FDG). As PDAC is associated with an overexpression of glucose transporter-1, it will result in an increased 18F-FDG uptake on PET-CT (Reske *et al.,* 1997). The limitation of using 18F-FDG uptake is false positive findings because glucose metabolism is not specific for malignant processes whereby physiological uptake can be found in inflammatory tissues and normal tissues (Yokoyama *et al.,* 2005).

Presently, circulating biomarkers are preferred due to their minimally invasive nature and ease of collection in clinical diagnoses(He and Yuan, 2014). The primarily employed serum marker, sialylated Lewis blood antigen CA19-9, is used to monitor responses to therapy and not as a diagnostic marker due to poor specificity and sensitivity (Chakraborty *et al.,* 2011). In addition, in patients with chronic pancreatitis, other gastrointestinal tract malignancies and benign diseases of the hepatobiliary system, CA19-9 can also be elevated (Xu *et al.,* 2011, Ulla Rocha et al., 2007). Another biomarker is the carcinoembryonic antigen (CEA), a serum biomarker that assists in diagnosing mucinous pancreatic tumours but does not provide accurate malignancy detection at early stages (Peng *et al.,* 2019). Both CA19-9 and CEA are biomarkers that are commonly used in clinical practice for certain medical conditions,

particularly in the context of cancer diagnosis, monitoring, and treatment. Therefore, due to a lack of reliable biomarkers, detecting PDAC early is difficult, creating an urgent need to identify more specific biomarkers (Meleady *et al.,* 2020).

1.2.5. Classification of PDAC

One of the most popular systems with criteria for the diagnosis and resection of PDAC is the National Comprehensive Cancer Network (NCCN). According to this system, PDAC is classified into the following categories (based on the tumour-vessel relationship): resectable, borderline resectable, locally advanced and metastatic (Kulkarni *et al.,* 2020). Resectable means the tumour may be removed by surgery. Borderline resectable is frequently described as a tumour that is restricted to the pancreatic bed with minimal involvement of adjacent vascular structures, allowing for the possibility of vascular restoration (Soweid, 2017). The borderline resectable category imaging tests are unable to consistently predict whether surgery is needed or not. Locally advanced is when cancer has spread to nearby organs, lymph nodes and tissues, and the chance of successful surgical resection is <5%. When cancer has spread to other parts of the body it is known as metastatic (Network, 2019). The main drivers of PDAC metastases formation are still insufficiently understood, mainly because the genetic make-up of most metastases closely resembles that of the corresponding primary tumour (Orth *et al.,* 2019).

1.2.6. Treatment of PDAC

For most PDAC patients, radical surgical resection is still the only possible curative therapy, but the recurrence rate can reach 85% (Digiacomo *et al.,* 2021). Whipple surgery, also known as pancreatoduodenectomy is the preferred method of choice for PDAC (resectable and borderline resectable) with mortality rates of <2% (Khachfe *et al.,* 2021). This complex surgical procedure involves the removal of the pancrease head, bile duct, gallbladder. Whipple surgery is a time-consuming and complex surgery with a high risk of complications like wound infections and postoperative pancreatic fistula (Shaeir *et al.,* 2022). Another form of treatment is radiation therapy where high-energy waves or particles are used to kill tumour cells.

Surgery or radiation therapy (RT) (a medical treatment modality that utilizes high-energy radiation to target and destroy abnormal cells or tumors within the body) are typically the main treatments for most cancers, with additional treatments like chemotherapy or hormone therapy sometimes used before or after the primary therapy. These additional treatments are

referred to as neoadjuvant therapy when given before the primary therapy, and adjuvant therapy when given after, as they serve to assist in the overall treatment plan (West and Jin, 2015). The ideal neoadjuvant/adjuvant treatment regimen for PDAC remains unclear, but typically involves folfirinox for neoadjuvant therapy and Gemcitabine-based regimens for adjuvant therapy in PDAC treatment (Principe *et al.,* 2021). Folfirinox is a combination therapy consisting of fluorouracil, leucovorin, oxaliplatin and irinotecan, used to kill or slow down the growth of cancer cells. Gemcitabine, also known as Gemzar or 2′, 2′-difluorodeoxycytidine, is a pyrimidine antimetabolite that functions as a specific analog of deoxycytidine. It is extensively utilized in the treatment of various solid tumors and neoplastic hematologic disorders (Xie et al., 2018). Gemcitabine is primarily metabolized within cells, and its major metabolite, gemcitabine triphosphate (dFdCTP), functions as a competitive substrate for deoxycytidine triphosphate (dCTP). This means that dFdCTP can be inserted into the growing DNA chain during replication, effectively impeding its elongation. As a consequence, the inhibition of DNA chain elongation leads to cell death through a process called apoptosis (Amrutkar and Gladhaug, 2017).

Several cancers are known to respond well to RT, however, the effectiveness of combining RT with chemotherapy for PDAC treatment is still an ongoing debate (Fujiwara *et al.,* 2020). Locally advanced PDAC is often treated with combined chemoradiation or combination chemotherapy which have shown limited improvement in survival thus far (Soweid, 2017). Despite some treatment modalities employed in the management of PDAC, they have not significantly improved patient outcomes and survival. Studies have shown that early detection of the disease may improve patient prognosis (He and Yuan, 2014). This reinforces the need for sensitive and specific biomarkers.

1.3. Biomarker discovery: The role of Proteomics

The use of proteomic technology for PDAC research (as well as other diseases) holds the potential to uncover early diagnostic biomarkers, track the progression of the disease, and facilitate the development of targeted therapies, complimenting existing genomic approaches (This includes techniques such as whole-genome sequencing, transcriptomics, and epigenomic (Meleady *et al.,* 2020). The proteome refers to the complete protein complement of cells, organs or organisms (Aslam *et al.,* 2017). Proteomics is the study of proteins including their function, expression, interactions, structure, localization and modifications (Domon and Aebersold, 2006). Additionally, it sheds light on the interplay between various

molecules that contribute to diseases, providing valuable insights that can help determine the underlying molecular mechanisms of diseases. (Meleady *et al.*, 2020).

The use of proteomics in cancer research has gained widespread popularity. By utilizing proteomics-based techniques, researchers can identify potential biomarkers and delineate patterns of protein expression that are linked to disease onset or progression (Vellan *et al.*, 2022). This information can be used to evaluate tumour prognosis, predict outcomes, classify tumours, and identify patients who may benefit from targeted therapies. Recent developments in proteomic technology have made it possible to conduct in-depth analysis at the cellular and molecular levels to better understand complex diseases like cancer (Vellan *et al.*, 2022). Human Epidermal Growth Factor Receptor 2 (HER2) is found to be overproduced in 15-30% of cases of invasive breast cancer, and this has important implications for both predicting the outcome and guiding treatment decisions (Burstein, 2005). In a study by Slamon *et al.*, 1987, the researchers demonstrated the prognostic value of HER2 amplification in 189 breast cancer patients. They discovered that an increase in the HER2 gene was a significant predictor for both overall survival ($P < 0.001$) and the time it took for the cancer to return ($P < 0.0001$) (Slamon et al., 1987).

Proteins are continually being demonstrated to have potential as biomarkers of PDAC. For example, studies (Li *et al.*, 2016, Lee *et al.*, 2017, Makawita *et al.*, 2013, Porterfield *et al.*, 2014) have previously identified regenerating family member alpha 1 (REG1A) and regenerating family member beta 1 (REG1B) from various biological samples such as serum, tissues, urine, and pancreatic ductal fluid as potential biomarkers (Li *et al.*, 2016, Lee et *al.*, 2017, Makawita *et al.*, 2013). In another study, the same group of researchers discovered that urinary REG1B levels (AUC: 0.93) exceeded in performance compared to REG1A (AUC: 0.90) in differentiating early-stage PDAC from chronic pancreatitis patients and healthy controls (Debernardi *et al.*, 2020). Nevertheless, to date, there have not been any newly developed biomarkers that have received clinical approval and demonstrated sufficient specificity and sensitivity for the detection of PDAC, and this study aims to identify proteins that could be potential biomarkers for the disease.

Another study employed normal phase liquid chromatography-mass spectrometry (NPLC-MS/MS) to identify alpha-1-acid glycoprotein as upregulated in PDAC tissues compared to normal pancreatic tissue, acquired from individuals with a benign pancreatic disease like mucinous cystadenoma, serous cystadenoma and pancreatic pseudotumor. The finding was

later confirmed by parallel reaction monitoring (PRM) (Zhou *et al.,* 2019). In an earlier study, Kim *et al.* (2014) identified fibrinogen as a possible biomarker in the serum of PDAC patients relative to healthy controls by use of matrix-assisted laser desorption/ionization-time of flight/mass spectrometry (MALDI-TOF/MS) (Kim *et al.,* 2014). The same protein was validated in PDAC patients, healthy controls and diabetic patients using ELISA and serum fibrinogen levels were found to be significantly elevated in PDAC patients relative to the healthy controls (Chung *et al.,* 2020).

The Clinical Proteomic Tumour Analysis Consortium (CPTAC) data portal provides access to key findings from in-depth proteomics studies of several tumours (Cao *et al.,* 2021). CPTAC has played a significant role in advancing proteogenomic studies aimed at promoting an understanding of the links between cancer genotype and phenotype, molecular pathogenesis, and response to treatment (Boys *et al.,* 2022). An in-depth analysis of 105 treatment-naïve, surgically resected PDAC tissue samples, including 67 paired normal adjacent tissues and nine normal pancreatic duct tissues revealed novel insights into pancreatic ductal adenocarcinoma. Following stromal and immune content correction, a total of 21 proteins were described as potential early detection and diagnostic targets. Specifically, out of these proteins, 12 are secreted, and they have the potential to function as markers for early detection in serum or pancreatic juice (González-Borja et al., 2019). Lysyl oxidase homolog 2 (LOXL2) abundance was associated with decreased survival ($P = 0.041$). It has been reported that thrombospondin 2 and laminin subunit gamma-2 levels were increased in the bloodstream of patients diagnosed with PDAC (Kim et al., 2017b). According to previous data from the Pancreatic Cancer Database, hexokinase 2, LOXL2, collagen type XII alpha 1 chain, chromosome 19 open reading frame 33, tetraspanin 1, and midkine were identified as being elevated. These findings were previously supported solely by evidence from RNA or proteomic analysis conducted on cell lines (Thomas et al., 2014).

Additionally, 14 of the 21 proteins were further validated via data-independent acquisition-mass spectrometry (DIA-MS). The researchers further conducted expression analysis of glycoproteins in tumours with different *KRAS* mutations and showed upregulation of both carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5) and carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6) in tumours with G12D, G12V, and Q61H *KRAS* mutations, which were suggested to possibly serve as potential targets for treatment (Cao *et al.,* 2021). In this study cohort, there were mainly individuals from Russia (7.9%), China (15%), Canada (13.6%), Poland (32.1%) and the

United states (20%) (Cao et al., 2021). Therefore, emphasizing the need for studies including patients of African ancestry.

In a previous study conducted in our laboratory, comparing proteomic profiles of tumours and normal tissue from PDAC patients of African ancestry using sequential window acquisition of all theoretical mass spectra (SWATH-MS), several significantly dysregulated proteins related to PDAC were identified (Nweke *et al.,* 2020). SWATH-MS is an advanced proteomics method that merges the advantages of data-independent acquisition (DIA) mass spectrometry with the precise identification and quantification of peptides. It operates by systematically collecting high-resolution mass spectrometry data across multiple mass-to-charge ratio (m/z) windows, encompassing the complete desired mass range. Most of these identified proteins (such as plectin, alpha-actinin-1, LIM and senescent cell antigen-like-containing domain protein 1, and integrin-linked kinase) are associated with extracellular matrix formation and cell proliferation, which play a major role in tumour progression (Nweke *et al.,* 2020).

Using tissue samples for biomarker discovery has its pros and cons. A primary advantage is that the biomarker will be at its highest concentration within the tissue (Blonder *et al.,* 2011). However, alternatiely blood is minimally invasive to collect and easily accessible (Peng *et al.,* 2019), making blood plasma a favourable biofluid for biomarker discovery studies. Proteome information from tissue and plasma samples may be linked as many proteins such as those involved in tumour progression are secreted from tumour cells and may be measured in the bloodstream. The secretome consists of proteins secreted by cells that are transiently expressed during organ development, response to different pathologies, maturation and environmental stimuli. Secreted proteins play critical roles in cell-cell signalling, growth and communication as well as provide insight into various pathological condition stages. They are an important source for discovering novel biomarkers (Stastna and Van Eyk, 2012).

1.4. Mass-spectrometry technologies

Mass spectrometry (MS)-based proteomics is an important technique in investigating alterations in the proteome (Peng *et al.,* 2019). The sensitivity, specificity and accuracy of protein-based diagnostics are strongly associated with the advancement of high-performance MS instruments for the measurements of analytes in clinical research (Ansari *et al.,* 2014). MS is used to identify the presence or abundance of proteins using their basic properties such as mass and net charge ratio (Sinha and Mann, 2020).

MS is a type of analytical method that allows for both structural and quantitative information to be obtained from analyte molecules. The process begins by introducing the molecules of interest into the instrument's ionization source, to produce positive or negative gas phase ions (Ho *et al.*, 2003). The resulting ions then move through the mass analyzer, with their mass/charge ratio determining their mobility. Upon reaching the detector, signals are generated and recorded by a computer system, which deconvolutes the signal to produce a mass spectrum, representing the intensity of ion signals as a function of their mass/charge ratio (Ho *et al.*, 2003).

Electrospray ionisation (ESI) is the most popular method of ionization used in proteomics analysis. Conventional ESI involves using a hollow tube-like structure, known as a capillary or spray needle, introducing the sample into the mass spectrometer under the influence of a high voltage and a flow of gas (Hu and Yao, 2022). A liquid that can undergo ionisation is passed through a spray needle that is maintained at a high voltage with a counter electrode. The voltage differential between the needle and the counter electrode is strong enough to overcome the surface tension of the liquid, resulting in the formation of a Taylor cone, which eventually releases charged droplets (Figure 1.2). These droplets are then subjected to further acceleration for entry into the mass analyzer(s) (Patrick, 2020).

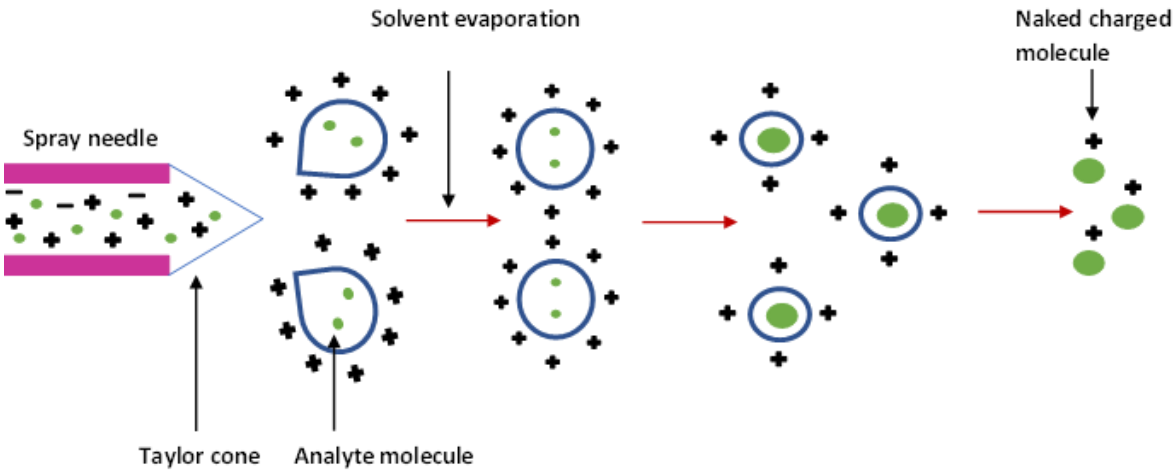


Figure 1. 2. Schematic representation of the electrospray ionization process. The sample is fed into the ionisation source at a specific flow rate. The analyte is then passed through an electrospray needle, which is charged with a high voltage. As the droplets travel from the needle tip to the cone, the solvent begins to evaporate, causing the droplets to shrink. When the surface tension can no longer support the charge, a phenomenon called a "Coulombic explosion" occurs, breaking apart the droplet into smaller droplets and naked charged analyte molecules. These smaller droplets can undergo the same process again.

One way to obtain sequence information of molecules involves isolating the precursor ions of interest through mass selection and further fragmenting them in a collision cell. Following this, the resulting fragment ions can be subjected to mass analysis using a second mass analyzer in a tandem mass spectrometer system (Ho *et al.,* 2003). A common tandem quadrupole system involves three quadrupoles arranged in a linear fashion, often referred to as a "triple-quad". The first quadrupole (Q1) is used to selectively isolate the precursor ion, based on its mass. Subsequently, the precursor ions are made to collide with a collision gas, in a second quadrupole collision cell (Q2) that uses only radio frequency (RF) waves. This collision activation process, also known as collision-induced dissociation (CID), results in further fragmentation of the precursor ions. The resulting daughter/fragment ions are indicative of the molecular structure of the precursor ions and can be analyzed using the third quadrupole mass analyzer (Q3) to provide structural information. In other systems, the Q3 is replaced with a more high-resolution mass analyzer such as a time-of-flight (TOF) device (often referred to as a “QTOF”), allowing for greater mass accuracy and proteome coverage. This approach of precursor selection followed by fragmentation and fragment ion analysis is typically referred to as MS/MS in literature (Ho *et al.,* 2003).

The development of the sequential window acquisition of all theoretical mass spectra (SWATH) has led to the reliable and precise label-free measurement of proteomes (Gillet *et al.,* 2012). The SWATH-MS technique involves a data-independent acquisition method, followed by an innovative targeted data extraction technique (Chen *et al.,* 2021). In summary, the method obtains full scan maps of peptide precursor ions without prior selection (MS1 scan). This is followed by MS2 scans with specific m/z isolation windows, that record the chromatographic elution traces of all detectable fragment ions (Figure 1.3). This analysis method utilises MS2-based quantification of the fragment ion data based on their intensities, providing consistent information for protein quantification (Chen *et al.,* 2021). SWATH-MS and has the added advantage of monitoring many more analytes in an unbiased manner. For SWATH-MS measurements as with other bottom-up proteomics approaches, trypsin is commonly used to digest non-labelled proteins and the peptides that result, are analyzed via liquid chromatography, typically reverse-phase, connected to a tandem mass spectrometer (typically a QTOF) running in data-independent acquisition mode.

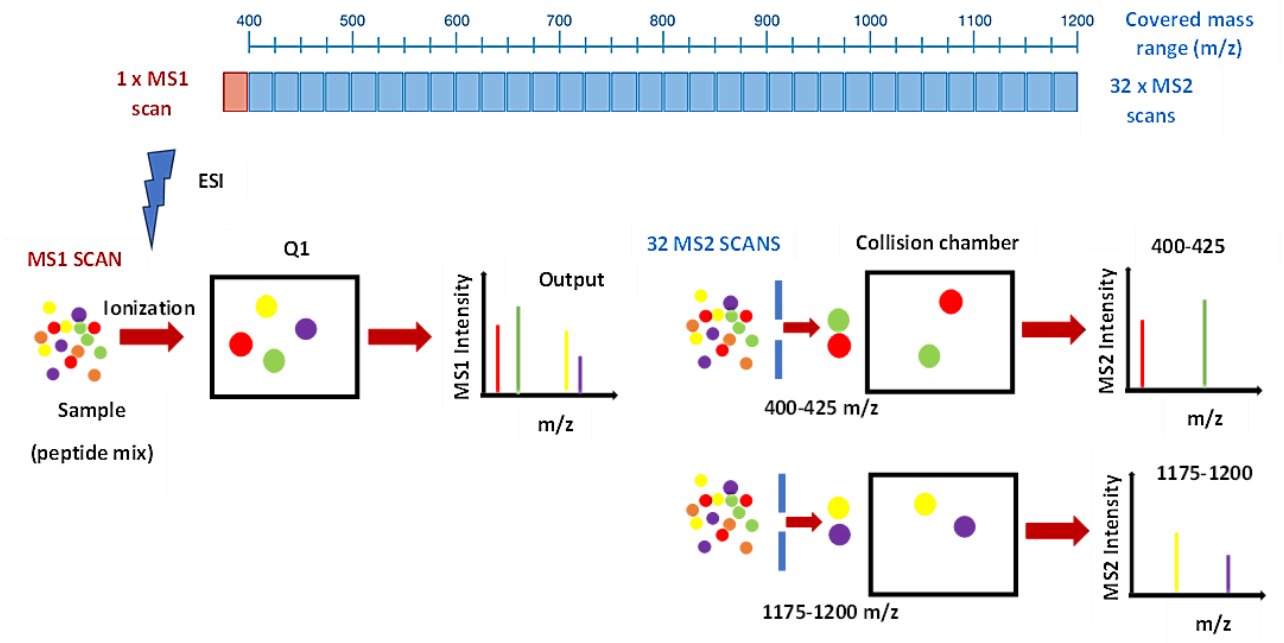


Figure 1. 3. Schematic diagram of SWATH-MS principle. The MS1 scan is comprehensive and detects all precursor ions without any initial selection. This is followed by 32 MS2 scans, for example, each with a defined increase of 25 m/z, starting from 400 m/z to 1200 m/z, adapted from (Chen et al., 2021).

1.5. Plasma proteomics, depletion, fractionation

Mass spectrometry-based plasma proteomics can provide solutions to many challenges such as early disease detection which depend entirely on more specific biomarkers (Ignjatovic *et al.*, 2019). One of the most accessible biofluids is plasma which has the potential to improve therapeutic monitoring and disease diagnosis.

Although plasma is easily accessible, they are often variably collected and processed during daily clinical routines. Both collection and processing have the potential to influence the outcomes of clinical studies, especially regarding biomarker identification (Geyer *et al.*, 2019). During the process of sample collection, processing and preparation for downstream analysis, a lack of standardized techniques can lead to contamination and inaccuracies in the results obtained (Geyer *et al.*, 2019). Furthermore, due to the prominence of high-abundance proteins in plasma in comparison to other clinical specimens, there are additional protocol considerations for in-depth biomarker studies (Kim et al., 2017a).

Plasma is a complex sample source and dynamic in range, whereby the concentrations of the least and most abundant proteins vary by approximately 10 orders of magnitude (Lee *et al.*, 2019). Prefractionation methods have been utilised in efforts to uncover the low-abundant proteins masked by highly-abundant ones through the reduction of protein complexity per

fraction (Kaur *et al.,* 2021). The detection of low-abundant proteins can also be improved through the depletion of highly-abundant proteins (Gianazza *et al.,* 2016). One of the most abundant plasma proteins in humans, constituting approximately 50% of all plasma proteins is albumin (Campos Munoz *et al.,* 2022).

Immunodepleting of highly-abundant proteins is a widely used sample preparation method that is applied to biofluids with a high dynamic range (e.g., plasma) (Garcia *et al.,* 2017). This process uses specific antibodies and their protein targets for the selective capture and removal of highly-abundant proteins. The majority of immunodepletion kits are currently available in column format wherein specific antibodies are bound onto affinity resins to retain high-abundant proteins while retrieving lower-abundant proteins for proteome analysis (Lee *et al.,* 2019). An alternative method to depletion for sampling low abundant proteins in the plasma proteome and increasing protein coverage is ion exchange chromatography. Proteins are fractionated based on their hydrophobicity and charge using ion exchange chromatography, with the elution rate being influenced by the salt concentration or pH of the mobile phase (Chutipongtanate *et al.,* 2017). Ion-exchange chromatography offers the advantage of elution occurring under moderate conditions, which allows the protein or peptide to retain its native structure. In addition, the cost is much lower compared to antibody-based methods (Zhu *et al.,* 2012).

1.6. Method standardisation for biomarker discovery

The preparation of plasma samples for protein quantification by liquid chromatography-MS/MS requires the following steps for processing: protein dilution, reduction, alkylation, enzymatic digestion and peptide clean-up (Krüger *et al.,* 2013, Dittrich *et al.,* 2015). Sample preparation, in particular, protease digestion is a major source of variability in LC-MS/MS, therefore it is vital to have a standardized and highly controlled digestion method that will meet the reproducibility and precision standards that are a requirement for reliably identifying and verifying biomarkers (Fu *et al.,* 2018). For improved accuracy and precision, each of the sample preparation steps must be conducted at a controlled temperature, start and end at a consistent time, and have accurate liquid transfers and good mixing for uniform reactions (Zhang and Rockwood, 2015).

Common sample preparation methods revolve around eliminating detergents that aid in extraction but hinder proteolytic digestion and LC-MS. However, multi-step preparations that

remove detergents often result in decreased reproducibility due to potential losses, biases, and contaminations, as well as requiring extensive time and labour (Doellinger *et al.,* 2020).

To establish a more robust pathway from discovering the biomarkers to validation, developing a simple and highly reproducible sample preparation workflow is important (Fu *et al.,* 2018).

1.7. Justification of study

With the increase in several high-throughput technologies for biomarker screening in recent years, there has been a plethora of potential novel biomarkers for the diagnosis of PDAC. However, there are still no new clinically approved ones with adequate specificity and sensitivity for the detection of PDAC. The application of MS-based plasma proteomics has the potential to address various challenges, including the identification of more precise biomarkers, leading to early detection and diagnosis of disease. However, the shortfall of standardized methods for preparing and processing samples may impede the identification of biomarkers. Additionally, a study by Nweke *et al.*,(2020) identified a number of differentially expressed proteins (potential biomarkers) from resectable African PDAC patient tissue samples. However, the collection of tissue is highly invasive and not easily accessible compared to blood. Therefore, the goal of this study was to develop a standardized method by assessing three parameters (ease of throughput, reproducibility and peptide identification) for plasma sample preparation. Importantly, this study uniquely utilizes samples from PDAC patients of African ancestry. This work will aid in further studies relating to the discovery of relevant biomarkers and understanding the disease progression in this group.

1.8. Aim

The study aims to develop a standard operating procedure (SOP), for the preparation of plasma samples to be used for discovery proteomics studies.

1.9. Objectives

1. To compare and select the optimal workflow for sample preparation clean up and protein digestion in plasma samples, by evaluating ease of throughput, reproducibility, and ability to detect more distinct peptides.
2. To generate an in-depth human plasma spectral library for the identification and quantification of proteins.
3. To formulate a standard operating procedure (SOP) for plasma sample preparation.

4. To implement the SOP on patient samples in a pilot study to determine its utility in profiling plasma proteins to identify key PDAC targets.

CHAPTER 2

METHODS AND MATERIALS

2.1. Experimental strategy

The experimental procedure included sample collection, sample preparation, selection of clean-up and digest method, comparison of depletion methods, building a spectral library, method optimization and a pilot study (Figure 2.1). Blood samples were collected from patients who provided written informed consent. SDS-PAGE was run to visualize the dynamic range (wide range of concentration of different proteins) of the plasma sample. Dilution and precipitation methods (25X dilution, 5X TEAB dilution and acetone precipitation) were compared by assessing reproducibility, ease of throughput and the number of peptides identified. Samples from each dilution and precipitation method were subjected to reduction, alkylation, clean-up and digestion using the hydrophilic interaction liquid chromatography (HILIC) method compared to protein aggregation capture (PAC) by evaluating reproducibility and the ability to detect a high number of peptides. Low-pH reverse-phase chromatography was applied to separate the peptides. SWATH-MS analysis was conducted to record all fragment ions of the detectable peptide precursors present in the samples. The proteins were identified by searching raw data against the reviewed human UniProt sequence database (30/07/2021) and a spectral library generated using the Pulsar search algorithm, and the dysregulated proteins were identified using Spectronaut version 14.

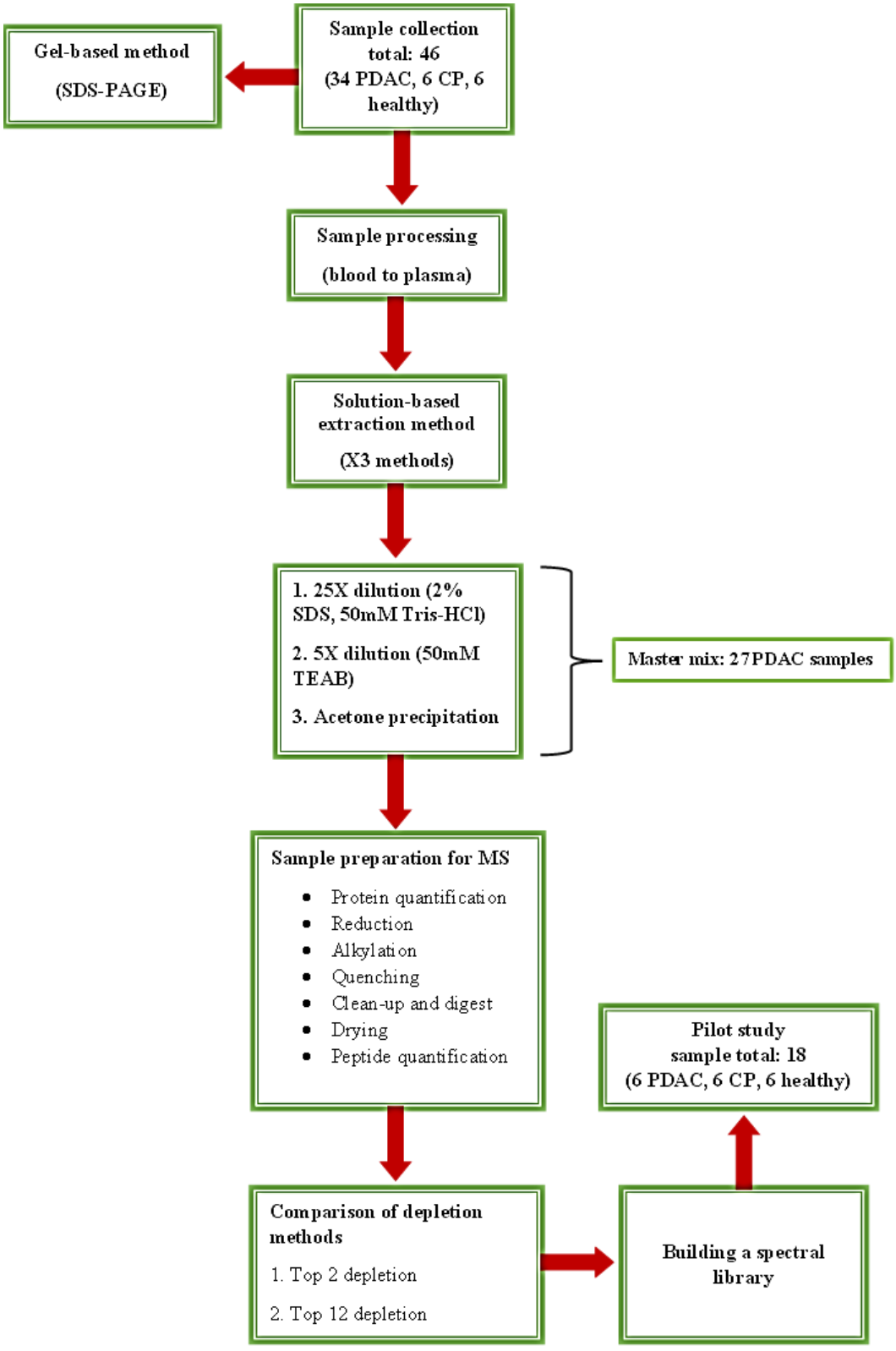


Figure 2. 1. Schematic overview of experimental strategy. The order of methods used are indicated by the arrows: sample collection from PDAC patients in the form of blood followed by sample preparation including preparation of plasma samples from blood; gel and solution-based protein extraction and preparing samples for MS analysis, comparison of depletion methods to remove highly-abundant proteins, building a spectral library, and pilot study to test the optimised methods.

2.2. Solutions, Buffers and Solvents

The solutions used in the experiments were prepared using ultra-purified water (Millipore Corporation, Billerica, MA, USA). All solvents and buffers used for MS analysis were prepared by use of MS-grade acetonitrile (ACN), formic acid and water (Table S1).

2.3. Sample collection and processing

A total of 46 blood samples (34 PDAC, 6 chronic pancreatitis and 6 healthy individuals) had been collected from consented individuals Chris Hani Baragwanath in Johannesburg South Africa. Written informed consent was obtained from individuals before sample collection. All patients were 18 years and older and of African ancestry. The diagnosis of PDAC patients recruited was clinically and histologically confirmed. From each patient, demographic and clinical information was collected and stored in an already-established REDCAP version 11.0.3 (Harris et al., 2009). The mean age of all recruited patients was 54 years (+/-13.4), and 74% of the patients were male. For PDAC patients, 22 had resectable pancreatic cancer (RPC), 8 had locally advanced pancreatic cancer (LAPC) and 4 had metastatic pancreatic cancer (MPC) (Table S2). Blood samples of 10 ml were collected by venipuncture in an EDTA Vacutainer tube, and the plasma separated within 2 hrs of blood collection which was then frozen at –80 °C. For plasma separation, the blood was allowed to stand for 60 mins and then centrifuged for 30 mins at 3000 rpm at 4 °C. The plasma samples were aliquoted (300-500 µl) into 0.5/1.5 ml protein Lobind tubes and stored by freezing at –80 °C to maintain sample integrity. A master mix of 27 PDAC plasma samples (the excluded 7 samples either had an unusual morphology or there was too little of the sample) was used for method optimization and 18 individual plasma samples (6 PDAC, 6 CP, 6 healthy) were used for the pilot. The Human Research Ethics committee of the University of Witwatersrand approved the study before any patient recruitment and sample collection (M211158).

2.4. Sample preparation

2.4.1. Sodium dodecyl-sulfate polyacrylamide gel electrophoresis

Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was conducted to visualize the dynamic range of plasma proteins. SDS-PAGE was conducted using a Invitrogen Bolt 4-12% Bis-Tris plus gels (Thermo Fisher, Massachusetts, USA) according to the manufacturer's instructions. Briefly, the plasma sample was pre-diluted 5-fold [10 µl of plasma and 40 µl of buffer (2% SDS, 50 mM Tris-HCl pH 8)]. From the prediluted sample,

10 µl of the sample was added to 5 µl of sample buffer (Bolt LDS Sample Buffer 4X) (Thermo Fisher Scientific) and 2 µl of reducing agent (Bolt/NuPage Reducing agent 10X, containing 500 mM DTT). The sample was then topped up with 3 µl of deionized water to a final volume of 20 µl, vortexed for 30 secs and boiled at 70 °C for 10 mins. The running buffer used was the NuPAGE MES SDS running buffer (diluted from 20X stock solution) pH 7.3 (Thermo Fisher Scientific). The gel was stained with colloidal Coomassie blue (0.1% Coomassie G-250, 20% ethanol and 3% ortho-phosphoric acid) and de-stained with distilled H_2O to visualize the proteins.

2.4.2. Strong Anion Exchange (SAX)

To improve protein coverage and enhance the spectral library, a fractionation technique (SAX) was used according to the manufacturer's instruction (ReSyn Biosciences, Pretoria, SA), and samples were evaluated on an SDS-PAGE.

Protein binding

The protein samples (containing either 0.25 mg or 0.5 mg total protein) were added to the equilibrated MagReSyn® SAX microparticles. The reaction volume was adjusted to a minimum of 250 µl with binding buffer and mixed thoroughly by vortexing for 3 secs. The proteins were allowed to bind to the microparticles for 5-10 mins at room temperature, with mixing to allow for good sample-microparticle interaction. The tubes were then placed on a magnetic separator, allowing the microparticles to clear. The supernatant was removed by pipette aspiration. The microparticles were washed by adding 250 µl of binding buffer (50 mM Tris, pH 8) and resuspended by vortexing for 3 secs. The microparticles were recovered on the magnetic separator. The magnet was inverted with the tubes still in place for the collection of microparticles that may have been stuck to the tube lid. The supernatant was then removed with a pipette. This was repeated twice, making a total of three washes.

Fractionation by increasing salt concentration

The microparticles with adsorbed protein mixture from the protein binding procedure were resuspended in 50 µl elution buffer (50 mM Tris pH 8, 50 mM NaCl) and desorption was allowed for 2 mins at room temperature. The microparticles were then recovered on the magnetic separator and the supernatant was aspirated with a pipette. The supernatant was kept for further analysis. this was repeated twice for better recovery of the fractionated proteins. The whole process was repeated using elution buffers consisting of increasing salt

concentrations (50 mM, 100 mM, 200 mM, 300 mM, 400 mM, and 2M). The eluted samples containing the fractionated proteins were analyzed on an SDS-PAGE gel as per the manufacturer's instructions. The fractionated plasma samples were further processed using the HILIC method for LC-MS analysis and library building.

2.4.3. Protein Depletion

Protein depletion was conducted using aPierce™ Albumin/IgG Removal kit and a Pierce™ Top 12 Abundant Protein Depletion Spin Columns according to the manufacturer's instructions (Thermo Scientific, Massachusetts, USA) to deplete highly abundant proteins.

Pierce Albumin/IgG Removal

Resin preparation: A spin column was placed into a microcentrifuge tube for each sample to be processed. The bottle with the immobilized Cibacron Blue/Protein A gel was swirled for an even suspension. The gel slurry (170 µl) was then added to the spin column. The column was placed in a microcentrifuge tube and centrifuged at 10 000 x *g* for 1 min, and the storage buffer was discarded. A plug was inserted at the bottom of the column containing the gel bed.

Sample processing: A plasma sample (10 µl) was diluted to 75 µl with the binding/wash buffer and added to the gel bed. The column was briefly vortexed forming a suspension and placed back into the microcentrifuge tube. The column containing the plasma sample was incubated at room temperature for 10 mins on an orbital shaker. The column cap was loosened, and the bottom plug was removed, the column was then centrifuged again at 10 000 x *g* for 1 min. The filtrate was recovered, and the resin bed was re-applied. The column was incubated at room temperature for 10 mins on an orbital shaker, the bottom plug and cap were then removed, and the column was centrifuged at 10 000 x *g* for 1 min. Binding/wash buffer (75 µl) was added to the gel bed and centrifuged at 10 000 x *g* for 1 min and collected wash buffer in a separate microcentrifuge tube.

Pierce Top 12 Abundant Protein Depletion Spin Column

The depletion spin column was equilibrated to room temperature, and 10 µl of the plasma samples was added to the column containing the resin slurry. The column was closed and inverted several times until the resin was suspended in the solution. The mixture was then incubated at room temperature for 60 mins. with gentle end-over-end mixing. The column was then placed in a 2 ml collection tube and centrifuged at 10 000 x *g* for 2 mins, and the

resin was discarded. The eluate was processed further using the HILIC method for LC-MS analysis and subsequent library building.

2.4.4. Solution-based method optimization

The plasma protein samples were quantified using the Pierce Bicinchoninic Acid Protein Assay kit (Thermo Fisher Scientific, Massachusetts, USA). Plasma sample dilution and precipitation methods were compared. Method 1: plasma diluted 25X with buffer (50 mM Tris-HCl pH 8 with 2% SDS), method 2: plasma diluted 5X with 50 mM TEAB, 2% SDS and method 3: acetone precipitation. For acetone precipitation, 1600 µl of ice-cold acetone was added to 50 µl of plasma sample. The plasma samples were then centrifuged at 12 000 x *g* for 30 mins at 4 °C. The supernatant was removed, and the pellet was dried for 1 min at 70 °C. Once the pellet had dried, 50 µl of 50 mM Tris-HCl, 2% SDS was added to it and sonicated for 5 mins. All methods were conducted over two different days.

Plasma samples from each method were incubated with 10 mM dithiothreitol (DTT), to reduce disulfide bonds in proteins formed by oxidation of sulfyhydryl groups on cysteine, at 60 °C for 30 mins and alkylated (to stabilize free sulfyhydryl groups and improve peptide yield) in the dark with 40 mM iodoacetamide (IAA) at room temperature for 30 mins. Alkylation was quenched with 10 mM of DTT. For protein aggregation capture (PAC), the magnetic beads were initially pre-equilibrated with 70% ACN twice for 30 secs at 25 RPM by mixing using a ELMI intelli-mixer RM-2L (ELMI, Riga, Latvia). The proteins were then bound to the MagReSyn® HILIC magnetic beads (Resyn Biosciences, Edenvale, SA) and digested with trypsin (1:20; enzyme:protein) at 47 °C for 4 hrs in a 96-deep well plate using a KingFisher™ Duo magnetic handling station (Thermo Fisher Scientific).

For the HILIC method, extracted proteins were bound to magnetic hydrophilic interaction liquid chromatography (HILIC) beads (ReSyn Biosciences, Edenvale, SA) and digested with trypsin (1:20; enzyme:protein) and LysC (1:100; enzyme:protein) at 47 °C for 4 hrs. Protein digestion was subsequently performed in a 96 deep-well plate (Figure 2.2), using a KingFisher™ Duo magnetic handling station (Thermo Fisher Scientific). However, for the pilot study, the KingFisher™ FLEX magnetic handling station (Thermo Fisher Scientific) was used (Figure 2.3). After digestion, the beads were removed, and the peptide samples from each clean-up method were transferred to 2 ml LoBind tubes. The samples were then dried overnight using the CentriVap speed vacuum dryer to remove the buffer through sublimation while still preserving the chemical and biological structure of the peptides. The

peptides were quantified by using a Pierce Quantitative Colorimetric Peptide Assay Kit (Thermo Fisher Scientific).

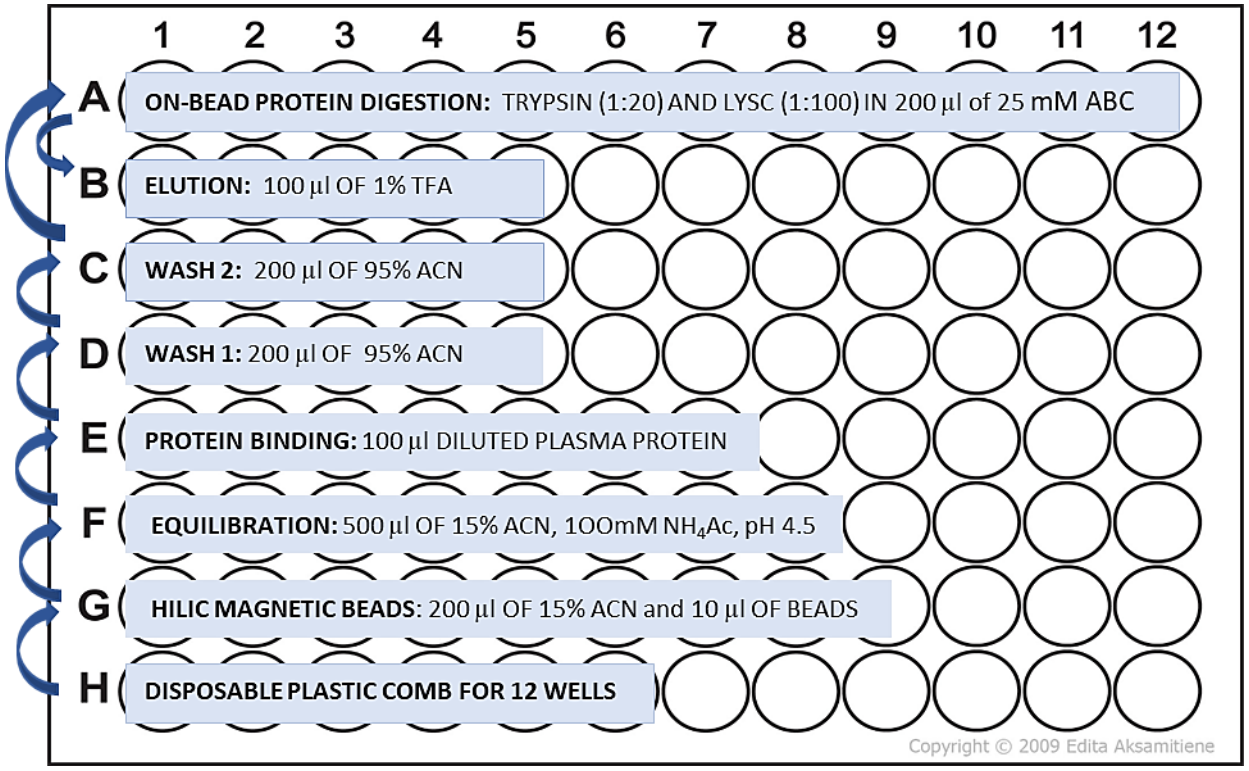


Figure 2. 2. Schematic representation of hydrophilic interaction liquid chromatography (HILIC)-based clean-up and digestion on an automated KingFisher™ Duo magnetic handling station (Thermo Scientific). 1. disposable comb collection (row H) 2. HILIC bead collection (row G), 3. Equilibration (row F) 4. Protein binding (row E) 5. First wash (row D) 6. Second wash (row C) 7. On-bead digestion with trypsin and LysC (row A) 8. Elution (row B).

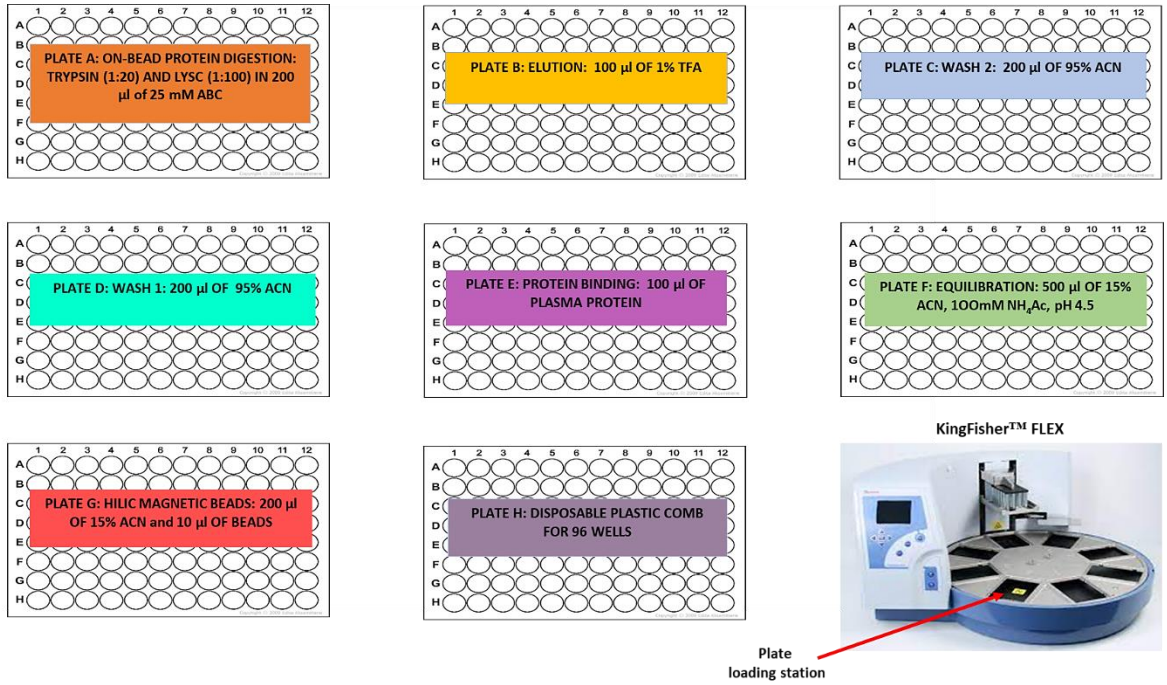


Figure 2. 3. Schematic representation of hydrophilic interaction liquid chromatography **(HILIC)-based clean-up and digestion on an automated KingFisher™ FLEX magnetic handling station (Thermo Scientific)**. For the KingFisher FLEX, 8 plates are prepared and loaded into the loading stations for HILIC-based sample preparation. Disposable plastic comb collection (plate H), HILIC microparticle collection (late G), HILIC microparticle equilibration (plate F), protein binding (plate E), wash 1 and wash 2 in plate D and plate C respectively, on-bead protein digestion (plate A) and elution (plate B).

2.5. Building the spectral library

The library was generated using Spectronaut Pulsar, raw data (.wiff files) from method optimization experiments (experiments 2-4, depletion and SAX fractionation) were searched against the reviewed human UniProt sequence database (30/07/2021) and a list of common contaminating proteins. Biognosys generic format default factory settings were applied. The QUANT 2.0 (SN standard) LQF method was selected, and no decoys were selected. Trypsin/P and LysC were chosen as the digestive enzymes. The search criteria also included carbamidomethylation of cysteine as a fixed modification.

2.6. Data acquisition for the spectral library building

Method optimization experiments, depleted and fractionated peptide samples were vacuum dried and re-suspended in 2% acetonitrile/0.2% formic acid. Analysis was conducted on a Dionex Ultimate 3000 RSLC system (Thermo Fisher Scientific) coupled to an AB Sciex 6600 TripleTOF mass spectrometer (MS) (Sciex, Massachusetts, USA). The online de-salting (removes excess salts and trace amounts of detergents which can interfere with peptide

ionization and cause chemical noise) of peptides was conducted on an Acclaim PepMap C18 trap column (75 µm × 2 cm). Separation was performed on the Acclain PepMap C18 nanoRSLC column (75 µm × 15 cm, 2 µm particle size). C18 columns are used with more polar solvents such as methanol, water and acetonitrile. The mobile phase is a polar liquid while the stationary phase is a non-polar hydrocarbon. Peptide elution was achieved by using a flowrate of 0.3 µl/min with a gradient: 6-35% solvent B in 60 min (Solvent A: MS grade water with 0.1% formic acid, Solvent B: 80% acetonitrile with 0.1% formic acid). The 6600 TripleTOF MS was operated in data-independent acquisition mode. SWATH-MS consisted of the acquisition of 60 MS2 scans of overlapping sequential precursor isolation windows (1 m/z overlap, variable m/z isolation width, high sensitivity mode) covering the mass range 400 to 900 m/z, and a single MS1 scan. For a total of 1 sec total cycle time, the accumulation time for the MS1 and MS2 scans was 100 ms and 15 ms, respectively.

2.7. Evotip loading and data acquisition for the pilot study

Peptide samples were loaded on evotips as per the manufacturer's instructions (Evosep, Biosystems). Solvent A (MS grade water with 0.1% formic acid), solvent B (MS grade acetonitrile with 0.1% formic acid) and MS grade 1-propanol were poured into their respective containers. A volume of 100 µl of 1-propanol was pipetted into each well of a 96-well plate MTP using a multichannel pipette. Solvent B (20 µl) was transferred to all evotips and placed in a centrifuge at 700 x *g* for 60 secs. The evotip adaptor rack was then placed on the MTP containing 1-propanol and soaked for 10 secs until the tips were pale white. Solvent A (20 µl) was transferred to each evotip and centrifuged at 700 x *g* for 60 secs. The samples (20 µl) were then added to each evotip centrifuged at 700 x *g* for 60 secs. Following sample loading, 20 µl of solvent A was added to each evotip and centrifuged at 700 x *g* for 60 secs. Lastly, 100 µl of solvent A was transferred to each evotip and centrifuged at 700 x *g* for 10 secs to keep the tips wet until MS analysis.

Analysis was conducted on the Evosep One LC system (Evosep Biosystems, Denmark) coupled to an AB Sciex 6600 TripleTOF mass spectrometer (MS) (Sciex). The online de-salting of peptides was conducted using a 60 samples per day Evosep (Evosep Biosystems) with a gradient length of 21 mins. The 6600 TripleTOF MS was operated in data-independent acquisition mode. SWATH-MS consisted of the acquisition of 60 MS2 scans of overlapping sequential precursor isolation windows (1 m/z overlap, variable m/z isolation width, high sensitivity mode) covering the mass range 400 to 900 m/z, and a single MS1 scan. For a total

of 1 sec total cycle time, the accumulation time for the MS1 and MS2 scans was 100 ms and 15 ms, respectively.

2.8. Pathway analysis and network interaction

The pathway analysis of differentially expressed proteins was conducted on Reactome version 82 (Croft et al., 2014). A cut-off value of 0.05 was used and the Benjamini-Hochberg correction test was applied for the adjusted p-value. A gene ontology biological process pathway analysis was also conducted on ShinyGO version 0.77 (Ge et al., 2020) at a false discovery rate (FDR) cut-off of 0.05. To show network interaction between the dysregulated proteins, the StringApp within Cytoscape version 3.9.1 (Doncheva et al., 2019, Shannon et al., 2003) was used with a confidence score of 0.7 as cut-off.

2.9. Statistical analysis

The distribution of the data was determined using the Shapiro-Wilks test. For the normally distributed data, with only two groups an unpaired t-test was used while for more than two groups a one-way ANOVA and posthoc analysis using Tukey's multiple comparison test was performed. The Kruskal-Wallis test was used for the analysis of non-parametric data. A p-value of < 0.05 was considered statistically significant and analysis was performed using GraphPad Prism version 8.3.0. The pilot data acquired from Spectronaut version 14 was transformed (log_2) in Perseus version 2.0.7.0 and a student's t-test was conducted to determine differentially expressed proteins between groups (chronic pancreatitis vs PDAC, PDAC vs healthy) at an FDR of 0.05 and S0 of 0.1. A power analysis was conducted using MSstats version 4.4.1 at a desired fold change of c (2,8), FDR = 0.01 and the number of samples = 6.

CHAPTER 3

RESULTS

The purpose of this study was to develop a standard operating procedure (SOP) for plasma proteomics LC-MS analysis that can be applied in-house or externally and is also adaptable for large study cohorts. A total of four experiments were conducted to optimise the sample preparation workflow that can be applied for screening plasma proteins. A mastermix containing 27 PDAC samples was used for the method optimization and 18 individual samples (6 PDAC, 6 CP, and 6 healthy) were used for the pilot study.

- Experiment 1 evaluated plasma sample dilution and precipitation methods (25X 50 mM Tris-HCl, 2% SDS pH 8 vs 5X 50 mM TEAB, 2% SDS vs 80% and 100% acetone precipitation).
- In experiment 2, the selected dilution methods and sample clean-up and digestion methods (HILIC vs PAC) were evaluated (25X 50 mM Tris-HCl with HILIC vs 25X 50 mM Tris-HCl with PAC vs 25X 50 mM TEAB, 2% SDS with HILIC)**.**
- Experiment 3 evaluated digestion conditions (25X 50 mM Tris-HCl with 4 hrs HILIC vs 16 hrs HILIC)
- Experiment 4 continued to evaluate digestion time together with digestion efficiency using two digestion enzymes: Trypsin and LysC (25x 50 mM Tris-HCl, 4 hrs HILIC with trypsin vs 16 hrs HILIC with trypsin vs 4 hrs HILIC with trypsin+lysC vs 16 hrs HILIC with trypsin+lysC vs 2 hrs HILIC with trypsin+lysC).

Section 1: Method optimization

3.1. Assessment of the dynamic range of study samples.

Plasma samples are highly complex and contain high abundant proteins that may mask the identification of low abundant proteins.

To view the dynamic range of plasma proteins, a study-specific sample pool was resolved using SDS-PAGE. From the gel the most abundant proteins were highlighted (e.g., albumin that migrated to approximately 66.5 kDa) (Figure 3.1). In an attempt to better detect low abundant proteins, the top 2 proteins (albumin and IgG) and top 12 proteins (albumin, IgG, α1-antitrypsin, α2-macroglobulin, apolipoprotein A-I, apolipoprotein A-II, fibrinogen,

haptoglobin, IgA, IgM, transferrin, and α1-acid glycoprotein) were depleted using two depletion kits (Pierce Albumin/IgG Removal kit and Pierce Top 12 Abundant Protein Depletion Spin Columns). Proteins were depleted as seen by the decrease in band intensity/thickness from crude plasma (Mix N) to the depleted plasma. The depleted samples have more defined bands near the molecular weights of the highly abundant proteins (Figure 3.2). Alternatively, fractionation was conducted on plasma samples using strong anion exchange (SAX) beads to separate the samples into fractions using increasing sodium chloride concentration for elution (Figure 3.3). Some fractions showed the removal of highly abundant proteins (lanes: 2, 5, 6, 7. 8. 9, 10 and, 11).

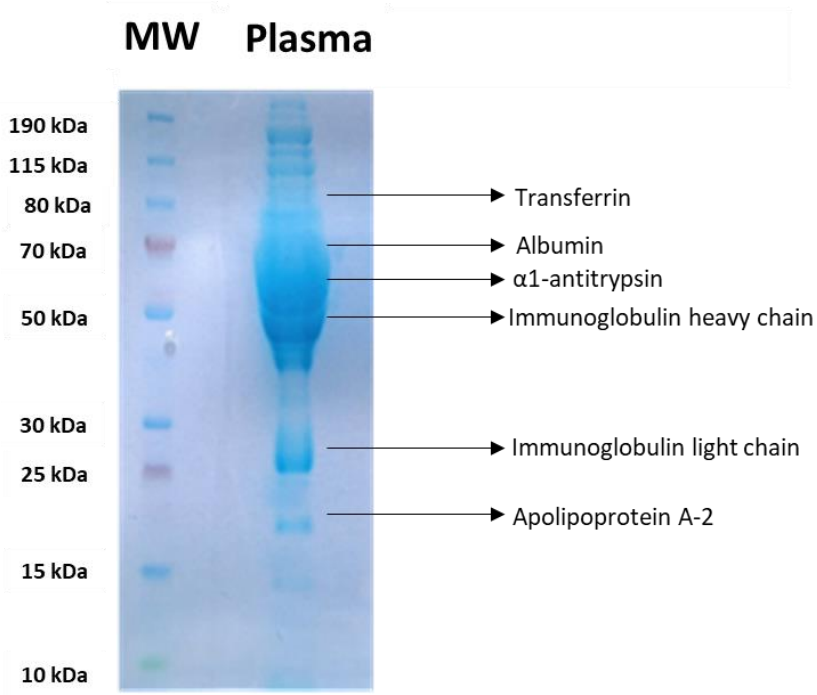


Figure 3. 1. Visualization of plasma dynamic range on SDS-PAGE. Gel depicting the high complexity of plasma samples relayed by the thickness and intensity of the bands. MW: molecular weight marker, lane 1: crude plasma sample. Some of the highly abundant proteins are depicted including Transferrin (MW: ~80 kDa), Albumin (MW: ~66.5 kDa), Immunoglobulin heavy chain (MW: ~50 kDa), α-1 antitrypsin (MW: ~52 kDa), Immunoglobulin light chain (MW: ~25 kDa) and Apolipoprotein A-2 (MW: ~17 kDa). kDa), α-1 antitrypsin (MW: ~52 kDa), Immunoglobulin light chain (MW: ~25 kDa) and Apolipoprotein A-2 (MW: ~17 kDa).

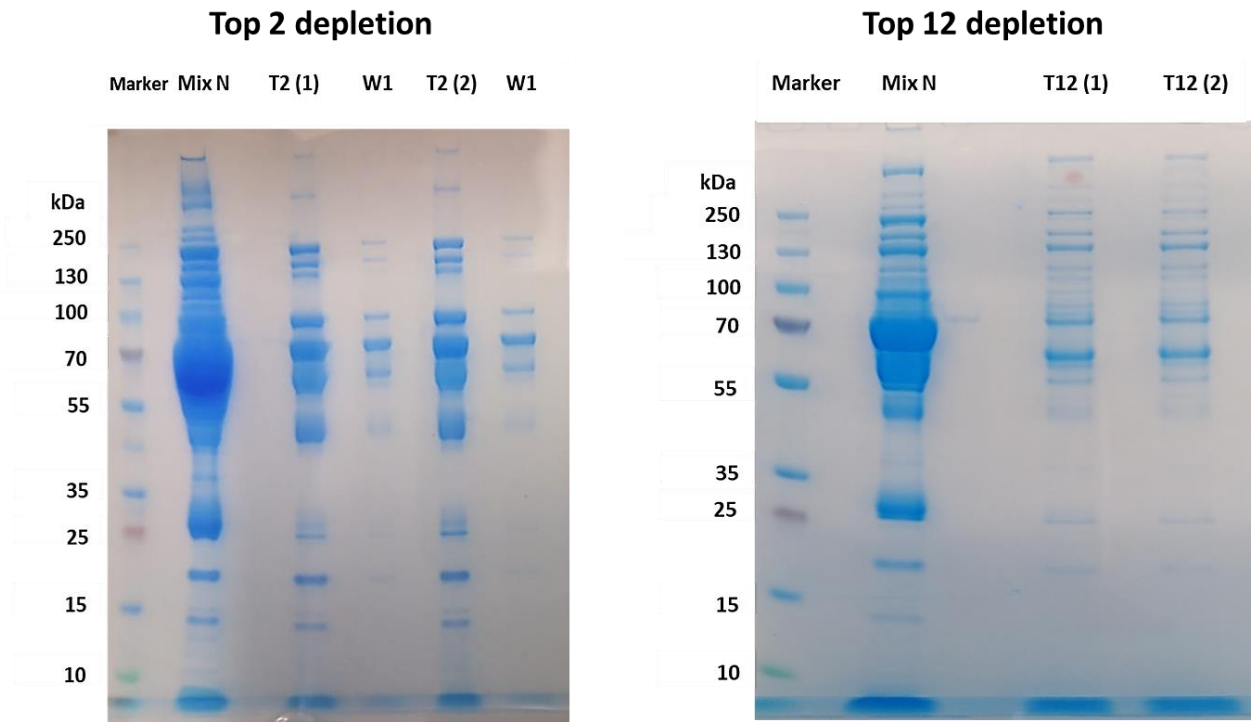


Figure 3. 2. Depleted plasma samples analyzed via SDS-PAGE. The gel on the left shows the Pierce Albumin/IgG Removal (Top 2 depletion)- lane 1: molecular weight marker, lane 2 (Mix N): crude plasma, lane 3: depleted sample 1, lane 4: wash 1, lane 5: depleted sample 2, lane 6: wash 2., and the gel on the right shows the Pierce Top 12 Abundant Protein Depletion (Top 12 depletion)- lane 1: molecular weight marker, lane 2: crude plasma, lane 3: depleted sample 1, lane 4: depleted sample 2.

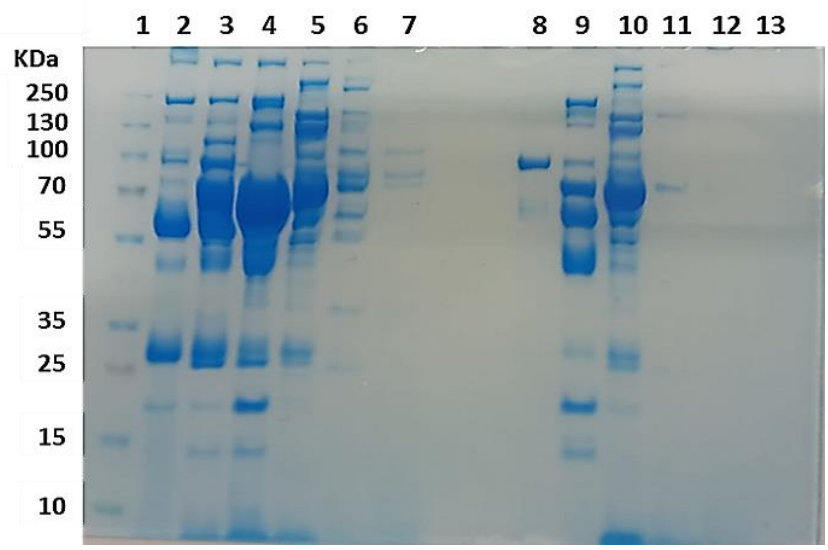


Figure 3. 3. Sample fractionation by SAX analyzed on SDS-PAGE. Plasma sample 1 with starting concentration of 0.25 mg (lane 2-7) was fractionated by increasing sodium chloride concentration: 1. Molecular weight marker 2. 50 mM, 3. 100 mM, 4. 200 mM, 5. 300 mM, 6. 400 mM and 7. 2M. Plasma sample 2 with starting concentration of 0.5 mg (lane 8-13) was fractionated by increasing sodium chloride concentration: 8. 50 mM, 9. 100 mM, 10. 200 mM, 11. 300 mM, 12. 400 mM and 13. 2 M.

3.2. Spectral library building

The library was built in-house using LC-MS data generated from depletion and fractionation experiments. The depletion experiment involved the use of top 2 and top 12 depletion kits (Figure 3.4). There was a significant difference in the identified peptides ($P < 0.0001$) and protein groups ($P < 0.0001$) between the top 2 and top 12 depletion. The fractionation experiments involved the elution of proteins that were bound to SAX beads using increasing concentrations of NaCl (Figure 3.5). Both the SAX individual fractions and depleted samples were subjected to reduction and alkylation, HILIC clean-up and subsequently LC-MS analysis. The results from the individual experiments were evaluated and then combined to form the spectral library. The spectral library identified 3 993 peptides matching 280 protein groups and 310 proteins (Table 3.1). The SAX plasma samples were fractionated at the protein level.

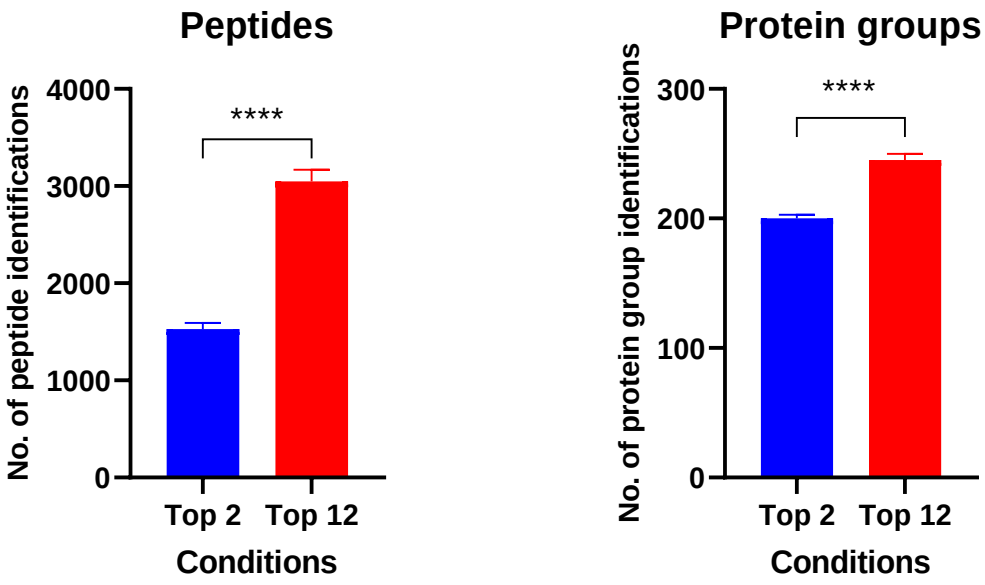


Figure 3. 4. Bar graphs of identified peptides and protein groups from depletion methods. The bar graph shows the number of identified peptides (left) and protein groups (right) from the Top 2 and Top 12 depletion kits. The unpaired t-test was used to compare the different methods and a p-value of < 0.05 was considered statistically significant. **** denotes $P < 0.0001$.

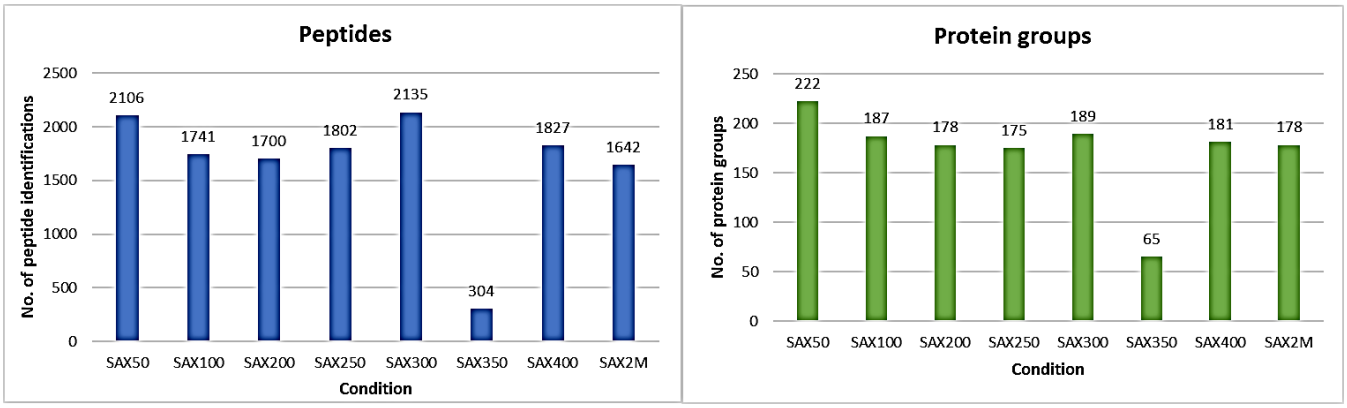


Figure 3. 5. Bar graph of identified peptides and protein groups from SAX fractionation. The bar graph shows the number of identified peptides (left) and protein groups (right) using fractionation by increasing sodium chloride concentration. The concentrations include 50mM (SAX50), 100 mM (SAX100), 200 mM (SAX200), 250 mM (SAX250), 300 mM (SAX300), 350 mM (SAX350), 400 mM (SAX400) and 2M (SAX2M) respectively.

Table 3. 1. Overview of parameters for the generated spectral library.

Parameter	**Spectral library identifications**
Peptides	3 993
Protein groups	280
Proteins	310

3.3. Experiment 1: Evaluation of plasma protein preparation methods

In experiment 1, precipitation methods were compared (80% and 100% acetone precipitation) with dilution methods (25X 50mM Tris-HCl, 2% SDS pH 8 and 5X 50mM TEAB). Each method consisted of four technical replicates and was repeated over two different days.

The total ion chromatograms (TICs) for the different methods in each experiment were evaluated in PeakView to assess the range of total signal across sample replicates and determine if the signal correlates with the amount of peptide loaded (Figures S1-S4). It provides valuable information about the distribution and abundance of ions present in a sample.

All samples were successfully loaded for MS analysis, and none showed under/overloading therefore the data were further analyzed on the Spectronaut 14 software where the data was normalized using global normalization (median) following data extraction. Boxplots of responses (total signal) for each of the individual runs before and after normalization are

shown in Figure 3.6. The objective of normalization was to enhance the comparability of data samples and ensure reliable results for downstream analysis by reducing errors and accounting for biases. This was accomplished by calculating the median intensity of each ion across all runs and determining the multiplication factor for each ion. This scaling technique allowed for the effective comparison of ion intensities.

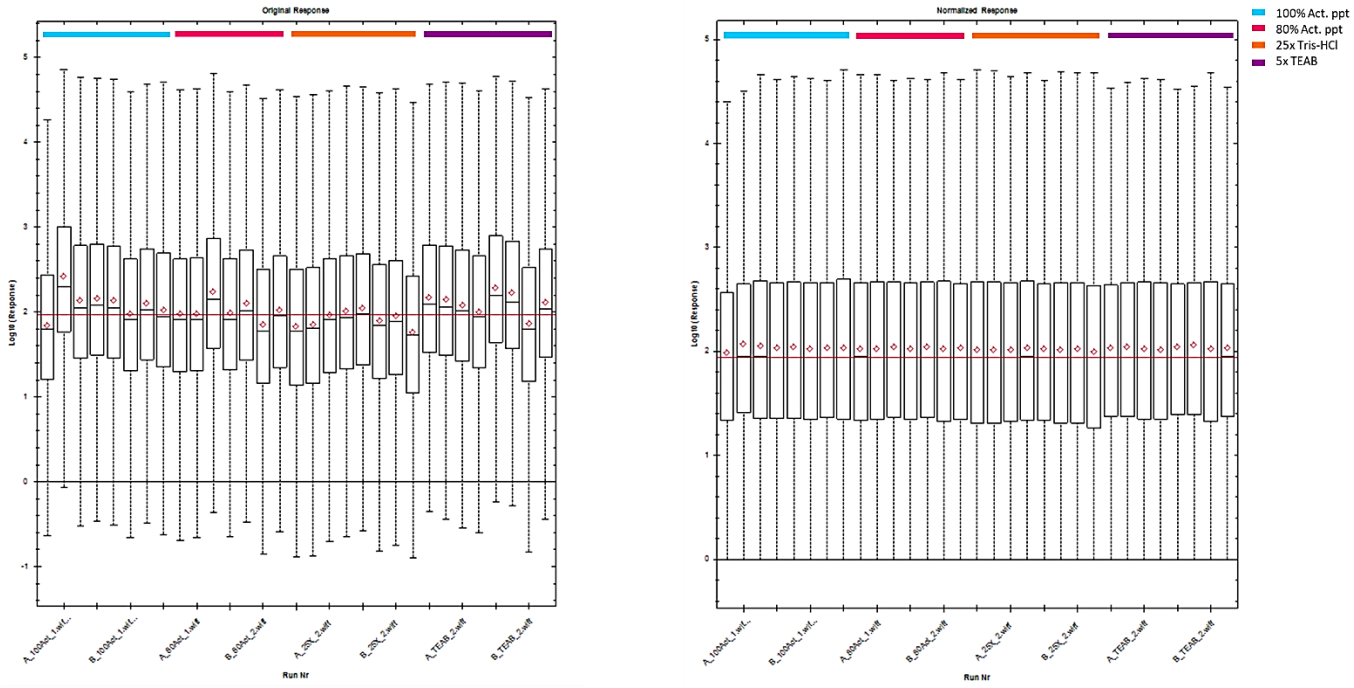


Figure 3. 6. Normalization plot for experiment 1. Box plots showing the normal response (signal response before normalization) of all precursor responses for each run (left) and the normalization of the same precursor responses for each run (right). Each method had 4 technical replicates and was conducted over two different days**.** Sample 2_80Act_A (80% acetone_day1) was an outlier and was removed.

The coefficient of variation (CV) measures any variation introduced between the two different days (inter-day variation) and across all replicates for both days and aids in determining if each method is reproducible. For the inter-day comparison (Figure 3.7), all methods show median CVs at the peptide level ranging from 15%-33% and for the intra-day comparison, all methods ranged from 23%-32% median CV (Figure 3.8).

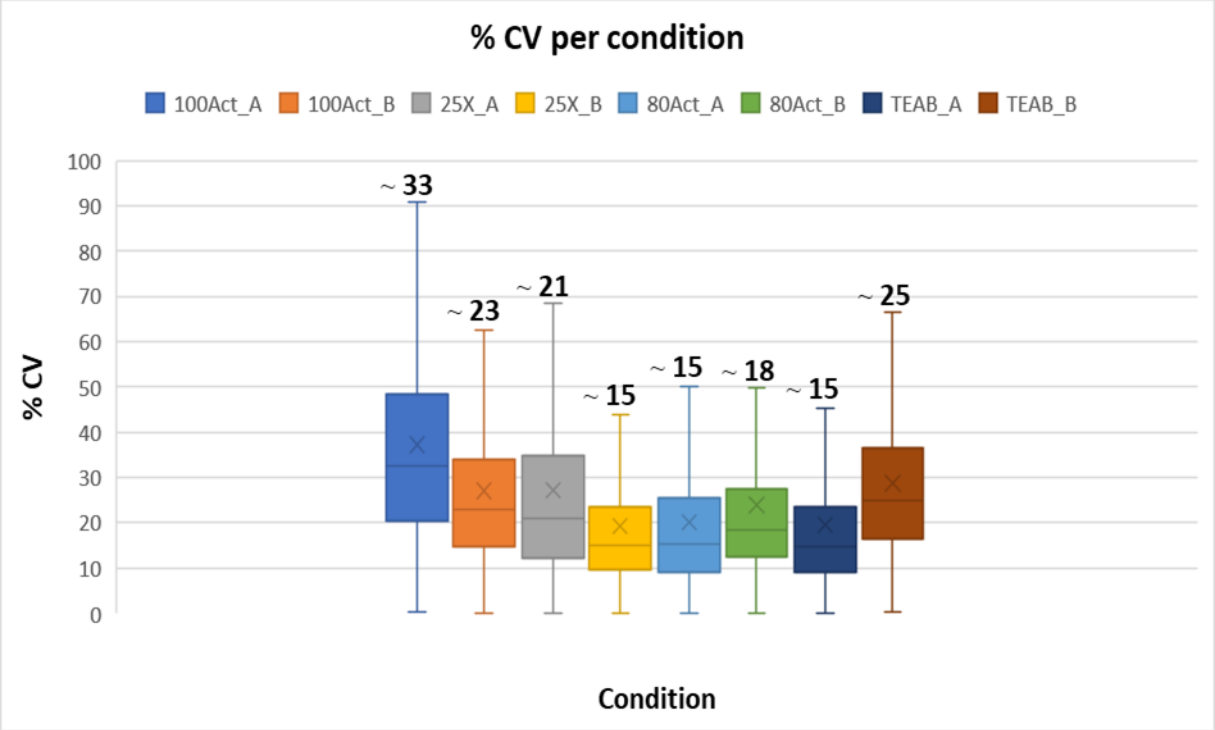


Figure 3. 7. Box plot indicating inter-day variability per condition over 2 days (A and B) in experiment 1. The box plot shows the median % CV of each plasma preparation condition over 2 days (A and B) in experiment 1. All methods range from 15%-33% variation. 100Act_A: 100% acetone precipitation_day 1, 100Act_B: 100% acetone precipitation_day 2, 25X_A: 25X 50 mM Tris-HCl, 2% SDS dilution_day 1, 25X_B: 25X 50 mM Tris-HCl, 2% SDS dilution_day 2, TEAB_A: 5X 50 mM TEAB, 2% dilution_day 1, and TEAB_B: 5X 50 mM TEAB, 2% dilution_day 2.

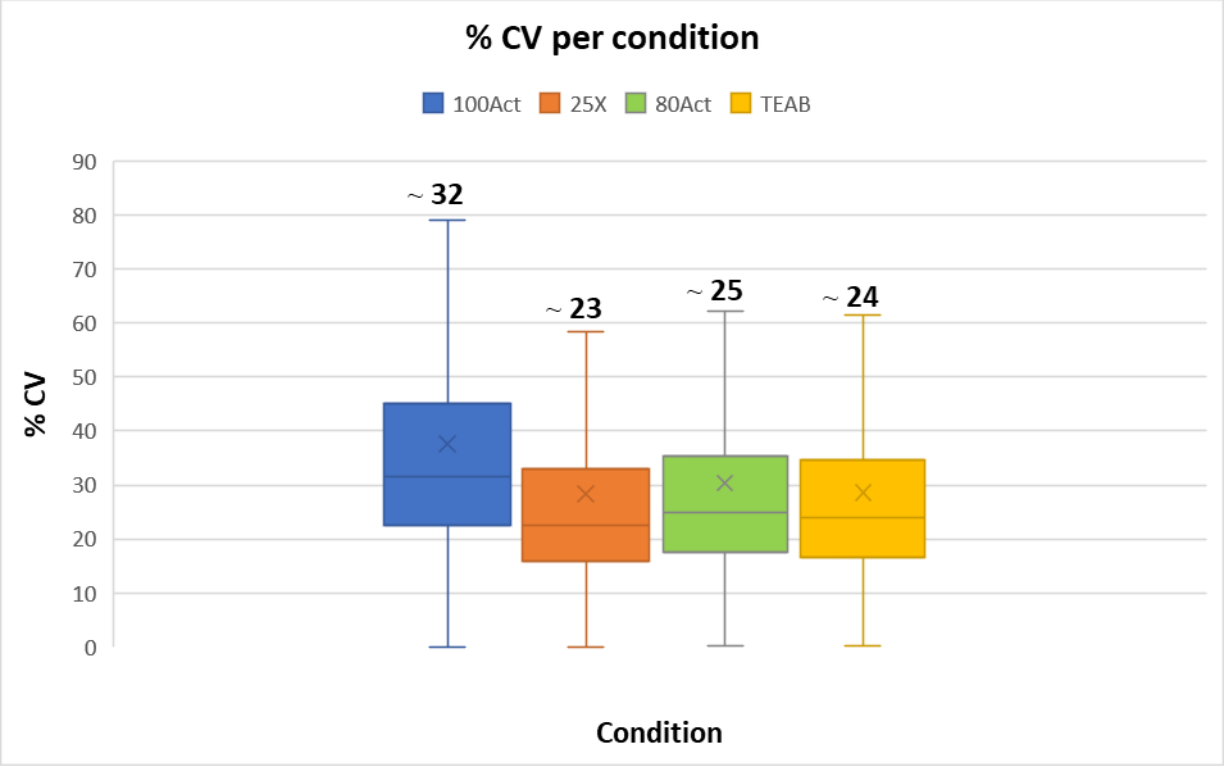


Figure 3. 8. Box plot indicating variability across all replicates per condition in experiment 1. The box plot shows the median % CV of each plasma preparation condition in experiment 1. All methods range from 23%-32% variation. 100Act: 100% acetone precipitation, 25X: 25X 50 mM Tris-HCl, 2% SDS dilution, 80Act: 80% acetone precipitation, and TEAB: 5X 50 Mm TEAB, 2% SDS dilution.

Evaluation of peptides and protein groups between the different experimental methods revealed there was a significant difference in the number of identified peptides between 25X Tris and 5X TEAB (P = 0.0007) dilution methods and between 80Act and 5X TEAB (P = 0.0220). For protein groups, there was a significant difference between the 25X Tris and 5X TEAB (P = 0.0264) dilution methods (Figure 3.9).

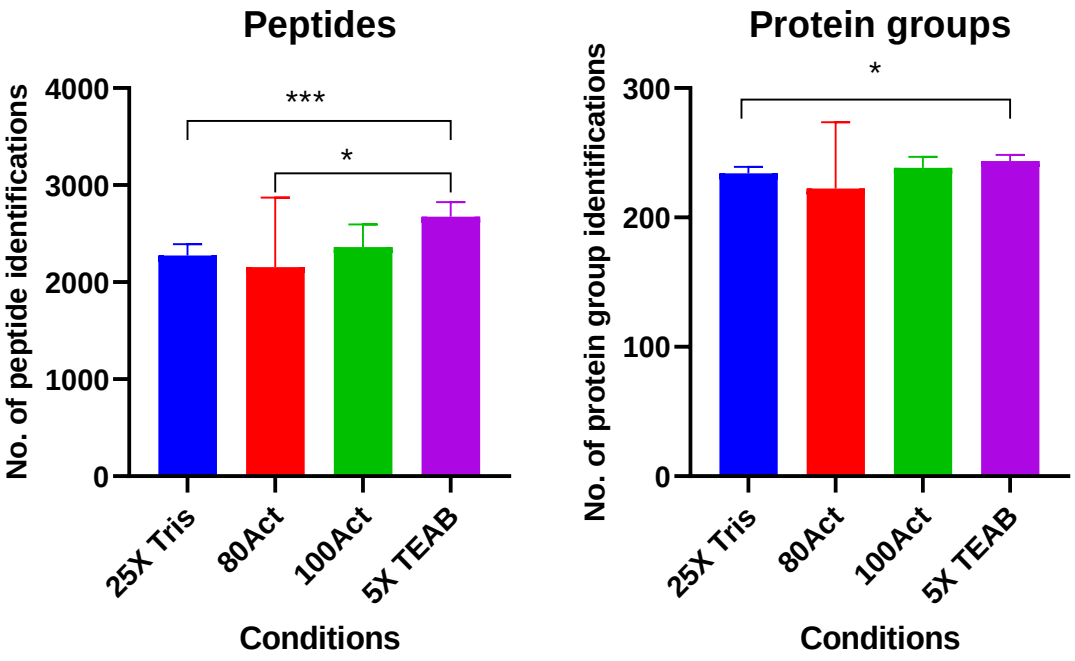


Figure 3. 9. Bar graph of identified peptides and protein groups from experiment 1. The bar graph shows the number of identified peptides (left) and protein groups (right), under the following conditions: 100% (100Act) and 80% (80Act) acetone precipitation, 25x Tris-HCl dilution (25X Tris) as well as a 5x TEAB dilution (5X TEAB), over 2 days (A and B). Sample 2_80Act_A was an outlier and was removed. The Kruskal-Wallis and Dunn's multiple comparison tests were used to compare the different methods and a p-value of < 0.05 was considered statistically significant. * denotes $p < 0.05$, *** denotes $p < 0.001$.

3.4. Experiment 2: evaluation of selected protein preparation methods and sample clean-up methods.

Selected protein dilution methods from experiment 1 and sample clean-up and digestion methods (HILIC vs PAC) were evaluated and compared. The methods included 25X 50mM Tris-HCl with HILIC (Tris_HILIC), 25X 50mM Tris-HCl with PAC (Tris_PAC), and 25X 50mM TEAB, 2% SDS with HILIC (TEAB_HILIC)**.** The selected methods were more reproducible, identified a higher number of peptides and are easier to automate. The dilution factor for TEAB buffer was increased from 5-fold to 25-fold to possibly improve protein coverage.

The total ion chromatograms (TICs) for the different methods were observed (Appendix, Figures S5-S6) and indicated no sample loading bias. The data was normalized in Spectronaut (Figure S7). Inter-day variation was observed where the median CVs between

the different days ranged from 16%-25% (Figure 3.10) and median CVs between the sample replicates ranged from 23%-29% (Figure 3.11).

There was a significant difference in the number of identified peptides between TEAB_HILIC and Tris_PAC (P = 0.0008) and between Tris_HILIC and Tris_PAC (P = 0.0439), but no significant difference in the protein groups (Figure 3.12).

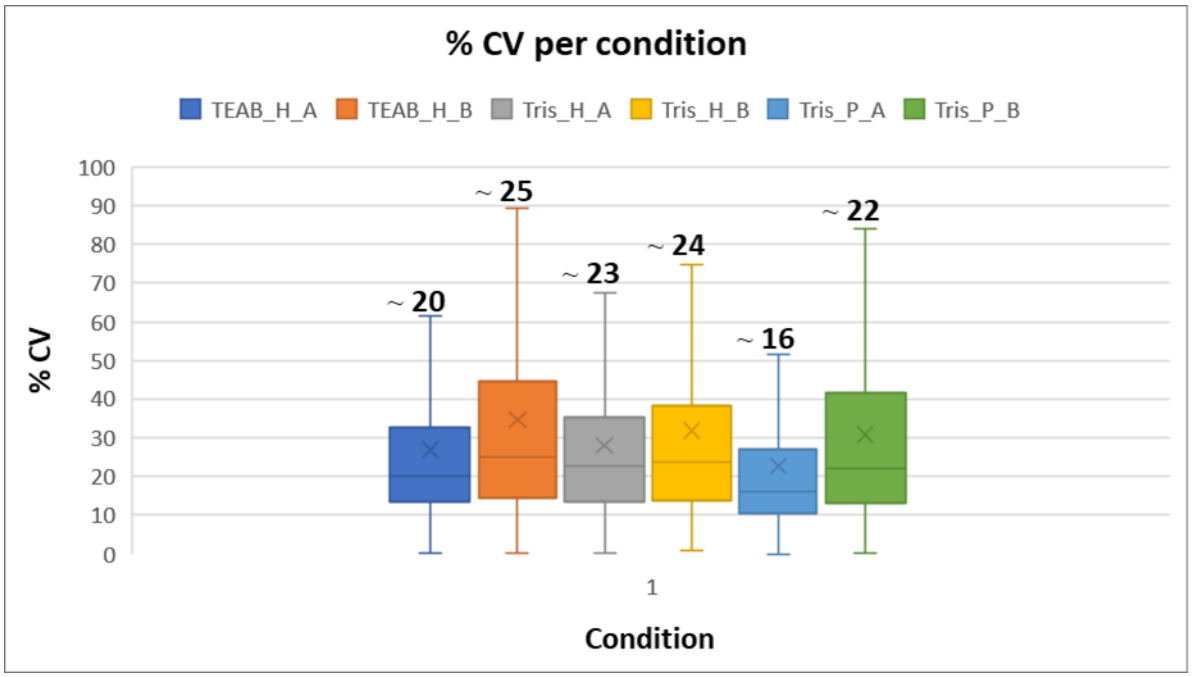


Figure 3. 10. Box plot indicating inter-day variability per condition over 2 days (A and B) in experiment 2. The box plot shows the median % CV of each condition and sample clean-up method over 2 days (A and B). the variation percentage between the different days in this experiment ranged from 16%-25%.TEAB_H_A: 25X TEAB_HILIC_day 1, TEAB_H_B: 25X TEAB_HILIC_day 2, Tris_H_A: 25X Tris_HILIC_day 1, Tris_H_B: 25X Tris_HILIC_day 2, Tris_P_A: 25X Tris_PAC_day 1, Tris_P_B: 25X Tris_PAC_day 2.

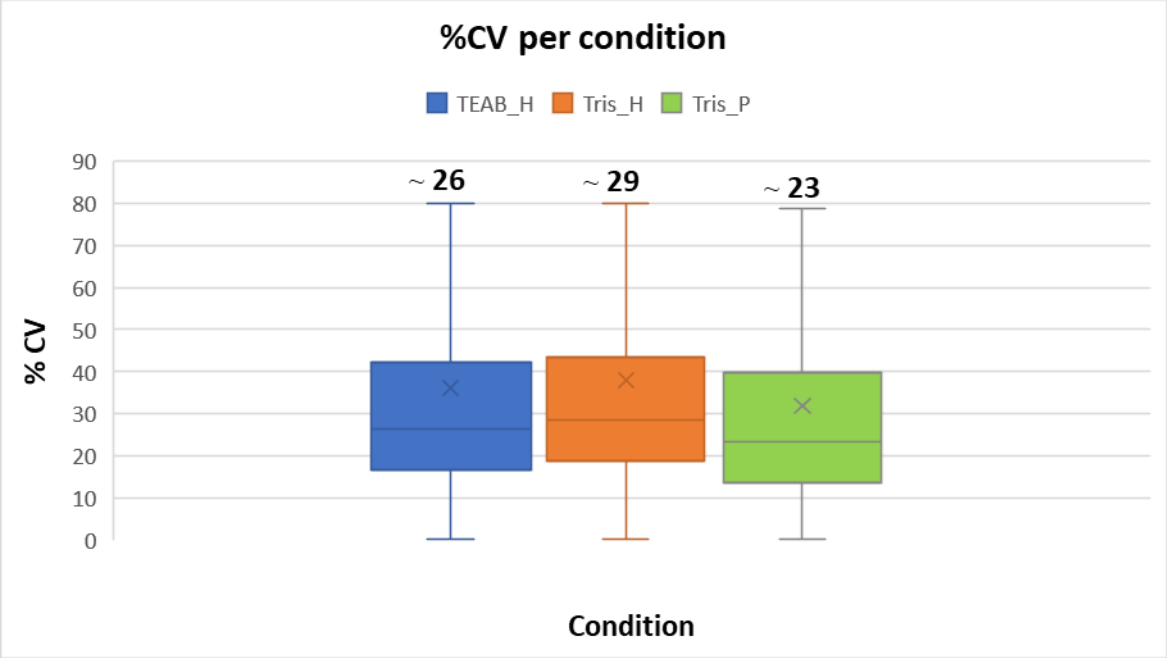


Figure 3. 11. Box plot indicating variability across all replicates per condition in experiment 2. The box plot shows the median % CV of each condition and sample clean-up method. All methods showed low intra-day variation, with % CVs ranging from 23%-29%. TEAB_H: 25X TEAB_HILIC, Tris_H: 25X Tris_HILIC, Tris_P: 25X Tris_PAC.

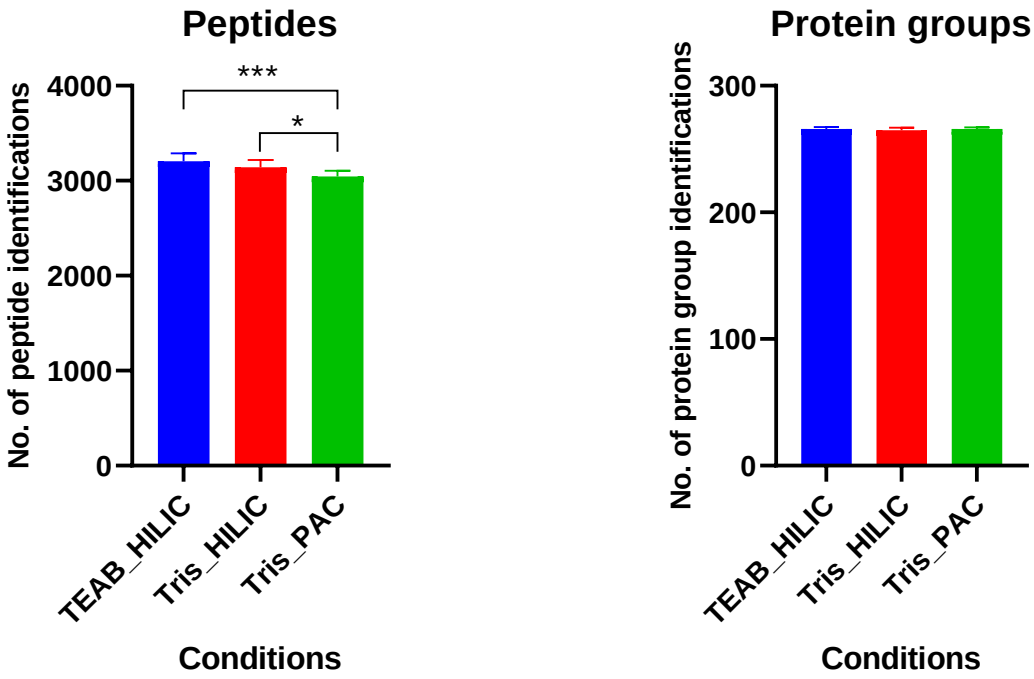


Figure 3. 12. Bar graph of identified peptides and protein groups from experiment 2. The bar graph shows the number of identified peptides (left) and protein groups (right), under the following conditions: 25x TEAB dilution with HILIC sample clean-up, 25x Tris-HCl dilution with HILIC and PAC sample clean-up, over 2 days (A and B). The one-way ANOVA and Tukey's multiple comparison posthoc test were used to compare the different methods and a p-value of < 0.05 was considered statistically significant. * denotes $p < 0.05$ and *** denotes $p < 0.001$.

3.5. Experiment 3: Evaluation of digestion conditions (time)

The HILIC clean-up method was re-evaluated. A 4 hrs HILIC digestion was compared to an overnight (16 hrs) HILIC digestion with trypsin. The digested time was extended to improve protein coverage.

The total ion chromatograms (TICs) for the different methods were assessed (Figures S10-S11), thereafter the data was normalized (Figure S12) before downstream analysis. Inter-day variation was observed where the median CV between the different days ranged from 15%-31% (Figure 3.13) and the median CV between the sample replicates ranged from 34%-36% (Figure 3.14).

Following the evaluation of peptide and protein group identification, there was a significant difference in the peptides identified (P = 0.0216), but no significant difference in protein group identifications between the 4 hrs and 16 hrs (O/N) digestions (Figure 3.15).

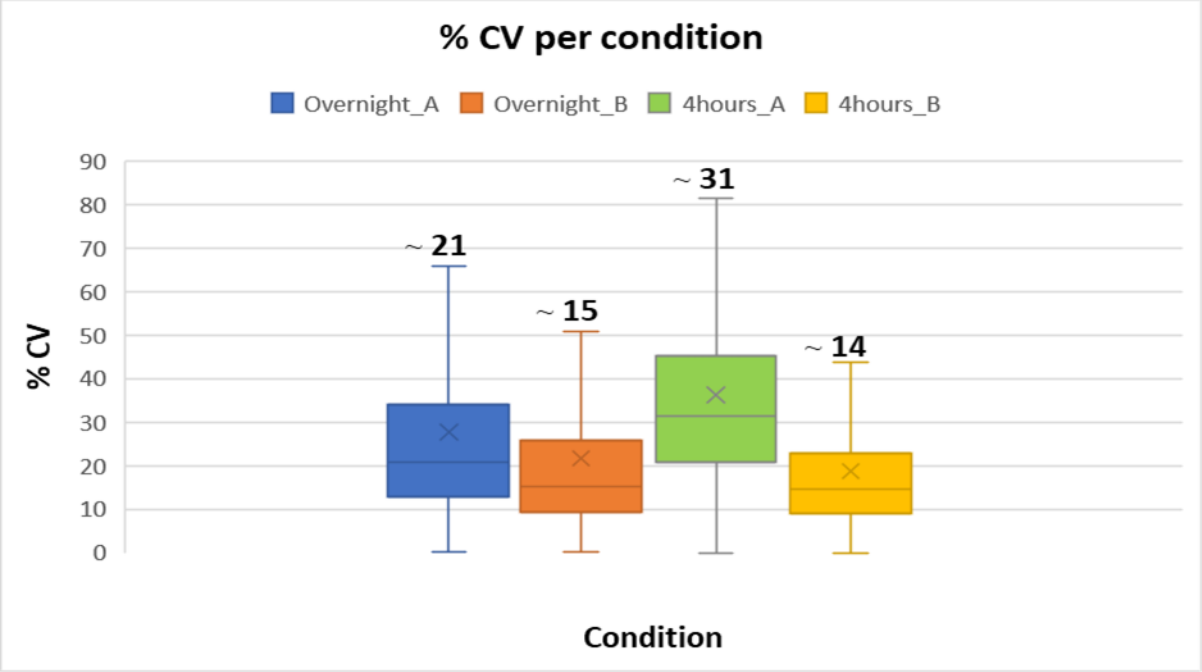


Figure 3. 13. Box plot indicating inter-day variability per condition over 2 days (A and B) in experiment 3. The box plot shows the median % CV of the overnight and 4-hour digestion condition and sample clean-up method over 2 days (A and B). In experiment 3, the variation between the different days ranges from 14%-31%. Overnight_A: 16 hrs HILC digestion day 1, Overnight_B: 16 hrs HILC digestion day 2, 4hours_A: 4 hrs HILIC digestion day 1, 4hours_B: 4 hrs HILIC digestion day 2.

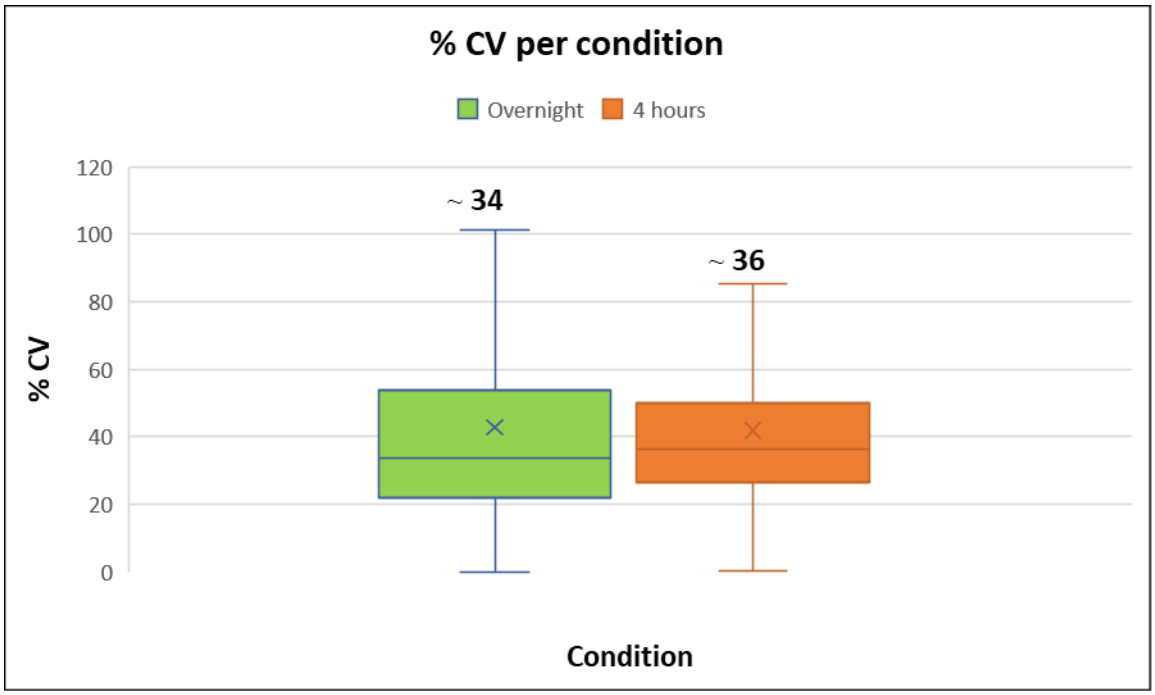


Figure 3. 14. Box plot indicating intra-day variability per condition in experiment 3. The box plot shows the median % CV of the 16 hrs (overnight) and 4 hrs (4 hours) HILIC digestion condition and sample clean-up. Both methods had slightly similar inter-day %CVs.

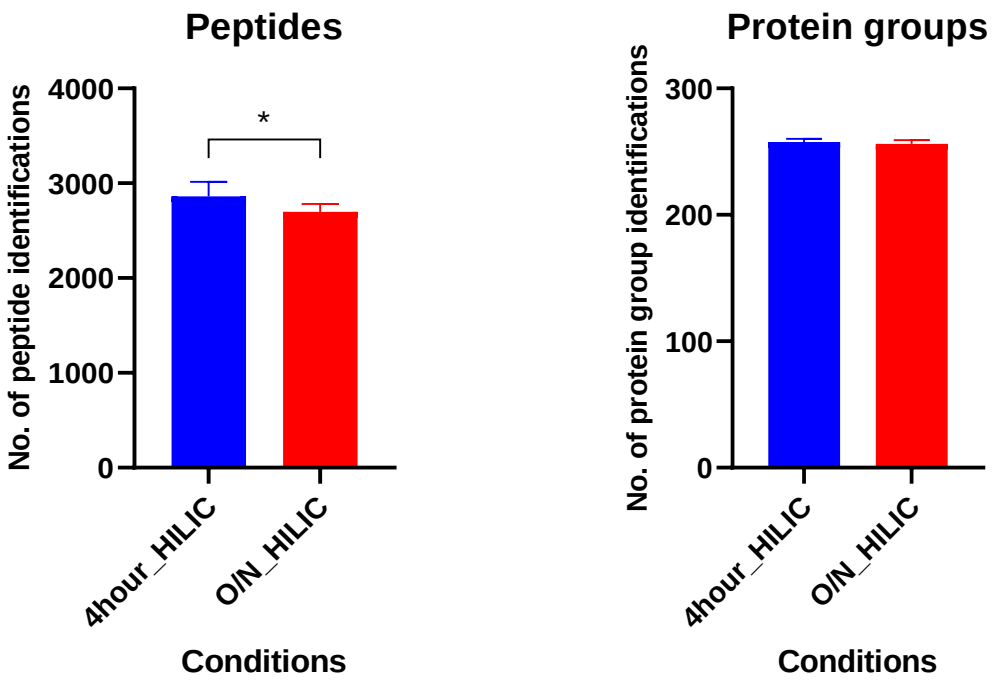


Figure 3. 15. Bar graph of identified peptides and protein groups from experiment 3. The bar graph shows the number of identified peptides, under the following conditions: 4-hrs digestion (left) and 16 hrs (O/N) digestion (right) at 47 °C, over 2 days (A and B). The unpaired t-test was used to compare the different methods and a p-value of < 0.05 was considered statistically significant. * denotes $p < 0.05$.

3.6. Experiment 4: Evaluation of digestion condition and enzyme efficiency

Digestion time was further evaluated and compared together with digestion efficiency using two digestion enzymes (Trypsin and LysC), the methods compared included 4 hrs HILIC with trypsin, 16 hrs HILIC with trypsin, 4 hrs HILIC with trypsin+lysC, 16 hrs HILIC with trypsin+lysC, and 2 hrs HILIC with trypsin+lysC.

The TICs (Figures S15-S19) were assessed and showed no sample loading bias. The data was normalized (Figure S20) before downstream analysis. Inter-day variation was observed where the median CV between the different days ranged from 15%-31% (Figure 3.16) and the median CV between the sample replicates ranged from 34%-36% (Figure 3.17).

There was a significant difference in the number of peptides in the following comparison: Trp_4h and Trp_ON (P = 0.0034), Trp_4h and Trp+LysC_ON (P < 0.0001), Trp+LysC_4h and Trp+LysC_ON (P = 0.0002), Trp_LysC_2h and Trp_ON (P = 0.0025), Trp+LysC_2h and Trp+LysC_ON (P < 0.0001), Trp_ON and Trp_LysC_ON (P = 0.0309). The protein

groups showed a significant difference only between the methods Trp_4h and Trp+LysC_ON (P = 0.0029) (Figure 3.18).

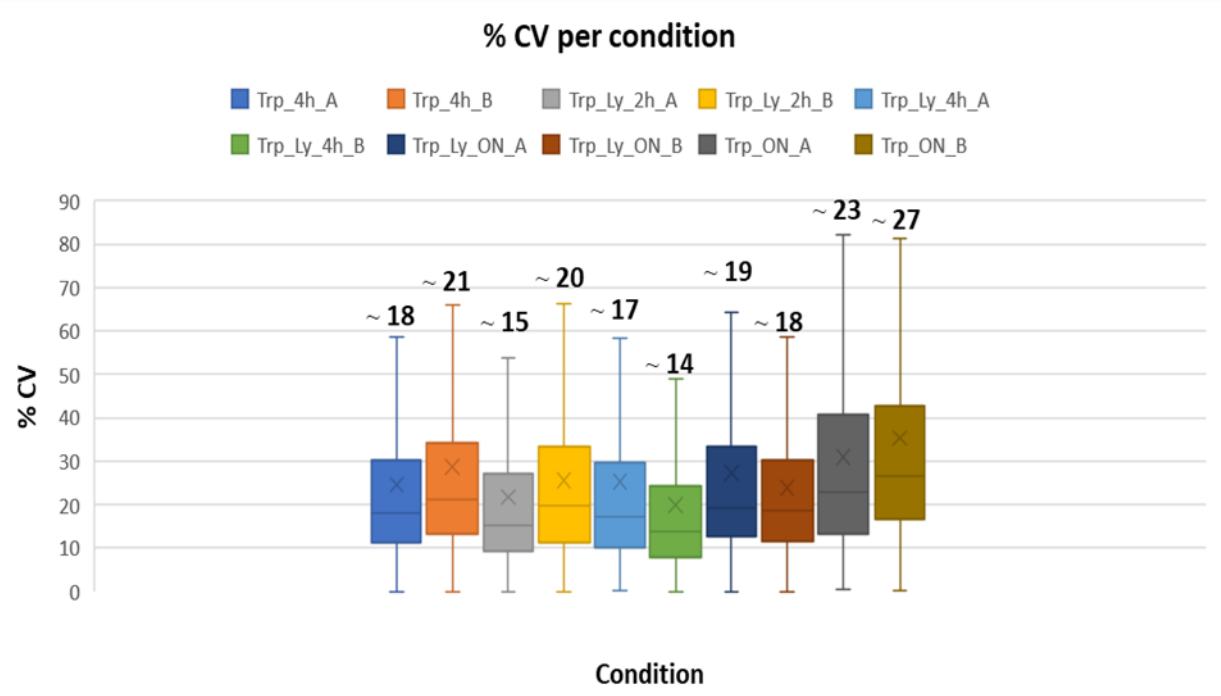


Figure 3. 16. Box plot indicating inter-day variability per condition over 2 days (A and B) in experiment 4. The box plot shows the median % CV of digestion condition over 2 days (A and B). The % CV between the two different days in this experiment ranges from 14%-27%. Trp_4h_A: 4 hrs HILIC digest with trypsin day 1, Trp_4h_B: 4 hrs HILIC digest with trypsin day 2, Trp_Ly_2h_A: 2 hrs HILIC digest with trypsin+LysC day 1, Trp_Ly_2h_B: 2 hrs HILIC digest with trypsin+LysC day 2, Trp_Ly_4h_A: 4 hrs HILIC digest with trypsin+LysC day 1, Trp_Ly_4h_B: 4 hrs HILIC digest with trypsin+LysC day 2, Trp_Ly_ON_A: 16 hrs HILIC digest with trypsin+LysC day 1, Trp_Ly_ON_B: 16 hrs HILIC digest with trypsin+LysC day 2, Trp_ON_A: 16 hrs HILIC digest with trypsin day 1, Trp_ON_B: 16 hrs HILIC digest with trypsin day 2.

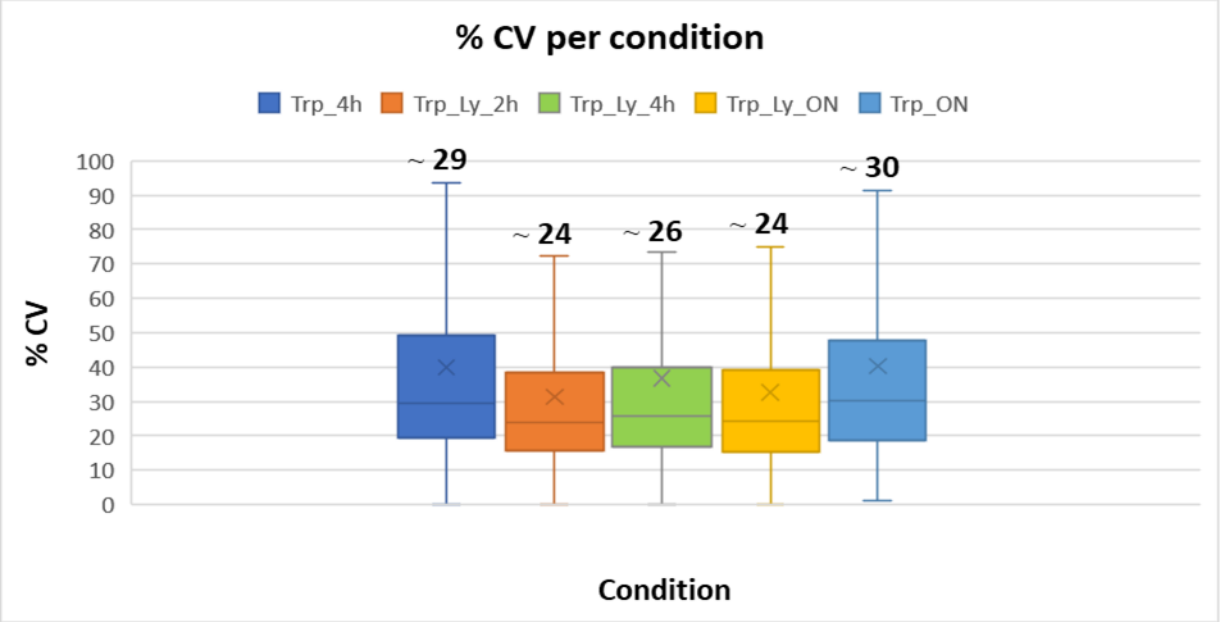


Figure 3. 17. Box plot indicating intra-day variability per condition in experiment 4. The box plot shows the median % CV of each digestion condition in experiment 4. All the different conditions in this experiment show slightly similar %CVs ranging from 24%-30%. Trp_4h: 4 hrs HILIC digest with trypsin, Trp_Ly_2h: 2 hrs HILIC digest with trypsin+LysC, Trp_Ly_4h: 4 hrs HILIC digest with trypsin+LysC, Trp_Ly_ON: 16 hrs HILIC digest with trypsin+LysC, Trp_ON: 16 hrs HILIC digest with trypsin.

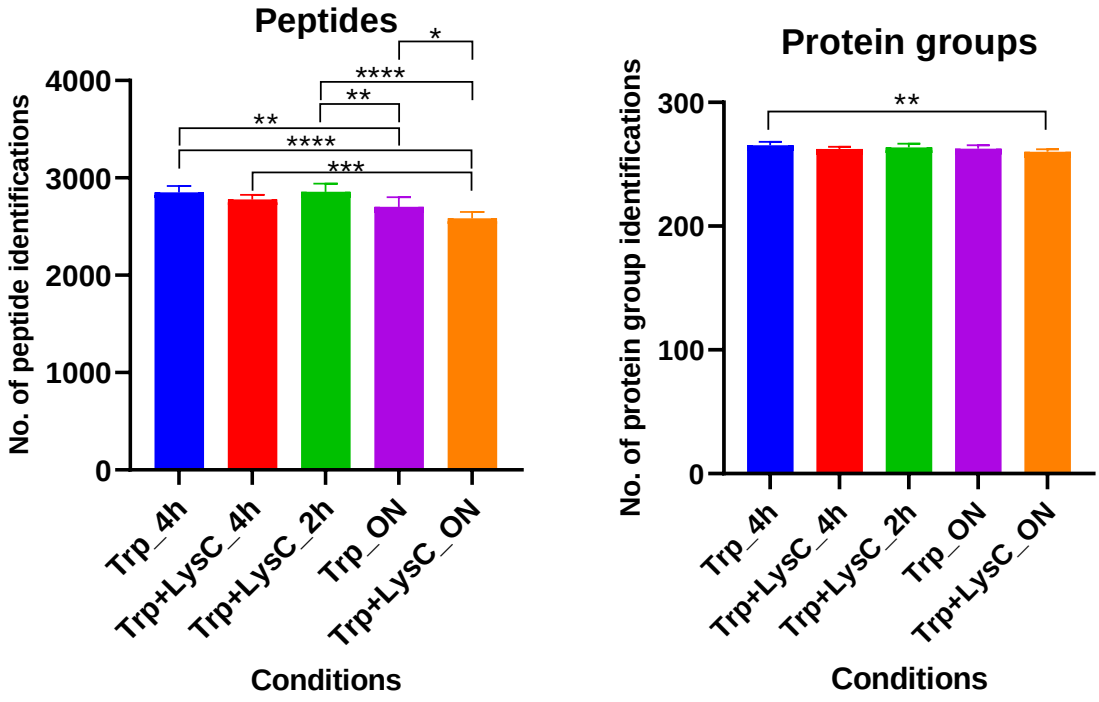


Figure 3. 18. Bar graph of identified peptides and protein groups from experiment 4. The bar graph shows the number of identified peptides (left) and protein groups (right) under the following conditions: 4 hrs tryptic digestion (Trp_4h), 2 hrs tryptic and LysC digestion (Trp+LysC_2h), 4 hrs tryptic and LysC digestion (Trp+LysC_4h), as well as a 16 hrs tryptic and LysC digestion (Trp+LysC_0N) and a 16 hrs tryptic digestion (Trp_ON) at 47 °C, over 2 days (A and B). Sample 1_Trp_Ly_4h_B was an outlier and was removed. The one-way ANOVA and Tukey's multiple comparison posthoc test were used to compare the different methods and a p-value of < 0.05 was considered statistically significant. *, **, ***, and ****, denotes $p < 0.05$, $p < 0.01$, $p < 0.001$, and $p < 0.0001$ respectively.

3.7. Final standard operating procedure

Blood samples are collected from patients, and plasma preparation is conducted from the blood samples. The plasma is diluted 25-fold with 50 mM Tris-HCl, 2% SDS pH 8. The plasma proteins are subjected to reduction with DTT at 60 ℃ and alkylated with IAA for 30 mins, purified of detergent and salts and then digested for 2 hrs at 47 ℃ with trypsin and LysC using the HILIC method. The resultant peptides are then analyzed via LC-MS analysis and the data is processed in the Spectronaut 14 software and further processed in the Perseus software. The full SOP can be found in the appendix 1.

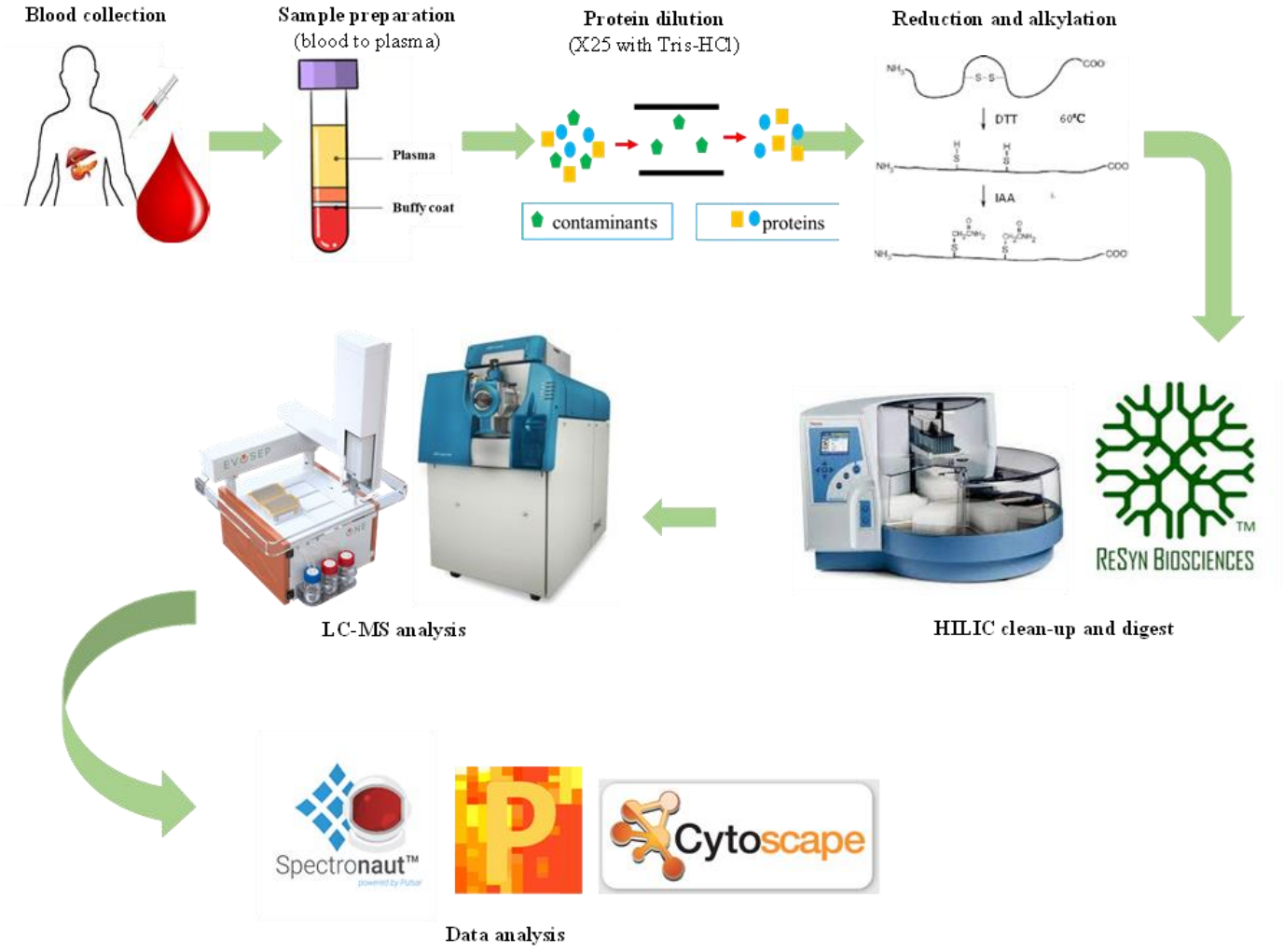


Figure 3. 19. Final standardized workflow. The final standard operating procedure includes the collection of blood samples from PDAC patients. The blood samples are then prepared to make plasma samples, and from the plasma samples proteins are extracted and subjected to reduction and alkylation. Thereafter the samples are cleaned up and digested using the automated HILIC method. Peptides are then analyzed via LC-MS and the data is processed using Spectronaut, Perseus and Cytoscape.

Section 2: Pilot study

3.8. Proteome profiling of plasma samples

The final standardized workflow was applied to a pilot study. The main aim of the pilot study was to evaluate the optimized preparation and digestion methods and determine their utility in profiling plasma proteins and identify key targets that are associated with PDAC. Before conducting the pilot study, a power analysis was conducted to aid in determining the power and fold change cut-off for a minimum of 6 samples (Table S3). Six samples per group were analyzed because that was the limited number of control samples that had been collected when the study was conducted. A second power analysis was conducted for future reference to determine the minimum number of samples and fold change cut-off that would be required for a power of 0.9 (Table S4). However, because this was a pilot study, only 18 patients (6 healthy individuals, 6 chronic pancreatitis and 6 PDAC patients) were evaluated.

Raw SWATH (.wiff) data files were extracted in Spectronaut 14 software and matched to the in-house generated spectral library. The t-test data (healthy vs PDAC and PDAC vs CP) generated were plotted as volcano plots on Perseus v2.0.7.0. Comparing PDAC to healthy patients, 25 proteins (Table 3.2) were found to be differentially expressed (22 downregulated and 3 upregulated). The top 3 (in abundance) upregulated proteins included Transthyretin, apolipoprotein A4 and tetranectin. The top 3 (in abundance) downregulated proteins included Haptoglobin-related protein, Complement C1q A chain, and Immunoglobulin heavy constant alpha 2 (Figure 3.20A; Table 3.2). The dysregulated proteins are further depicted in a table with respective cancer prognostic biomarker status and expression pattern (Table 3.2).

A total of two proteins (Afamin and Immunoglobulin heavy variable 3-43**)** were downregulated in comparing PDAC to CP patients (Figure 3.20B).

Table 3. 2. differentially expressed proteins measured in group comparisons (healthy patients versus patients with PDAC).

Unit accession no.	Protein description	Cancer prognostic biomarker status	-log p	Expression pattern
P00739	Haptoglobin-related protein	Not prognostic	3.26	Downregulated
P00751	Complement factor B	Prognostic marker in renal cancer (unfavourable), breast cancer	2.84	Downregulated

		(favourable)		
P01009	Serpin family A member 1	Prognostic marker in breast cancer (unfavourable) and colorectal cancer (favourable)	2.30	Downregulated
P01011	Serpin family A member 3	Prognostic marker in breast cancer (favourable)	4.91	Downregulated
P01019	Angiotensinogen	Not prognostic	2.62	Downregulated
P01024	Complement C3	Prognostic marker in renal cancer (unfavourable) and liver cancer (favourable)	2.98	Downregulated
P01701	Immunoglobulin lambda variable 1-51	Prognostic marker in renal cancer (unfavourable), head and neck cancer (favourable), breast cancer (favourable)	1.84	Downregulated
P01877	Immunoglobulin heavy constant alpha 2	Prognostic marker in breast cancer (favourable), renal cancer (unfavourable), head and neck cancer (favourable)	2.03	Downregulated
P02745	Complement C1q A chain	Prognostic marker in renal cancer (unfavourable)	1.78	Downregulated
P02748	Complement C9	Not a prognostic marker	2.40	Downregulated
P02750	Leucine-rich alpha-2-glycoprotein 1	Prognostic marker in renal cancer (unfavourable) and thyroid cancer (favourable)	2.48	Downregulated
P02766	Transthyretin	Not prognostic	2.07	Upregulated
P02774	GC vitamin D binding protein	Prognostic marker in renal cancer (unfavourable) and liver cancer (favourable)	4.07	Downregulated
P03951	Coagulation factor XI	Prognostic marker in liver cancer (favourable)	1.83	Downregulated
P05155	Serpin family G member 1	Prognostic marker in renal cancer (unfavourable), ovarian cancer (favourable), colorectal cancer (unfavourable)	3.21	Downregulated
P05452	C-type lectin domain family 3	Prognostic marker in liver cancer (favourable), head and neck cancer (favourable), pancreatic cancer (favourable)	2.61	Upregulated

P05543	Serpin family A member 7	Not prognostic	2.16	Downregulated
P06727	Apolipoprotein A4	Not prognostic	2.43	Upregulated
P08185	Serpin family A member 6	Prognostic marker in endometrial cancer (favourable)	2.26	Downregulated
P08697	Serpin family F member 2	Prognostic marker in liver cancer (favourable)	2.21	Downregulated
P10909	Clusterin	Prognostic marker in thyroid cancer (favourable)	2.08	Downregulated
Q06033	Inter-alpha-trypsin inhibitor heavy chain 3	Prognostic marker in renal cancer (unfavourable)	3.19	Downregulated
A0A0C4DH34	Immunoglobulin heavy variable 4-28	Prognostic marker in renal cancer (unfavourable), head and neck cancer (favourable)	3.63	Downregulated
A0A0C4DH41	Immunoglobulin heavy variable 4-61	Prognostic cancer in renal cancer (unfavourable), head and neck (favourable)	2.75	Downregulated
P01031	Complement C5	Prognostic marker in liver cancer (favourable)	3.91	Downregulated

all values were transformed at log2 (x) fold change. The cancer prognostic status of each protein was identified on The Human Protein Atlas. Unfavourable: high expression of gene is unfavourable to the specific cancer and favourable: high expression of gene is favourable to the specific cancer.

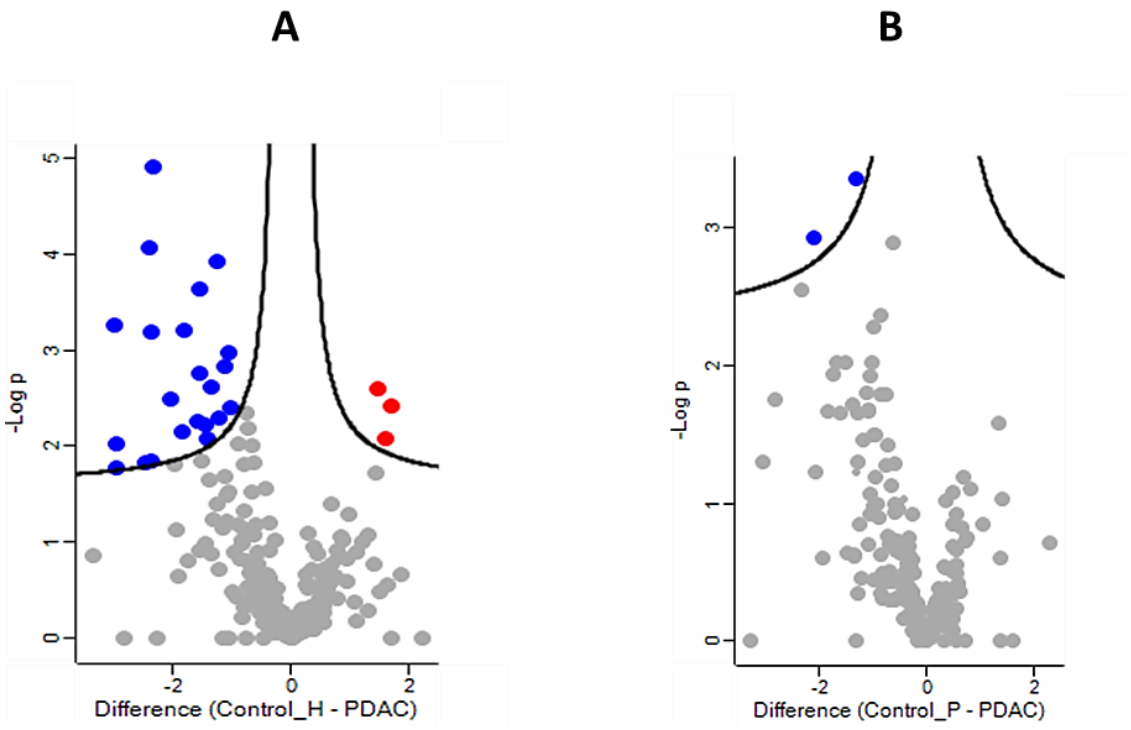


Figure 3. 20. Volcano plots displaying dysregulated proteins between (A) PDAC and healthy patients and (B) PDAC and CP patients. The volcano plot shows the comparison between healthy patients (Control_H) and PDAC (A). The blue and red nodes (circles) display 22 downregulated and 3 upregulated proteins, respectively. Comparison between CP (Control_P) and PDAC (B). The blue nodes display 2 downregulated proteins (P43652-Afamin and A0A0B4J1X8-Immunoglobulin heavy variable 3-43). Proteins were visually represented by fold change (difference) and significance (-log p) using a false discovery rate of 0,01 and s0 of 0,1 using Perseus software.

ShinyGO v0.76.3 and Reactome were used to identify pathways enriched by the proteins dysregulated in PDAC patients compared to healthy patients. Some of these pathways are associated with carcinogenesis including those related to the complement system, immune response, platelet degranulation, and regulated exocytosis.

Table 3. 3. The enriched pathways from the dysregulated proteins between the healthy and PDAC patients.

Pathways	Enrichment FDR	Number of Genes from the dataset	Protein names	Pathway genes	Fold enrichment
Complement activation, classical pathway	2.0E-09	6	C5, C9, CLU, C3, SERPING1, C1QA	51	149
Humoral immune response mediated by circulating immunoglobulin	4.1E-09	6	C5, C9, CLU, C3, SERPING1, C1QA	64	118.7
Platelet degranulation	7.5E-08	7	CLU, SERPING1, ITIH3, CLEC3B, SERPINA3, SERPINA1	137	55.5
B cell-mediated immunity	1.1E-07	6	C5, C9, CLU, C3, SERPING1, C1QA	148	51.3
Negative regulation of endopeptidase activity	4.4E-09	8	C5, SERPINA7, C3, AGT, SERPING1, ITIH3,	273	37.1

			SERPINA3, SERPINA1		
Regulated exocytosis	5.3E-07	9	TTR, CLU, C3, SERPING1, ITIH3, CLEC3B, LRG1, SERPINA3, SERPINA1	903	12.6
Immune effector process	4.6E-07	10	C5, C9, TTR, CLU, C3, SERPING1, C1QA, LRG1, SERPINA3, SERPINA1	1245	10.2

For a better understanding of the possible interactions of the identified differentially expressed proteins, STRING was used to observe the functional and physical protein-protein interaction networks (Figure 3.15). A high confidence level of 70% in STRING was used in searching these interactions. The network analysis confirmed the pathway analysis whereby some proteins were enriched in pathways linked to immune response, platelet degranulation and the complement system. STRING software was unable to detect specific proteins such as immunoglobulin lambda variable 1-51, immunoglobulin heavy constant alpha 2, immunoglobulin heavy variable 4-28, immunoglobulin heavy variable 4-61.

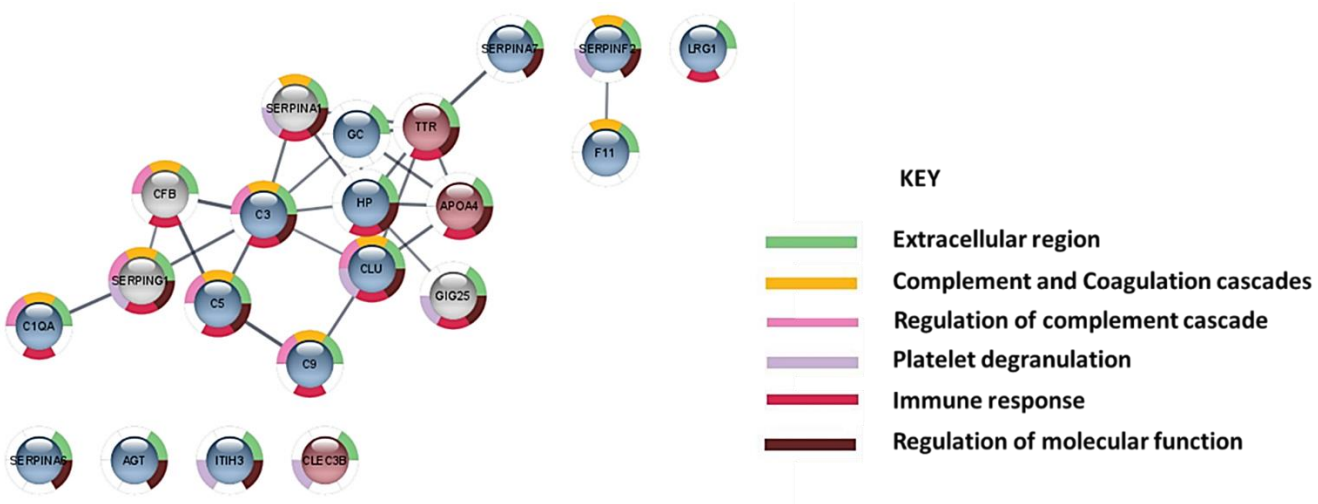
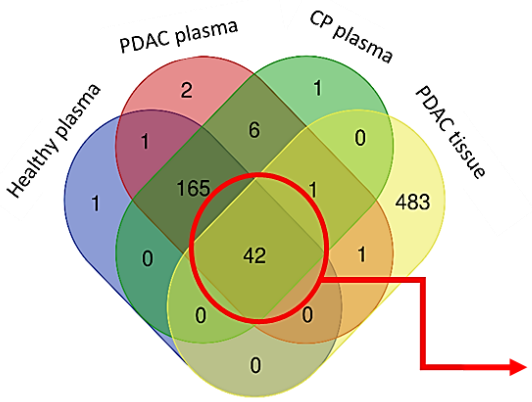


Figure 3. 21. Protein-protein network analysis. Protein network of differentially expressed proteins in comparing PDAC to healthy patients showing their functional and physical networks and the enriched pathways they are involved in. The blue and red nodes (circles) represent downregulation and upregulation, respectively.

In a prior study wherein tissue samples from resectable African PDAC patients were profiled, a total of 527 candidate dysregulated proteins were identified (Nweke et al., 2020). A Venn diagram was conducted to determine if there was any overlap between all quantified proteins found in this pilot study and dysregulated proteins observed in the tissue study, to determine if tissue proteins could be secreted into plasma. A total of 42 dysregulated proteins found in tissue were observed across all the plasma conditions (Healthy, CP, PDAC) (Figure 3.16).



Unit accession number	Protein description
P02656	Apolipoprotein C-III
P02675	Fibrinogen beta chain
P02647	Apolipoprotein A-I
P01023	Alpha-2-macroglobulin
P07225	Vitamin k-dependent protein S
CAH1_HUMAN	Carbonic anhydrase I
P05160	Coagulation factor XIII B chain
P01011	Alpha-1-antichymotrypsin
P02042	Hemoglobin subunit delta
P23142	Fibulin 1
P04004	Vitronectin
P02654	Apolipoprotein C-I
P02655	Apolipoprotein C-II
P04114	Apolipoprotein B-100
P00739	Haptoglobin-related protein
P00734	Prothrombin
LYSC_HUMAN	Lysozyme C
P02775	Platelet basic protein
P02649	Apolipoprotein E
P02671	Fibrinogen alpha chain
P35908	Keratin, type II cytoskeletal 2 epidermal
P0DJI8	Serum amyloid A-1 protein
O75636	Ficolin-3
P01602	Immunoglobulin kappa variable 1-5
P04003	C4b-binding protein alpha chain
O14791	Apolipoprotein L1
P02652	Apolipoprotein A-II
O95445	Apolipoprotein M
P27169	Serum paraoxonase/arylesterase 1
P02776	Platelet factor 4
P00742	Coagulation factor X
P06702	Protein S100-A9
P10909	Clusterin
P08697	Alpha-2-antiplasmin
P00747	Plasminogen
P22792	Carboxypeptidase N subunit 2
Q14624	Inter-alpha-trypsin inhibitor heavy chain H4
P02679	Fibrinogen gamma chain
P35542	Serum amyloid A-4 protein
P55056	Apolipoprotein C-IV
P05546	Heparin cofactor 2
P80108	Phosphatidylinositol-glycan-specific phospholipase D

Figure 3. 22. Venn diagram and table of candidate proteins from tissue and plasma study. The Venn diagram (left) shows the number of candidate proteins identified in healthy plasma, PDAC plasma, CP plasma and tissue plasma. The table (right) shows the 42 similar proteins found in both the plasma and tissue studies. The plasma pilot consists of all candidate proteins and the tissue study consists of dysregulated proteins.

CHAPTER 4

DISCUSSION

The lack of optimization of methods for sample process and downstream analysis can limit the discovery of biomarkers for PDAC. In this study, an optimal preparation and digestion method was determined, and its utility was demonstrated in profiling and identifying differentially expressed proteins in PDAC plasma from patients of African ancestry. The identified proteins were also shown to be involved in pathways that are linked to carcinogenesis.

Section 1: Method optimization

4.1. Plasma contains highly abundant proteins

The SDS-PAGE (Figure 3.1) conducted on a study-specific sample pool revealed some of the well- known high abundance proteins in plasma (albumin, transferrin, α-1 antitrypsin, immunoglobulin heavy and light chain and apolipoprotein A-2) (Nanjappa *et al.,* 2014). These highly abundant proteins tend to mask the low abundant proteins, which could lead to them being undetected. The low-abundant proteins are potentially key targets for the diagnosis of diseases including PDAC. The top 12 abundant proteins in plasma include albumin, Immunoglobulin G (IgG), α1-antitrypsin, α2-macroglobulin, apolipoprotein A-I, apolipoprotein A-II, fibrinogen, haptoglobin, IgA, IgM, α1-acid glycoprotein and transferrin (Table 4.1) and may comprise up to 99% of total protein abundance (Nanjappa *et al.,* 2014). Albumin is synthesized by the liver and translated from a single gene into preproalbumin in hepatocytes (van de Wouw and Joles, 2022). However, little albumin is stored in the liver as most of it is rapidly distributed through the body (Jagdish et al., 2021) and is therefore required in high amounts in the body. Low levels of this protein may be indicative of diseases of the liver and kidney, infections, malnutrition, and cancer (Sheinenzon *et al.,* 2021).

IgG is an antibody that provides immunity against pathogens by enhancing phagocytosis, neutralizing viral/bacterial toxins and activating the complement system (Megha and Mohanan, 2021). Immunoglobulin A (IgA) prevents the colonization and growth of bacteria (Sathe and Cusick, 2021). α1-antitrypsin protects the lungs from proteolytic damage by

hindering neutrophil elastase activity (Abdulkarim and Craig, 2018). Apolipoprotein A-I is responsible for cellular cholesterol (Mangaraj *et al.,* 2016), while apolipoprotein A-II plays an important role in high-density lipoprotein particle synthesis (Xepapadaki *et al.,* 2019).

Transferrin plays a role in the transportation of iron through the blood to the spleen, bone marrow and liver (Ogun and Adeyinka, 2021). Haptoglobin prevents renal damage and iron loss by binding haemoglobin (Wassell, 2000, Liz *et al.,* 2020). Fibrinogen serves as a substrate for three primary enzymes: plasmin, thrombin and factor XIIIA. Through its multiple functional interactions, fibrinogen is involved in the process of homeostasis and aids in facilitating platelet regulation (Kaur *et al.,* 2021). Alpha-2-macroglobulin (α2M) acts as a protective shield against pathogens in the blood and tissues of vertebrates by binding to both host and foreign particles and peptides (Borth, 1992). The primary role of α1-acid glycoprotein is to regulate the immune system by modulating cytokine secretion (Ruiz, 2021). However, because of these proteins' high abundance in the plasma, it is difficult to determine the difference between the diseased and healthy, making them unreliable diagnostic markers for PDAC.

Due to the high dynamic range of plasma samples, there was a need to deplete some of the highly abundant proteins and ultimately improve plasma proteome coverage. This would allow for an increased chance of identifying and quantifying key targets involved in PDAC. Protein depletion was conducted using 2 depletion kits (Pierce Albumin/IgG Removal kit and Pierce Top 12 Abundant Protein Depletion Spin Columns). After depletion, the proteins were visualized using SDS-PAGE (Figure 3.2). Alternative to protein depletion, protein level fractionation was also conducted using the strong anion exchange (SAX) method and some of the lanes (2, 3, 5, 6, 8, 9, 10, 11) showed the depletion of high abundant proteins (Figure 3.3).

Table 4.1. Top 12 abundant proteins and their amounts in blood plasma.

Protein	**Relative protein amount in plasma (%)**
Albumin	50-60
IgG	10-20
IgA	10-15
IgM	5-10
Fibrinogen	4
A2-macroglobulin	3-5
A1-antitrypsin	0.5-1.5
Apolipoprotein A-I	10-20
Apolipoprotein A-II	20-25

Haptoglobin	0.2-0.5
A1-acid glycoprotein	1-3
Transferrin	30-40

4.2. Spectral library building: Depletion and fractionation of abundant proteins

The depletion and fractionation methods were performed to reduce highly abundant proteins. The resultant eluates/fractions were further subjected to reduction and alkylation, hydrophilic interaction liquid chromatography (HILIC) clean-up and digestion, and mass spectrometry analysis. The resulting data from both methods were combined to build a spectral library consisting of 3993 peptides and 280 protein groups (Table 3.1). For data-independent analysis, obtaining a suitable spectral library for the identification and quantification of peptides is one of the most important requisites for protein identification (Barkovits *et al.,* 2020). Spectral libraries are a compilation of known compound reference spectra that have been formerly observed (Wang, 2017). Additionally, the use of an in-house generated library allowed for better ion pattern matching and retention time alignment, which assures increased peptide identification due to instrument and sample specificity (Govaert *et al.,* 2017). The value of a spectral library is generally evaluated based on the library size (Barkovits *et al.,* 2020).

A surplus of highly abundant peptides limits the measurable dynamic range during mass spectrometry analysis, reduces chromatographic resolution and causes identification bias to highly abundant proteins in shotgun proteomics (Fonslow *et al.*, 2013). Affinity depletion and pre-fractionation of proteins/peptides are common ways to address this (Fonslow *et al.,* 2013). Affinity depletion specifically captures target proteins using immobilized antibodies or other molecules (for example cibacron blue and protein A/G) with increased specificity and affinity for the targets (Jankovska *et al.,* 2019).

Two depletion kits were used and the unpaired t-test comparing the top 2 and top 12 methods, displayed a significant difference in the number of peptides ($P < 0.0001$) and protein groups ($P < 0.0001$) identified by the two methods (Figure 3.4). The top 12 depletion identified a higher number of peptides (mean: 3050) and protein groups (mean: 245) meaning the method removed more highly-abundant proteins compared to the top 2 depletion (peptide mean: 1529, protein group mean: 200). This is further corroborated by a study conducted by Pisanu et al. (2018), comparing seven commercial products for depleting proteins in human serum.

The top 2 and top 12 depletion spin columns identified the highest number of peptides compared to other products, obtaining 40.7% and 60.8% more peptides than untreated serum respectively (Pisanu *et al.*, 2018). The Pierce™ Albumin/IgG Removal kit and the Pierce™ Top 12 Abundant Protein Depletion spin columns contained a resin with highly specific immobilized antibodies for the removal of proteins. Both methods can remove up to 95% of the respective high abundant proteins. However, the depletion of highly abundant proteins may also remove low- abundant proteins of interest (Lee *et al.*, 2014). The difference in performance between the two depletion methods may be due to binding capacity. The top 2 depletion method is designed to only deplete the two most abundant proteins, while the top 12 depletion method targets a broader range of abundant proteins. The complexity of the sample being processed can impact the performance of the columns. If the sample contains a high number of different abundant molecules, the top 12 depletion spin column, with a broader specificity, may be more effective in depleting a larger number of targets

The plasma samples were fractionated at the protein level by increasing NaCl concentration. The SAX method independently identified an average of 1657 peptides and an average of 172 protein groups (Figure 3.5). The results from each fraction were combined with the results from the depleted and non-depleted plasma samples, to enhance the spectral library. Sodium chloride is commonly utilized as a mild eluent for protein separation since it does not significantly impact the structure of proteins, making it the most extensively used option (Acikara, 2013). SAX enables the capture, purification, and recovery of plasma proteins by utilizing the differences in net charge to reversibly adsorb oppositely charged molecules. A salt gradient was utilized to displace the bound analytes on the column surface as the ions in the elution buffer compete for the charged sites. Lower surface charge density analytes are eluted at lower salt concentrations, whereas higher surface charge density analytes are eluted as the salt concentration is increased (Kailasam).

4.3. Plasma sample preparation for protein analysis

Two methods were compared and evaluated including protein dilution (using Tris-HCl with 2% SDS, and TEAB, 2% SDS buffer) and acetone precipitation (100% acetone and 80% acetone). Before downstream analysis, the total ion chromatograms (TICs) (Figure S1-S4) were evaluated and thereafter the data was normalized (Figure 3.6). TICs were evaluated to determine the total signal range across different samples and replicates. This evaluation was conducted to establish if there was a correlation between the signal and the number of loaded

peptides. The samples showed no over/underloading by observing the overlay of the replicate signals. The use of TICs provided significant benefits in data processing, as it eliminated the need for time-consuming compound identification and the use of complex libraries or standards. Moreover, TIC analysis does not require specialized knowledge from an analyst (Barea-Sepúlveda et al., 2022). The significant advantage of this approach was that samples with lower ion intensities could still be included in statistical tests without being identified as outliers, ensuring their fair consideration (O'Rourke et al., 2019).

Acetone precipitation is commonly used for the concentration and precipitation of proteins (Damerval *et al.,* 1986). Acetone precipitation minimizes protein degradation, protease activity and reduction of contaminants in samples (Shaw and Riederer, 2003). There was a significant difference in the peptide identifications between the Tris-HCl with 2% SDS and TEAB with 2% SDS (P = 0.0007) and between 80% acetone precipitation and TEAB with 2% SDS (P = 0.0220) dilution methods (Figure 3.9). For the protein groups, a significant difference was observed between the Tris-HCl with 2% SDS and TEAB, 2% SDS (P = 0.0264) (Figure 3.9). However, all the dilution methods used identified similar numbers of peptide and protein groups. Due to the sample replicates being processed over two different days, it was important to observe CVs to determine the reproducibility of the methods. The 100% and 80% acetone precipitation yielded higher median CVs (32% and 25% respectively) across its replicates compared to the dilution methods (Figure 3.8). The acetone precipitation methods, therefore, showed less reproducibility.

The statistical metric known as the coefficient of variation (CV) was used to measure the spread of data points in the data set around their mean. It was calculated by dividing the standard deviation by the mean (Reed et al., 2002) and was a valuable tool for comparing the level of variation between two different data sets, even when their means differ significantly. Acetone precipitation was not selected as the method of choice because it provided no observable benefit for the increased number of steps in its protocol. Furthermore, acetone precipitation requires a prolonged centrifugation step which is not readily automatable. With sample preparation already being time-consuming and automation being the goal, more steps in the protocol are unfavourable and will not be beneficial for large-scale studies.

Proper sample preparation is an important step for mass spectrometry (MS)-based proteomics because it can be simultaneously time-consuming and variable. MS results are highly impacted by the reproducibility and quality of sample preparation and extraction. (Luque-

Garcia and Neubert, 2007). Contaminants and interfering species from samples may negatively impact ionization during the electrospray process, peptide chromatography and mass spectrometry sequencing (Batth *et al.,* 2019). Therefore, it is essential to have a standardized sample preparation method that can remove these impurities for plasma profiling to attain reliable biomarkers. Small changes in a sample preparation technique can lead to unreproducible and variable results (Villanueva *et al.,* 2004).

4.4. Sample dilution, clean-up and protein digestion

The different dilution buffers (25X dilution with 50 mM Tris-HCl, 2% SDS pH 8 and 25X dilution with 50 mM TEAB, 2% SDS pH 8) were further compared and evaluated in addition to the comparison of hydrophilic interaction liquid chromatography (HILIC) and protein aggregation capture (PAC) clean-up methods. TEAB dilution with HILIC had significantly higher peptides identified (mean: 3206) compared to Tris dilution with PAC (mean: 3046) depicted by P = 0.0008. Additionally, Tris dilution with HILIC identified significantly higher peptides (mean: 3141) in comparison to Tris dilution with PAC (mean: 3046) represented by P = 0.0439 (Figure 3.12). The 25X dilution with 50 mM Tris-HCl, 2% SDS was selected as the standard dilution procedure as this method showed to be reproducible (Figure 3.10; Figure 3.11) and Tris (hydroxymethyl) aminomethane was more affordable and readily available in our laboratory. however, a 25X dilution with 50 mM TEAB, 2% SDS dilution is an alternative. The solubilization of proteins can be difficult to attain and frequently requires the use of strong salts and/or detergents that are incompatible with subsequent MS analysis (Yeung *et al.,* 2008). One of the most widely used reagents for the solubilization of biological components is SDS. This ionic detergent can attach to proteins via hydrophobic and ionic interactions. SDS solubilizes proteins by adding a negative charge to amino acids thereby changing their secondary and tertiary structures, and since like charges repel the proteins begin to linearize and become inactive (Figure 4.4.1) (Andersen *et al.,* 2009).

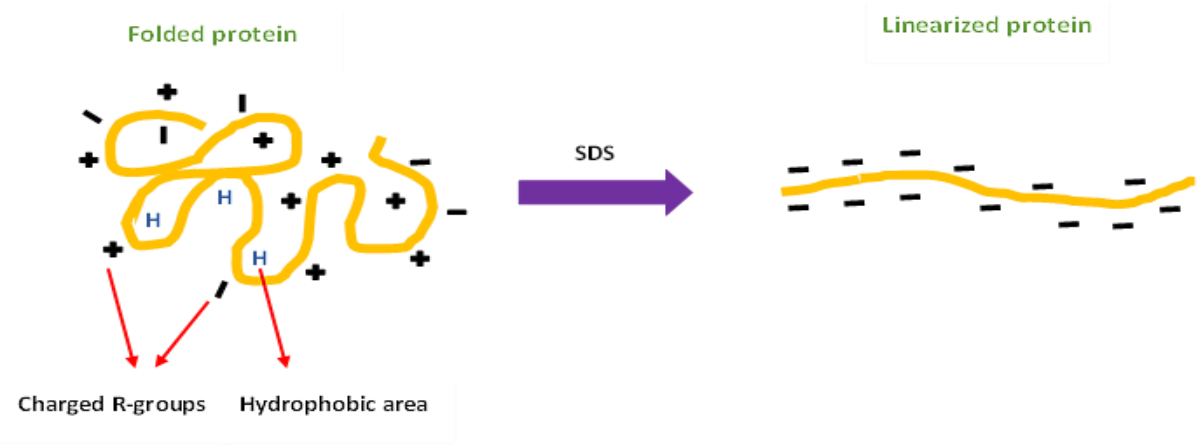


Figure 4.4. 1. Protein denaturation by SDS. The folded protein initially consists of positive and negative charges due to the charged R-groups. The hydrophobic domains are where the non-polar R-groups collect to escape the polar water surrounding the protein. SDS disrupts the hydrophobic areas and confers a net negative charge. The protein is denatured and results in a linearized protein (primary structure).

Studies on aggregated proteins and membrane proteins, which are typically challenging to solubilize by conventional reagents, benefit from the use of SDS (Zhou *et al.*, 2012). The most widely used form of extraction for plasma proteomics studies are dilutions with buffering salts and/or detergents and precipitation methods. The Tris-HCl buffer (pH range 7.0-9.0) can extract soluble cytoplasmic proteins, diluting out sample contaminants and maintaining a pH environment over a wide temperature range.

TEAB buffer is a volatile buffer system that is often used in mass spectrometry-based proteomics workflows for protein dilution and digestion. TEAB buffer is particularly useful because it is volatile and can be removed easily by evaporation, which is important for downstream mass spectrometry analysis. It has a pH range of 7.5-8.5, which is optimal for trypsin digestion. The only limitation of TEAB is it is more costly than Tris-HCl.

Following protein dilution, it is crucial to remove detergents (e.g., SDS) and contaminants as some might interfere with the enzymatic digestion and mass spectrometry analysis (Zhou et al., 2012). SDS causes significant analyte suppression in electrospray MS (Botelho *et al.*, 2010). A comparison of HILIC and PAC clean-up and digestion methods was conducted to remove contaminants and improve digestion efficiency.

HILIC and PAC are two methods that can be used for the removal of contaminants before protein digestion and MS analysis. The PAC method utilizes the principle of non-specifically binding precipitated and aggregated proteins to sub-micron particles regardless of their

surface chemistry (Batth *et al.,* 2019). In the context of protein binding, HILIC was used to separate proteins based on their surface charge and hydrophilicity. In HILIC, a hydrophilic stationary allowed for interactions between the stationary phase and the hydrophilic regions of the protein. The mobile phase, which was a mixture of water and an organic solvent (acetonitrile) which interacts with the hydrophobic regions of the protein (Boersema *et al.,* 2008) and trifluoroacetic acid used to elute the protein of interest. Both HILIC and PAC were performed using magnetic beads. (Batth et al., 2019)Magnetic beads were used because they have the following benefits: easy to handle, linear scalability and high-throughput compatibility. Automation of workflows is easily achieved by either using a magnetic bead handling system (KingFisher™) or integrating a magnetic stand into different liquid handling stations.

Although HILIC yielded significantly higher peptide identifications compared to PAC (Figure 3.12), both methods showed to be reproducible in terms of low CVs between the different days and between technical replicates (Figures 3.10-3.11). Therefore, either method is suitable for plasma proteomics analysis. The HILIC method was selected because our laboratory has more historic data using this method, including the previous study on PDAC tissue proteomics.

Digestion time was compared and evaluated to improve protein coverage. The 4 hrs HILIC digestion obtained significantly higher peptides (mean: 2860) compared to the 16 hrs digestion (mean: 2699) depicted by P = 0.0216 (Figure 3.15). However, there was still room for improvement, therefore digestion times (2 hrs vs 4 hrs vs 16 hrs) were further compared in addition to enzyme efficiency (trypsin vs trypsin+LysC). Trypsin is one of the commonly used enzymes for MS due to its availability, high cleavage specificity and low cost. Trypsin cleaves at the C-terminus of arginine or lysine residues unless there is a proline residue present (Siepen *et al.,* 2007). LysC addresses trypsin proteolytic inefficiency at lysine sites. The benefits of the trypsin+LysC mixture for protein mass spectrometry include allowing more accurate protein quantitation, increasing the number of identified proteins and peptides of complex protein samples and improving analytical reproducibility (Saveliev *et al.,* 2013). This is further supported by results from this study where the trypsin+LysC 2 hrs digestion had significantly higher peptides (mean: 2856) compared to trypsin only 16 hrs digestion (mean: 2700) with a P = 0.0025 (Figure 3.18).

The use of trypsin combined with alternative proteases (e.g., LysC, Asp-N, chymotrypsin, Lys-N, LysargiNase and Glu-C) was reported to improve the number of peptides per protein and sequence coverage, mostly for low abundant proteins (Giansanti *et al.*, 2016, Swaney *et al.*, 2010). Saveliev et al. (2013) conducted a model experiment in which they digested HeLa protein extract using either trypsin or a trypsin+LysC mix. The trypsin digest resulted in a decrease of about 11% in the number of identified proteins compared to the trypsin+LysC digest (Saveliev *et al.*, 2013).

Shorter digestion times may be more effective than longer ones for identifying peptides, but the reason for this improvement is uncertain. Autolysis of trypsin can also diminish its enzyme activity and site-specificity over time (Klammer and MacCoss, 2006). Prolonged trypsin digestion methods are time-consuming and are not suitable for high-throughput or automated sample preparation and analysis. Additionally, it results in a reduction in amino acid coverage of proteins and consequently loss of tryptic peptides (Hildonen *et al.*, 2014). Therefore, the final standard operating procedure (Figure 3.13) using 50 mM Tris-HCl, 2% SDS, pH 8 to dilute plasma 25 times followed by HILIC clean-up and 2 hrs on-bead digestion at 47 ℃ with a trypsin+LysC mix was tested in a pilot study.

Section 2: Pilot study

4.5. Dysregulated proteins may be involved in PDAC development

In the pilot study, proteomic profiling was performed to test the utility of the SOP. Differentially expressed proteins were identified in PDAC plasma samples obtained from patients of African ancestry. Comparing PDAC to healthy patients, the top 3 (in abundance) upregulated proteins included Transthyretin, Apolipoprotein A4, and tetranectin. The top 3 (in abundance) downregulated proteins included Haptoglobin-related protein, Complement C1q A chain, and Immunoglobulin heavy constant alpha 2 (Figure 3.20A, Table 3.2). The top 2 upregulated proteins (transthyretin and apolipoprotein A4) have been identified in other literature as potential biomarkers for PDAC development (Harsha *et al.*, 2009, Litman-Zawadzka *et al.*, 2019) validating that the optimized protocol succeeded in identifying proteins associated with PDAC. The identification of these proteins using the SOP developed suggests it can be used to profile plasma proteins and possibly identify key targets associated with PDAC. Some of the top dysregulated proteins that have been linked to PDAC or cancer development are discussed below.

Transthyretin is crucial for the binding and distribution of thyroid hormones, maintaining blood levels of vitamin A and for the growth of the liver and pancreas (Schweigert *et al.,* 2004). Lv et al. (2011) identified a >2-fold upregulation of this protein in pancreatic juice compared to controls and immunohistochemistry staining demonstrated an overexpression of transthyretin in PDAC tissues (Lv *et al.,* 2011). However, in other studies, the levels of this protein were found to be downregulated in PDAC, but these discrepancies may be explained by differences in biosamples used and sample population (Chen *et al.,* 2013). In another study, it was discovered that levels of transthyretin were upregulated in sera of patients with ovarian cancer (early stage) using the ProteinChip Biomarker system (Chen *et al.,* 2013).

Studies have discovered that mutant apolipoproteins can cause aberrant blood lipid utilization and metabolism which plays significant roles in the onset and progression of atherosclerosis, hyperlipidaemia, cardiovascular diseases, and malignancies (Ren *et al.,* 2019). The authors suggest that these proteins may serve as therapeutic targets (by inhibiting expression/ using corresponding mimetic peptides) as well as aid in cancer diagnosis and prognosis (Ren *et al.,* 2019). For example, the utility of ApoA2 as an early marker of pancreatic cancer was demonstrated when compared to healthy individuals (Sato *et al.,* 2020).

An increase in the expression of C1q, the initial component of the complement classical pathway, has been suggested as a prognostic marker in PDAC patients, emphasizing the significance of the complement system in the advancement of tumours (Speth *et al.,* 2022). It can be produced by different cells within the tumour microenvironment including tumour cells, macrophages and dendritic cells (Ain *et al.,* 2021). Complement C1q has several immunoregulatory functions, such as enhancing phagocytosis, regulating cytokine production, and promoting T-lymphocyte maturation. However, recent studies have shown that C1q can also stimulate cancer development and proliferation by promoting pro-tumorigenic processes such as angiogenesis (Ain *et al.,* 2021).

This study identified only 2 (Afamin and Immunoglobulin heavy variable 3-43) proteins that were differentially expressed in PDAC compared to CP patients (Figure 3.20B). This may be due to similarities in some molecular mechanisms associated with both diseases as CP is a risk factor for PDAC development. In individuals with CP, prolonged inflammation contributes to the proliferation of stellate cells and cell turnover, resulting in the formation of a microenvironment in the pancreas that facilitates carcinogenesis (Gandhi *et al.,* 2022). If dysregulated proteins cannot be accurately distinguished from those present in chronic

pancreatitis (CP), it may lead to delayed diagnosis and the missed opportunity for early intervention and potentially curative treatments. Additionally, it may limit the accuracy and reliability of prognostic indicators, making it challenging to predict disease progression and response to therapy. Afamin was reported to be decreased in ovarian cancer and has been proposed to be a potential biomarker (Jackson *et al.,* 2007). IGHV mutations have been observed to have a prognostic impact on cancer patients including chronic lymphocytic leukemia (Kaufman *et al.,* 2022).

4.6. Dysregulated proteins are enriched in key pathways involved in carcinogenesis

Pathway enrichment analysis was performed on dysregulated proteins using ShinyGo and Reactome software. Several pathways were found to be enriched including those related to the complement system, immune response, platelet degranulation, regulated exocytosis and negative regulation of endopeptidase. The enrichment of the complement system and immune response-related pathways were due to the downregulation of specific proteins including C5, C9, CLU, C3, serpin family G member 1 (SERPING1) and C1QA. The complement system serves as one of the initial defence mechanisms against stressed host cells or invading pathogens. It comprises three independent but interconnected pathways (alternative pathway, lectin pathway and classical pathway) and operates as a coordinated and highly controlled biological system (Ain *et al.,* 2021). The activation of the classical pathway occurs following the attachment of the C1 complex to targets such as an antigen-antibody (IgG or IgM) immune complex or an apoptotic cell. The C1 complex comprises C1q molecules and two distinct serine proteases (Revel *et al.,* 2020). Upon analysing complement gene expression in 30 different types of cancer, several genes responsible for the classical pathway were upregulated (Roumenina *et al.,* 2019). The mechanisms by which the complement system is activated in tumours remain unclear, although various models suggest that both classical and alternative pathway components are present in higher concentrations within the tumour microenvironment (Revel *et al.,* 2020).

The involvement of serine proteases and their inhibitors is frequently observed in different types of cancers, as well as in stromal cell behaviours that promote the growth of primary tumours (Zheng *et al.,* 2013). In a study by Peng et al. (2018), SERPING1 was found to be reduced in prostate cancer tissues compared to non-cancerous tissues (Peng *et al.,* 2018), and reported to be a potential early diagnostic marker for lung and breast cancer (Lee *et al.,* 2016, P *et al.,* 2004). *SERPING1* encodes for the C1-inhibitor, which can hinder the activation of

C1r and C1s, inhibiting the formation of the C1 complex, and consequently decreasing the activation of the classical complement pathway (Carter *et al.,* 1991). A deficiency of complement C5 effectively inhibits the progression of colorectal cancer by attracting myeloid-derived suppressor cells to the inflamed colorectal region and weakening the activity of CD8 (+) regulatory T (Treg) cells (Ding *et al.,* 2020). Most cell types within a tumour may express complement C3aR receptors, indicating that C3a may potentially be utilized by tumour cells to facilitate tumour growth (Revel *et al.,* 2020). The function of C9 in complement-dependent cytotoxicity-induced tumour cell death is essential, and the downregulation of C9 expressed by tumour-associated macrophages due to hypoxia contributes to the advancement of non-small cell lung cancer (Zhang *et al.,* 2019).

The identification of SNPs rs17091162 in SERPINA3 has revealed a potential correlation with the survival rates of pancreatic cancer patients, indicating the significant impact of SERPINA3 on disease prognosis in cancer (Zeng *et al.,* 2011). Tumour cells produce and secrete SERPINA1, which is implicated in several physiological and pathological mechanisms such as angiogenesis, tumour invasion, and metastasis (Chantada-Vázquez *et al.,* 2022).

The complement system is a vital component of innate immunity, with a key role in initiating the inflammatory response and regulating the function of T and B cells. Additionally, once pathogen-specific antibodies are produced, the complement system aids in removing immune complexes and eliminating the pathogen (Merle *et al.,* 2015). The microenvironment of PDAC comprises a substantial number of immune cells however, patients with this condition exhibit immune dysfunction. This can be attributed to an immunosuppressive tumour microenvironment, which hinders the activation or effectiveness of immune effectors (Chang *et al.,* 2016). In this study, one of the proteins enriching the immune-response pathway was CLU which was observed to be downregulated. CLU has been demonstrated to play significant roles in several physiopathological processes and observed to be altered in different human carcinomas (Redondo *et al.,* 2009). Its downregulation is corroborated by studies that similarly observed this trend (Sala *et al.,* 2009). CLU may contribute to the development of cancer by inhibiting apoptosis, activating autophagy, and affecting several signalling pathways including IGF-1/IGFR, EGFR, NF-kB, PI3K/AKT, TGFβ, and specific miRNAs (Sala et al., 2009). A study found that the signalling pathways facilitated by C5AR1 promoted the conversion of the tumour microenvironment into a conducive environment for

tumour growth by activating innate immune cells, suppressing effector T cells, and releasing factors that support tumour growth (Gadwa *et al.,* 2021).

Platelet degranulation-related pathways were attributed to the upregulation of CLEC3B and downregulation of proteins including CLU, SERPING1, ITIH3, SERPINA3, and SERPINA1. When platelets are activated, they release growth factors and bioactive lipids into the bloodstream, which recruits and aggregates more platelets and initiates thrombogenesis. This process is essential for wound healing and detecting abnormal cells in the bloodstream. In the case of cancer cells, platelet-mediated recognition is triggered and amplified by various factors, leading to suppression of immune recognition and elimination, arrest at endothelium, or entrapment in the microvasculature. These interactions support cancer cell survival and spread, leading to the establishment of secondary lesions, which are important targets for prevention and therapy (Menter *et al.,* 2014). In general, pancreatic tumours tend to promote platelet aggregation (Mai and Inkielewicz-Stepniak, 2021), leading to the formation of blood clots. PDAC is recognized as one of the most thrombogenic types of cancer and interacts with platelets in both direct and indirect ways to stimulate their activation, degranulation, and aggregation. These processes, in turn, facilitate tumour proliferation and metastasis (Palacios-Acedo *et al.,* 2022). Nweke et al. (2020) observed elevated levels of platelet degranulation-related pathways (Nweke *et al.,* 2020). Although limited information is available on the involvement of ITIH3 in platelet function and thrombosis, this protein has been demonstrated to have important roles in inflammation and inhibiting cancer metastasis (de Almeida *et al.,* 2022).

The enrichment of exocytosis-related pathways was due to the downregulation of TTR, CLU, C3, SERPING1, ITIH3, LRG1, SERPINA3, and SERPINA1 and the upregulation of CLEC3B. As cancer advances, lysosomes tend to shift towards the plasma membrane due to alterations in either the cytoskeletal network or the trafficking and exocytosis of lysosomes (Machado *et al.,* 2021). Cancer cells exploit lysosomal exocytosis to become aggressive and resistant to treatment. This process empowers cancer cells in various ways, including altering the plasma membrane composition to influence signalling events, secreting active enzymes to remodel the surrounding matrix and activate stroma cells, enhancing the release of exosomes to propagate signalling molecules to neighbouring cells, and promoting the efflux of lysosomotropic chemotherapeutics to confer drug resistance (Machado *et al.,* 2021). A study by Kwan et al. (2021) revealed that LRG1 controls several facets of melanoma metastasis by modulating EGFR/STAT3 signalling (Kwan *et al.,* 2021). This protein has been suggested to

be a potential prognostic marker for pancreatic cancer (Furukawa *et al.,* 2015). In recent reports, LRG1 has been found to play a role in tumorigenesis beyond angiogenesis, specifically in regulating epithelial-mesenchymal transition by activating RUNX1 (Lu et al., 2020) and TGFβ/Smad2,3 signalling pathways (Zhang *et al.,* 2020).

4.7. Significantly dysregulated proteins in plasma may be secreted by tumours

To identify any similar proteins in the datasets of plasma obtained from the current study and tissue from a previous study by Nweke et al. (2020), a Venn diagram was used (Figure 3.22). The expression of similar proteins at the tissue and protein level may reinforce their possible role in pancreatic tumourigenesis. A total of 42 differentially expressed tissue proteins were observed across all plasma samples (PDAC, CP, healthy individuals). Some of the proteins were observed to be expressed in PDAC in other studies. Amongst these are members of the Apolipoprotein family (APOA1, APOA2, APOC1, APOC2, and APOE). The human apolipoprotein gene has 22 members which are characterized into 10 subfamilies (APOA-APOJ). Apolipoproteins can bind to lipids forming lipoproteins because they function as lipid carriers, these proteins act as ligands for structural components of lipoproteins, cell membrane receptors and cofactors of enzymes (Kastelein *et al.,* 2008). Apolipoproteins are components of plasma lipoproteins that bind and transport blood lipids to various body tissues where they are used for metabolism (Di Angelantonio *et al.,* 2009). They are known to be associated with cancers and have important functions in this disease through multiple mechanisms., For example, a previous study reported increased levels of Apolipoprotein CI (APOC1), Apolipoprotein CII (APOC2), and Apolipoprotein E (APOE) in serum and tissue of pancreatic cancer patients (Ren *et al.,* 2019).

Several proteins that are mainly produced by the liver and have been implicated in carcinogenesis were commonly dysregulated in this study and the tissue dataset (Figure 3.22). They include fibrinogen (FGB), alpha-2-macroglobulin (A2M), alpha-1-antichymotrypsin (SERPINA3), C4b-binding protein alpha chain (C4BPA), haptoglobin (HP) and serum amyloid (SAA1). Fibrinogen is a glycoprotein comprising two sets of three polypeptide chains (a, b, c), and is responsible for the clotting and circulation of blood through platelet interaction (Ren *et al.,* 2019). There has been recent evidence indicating plasma fibrinogen to have a critical function in tumour progression (Ren *et al.,* 2019). The fibrinogen beta and gamma chains were reported to be uniquely and notably expressed in pancreatic cancer sera (Bloomston *et al.,* 2006). The C4b-binding protein alpha chain is a component of the classical

pathway and controls inflammation due to bacterial infections and has been reported to be a potential serum marker in differentiating PDAC from CP (Sogawa *et al.*, 2016, Sasaki *et al.*, 2021). In addition, haptoglobin and serum amyloid have been observed to be elevated in PDAC (Firpo *et al.*, 2009). Haptoglobin was found to be upregulated in the plasma dataset and serum amyloid was downregulated in both the tissue and plasma data. It's expected that proteins produced by the liver are aberrant in PDAC as liver functioning is impaired, and almost half of PDAC patients have liver metastases (Saki *et al.*, 2021).

A2M is one of the highly abundant serum proteins in humans. A2M is a protease inhibitor and has been observed to hinder tumour cell proliferation when binding to low-density lipoprotein receptor-related protein 1 (Lin *et al.*, 2014). This protein was found to be elevated and downregulated in the plasma and tissue data respectively (Figure 3.20). Cancer cells can secrete proteins into the bloodstream, leading to their increased presence in plasma. Proteolytic enzymes, such as matrix metalloproteinases, can degrade proteins within the tumor microenvironment. This degradation may lead to a decrease in the levels of certain proteins within the tissue while fragments or cleaved forms of these proteins may be detected in the plasma (Kessenbrock et al., 2010). However this does not discredit the protein as being a potential biomarker. The protein can be validated in independent patient cohorts. Alpha-1-antichymotrypsin, a protease inhibitor, is also produced in the brain and aorta and secreted in astrocytes. It functions in inflammation, acute phase response and proteolysis and may be a potential biomarker for the diagnosis of pancreatic cancer (Jin *et al.*, 2022).

Proteins linked to the extracellular matrix were also found to be dysregulated in the tissue and plasma datasets including fibulin and vitronectin. Fibulin proteins are part of the extracellular matrix proteins and regulate cell migration and proliferation, and fibulin 1 has been reported to be associated with tumour growth (Gong *et al.*, 2020). Another glycoprotein known as vitronectin can link cells and the extracellular matrix via ligands such as plasminogen activator inhibitor 1, integrin and urokinase plasminogen activator receptor (Burgos-Panadero *et al.*, 2019). By binding to these ligands, vitronectin can promote matrix degradation and cell adhesion in the matrix of tumours (Madsen *et al.*, 2007). The protein is found in plasma and the extracellular matrix of some tissues and is produced mainly in the liver. Tumour cells secrete vitronectin which is known to be associated with metastasis, tumour invasion and angiogenesis in some humans (Burgos-Panadero *et al.*, 2019).

CHAPTER 5

Conclusion and future work

The study identified an optimal protein preparation, clean-up and digestion method for patient samples, which were reproducible, be easily automated, increased the number of proteins being identified by LC-MS analysis and allowed for comprehensive proteome profiling of the pilot PDAC cohort. The plasma dilution methods produced similar results to the precipitation methods tested and were preferred over the latter as they require less additional manual processing steps. The pilot study which used the optimized method identified differentially expressed proteins associated with PDAC. However between the CP and PDAC patients there was a lack of distinctive differentially expressed proteins. This may pose challenges in accurately differentiating between these two conditions based solely on protein expression patterns. Further highlighting the need for additional diagnostic markers. The dysregulated proteins showed enrichment of pathways associated with tumorigenesis, suggesting that the method could be reliable in identifying proteins associated with PDAC, especially in our population group. Due to a low sample size for the pilot study, future studies utilizing this SOP should involve a larger sample size. Additionally, future large-scale studies should aim to automate the entire sample preparation process which will be critical for reducing human error and variation. Although a significant number of proteins associated with PDAC were identified, other new possible approaches can be explored to increase protein coverage, such as studying extracellular vesicles (EV) from plasma. Profiling EVs for the identification of biomarkers is important for non-invasive, specific, and reliable diagnosis, monitoring, and treatment of various diseases such as cancer.

REFERENCES

ABDULKARIM, A. & CRAIG, T. J. 2018. Alpha 1 Antitrypsin Mutation.

ACIKARA, Ö. B. 2013. Ion-exchange chromatography and its applications. *Column chromatography,* 10**,** 55744.

ADAMSKA, A., DOMENICHINI, A. & FALASCA, M. 2017. Pancreatic Ductal Adenocarcinoma: Current and Evolving Therapies. *Int J Mol Sci,* 18.

AIER, I., SEMWAL, R., SHARMA, A. & VARADWAJ, P. K. 2019. A systematic assessment of statistics, risk factors, and underlying features involved in pancreatic cancer. *Cancer Epidemiol,* 58**,** 104-110.

AIN, D., SHAIKH, T., MANIMALA, S. & GHEBREHIWET, B. 2021. The role of complement in the tumor microenvironment. *Fac Rev,* 10**,** 80.

ALLOTT, E. H. & HURSTING, S. D. 2015. Obesity and cancer: mechanistic insights from transdisciplinary studies. *Endocr Relat Cancer,* 22**,** R365-86.

AMRUTKAR, M. & GLADHAUG, I. P. 2017. Pancreatic cancer chemoresistance to gemcitabine. *Cancers,* 9**,** 157.

ANDERSEN, K. K., OLIVEIRA, C. L., LARSEN, K. L., POULSEN, F. M., CALLISEN, T. H., WESTH, P., PEDERSEN, J. S. & OTZEN, D. 2009. The role of decorated SDS micelles in sub-CMC protein denaturation and association. *J Mol Biol,* 391**,** 207-26.

ANSARI, D., ARONSSON, L., SASOR, A., WELINDER, C., REZELI, M., MARKO-VARGA, G. & ANDERSSON, R. 2014. The role of quantitative mass spectrometry in the discovery of pancreatic cancer biomarkers for translational science. *J Transl Med,* 12**,** 87.

ASLAM, B., BASIT, M., NISAR, M. A., KHURSHID, M. & RASOOL, M. H. 2017. Proteomics: Technologies and Their Applications. *J Chromatogr Sci,* 55**,** 182-196.

BAREA-SEPÚLVEDA, M., DUARTE, H., ALIAÑO-GONZÁLEZ, M. J., ROMANO, A. & MEDRONHO, B. 2022. Total Ion Chromatogram and Total Ion Mass Spectrum as Alternative Tools for Detection and Discrimination (A Review). *Chemosensors,* 10**,** 465.

BARKOVITS, K., PACHARRA, S., PFEIFFER, K., STEINBACH, S., EISENACHER, M., MARCUS, K. & USZKOREIT, J. 2020. Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition. *Mol Cell Proteomics,* 19**,** 181-197.

BATTH, T. S., TOLLENAERE, M. X., RÜTHER, P., GONZALEZ-FRANQUESA, A., PRABHAKAR, B. S., BEKKER-JENSEN, S., DESHMUKH, A. S. & OLSEN, J. V. 2019. Protein Aggregation Capture on Microparticles Enables Multipurpose Proteomics Sample Preparation. *Mol Cell Proteomics,* 18**,** 1027-1035.

BECKER, A. E., HERNANDEZ, Y. G., FRUCHT, H. & LUCAS, A. L. 2014. Pancreatic ductal adenocarcinoma: risk factors, screening, and early detection. *World J Gastroenterol,* 20**,** 11182-98.

BENJAMIN, O. & LAPPIN, S. L. 2022. Chronic Pancreatitis. *StatPearls.* Treasure Island (FL): StatPearls Publishing

Copyright © 2022, StatPearls Publishing LLC.

BENJAMIN, O. & LAPPIN, S. L. 2023. Chronic Pancreatitis. *StatPearls.* Treasure Island (FL) ineligible companies. Disclosure: Sarah Lappin declares no relevant financial relationships with ineligible companies.: StatPearls Publishing

Copyright © 2023, StatPearls Publishing LLC.

BLONDER, J., ISSAQ, H. J. & VEENSTRA, T. D. 2011. Proteomic biomarker discovery: it's more than just mass spectrometry. *Electrophoresis,* 32**,** 1541-8.

BLOOMSTON, M., ZHOU, J. X., ROSEMURGY, A. S., FRANKEL, W., MURO-CACHO, C. A. & YEATMAN, T. J. 2006. Fibrinogen gamma overexpression in pancreatic cancer identified by large-scale proteomic analysis of serum samples. *Cancer Res,* 66**,** 2592-9.

BOERSEMA, P. J., MOHAMMED, S. & HECK, A. J. 2008. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. *Anal Bioanal Chem,* 391**,** 151-9.

BORTH, W. 1992. Alpha 2-macroglobulin, a multifunctional binding protein with targeting characteristics. *Faseb j,* 6**,** 3345-53.

BOTELHO, D., WALL, M. J., VIEIRA, D. B., FITZSIMMONS, S., LIU, F. & DOUCETTE, A. 2010. Top-down and bottom-up proteomics of SDS-containing solutions following mass-based separation. *J Proteome Res,* 9**,** 2863-70.

BOYS, E. L., LIU, J., ROBINSON, P. J. & REDDEL, R. R. 2022. Clinical applications of mass spectrometry-based proteomics in cancer: Where are we? *Proteomics***,** e2200238.

BRUNE, K. A., LAU, B., PALMISANO, E., CANTO, M., GOGGINS, M. G., HRUBAN, R. H. & KLEIN, A. P. 2010. Importance of age of onset in pancreatic cancer kindreds. *J Natl Cancer Inst,* 102**,** 119-26.

BURGOS-PANADERO, R., NOGUERA, I., CAÑETE, A., NAVARRO, S. & NOGUERA, R. 2019. Vitronectin as a molecular player of the tumor microenvironment in neuroblastoma. *BMC Cancer,* 19**,** 479.

BURSTEIN, H. J. 2005. The distinctive nature of HER2-positive breast cancers. *N Engl J Med,* 353**,** 1652-4.

CAMPOS MUNOZ, A., JAIN, N. K. & GUPTA, M. 2022. Albumin Colloid. *StatPearls.* Treasure Island (FL): StatPearls Publishing

Copyright © 2022, StatPearls Publishing LLC.

CAO, L., HUANG, C., CUI ZHOU, D., HU, Y., LIH, T. M., SAVAGE, S. R., KRUG, K., CLARK, D. J., SCHNAUBELT, M., CHEN, L., DA VEIGA LEPREVOST, F., EGUEZ, R. V., YANG, W., PAN, J., WEN, B., DOU, Y., JIANG, W., LIAO, Y., SHI, Z., TEREKHANOVA, N. V., CAO, S., LU, R. J., LI, Y., LIU, R., ZHU, H., RONNING, P., WU, Y., WYCZALKOWSKI, M. A., EASWARAN, H., DANILOVA, L., MER, A. S., YOO, S., WANG, J. M., LIU, W., HAIBE-KAINS, B., THIAGARAJAN, M., JEWELL, S. D., HOSTETTER, G., NEWTON, C. J., LI, Q. K., ROEHRL, M. H., FENYÖ, D., WANG, P., NESVIZHSKII, A. I., MANI, D. R., OMENN, G. S., BOJA, E. S., MESRI, M., ROBLES, A. I., RODRIGUEZ, H., BATHE, O. F., CHAN, D. W., HRUBAN, R. H., DING, L., ZHANG, B. & ZHANG, H. 2021. Proteogenomic characterization of pancreatic ductal adenocarcinoma. *Cell,* 184**,** 5031-5052.e26.

CARTER, P. E., DUPONCHEL, C., TOSI, M. & FOTHERGILL, J. E. 1991. Complete nucleotide sequence of the gene for human C1 inhibitor with an unusually high density of Alu elements. *Eur J Biochem,* 197**,** 301-8.

CHAKRABORTY, S., BAINE, M. J., SASSON, A. R. & BATRA, S. K. 2011. Current status of molecular markers for early detection of sporadic pancreatic cancer. *Biochim Biophys Acta,* 1815**,** 44-64.

CHANG, J. H., JIANG, Y. & PILLARISETTY, V. G. 2016. Role of immune cells in pancreatic cancer from bench to clinical application: An updated review. *Medicine (Baltimore),* 95**,** e5541.

CHANG, K. J., PARASHER, G., CHRISTIE, C., LARGENT, J. & ANTON-CULVER, H. 2005. Risk of pancreatic adenocarcinoma: disparity between African Americans and other race/ethnic groups. *Cancer,* 103**,** 349-57.

CHANTADA-VÁZQUEZ, M. D. P., CONDE-AMBOAGE, M., GRAÑA-LÓPEZ, L., VÁZQUEZ-ESTÉVEZ, S., BRAVO, S. B. & NÚÑEZ, C. 2022. Circulating Proteins Associated with Response and

Resistance to Neoadjuvant Chemotherapy in HER2-Positive Breast Cancer. *Cancers (Basel),* 14.

CHEN, J., CHEN, L. J., XIA, Y. L., ZHOU, H. C., YANG, R. B., WU, W., LU, Y., HU, L. W. & ZHAO, Y. 2013. Identification and verification of transthyretin as a potential biomarker for pancreatic ductal adenocarcinoma. *J Cancer Res Clin Oncol,* 139**,** 1117-27.

CHEN, M. X., ZHANG, Y., FERNIE, A. R., LIU, Y. G. & ZHU, F. Y. 2021. SWATH-MS-Based Proteomics: Strategies and Applications in Plants. *Trends Biotechnol,* 39**,** 433-437.

CHUNG, K. H., LEE, J.-C., LEE, J., CHO, I.-K., KIM, J., JANG, W., YOO, B. C. & HWANG, J.-H. 2020. Serum fibrinogen as a diagnostic and prognostic biomarker for pancreatic ductal adenocarcinoma. *Pancreatology,* 20**,** 1465-1471.

CHUTIPONGTANATE, S., CHATCHEN, S. & SVASTI, J. 2017. Plasma prefractionation methods for proteomic analysis and perspectives in clinical applications. *Proteomics Clin Appl,* 11.

CROFT, D., MUNDO, A. F., HAW, R., MILACIC, M., WEISER, J., WU, G., CAUDY, M., GARAPATI, P., GILLESPIE, M., KAMDAR, M. R., JASSAL, B., JUPE, S., MATTHEWS, L., MAY, B., PALATNIK, S., ROTHFELS, K., SHAMOVSKY, V., SONG, H., WILLIAMS, M., BIRNEY, E., HERMJAKOB, H., STEIN, L. & D'EUSTACHIO, P. 2014. The Reactome pathway knowledgebase. *Nucleic Acids Res,* 42**,** D472-7.

CUI, Y. & ANDERSEN, D. K. 2012. Diabetes and pancreatic cancer. *Endocr Relat Cancer,* 19**,** F9-f26.

DAMERVAL, C., DE VIENNE, D., ZIVY, M. & THIELLEMENT, H. 1986. Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. *Electrophoresis,* 7**,** 52-54.

DE ALMEIDA, L. G. N., YOUNG, D., CHOW, L., NICHOLAS, J., LEE, A., POON, M. C., DUFOUR, A. & AGBANI, E. O. 2022. Proteomics and Metabolomics Profiling of Platelets and Plasma Mediators of Thrombo-Inflammation in Gestational Hypertension and Preeclampsia. *Cells,* 11.

DEBERNARDI, S., O'BRIEN, H., ALGAHMDI, A. S., MALATS, N., STEWART, G. D., PLJEŠA-ERCEGOVAC, M., COSTELLO, E., GREENHALF, W., SAAD, A. & ROBERTS, R. 2020. A combination of urinary biomarker panel and PancRISK score for earlier detection of pancreatic cancer: A case–control study. *PLoS Medicine,* 17**,** e1003489.

DESAI, N., KAURA, T., SINGH, M., WILLINGHAM, F. F., RANA, S. & CHAWLA, S. 2022. Epidemiology and Characteristics of Chronic Pancreatitis-Do the East and West Meet? *Gastro Hep Advances*.

DI ANGELANTONIO, E., SARWAR, N., PERRY, P., KAPTOGE, S., RAY, K. K., THOMPSON, A., WOOD, A. M., LEWINGTON, S., SATTAR, N., PACKARD, C. J., COLLINS, R., THOMPSON, S. G. & DANESH, J. 2009. Major lipids, apolipoproteins, and risk of vascular disease. *Jama,* 302**,** 1993-2000.

DIAZ, K. E. & LUCAS, A. L. 2019. Familial Pancreatic Ductal Adenocarcinoma. *Am J Pathol,* 189**,** 36-43.

DIGIACOMO, G., VOLTA, F., GARAJOVA, I., BALSANO, R. & CAVAZZONI, A. 2021. Biological Hallmarks and New Therapeutic Approaches for the Treatment of PDAC. *Life (Basel),* 11.

DING, P., LI, L., LI, L., LV, X., ZHOU, D., WANG, Q., CHEN, J., YANG, C., XU, E., DAI, W., ZHANG, X., WANG, N., WANG, Q., ZHANG, W., ZHANG, L., ZHOU, Y., GU, H., LEI, Q., ZHOU, X. & HU, W. 2020. C5aR1 is a master regulator in Colorectal Tumorigenesis via Immune modulation. *Theranostics,* 10**,** 8619-8632.

DITTRICH, J., BECKER, S., HECHT, M. & CEGLAREK, U. 2015. Sample preparation strategies for targeted proteomics via proteotypic peptides in human blood using liquid chromatography tandem mass spectrometry. *Proteomics Clin Appl,* 9**,** 5-16.

DOELLINGER, J., SCHNEIDER, A., HOELLER, M. & LASCH, P. 2020. Sample Preparation by Easy Extraction and Digestion (SPEED) - A Universal, Rapid, and Detergent-free Protocol for Proteomics Based on Acid Extraction. *Mol Cell Proteomics,* 19**,** 209-222.

DOMON, B. & AEBERSOLD, R. 2006. Mass spectrometry and protein analysis. *Science,* 312**,** 212-7.

DONCHEVA, N. T., MORRIS, J. H., GORODKIN, J. & JENSEN, L. J. 2019. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res,* 18**,** 623-632.

ELBANNA, K. Y., JANG, H. J. & KIM, T. K. 2020. Imaging diagnosis and staging of pancreatic ductal adenocarcinoma: a comprehensive review. *Insights Imaging,* 11**,** 58.

FIRPO, M. A., GAY, D. Z., GRANGER, S. R., SCAIFE, C. L., DISARIO, J. A., BOUCHER, K. M. & MULVIHILL, S. J. 2009. Improved diagnosis of pancreatic adenocarcinoma using haptoglobin and serum amyloid A in a panel screen. *World J Surg,* 33**,** 716-22.

FONSLOW, B. R., STEIN, B. D., WEBB, K. J., XU, T., CHOI, J., PARK, S. K. & YATES, J. R., 3RD 2013. Digestion and depletion of abundant proteins improves proteomic coverage. *Nat Methods,* 10**,** 54-6.

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**,** 420-428.

FUJIWARA, K., SAUNG, M. T., JING, H., HERBST, B., ZARECKI, M., MUTH, S., WU, A., BIGELOW, E., CHEN, L., LI, K., JURCAK, N., BLAIR, A. B., DING, D., WICHROSKI, M., BLUM, J., CHEADLE, N., KOENITZER, J. & ZHENG, L. 2020. Interrogating the immune-modulating roles of radiation therapy for a rational combination with immune-checkpoint inhibitors in treating pancreatic cancer. *J Immunother Cancer,* 8.

FURUKAWA, K., KAWAMOTO, K., EGUCHI, H., TANEMURA, M., TANIDA, T., TOMIMARU, Y., AKITA, H., HAMA, N., WADA, H., KOBAYASHI, S., NONAKA, Y., TAKAMATSU, S., SHINZAKI, S., KUMADA, T., SATOMURA, S., ITO, T., SERADA, S., NAKA, T., MORI, M., DOKI, Y., MIYOSHI, E. & NAGANO, H. 2015. Clinicopathological Significance of Leucine-Rich α2-Glycoprotein-1 in Sera of Patients With Pancreatic Cancer. *Pancreas,* 44**,** 93-8.

GADWA, J., BICKETT, T. E., DARRAGH, L. B., KNITZ, M. W., BHATIA, S., PIPER, M., VAN COURT, B., BHUVANE, S., NGUYEN, D., NANGIA, V., KLECZKO, E. K., NEMENOFF, R. A. & KARAM, S. D. 2021. Complement C3a and C5a receptor blockade modulates regulatory T cell conversion in head and neck cancer. *J Immunother Cancer,* 9.

GANDHI, S., DE LA FUENTE, J., MURAD, M. H. & MAJUMDER, S. 2022. Chronic Pancreatitis Is a Risk Factor for Pancreatic Cancer, and Incidence Increases With Duration of Disease: A Systematic Review and Meta-analysis. *Clin Transl Gastroenterol,* 13**,** e00463.

GARCIA, S., BALDASSO, P. A., GUEST, P. C. & MARTINS-DE-SOUZA, D. 2017. Depletion of Highly Abundant Proteins of the Human Blood Plasma: Applications in Proteomics Studies of Psychiatric Disorders. *Methods Mol Biol,* 1546**,** 195-204.

GE, S. X., JUNG, D. & YAO, R. 2020. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics,* 36**,** 2628-2629.

GEYER, P. E., VOYTIK, E., TREIT, P. V., DOLL, S., KLEINHEMPEL, A., NIU, L., MÜLLER, J. B., BUCHHOLTZ, M. L., BADER, J. M., TEUPSER, D., HOLDT, L. M. & MANN, M. 2019. Plasma Proteome Profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol Med,* 11**,** e10427.

GIANAZZA, E., MILLER, I., PALAZZOLO, L., PARRAVICINI, C. & EBERINI, I. 2016. With or without you - Proteomics with or without major plasma/serum proteins. *J Proteomics,* 140**,** 62-80.

GIANSANTI, P., TSIATSIANI, L., LOW, T. Y. & HECK, A. J. 2016. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat Protoc,* 11**,** 993-1006.

GILLET, L. C., NAVARRO, P., TATE, S., RÖST, H., SELEVSEK, N., REITER, L., BONNER, R. & AEBERSOLD, R. 2012. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics,* 11**,** O111.016717.

GONG, J., JIE, Y., XIAO, C., ZHOU, W., LI, X., CHEN, Y., WU, Y., CAO, J., ZHANG, Q., GAO, Z., HU, B. & CHONG, Y. 2020. Increased Expression of Fibulin-1 Is Associated With Hepatocellular

Carcinoma Progression by Regulating the Notch Signaling Pathway. *Front Cell Dev Biol,* 8**,** 478.

GONZÁLEZ-BORJA, I., VIÚDEZ, A., GOÑI, S., SANTAMARIA, E., CARRASCO-GARCÍA, E., PÉREZ-SANZ, J., HERNÁNDEZ-GARCÍA, I., SALA-ELARRE, P., ARRAZUBI, V. & OYAGA-IRIARTE, E. 2019. Omics approaches in pancreatic adenocarcinoma. *Cancers,* 11**,** 1052.

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.

HAN, M., TRAN, T. P. T. & OH, J. K. 2022. Chronic pancreatitis and cancer risk in a matched cohort study using national claims data in South Korea. *Sci Rep,* 12**,** 5545.

HARRIS, P. A., TAYLOR, R., THIELKE, R., PAYNE, J., GONZALEZ, N. & CONDE, J. G. 2009. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform,* 42**,** 377-81.

HARSHA, H. C., KANDASAMY, K., RANGANATHAN, P., RANI, S., RAMABADRAN, S., GOLLAPUDI, S., BALAKRISHNAN, L., DWIVEDI, S. B., TELIKICHERLA, D., SELVAN, L. D., GOEL, R., MATHIVANAN, S., MARIMUTHU, A., KASHYAP, M., VIZZA, R. F., MAYER, R. J., DECAPRIO, J. A., SRIVASTAVA, S., HANASH, S. M., HRUBAN, R. H. & PANDEY, A. 2009. A compendium of potential biomarkers of pancreatic cancer. *PLoS Med,* 6**,** e1000046.

HE, X. Y. & YUAN, Y. Z. 2014. Advances in pancreatic cancer research: moving towards early detection. *World J Gastroenterol,* 20**,** 11241-8.

HELLER, D. R., NICOLSON, N. G., AHUJA, N., KHAN, S. & KUNSTMAN, J. W. 2020. Association of Treatment Inequity and Ancestry With Pancreatic Ductal Adenocarcinoma Survival. *JAMA Surg,* 155**,** e195047.

HEZEL, A. F., KIMMELMAN, A. C., STANGER, B. Z., BARDEESY, N. & DEPINHO, R. A. 2006. Genetics and biology of pancreatic ductal adenocarcinoma. *Genes Dev,* 20**,** 1218-49.

HILDONEN, S., HALVORSEN, T. G. & REUBSAET, L. 2014. Why less is more when generating tryptic peptides in bottom-up proteomics. *Proteomics,* 14**,** 2031-41.

HO, C. S., LAM, C. W., CHAN, M. H., CHEUNG, R. C., LAW, L. K., LIT, L. C., NG, K. F., SUEN, M. W. & TAI, H. L. 2003. Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin Biochem Rev,* 24**,** 3-12.

HU, B. & YAO, Z. P. 2022. Electrospray ionization mass spectrometry with wooden tips: A review. *Anal Chim Acta,* 1209**,** 339136.

IGNJATOVIC, V., GEYER, P. E., PALANIAPPAN, K. K., CHAABAN, J. E., OMENN, G. S., BAKER, M. S., DEUTSCH, E. W. & SCHWENK, J. M. 2019. Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data. *J Proteome Res,* 18**,** 4085-4097.

JACKSON, D., CRAVEN, R. A., HUTSON, R. C., GRAZE, I., LUETH, P., TONGE, R. P., HARTLEY, J. L., NICKSON, J. A., RAYNER, S. J., JOHNSTON, C., DIEPLINGER, B., HUBALEK, M., WILKINSON, N., PERREN, T. J., KEHOE, S., HALL, G. D., DAXENBICHLER, G., DIEPLINGER, H., SELBY, P. J. & BANKS, R. E. 2007. Proteomic profiling identifies afamin as a potential biomarker for ovarian cancer. *Clin Cancer Res,* 13**,** 7370-9.

JAGDISH, R. K., MARAS, J. S. & SARIN, S. K. 2021. Albumin in Advanced Liver Diseases: The Good and Bad of a Drug! *Hepatology,* 74**,** 2848-2862.

JANKOVSKA, E., SVITEK, M., HOLADA, K. & PETRAK, J. 2019. Affinity depletion versus relative protein enrichment: a side-by-side comparison of two major strategies for increasing human cerebrospinal fluid proteome coverage. *Clin Proteomics,* 16**,** 9.

JIN, Y., WANG, W., WANG, Q., ZHANG, Y., ZAHID, K. R., RAZA, U. & GONG, Y. 2022. Alpha-1-antichymotrypsin as a novel biomarker for diagnosis, prognosis, and therapy prediction in human diseases. *Cancer Cell Int,* 22**,** 156.

KAILASAM, S. Exploring the Principle of Ion Exchange Chromatography and Its Applications Article Published: January 14, 2022.

KAMISAWA, T., WOOD, L. D., ITOI, T. & TAKAORI, K. 2016. Pancreatic cancer. *The Lancet,* 388**,** 73-85.

KANDA, M., MATTHAEI, H., WU, J., HONG, S. M., YU, J., BORGES, M., HRUBAN, R. H., MAITRA, A., KINZLER, K., VOGELSTEIN, B. & GOGGINS, M. 2012. Presence of somatic mutations in most early-stage pancreatic intraepithelial neoplasia. *Gastroenterology,* 142**,** 730-733.e9.

KASTELEIN, J. J., VAN DER STEEG, W. A., HOLME, I., GAFFNEY, M., CATER, N. B., BARTER, P., DEEDWANIA, P., OLSSON, A. G., BOEKHOLDT, S. M., DEMICCO, D. A., SZAREK, M., LAROSA, J. C., PEDERSEN, T. R. & GRUNDY, S. M. 2008. Lipids, apolipoproteins, and their ratios in relation to cardiovascular events with statin treatment. *Circulation,* 117**,** 3002-9.

KAUFMAN, M., YAN, X. J., LI, W., GHIA, E. M., LANGERAK, A. W., RASSENTI, L. Z., BELESSI, C., KAY, N. E., DAVI, F., BYRD, J. C., POSPISILOVA, S., BROWN, J. R., CATHERWOOD, M., DAVIS, Z., OSCIER, D., MONTILLO, M., TRENTIN, L., ROSENQUIST, R., GHIA, P., BARRIENTOS, J. C., KOLITZ, J. E., ALLEN, S. L., RAI, K. R., STAMATOPOULOS, K., KIPPS, T. J., NEUBERG, D. & CHIORAZZI, N. 2022. Impact of the Types and Relative Quantities of IGHV Gene Mutations in Predicting Prognosis of Patients With Chronic Lymphocytic Leukemia. *Front Oncol,* 12**,** 897280.

KAUR, G., POLJAK, A., ALI, S. A., ZHONG, L., RAFTERY, M. J. & SACHDEV, P. 2021. Extending the Depth of Human Plasma Proteome Coverage Using Simple Fractionation Techniques. *J Proteome Res,* 20**,** 1261-1279.

KESSENBROCK, K., PLAKS, V. & WERB, Z. 2010. Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell,* 141**,** 52-67.

KHACHFE, H. H., CHAHROUR, M. A., HABIB, J. R., YU, J. & JAMALI, F. R. 2021. A Quality Assessment of the Information Accessible to Patients on the Internet About the Whipple Procedure. *World J Surg,* 45**,** 1853-1859.

KHAWJA, S. N., MOHAMMED, S., SILBERFEIN, E. J., MUSHER, B. L., FISHER, W. E. & VAN BUREN, G., 2ND 2015. Pancreatic cancer disparities in African Americans. *Pancreas,* 44**,** 522-7.

KHORANA, A. A., MANGU, P. B., BERLIN, J., ENGEBRETSON, A., HONG, T. S., MAITRA, A., MOHILE, S. G., MUMBER, M., SCHULICK, R., SHAPIRO, M., URBA, S., ZEH, H. J. & KATZ, M. H. G. 2017. Potentially Curable Pancreatic Cancer: American Society of Clinical Oncology Clinical Practice Guideline Update. *J Clin Oncol,* 35**,** 2324-2328.

KIM, H., PARK, J., WANG, J. I. & KIM, Y. 2017a. Recent advances in proteomic profiling of pancreatic ductal adenocarcinoma and the road ahead. *Expert Rev Proteomics,* 14**,** 963-971.

KIM, J., BAMLET, W. & OBERG, A. 2017b. Detection of early pancreatic ductal adenocarcinoma with thrombospondin-2 and CA19-9 blood markers. Sci. Transl. Med.

KIM, K., AHN, S., LIM, J., YOO, B. C., HWANG, J.-H. & JANG, W. 2014. Detection of pancreatic cancer biomarkers using mass spectrometry. *Cancer informatics,* 13**,** CIN. S16341.

KLAMMER, A. A. & MACCOSS, M. J. 2006. Effects of modified digestion schemes on the identification of proteins from complex mixtures. *J Proteome Res,* 5**,** 695-700.

KOPP, J. L., VON FIGURA, G., MAYES, E., LIU, F. F., DUBOIS, C. L., MORRIS, J. P. T., PAN, F. C., AKIYAMA, H., WRIGHT, C. V., JENSEN, K., HEBROK, M. & SANDER, M. 2012. Identification of Sox9-dependent acinar-to-ductal reprogramming as the principal mechanism for initiation of pancreatic ductal adenocarcinoma. *Cancer Cell,* 22**,** 737-50.

KRAMER-MAREK, G., GORE, J. & KORC, M. 2013. Molecular imaging in pancreatic cancer--a roadmap for therapeutic decisions. *Cancer Lett,* 341**,** 132-8.

KRÜGER, T., LEHMANN, T. & RHODE, H. 2013. Effect of quality characteristics of single sample preparation steps in the precision and coverage of proteomic studies--a review. *Anal Chim Acta,* 776**,** 1-10.

KULKARNI, N. M., SOLOFF, E. V., TOLAT, P. P., SANGSTER, G. P., FLEMING, J. B., BROOK, O. R., WANG, Z. J., HECHT, E. M., ZINS, M., BHOSALE, P. R., ARIF-TIWARI, H., MANNELLI, L., KAMBADAKONE, A. R. & TAMM, E. P. 2020. White paper on pancreatic ductal adenocarcinoma from society of abdominal radiology's disease-focused panel for pancreatic

ductal adenocarcinoma: Part I, AJCC staging system, NCCN guidelines, and borderline resectable disease. *Abdom Radiol (NY),* 45**,** 716-728.

KWAN, Y. P., TEO, M. H. Y., LIM, J. C. W., TAN, M. S., ROSELLINNY, G., WAHLI, W. & WANG, X. 2021. LRG1 Promotes Metastatic Dissemination of Melanoma through Regulating EGFR/STAT3 Signalling. *Cancers (Basel),* 13.

LEE, C. S., TAIB, N. A., ASHRAFZADEH, A., FADZLI, F., HARUN, F., RAHMAT, K., HOONG, S. M., ABDUL-RAHMAN, P. S. & HASHIM, O. H. 2016. Unmasking Heavily O-Glycosylated Serum Proteins Using Perchloric Acid: Identification of Serum Proteoglycan 4 and Protease C1 Inhibitor as Molecular Indicators for Screening of Breast Cancer. *PLoS One,* 11**,** e0149551.

LEE, M.-H., THOMAS, J. L., LIAO, C.-L., JURCEVIC, S., CRNOGORAC-JURCEVIC, T. & LIN, H.-Y. 2017. Polymers imprinted with three REG1B peptides for electrochemical determination of Regenerating Protein 1B, a urinary biomarker for pancreatic ductal adenocarcinoma. *Microchimica Acta,* 184**,** 1773-1780.

LEE, M. J., NA, K., JEONG, S. K., LIM, J. S., KIM, S. A., LEE, M. J., SONG, S. Y., KIM, H., HANCOCK, W. S. & PAIK, Y. K. 2014. Identification of human complement factor B as a novel biomarker candidate for pancreatic ductal adenocarcinoma. *J Proteome Res,* 13**,** 4878-88.

LEE, P. Y., OSMAN, J., LOW, T. Y. & JAMAL, R. 2019. Plasma/serum proteomics: depletion strategies for reducing high-abundance proteins for biomarker discovery. *Bioanalysis,* 11**,** 1799-1812.

LI, Q., WANG, H., ZOGOPOULOS, G., SHAO, Q., DONG, K., LV, F., NWILATI, K., GUI, X. Y., CUGGIA, A., LIU, J. L. & GAO, Z. H. 2016. Reg proteins promote acinar-to-ductal metaplasia and act as novel diagnostic and prognostic markers in pancreatic ductal adenocarcinoma. *Oncotarget,* 7**,** 77838-77853.

LIN, Z., YIN, H., LO, A., RUFFIN, M. T., ANDERSON, M. A., SIMEONE, D. M. & LUBMAN, D. M. 2014. Label-free relative quantification of alpha-2-macroglobulin site-specific core-fucosylation in pancreatic cancer by LC-MS/MS. *Electrophoresis,* 35**,** 2108-15.

LITMAN-ZAWADZKA, A., ŁUKASZEWICZ-ZAJĄC, M. & MROCZKO, B. 2019. Novel potential biomarkers for pancreatic cancer - A systematic review. *Adv Med Sci,* 64**,** 252-257.

LIZ, M. A., COELHO, T., BELLOTTI, V., FERNANDEZ-ARIAS, M. I., MALLAINA, P. & OBICI, L. 2020. A Narrative Review of the Role of Transthyretin in Health and Disease. *Neurol Ther,* 9**,** 395-402.

LUQUE-GARCIA, J. L. & NEUBERT, T. A. 2007. Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry. *J Chromatogr A,* 1153**,** 259-76.

LV, S., GAO, J., ZHU, F., LI, Z., GONG, Y., XU, G. & MA, L. 2011. Transthyretin, identified by proteomics, is overabundant in pancreatic juice from pancreatic carcinoma and originates from pancreatic islets. *Diagn Cytopathol,* 39**,** 875-81.

MACHADO, E. R., ANNUNZIATA, I., VAN DE VLEKKERT, D., GROSVELD, G. C. & D'AZZO, A. 2021. Lysosomes and Cancer Progression: A Malignant Liaison. *Front Cell Dev Biol,* 9**,** 642494.

MADSEN, C. D., FERRARIS, G. M. S., ANDOLFO, A., CUNNINGHAM, O. & SIDENIUS, N. 2007. uPAR-induced cell adhesion and migration: vitronectin provides the key. *The Journal of cell biology,* 177**,** 927-939.

MAI, S. & INKIELEWICZ-STEPNIAK, I. 2021. Pancreatic Cancer and Platelets Crosstalk: A Potential Biomarker and Target. *Front Cell Dev Biol,* 9**,** 749689.

MAKAWITA, S., DIMITROMANOLAKIS, A., SOOSAIPILLAI, A., SOLEAS, I., CHAN, A., GALLINGER, S., HAUN, R. S., BLASUTIG, I. M. & DIAMANDIS, E. P. 2013. Validation of four candidate pancreatic cancer serological biomarkers that improve the performance of CA19.9. *BMC Cancer,* 13**,** 404.

MALKA, D., HAMMEL, P., MAIRE, F., RUFAT, P., MADEIRA, I., PESSIONE, F., LEVY, P. & RUSZNIEWSKI, P. 2002. Risk of pancreatic adenocarcinoma in chronic pancreatitis. *Gut,* 51**,** 849-852.

MANGARAJ, M., NANDA, R. & PANDA, S. 2016. Apolipoprotein A-I: A Molecule of Diverse Function. *Indian J Clin Biochem,* 31**,** 253-9.

MATSUO, K., ITO, H., WAKAI, K., NAGATA, C., MIZOUE, T., TANAKA, K., TSUJI, I., TAMAKOSHI, A., SASAZUKI, S., INOUE, M. & TSUGANE, S. 2011. Cigarette smoking and pancreas cancer risk: an evaluation based on a systematic review of epidemiologic evidence in the Japanese population. *Jpn J Clin Oncol,* 41**,** 1292-302.

MEGHA, K. B. & MOHANAN, P. V. 2021. Role of immunoglobulin and antibodies in disease management. *Int J Biol Macromol,* 169**,** 28-38.

MELEADY, P., ABDUL RAHMAN, R., HENRY, M., MORIARTY, M. & CLYNES, M. 2020. Proteomic analysis of pancreatic ductal adenocarcinoma. *Expert Rev Proteomics,* 17**,** 453-467.

MENTER, D. G., TUCKER, S. C., KOPETZ, S., SOOD, A. K., CRISSMAN, J. D. & HONN, K. V. 2014. Platelets and cancer: a casual or causal relationship: revisited. *Cancer Metastasis Rev,* 33**,** 231-69.

MERLE, N. S., NOE, R., HALBWACHS-MECARELLI, L., FREMEAUX-BACCHI, V. & ROUMENINA, L. T. 2015. Complement System Part II: Role in Immunity. *Front Immunol,* 6**,** 257.

NANJAPPA, V., THOMAS, J. K., MARIMUTHU, A., MUTHUSAMY, B., RADHAKRISHNAN, A., SHARMA, R., AHMAD KHAN, A., BALAKRISHNAN, L., SAHASRABUDDHE, N. A., KUMAR, S., JHAVERI, B. N., SHETH, K. V., KUMAR KHATANA, R., SHAW, P. G., SRIKANTH, S. M., MATHUR, P. P., SHANKAR, S., NAGARAJA, D., CHRISTOPHER, R., MATHIVANAN, S., RAJU, R., SIRDESHMUKH, R., CHATTERJEE, A., SIMPSON, R. J., HARSHA, H. C., PANDEY, A. & PRASAD, T. S. 2014. Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic Acids Res,* 42**,** D959-65.

NETWORK, N. C. C. 2019. NCCN guidelines for patients: Pancreatic Cancer. *National Comprehensive Cancer Network*.

NWEKE, E. E., NAICKER, P., ARON, S., STOYCHEV, S., DEVAR, J., TABB, D. L., OMOSHORO-JONES, J., SMITH, M. & CANDY, G. 2020. SWATH-MS based proteomic profiling of pancreatic ductal adenocarcinoma tumours reveals the interplay between the extracellular matrix and related intracellular pathways. *PLoS One,* 15**,** e0240453.

O'ROURKE, M. B., TOWN, S. E. L., DALLA, P. V., BICKNELL, F., KOH BELIC, N., VIOLI, J. P., STEELE, J. R. & PADULA, M. P. 2019. What is Normalization? The Strategies Employed in Top-Down and Bottom-Up Proteome Analysis Workflows. *Proteomes,* 7.

OGUN, A. S. & ADEYINKA, A. 2021. Biochemistry, transferrin. *StatPearls [Internet].* StatPearls Publishing.

OHTANI, H. 1998. Stromal reaction in cancer tissue: pathophysiologic significance of the expression of matrix-degrading enzymes in relation to matrix turnover and immune/inflammatory reactions. *Pathol Int,* 48**,** 1-9.

ORTH, M., METZGER, P., GERUM, S., MAYERLE, J., SCHNEIDER, G., BELKA, C., SCHNURR, M. & LAUBER, K. 2019. Pancreatic ductal adenocarcinoma: biological hallmarks, current status, and future perspectives of combined modality treatment approaches. *Radiat Oncol,* 14**,** 141.

P, O. C., RUSCH, V., TALBOT, S. G., SARKARIA, I., VIALE, A., SOCCI, N., NGAI, I., RAO, P. & SINGH, B. 2004. Casein kinase II alpha subunit and C1-inhibitor are independent predictors of outcome in patients with squamous cell carcinoma of the lung. *Clin Cancer Res,* 10**,** 5792-803.

PALACIOS-ACEDO, A. L., LANGIU, M., CRESCENCE, L., MÈGE, D., DUBOIS, C. & PANICOT-DUBOIS, L. 2022. Platelet and Cancer-Cell Interactions Modulate Cancer-Associated Thrombosis Risk in Different Cancer Types. *Cancers (Basel),* 14.

PARK, J., MORLEY, T. S., KIM, M., CLEGG, D. J. & SCHERER, P. E. 2014. Obesity and cancer--mechanisms underlying tumour progression and recurrence. *Nat Rev Endocrinol,* 10**,** 455-465.

PATRICK, A. L. 2020. Electrospray ionization enters the final frontier: Mass spectrometry's role in understanding electrospray thrusters and their plumes. *Rapid Commun Mass Spectrom,* 34**,** e8587.

PENG, H., CHEN, R. & PAN, S. 2019. Proteomics Profiling of Pancreatic Cancer: Roles in Biomarker Discovery. *Oncogenomics.* Elsevier.

PENG, S., DU, T., WU, W., CHEN, X., LAI, Y., ZHU, D., WANG, Q., MA, X., LIN, C., LI, Z., GUO, Z. & HUANG, H. 2018. Decreased expression of serine protease inhibitor family G1 (SERPING1) in prostate cancer can help distinguish high-risk prostate cancer and predicts malignant progression. *Urol Oncol,* 36**,** 366.e1-366.e9.

PISANU, S., BIOSA, G., CARCANGIU, L., UZZAU, S. & PAGNOZZI, D. 2018. Comparative evaluation of seven commercial products for human serum enrichment/depletion by shotgun proteomics. *Talanta,* 185**,** 213-220.

PORTERFIELD, M., ZHAO, P., HAN, H., CUNNINGHAM, J., AOKI, K., VON HOFF, D. D., DEMEURE, M. J., PIERCE, J. M., TIEMEYER, M. & WELLS, L. 2014. Discrimination between adenocarcinoma and normal pancreatic ductal fluid by proteomic and glycomic analysis. *J Proteome Res,* 13**,** 395-407.

PRINCIPE, D. R., UNDERWOOD, P. W., KORC, M., TREVINO, J. G., MUNSHI, H. G. & RANA, A. 2021. The Current Treatment Paradigm for Pancreatic Ductal Adenocarcinoma and Barriers to Therapeutic Efficacy. *Front Oncol,* 11**,** 688377.

RAWLA, P., SUNKARA, T. & GADUPUTI, V. 2019. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. *World J Oncol,* 10**,** 10-27.

REBOURS, V., BOUTRON-RUAULT, M. C., SCHNEE, M., FÉREC, C., MAIRE, F., HAMMEL, P., RUSZNIEWSKI, P. & LÉVY, P. 2008. Risk of pancreatic adenocarcinoma in patients with hereditary pancreatitis: a national exhaustive series. *Am J Gastroenterol,* 103**,** 111-9.

REDONDO, M., TELLEZ, T. & ROLDAN, M. J. 2009. The role of clusterin (CLU) in malignant transformation and drug resistance in breast carcinomas. *Adv Cancer Res,* 105**,** 21-43.

REED, G. F., LYNN, F. & MEADE, B. D. 2002. Use of coefficient of variation in assessing variability of quantitative assays. *Clin Diagn Lab Immunol,* 9**,** 1235-9.

REN, L., YI, J., LI, W., ZHENG, X., LIU, J., WANG, J. & DU, G. 2019. Apolipoproteins and cancer. *Cancer Med,* 8**,** 7032-7043.

RESKE, S. N., GRILLENBERGER, K. G., GLATTING, G., PORT, M., HILDEBRANDT, M., GANSAUGE, F. & BEGER, H. G. 1997. Overexpression of glucose transporter 1 and increased FDG uptake in pancreatic carcinoma. *J Nucl Med,* 38**,** 1344-8.

REVEL, M., DAUGAN, M. V., SAUTÉS-FRIDMAN, C., FRIDMAN, W. H. & ROUMENINA, L. T. 2020. Complement System: Promoter or Suppressor of Cancer Progression? *Antibodies (Basel),* 9.

ROUMENINA, L. T., DAUGAN, M. V., PETITPREZ, F., SAUTÈS-FRIDMAN, C. & FRIDMAN, W. H. 2019. Context-dependent roles of complement in cancer. *Nat Rev Cancer,* 19**,** 698-715.

RUIZ, M. 2021. Into the Labyrinth of the Lipocalin α1-Acid Glycoprotein. *Front Physiol,* 12**,** 686251.

SAKI, K., MANSOURI, V., ASRI, N., FATHI, M. & RAZZAGHI, Z. 2021. Common and differential features of liver and pancreatic cancers: molecular mechanism approach. *Gastroenterol Hepatol Bed Bench,* 14**,** S87-s93.

SALA, A., BETTUZZI, S., PUCCI, S., CHAYKA, O., DEWS, M. & THOMAS-TIKHONENKO, A. 2009. Regulation of CLU gene expression by oncogenes and epigenetic factors implications for tumorigenesis. *Adv Cancer Res,* 105**,** 115-32.

SASAKI, K., TAKANO, S., TOMIZAWA, S., MIYAHARA, Y., FURUKAWA, K., TAKAYASHIKI, T., KUBOKI, S., TAKADA, M. & OHTSUKA, M. 2021. C4b-binding protein α-chain enhances antitumor immunity by facilitating the accumulation of tumor-infiltrating lymphocytes in the tumor microenvironment in pancreatic cancer. *J Exp Clin Cancer Res,* 40**,** 212.

SATHE, A. & CUSICK, J. K. 2021. Biochemistry, Immunoglobulin M. *StatPearls [Internet].* StatPearls Publishing.

SATO, Y., KOBAYASHI, T., NISHIUMI, S., OKADA, A., FUJITA, T., SANUKI, T., KOBAYASHI, M., ASAHARA, M., ADACHI, M., SAKAI, A., SHIOMI, H., MASUDA, A., YOSHIDA, M., TAKEUCHI, K., KODAMA, Y., KUTSUMI, H., NAGASHIMA, K. & HONDA, K. 2020. Prospective Study Using Plasma Apolipoprotein A2-Isoforms to Screen for High-Risk Status of Pancreatic Cancer. *Cancers (Basel),* 12.

SAVELIEV, S., BRATZ, M., ZUBAREV, R., SZAPACS, M., BUDAMGUNTA, H. & URH, M. 2013. Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis. *Nature Methods,* 10**,** i-ii.

SCHMIDT, C. M., WHITE, P. B., WATERS, J. A., YIANNOUTSOS, C. T., CUMMINGS, O. W., BAKER, M., HOWARD, T. J., ZYROMSKI, N. J., NAKEEB, A., DEWITT, J. M., AKISIK, F. M., SHERMAN, S., PITT, H. A. & LILLEMOE, K. D. 2007. Intraductal papillary mucinous neoplasms: predictors of malignant and invasive pathology. *Ann Surg,* 246**,** 644-51; discussion 651-4.

SCHWEIGERT, F. J., WIRTH, K. & RAILA, J. 2004. Characterization of the microheterogeneity of transthyretin in plasma and urine using SELDI-TOF-MS immunoassay. *Proteome Sci,* 2**,** 5.

SHAEIR, S. S., ABDELSHAHEED, F., ELSEBAIY, A., SOBEIH, M. E., DESOKY, E., OWIS, A. S., ELTOHAMY, M. I. & OTHMAN, H. F. 2022. Postoperative outcome of pancreaticoduodenectomy for pancreatic ductal adenocarcinoma and predictors of 1st year mortality.

SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res,* 13**,** 2498-504.

SHAW, M. M. & RIEDERER, B. M. 2003. Sample preparation for two-dimensional gel electrophoresis. *Proteomics,* 3**,** 1408-17.

SHEINENZON, A., SHEHADEH, M., MICHELIS, R., SHAOUL, E. & RONEN, O. 2021. Serum albumin levels and inflammation. *Int J Biol Macromol,* 184**,** 857-862.

SIEGEL, R. L., MILLER, K. D., WAGLE, N. S. & JEMAL, A. 2023. Cancer statistics, 2023. *CA Cancer J Clin,* 73**,** 17-48.

SIEPEN, J. A., KEEVIL, E. J., KNIGHT, D. & HUBBARD, S. J. 2007. Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *J Proteome Res,* 6**,** 399-408.

SILVERMAN, D. T., HOOVER, R. N., BROWN, L. M., SWANSON, G. M., SCHIFFMAN, M., GREENBERG, R. S., HAYES, R. B., LILLEMOE, K. D., SCHOENBERG, J. B., SCHWARTZ, A. G., LIFF, J., POTTERN, L. M. & FRAUMENI, J. F., JR. 2003. Why do Black Americans have a higher risk of pancreatic cancer than White Americans? *Epidemiology,* 14**,** 45-54.

SINHA, A. & MANN, M. 2020. A beginner's guide to mass spectrometry–based proteomics. *The Biochemist,* 42**,** 64-69.

SLAMON, D. J., CLARK, G. M., WONG, S. G., LEVIN, W. J., ULLRICH, A. & MCGUIRE, W. L. 1987. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science,* 235**,** 177-82.

SOCIETY, A. C. 2014. *Cancer facts & figures 2014*, American Cancer Society.

SOGAWA, K., TAKANO, S., IIDA, F., SATOH, M., TSUCHIDA, S., KAWASHIMA, Y., YOSHITOMI, H., SANDA, A., KODERA, Y., TAKIZAWA, H., MIKATA, R., OHTSUKA, M., SHIMIZU, H., MIYAZAKI, M., YOKOSUKA, O. & NOMURA, F. 2016. Identification of a novel serum biomarker for pancreatic cancer, C4b-binding protein α-chain (C4BPA) by quantitative proteomic analysis using tandem mass tags. *Br J Cancer,* 115**,** 949-956.

SOWEID, A. M. 2017. The borderline resectable and locally advanced pancreatic ductal adenocarcinoma: Definition. *Endosc Ultrasound,* 6**,** S76-s78.

SPETH, C., BELLOTTI, R., SCHÄFER, G., RAMBACH, G., TEXLER, B., THURNER, G. C., ÖFNER, D., LASS-FLÖRL, C. & MAGLIONE, M. 2022. Complement and Fungal Dysbiosis as Prognostic Markers and Potential Targets in PDAC Treatment. *Curr Oncol,* 29**,** 9833-9854.

STASTNA, M. & VAN EYK, J. E. 2012. Secreted proteins as a fundamental source for biomarker discovery. *Proteomics,* 12**,** 722-35.

SWANEY, D. L., WENGER, C. D. & COON, J. J. 2010. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res,* 9**,** 1323-9.

TAKAORI, K., BASSI, C., BIANKIN, A., BRUNNER, T. B., CATALDO, I., CAMPBELL, F., CUNNINGHAM, D., FALCONI, M., FRAMPTON, A. E., FURUSE, J., GIOVANNINI, M., JACKSON, R., NAKAMURA, A.,

NEALON, W., NEOPTOLEMOS, J. P., REAL, F. X., SCARPA, A., SCLAFANI, F., WINDSOR, J. A., YAMAGUCHI, K., WOLFGANG, C. & JOHNSON, C. D. 2016. International Association of Pancreatology (IAP)/European Pancreatic Club (EPC) consensus review of guidelines for the treatment of pancreatic cancer. *Pancreatology,* 16**,** 14-27.

THOMAS, J. K., KIM, M.-S., BALAKRISHNAN, L., NANJAPPA, V., RAJU, R., MARIMUTHU, A., RADHAKRISHNAN, A., MUTHUSAMY, B., KHAN, A. A. & SAKAMURI, S. 2014. Pancreatic Cancer Database: an integrative resource for pancreatic cancer. *Cancer biology & therapy,* 15**,** 963-967.

TORIOLA, A. T., LUO, S., THOMAS, T. S., DRAKE, B. F., CHANG, S. H., SANFILIPPO, K. M. & CARSON, K. R. 2020. Metformin Use and Pancreatic Cancer Survival among Non-Hispanic White and African American U.S. Veterans with Diabetes Mellitus. *Cancer Epidemiol Biomarkers Prev,* 29**,** 169-175.

TRAMACERE, I., SCOTTI, L., JENAB, M., BAGNARDI, V., BELLOCCO, R., ROTA, M., CORRAO, G., BRAVI, F., BOFFETTA, P. & LA VECCHIA, C. 2010. Alcohol drinking and pancreatic cancer risk: a meta-analysis of the dose-risk relation. *Int J Cancer,* 126**,** 1474-86.

TUMMERS, W. S., WILLMANN, J. K., BONSING, B. A., VAHRMEIJER, A. L., GAMBHIR, S. S. & SWIJNENBURG, R. J. 2018. Advances in Diagnostic and Intraoperative Molecular Imaging of Pancreatic Cancer. *Pancreas,* 47**,** 675-689.

ULLA ROCHA, J. L., ALVAREZ SANCHEZ, M. V., PAZ ESQUETE, J., FERNANDEZ SALGADO, E., ALVAREZ ALVAREZ, C., VAZQUEZ SANLUIS, M. J., LEDO BARRO, L. & VAZQUEZ ASTRAY, E. 2007. Evaluation of the bilio-pancreatic region using endoscopic ultrasonography in patients referred with and without abdominal pain and CA 19-9 serum level elevation. *Jop,* 8**,** 191-7.

USHIO, J., KANNO, A., IKEDA, E., ANDO, K., NAGAI, H., MIWATA, T., KAWASAKI, Y., TADA, Y., YOKOYAMA, K., NUMAO, N., TAMADA, K., LEFOR, A. K. & YAMAMOTO, H. 2021. Pancreatic Ductal Adenocarcinoma: Epidemiology and Risk Factors. *Diagnostics (Basel),* 11.

VACHIRANUBHAP, B., KIM, Y. H., BALCI, N. C. & SEMELKA, R. C. 2009. Magnetic resonance imaging of adenocarcinoma of the pancreas. *Top Magn Reson Imaging,* 20**,** 3-9.

VAN DE WOUW, J. & JOLES, J. A. 2022. Albumin is an interface between blood plasma and cell membrane, and not just a sponge. *Clin Kidney J,* 15**,** 624-634.

VELLAN, C. J., JAYAPALAN, J. J., YOONG, B. K., ABDUL-AZIZ, A., MAT-JUNIT, S. & SUBRAMANIAN, P. 2022. Application of Proteomics in Pancreatic Ductal Adenocarcinoma Biomarker Investigations: A Review. *Int J Mol Sci,* 23.

VILLANUEVA, J., PHILIP, J., ENTENBERG, D., CHAPARRO, C. A., TANWAR, M. K., HOLLAND, E. C. & TEMPST, P. 2004. Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. *Anal Chem,* 76**,** 1560-70.

WADDELL, N., PAJIC, M., PATCH, A. M., CHANG, D. K., KASSAHN, K. S., BAILEY, P., JOHNS, A. L., MILLER, D., NONES, K., QUEK, K., QUINN, M. C., ROBERTSON, A. J., FADLULLAH, M. Z., BRUXNER, T. J., CHRIST, A. N., HARLIWONG, I., IDRISOGLU, S., MANNING, S., NOURSE, C., NOURBAKHSH, E., WANI, S., WILSON, P. J., MARKHAM, E., CLOONAN, N., ANDERSON, M. J., FINK, J. L., HOLMES, O., KAZAKOFF, S. H., LEONARD, C., NEWELL, F., POUDEL, B., SONG, S., TAYLOR, D., WADDELL, N., WOOD, S., XU, Q., WU, J., PINESE, M., COWLEY, M. J., LEE, H. C., JONES, M. D., NAGRIAL, A. M., HUMPHRIS, J., CHANTRILL, L. A., CHIN, V., STEINMANN, A. M., MAWSON, A., HUMPHREY, E. S., COLVIN, E. K., CHOU, A., SCARLETT, C. J., PINHO, A. V., GIRY-LATERRIERE, M., ROOMAN, I., SAMRA, J. S., KENCH, J. G., PETTITT, J. A., MERRETT, N. D., TOON, C., EPARI, K., NGUYEN, N. Q., BARBOUR, A., ZEPS, N., JAMIESON, N. B., GRAHAM, J. S., NICLOU, S. P., BJERKVIG, R., GRÜTZMANN, R., AUST, D., HRUBAN, R. H., MAITRA, A., IACOBUZIO-DONAHUE, C. A., WOLFGANG, C. L., MORGAN, R. A., LAWLOR, R. T., CORBO, V., BASSI, C., FALCONI, M., ZAMBONI, G., TORTORA, G., TEMPERO, M. A., GILL, A. J., ESHLEMAN, J. R., PILARSKY, C., SCARPA, A., MUSGROVE, E. A., PEARSON, J. V., BIANKIN, A. V. &

GRIMMOND, S. M. 2015. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature,* 518**,** 495-501.

WANG, M. 2017. *Spectral Library Construction and Matching of MS/MS Spectra at Repository Scales*, University of California, San Diego.

WASSELL, J. 2000. Haptoglobin: function and polymorphism. *Clin Lab,* 46**,** 547-52.

WEST, H. & JIN, J. 2015. Neoadjuvant Therapy. *JAMA Oncology,* 1**,** 550-550.

WINTER, K., TALAR-WOJNAROWSKA, R., DĄBROWSKI, A., DEGOWSKA, M., DURLIK, M., GĄSIOROWSKA, A., GŁUSZEK, S., JURKOWSKA, G., KACZKA, A., LAMPE, P., MAREK, T., NASIEROWSKA-GUTTMEJER, A., NOWAKOWSKA-DUŁAWA, E., RYDZEWSKA, G., STRZELCZYK, J., ŚLEDZIŃSKI, Z. & MAŁECKA-PANAS, E. 2019. Diagnostic and therapeutic recommendations in pancreatic ductal adenocarcinoma. Recommendations of the Working Group of the Polish Pancreatic Club. *Prz Gastroenterol,* 14**,** 1-18.

XEPAPADAKI, E., KALOGEROPOULOU, C., ZVINTZOU, E., FILOU, S. & KYPREOS, K. 2019. High density lipoprotein: the role of apolipoprotein A2. *Hellenic Journal of Atherosclerosis,* 9.

XIE, Z., ZHANG, Y., JIN, C. & FU, D. 2018. Gemcitabine-based chemotherapy as a viable option for treatment of advanced breast cancer patients: a meta-analysis and literature review. *Oncotarget,* 9**,** 7148-7161.

XU, Q., ZHANG, T. P. & ZHAO, Y. P. 2011. Advances in early diagnosis and therapy of pancreatic cancer. *Hepatobiliary Pancreat Dis Int,* 10**,** 128-35.

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**,** 135-7.

YOKOYAMA, Y., NAGINO, M., HIROMATSU, T., YUASA, N., ODA, K., ARAI, T., NISHIO, H., EBATA, T. & NIMURA, Y. 2005. Intense PET signal in the degenerative necrosis superimposed on chronic pancreatitis. *Pancreas,* 31**,** 192-4.

ZENG, H., YU, H., LU, L., JAIN, D., KIDD, M. S., SAIF, M. W., CHANOCK, S. J., HARTGE, P. & RISCH, H. A. 2011. Genetic effects and modifiers of radiotherapy and chemotherapy on survival in pancreatic cancer. *Pancreas,* 40**,** 657-63.

ZHANG, N., REN, Y., WANG, Y., ZHAO, L., WANG, B., MA, N., GAO, Z. & CAO, B. 2020. LRG1 Suppresses Migration and Invasion of Esophageal Squamous Cell Carcinoma by Modulating Epithelial to Mesenchymal Transition. *J Cancer,* 11**,** 1486-1494.

ZHANG, R., LIU, Q., LI, T., LIAO, Q. & ZHAO, Y. 2019. Role of the complement system in the tumor microenvironment. *Cancer Cell Int,* 19**,** 300.

ZHANG, Y. V. & ROCKWOOD, A. 2015. Impact of automation on mass spectrometry. *Clin Chim Acta,* 450**,** 298-303.

ZHENG, D., CHEN, H., DAVIDS, J., BRYANT, M. & LUCAS, A. 2013. Serpins for diagnosis and therapy in cancer. *Cardiovasc Hematol Disord Drug Targets,* 13**,** 123-32.

ZHOU, J. Y., DANN, G. P., SHI, T., WANG, L., GAO, X., SU, D., NICORA, C. D., SHUKLA, A. K., MOORE, R. J., LIU, T., CAMP, D. G., 2ND, SMITH, R. D. & QIAN, W. J. 2012. Simple sodium dodecyl sulfate-assisted sample preparation method for LC-MS-based proteomics applications. *Anal Chem,* 84**,** 2862-7.

ZHOU, Q., ANDERSSON, R., HU, D., BAUDEN, M., SASOR, A., BYGOTT, T., PAWŁOWSKI, K., PLA, I., MARKO-VARGA, G. & ANSARI, D. 2019. Alpha-1-acid glycoprotein 1 is upregulated in pancreatic ductal adenocarcinoma and confers a poor prognosis. *Translational Research,* 212**,** 67-79.

ZHOU, Q. & MELTON, D. A. 2018. Pancreas regeneration. *Nature,* 557**,** 351-358.

ZHU, S., ZHANG, X., GAO, M., HONG, G., YAN, G. & ZHANG, X. 2012. Developing a strong anion exchange/RP (SAX/RP) 2D LC system for high-abundance proteins depletion in human plasma. *Proteomics,* 12**,** 3451-63.

APPENDIX

Table S 1. A list of all buffer solutions and solvents with their respective components and company manufacturer.

Buffer solution	**Solvents/components**	**Company**
Running MES buffer	1. MES	1. Thermo-Fisher Scientific
50 mM Tris-HCl, 2% SDS, pH 8	2. SDS (Sodium dodecyl sulfate) 3. Tris-HCl (Tris hydrochloride)	1. SIGMA 2. SIGMA
50 mM TEAB	4. TEAB	3. SIGMA
95% ACN	5. ACN (Acetonitrile) 6. NH_4Ac (Ammonium acetate)	4. SIGMA 5. SIGMA
30% ACN, 200 mM NH_4Ac, pH 4.5		
15% ACN, 200 mM NH_4Ac, pH 4.5		
1% TFA	7. TFA (Trifluoroacetic acid)	6. SIGMA
70% EtOH	8. EtOH (Ethanol)	7. SIGMA
2% ACN, 0.2% FA	9. ACN (Acetonitrile) 10. FA (Formic acid)	8. SIGMA 9. SIGMA

Table S 2. Patient demographic characteristics and clinical stage of disease.

	Sample ID	**Clinical staging**	**Age**	**Gender**

PDAC	P1	RPC	64	Male
	P2	RPC	73	Male
	P3	RPC	68	Female
	P4	RPC	63	Male
	P5	RPC	55	Male
	P6	LAPC	64	Female
	P7	RPC	68	Male
	P8	RPC	44	Male
	P9	MPC	66	Male
	P10	Borderline RPC	41	Female
	P11	RPC	44	Male
	P12	LAPC	57	Male
	P13	RPC	50	Female
	P14	RPC	65	Male
	P15	LAPC	39	Male
	P16	RPC	62	Male
	P17	MPC	81	Female
	P18	RPC	46	Male
	P19	RPC	72	Male
	P20	LAPC	62	Male
	P21	RPC	66	Male
	P22	RPC	65	Male
	P23	MPC	46	Male
	P24	RPC	68	Male
	P25	LAPC	68	Male

	P26	LAPC	56	Female
	P27	RPC	66	Female
	P28	LAPC	29	Male
	P29	LAPC	51	Male
	P30	MPC	40	Male
	P31	RPC	38	Female
	P32	RPC	52	Female
	P33	RPC	51	Male
	P34	RPC	69	Female
Chronic pancreatitis	CP1	*	45	Male
	CP2	*	46	Male
	CP3	*	60	Male
	CP4	*	58	Male
	CP5	*	47	Male
	CP6	*	55	Male
Healthy	HC1	*	45	Female
	HC2	*	56	Male
	HC3	*	28	Female
	HC4	*	29	Male
	HC5	*	28	Male
	HC6	*	33	Male

* Denotes disease/patients with no clinical staging. RPC: Resectable pancreatic cancer,

MPC: Metastatic pancreatic cancer, LAPC: Locally advanced pancreatic cancer, HC: Health control, CP: Chronic pancreatitis, The highlighted PDAC patients were selected for the pilot study. PDAC patients highlighted in green were selected for the pilot study.

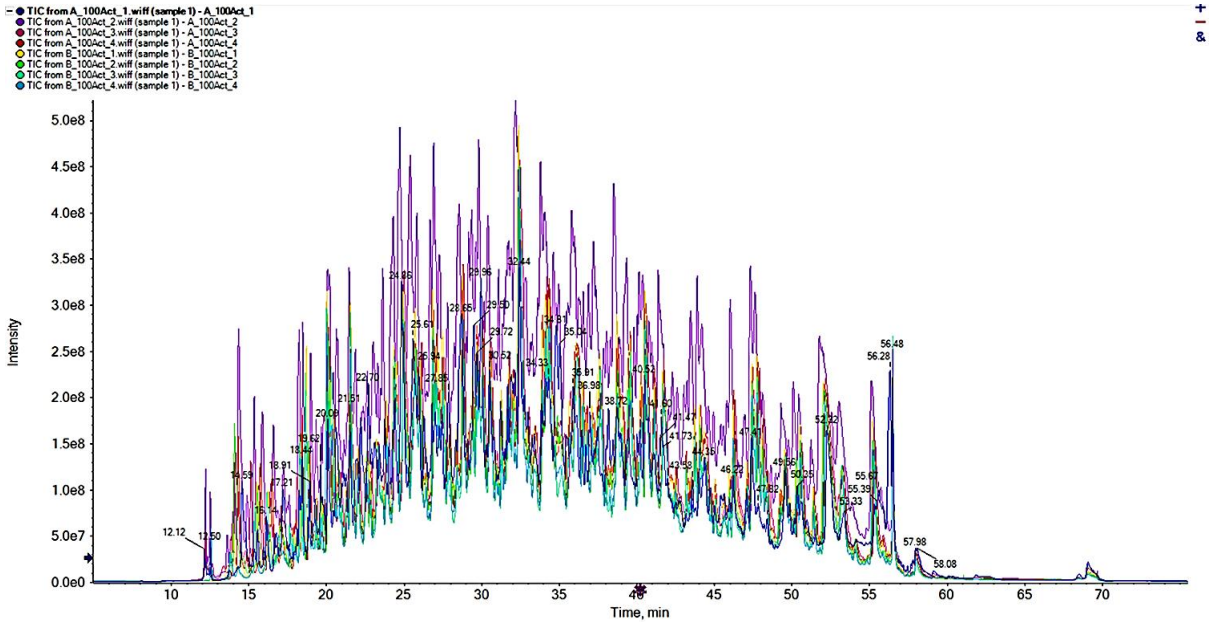


Figure S 1. MS1 total ion chromatograms (TICs) of 100% acetone precipitation replicates overlaid using Peak View. All 100% acetone precipitation TICs from batches A and B overlaid. LC-MS: 500 ng load, 45 min gradient, SWATH 60VW, 15 ms cycle time, 75 min run.

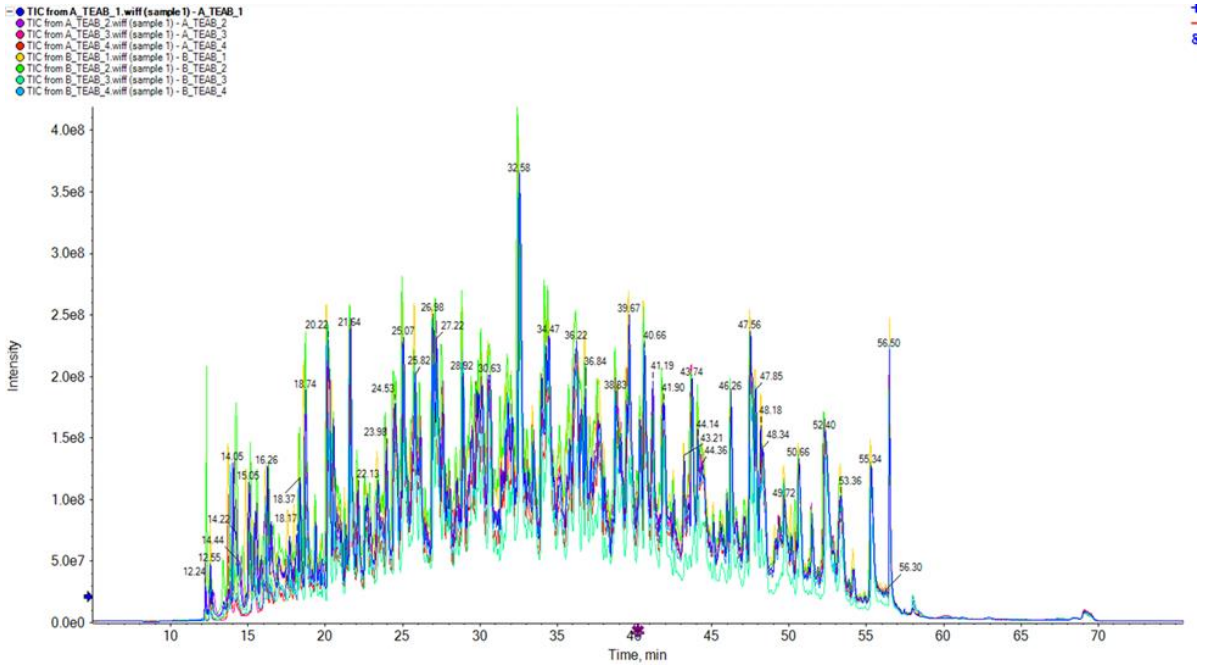


Figure S 2. MS1 total ion chromatograms (TICs) of 5x TEAB replicates overlaid using Peak View. All 5x TEAB TICs from batches A and B overlaid. LC-MS: 500 ng load, 45 min gradient, SWATH 60VW, 15 ms cycle time, 75 min run.

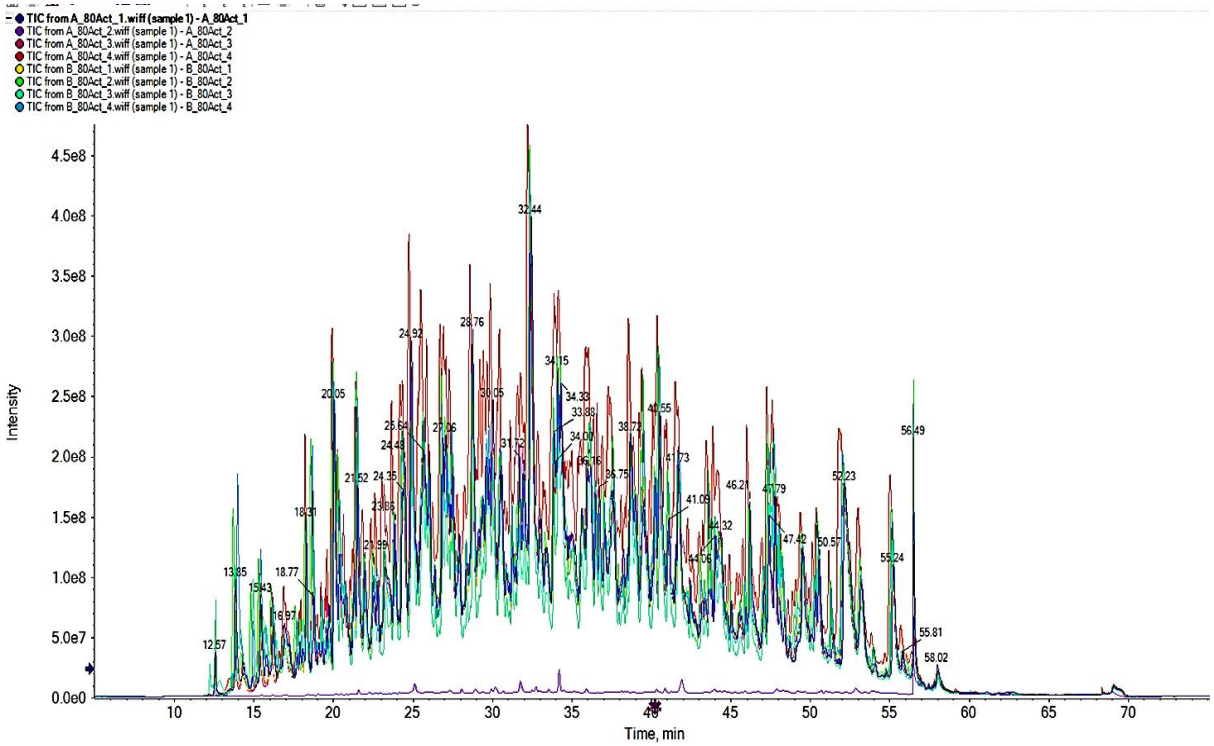


Figure S 3. MS1 total ion chromatograms (TICs) of 80% acetone precipitation replicates overlaid using Peak View. All 80% acetone precipitation TICs from batch A and B overlaid. LC-MS: 500 ng load, 45 min gradient, SWATH 60VW, 15 ms cycle time, 75 min run.

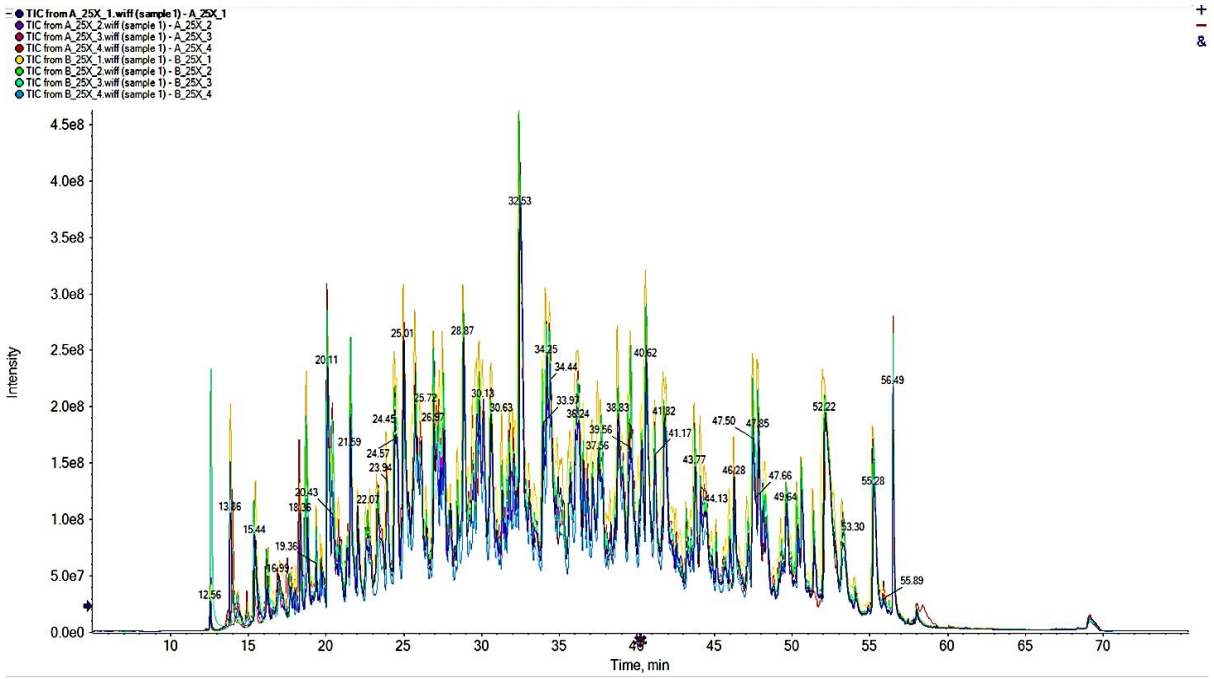


Figure S 4. MS1 total ion chromatograms (TICs) of 25x Tris-HCl replicates overlaid using Peak View. All 25x Tris-HCl TICs from batches A and B overlaid. LC-MS: 500 ng load, 45 min gradient, SWATH 60VW, 15 ms cycle time, 75 min run.

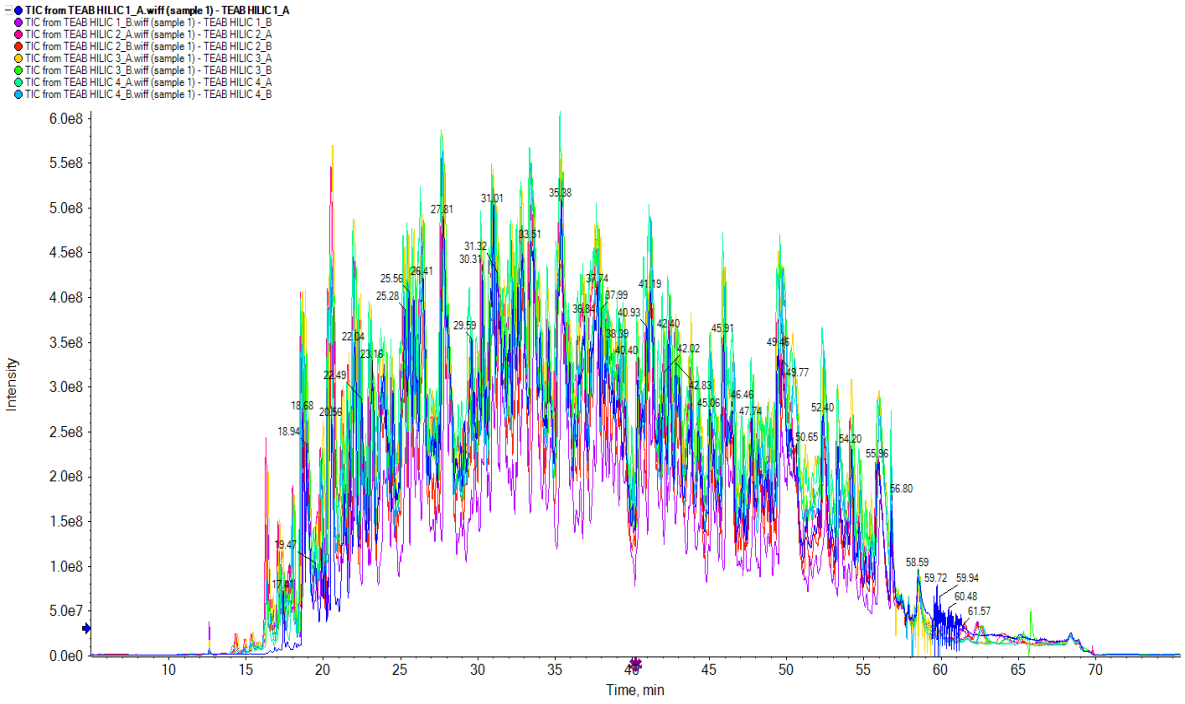


Figure S 5. MS1 total ion chromatograms (TICs) of TEAB_HILIC (TEAB_H) replicates overlaid using Peak View. All TEAB_H TICs from batches A and B overlaid. LC-MS: 500 ng load, 45 min gradient, SWATH 60VW, 15 ms cycle time, 75 min run.

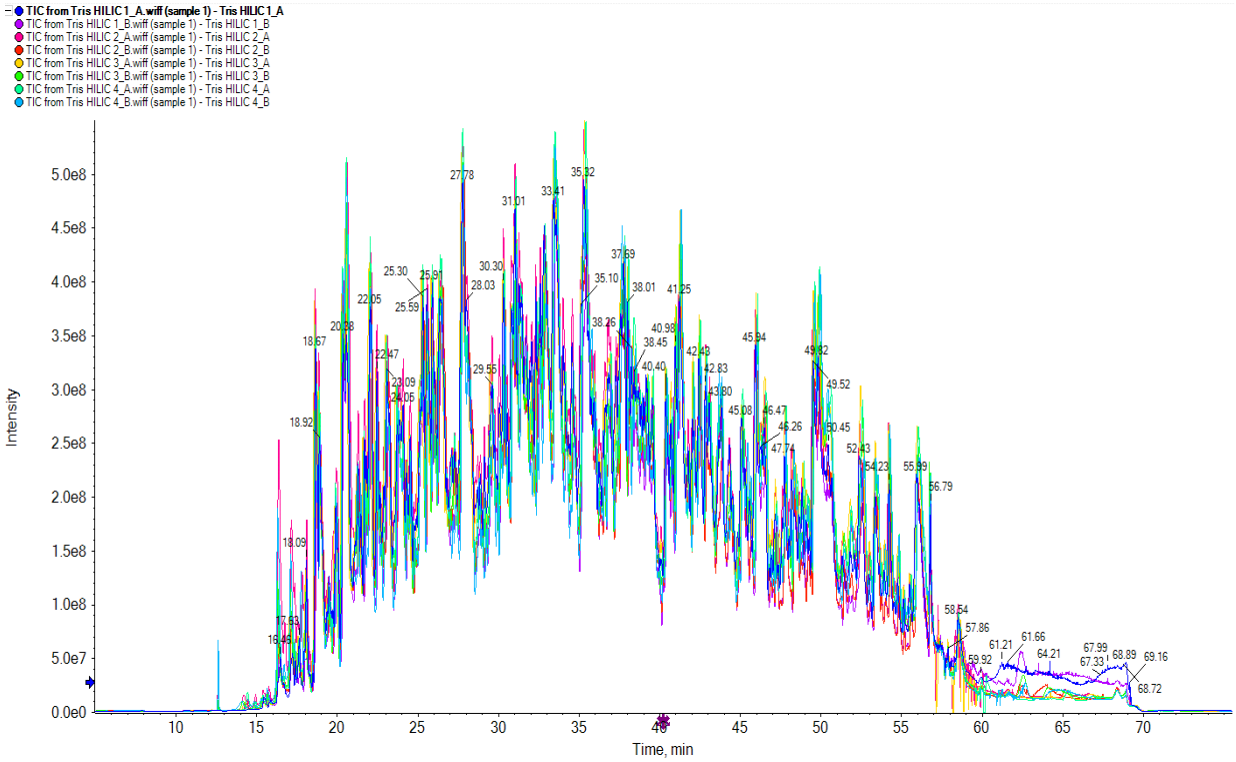


Figure S 6. MS1 total ion chromatograms (TICs) Tris_HILIC (Tris_H) replicates overlaid using Peak View. All Tris_H TICs from batches A and B overlaid. LC-MS: 500 ng load, 45 min gradient, SWATH 60VW, 15 ms cycle time, 75 min run.

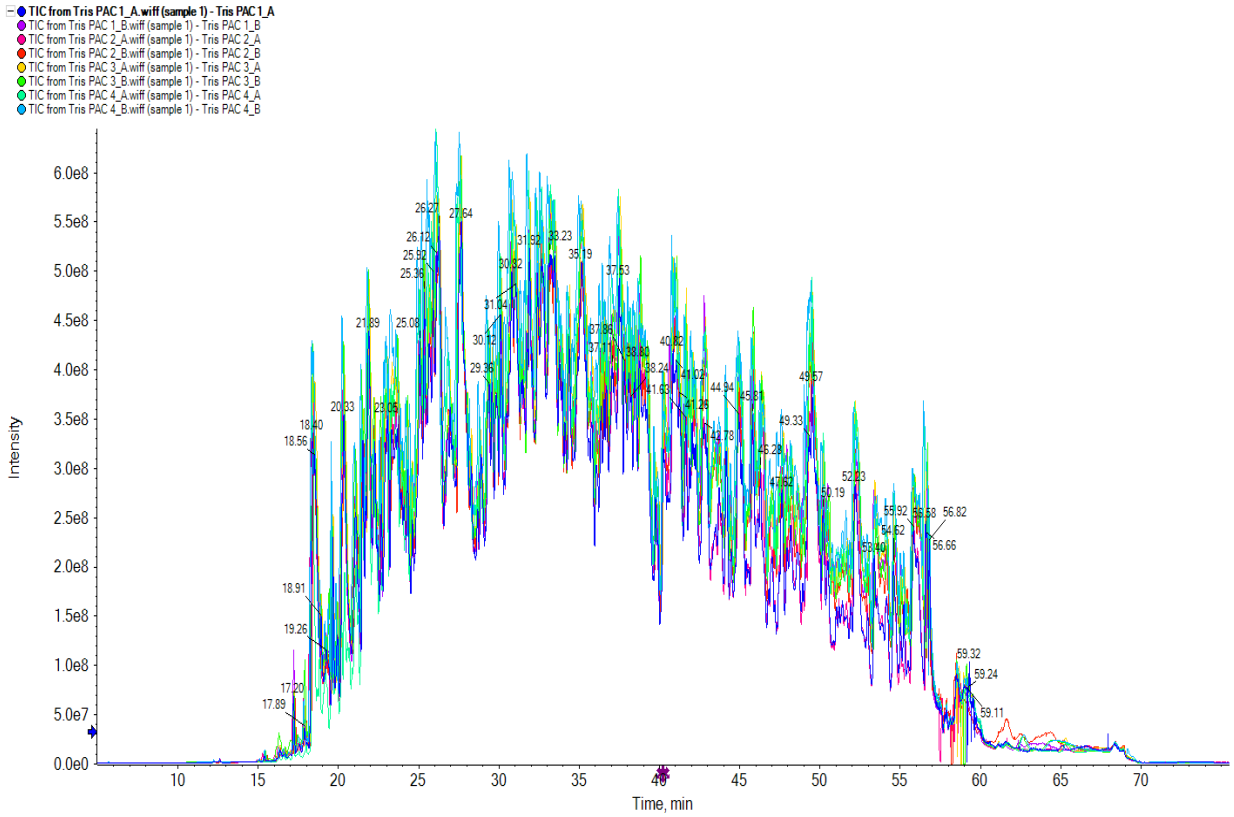


Figure S 7. MS1 total ion chromatograms (TICs) of Tris_PAC (Tris_P) replicates overlaid using Peak

View. All Tris_P TICs from batches A and B overlaid. LC-MS: 500 ng load, 45 min gradient, SWATH 60VW, 15 ms cycle time, 75 min run.

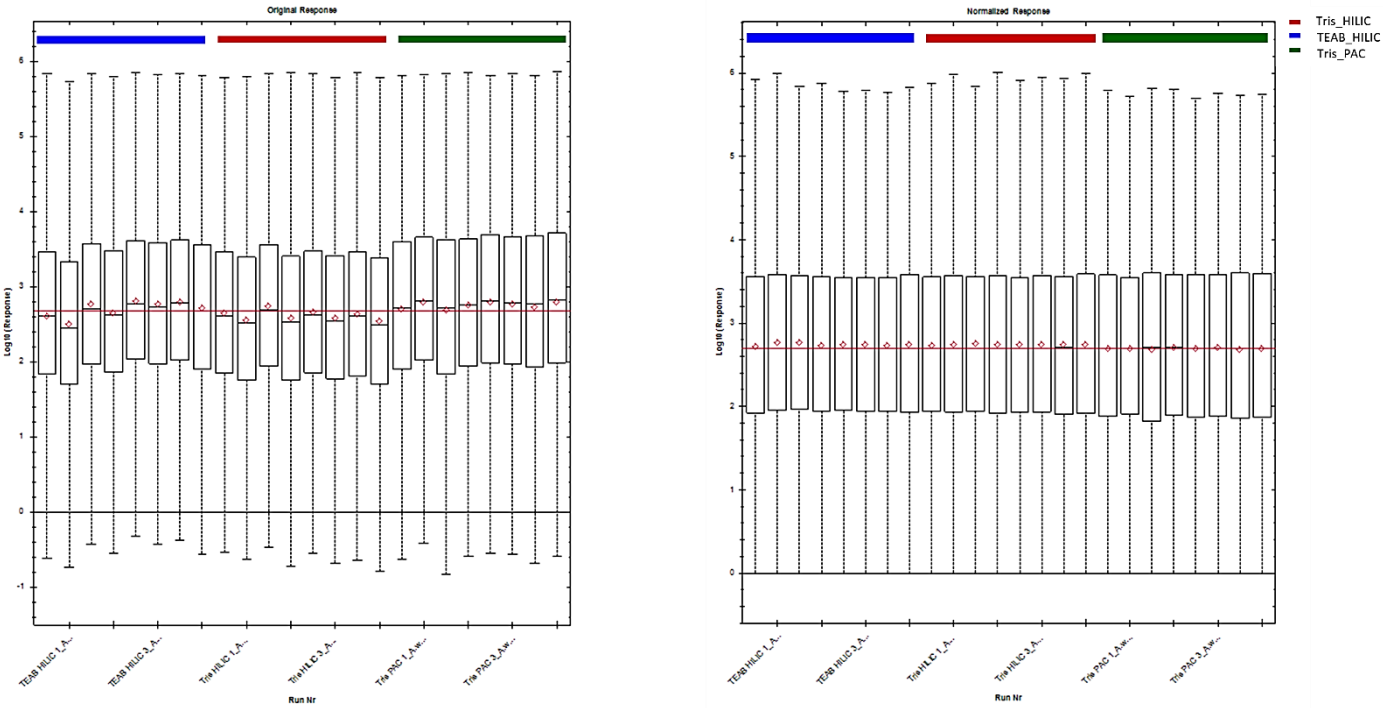


Figure S 8. Normalization plot for experiment 2. Box plots show the normal response (signal response before normalization) of all precursor ion quantities for each run (left) and the normalization of the same precursor ion quantities for each run (right).

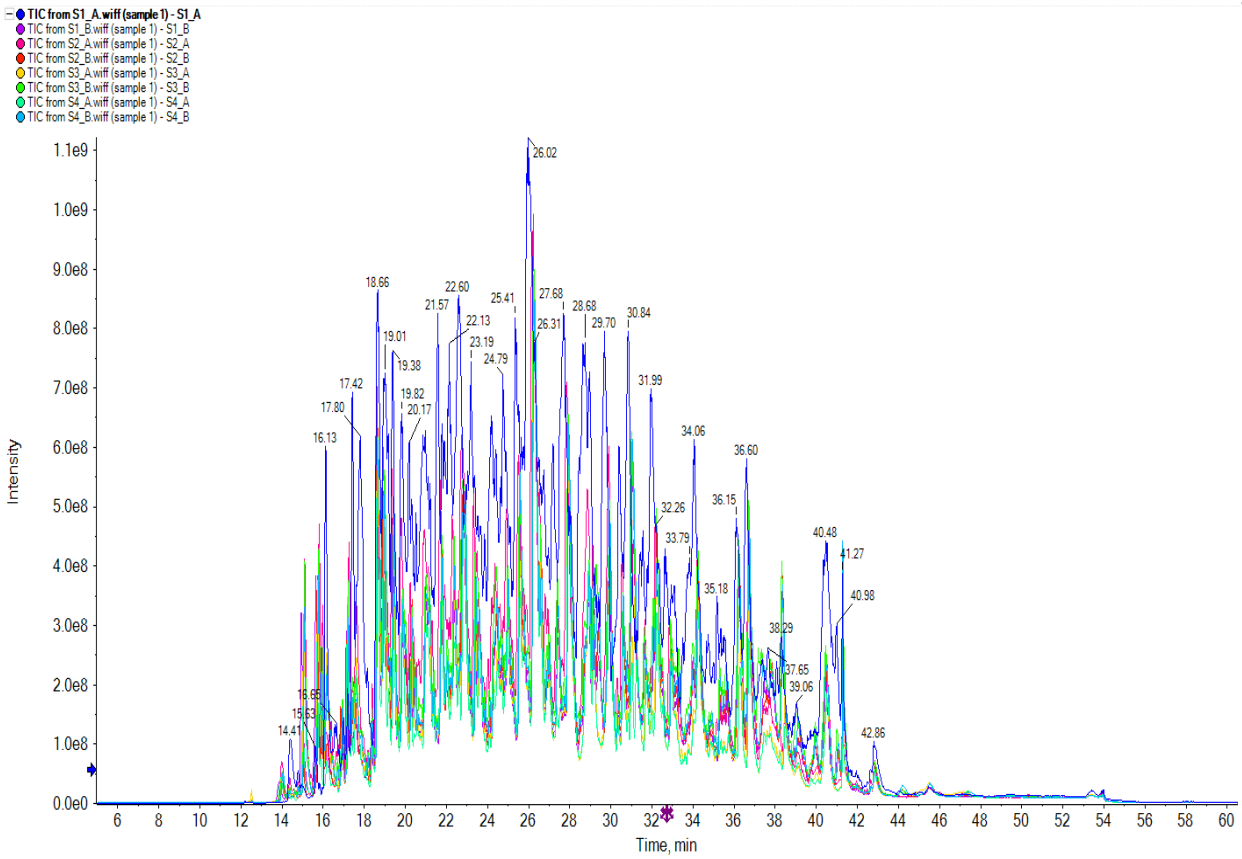


Figure S 9. MS1 total ion chromatograms (TICs) of 4 hrs digestion replicates overlaid using Peak View. All 4 hrs digestion TICs from batches A and B overlaid. LC-MS: 500 ng load, 60 min gradient, SWATH 60VW, 15 ms cycle time.

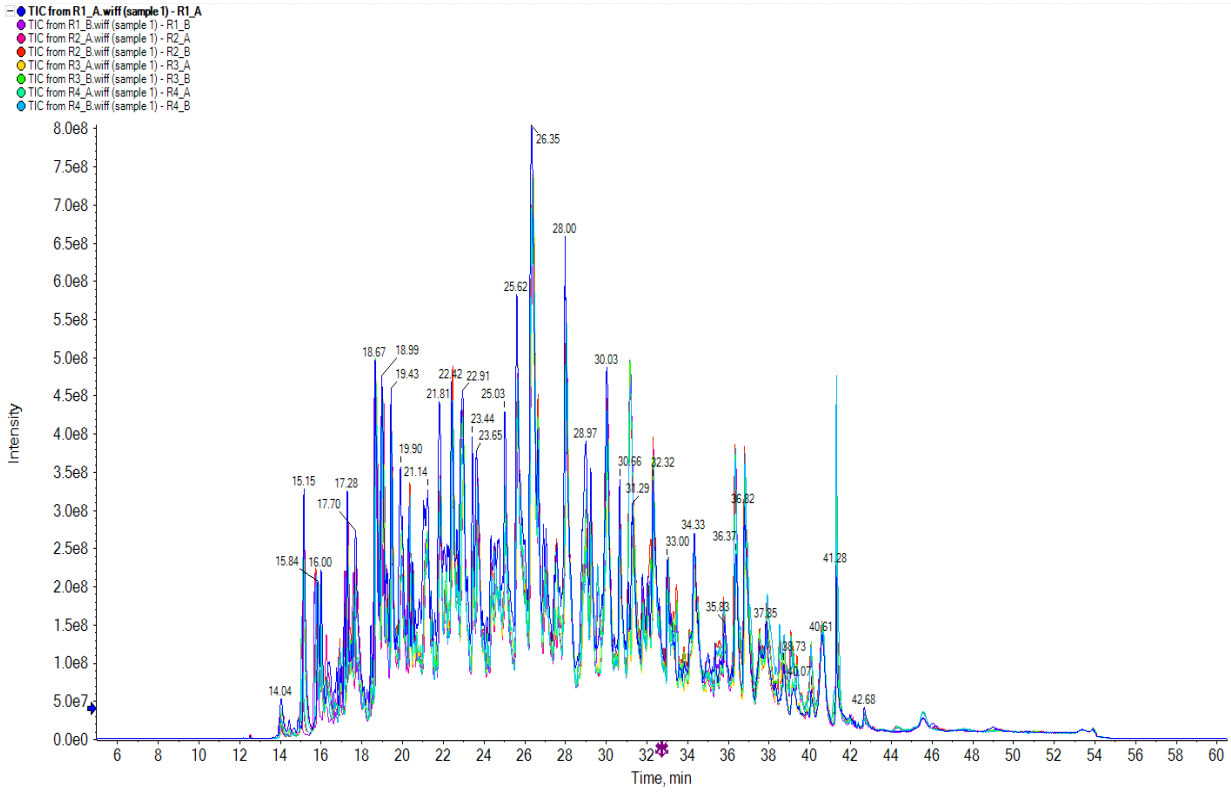


Figure S 10. MS1 total ion chromatograms (TICs) of 16 hrs digestion replicates overlaid using Peak View. All 16 hrs digestion TICs from batches A and B overlaid. LC-MS: 500 ng load, 60 min gradient, SWATH 60VW, 15 ms cycle time.

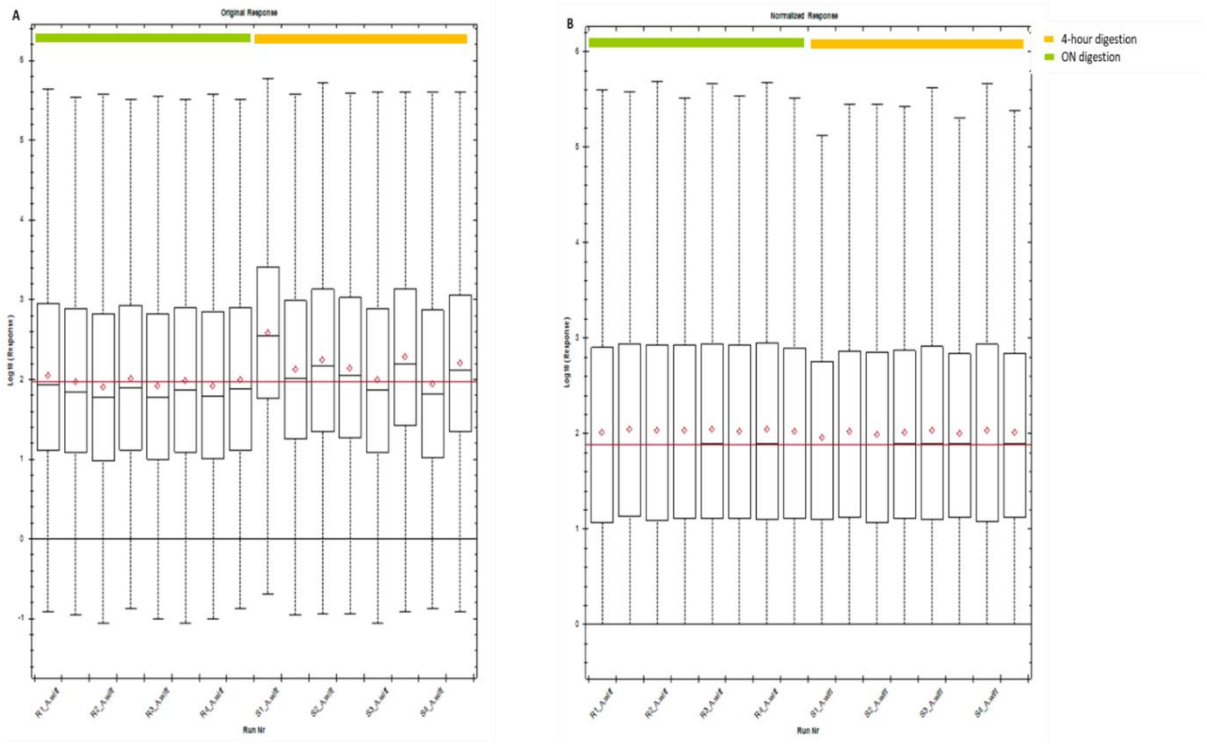


Figure S 11. Normalization plot for experiment 3. Box plots showing the normal response (signal response before normalization) of all precursor ion quantities for each run (left) and the normalization of the same precursor ion quantities for each run (right).

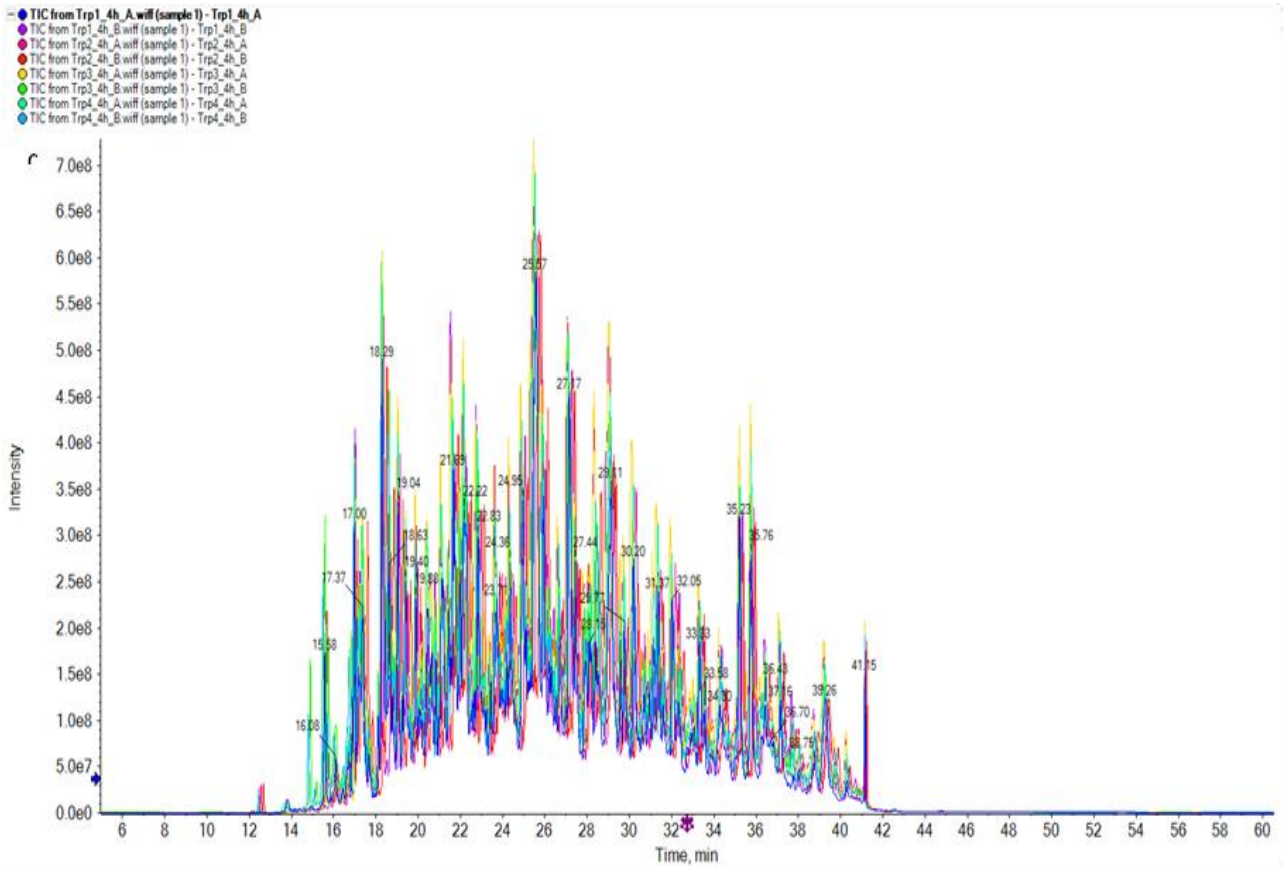


Figure S 12. MS1 total ion chromatograms (TICs) of 4 hrs trypsin digestion replicates overlaid using Peak View. All 4 hrs trypsin digestion TICs from batch A and B overlaid. LC-MS: 500 ng load, 60 min gradient, SWATH 60VW, 15 ms cycle time.

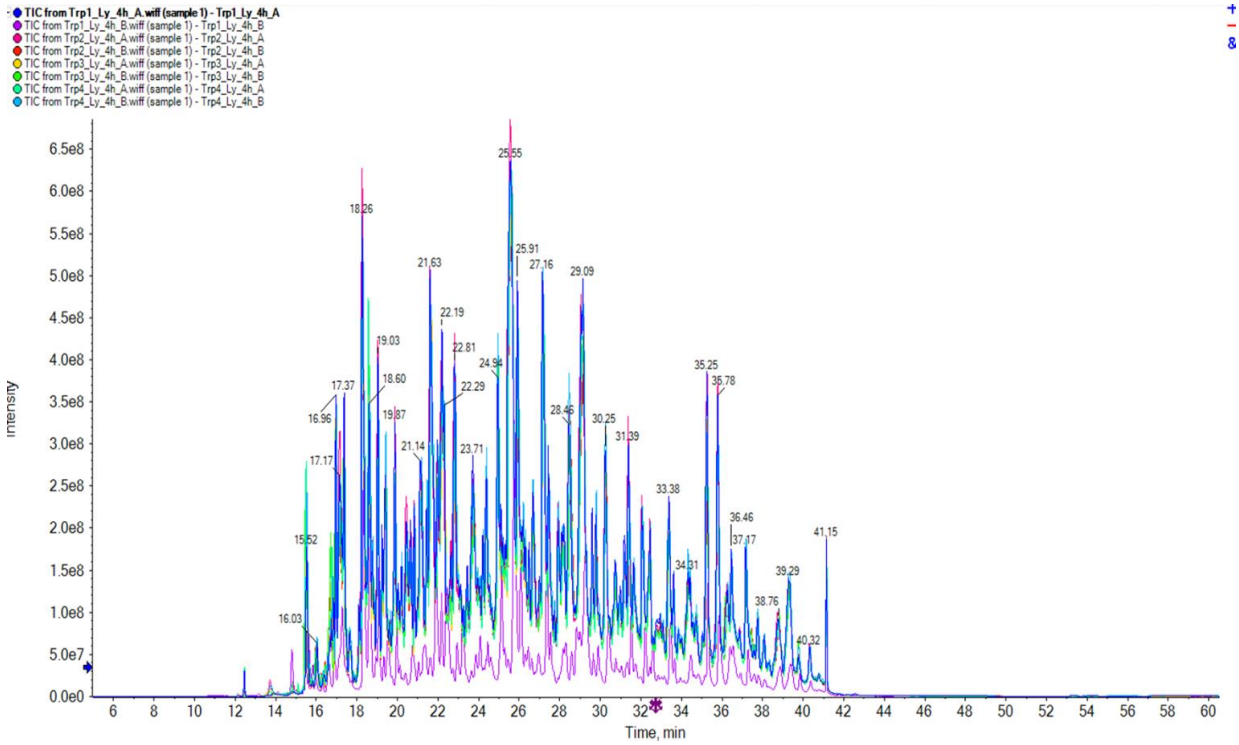


Figure S 13. MS1 total ion chromatograms (TICs) of 4-hour trypsin-LysC digestion replicates overlaid using Peak View. All 4-hour trypsin-LysC digestion TICs from batches A and B overlaid. LC-MS: 500ng load, 60 min gradient, SWATH 60VW, 15 ms cycle time.

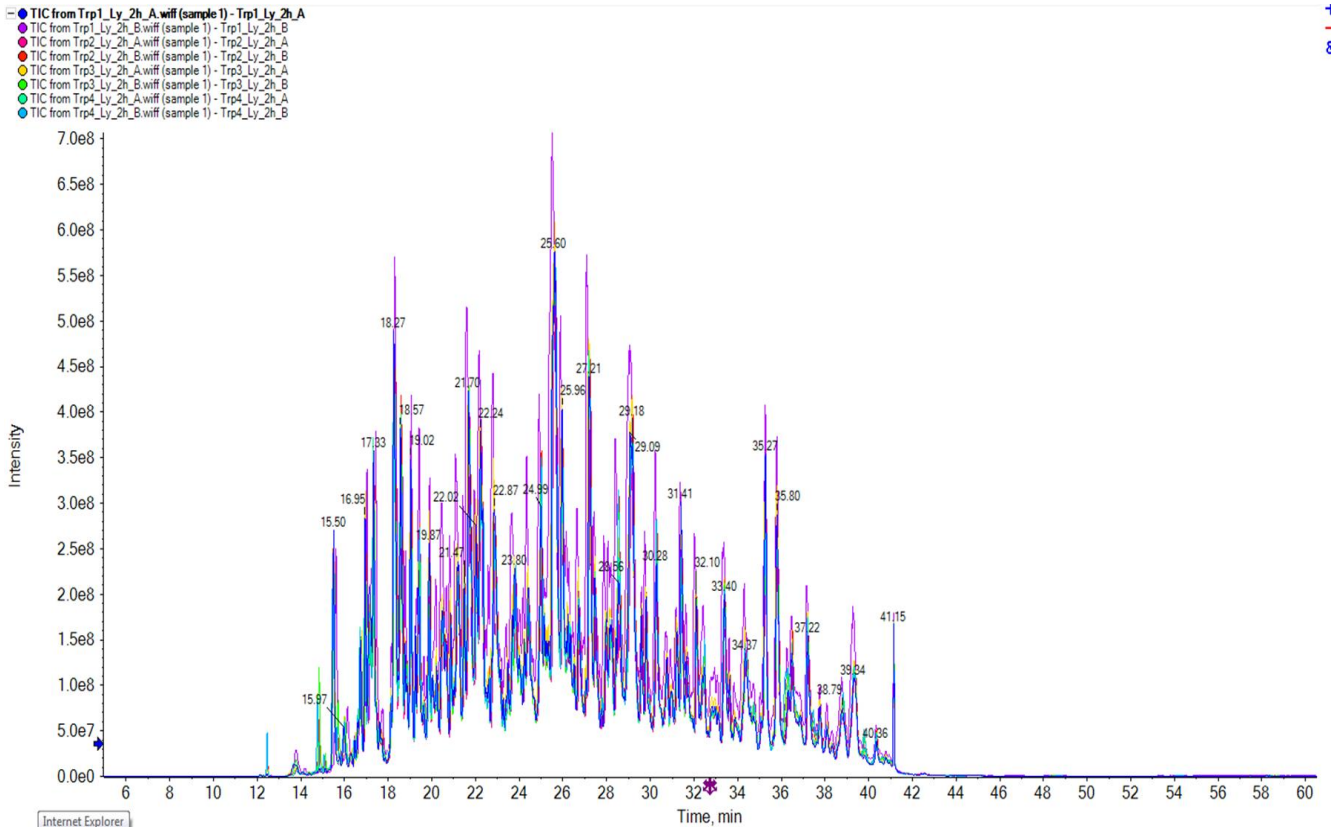


Figure S 14. MS1 total ion chromatograms (TICs) of 2 hrs trypsin-LysC digestion replicates overlaid using Peak View. All 2 hrs trypsin-LysC digestion TICs from batches A and B overlaid. LC-MS: 500 ng load, 60 min gradient, SWATH 60VW, 15 ms cycle time.

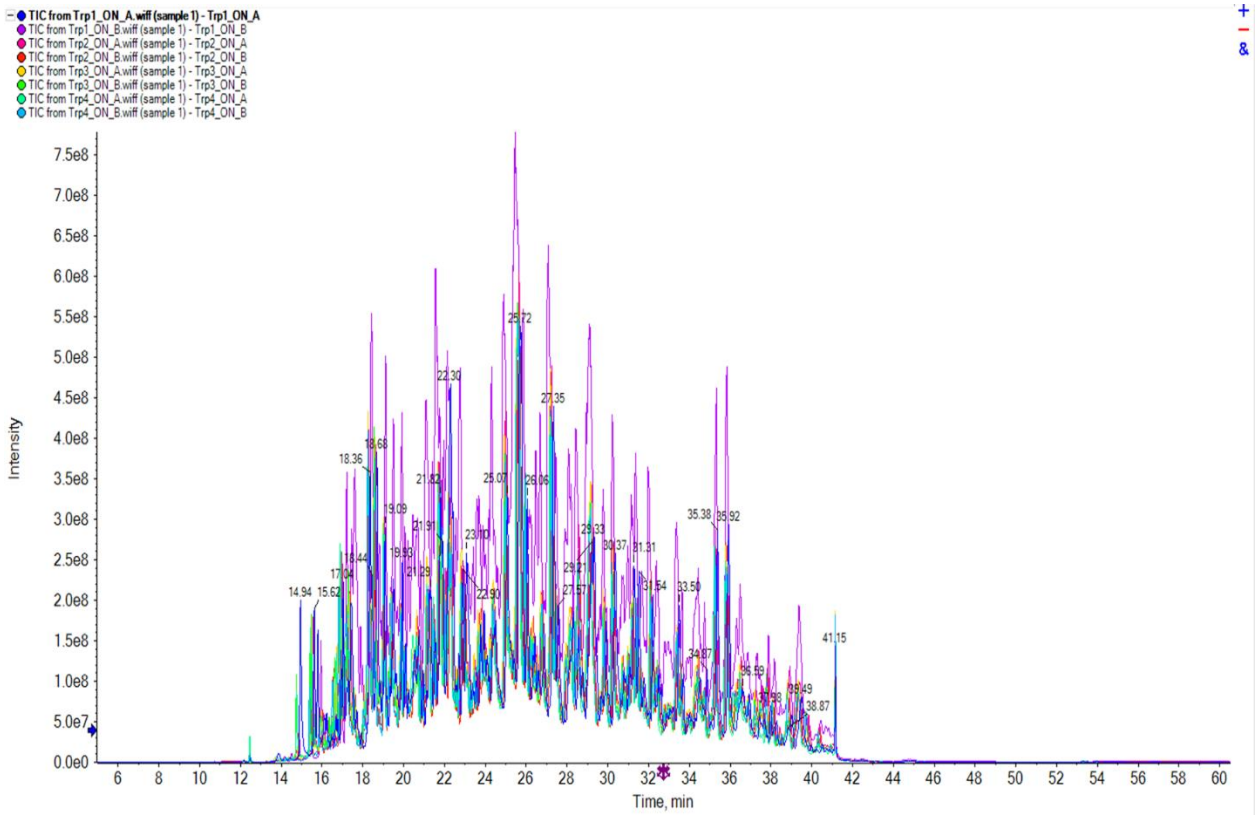


Figure S 15. MS1 total ion chromatograms (TICs) of 16 hrs trypsin digestion replicates overlaid using

Peak View. All 16 hrs trypsin digestion TICs from batches A and B overlaid. LC-MS: 500 ng load, 60 min gradient, SWATH 60VW, 15 ms cycle time.

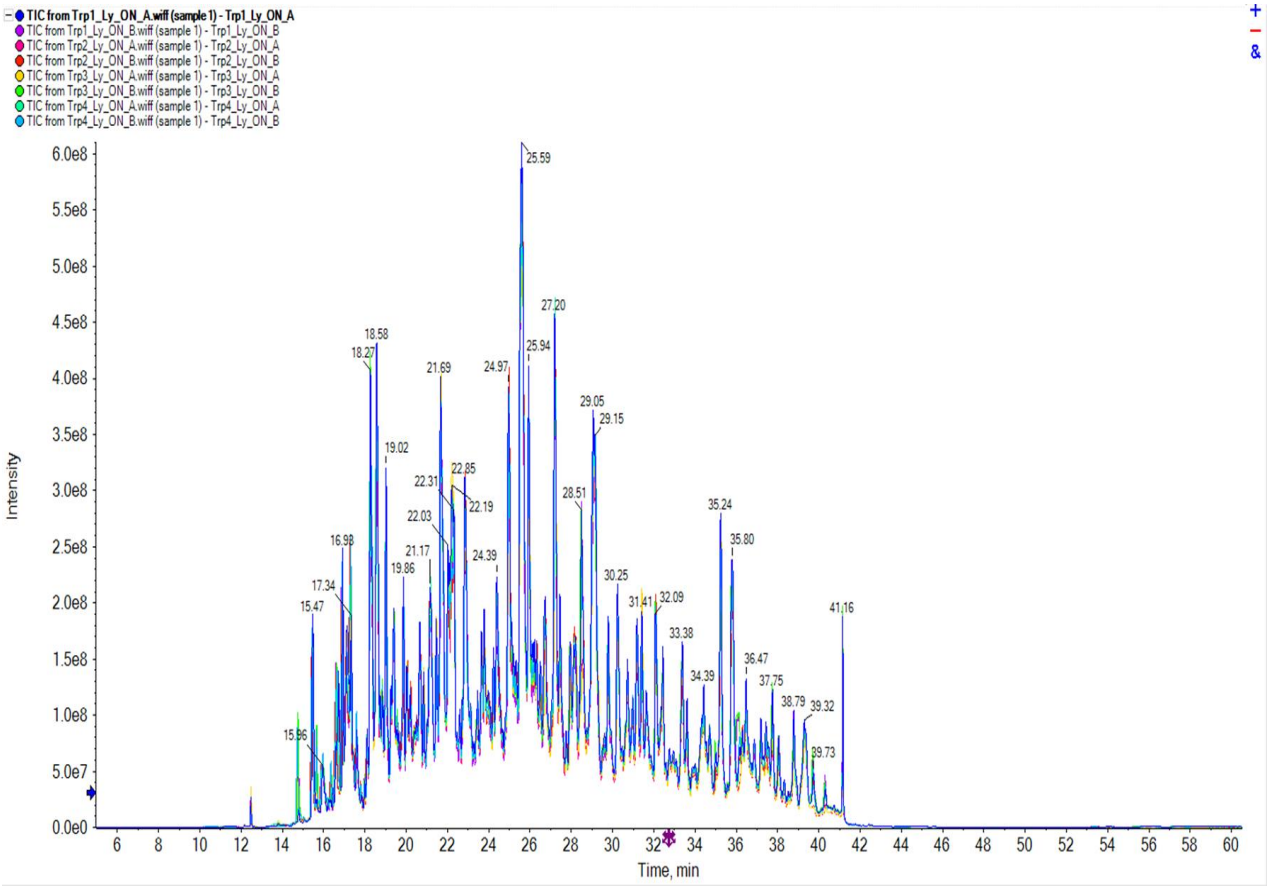


Figure S 16. MS1 total ion chromatograms (TICs) of 16 hrs trypsin-LysC digestion replicates overlaid using Peak View. All 16 hrs trypsin-LysC digestion TICs from batches A and B overlaid. LC-MS: 500 ng load, 60 min gradient, SWATH 60VW, 15 ms cycle time.

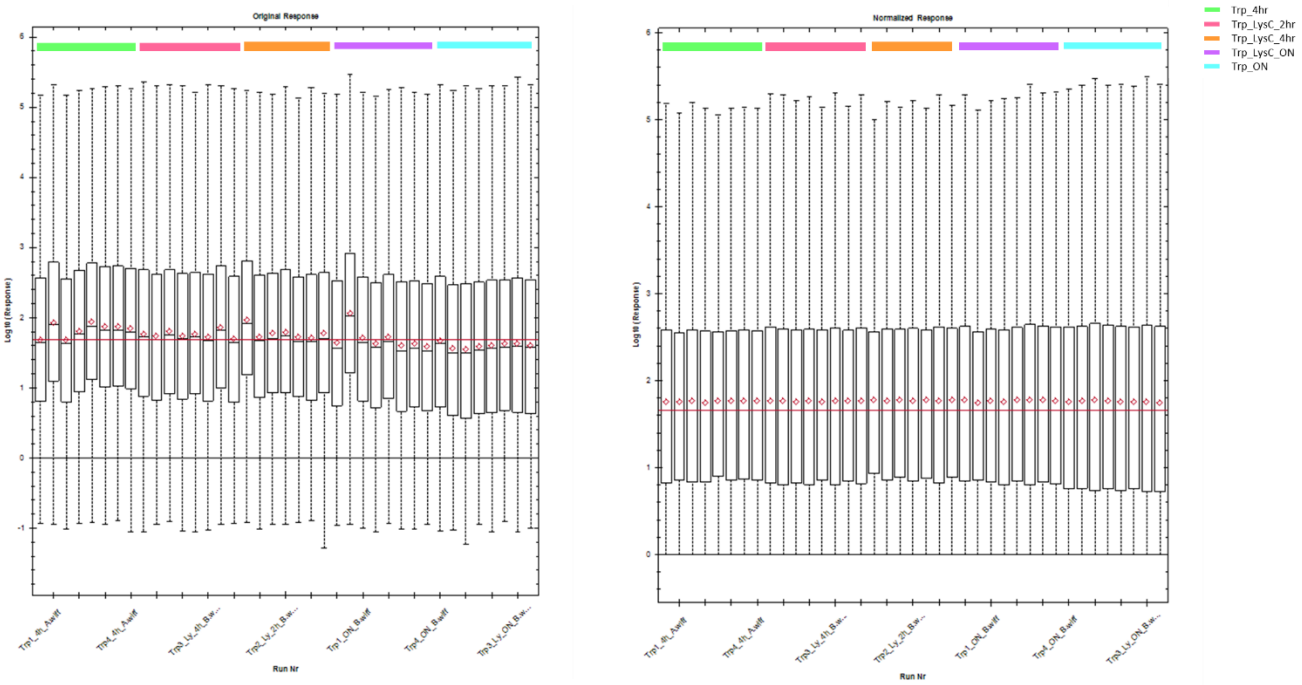


Figure S 17. Normalization plot for experiment 4. Box plots showing the normal response (signal response before normalization) of precursor ion quantities for each run (left) and the normalization of the same precursor ion quantities for each run (right).

Table S 3. Power analysis for a minimum of 6 samples.

desiredFC	numSample	FDR	power	CV
3.150	6	0.01	0.78	0.039
3.175	6	0.01	0.79	0.039
3.200	6	0.01	0.80	0.038
3.225	6	0.01	0.81	0.038
3.250	6	0.01	0.82	0.038
3.275	6	0.01	0.83	0.038
3.300	6	0.01	0.84	0.037
3.325	6	0.01	0.84	0.037
3.350	6	0.01	0.85	0.037
3.375	6	0.01	0.86	0.036
3.400	6	0.01	0.87	0.036
3.425	6	0.01	0.87	0.036
3.450	6	0.01	0.88	0.036
3.475	6	0.01	0.89	0.035

3.500	6	0.01	0.89	0.035
3.525	6	0.01	0.90	0.035
3.550	6	0.01	0.90	0.035

Table S 4. Power analysis identifying the minimum number of samples required for a power of 0.9 and a fold change of 1.5.

desiredFC	**numSample**	**FDR**	**power**	**CV**
1.500	58	0.01	0.9	0.008
1.525	54	0.01	0.9	0.009
1.550	50	0.01	0.9	0.010
1.575	46	0.01	0.9	0.010
1.600	43	0.01	0.9	0.011
1.625	41	0.01	0.9	0.011
1.650	38	0.01	0.9	0.012
1.675	36	0.01	0.9	0.012
1.700	34	0.01	0.9	0.013
1.725	32	0.01	0.9	0.013
1.750	31	0.01	0.9	0.014
1.775	29	0.01	0.9	0.014
1.800	28	0.01	0.9	0.015
1.825	26	0.01	0.9	0.016
1.850	25	0.01	0.9	0.016

1. Final SOP

INTRODUCTION: Diluting samples in Tris-HCL helps in efficiently removing contaminants that may be found in the sample. SDS mainly aids in solubilizing the proteins. However, SDS interferes with LC-MS analysis and therefore must be removed prior to the analysis using the MagResyn HILIC workflow described below.

This protocol describes the efficient sample preparation of plasma with clean-up using the MagResyn HILIC for bottom-up proteomics applications.

MATERIALS:

- MagResyn® HILIC – Catalogue Number MR-HL002
- Eppendorf LoBind microcentrifuge tubes (0.5 and 2ml)
- Magnetic separator or Magnetic bead handling station (e.g., KingFisher™)
- Pipettes
- Sodium dodecyl sulphate (SDS)
- Tis (hydroxymethyl) aminomethane (Tris-HCL)
- Iodoacetamide (IAA)
- Dithiothreitol (DTT)
- Acetonitrile (ACN)
- Ammonium acetate (NH_4Ac)
- Ammonium bicarbonate ($NH_4Bicarb$)
- Protein quantification method (BCA Assay)
- MS grade water

REAGENT PREPARATION:

- DTT Stock: 1M DTT (prepare fresh)
- IAA Stock: 1M IAA (prepare fresh, light-sensitive)
- Plasma dilution buffer: 2% SDS in 50mM Tris-HCL pH 8
- Equilibration Buffer: 15% ACN, 100mM NH_4Ac pH 4.5 (diluted from 1M NH_4Ac stock)
- Binding Buffer: 30% ACN, 200 NH_4Ac pH 4.5
- Wash Buffer; 95% Acetonitrile
- Trypsin Digestion Buffer: 25mM $NH_4Bicarb$

METHOD:

Notes:

- *The current methods are suitable for manual preparation or may be fully automated on magnetic bead handling systems such as KingFisher™ or similar*
- *The current protocol is sufficient for the clean-up of 20µg of plasma sample*

DILUTION AND SOLUBILIZATION OF PLASMA PROTEINS

1**.** Thaw the plasma sample on ice slowly

2. Transfer 20µl of sample to a 2ml Eppendorf® LoBind tube

3. Add 480µl of 50mM Tris-HCL, 2% SDS (25-fold dilution)

4. Quantify diluted plasma using the Pierce BCA Protein Assay kit from Thermo Fisher scientific

5. Spike 250ug of sample with 0,8ug of casein and 0,2ug of enolase

6. Transfer 20ug of sample to a 2ml Eppendorf® LoBind tube

7. Chemically reduce the sample using a final concentration of 10mM DTT for 30min at 60°C

8. Alkylate sample using a final concentration of 40mM IAA for 30min at room temperature in the dark

9. Quench the residual IAA with 10mM DTT

10. The sample is now ready for on-bead protein capture, clean-up and digestion using HILIC protocol

MICROPARTICLE EQUILIBRATION

11**.** Transfer 10µl MagResyn® HILIC beads (completely re-suspended) to a 2ml LoBind tube

12. Place the tube on a magnetic separator and allow 5-10 sec for the microparticles to clear

13. Remove the shipping solution by aspiration with a pipette and discard

14. wash the microparticles by re-suspending in 200µl of equilibration buffer with agitation (e.g. gentle vortex mixing) for 15-30 sec.

15. Place the tube on the magnetic separator and allow the microparticles to clear.

16. Remove the equilibration solution by aspiration with a pipette and discard.

17. Repeat steps 14-15.

Note: only remove the 2nd equilibration solution from the microparticles once the sample is ready to be added (see step 18 below). This will ensure that microparticles do not dry and can be easily resuspended when the sample is added.

SAMPLE BINDING AND WASHING

18. Mix 20µg (for automated protocol, makeup with Equilibration buffer to 50µl) of reduced and alkylated plasma sample with 50µl binding buffer**.**

19. Add the plasma-bind buffer mix to the equilibrated microparticles.

20. Incubate for 30min at room temperature in intellimixer to ensure adequate sample and microparticle interaction.

21. Place the tube on the magnetic separator and allow the microparticles to clear. Remove and discard the unbound fraction by aspiration with a pipette.

22. Wash the microparticles with 200µl of wash buffer and mix for 60 sec with gentle agitation.

23. Place the tube on a magnetic separator and allow 5-10 sec for the microparticles to clear. Remove the supernatant and discard.

24. Repeat steps 22-23.

ON-BEAD PROTEIN DIGESTION

25. Perform on-bead digestion by adding 200µl of 25mM Ammonium Bicarbonate 0.2µg sequencing grade LysC (1:100 enzyme:protein ratio) and 1µg sequencing grade Trypsin (1:20 enzyme: protein ration) and incubate for 2 hrs at 47 ℃. Ensure sufficient mixing to keep the particles in solution during digestion.

26. Place the tube on a magnetic separator and allow 5-10 sec for the microparticles to clear.

27. Remove peptide solution and place in a 0.5 ml Eppendorf LoBind tube.

28. Incubate microparticles with 100µl of 1% TFA for 5 min whilst mixing and subsequently to pool the supernatant with digest solution from step 27.

29. Vacuum dry samples and re-suspend in 40µl LCMS solvent e.g.2% ACN with 0.2% formic acid.

96 DEEP-WELL PLATE SET-UP

ROW A	300µl of 25mM Ammonium Bicarbonate pH 8.0 containing Trypsin and LysC
ROW B	100µl of 1% TFA
ROW C	200µl of HILIC wash buffer: 95% ACN
ROW D	200µl of HILIC wash buffer: 95% ACN
ROW E	Equal volumes (minimum 50µl) of starting protein solution and HILIC binding buffer: 30% CAN in 200mM NH_4Ac pH 4.5
ROW F	500µl of HILIC equilibration buffer: 15% ACN in 100mM NH_4Ac pH 4.5
ROW G	200µl of HILIC equilibration buffer: 15% ACN in 100mM NH_4Ac pH 4.5 and 10l of HILIC beads
ROW H	Disposable plastic combs

30. Perform a Peptide Assay to determine peptide recovery and adjust the loads for LC-MS analysis.

31. Analyze digest by LC-MSMS

2. Ethics approval certificate



ethics approval.pdf

3. Plagiarism report



Turnitin plagiarism report.docx.pdf

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

Office of the Deputy Vice-Chancellor (Research and Innovation)

TO: Ms S Dubazana
School of Clinical Medicine
Department of Surgery
Medical School
University

E-mail: snedubazana@gmail.com

CC: Supervisor: Dr E Nweke
<Ekene.Nweke@wits.ac.za>
and <HREC-Medical Research Office@wits.ac.za>

FROM: Mr Iain Burns
Human Research Ethics Committee (Medical)
Tel: 011 717 1252

E-mail: Iain.Burns@wits.ac.za

DATE: 2022/03/28

REF: R14/49

PROTOCOL NO: **M211158** (This is your ethics application reference number. Please quote it in all enquiries, oral or written, relating to this study.)

PROJECT TITLE: *Method optimization for verification of candidate pancreatic ductal adenocarcinoma reporter proteins in plasma*

Please find attached the Clearance Certificate for the above project. I hope it goes well and that an article in a recognized publication comes out of it. This will reflect well on your professional standing and contribute to Government funding of the University.

MSWorks2000/Iain0007/Clearscan.wps

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

R49 Ms S Dubazana

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M211158

NAME: **(Principal Investigator)**	Ms S Dubazana
DEPARTMENT:	School of Clinical Medicine Department of Surgery Medical School University
PROJECT TITLE:	*Method optimization for verification of candidate pancreatic ductal adenocarcinoma reporter proteins in plasma*
DATE CONSIDERED:	2021/11/26
DECISION:	Approved unconditionally
CONDITIONS:	Sub-study under M180284
NOTE:	If contact information regarding student study participants is required, please contact the Registrar's office - <Nicoleen.Potgieter@wits.ac.za>
SUPERVISOR:	Dr E Nweke
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2022/03/28

This Clearance Certificate is valid for 5 years from the date of approval. An extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office secretariat on the 3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in **November** and therefore reports and re-certification will be due in the month of **November** each year. Unreported changes to the study may invalidate the clearance given by the HREC (Medical).

s.dubazana
Signature of Principal Investigator

29/03/2022
Date