

https://dspace.org/

Potential protein biomarker discovery for gallbladder cancer in a South African cohort

Baichan, Pavan

2024
https://hdl.handle.net/10539/37595

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

Potential protein biomarker discovery for gallbladder cancer in a South African cohort



WITS UNIVERSITY

Pavan Baichan

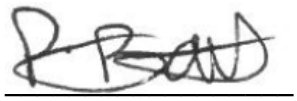
A thesis submitted in fulfilment of the requirement for the degree of Doctor of Philosophy

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

JUNE 2023

DECLARATION

I, Pavan Baichan, declare that this thesis entitled “Potential protein biomarker discovery for gallbladder cancer in a South African cohort” is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy at the University of Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other university.



Pavan Baichan

9 June 2023

Date

DEDICATION

To my parents and my love, thank you for standing by me through all the years of studying and hardship.

My Protectors.

We finally made it.

PUBLICATIONS AND PRESENTATIONS

Publications:

1. Baichan, P., Naicker, P., Augustine, T.N., Smith, M., Candy, G.P., Devar, J.W.S., and Nweke, E.E. 2023. Proteomic Analysis Identifies Dysregulated Proteins and Associated Molecular Pathways in A Cohort of Gallbladder Cancer Patients of African Ancestry. *Clinical Proteomics, 20(8),* pp.1-14.

Contribution: I, as the candidate and first author, acquired ethical clearance, performed sample collection, with the aid of surgical clinicians, and all described methodologies. Thereafter, I analysed all data and built all relevant figures and tables, with the aid of the authors listed, to compile the results section. Finally, I performed the writing and editing of the manuscript with the aid of all other authors for final publication.

2. Baichan, P., Naicker, P., Devar, J., Smith, M. and Nweke, E.E., 2023. A Look at Emerging Therapeutic Targets for Gallbladder Cancer: A Multi-Omics Approach. *In Gallbladder Cancer: Current Treatment Options* (pp. 161-175). Singapore: Springer Nature Singapore.
3. Baichan, P., Naicker, P., Devar, J.W.S., Smith, M., Candy, G.P. and Nweke, E., 2020. Targeting gallbladder cancer: a pathway based perspective. *Molecular Biology Reports*, *47*(3), pp.2361-2369.

Contribution: For publications two and three, I, as the candidate and first author, performed the writing and editing of the manuscripts with the aid of all listed authors.

Published Abstracts

International Conferences – Presentations:

1. Baichan, P., Naicker, P., Devar, J.W.S., Smith, M., Candy, G.P., and Nweke, E.E. 2022. Proteome Analysis Distinguishes Gallbladder Cancer from Gallstone Disease in Patients of African Ancestry. HPB. 24: S28.
2. Baichan, P., Naicker, P., Augustine, T.N., Smith, M., Candy, G.P., Devar, J.W.S., and Nweke, E.E. Proteomic Analysis Distinguishes Gallbladder Cancer from Gallstone Disease in Patients of African Ancestry. 15th Congress of the International Hepato-Pancreato Biliary Association World Congress, 30th March – 2nd April 2022, New York City.

International Conferences – Poster Presentations:

1. Baichan, P., Naicker, P., Devar, J.W.S., Smith, M., Candy, G.P., and Nweke, E.E. Proteome Analysis Distinguishes Gallbladder Cancer from Gallstone Disease in Patients of African Ancestry. 14th Congress of the European-African Hepato-Pancreato Biliary Association Virtual Congress, 15th – 17th September 2021, Spain.

Local Conferences – Oral Presentations:

1. Baichan, P., Naicker, P., Devar, J.W.S., Smith, M., Candy, G.P., and Nweke, E.E. Proteomic Analysis Identifies Dysregulated Proteins and Associated Molecular Pathways in A Cohort of Gallbladder Cancer Patients of African Ancestry. Bert Myburgh Research Forum, Department of Surgery, University of Witwatersrand, 30th November 2022.
2. Baichan, P., Naicker, P., Devar, J.W.S., Smith, M., Candy, G.P., and Nweke, E.E. Proteomic Analysis Identifies Dysregulated Proteins and Associated Molecular Pathways in A Cohort of Gallbladder Cancer Patients of African Ancestry. Faculty of Health Sciences Research Day, Wits Medical School, University of the Witwatersrand, 15th September 2022.
3. Baichan, P., Naicker, P., Devar, J.W.S., Smith, M., Candy, G.P., and Nweke, E.E. Proteome Analysis Distinguishes Gallbladder Cancer from Gallstone Disease in Patients of African Ancestry. Bert Myburgh Research Forum, Department of Surgery, University of Witwatersrand, 24th November 2021.
4. Baichan, P., Naicker, P., Devar, J.W.S., Smith, M., Candy, G.P., and Nweke, E.E. Proteome Analysis Distinguishes Gallbladder Cancer from Gallstone Disease in Patients of African Ancestry. School of Clinical Medicine (SOCM) Biennial Research Day. 30th September 2021.
5. Baichan, P., Naicker, P., Devar, J.W.S., Smith, M., Candy, G.P., and Nweke, E.E. Proteome Analysis Distinguishes Gallbladder Cancer from Gallstone Disease in Patients of African Ancestry. 48th Surgical Research Society of Southern Africa, Virtual Congress, 9-10th July 2021.

ACKNOWLEDGEMENTS

- To my supervisors, Dr Ekene Emmanuel Nweke and Dr Previn Naicker, for your continued support, mentorship, and guidance throughout this journey.
- Dr Nweke, your hands-on and direct involvement in my research was invaluable. You kept me honest and on my toes to make sure I performed to the best of my abilities and guided me in dealing with the complexities with working both independently and in a team, which I may not have been able to navigate on my own. Beyond that, you took time out to suggest external courses to help me hone my skills beyond my project so I may become a better-rounded researcher and scientist. During the hardest time of my tenure as a PhD candidate, you still took the time to go through my work with a fine tooth comb to make sure that I produced a manuscript we all could be proud of. I will be eternally thankful to you for helping me achieve a lifelong dream.
- Dr Naicker, you were always on hand to aid me with hands-on experimental guidance and expertise at the CSIR. You guided me meticulously through new techniques which I previously had not encountered. Through every repetitive question, you were patient and taught me like a true leader and mentor. Even after my time at the CSIR came to an end, you were always available for any information or piece of advice I needed through the writing of my manuscript and more so always ready to edit and give more input and guidance To you too I will be eternally grateful in helping me achieve this dream.
- To Prof. Geoffrey Candy, even though our time together was brief you provided me with essential guidance when I first joined your team. Even beyond that, you took the time to provide me with invaluable expertise and experience so that my manuscript may reflect the time, effort, and work that went into the entire research project.
- Dr John Devar for his dedication to my project in aiding me in collecting clinical samples. Thank you Dr Devar for always making sure my precious samples were always collected, even in my absence, even if you had to perform the procedures yourself.
- My friends and colleagues in the Department of Surgery, I appreciate all you have done during this journey, whether academic or personal.
- To my loving parents and girlfriend, your endless love and support brought me to this point after long hard years. This is your achievement as much as it is mine.

TABLE OF CONTENTS

SUMMARY **1**

1 CHAPTER 1: INTRODUCTION **2**

1.1 The gallbladder **2**

1.2 Diseases of the gallbladder **3**

1.2.1 Porcelain gallbladder 3

1.2.2 Gallstone disease 4

1.2.3 Cholecystitis 5

1.3 Molecular mechanisms associated with inflammation of the gallbladder and linked with gallbladder cancer **6**

1.4 Gallbladder cancer and epidemiology **9**

1.4.1 Epidemiology 9

1.4.2 Risk factors 9

1.4.3 Types of gallbladder cancer 11

1.5 Clinical presentation and diagnosis **11**

1.6 Gallbladder cancer staging and differentiation **13**

1.7 Molecular mechanisms of gallbladder cancer progression **14**

1.7.1 P53 pathway 14

1.7.2 ERBB2/3 and the PI3K/Akt/mTOR signalling pathway 15

1.7.3 Cyclooxygenase-2 15

1.7.4 MAPK/ERK signalling pathway 15

1.7.5 KRAS 15

1.7.6 PI3K/Akt/mTOR signalling pathway 16

1.7.7 Hedgehog signalling pathway 17

1.7.8 EGFR signalling pathway 18

1.8 Protein biomarkers in gallbladder cancer **20**

1.9 Biomarker discovery using proteomic tools **21**

1.9.1 Two-dimensional differential gel electrophoresis (2D-DIGE) 21

1.9.2 Protein arrays 21

1.9.3 Liquid chromatography-mass spectrometry (LC-MS) 23

1.9.4 Bottom-up proteomics 23

1.9.5 Sequential window of all theoretical mass spectra (SWATH-MS) 24

1.10 Rationale **25**

1.11 Aims and Objectives **26**

1.11.1 Aim 26

1.11.2 Objectives 26

2 CHAPTER 2: MATERIALS AND METHODS **27**

2.1 Summary of experimental strategy **27**

2.2 Ethics statement **29**

2.3 Sample and clinical data collection **29**

Tissue patient sample population 29

Plasma patient sample population 29

2.4 Tissue homogenisation and protein extraction 30
2.5 2-D Protein quantification assay 30
2.6 Protein aggregation capture 32
2.7 Hydrophilic interaction liquid chromatography 33
2.8 Pierce quantitative colourimetric peptide assay 34
2.9 High pH reversed-phase fractionation 35
2.10 Liquid chromatography-mass spectrometry (LC-MS) data acquisition 37
2.11 Liquid chromatography-mass spectrometry (LC-MS) data processing 38
2.12 Retrospective power analysis 39
2.13 Pathway and network analyses 39
2.14 Statistical Analysis 39
3 CHAPTER 3: RESULTS 41
3.1 Demographics, clinical inflammatory markers and liver function tests are significantly altered in GBC patients 41
3.2 Custom-built project-specific SWATH spectral library 45
3.3 Quality control of liquid chromatography-mass spectrometry setup 45
3.4 Sequential window of all theoretical mass spectra analysis identified differentially abundant proteins in gallbladder cancer patients compared to normal, gallstone disease, and benign biliary pathology controls 49
3.4.1 Relatively abundant proteins detected in the tissue sample population 49
3.4.2 Pathway enrichment by the dysregulated proteins in the tissue sample population 51
3.4.3 Altered proteins identified in the plasma sample sample population 56
3.4.4 Pathway enrichment by the dysregulated proteins in the plasma sample population 57
3.5 Commonly dysregulated proteins in tissue and plasma sample sample populations 59
3.6 Quality control of the commonly dysregulated proteins 61
3.7 Hierarchical clustering, principal component analysis, and Spearman's rank correlation test of the commonly dysregulated proteins 64
3.8 Summary of results 66
4 CHAPTER 4: DISCUSSION 67
4.1 Elevated clinical parameters in gallbladder cancer patients 67
4.2 Pathways dysregulated in the tissue sample population 69
4.2.1 Enriched signalling pathways by the dysregulated proteins in the GBC/Normal comparison 69
4.2.2 Pathways enriched by proteins dysregulated in gallbladder cancer patients compared to gallstone disease patients 73
4.2.3 Enriched signalling pathways by the dysregulated proteins in the GD/Normal group 74
4.2.4 Proteins similarly dysregulated in gallbladder cancer compared to gallstone disease and normal gallbladders 75
4.3 Enriched pathways by the dysregulated proteins in the plasma sample population 76

4.4 Commonly dysregulated proteins may be linked to gallbladder cancer in the patient sample populations 78
4.4.1 Functions and pathways enriched by the identified commonly dysregulated proteins 78
4.4.2 Commonly dysregulated proteins significantly correlate with each other and can distinguish gallbladder cancer from control groups 81
4.5 Proteins associated with a history of gallstones and tumour staging 82
4.6 Dysregulated proteins observed in a sample population of South African gallbladder cancer patients compares to other populations 83
5 CHAPTER 5: CONCLUSION AND FUTURE DIRECTION 85
6 CHAPTER 6: STUDY LIMITATIONS 87
REFERENCES 88
SUPPLEMENTARY DATA 108
APPENDIX A: TURNITIN REPORT 144
APPENDIX B: ETHICS CLEARANCE CERTIFICATE 146
APPENDIX C: STUDY INFORMATION 147
APPENDIX D: INFORMED CONSENT (GALLBLADDER CANCER PATIENTS) 149
APPENDIX E: INFORMED CONSENT (GALLSTONE DISEASE PATIENTS) 150
APPENDIX F: INFORMED CONSENT (NORMAL GALLBLADDER PATIENTS) 151
APPENDIX G: PERMISSION TO CONDUCT RESEARCH 152

LIST OF FIGURES

Figure 1.1: The location of the gallbladder and depiction of gallstones....3
Figure 1.2: The MAPK/ERK and PI3K/Akt/mTOR signalling pathways....17
Figure 1.3: Hedgehog and epidermal growth factor receptor (EGFR) signalling pathways ...19
Figure 2.1: Flow chart of the summary of methods used 28
Figure 2.2: Elution profile for high pH reverse phase (RP) fractionation and fraction concatenation 36
Figure 3.1: Tissue HeLa cell lysate quality control 47
Figure 3.2: HeLa Cell lysate quality control for the plasma dataset....48
Figure 3.3: Venn diagram for common proteins dysregulated across tissue sample group comparisons 51
Figure 3.4: Network and pathway analysis of dysregulated proteins in GBC tissues compared to normal tissues 52
Figure 3.5: Network and pathway analysis of dysregulated proteins in GBC compared to GD tissue samples....53
Figure 3.6: Network and pathway analysis of dysregulated proteins in GD compared to normal tissue samples....54
Figure 3.7: Pathway and network analyses for the commonly dysregulated proteins between GBC/Normal and GBC/GD tissue groups 55
Figure 3.8: Annotated molecular functions for dysregulated proteins in tissue groups 56
Figure 3.9: Pathway and network analysis of the dysregulated proteins in the plasma group 58
Figure 3.10: Annotated molecular functions for plasma dysregulated proteins 59
Figure 3.11: Venn diagram for common dysregulated proteins among sample group comparisons 60
Figure 3.12: Network and pathway analysis for the identified Commonly Dysregulated Proteins between the tissue and plasma groups 61
Figure 3.13: Representative MS2 _DLATVYVDVLK_.2 precursor chromatogram overview for APOA1 in the tissue and plasma sample populations....63
Figure 3.14: Principal component analysis and Spearman's rank correlation test for commonly dysregulated proteins (CDPs). 65
Figure 4.1: Summary of the roles of the ECM and lipid metabolism pathways....72
Figure S3.1: Log2 quantities for the CDPs identified in GBC/GD and GBC/BBP comparisons per patient……...108
Figure S3.2: The Log2 quantities for the CDPs identified in GBC/Normal, GBC/GD, and GBC/BBP Comparisons……109
Figure S3.3: Representative MS2 _VKSPELQAEAK_.3 precursor chromatogram overview for APOA2 in tissue and plasma cohorts……..110
Figure S3.4: Representative MS2 _YWGVASFLQK_.2 precursor chromatogram overview for RET4 in tissue and plasma cohorts……..111
Figure S3.5: Representative MS2 _AADDTWEPFASGK_.2 precursor chromatogram overview for TTR in tissue and plasma cohorts……..112
Figure S3.6: Representative MS2 _NFPSPVDAAFR_.2 precursor chromatogram overview for HEMO in tissue and plasma cohorts…….……..…113

Figure S3.7: Representative MS2 _VLGAFSDGLAHLDNLK_.3 precursor chromatogram overview for HBB in tissue and plasma cohorts……..114
Figure S3.8: Representative MS2 _TYFPHFDLSHGSAQVK_.3 precursor chromatogram overview for HBA in tissue and plasma cohorts……..115
Figure S3.9: Representative MS2 _AFVNC[Carbamidomethyl (C)]DENSR_.2 precursor chromatogram overview for PIGR in tissue and plasma cohorts…….116
Figure S3.10: Representative MS2 _ALMDETMK_.2 precursor chromatogram overview for APOE in tissue and plasma cohorts…….117
Figure S3.11: Hierarchical cluster analysis for the differentially expressed proteins……..118
Figure S3.12: Spearman's Rho values and p-values for correlation of CDPs……..119

LIST OF TABLES

Table 1.1: Gallbladder cancer staging as per the AJCC 13
Table 2.1: Preparation of the standard curve for the 2-D protein quantification assay 31
Table 2.2: 96 deep well plate setup for protein aggregation capture 33
Table 2.3: Standard preparation for Pierce™ quantitative colourimetric peptide assay 35
Table 3.1: Clinical and demographic characteristics of gallbladder cancer and gallstone disease tissue sample population patients.....42
Table 3.2: Clinical and demographic characteristics of gallbladder cancer and benign biliary pathology plasma sample population patients 44
Table 4.1: Commonly Dysregulated Proteins (CDPs) and their functions.....79
Table S3.1: TNM staging and gallstone disease history for GBC plasma patients……..120
Table S3.2: The total precursors, peptides, and protein groups identified for each individual patient for the GBC versus normal comparison……...121
Table S3.3: All dysregulated proteins in GBC compared to normal……122
Table S3.4: The total precursors, peptides, and protein groups identified for each individual patient for the GBC versus GD comparison……125
Table S3.5: All differentially abundant proteins in GBC compared to GD……126
Table S3.6: The total precursors, peptides, and protein groups identified for each individual patient for the GD versus normal comparison……131
Table S3.7: All differentially expressed proteins in GD compared to normal……132
Table S3.8: Proteins commonly dysregulated in GBC compared to GD and normal with relative average Log2 fold changes……135
Table S3.9: The total precursors, peptides, and protein groups identified for each individual patient for the GBC versus BBP plasma comparison……..137
Table S3.10: All differentially abundant proteins in GBC compared to BBP plasma……..139
Table S3.11: Proteins associated with non-metastatic vs metastatic staging for GBC plasma patients ……...140
Table S3.12: Commonly dysregulated proteins across group comparisons with relative average Log_2 fold changes……141
Table S3.13: Protein coverage and number of samples identified in for the CDPs……..143

LIST OF ABBREVIATIONS

2D-DIGE	Two-dimensional differential gel electrophoresis
60SPD	60 samples per day
A1AT	Alpha-1-antitrypsin
AACT	Alpha-1-antichymotrypsin
ABCG5	Adenosine triphosphate-binding cassette sub-family G member 5
ABCG8	Adenosine triphosphate -binding cassette sub-family G member 8
ACN	Acetonitrile
ACTA	Actin alpha 2
AJCC	American Joint Committee on Cancer
AK1C1	Aldo-keto reductase family 1 member C1
AK1C2	Aldo-keto reductase family 1 member C2
AK1C3	Aldo-keto reductase family 1 member C3
AKT	Protein kinase B
ALBU	Albumin
ALDOB	Fructose-bisphosphate aldolase B
ALP	Alkaline phosphatase
ALT	Alanine transaminase
APDJ	Anomalous pancreaticobiliary ductal junction
APOA1	Apolipoprotein A-I
APOA2	Apolipoprotein A-II
APOE	Apolipoprotein E
ARK73	Aflatoxin B1 aldehyde reductase member 3
AST	Aspartate transaminase
BBP	Benign biliary pathology
BCA assay	Bicinchoninic acid assay
BSA	Bovine serum albumin
CA 125	Carbohydrate antigen 125
CA 19-9	Carbohydrate antigen 19-9
CAH1/2	Carbonic anhydrase 1/2
CATD	Cathepsin D

CAV1	Caveolin-1
CCK	Cholecystokinin
CCKAR	Cholecystokinin A receptor
CCKR	Cholecystokinin receptor
CDPs	Commonly dysregulated proteins
CEA	Carcinoembryonic antigen
CH60	60 kDa heat shock protein
CHBAH	Chris Hani Baragwanath Academic Hospital
CO1A1	Collagen alpha-1(I) chain
COCA1	Collagen alpha-1(XII) chain
COX2	Cyclooxygenase-2
CRP	C-reactive protein
CT	Computed tomography
CV	Coefficient of variation
Cy3/5	Cyanin 3/5
CYFRA 21-1	Soluble fragment of cytokeratin 19
DAMP	Damage-associated molecular pattern
DDA	Data-dependent acquisition
DESM	Desmin
DIA	Data-independent acquisition
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E2	17β-oestradiol
ECHP	Peroxisomal bifunctional enzyme
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
ELISA	Enzyme-linked immunosorbent assay
EMT	Epithelial-to-mesenchymal transition
ERBB	Receptor tyrosine-protein kinase
ERCP	Endoscopic retrograde cholangiopancreatography
ERK	Extracellular signal-regulated kinase

ESR	Oestrogen receptor
EST1	Liver carboxylesterase 1
EUS	Endoscopic ultrasonography
EUS-FNA	Endoscopic ultrasound-guided fine needle aspiration
FABP5	Fatty acid-binding protein 5
FABPL	Fatty acid binding protein, liver
FC	Fold change
FDA	Food and Drug Administration
FDR	False-discovery rate
FETUA	Alpha-2-HS-glycoprotein
FINC	Fibronectin
g	Relative centrifugal force
G6PD	Glucose-6-phosphate 1-dehydrogenase
GBC	Gallbladder cancer
GD	Gallstone disease
GGT	Gamma-glutamyl transferase
Gli	Glioma-associated oncogene
GSTM3	Glutathione S-transferase Mu 3
GTPase	Guanosine triphosphatase
H2A1B	Histone H2A type 1-B/E
H2A1C	Histone H2A type 1-C
H2A3	Histone H2A type 3
Hb	Haemoglobin
HBA	Haemoglobin subunit alpha
HBB	Haemoglobin subunit beta
HDL	High-density lipoprotein
HEMO	Hemopexin
Hh	Hedgehog
HILIC	Hydrophilic interaction chromatography
HPB	Hepatopancreaticobiliary
IAA	Iodoacetamide
IEF	Isoelectric focusing
IFN	Interferon
IL	Interleukin

IQR	Interquartile range
ITAM	Integrin alpha-M
ITIH3	Inter-alpha-trypsin inhibitory heavy chain H3
iTRAQ	Isobaric tags for relative and absolute quantification
JAK2	Tyrosine-protein kinase JAK2
JIP4	C-jun-amino-terminal kinase-interacting protein 4
KRAS	Kirsten rat sarcoma viral oncogene homolog
LAMA5	Laminin subunit alpha-5
LAMB2	Laminin subunit beta-2
LAMC1	Laminin subunit gamma-1
LC-MS	Liquid chromatography-mass spectrometry
LDL	Low-density lipoprotein
LFT	Liver function test
Lith 1/2	Lithogenic 1/2
m/z	Mass-to-charge
MALAT1	Metastasis associated lung adenocarcinoma transcript 1
MAPK	Mitogen-activated protein kinase
MEK	Mitogen-activated protein kinase kinase
MMP	Matrix metalloproteinase
MRCP	Magnetic resonance cholangiopancreatography
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MS	Mass spectrometry
mTOR	Mammalian target of rapamycin
ML12A/B	Myosin regulatory light chain 12A/B
MYH11	Myosin-11
MYL9	Myosin regulatory light polypeptide 9
NFκB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NSCLC	Non-small cell lung carcinoma
p-mTOR	Phosphorylated-mammalian target of rapamycin
P53	Tumour protein p53
PAC	Protein aggregation capture
PAMP	Pattern-associated molecular pattern

PC	Principal component
PCA	Principal component analysis
PDAC	Pancreatic ductal adenocarcinoma
PDGF	Platelet-derived growth factor
PGB	Porcelain gallbladder
PI3K	Phosphoinositide 3-kinase
PIGR	Polymeric immunoglobulin receptor
PLF4	Platelet factor 4
POSTN	Periostin
PRR	Pattern recognition receptors
Ptch1	Protein patched homolog 1
QC	Quality control
R-RAS	Ras-related protein R-Ras
RAC2	Ras-related C3 botulinum toxin substrate 2
Ras	Rat sarcoma
RET4	Retinol-binding protein 4
RNA	Ribonucleic acid
RPLC	Reverse-phase liquid chromatography
RRPA	Reverse-phase protein array
RSLC	Rapid separation liquid chromatography
SDS	Sodium dodecyl sulphate
Shh	Sonic Hedgehog
SLC25A22	Solute carrier family 25 member 22
Smo	Smoothened
SPOCK1	sparc/osteonectin, cwcv and kazal-like domains proteoglycan 1
STAT	Signal transducer and activator of transcription
STRA6	Retinoic acid 6
SWATH-MS	Sequential window of all theatrical mass spectra
TENA	Tenascin
TETN	Tetranectin
TFA	Trifluoroacetic acid
TGF	Transforming growth factor
THIL	Acetyl-CoA acetyltransferase
THIM	3-Ketoacyl-CoA thiolase

TIC	Total ion chromatogram
TNF	Tumour necrosis factor
TNM	Tumour, Nodes, and Metastasis
TOF	Time-of-flight
TPM1	Tropomyosin alpha-1 chain
TPM2	Tropomyosin beta chain
Tris-HCl	Trisaminomethane-hydrochloric acid
TSP1	Thrombospondin-1
TTR	Transthyretin
TYMP	Thymidine phosphorylase
USA	United States of America
VTNC	Vitronectin
WCC	White cell count

SUMMARY

Gallbladder cancer (GBC) has a poor prognosis with the prevalence of GBC varying according to geographical location. The prevalence of GBC in South Africa is poorly tracked, and the molecular mechanisms associated with GBC in African patients are inadequately understood. This study aimed to determine dysregulated proteins in tissue and blood plasma in South African GBC patients to identify potential molecular mechanisms of disease progression and plausible biomarkers. Following ethical approval, tissue from 27 GBC, 13 gallstone disease (GD), and five normal tissues were obtained. Blood plasma was collected from 54 GBC and 73 benign biliary pathology (BBP) patients who consented to the study. A bottom-up proteomics approach was undertaken, using PAC and HILIC digestion methods, and SWATH-MS for quantitative proteomic profiling. Hierarchical cluster analysis, PCA analysis, and Spearman's rank correlation analysis were performed. Furthermore, pathway and network analyses were conducted. There were 62, 194, and 105 dysregulated proteins in the GBC/Normal, GBC/GD, and GD/Normal group comparisons, respectively, and 33 dysregulated proteins in the GBC/BBP plasma comparison. The dysregulated proteins in GBC patients enriched pathways involved in smooth muscle contraction, metabolism, extracellular matrix organisation and interactions, innate immunity, and platelet and neutrophil degranulation. Further analysis showed that S100A8 and S100A9 were downregulated in GBC plasma patients with GD history compared to those with no GD history. Additionally, APOE and ITIH3 were elevated in non-metastatic staging GBC patients. Seven proteins were found to be commonly dysregulated in GBC/GD and GBC/BBP comparisons and another two proteins were commonly dysregulated in the GBC/Normal, GBC/GD, and GBC/BBP comparisons, termed "Commonly dysregulated proteins (CDPs)". Quality control assessment of the MS2 fragment ion chromatograms of the CDPs indicated strong signal-to-noise ratios and correct fragment-to-precursor matching. The CDPs could distinguish between GBC and controls and the Spearman's rank correlation test showed significant correlations between the expression of the CDPs. The identified dysregulated proteins aid in further understanding the molecular mechanisms associated with GBC in patients with African ancestry. The alteration of specific proteins in tissue and plasma samples suggests their potential use as biomarkers for GBC patients in this sample cohort.

1 CHAPTER 1: INTRODUCTION

1.1 The gallbladder

The gallbladder is a small, pear-shaped organ underneath the liver and is an accessory organ of the digestive tract. The gallbladder stores and concentrates bile produced and released by the liver (Figure 1.1A). The primary function of bile is to aid in the digestion of fats (1,2). It is attached to the liver via loose connective tissue, small veins, and the lymphatic system (3). The gallbladder is connected to the cystic duct that drains into the common bile duct. The cystic duct is 3 – 4 cm long and links to the common hepatic duct to form the common bile duct (3). The gallbladder comprises three main structures: the fundus, the body, and the neck/infundibulum (Figure 1.1B) (2). The fundus structure is the rounded portion of the gallbladder which faces the abdominal wall. The body of the gallbladder lies in a shallow depression in the right lobe of the liver and contains the bile. The neck/infundibulum is the tapering portion forming part of the cystic duct leading into the common hepatic duct and finally the common bile duct (4). The human gallbladder ranges from 7 – 10 cm long and 3 – 4 cm wide with a wall thickness of roughly 2 mm (varying with distension), and can contain approximately 50 mL of bile (3). Various gallbladder conditions are considered diseased states and may require clinical or healthcare intervention.

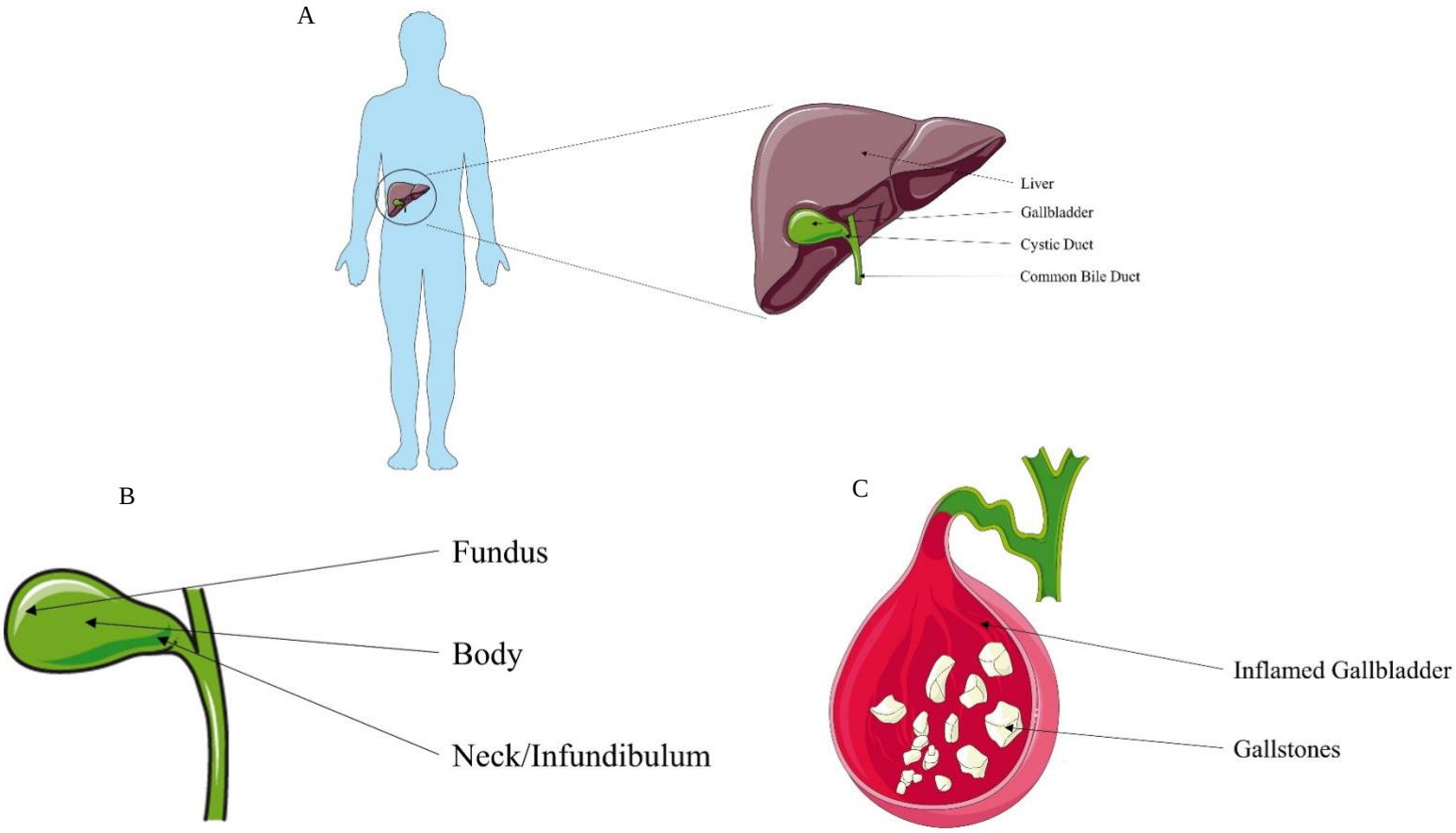


Figure 1.1: The location of the gallbladder and depiction of gallstones. (A) The gallbladder is located underneath the liver in the hepatobiliary duct system. It is a small, hollow organ where bile is stored before being secreted into the small intestine. (B) The anatomy of the gallbladder consists of three main structures: the fundus (the rounded base), the body (where bile collects), and the neck/infundibulum (the opening of the gallbladder). (C) Gallstones comprise various salts and bilirubin, all constituents of bile that collect in the fundus. (Figure 1.1 modified with text and annotation after adaptation of "Digestive System" from Servier Medical Art by Servier, licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/).

1.2 Diseases of the gallbladder

There are various diseases associated with the gallbladder, namely porcelain gallbladder, gallstone disease (cholelithiasis), inflammation (cholecystitis), and gallbladder cancer (5).

1.2.1 Porcelain gallbladder

Porcelain gallbladder (PGB) is a disease where the gallbladder develops a blue discolouration and becomes fragile due to excessive calcium deposits (6). While the specific pathogenesis of PGB is controversial, it is hypothesised that it is associated with chronic inflammation and gallstones. The clinical presentation of many patients suffering from PGB is asymptomatic and is usually detected through incidental radiographic techniques. PGB is detected mainly with ultrasonography and abdominal radiography, with confirmation via surgical procedures (such as a laparotomy) or pathology tests (7). PGB is exceedingly rare and is detected in <1% of patients with gallbladder diseases. However, this disease is most prevalent in females and individuals over 60 years. Gallstones are involved in 60 – 90% of PGB cases (7).

Porcelain gallbladder is also considered a precursor disease for gallbladder cancer (GBC), and it has been observed that 12 – 33% of patients with PGB develop GBC (7). Some studies have reported an association of 10%, and others indicated an association of 2 – 3%. Patients with mucosal calcification, or partial mural calcification, are believed to be at increased risk (8,9).

1.2.2 Gallstone disease

Gallstone disease (GD), or cholelithiasis, is characterised by the formation of crystallised gallstones formed from cholesterol or bilirubin. Gallstones can form in the gallbladder or, less frequently, in the bile duct (Figure 1.1C) (10). Depending on the location of the gallstones, various secondary diseases may arise, including cholecystitis, obstructive jaundice, and pancreatitis (11). Gallstones can be categorised as either cholesterol or pigmented gallstones. More than 80% of gallstones are estimated to comprise cholesterol, indicating a defect in the biliary tree. Pigmented gallstones can be further subdivided into brown and black stones. Brown stones are typically associated with bacterial infection and are located in the biliary tree rather than inside the gallbladder. Black stones indicate haemolytic conditions or cirrhosis and are located inside the gallbladder (12). Some risk factors for GD include being female, advanced age, poor nutrition, and genetic risk factors (13).

Gallbladder hypomotility contributes to the formation of gallstones, whereby excessive cholesterol and bile accumulate within the smooth muscle cells of the gallbladder wall which leads to stiffening. This prolongs the duration of bile in the gallbladder which induces crystallisation and gallstone formation (10). Most patients suffering from GD do not present with noticeable physical symptoms. In the case of symptomatic GD patients, complete surgical removal of the gallbladder (cholecystectomy) is the first-line treatment (14). The rates of cholecystectomies performed vary by region; high-incidence regions can perform up to 220 per 100,000 whereas some lower-incidence regions perform 65 per 100,000. Moreover, Latin American countries have the highest GD prevalence (15).

Various imaging techniques are employed to diagnose GD due to the non-specific symptoms of this condition. Ultrasonography confirms GD with a success rate of >90%; however, bile duct stones may be missed in >50% of cases. Other techniques to diagnose GD are magnetic resonance cholangiopancreatography (MRCP), endoscopic ultrasound (EUS), and endoscopic retrograde cholangiopancreatography (ERCP) (16). Gallstone disease is considered to be a significant risk factor for gallbladder cancer onset and progression. Up to 85% of patients diagnosed with GBC have a history of gallstones. Interestingly, only 0.5 – 3% of patients

suffering from GD develop GBC (9,17). Although little is known about the mechanisms associated with GD progression to GBC, it is thought that chronic epithelial irritation and mucosal damage by gallstones are involved in the process. Additionally, patients with gallstones >3 cm in size have an increased risk of up to 10-fold (9) of developing GBC compared to stones <1 cm. Symptomatic GD is also linked to GBC occurrence compared to asymptomatic disease, and cholesterol gallstones are linked to more symptomatic cases of GD (18).

The prevalence of GD varies from males to females and in geographical regions. The North American Indian incidence for GD is 64.1% in females and 29.5% in males, Mapuche Chileans have an incidence of 49.4% in females and 12.6% in males, white Americans have a prevalence of 16.6% in females and 8.6% in males, and Africans have a low prevalence of <5% for both sexes (19). There is a significant lack of current information on the prevalence of GD in South Africans. Studies completed in the 1980s indicated that GD prevalence in black South Africans was “low”; however, this information is outdated and needs reassessment (20). In this study, patients diagnosed with GD were used as a control population because this disease is a risk factor for GBC.

1.2.3 Cholecystitis

Cholecystitis is inflammation of the gallbladder and can be categorised as acute or chronic. Acute cholecystitis is generally induced by blockage of the cystic/bile ducts by gallstones, and >90% of acute cholecystitis patients are associated with GD (21). Other causes of cholecystitis are a mass in the gallbladder (primary tumour or polyps) and parasites. Causes independent of gallstones are known as acalculous cholecystitis (12). In chronic cholecystitis (repeated bouts of acute cholecystitis), nearly 100% of cases described gallstone involvement (22). The symptoms of acute cholecystitis include upper right quadrant pain, nausea, vomiting, and anorexia. Various imaging techniques, such as ultrasonography and magnetic resonance imaging (MRI), are used to diagnose cholecystitis. Similar to GD, the gold standard treatment for cholecystitis is cholecystectomy (12). PGB, GD, and cholecystitis are significant precursors for diseased gallbladders developing into malignant gallbladder cancer. Approximately 85% of GD and cholecystitis cases and roughly 25% of PGB cases develop into gallbladder cancer. The mechanism of cholecystitis-inducing transformation is described in subsequent sections.

1.3 Molecular mechanisms associated with inflammation of the gallbladder and linked with gallbladder cancer

As previously indicated, inflammation of the gallbladder, in the presence and absence of gallstones, is considered a precursor for gallbladder cancer. Dysfunction of the gallbladder motility and bile-secretion pathways may result in gallstone formation. Aberrations in genes and associated pathways have been implicated in the pathogenesis of gallstones (23,24).

Murine model studies have been used to understand the genetic effects on gallstone formation better. Gallstone-susceptible (C57L/J) and gallstone-resistant (AKR/J) mice were used to demonstrate that lithogenic genes 1 and 2 (*Lith1* and *Lith2*) may play a role in the formation of gallstones. *Lith1* plays a role in determining liver cholesterol hypersecretion, and *Lith2* regulates the salt-dependent flow of bile (25). The human functional counterparts of *Lith1* and *Lith2* are *ABCG5* (Adenosine triphosphate-binding cassette sub-family G member 5) and *ABCG8* (Adenosine triphosphate-binding cassette sub-family G member 8), respectively. In hepatocytes, these proteins transport neutral sterols into bile or promote the active efflux of cholesterol from the enterocyte into the intestinal lumen for faecal excretion (26). The inactivation of ABCG5/8 proteins causes a reduction in cholesterol secretion into bile. Conversely, overexpression of these proteins results in increased cholesterol content in bile and the gallbladder and increased propensity of cholesterol crystal precipitation (24,27). ABCG5/8 were associated with cholesterol gallstone formation. Two gene variants of these proteins (*ABCG5-R50C* and *ABCG8-D19H*) were identified in several population groups, including Germans, Chileans, Chinese, and Indians (26).

The oversaturation of the bile may also induce cholelithogenesis, which is the process of gallstone formation. Under normal physiological conditions, varied gallbladder absorption rates for biliary cholesterol, bile salts, and phospholipids affect the concentration and saturation of cholesterol in bile. Therefore, abnormal lipid absorption may enhance cholelithogenesis due to the persistent supersaturation of cholesterol in bile during the process of bile concentration for storage (28). Another approach to cholelithogenesis is gallbladder motility dysfunction. Cholecystokinin A receptor (CCKAR) is a receptor protein that regulates the motility of the smooth muscle of the gallbladder. It was observed in a murine model that deletion of the *CCKAR* gene leads to dysfunction of the smooth muscle contractility of the gallbladder. It is still unclear whether variants in the *CCKAR* gene are associated with increased gallstone prevalence in human patients (29).

Studies have shown that females develop gallstones more frequently than males (19,30). Oestrogen, such as 17β-oestradiol (E2), is a female hormone with liposoluble properties allowing it to diffuse passively into cells and act as a transcription factor. When E2 diffuses into the liver, it induces cholesterol secretion into bile and increases cholesterol saturation, elevating the risk of cholelithiasis (31). Moreover, E2 may bind oestrogen receptors ESR1 (oestrogen receptor 1) and ESR2 (oestrogen receptor 2), which become active complexes that regulate transcription. Furthermore, oestrogen-activated ESR1 also stimulates the activity of ABCG5/8 (24,32), further driving the cholelithogenesis process.

Gallstones, as previously mentioned, are a risk factor for gallbladder inflammation which aids in the tumourigenic process. Murine model studies have also indicated that cholesterol crystals, which initiate gallstone formation, induce gallbladder changes such as a thickened mucus layer, and interleukin-1 (IL-1) activity in the gallbladder wall are hallmarks of gallbladder inflammation. Additionally, gallstones cause repeated trauma to the gallbladder mucosa inducing chronic inflammation (33,34). Gallbladder cancer is known to be influenced by inflammation (35). Inflammation is the activated response due to tissue injury and is beneficial as it aims to restore homeostasis. However, unregulated inflammation may become chronic and induce transformation from normal to malignant cells. Chronic inflammation is important in initiating, sustaining, and advancing tumour growth (36). Acute inflammatory processes are limited in response time; however, many cases evolve into chronic inflammation associated with tumourigenesis. Two pathways, NFκB (nuclear factor kappa B) and STAT3 (signal transducer and activator of transcription 3), have been identified as crucial regulators of pro-inflammatory cytokines vital for chronic inflammation persistence and tumour proliferation (37).

The inflammasome is a protein complex comprised of danger-sensing proteins which promote the inflammatory response. The NFκB signalling pathway is involved in various cellular processes such as apoptosis, proliferation, response to infection, and inflammation. The NFκB pathway activates and forms inflammasomes which then promote the production of more inflammatory proteins (38). During inflammation, large amounts of free radicals are released, which cause DNA damage by introducing mutations into the DNA sequence. The cells undertake DNA damage repair, or apoptosis, to prevent cancer cell formation (39). NFκB signalling is persistent during chronic inflammation inhibiting DNA damage repair and apoptotic mechanisms. This causes many abnormal cells to aggregate and become cancer cells

(38). Cancer cells can essentially hijack the normal regulation of the NFκB pathway. Under normal physiological conditions, the NFκB pathway is expressed in certain cell types, such as neurons, and inactive in other cell types but is consistently active in many types of cancers (40).

The STAT signal transduction pathway functions to regulate various physiological processes such as the growth, survival, and differentiation of cells. Specifically, STAT3 induces a carcinogenic inflammatory environment, promotes oncogenesis and is typically activated after IL-6 activation (41). The STAT3-mediated oncogenesis is regulated by the oncogene *Src* which has been directly linked to human cancers (42). Additionally, chronic inflammation has been linked to STAT3-mediated malignant properties. Moreover, STAT3 regulates various activators of the carcinogenic process, such as chemokines and cytokines. This regulation induces an autocrine feedback loop to activate STAT3 further (41). The NFκB and STAT3 pathways are important in transformation processes such as growth, survival, and apoptosis. However, these pathways are not directly mutated; the upstream chemokines and cytokines become dysregulated, inducing aberrant activation of these pathways (41).

Cytokines can be divided into two main categories: pro-inflammatory and anti-inflammatory. Some proinflammatory cytokines include IL-1, IL-6, IL-8, interferon-gamma (IFNγ), and tumour necrosis factor-alpha (TNFα). Anti-inflammatory cytokines include IL-4, IL-10, transforming growth factor (TGFβ), IFNα, and IFNβ (43). TNFα and IL-1β are vital in initiating chronic inflammation and activating the NFκB signalling pathway. IL-1β overexpression has been shown to induce gastric cancer formation. TGFβ plays a role in tumour progression by inducing cancer-associated inflammation and microenvironment remodelling (43,44). IL-6 independently has been observed to promote chronic inflammation and carcinogenesis of the stomach and colon via the activation of the STAT3 pathway (45). Inhibitors of these cytokines and the downstream signal transduction pathways (such as the STAT3 and NFκB pathways) are promising targets for inflammatory and, subsequently, cancer therapy (43).

The following sections discuss several aspects of gallbladder cancer, such as epidemiology, risk factors, histology, clinical presentation, and molecular and signalling mechanisms.

1.4 Gallbladder cancer and epidemiology

1.4.1 Epidemiology

Gallbladder cancer (GBC) is the most prevalent cancer of the biliary tract, accounting for 80 – 95% of cases (46,47). About 80% of patients are diagnosed at an advanced or metastasized stage; hence GBC has a 5-year survival rate of ~19% (15). The Globocan 2020 statistics database estimated that GBC had approximately 115 949 newly diagnosed cases (41 062 males and 74 887 females) annually and accounted for 84 695 of all cancer deaths (30 265 males and 54 430 females) globally (47). In the USA, by 2023, it is estimated there will be 12 220 new cases (5 750 males and 6 470 females) and 4 510 deaths (1 900 males and 2 610 females) (48). The global incidence trends of GBC vary significantly according to geographical location. Chile has reported the highest rates of GBC (27 cases per 100 000), followed by Northern India regions (21.5/100 000). Other high-risk areas include Poland (14/100 000), South Pakistan (11.3/100 000), Japan (7/100 000), and Israel (5/100 000) (17). Latin America and South Asia have the highest incidence of GBC worldwide (15,49). Conversely, European ancestry regions, including the United States, Australia, Canada, the United Kingdom, and New Zealand, show a lower incidence of GBC. Notably, the United States has a diverse population with varying ancestries, each with its own incidence; for example, Native Americans and Hispanic Americans show the highest incidence of GBC compared to Caucasians (17).

The GBC prevalence in sub-Saharan Africa is poorly tracked. In South Africa, all GBC cases documented by the National Cancer Registry are based solely on histologically confirmed GBC cases. Clinically and radiologically suspected GBC patients, either independently or in conjunction, are excluded by the National Cancer Registry if the cases are not histologically confirmed (50). In 2018, there were 574 histologically confirmed new cases and 287 deaths (51). However, a recent study evaluating records from 2003 to 2015 has suggested a higher incidence of GBC in South Africa than reported (50).

1.4.2 Risk factors

Several demographic factors are associated with an increased risk of GBC. GBC is age-dependent, with the highest rates occurring during the seventh decade of life. Moreover, females are much more predisposed to developing GBC, with a two to six-fold more likely chance of developing GBC than males (9). While the exact mechanisms driving females to be more predisposed to GBC are not certain, it is speculated that female hormones increase the risk by inducing gallstone formation (30). Additionally, the use of hormone replacement

therapy and oral contraceptives also increases the risk for GBC. Other risk factors unique to females include a young age for the first menstrual cycle, an advanced age for the first pregnancy, and late menopause (52,53). Various other conditions predispose patients to GBC onset and progression, such as exposure to carcinogens and toxins.

Carcinogens and environmental exposure to toxins have also been linked to GBC. Persons working in various industries, such as oil, textiles and paper and miners exposed to radon have been reported to have an increased propensity of developing GBC (54). Some environmental toxins, such as fungal aflatoxins and ochratoxins, are substantial risk factors associated with GBC. A wide range of fungal species produces aflatoxins. Although these compounds' exact mechanism of action is not fully understood, there are postulated mechanisms of cause for these carcinogens. The presumed mechanism of action of aflatoxins is continued exposure of gallbladder epithelium to the carcinogenic aflatoxin metabolites excreted from the liver into the bile (55). Ochratoxin A is another compound produced by fungi that have been linked with GBC development. Red chilli peppers from Southern Latin America contain high levels of ochratoxin A. It has been speculated that GBC development associated with ochratoxin A may be stronger than that of aflatoxins (56).

Other gallbladder-related abnormalities, such as polyps, infection, and congenital biliary tree defects, have also been observed to be risk factors for GBC onset. Gallbladder polyps are benign growths which usually form in the mucosal lining of the gallbladder. Polyps are often identified via routine ultrasound examination or incidentally after cholecystectomy. The most common benign polyps are cholesterol polyps (60%) and adenomyoma polyps (25%). Adenocarcinoma polyps (80%) are the most prevalent malignant polyps (9,57). Chronic bacterial infections can also stimulate gallbladder irritation and inflammation; *Salmonella typhi* and *Helicobacter bilis* are two types of bacteria known to induce GBC. Patient carriers of *S typhi* can have up to 12-fold increased risk of GBC; infection with *S. typhi* is twice as common in females than in males (9,17,58). *Helicobacter bilis* has been associated with the formation of cholesterol gallstones. Moreover, chronic bacterial infection induces inflammation of the gallbladder and biliary tree; this causes DNA damage which is a critical factor in the tumourigenic process (17,59).

An anomalous pancreaticobiliary ductal junction (APDJ) is a rare congenital structural defect detected in the biliary tree. This is a rare congenital defect whereby the biliary duct and pancreatic duct junction are outside the duodenal wall. This causes the reflux of pancreatic

juice and bile into the ducts (60,61). The pancreatic juice constantly encompasses the biliary mucosa causing injury and inflammation, predisposing the mucosa to carcinogenesis. It is estimated that approximately 10% of GBC patients have APDJ; however, it was noted that Asian populations have a more prevalent incidence of APDJ (62).

1.4.3 Types of gallbladder cancer

Cancers are classified into several types depending on the tissue. The most common variants are adenocarcinoma, squamous cell carcinoma, and adenosquamous carcinoma. Neuroendocrine cell tumours, small cell carcinoma, and anaplastic carcinomas are rarer variants (63). Gallbladder cancers are usually adenocarcinomas (up to 97%), with the other variants accounting for 3 – 20% of cases (64).

Adenocarcinomas can be further divided into subtypes: papillary, tubular, and nodular variants (5). Adenocarcinomas originate from the glandular cells of the gallbladder. Gallbladder papillary adenocarcinomas are observed to be less aggressive than the other subtypes (5). Squamous cell carcinomas originate from epithelial tissues, which can be described as stratified squamous epithelium. Adenosquamous cell carcinoma is a combination between adenocarcinoma and squamous cell carcinoma; however, the exact amount of each component is unclear (65,66). Neuroendocrine cell tumours are derived from specialised neuroendocrine cells that function as chemo- and mechanosensors (67). Small cell carcinomas of the gallbladder are defined as poorly differentiated neuroendocrine tumours and are associated with a poor prognosis (68).

The specific symptoms and techniques used to identify GBC in patients are described in detail in the following section.

1.5 Clinical presentation and diagnosis

The detection of gallbladder cancer often occurs as an incidental finding during surgical procedures intended for other causes/diseases or on imaging techniques. Early-stage GBC is usually detected after cholecystectomy and a pathological review of the gallbladder (69). GBC patients are often asymptomatic but can present with vague and non-specific symptoms such as abdominal pain, nausea or vomiting, weight loss, weakness, and jaundice. These symptoms are all similar to cholecystitis which may confuse the diagnosis (57,69). Additionally, some tumour markers are employed to diagnose and evaluate GBC, such as carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA 19-9). However, these markers lack

specificity (79.2% for CEA and 92.7% for CA 19-9) and sensitivity (79.4% for CEA and 50% for CA 19-9) (69). Due to the lack of specificity and sensitivity, other techniques have been explored to identify GBC, including the cytological examination of gallbladder tissues. Endoscopic ultrasound-guided fine needle aspiration (EUS-FNA) is a technique used to obtain a minute tissue sample of GBC for cytological observation. It results in minimal complications, is minimally invasive, and is highly sensitive and specific. However, this technique is not included in current standardised guidelines for diagnosing and managing GBC (57,70,71).

Some surgical techniques are also employed for the diagnosis and possibly as a treatment for GBC. Endoscopic retrograde cholangiopancreatography (ERCP) is an invasive procedure to evaluate patients with biliary pathologies. A side-viewing duodenoscope is introduced through the patient's mouth, down the oesophagus, and into the duodenum. A secondary tube is then inserted to reach the bile duct (72). However, due to the invasive nature of ERCPs, this procedure can cause complications such as sepsis, haemorrhage, pancreatitis, and bile leakage. Despite these complications, ERCP is still the gold standard for investigating the hepatopancreaticobiliary (HPB) region and can be used for diagnosis and treatment (73). Magnetic resonance cholangiopancreatography (MRCP) is another diagnostic technique used to identify GBC, and it is a non-invasive alternative to ERCP. MRCP may assist with assessing the disease extent more accurately by providing better visualisation of the HPB system and can identify patients where gallbladder resection would not be undertaken. However, in contrast to ERCP, MRCP has no therapeutic avenue, although if no therapeutic intervention is necessary, MRCP avoids the complications associated with ERCP (69,73).

Various imaging techniques for the diagnosis of GBC are also used; these include ultrasonography, computed tomography (CT), and magnetic resonance imaging (MRI) scans (5). Ultrasonography may be inadequate in distinguishing between cholecystitis-associated wall thickening and GBC and has an accuracy of 68.8%. CT scans are the imaging technique of choice for detecting and staging GBC (63). They have a reported success rate of 84% for tumour staging and 85% for liver and lymph node involvement and metastasis (74). MRI scans are used less frequently than CT scans to stage GBC. However, the high contrast resolution can identify anomalies. When used in conjunction with MRCP, MRI scans improve the accuracy of diagnosis and disease progression (tumour stage with 84% accuracy, 100% accuracy for liver invasion, and 92% for lymph node involvement) (63). As with any other cancer type, the progression of GBC may be defined by various stages. Additionally, the cell

phenotype (cellular differentiation) may also be classified to assess disease severity. The specific GBC staging and differentiation grades are described in the next section.

1.6 Gallbladder cancer staging and differentiation

Staging the cancer is essential to understand the progression and severity of the disease. Staging classifications are determined based on the location of cancer: local tumours, locally advanced or distant metastasis, and lymph node invasion (5). The most widely used staging system is the TNM system by the American Joint Committee on Cancer (AJCC) (Table 1.1).

Table 1.1: Gallbladder cancer staging as per the AJCC (75)

T Category	**Definition**
Tis	Carcinoma *in situ*
T1a	Tumour invades lamina propria
T1b	Tumour invades muscular layer
T2	Tumour invades perimuscular connective tissue with no extension into the liver
T2a	Invasion of the perimuscular connective tissue on the peritoneal side
T2b	Invasion of the perimuscular connective tissue on the hepatic side
T3	A combination of: perforation of the serosa or directly invades the liver or one other adjacent organ or structure
T4	Tumour invades the main portal vein or hepatic artery or invades two more extrahepatic organs or structure
N Category	**Definition**
N0	No regional lymph node involvement
N1	Metastasis to 1-3 regional lymph nodes
N2	Metastasis to ≥ 4 regional lymph nodes
AJCC Stage Groups	**Definition**
I	T1 N0 M0
IIA	T2a N0 M0
II	T2 N0 M0
IIB	T2b N0 M0
IIIA	T3 N0 M0
IIIB	T1-3 N1 M0
IVA	T4 N0-1 M0
IVB	Any T N2 M0, any T, any N M1

T = the advancement of the tumour and any invasion into nearby and distant tissues.
N = lymph node involvement.
M = metastasis of the tumour to any distant tissues, organs or both.

Tumours are also graded according to tumour cell differentiation. Differentiation relates to the phenotypical similarity to normal cells and is defined from grade 1, well-differentiated, to the undifferentiated grade 4 (5). Well-differentiated cancer cells resemble normal cells, and undifferentiated cells do not. Grade 3 tumours (poorly differentiated cells) are the most common in GBC. However, the cell grade has no prognostic impact on the GBC tumour (5).

1.7 Molecular mechanisms of gallbladder cancer progression

Gallbladder cancer is considered rare. Over the years, several studies have been conducted in some population groups to decipher the molecular mechanisms linked to this disease. Molecular changes, such as gene aberrations and protein dysregulation, play a significant role in GBC onset and progression (76). Current data suggest that molecular changes associated with GBC may vary across different geographical and ethnic groups, highlighting the need for investigating these changes across different groups (17). Furthermore, these changes can be interrogated in the context of enriched signalling pathways (77–80). Signalling pathways govern the fundamental cellular processes, such as proliferation, via a signal transduction cascade. The dysregulation of these signalling pathways is a hallmark of carcinogenesis (77). Some of the pathways dysregulated in GBC include those associated with p53, receptor tyrosine-protein kinase 2/3 (ERBB2/3) (76), Kirsten rat sarcoma viral oncogene homolog (KRAS) (81), epidermal growth factor receptor (EGFR) (82), hedgehog (Hh) (83), and cyclooxygenase-2 (COX2) (84) among others.

1.7.1 P53 pathway

Mutations in p53 are common and identified across several cancers. P53 functions as a tumour suppressor to prevent cancer onset and progression by inducing apoptosis in dysfunctional cells (85). During GBC tumourigenesis, alterations in the p53 gene are commonly observed in exons 5, 6, 7, 8, and 9. However, the frequency of exon mutations seems largely conditional on geographical location and ethnicity. Patients from the USA and Chile show more frequent mutations in exons 5 and 6 compared to exons 7, 8, and 9 (86). However, Peruvian patients have more frequent mutations in exons 7 and 8, compared to exons 5 and 6 in Chilean patients. Similar to the Peruvian population, Japan (50%) and Hungary (33%) showed mutations in exons 7 and 8. These variations in mutational signatures further highlight the differences in molecular signatures of GBC patients associated with different populations and ethnicities (86).

1.7.2 ERBB2/3 and the PI3K/Akt/mTOR signalling pathway

The proteins ERBB2/3 are protein-tyrosine kinases crucially involved in cell cycle regulation via the phosphoinositide 3-kinase (PI3K)/protein kinase B (Akt)/mammalian target of rapamycin (mTOR) pathway (79,87). Upregulation of *ERBB2* has been observed in various human cancers and is dysregulated in 33 – 64% of GBC cases (77,88). ERBB proteins do not necessarily bind ligands but rather form heterodimers which become hyperphosphorylated, causing hyperactivity and inducing tumour initiation and progression (76,77,87). Three antibody treatments are currently approved by the US Food and Drug Administration (FDA) for treating ERBB2-driven tumours: trastuzumab, ado-trastuzumab, and emtansine (89).

1.7.3 Cyclooxygenase-2

Cyclooxygenase-2 (COX2) is an enzyme involved in tissue prostaglandin synthesis and has been observed to play a role in neoplastic conditions. Normal gallbladder tissue does not express COX2, while aberrant GBC tissue shows elevated COX2 expression (90,91). Inhibition of COX2 indicated a decrease in mitogenesis and an increase in apoptosis (92).

1.7.4 MAPK/ERK signalling pathway

The MAPK/ERK signalling pathway regulates the cell cycle involved in cellular proliferation and apoptosis (82). In GBC, the MAPK/ERK pathway was observed to be directly involved in the onset and progression of the tumour (Figure 1.2) (78,82). It is activated via various signalling molecules such as cytokines, EGF, and platelet-derived growth factor (PDGF) (93,94). Various proteins involved in this pathway were found to be either overexpressed or underwent a gain-of-function mutation. These proteins include ERK1/2, Ras, MALAT1, and SLC25A22. MALAT1 and SLC25A22 induce epithelial-to-mesenchymal transition (EMT), which aids in the migration of tumour cells (95–97). Furthermore, within the Ras family of oncogenes is the *KRAS* oncogene which has been observed to be mutated in GBC (86).

1.7.5 KRAS

The oncogene *KRAS* encodes the protein KRAS, a well-established protein in cancer and the tumourigenic process. Overexpression of this protein induces the MAPK signalling pathway and cellular proliferation (86,98). Mutations in the *KRAS* gene in western countries are identified in 0 – 10% of GBC cases and, in eastern countries, can range from 2 – 59%. The mutations can arise in various codons of the gene, namely codons 12, 13, and 61 (86). Additionally, *KRAS* mutations have been identified in other HPB system cancers, such as

pancreatic ductal adenocarcinoma (PDAC). Mutations in *KRAS* have been observed in >90% of PDAC patients. Recently, KRAS inhibitors have been employed in murine and cell-based studies and show full anti-tumour effects, making this area of research promising for targeting *KRAS* mutations in cancer (99). While it is crucial to understand the proteins involved in GBC carcinogenesis, the pathways in which these proteins are involved are essential in understanding the processes implicated in cancer.

1.7.6 PI3K/Akt/mTOR signalling pathway

The PI3K/Akt/mTOR pathway has been identified to be dysregulated in GBC. This pathway is involved in the cell cycle and cellular survival (Figure 1.2). Phosphorylated mTOR (p-mTOR) has been associated with poor prognosis in GBC patients and is highly active in early GBC, which significantly increases as the tumour progresses (79,100). PI3K is regulated by the gene *PI3KCA*, and mutations in this gene are associated with various cancer types, including GBC. Moreover, PI3K is activated via receptor tyrosine kinases to induce Akt activation downstream (77,101,102). Various accessory proteins also activate and are activated by this pathway, including SPOCK1 (sparc/osteonectin, cwcv and kazal-like domains proteoglycan 1) and fibronectin (103,104). SPOCK1 has been indicated to have higher expression levels in GBC tumours than in normal tissues. SPOCK1 functions to phosphorylate Bax, a pro-apoptotic protein, inactivating it and allowing for increased survival (105,106). These proteins promote further functioning of the PI3K/Akt/mTOR pathway.

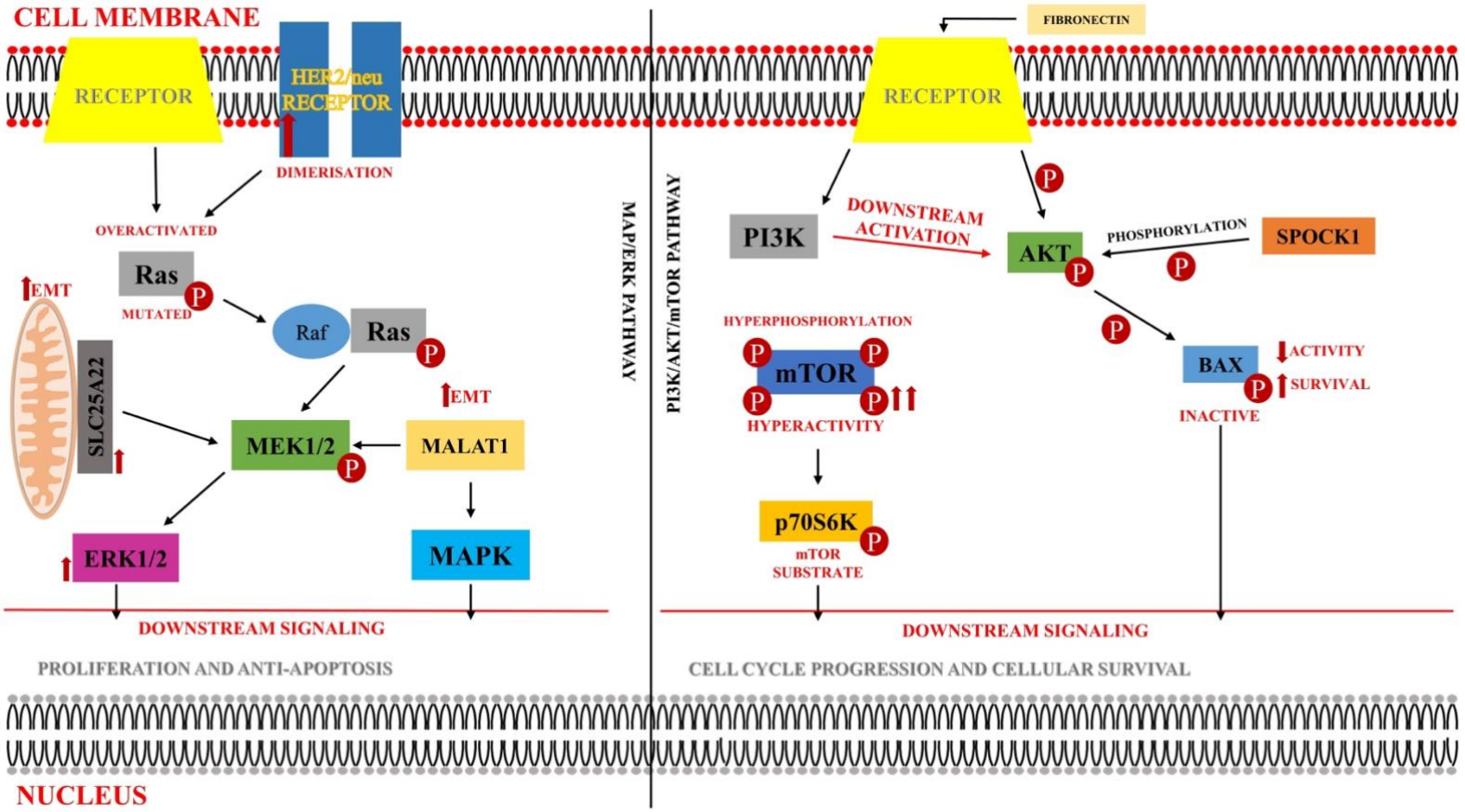


Figure 1.2: The MAPK/ERK and PI3K/Akt/mTOR signalling pathways (77). The MAPK/ERK pathway induces a downstream signalling cascade and is activated by various signalling biomolecules. Upon activation, Ras becomes over-activated due to mutations and phosphorylation. Raf is then recruited to form a complex with Ras, this complex activates MEK1/2 via phosphorylation. However, MEK1/2 is also activated via MALAT1 and SLC25A22; both MALAT1 and SLC25A22 are involved in EMT. Additionally, MALAT1 also induces activation of MAPK which induces a downstream signalling cascade. Increased MEK1/2 and ERK1/2 activations induce a downstream activation causing increased proliferation and anti-apoptosis. PI3K activates Akt downstream via phosphorylation. SPOCK1 also induces Akt activation via phosphorylation, which then causes the phosphorylation of SPOCK1. Phosphorylated SPOCK1 then phosphorylates Bax, a pro-apoptotic protein which becomes inactive after phosphorylation. Increased survival is achieved via the PI3K/Akt/mTOR pathway. Upstream effectors lead to hyperphosphorylation of mTOR, which subsequently phosphorylates p70S6K. Phospho-p70S6K increases protein translation via phosphorylation of the 40S ribosomal protein S6. Downstream, this pathway leads to cell cycle progression and survival.

1.7.7 Hedgehog signalling pathway

The Hedgehog (Hh) pathway is essential in several human developmental processes and is involved in cancer onset and progression (Figure 1.3) (77,83,107). There are three main Hh ligands: Sonic, Indian, and Desert (83,108). Once one of these ligands binds to the receptor, Smoothened (Smo), a transmembrane protein, is activated, transducing a downstream cascade of activation of transcription factors such as Gli1, Gli2, and Gli3. Gli1 has been suggested as a potential biomarker due to its strong activation of downstream targets. It has been observed that the suppression of Smo decreases GBC invasiveness by inhibiting the activity of matrix metalloproteinase-2 and -9 (MMP-2 and MMP-9) (83,108). MMPs degrade the basal membrane allowing for more significant invasive potential and inducing EMT transition

(77,109). Immunohistochemistry techniques identified that some Hh signalling proteins, including Sonic Hedgehog (Shh), Patched (Ptch1), and Gli1, were significantly higher in GBC compared to normal tissues and correlated to tumour staging and lymph node metastasis (110). GBC patients also showed higher expression levels of Ptch1 and Gli1 when compared to chronic cholecystitis patients. In contrast, levels of Shh were higher in chronic cholecystitis patients compared to GBC patients. Since Gli1 is a nuclear transcriptional regulator, it is speculated that aberrant expression of Gli1 may cause extended and altered expression of Shh, causing progression from chronic cholecystitis to gallbladder carcinogenesis. Furthermore, the upregulation of Gli1 led to reduced survival outcomes of GBC patients, suggesting its potential use as a positive prognostic biomarker (77,110,111).

1.7.8 EGFR signalling pathway

Epidermal growth factor receptor (EGFR) is known to be associated with various cancers. It activates signalling pathways such as the MAPK/ERK and PI3K/Akt/mTOR pathways (112). This indicates EGFRs involvement in cellular proliferation, angiogenesis, adhesion, and metastasis (113). Overexpression of EGFR was observed in 44 – 77% of GBC patients in various studies. This elevated expression correlated with poor prognosis. However, the frequency of somatic mutations of EGFR was low, ranging from 2.5 – 3.9%, indicating a different modality of overexpression (76,114).

Moreover, EGFR is implicated in its signalling pathway, the EGFR signalling pathway, and regulates similar cellular processes as the MAPK/ERK and PI3K/Akt/mTOR pathways (Figure 1.3) (115). This pathway has been targeted in several therapies, and EGFR is observed to be expressed in up to 93.33% of GBC cases (116). The EGFR pathway also interacts with and activates the PI3K/Akt/mTOR pathway, indicating crosstalk between pathways involved in the GBC carcinogenic process. The PI3K subunit p110α binds EGFR and induces a signalling cascade. Also, a mutation in p110α increases EGFR binding affinity to Akt, increasing downstream activation and, subsequently, GBC progression (117). Moreover, it is possible to target EGFR to suppress Akt and ERK signalling. Gefitinib, an EGFR tyrosine kinase inhibitor, prevents EGFR autophosphorylation. After Gefitinib treatment, phosphorylation of EGFR, Akt, and ERK was significantly reduced, resulting in the induction of apoptosis by increasing the expression of the pro-apoptotic protein, Bax (77,118).

Identifying and understanding dysregulated signalling pathways in GBC have helped to delineate the molecular mechanisms associated with disease progression. Moreover, p53

overexpression, *KRAS* mutations, amplification of ERBB2/3, and EGFR overexpression have been identified as molecular prognostic markers for GBC (119). However, these proteins are present and aberrant in various cancers, so there is still a need to discover relevant biomarkers that could help with GBC disease detection at the early stages of the disease.

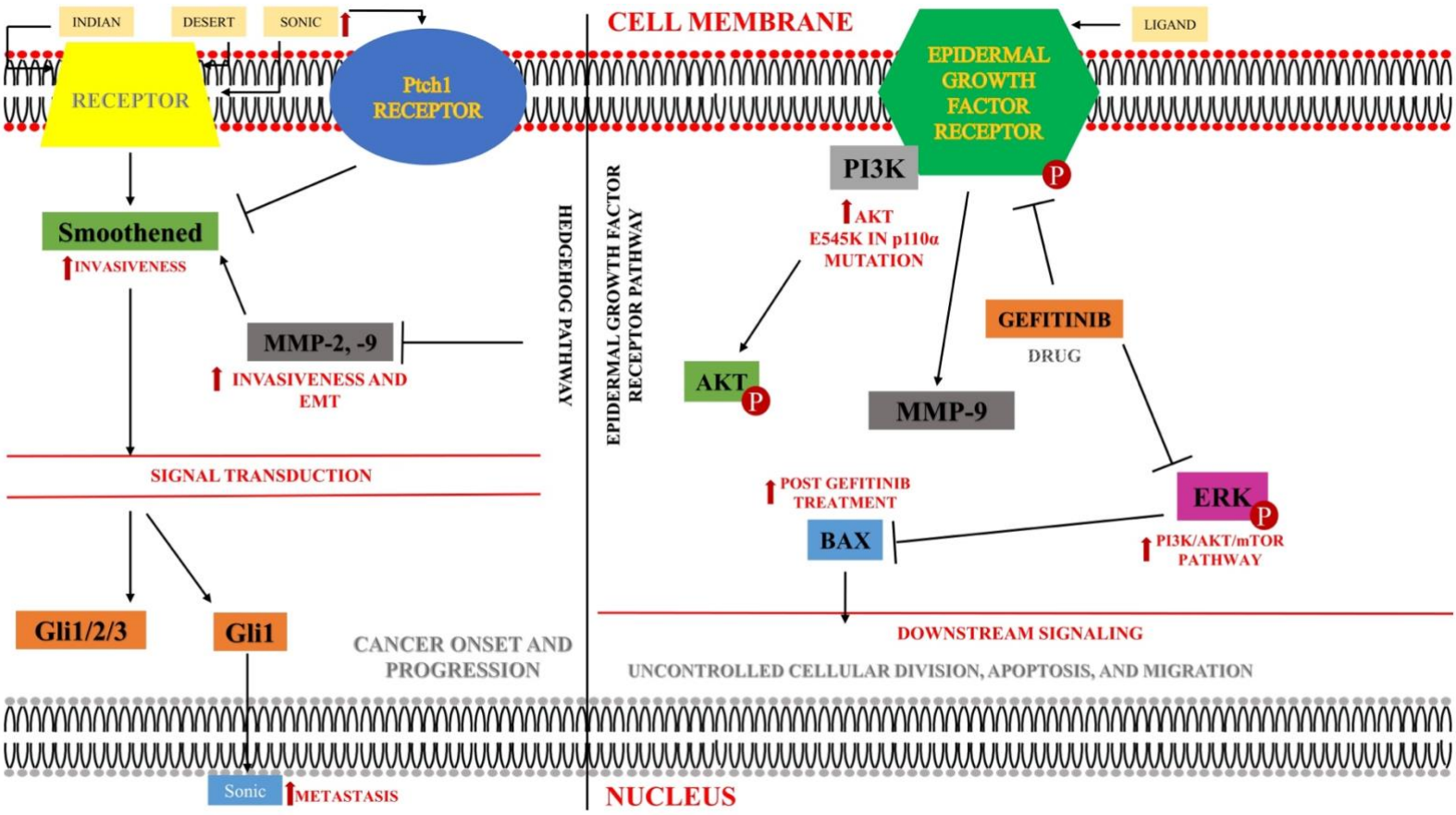


Figure 1.3: Hedgehog and epidermal growth factor receptor (EGFR) signalling pathways (77). The Hedgehog (Hh) signalling pathway is activated by various biomolecules: Indian, Sonic, and Desert. The Indian and Desert biomolecules bind to the receptors to activate Smoothened (Smo) and activated Smo increases the cellular invasive potential. Conversely, Sonic serves to inhibit Smo functioning after binding to the Ptch1 receptor. Furthermore, Smo can also be blocked by inhibiting the activity of MMP-2 and -9. Activated Smo induces a signalling cascade downstream to activate Gli1/2/3. Gli1 functions as a nuclear transcriptional regulator and may induce altered and extended expression of Sonic. This altered expression of Sonic increases the metastatic potential. Ultimately, the Hh pathway is involved in cancer onset and progression. The EGFR pathway is activated when various signalling biomolecules bind to EGFR, which results in phosphorylation. EGFR also activates the PI3K/Akt/mTOR pathway by binding PI3K which induces Akt activation. Moreover, Gefitinib is a drug which blocks EGFR functioning. Gefitinib also blocks phospho-ERK, which functions to block Bax. Therefore, Gefitinib treatment induces apoptosis by increasing the expression of Bax. Furthermore, blocking EGFR also blocks MMP-9 expression. The EGFR signalling pathway is involved in the uncontrolled cell division, apoptosis, and migration.

1.8 Protein biomarkers in gallbladder cancer

Biomarkers can aid in the detection of diseases. Diagnostic markers can detect and monitor disease, while prognostic biomarkers predict disease progression (120). A good biomarker should have sufficient sensitivity and specificity (>90%) for the disease and be easy to detect, for example, those secreted into biofluids such as blood and urine. Several biomarkers have been identified for use in GBC diagnosis, including carbohydrate antigen (CA 19-9), carcinoembryonic antigen (CEA) (121), carbohydrate antigen 125 (CA 125) (122), and the soluble fragment of cytokeratin 19 (CYFRA 21-1) (123).

Carbohydrate antigen 19-9 (CA 19-9) is a cell surface glycoprotein complex produced by the biliary system ductal cells. The presence of CA 19-9 depends on the Lewis antigen; however, up to 22% of the population cannot produce this antigen and, consequently, cannot express CA 19-9 (124). CA 19-9 also has a sensitivity and specificity for GBC of only 50% and 92.7% (69). This protein is also elevated in benign gallbladder diseases such as cholelithiasis (125). Due to this, it is not recommended to use this protein solely in confirming the diagnosis of GBC but instead in conjunction with other markers or screening tests (124). Like CA 19-9, carcinoembryonic antigen (CEA) is a glycoprotein intracellular adhesion molecule expressed by epithelial tumour cells. Increased serum CEA levels have been detected in various cancers, such as pancreatic and colorectal cancers. However, CEA levels may also increase in benign diseases such as pancreatitis and liver disease (126). The sensitivity and specificity of CEA vary widely in various studies making this protein not ideal as an individual diagnostic protein for GBC but rather to be used in conjunction with other proteins and clinical markers (127). It has also been observed to distinguish GBC from benign diseases with 79.4% and 79.2% sensitivity and specificity (69). Combined with other markers, this can increase to 87.5% sensitivity and 85.7% specificity (125).

The soluble fragment of cytokeratin 19 (CYFRA 21-1) is another strong candidate as a GBC marker. Cytokeratin 19 is an intermediate filament constituent widely expressed in epithelial cells in various organs. CYFRA 21-1 is detected in peripheral blood in minimal quantities, although abnormal increases of serum CYFRA 21-1 has been reported in various cancers, such as cholangiocarcinoma (128), lung cancer (129), and head and neck cancer (130). Studies have indicated that, in GBC patients, CYFRA 21-1 has 74.6% sensitivity and 84.6% specificity and is correlated with cancer aggressiveness (119,128).

Various techniques have been developed for the identification of specific protein biomarkers. These techniques exploit the expression profile of proteins in diseases. The following section elaborates on the various techniques in protein expression profiling, specifically the relevance of mass spectrometry.

1.9 Biomarker discovery using proteomic tools

Proteomics is the wide-scale study of proteins and their expression profiles (131) and has extensive utility in biomarker discovery studies. Proteins are extensively used as biomarkers for disease due to their stability and role as effectors of biological processes (132). The landscape of the proteome can potentially detect new protein biomarkers, which may be helpful in disease diagnosis and drug discovery (133,134). Various techniques have been developed and are used in the large-scale detection of proteins, such as 2-dimensional differential gel electrophoresis (2D-DIGE), protein arrays, and liquid chromatography-mass spectrometry (LC-MS).

1.9.1 Two-dimensional differential gel electrophoresis (2D-DIGE)

Two-dimensional differential gel electrophoresis (2D-DIGE) is used to identify thousands of proteins and their abundance. The two-dimensional separation of proteins is according to charge (first dimension) or isoelectric focusing (IEF) and molecular weight (second dimension) (135). This technique can be multiplexed by separating more than one sample on 2D-DIGE using fluorescent cyanine dyes known as Cy3 and Cy5. The separated samples can then be visualised on the same 2D gel using fluorescent imaging techniques according to the specific excitation wavelengths of the dyes (136). The main advantage of 2D-DIGE is its ability to analyse multiple pre-labelled samples on the same 2D gel compared to traditional Coomassie and silver staining methods. This limits gel-to-gel variability and increases reliability and reproducibility (137,138). A significant limitation of 2D-DIGE is the lack of suitable dyes for protein detection. Furthermore, low-abundant proteins become difficult to detect. Moreover, spots may comprise multiple proteins, or a singular protein may be present on multiple spots due to differential digestions or isoforms. Due to this, acid and basic proteins are not well represented (137,139).

1.9.2 Protein arrays

Protein arrays are another technique which has been used in the detection of cancer biomarkers. They exploit the various binding and structural properties of proteins and are useful in detecting

protein biomarkers (140,141). Protein arrays may be divided into analytical, functional, and reverse-phase protein arrays and may be considered multiplexed versions of enzyme-linked immunosorbent assays (ELISAs).

1.9.2.1 Analytical and functional arrays

Analytical arrays fix an antibody or lectin onto a solid matrix. After that, there are various approaches to detecting protein biomarkers. First, the targeted protein can be directly labelled with a fluorophore, bind directly to the fixed antibody/lectin on the matrix and detect the emitted signal. The other approach is the sandwich format, whereby the targeted protein is introduced to bind to the array matrix. A secondary antibody is introduced to bind to the target protein. The secondary antibody contains the fluorophore for signal detection (141).

Functional arrays are distinctive from analytical arrays as individually purified recombinant proteins are spotted onto the matrix (142). Functional arrays are usually employed to elucidate the binding properties of proteins such as protein-protein, protein-peptide, protein-nucleic acid, protein-glycan, and protein-lipid interactions (141) and are used to identify posttranslational protein modifications. The main advancement of functional arrays is the identification of serological biomarkers, as the proteins can be scrutinised for immunogenic activity and potential biomarker use (143).

1.9.2.2 Reverse-phase protein array (RRPA)

Reverse-phase protein array (RRPA) differs from analytical and functional arrays because the whole sample lysate is immobilised on the matrix (141,144). Many sample lysates can be immobilised on a single matrix, allowing for high-throughput analysis (145). Antibodies with known binding properties are individually incubated on the matrix to examine protein sequences or structures. RRPA has been used to study differentially regulated signalling networks in clinical samples (146).

Protein arrays are robust and ideal for detecting biomarkers due to their miniaturised features, high-throughput capabilities, and sensitive detection (141,147). Unfortunately, the use of protein arrays is limited due to the high cost of producing antibodies with a high specificity and affinity for their targets. Moreover, preserving the tertiary, functionally active protein structure before analysis limits using protein arrays as a proteomic strategy (148).

1.9.3 Liquid chromatography-mass spectrometry (LC-MS)

Liquid chromatography-mass spectrometry (LC-MS) identifies proteins based on their mass-to-charge ratio (*m/z*). Proteins are typically prepared via proteolytic digestion by endoproteinases such as trypsin. Reverse-phase liquid chromatography (RPLC) is a commonly used LC strategy when coupled to MS for proteomics. RPLC induces peptide binding to the stationary phase using a polar aqueous mobile phase. Hydrophilic peptides will be eluted first, and then the more hydrophobic peptides using a mobile phase gradient with increasing organic solvent. This fractionation of elution deepens the coverage of the proteome (149–151). The resulting eluted peptides are then ionised and detected on a mass spectrometer. Tandem MS (MS/MS) operates by firstly acquiring precursor ion spectra (MS1) and then secondly acquiring fragment ion spectra (MS2) (152). The acquired MS/MS spectra can be matched to determine the peptide sequence information using database searching tools. They can resolve the identity and abundance of matching proteins in a sample. Database searching depends on the search criteria, such as proteolytic enzyme constraints, post-translational modifications allowed, and *m/z* ranges (152).

MS has many advantages, such as in-depth coverage of the proteome and provides additional data, such as structural information and proteoform detection. Moreover, MS does not require antibodies for protein detection. However, protein-protein interactions may be studied using affinity purification-mass spectrometry (156). MS, similar to protein arrays, is also rapid and high-throughput, as it can analyse large sample numbers quickly and precisely (157). However, some of the limitations of MS are time-consuming sample preparation, especially in large sample cohorts, and it is a highly specialised technique that requires skilled technicians to operate. It also cannot differentiate between isomers of the same *m/z* ratio (158). Considering the limitations of MS, it is still a robust and reliable technique for identifying sample proteomes. There are several approaches to MS for identifying proteins, including bottom-up proteomics and label-free approaches such as the SWATH-MS detection method.

1.9.4 Bottom-up proteomics

Bottom-up proteomics describes the general approach used to identify proteins and characterise the amino acid sequences by proteolytic digestion before MS analysis (159). Bottom-up proteomic approaches use physical property advantages that peptides have compared to proteins. They are more easily separated in liquid chromatography (LC), ionise well, and fragment more predictably. This allows for a more robust methodology and the identification

and quantification of thousands of proteins from complex lysates (160). Bottom-up proteomics may use a variety of proteolytic enzymes, with trypsin being the most common. Trypsin predictably cleaves arginine and lysine residues producing a positively charged C-terminus of the peptides, suitable for positive mode electrospray ionisation MS. Although trypsin efficiently digests peptides, approximately 56% are ≤ 6 residues long and too small to be effectively identified by MS (161,162). However, the main drawback of only using trypsin is that it has a lower digestion efficiency of cleaving lysine residues than arginine residues, hampering the accuracy and sensitivity of the peptide quantification (163,164). This shortcoming can be overcome by using endoproteinase Lys-C, a protease that specifically cleaves at the C-terminus of lysine residues, reducing the missed cleavage rate (165). Bottom-up proteomics may be implemented using either labelling or label-free strategies. This study used a label-free approach (SWATH-MS) for protein detection.

1.9.5 Sequential window of all theoretical mass spectra (SWATH-MS)

A relatively new strategy for LC-MS-based detection of proteins is the sequential window of all theoretical mass spectra (SWATH-MS) (166). In SWATH-MS, non-labelled protein samples are digested and analysed in a data-independent acquisition (DIA) manner. In SWATH-MS, complete ion precursor maps of all co-eluting precursor peptide ions within an *m/z* range (MS1) are first obtained. After that, a series of MS2 scans in set *m/z* isolation windows are used to generate corresponding fragment ion spectra. The precursor window is adjusted stepwise using predefined increments until the defined total precursor ion mass range is covered (167). Unlike data-dependent acquisition (DDA), which generates MS2 scans on the highest intensity MS1 precursor scans, SWATH generates MS2 scans regardless of the precursor ions' MS1 scan intensity, allowing for better coverage of lower-abundant peptides in a given sample (167,168). Quantitative analysis of peptides by SWATH-MS is typically based on the "peptide-centric scoring" strategy. The dataset is queried for the presence of specific peptides using query parameters that relate to the peptides; this is conventionally achieved using generated spectral libraries acquired from DDA methods (167). A significant advantage of SWATH-MS is the detection of peptides covering thousands of proteins with great quantitative consistency, reproducibility, and accuracy, which is ideal for biomarker studies. Various studies have established the utility of SWATH-MS in proteomic profiling in solid tumours to identify potential biomarkers (169–172).

1.10 Rationale

This study aimed to identify potential biomarkers for detecting GBC in a cohort of South African GBC patients. GBC is usually diagnosed and detected at an advanced stage due to non-specific clinical presentation. The prognosis at this stage is poor, and the survival rate is ~19%. The repercussions of late detection are catastrophic. Apart from the primary tumour, metastases can cause many other debilitating symptoms that disrupt daily functioning. Treatment options are minimal and can contribute to the use of large amounts of healthcare resources to facilitate, at a minimum, symptomatic relief to patients and other social supports where needed. The impact of this study on the socioeconomic paradigm of the community would be paramount in allowing a potentially more efficient diagnosis of this disease. This would potentially alleviate the burden on social grants administered as well as reduce the large costs incurred to the healthcare system. Current clinical biomarkers used in diagnosing GBC are not ideal due to their poor sensitivities and specificities. Additionally, current biomarkers utilised in the diagnosis of GBC have been developed using international populations, such as European and North American patients. Molecular variations in these populations may be linked with varied prevalence of GBC. This suggests the vital importance of investigating a South African GBC population due to the variability of genetics and environment compared to those of European and North American ancestry, which can contribute to changes in some biological mechanisms (173). This suggests the need to discover novel potential biomarkers to detect GBC. Furthermore, elucidating the molecular mechanisms of GBC, such as the various dysregulated pathways, is vital in understanding the progression of this disease. Proteins are suitable as potential biomarkers as these biomolecules are the effectors of disease and their aberrant expression may indicate disease onset and progression. Many of these proteins are directly involved in the signal transduction pathways and studying them allows a better understanding of the molecular mechanisms of the disease. This makes protein biomarkers ideal candidates for identifying and understanding GBC. Additionally, molecular studies on GBC patients of African ancestry are lacking, and studies have suggested some distinct signatures across different ethnicities and populations. This study undertook a high-throughput SWATH-MS proteomics approach and, crucially, used independent tissue and plasma sample groups biosamples from South African GBC patients to demonstrate novel proteomic signatures to validate the findings in the tissue group.

1.11 Aims and Objectives

1.11.1 Aim

To identify proteomic signatures in a cohort of South African gallbladder cancer patients using high-throughput SWATH-Mass spectrometry.

1.11.2 Objectives

1. To determine the demographic and clinicopathological characteristics in a South African gallbladder cancer patient cohort.
2. To identify relatively abundant proteins in tissues and plasma of gallbladder cancer patients compared to patients with benign biliary pathologies (such as gallstone disease) and normal gallbladders.
3. To identify potential molecular mechanisms associated with gallbladder cancer progression using bioinformatic tools for network and pathway analyses.
4. To evaluate the quality of the identified target proteins for their utility in identifying gallbladder cancer.
5. To demonstrate the effectiveness of the target proteins in distinguishing between gallbladder cancer and benign disease conditions using statistical tests such as principal component analysis.

2 CHAPTER 2: MATERIALS AND METHODS

2.1 Summary of experimental strategy

Ethical approval was obtained from the Human Research Ethics Committee of the University of the Witwatersrand (Medical). Samples were obtained from surgical patients admitted to Surgical Unit 1 at Chris Hani Baragwanath Academic Hospital. Tissue, blood, and available patient clinical information were collected. Tissue samples were collected using Tru-cut ultrasound-guided biopsies, cholecystectomies, and liver transplant donors. Blood samples were collected in 10 mL EDTA blood vials and processed within six hours of collection. After sample collection, the protein was extracted from the tissue samples. For plasma samples, proteins were diluted prior to downstream preparation. The proteins were purified by protein aggregation capture (PAC) and hydrophilic interaction liquid chromatography (HILIC) preparation methods using trypsin and Lys-C endoproteinases to form smaller peptides. The digested peptides were injected into the mass spectrometer using data-dependent acquisition (DDA) and data-independent acquisition (DIA) detection methods to identify relatively dysregulated proteins. The dysregulated proteins in the tissue and plasma datasets were visualised for any commonly dysregulated proteins between the sample groups using Venny. Subsequently, quality control of the proteins found to be commonly dysregulated across the tissue and plasma sample groups was performed by investigating their MS2 fragment chromatograms. Network and pathway analysis was performed using Cytoscape and stringApp to elucidate the molecular mechanisms and pathways involved by the relatively dysregulated proteins. Hierarchical cluster and principal component analyses were performed to determine whether the differentially expressed proteins in each group comparison could distinguish between diseased and control states. The Spearman's rank correlation test was done to critically analyse the detected commonly dysregulated proteins to determine if there were correlations between the expression of these proteins. A summary of the methods used is shown in a flowchart represented in Figure 2.1.

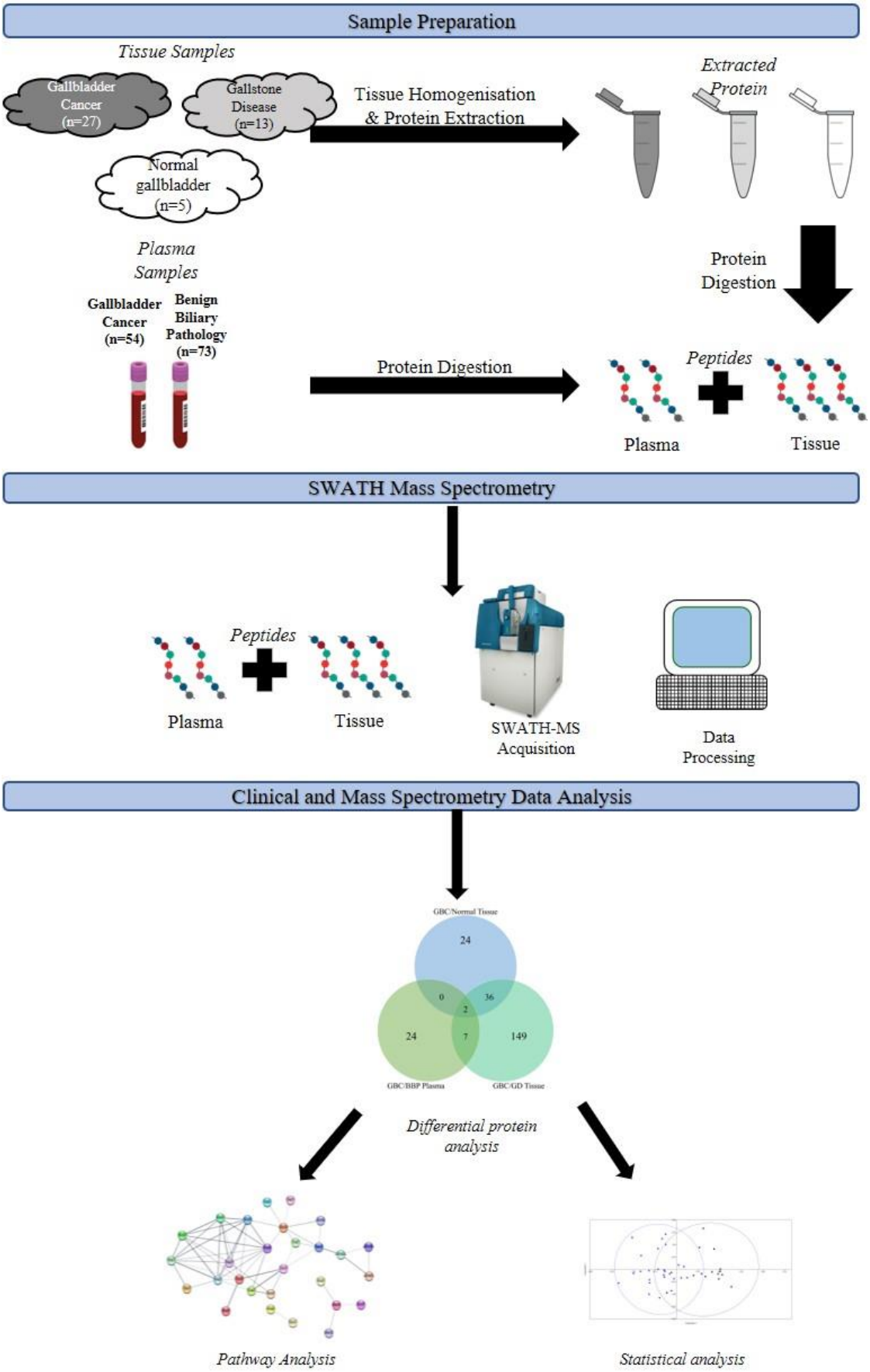


Figure 2.1: Flow chart of the summary of methods used. Sequential window of all theoretical mass spectra (SWATH-MS) analysis was used to profile biosamples from patients with gallbladder cancer and benign biliary pathologies, such as gallstone disease. Proteins dysregulated in gallbladder cancer patients were identified. Subsequently, pathway and statistical analyses were conducted.

2.2 Ethics statement

Ethical approval (M190555) was obtained from the Human Research Ethics Committee (Medical) of the University of the Witwatersrand, Johannesburg (Appendix B). All patients recruited provided written informed consent to be enrolled in the study.

2.3 Sample and clinical data collection

Tissue patient sample group

Patients who presented to Surgical Unit 1 at Chris Hani Baragwanath Academic Hospital (CHBAH), Johannesburg, South Africa, were recruited between April 2019 and December 2020. A total of 27 GBC tumours, 13 gallstone disease (GD) tissues, and five normal gallbladder tissues were included in this study. The inclusion criteria for this study were patients over 18 years of age, of African ancestry, and clinically and histologically confirmed primary diagnosis of GBC or GD. All GBC tissues collected were identified as advanced-stage (Stage IV) tumours according to the American Joint Committee on Cancer (AJCC) on Cancer Staging Manual 8th Edition (75). Exclusion criteria were patients with additional primary HPB disease diagnoses. A core sample of the GBC tumour was obtained from Tru-cut ultrasound-guided biopsy at the liver metastasis site, and GD tissue was obtained via laparoscopic cholecystectomy. Non-diseased gallbladder tissues obtained from liver transplant donors were used as normal samples. All tissue samples obtained were stored in approximately 700 µL of RNAlater™ (Sigma-Aldrich, Germany) and placed in a -80 °C freezer until further processing.

Plasma patient sample population

A separate sample population of patients who presented to CHBAH, Johannesburg, South Africa, were recruited between February 2021 and October 2021. A separate sample population of patients for plasma collection was utilised to serve as an independent validation for the tissue sample population. A total of 54 GBC and 73 benign biliary pathologies (BBP) patients were included in this study. The inclusion criteria for the GBC group were the same as those recruited for the tissue sample group. The inclusion criteria for the BBP sample group required patients to be over 18 years of age and clinically diagnosed with GD or cholecystitis, an inflamed gallbladder with or without gallstones. The exclusion criteria for both sample groups were patients diagnosed with any other HPB diseases.

Blood samples were collected in 10 mL EDTA blood vials and processed within six hours. Processing included separating blood into whole blood and plasma by allowing the blood vial

to stand upright. The plasma was then carefully transferred to a fresh 15 mL Falcon tube and centrifuged at 4000*g* for 30 minutes (Beckman Coulter Allegra™ X-22R benchtop centrifuge, Brea, California, USA) to remove any debris. After that, the plasma was aliquoted and stored at -80°C until required. There were missing data for the GBC patients, 23 patients (38.89%) for TNM staging and 14 patients (25.9%) for historic gallstone disease status.

Each patient's demographic and clinical information was captured on the day of collection and stored in REDCap (v11.3.4, Vanderbilt University, Nashville, Tennessee, USA).

2.4 Tissue homogenisation and protein extraction

Between 15 – 20 mg of the tissue was resuspended in 500 µL ATL Lysis Buffer (Qiagen, Hilden, Germany) and homogenised using the Tissue Ruptor (Qiagen, Hilden, Germany) until all the tissue was visibly solubilised. The total volume was determined, and 4x volume of cold acetone (LC-MS grade stored at -20°C) was added and incubated at -20°C for 60 minutes. The resulting precipitant was centrifuged at >14,000*g* for 10 minutes (Beckman Coulter Microfuge™ 18 benchtop centrifuge, Brea, California, USA). The pellet was washed with 100 µL ice-cold ethanol (LC-MS grade) and dried for approximately 1 minute. The pellet was resuspended in 200 µL 2% SDS in 50 mM Tris-HCl pH 8 supplemented with PhosSTOP phosphatase inhibitors (Roche, Basel, Switzerland). The solution was sonicated using probe sonication: 9 cycles of 10-second sonication followed by 10 seconds on ice at 70% power. The solution was centrifuged at >14 000*g* for 10 minutes (Beckman Coulter Microfuge™ 18 benchtop centrifuge, Brea, California, USA), and the supernatant was transferred to an 0.5 mL Eppendorf tube. The centrifugation was repeated, and the protein was quantified using the 2-D protein quantification kit (Cytiva, Massachusetts, USA).

2.5 2-D Protein quantification assay

Principle: The 2-D protein quantification assay measures protein concentrations in a sample using a known concentration protein standard curve solution of bovine serum albumin (BSA). The protein solution is precipitated using the provided precipitant and co-precipitant solutions. A diluted copper solution is then added to the protein precipitant. The Cu^{2+} ions chelate the peptide bonds of the proteins reducing the Cu^{2+} ions. The absorbance of the solution is then determined at 480 nm wavelength. The absorbance of the solution is inversely proportional to the protein concentration. The absorbance of the protein solution can be plotted against the BSA standard curve, and protein concentration is determined

(https://www.sigmaaldrich.com/ZA/en/technical-documents/protocol/protein-biology/protein-quantitation/protein-determination-using-2-d-quant-kit).

Protocol: Before performing the assay, the working colour solution was prepared by mixing 100 parts of Colour Reagent A with 1 part of Colour Reagent B. Each assay required 1 mL of working colour solution. The standard curve solutions were prepared as described (Table 2.1). The useful range of this assay is 0.5 – 50 µg protein.

Table 2.1: Preparation of the standard curve for the 2-D protein quantification assay

Tube Number	Volume of 2 µg/µL bovine serum albumin standard solution (µL)	Protein Quantity (µg)
1	0	0
2	5	10
3	10	20
4	15	30
5	20	40
6	25	50

Sample tubes were prepared by adding 10 µL of sample into separate fresh tubes. After that, 500 µL of the precipitant solution was added to each tube, vortex-mixed briefly, and incubated for 2 – 3 minutes at room temperature. Then, 500 µL of co-precipitant was added to each tube and vortex-mixed briefly. The tubes were centrifuged at 10 000*g* for 5 minutes (Beckman Coulter Microfuge™ 18 benchtop centrifuge, Brea, California, USA) to sediment the protein. As soon as the centrifugation was complete, the supernatant was removed rapidly to avoid resuspension of the pellet. Any remaining supernatant was removed after a brief centrifugation pulse. Subsequently, 100 µL of copper solution and 400 µL of distilled/de-ionised water was added to each tube and vortex-mixed to resuspend the protein pellet. After that, 1 mL of the prepared working colour solution was added to each tube and mixed. The tubes were incubated for 20 minutes at room temperature, and the absorbance was measured at 480 nm (BioTek PowerWave 340 UV/Vis plate reader, Winooski, Vermont, USA). The absorbance at 480 nm is inversely correlated to protein concentration; if the protein concentration is high, then colour development is low, and vice versa. Therefore, protein quantitation is determined from a linear regression model of the standard curve.

2.6 Protein aggregation capture

Principle: Protein aggregation capture (PAC) performs protein purification via solid-phase extraction on magnetic beads followed by on-bead protein digestion. Before protein aggregation onto the microbeads occurs, reduction and alkylation of the proteins are performed to disrupt the tertiary protein structure (linearise the protein) and break the disulphide bonds of cysteine residues to prevent the protein from re-folding. The linearised proteins are then aggregated onto the microbeads. The addition of acetonitrile, as the mobile phase, forces the proteins to aggregate on the beads, the stationary phase. Protein digestion uses endoproteinases, such as trypsin, resulting in a positive charge at the peptide C-terminus. Trypsin cleaves proteins at arginine and lysine residues (161,162). The digestion of the proteins is then terminated by lowering the pH of the solution. The small peptides are quantified preceding mass spectrometry analysis for protein identification (174).

Protocol: PAC was performed on all tissue samples (GBC, GD, and normal tissues). First, protein reduction was done by adding 10 mM dithiothreitol (DTT) to samples and incubating for 30 minutes at 37°C. Protein alkylation was done by adding 20 mM iodoacetamide (IAA) (final concentrations of DTT and IAA from 1 M stock solutions). The samples were incubated for 30 minutes at room temperature in the dark.

PAC was performed as previously described (174) with modifications (Table 2.2 for 96 deep well plate layouts). MagResyn™ Hydroxyl beads (ReSyn Biosciences, Edenvale, South Africa) were used for protein capture. A protein:bead ratio of 1:4 (by weight) was used for PAC; 20 µg of protein was used per sample, and trypsin was used in a ratio of 1:10 (protease:protein) for digestion (Table 2.2, Row G). A final concentration of 70% acetonitrile (ACN, LC-MS grade) was used for on-bead protein aggregation, which was allowed to occur for 10 minutes without agitation (Table 2.2, Row G). The PAC protocol, including on-bead digestion, was automated using a KingFisher™ Duo (Thermo Fisher Scientific, Massachusetts, USA) purification system. The protein was digested for four hours at 37°C with gentle agitation. Once the digestion was completed, the plate was transferred to a magnetic rack to recover digested peptides. The peptides were transferred to fresh 0.5 mL protein LoBind tubes (Eppendorf, Hamburg, Germany), and recovered volumes were determined. Digestion was terminated by adding trifluoroacetic acid (TFA, LC-MS grade) to a final concentration of 0.5%. The samples were frozen at -80°C and dried at -4°C using a CentriVap vacuum concentrator (Labconco, Missouri, USA). The peptides were resuspended in 2% ACN and 0.2% formic acid

(LC-MS grade). The resuspended peptides were quantified using the Pierce™ quantitative colourimetric peptide assay (Thermo Fisher Scientific, Massachusetts, USA).

Table 2.2: 96 deep well plate setup for protein aggregation capture

Row A	50 mM ammonium bicarbonate pH 8 with Trypsin – 200 µL total
Row B	70% Ethanol – 500 µL
Row C	70% Ethanol – 500 µL
Row D	95% Acetonitrile – 500 µL
Row E	95% Acetonitrile – 500 µL
Row F	95% Acetonitrile – 500 µL
Row G	Beads (minimum 5 µL)/Protein (up to 25 µL)/Acetonitrile (30% volume of beads + protein/70% of 100% Acetonitrile) – minimum 100 µL
Row H	Disposable plastic comb

2.7 Hydrophilic interaction liquid chromatography

Principle: Hydrophilic interaction liquid chromatography (HILIC) performs protein purification via solid-phase extraction on magnetic beads followed by on-bead protein digestion. The protein solution interacts with a stationary phase (magnetic beads) in an aqueous mobile phase. Similar to PAC, reduction and alkylation were performed to linearise the protein and break the disulphide bonds to prevent protein folding. HILIC is performed by first equilibrating the magnetic beads. Thereafter, the protein samples are allowed to aggregate onto the magnetic beads. Protein interactions with the proprietary functional groups on the magnetic beads were induced by adding ACN to the protein solution. Any unbound proteins are washed away, and protein cleavage is performed using the endoproteinases trypsin and Lys-C. Trypsin cleaves proteins at arginine and lysine residues, and Lys-C cleaves at lysine residues only (161,162). Lastly, the digestion is terminated by lowering the pH of the solution. The resulting peptides are then quantified preceding mass spectrometry analysis for protein identification.

Protocol: HILIC was performed on all plasma samples. Firstly, protein reduction was performed by adding 10 mM DTT and incubating for 30 minutes at 37°C. Alkylation was conducted by adding 20 mM IAA and incubating for 30 minutes at room temperature in the dark. A total of 30 µg of protein was reduced, alkylated and stored in HILIC binding buffer in a 1:1 ratio (200 mM NH_4Ac; 30% ACN, pH 4.5).

MagResyn™ HILIC beads (ReSyn Biosciences, Edenvale, South Africa) at a protein:bead ratio of 1:4 by weight were used for protein capture. The beads were equilibrated using an equilibration buffer (100 mM NH_4Ac; 15% ACN, pH 4.5), and on-bead aggregation was performed for 30 minutes to bind the protein to the HILIC beads. After that, two washes using 95% CAN (LC-MS grade) were performed. Following the washes, digestion was done using trypsin and endoproteinase Lys-C (1:20 and 1:100 ratio of protease:protein, respectively). Digestion occurred for 2 hours at room temperature with gentle agitation. Once digestion was complete, it was terminated using 1% TFA. The HILIC protocol was performed using an automated KingFisher™ Flex purification system (Thermo Fisher Scientific, Massachusetts, USA). Each step of the HILIC protocol was performed in a separate, fresh 96 deep well plate. Once the digestion was completed and terminated, the peptides were recovered using a magnetic rack and transferred to new 0.5 mL protein LoBind tubes (Eppendorf, Hamburg, Germany). The samples were frozen at -80°C and then dried at -4°C using a CentriVap vacuum concentrator (Labconco, Missouri, USA). The peptides were resuspended in 2% ACN and 0.2% formic acid. The resuspended peptides were quantified using the Pierce™ quantitative colourimetric peptide assay (Thermo Fisher Scientific, Massachusetts, USA).

2.8 Pierce™ quantitative colourimetric peptide assay

Principle: The Pierce™ quantitative colourimetric peptide assay kit measures the concentration of peptides relative to a known peptide digest standard. The proprietary solutions provided in the kit serve as modified BCA reagents that reduce Cu^{2+} ions into Cu^{1+}. The peptide amide backbone reduces the copper ions, followed by the chelator coupling to the Cu^{1+} ions producing a bright red complex which absorbs at 480 nm. Peptide concentrations of unknown samples can be interpolated using the standard curve of absorbance versus concentration values of the peptide digest standard (https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2F23275_quantpeptide_color_UG.pdf).

Protocol: The Pierce™ quantitative colourimetric peptide assay kit (Thermo Fisher Scientific, Massachusetts, USA) was used to determine the concentration of the digested peptides. Before the assay was performed, a protein standard curve was prepared (Table 2.3). The diluent used for this assay was 2% ACN and 0.2% formic acid (LC-MS grade).

Table 2.3: Standard preparation for Pierce™ quantitative colourimetric peptide assay

Tube Label	Diluent (µL)	Peptide Digest (µL)	Final Concentration (µg/mL)
Blank	50		0
A	0	100 of stock digest	1000
B	50	50 of Vial A	500
C	50	50 of Vial B	250
D	50	50 of Vial C	125
E	50	50 of Vial D	62.5
F	50	50 of Vial E	31.3
G	50	50 of Vial F	15.6

After that, 180 µL of working reagent per well was prepared: 50 parts Reagent A, 48 parts Reagent B, and two parts Reagent C per 180 µL. Peptide digest standards (20 µL) were added in triplicate to a 96-well plate. Subsequently, 5 µL of sample was added in triplicate and diluted with 15 µL 2% ACN and 0.2% formic acid. Next, 180 µL of the working reagent was added to each peptide digest standard and sample well and mixed thoroughly. The 96-well plate was then incubated at 37°C for 15 minutes, and absorbance was measured at 480 nm (BioTek PowerWave 340 UV/Vis plate reader, Winooski, Vermont, USA).

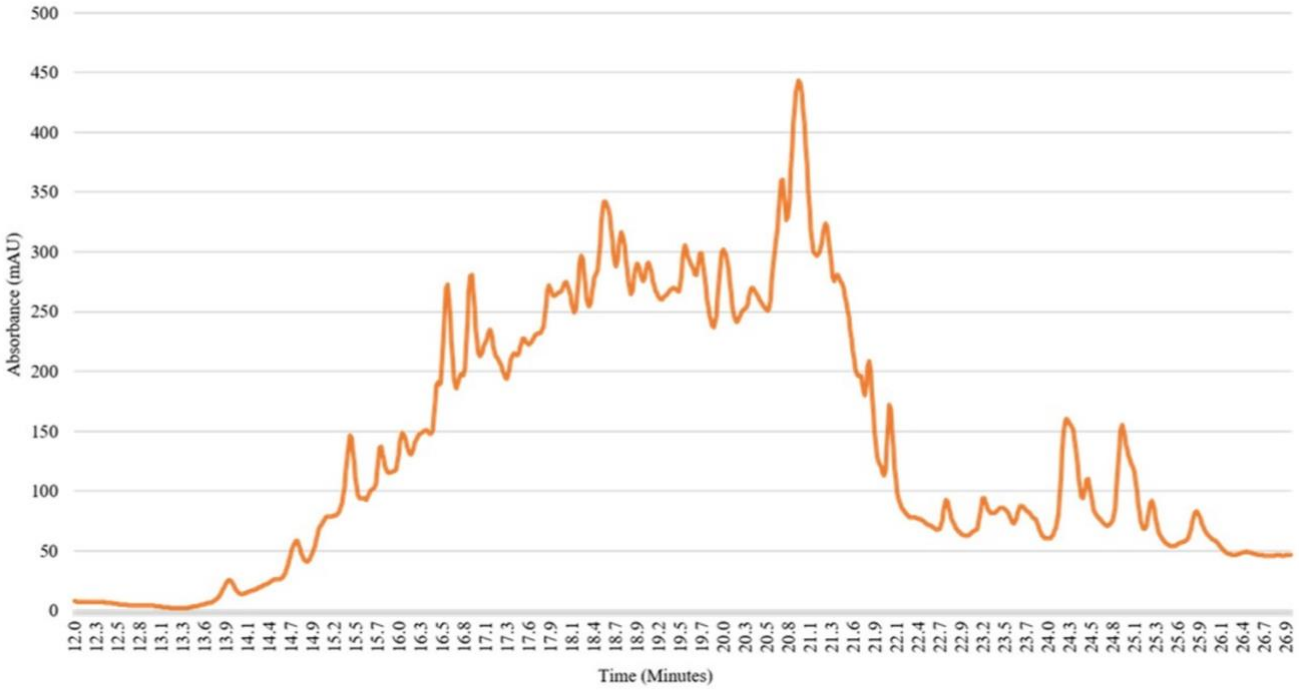
2.9 High pH reversed-phase fractionation

Principle: High pH reversed-phase fractionation divides the peptides in a sample based on the hydrophobicity. This allows more comprehensive protein identification in a given sample to build the spectral library. The peptides are bound to a hydrophobic resin under aqueous conditions. The peptides are then eluted using increasing concentrations of ACN into multiple fractions. The fractions are then vacuum dried to increase peptide concentration before mass spectrometry analysis. During mass spectrometry analysis, the fractions are separated again using a low-pH gradient which further reduces the sample complexity and allows the identification of low-abundant proteins.

Protocol: For high pH reversed-phase (RP) fractionation, a single aliquot of each prepared GBC and GD sample was pooled together in their respective groups. Approximately 30 µg of each pool was used for high pH RP fractionation. A linear gradient of 5 – 45% of Reagent B (Reagent A: 20 mM NH_4OH; Reagent B: 20 mM NH_4OH; 80% ACN) over 10 minutes at a flow rate of 75 µL.minute^{-1} was employed on a Hypersil GOLD C18 column (1 mm x 15 cm, 3 µm particle size) and maintained at 50°C to fractionate the pooled samples (Figure 2.2A). Fractions were collected at 30-second intervals between 13 – 23 minutes of fractionation. Appropriate fractions were collected and mixed (Figure 2.2B). Fractions were concentrated

using a CentriVap vacuum concentrator (Labconco, Missouri, USA) and resuspended before liquid chromatography-mass spectrometry.

A



B

Concatenation	Fractions
1	F2-4 + F12
2	F5 + F13
3	F6 + F14
4	F7 + F15
5	F8 + F16
6	F9 + F17
7	F10 + F18
8	F11 + F19-21

Figure 2.2: Elution profile for high pH reverse phase (RP) fractionation and fraction concatenation. (A) Fractionation profile of eluted peptides using a gradient of 20 mM NH_4OH and 20 mM NH_4OH/80% Acetonitrile using a Hypersil GOLD C18 column (1 mm × 15 cm, 3 µm particle size) maintained at 50°C over approximately 15 minutes. Fractions were collected at 30-second intervals between 13 – 23 minutes. (B) The fraction concatenation profile of collected fractions.

2.10 Liquid chromatography-mass spectrometry (LC-MS) data acquisition

Principle: Mass spectrometry (MS) identifies digested proteins based on their mass-to-charge ratio (*m/z*) and tandem MS/MS scans. Liquid chromatography elutes proteins based on their hydrophobicity, allowing deeper proteome coverage. There are two common label-free detection methods for MS, data-dependent acquisition (DDA) and data-independent acquisition (DIA). In the DDA mode, the most intense peptide ions from the first MS detection (MS1) are fragmented for a second MS detection (MS2). In DIA mode, the MS2 scan is performed for an *m/z* precursor range regardless of the intensity of the MS1 scan. The DDA detection method may be employed to build a spectral library for matching DIA data generated for individual study samples. Sequential window acquisition of all theoretical mass spectra (SWATH-MS) is a DIA detection method that unbiasedly fragments all peptides in a given mass-range window (167,168).

Protocol: For DDA of the concatenated fractions, ~500 ng of tryptic peptides of each sample were analysed using a Dionex Ultimate 3000 RSLC system coupled to an AB Sciex 6600 Triple TOF mass spectrometer. Peptide samples were inline desalted using an Acclaim PepMap C18 trap column (75 µm x 2 cm; 2 minutes at 5 $\mu L.minute^{-1}$ using 2% ACN/0.2% formic acid). Trapped peptides were gradient eluted and separated on a Waters Acquity CSH C18 NanoEase column (75 µm x 25 cm; 1.7 µm particle size) maintained at 45°C at a flow rate of 0.3 $\mu L.minute^{-1}$ with a linear gradient of 4 – 40% Solvent B over 45 minutes. Precursor (MS) scans were acquired from *m/z* 400 – 1500 (2^+ – 5^+ charge states) using an accumulation time of 200 milliseconds (ms) followed by 40 fragment ion (MS/MS) scans, acquired from *m/z* 100 – 1800 with 20 ms accumulation time each.

The Evotip C18 (Evosep Biosystems, Odense, Denmark) trap columns were employed for SWATH-MS. First, the Evotip C18 trap columns were conditioned before loading the peptides according to the manufacturer's instructions. After the trap columns were conditioned, 20 µL of the sample was added to the Evotips. The Evotips were kept in the storage boxes with a solution of 0.1% formic acid in MS water to keep them submerged for preservation until analysis. For SWATH-MS detection, tryptic peptides (~500 ng) of each sample were analysed using an Evosep One LC System (Evosep Biosystems, Odense, Denmark) coupled to an AB Sciex 6600 Triple TOF mass spectrometer (AB Sciex, Massachusetts, USA). Peptide samples were separated on an Evosep performance column (8 cm x 150 µm) packed with 1.5 µm Dr Maisch C18 beads. The column was maintained at 35°C, and peptides were eluted over 21

minutes (60SPD method) with a gradient of 0 – 35% Solvent B (Solvent A: 0.1% formic acid; Solvent B: 0.1% formic acid; 100% ACN). Precursor scans ranged from *m/z* 400 – 900 using an accumulation time of 100 minutes, and fragment ions were acquired from *m/z* 100 – 1800 with a 15 minutes accumulation time per window across 60 variable-width windows that overlapped by 0.5 Da.

Quality control was performed using HeLa cell-extract digest (Thermo Fisher Scientific, Massachusetts, USA) and conducted as described above.

2.11 Liquid chromatography-mass spectrometry (LC-MS) data processing

A spectral library was built in Spectronaut v16 (Biognosys, Schlieren, Switzerland) using the Pulsar search algorithm. The data files from the high pH RP fractionation (DDA) and SWATH-MS (DIA) of the tissue dataset were used to build the spectral library. Specific trypsin digestion was used for the enzyme setting. A peptide length of 7 – 52 was used, and two missed cleavages per peptide were allowed. Carbamidomethylation was added as a fixed modification, and N-terminal acetylation and methionine oxidation were added as variable modifications. A Swiss-Prot Human FASTA file (https://www.uniprot.org/uniprotkb?facets=reviewed:true&query=(taxonomy_id:9606)) downloaded on 12 June 2021), including common contaminating proteins, was used as the search database. For DIA analysis, the standard identification and quantification settings were used for data processing. However, data filtering was set at a q-value percentile (0.5 fraction) without imputation (precursors meeting the q-value cut-off need to be identified in at least 50% of runs to be included in the analysis). A q-value ≤ 0.01 cut-off was applied at the precursor and protein levels. Quantification was performed at the MS2 level. Label-free cross-run normalisation was employed using a global normalisation strategy.

The following various filters were applied to identify candidate dysregulated proteins: a $q \leq 0.01$, absolute Log_2 fold change (FC) dependent on a retrospective power analysis, present in ≥ 50% of samples for each condition, and ≥ 2 unique peptides.

The generated datasets during this study are available on the ProteomeXchange Consortium (175) repository with the dataset identifier PXD029877.

2.12 Retrospective power analysis

A retrospective power analysis was performed using the MSstats package using R to determine the appropriate FC cut-off (Northeastern University, MSstats v4.4.1) (Bioconductor version: 3.15, R v4.2.0, and R Studio v1.4.1717). The dataProcess function was performed first to normalise the output data from Spectronaut, fragment level peak area of all identified proteins. After that, the groupComparison function was used to compare the protein changes between GBC, GD, and normal for tissue samples and GBC and BBP for plasma samples. Finally, the designSampleSize function was applied to determine the minimum number of replicates required to achieve the desired statistical power. The parameters were a false discovery rate (FDR) = 0.05 and n = the minimum number of samples for each comparison: 5 for GBC vs normal and GD vs normal, 13 for GBC vs GD, and 54 for GBC vs BBP plasma. At power = 0.8, proteins that show fold changes of ≥ 5.2, ≥ 2.775, ≥ 4.65, and ≥ 1.6 are significantly dysregulated for GBC vs normal, GBC vs GD, GD vs normal, and GBC vs BBP plasma comparisons, respectively. A subset analysis was performed on the GBC plasma sample population to determine if GD history versus no GD history affected protein expression. The power analysis indicated that proteins with a fold change of ≥ 2.45 were significantly dysregulated at a power of 0.8 (n = 14).

2.13 Pathway and network analyses

Pathway and network analyses were performed on all the significantly dysregulated proteins identified to determine enriched pathways and interactions. Network analysis and visualisation were performed using Cytoscape v3.8.2 (176). The stringApp v1.7.0 (177) and Reactome v8.0.5 (178) plugins on Cytoscape were used to determine the pathways enriched by the significantly dysregulated proteins and the top 10 significantly ($p < 0.05$) enriched pathways were selected. The proteins were queried with filters including species: *Homo sapiens* as the reference species and zero additional interactors. Within the network, single non-interacting proteins were excluded. The identified dysregulated proteins were also submitted to the PANTHER™ Classification System v17.0 to identify the molecular functions of target proteins (179).

2.14 Statistical Analysis

The demographic and clinical characteristics were analysed using R (v4.0.2 and R Studio v1.4.1717). All data were non-parametric, and a p-value < 0.05 was considered significant. The categorical and continuous data were analysed using Fisher's Exact and Mann-Whitney U tests,

respectively. The Mann-Whitney U test was also performed on all the plasma sample group dysregulated proteins' raw values to determine whether there were any significant changes to protein expression in non-metastatic versus metastatic disease. Venny v2.1.0 was used to visualise uniquely and commonly dysregulated proteins in the tissue and plasma sample groups (GBC/normal tissue, GBC/GD tissue, and GBC/BBP plasma).

A Kruskal-Wallis ANOVA by Ranks with post-hoc analysis was performed using Statistica v13.5 to determine any associations between commonly dysregulated protein expression and sex and age ranges. An unsupervised principal component analysis (PCA) was performed using PAST v4.07b on the commonly dysregulated proteins between GBC/Normal tissue, GBC/GD tissue, and GBC/BBP plasma groups. For the PCA analysis, a correlation matrix was used to explain the maximal variance in samples that would permit the delineation of the disease contexts. Significant loadings for the PCA analysis were determined using the following equation:

$$\frac{1}{\sqrt{n(variables) - 1}}$$

The values generated by the equation were used to determine the positive and negative significant loadings for PCA analysis. Finally, a Spearman's rank correlation test was conducted to determine whether protein expression correlated across the sample types. Hierarchical clustering analysis was performed on all the differentially expressed proteins for each group comparison. The hierarchical clustering analysis generated on Spectronaut was used for this analysis.

3 CHAPTER 3: RESULTS

3.1 Demographics, clinical inflammatory markers and liver function tests are significantly altered in GBC patients

The clinical and demographic information captured for the tissue and plasma patient sample groups were described and compared between the groups (Tables 3.1 and 3.2). Females were the most represented across all patient groups: GBC (69.23% for tissue patients and 53.70% for plasma patients), GD (86.76%), and BBP (82.19%). This was significant in the plasma but not the tissue sample group. Both the GBC tissue and plasma sample group showed a significant increase in the age of the patients. Several of the clinical parameters showed significant differences between the patient groups. White cell count (WCC) and C-reactive protein (CRP) (known inflammatory markers) are significantly elevated in GBC tissue and plasma patients compared to GD and BBP patients. Conversely, haemoglobin (Hb) was significantly decreased in both GBC patient sample groups.

In both GBC patient sample groups, total and direct bilirubin, alkaline phosphatase (ALP), gamma-glutamyl transferase (GGT), and aspartate transaminase (AST) were all significantly elevated, whereas albumin was decreased. Moreover, total protein was significantly decreased in GBC plasma patients compared to BBP patients. The tumour marker CA 19-9 was significantly elevated in GBC tissue but not in plasma patients. However, it is worth noting there was a large percentage of missing CA 19-9 expression data for the tissue and plasma (66.67% missing data for both) sample groups. Additionally, creatinine and alanine transaminase (ALT) were not significantly altered in GBC patients. The most common tumour type presented in both GBC patient sample groups was adenocarcinoma.

Concerning the GBC tissue sample group specifically, there were significantly more GBC patients presenting with jaundice (83.33%) than GD patients (30.77%). Cholangitis status was not significantly altered in GBC tissue patients; however, there were some missing data for GBC (41.67%) and GD (53.85%) patients. While there may be missing patient data for tumour staging and subtype in the GBC plasma patients, Stage IV (23.21%) and adenocarcinoma (33.33%) were the most common stage and tumour subtype, respectively. The TNM staging and GD history for each GBC plasma patient are shown in Table S3.1. A total of 14 (25.9%) patients had a history of GD, 26 (48.2%) patients had no history of GD, and 14 (25.9%) patients had an unknown GD history.

Table 3.1: Clinical and demographic characteristics of patients with gallbladder cancer and gallstone disease tissue sample groups

Parameter	Normal Ranges*	Gallstone Disease (n=13) Median [IQR]	Gallbladder Cancer (n=24**) Median [IQR]	p-value
Sex				
Female, n (%)	-	11 (84.62%)	16 (33.33%)	0.4395[#]
Male, n (%)		2 (15.38%)	8 (66.67%)	
Age (years)	-	43 [36 – 62]	59.5 [55 – 69]	0.0078
White Cell Count (x10^9/L)	3.9 – 12.6	7.48 [5.83 – 9.27]	10.975 [10.10 – 17.06]	<0.001
Haemoglobin (g/dL)	12.4 – 16.7	11.9 [10.5 – 12.6]	10.15 [9.33 – 11.51]	0.0467
C-reactive protein (mg/L)	<5	11 [1 – 142]	129 [61.5 – 192] (n=23)	0.0250
Creatinine (µmol/L)	44 – 80	65 [59 – 67]	71.5 [59 – 87.75]	0.1918
Liver Function Tests				
Total bilirubin (µmol/L)	2 – 21	9 [8 – 21]	114.5 [18.75 – 191.50]	<0.001
Direct bilirubin (µmol/L)	<3.4	4 [3 – 14]	85 [13 – 179]	0.0011
Total protein (g/L)	60 – 80	72 [69 – 75]	70 [60 – 75.25]	0.5240
Albumin (g/L)	35 - 52	38 [34 – 73]	31.5 [28.75 – 34.25]	0.0168
Alkaline phosphatase (U/L)	42 – 98	113 [67 – 172]	615.5 [360.50 – 821.25]	<0.001
Gamma-glutamyl transferase (U/L)	0 – 38	76 [22 – 371]	536 [313 – 740.5]	0.0044
Alanine transaminase (U/L)	7 – 35	49 [24 – 91]	66 [24.75 – 93.75]	0.6331
Aspartate transaminase (U/L)	13 – 35	39 [26 – 69]	76.5 [43.25 – 104.50]	0.0433
Jaundice				
Yes, n (%)	-	4 (30.77%)	20 (83.33%)	0.0030[#]
No, n (%)		9 (69.23%)	4 (16.67%)	
Cholangitis				
Yes, n (%)	-	0 (0%)	1 (4.17%)	0.825[#]
No, n (%)		6 (46.15%)	13 (54.17%)	
Data missing		7 (53.85%)	10 (41.67%)	
Tumour Type				
Adenocarcinoma			17 (70.83%)	
Small neuroendocrine carcinoma	-	-	1 (4.17%)	-
Squamous cell carcinoma			2 (8.33%)	
Unknown			4 (16.67%)	
Carbohydrate antigen 19-9 (U/mL)	0 – 37	1.935	410	0.0081

		[1.593 – 2.195] (n=4)	[35.75 – 1204.75] (n=8)

[#]The Fisher's Exact Test was performed on the sex, jaundice, and cholangitis data. The Mann-Whitney U Test was performed for clinical data and liver function tests.

*Normal physiological ranges obtained from Bio Analytical Research Corporation South Africa (retrieved from https://www.barcsa.co.za/test-directory/test-reference-ranges/, accessed on 21/11/2021, based on female physiological ranges due to majority sample numbers).

**Three patients from the gallbladder cancer group were excluded due to missing clinical data information.

Hyphen (-) indicates the absence of data.

Table 3.2: Clinical and demographic characteristics of patients with gallbladder cancer and benign biliary pathology plasma sample group

Parameter	Normal Ranges*	Benign Biliary Pathology (n=73) Median [IQR]	Gallbladder Cancer (n=54) Median [IQR]	p-value
Sex				
Female, n (%)	-	60 (82.19%)	29 (53.70%)	<0.001[#]
Male, n (%)		13 (17.81%)	25 (46.30%)	
Age (years)	-	39 [31 – 54.5]	57.5 [47.25 – 66.75]	<0.001
White Cell Count (x10^9/L)	3.9 – 12.6	7.71 [6.53 – 9.09] (n=60)	10.47 [7.55 – 13.60] (n=51)	0.0182
Haemoglobin (g/dL)	12.4 – 16.7	12.25 [10.70 – 13.70] (n=60)	10.2 [8.65 – 12.40] (n=51)	<0.001
C-reactive protein (mg/L)	<5	15 [8 – 65] (n=69)	106 [40 – 193] (n=54)	<0.001
Creatinine (µmol/L)	44 – 80	70 [62 – 82] (n=61)	62 [51.5 – 88] (n=51)	0.0531
Liver Function Tests				
Total bilirubin (µmol/L)	2 – 21	31 [11 – 97]	141 [44.50 – 281.75]	<0.001
Direct bilirubin (µmol/L)	<3.4	23 [4 – 81]	122 [46 – 239.5]	<0.001
Total protein (g/L)	60 – 80	73 [68 – 79] (n=55)	63.5 [58 – 72.25] (n=48)	<0.001
Albumin (g/L)	35 - 52	39 [35.5 – 42.0] (n=55)	30 [26 – 34.5] (n=48)	<0.001
Alkaline phosphatase (U/L)	42 – 98	192 [112 – 393]	516.5 [281.25 – 908.75]	<0.001
Gamma-glutamyl transferase (U/L)	0 – 38	166 [80 – 419]	435.5 [238.5 – 860]	<0.001
Alanine transaminase (U/L)	7 – 35	45 [17 – 102] (n=55)	57.5 [35.75 – 103.25] (n=48)	0.2138
Aspartate transaminase (U/L)	13 – 35	53 [28 – 90] (n=55)	104.5 [44.75 – 185.50] (n=48)	0.0060
Tumour Staging				
I	-	-	8 (14.29%)	-
II			5 (8.93%)	
III			7 (12.50%)	
IV			13 (23.21%)	
Unknown				

			23 (38.89%)	
Tumour Type				
Adenocarcinoma			18 (33.33%)	
Small neuroendocrine carcinoma	-	-	0 (0%)	-
Squamous cell carcinoma			1 (1.85%)	
Unknown			35 (64.81%)	
Carbohydrate antigen 19-9 (U/mL)	0 – 37	113 [70.5 – 184.5] (n=3)	364 [25 – 1479] (n=18)	0.6509

[#]The Fisher's Exact Test was performed on the sex data. The Mann-Whitney U Test was performed for clinical data and liver function tests.

*Normal physiological ranges obtained from Bio Analytical Research Corporation South Africa (retrieved from https://www.barcsa.co.za/test-directory/test-reference-ranges/, accessed on 21/11/2021, based on female physiological ranges due to majority sample numbers).

Hyphen (-) indicates the absence of data.

3.2 Custom-built project-specific SWATH spectral library

The project-specific spectral library was built in Spectronaut v16 using the Pulsar search algorithm. The default library identification and quantifications were used except for the data filtering, as described in Section 2.11. The spectral library was matched to the Swiss-Prot Human FASTA File. The project-specific spectral library matched and returned precursors, peptides, proteins, and protein groups. For the tissue sample group, 87 341 precursors, 65 725 modified peptides, 62 204 peptides, and 57 435 proteolytic peptides were identified in this library. These precursors and peptides matched 6362 proteins in 6204 protein groups. The plasma library showed 5944 precursors, 4011 modified peptides, 3471 peptides, and 2803 proteolytic peptides for the plasma sample group, which matched to 346 proteins and 313 protein groups.

3.3 Quality control of liquid chromatography-mass spectrometry setup

Commercial HeLa cell-extract digest was used as quality control to verify the performance of the LC-MS setup during the complete dataset acquisition. Although protein extraction demonstrates the most variation in proteomics due to human error and the variations in the amount and subtypes of tissue present, it is essential to observe any instrument drift and deficiencies. Three injections of the HeLa digest were performed at the beginning, middle, and end of the tissue and plasma datasets to determine any time-dependent instrument variation. The metrics investigated were the total ion chromatogram (TIC), run identifications, and coefficient of variation (CV). The TIC overlay plot for the HeLa quality control for the tissue and plasma datasets shows substantial overlap for each HeLa lysate injection, indicating

consistent ion detection and instrument sensitivity (Figures 3.1A and 3.2A). This shows consistent instrument functioning with minimal experimental drift throughout the tissue and plasma runs. Additionally, the run identifications show the number of precursors identified per HeLa digest injection (Figures 3.1B and 3.2B). The number of precursors identified across the HeLa injections is consistent and demonstrates minimal instrument drift and variation. The coefficient of variation (CV) is a commonly used metric in proteomics for comparing agreement between replicates. As expected for quality control, the CV of the HeLa cell digest was 16.6% and 16.5% at the precursor level for the tissue and plasma dataset, which is below the established threshold of 20% for this specific LC-MS setup (Figures 3.1C and 3.2C).

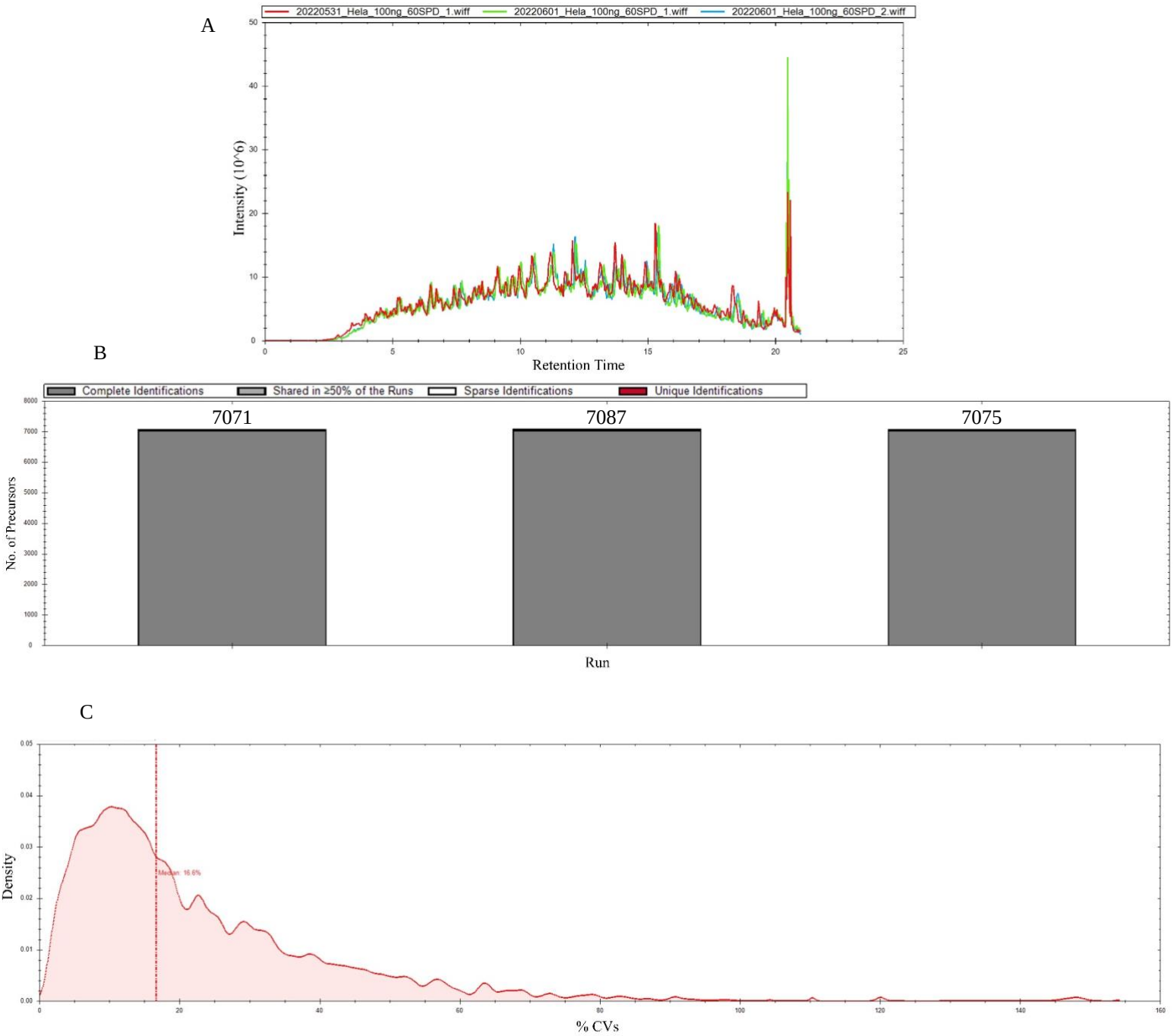


Figure 3.1: Tissue HeLa cell lysate quality control. (A) The total ion chromatogram (TIC) overlap plot for each injection of HeLa cell lysate as quality control during the tissue sample experimentation. (B) The total run identifications indicate that all identified precursors are present across all HeLa cell lysate injections with no unique identifications. (C) The median coefficient of variance (CV) indicates minimal variance around the median, with a CV of 16.6%.

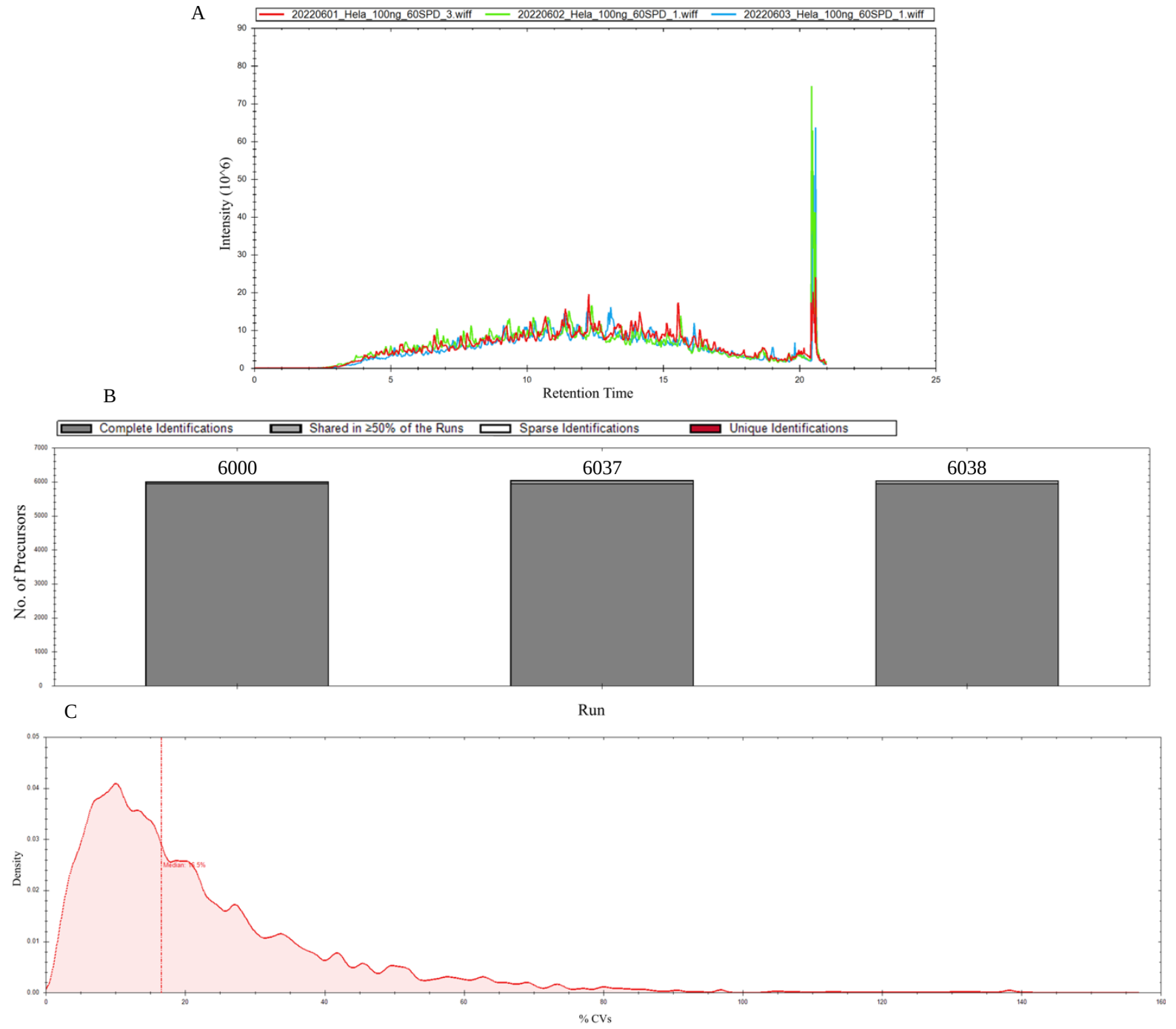


Figure 3.2: HeLa Cell lysate quality control for the plasma dataset. (A) The total ion chromatogram overlap plot for each injection of HeLa cell lysate as quality control during the plasma sample experimentation. (B) The total run identifications indicate that all identified precursors are present across all HeLa cell lysate injections with no unique identifications. (C) The median coefficient of variance (CV) indicates minimal variance around the median, with a CV of 16.5%.

3.4 Sequential window of all theoretical mass spectra analysis identified differentially abundant proteins in gallbladder cancer patients compared to normal, gallstone disease, and benign biliary pathology controls

3.4.1 Relatively abundant proteins detected in the tissue sample group

Protein abundance for the tissue and plasma samples was measured using SWATH-MS. Relative protein abundance was measured between the tissue groups using an unpaired Student's t-test.

For the GBC/Normal comparison (24 GBC tissues and 5 normal tissues), SWATH-MS identified 36 606 peptides matching 2955 protein groups across all runs. There were 12 527 peptides matching 1834 protein groups identified in 50% of the total runs. A total of 445 proteins were detected in full profiles and 3038 proteins in sparse profiles. Sparse profiles refer to all proteins that were detected, regardless of the number of samples the proteins were detected in. The details of precursors, peptides, and protein groups for each GBC and normal individual are described in Table S3.2. There was an average of 35% missed cleavages, 35.9% total data completeness, 4.5 and 4.4 data points per peak for GBC and normal patient groups, respectively, and median peak widths of approximately 0.11 min for both GBC and normal groups. The retrospective power analysis indicated that proteins with a Log_2 FC $\geq$ 5.2 for the GBC/Normal comparison were considered significant due to the limiting sample numbers of the normal sample group. Additionally, 62 proteins were significantly dysregulated in the GBC/Normal comparison (Table S3.3); 38 were upregulated, and 24 were downregulated. The top upregulated protein was periostin (POSTN), and the top downregulated protein was c-jun-amino-terminal kinase-interacting protein 4 (JIP4).

In the GBC/GD comparison (24 GBC tissues and 13 GD tissues), there were 36 830 peptides matching 2951 protein groups across all runs. A total of 11 261 peptides that matched 1674 protein groups were identified in 50% of the runs. Altogether, 300 proteins were detected in full profiles and 3035 in sparse profiles. The details of precursors, peptides, and protein groups for each of the GBC and GD patients are described in Table S3.4. The missed cleavage average was 35%, with 33.2% data completeness. There were 4.5 and 4.4 data points per peak for GBC and GD, respectively, and the median peak widths were approximately 0.11 min for both groups. Proteins with a Log_2 FC $\geq$ 2.775 were considered significant for the GBC/GD comparison, as revealed by the retrospective power analysis. There were 194 dysregulated proteins (Table S3.5); 88 were upregulated, and 106 were downregulated. The top upregulated

and downregulated proteins in this group are aldo-keto family 1 member C2 (AK1C2) and desmin (DESM), respectively.

For the GD/Normal comparison (13 GD tissues and 5 normal tissues), 24 783 peptides matched to 2161 protein groups were identified. In 50% of the runs, there were 9546 peptides matching 1408 protein groups. Altogether, 446 proteins were detected in full profiles and 2236 in sparse profiles. The details of precursors, peptides, and protein groups for each of the GD and normal patients are described in Table S3.6. The average number of missed cleavages was 35%, 39.9% data completeness, 4.4 data points per peak for both GD and normal, and 0.11 min median peak width for both GD and normal groups. Proteins with a Log_2 FC $\geq$ 4.65 were considered significant for the GD/Normal comparison, as shown by the retrospective power analysis. There were 105 dysregulated proteins; 42 were upregulated, and 63 were downregulated (Table S3.7). The proteins with the most significant increased fold change were histones class H2A (H2A1B, H2A3, and H2A1C), and Aflatoxin B1 aldehyde reductase member 3 (ARK73) was the top downregulated protein.

There are common dysregulated proteins across the tissue group comparisons. There were 32 proteins commonly dysregulated between GBC/Normal and GBC/GD, ten proteins common between GBC/Normal and GD/Normal, and 25 proteins common between GD/Normal and GBC/GD (Table S3.8). Six proteins were also commonly dysregulated across all three tissue comparisons (Figure 3.3).

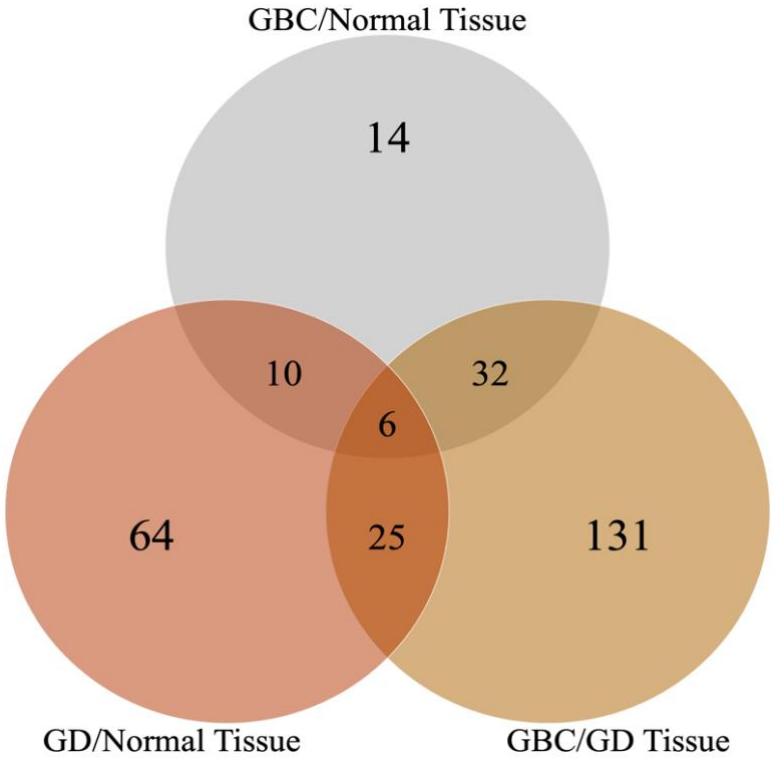
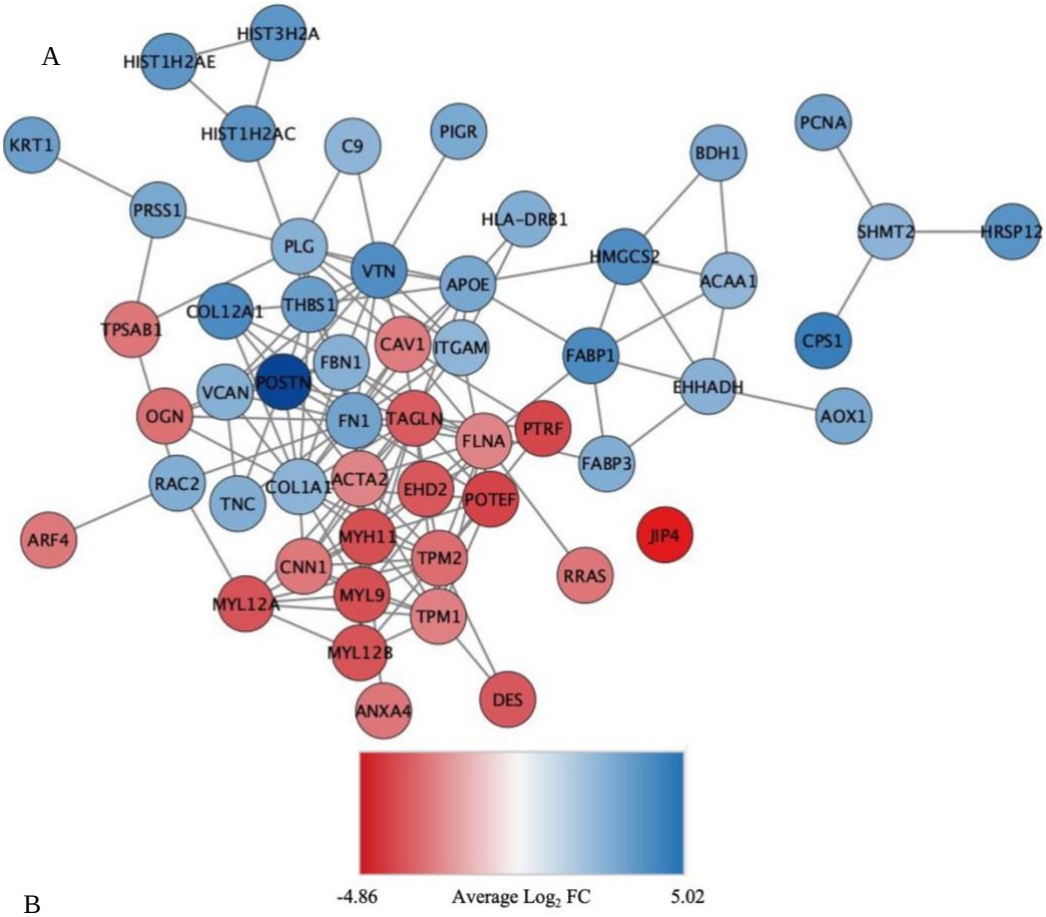


Figure 3.3: Venn diagram for common proteins dysregulated across tissue sample group comparisons. There are 62, 194, and 105 dysregulated proteins across GBC/Normal (grey), GBC/GD (brown), and GD/Normal (orange) groups, respectively. There are commonly dysregulated proteins, as indicated by the overlaps between the groups. The Venn diagram was generated using Venny (v2.1.0).

3.4.2 Pathway enrichment by the dysregulated proteins in the tissue sample group

Network and pathway analyses were performed on the differentially expressed proteins to determine their interactions and enriched biological pathways (Figures 3.4, 3.5, and 3.6). The top downregulated pathway for GBC/Normal is smooth muscle contraction. The top upregulated pathways are integrin cell surface interactions, extracellular matrix (ECM) organisation, and metabolism (Figure 3.4B). Like the GBC/Normal comparison, the top downregulated pathways for GBC/GD are smooth muscle contraction and cell-ECM interactions (Figure 3.5B). Both the upregulated and downregulated proteins in the GBC/GD group showed enrichment in metabolism and ECM organisation pathways. For the GD/Normal group, neutrophil degranulation was shown to be the top upregulated pathway and metabolism the top downregulated pathway. Conversely, the metabolism pathways were upregulated in the GBC/Normal and GBC/GD comparisons (Figure 3.6B). The innate immune system pathway was enriched by proteins identified to be both upregulated and downregulated.

A

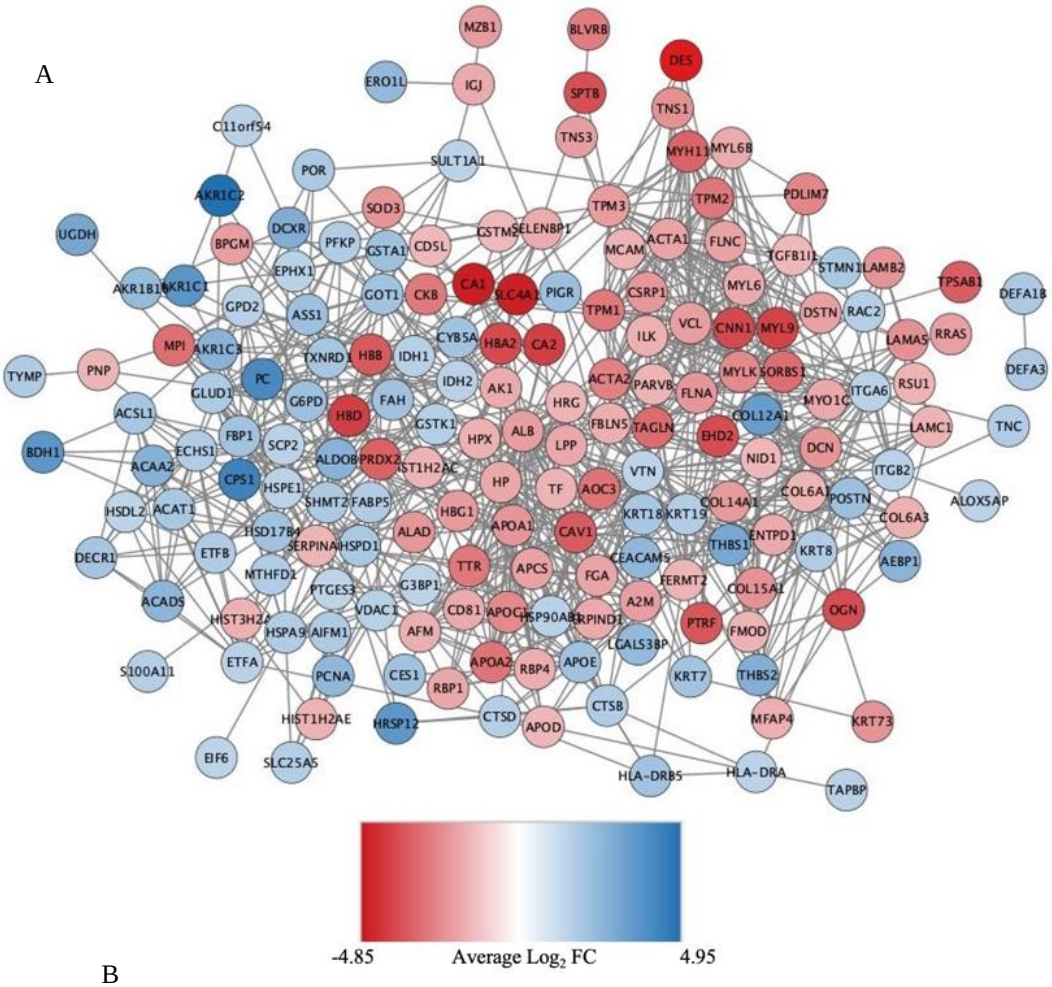


B

No. of Genes	Pathway	FDR Value	Genes Involved
7	Smooth Muscle Contraction	1.98E-07	MYL12A|MYL9|TPM1|TPM2|MYH11|ACTA2|MYL12B
12	Extracellular matrix organization	2.52E-07	COL1A1|VTN|THBS1|VCAN|TNC|PRSS1|PLG|COL12A1|FBN1|TPSAB1|FN1|ITGAM
7	Integrin cell surface interactions	1.14E-05	COL1A1|VTN|THBS1|TNC|FBN1|FN1|ITGAM
5	Syndecan interactions	2.73E-05	COL1A1|VTN|THBS1|TNC|FN1
4	RHO GTPases activate PAKs	3.40E-04	MYL9|FLNA|MYH11|MYL12B
6	Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	6.70E-04	APOE|VCAN|TNC|PLG|FBN1|FN1
5	ECM proteoglycans	0.001	COL1A1|VTN|VCAN|TNC|FN1
19	Metabolism	0.0027	EHHADH|APOE|HRSP12|OGN|VCAN|CA2|FABP1|DCXR|PRSS1|ACAA1|SHMT2|CAV1|HMGCS2|FABP3|AOX1|AKR1C2|BDH1|TYMP|CPS1
5	Post-translational protein phosphorylation	0.0043	APOE|VCAN|TNC|FBN1|FN1
9	Haemostasis	0.0118	COL1A1|RAC2|THBS1|EHD2|PLG|CAV1|FN1|FLNA|ITGAM

Figure 3.4: Network and pathway analysis of dysregulated proteins in GBC tissues compared to normal tissues. (A) Network analysis indicating protein interactions between dysregulated proteins discovered in GBC compared to normal. The colour change is determined by the average Log_2 fold change of the protein; blue indicates upregulated, and red indicates downregulated. (B) Dysregulated pathways of GBC compared to normal. The top upregulated pathways are ECM organisation, integrin cell surface interactions, and metabolism. The top downregulated pathway is smooth muscle contraction.

A

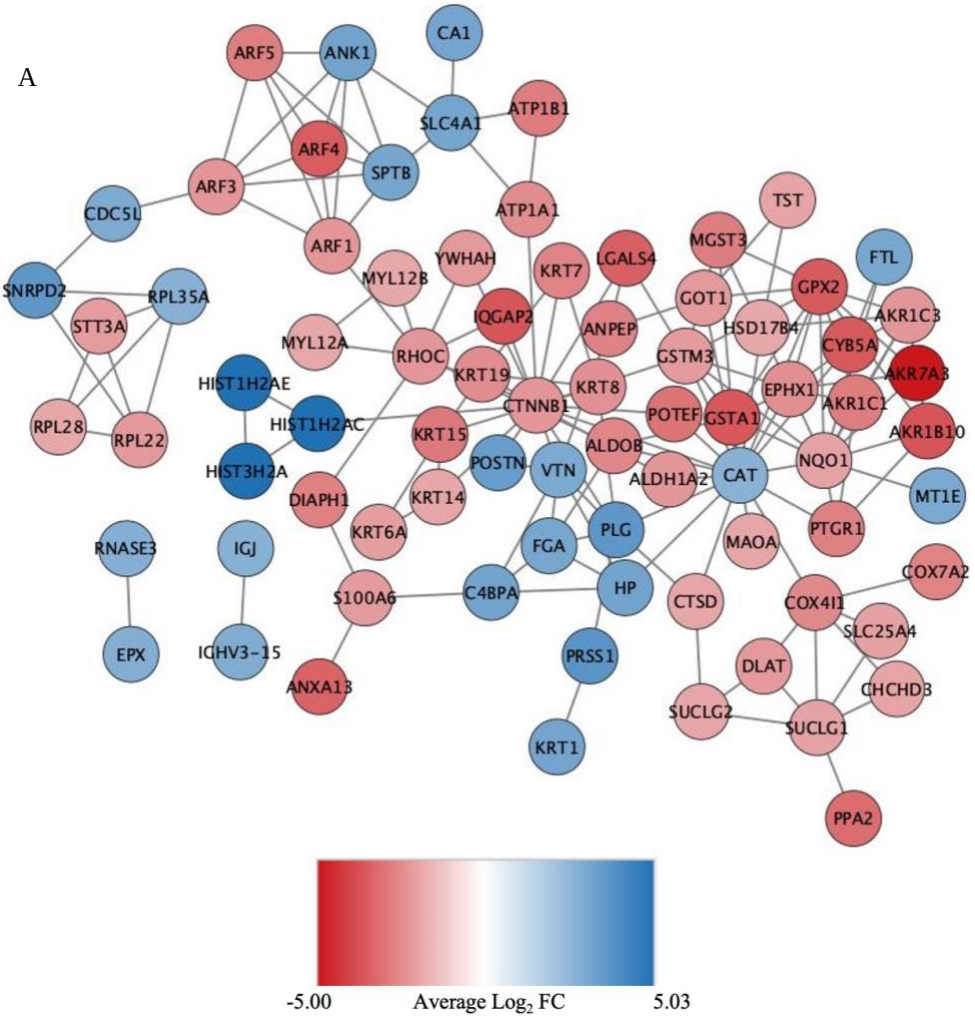


B

No. of Genes	Pathway	FDR Value	Genes Involved
70	Metabolism	1.62E-17	DCN|MTHFD1|DECR1|RBP1|APOA1|TTR|GSTM2|ACADS|APOE|HRSP12|OGN|BLVRB|VDAC1|ACAT1|GLUD1|ACAA2|CA2|ALB|FABP5|CKB|DCXR|GPD2|AOC3|UGDH|IDH2|SHMT2|GSTA1|CAV1|CYB5A|SERPINA6|ETFB|FMOD|AKR1B10|CES1|PNP|APOA2|ECHS1|GOT1|ENTPD1|RBP4|SCP2|HSP90AB1|ASS1|AK1|ALDOB|SLC3A2|AKR1C3|AKR1C2|AKR1C1|PFKP|BDH1|BPGM|G6PD|PC|SULT1A1|TYMP|FAH|ALAD|IDH1|CPS1|FBP1|POR|HSD17B4|ACSL1|CA1|TXNRD1|ETFA|ALOX5AP|EPHX1|PTGES3
24	Extracellular matrix organization	1.34E-11	DCN|VTN|CTSD|TTR|LAMA5|LAMC1|THBS1|NID1|TNC|COL6A3|COL14A1|FGA|A2M|COL12A1|TPSAB1|FBLN5|CTSB|FMOD|COL6A1|COL15A1|MFAP4|ITGB2|ITGA6|LAMB2
11	Smooth Muscle Contraction	8.37E-10	VCL|MYL9|TPM1|MYLK|TPM3|SORBS1|TPM2|MYH11|ACTA2|MYL6|MYL6B
6	Cell-extracellular matrix interactions	2.19E-05	FLNC|FERMT2|FLNA|RSU1|ILK|PARVB
9	ECM proteoglycans	2.76E-05	DCN|VTN|LAMA5|LAMC1|TNC|COL6A3|FMOD|COL6A1|LAMB2
10	Post-translational protein phosphorylation	3.90E-05	SERPIND1|APOA1|APOE|LAMC1|TNC|ALB|FGA|APOA2|TF|LAMB2
8	Non-integrin membrane-ECM interactions	4.59E-05	VTN|TTR|LAMA5|LAMC1|THBS1|TNC|ITGA6|LAMB2
8	Integrin cell surface interactions	3.60E-04	VTN|THBS1|TNC|COL6A3|FGA|COL6A1|ITGB2|ITGA6
6	Bile acid and bile salt metabolism	6.20E-04	ALB|SCP2|AKR1C3|AKR1C2|AKR1C1|HSD17B4
10	Fatty acid metabolism	0.001	DECR1|ACADS|ACAA2|ECHS1|SCP2|AKR1C3|HSD17B4|ACSL1|ALOX5AP|PTGES3

Figure 3.5: Network and pathway analysis of dysregulated proteins in GBC compared to GD tissue samples. (A) Network analysis showing the interactions of the dysregulated proteins in GBC compared to GD. The colour change is determined by the average Log_2 fold change of the protein; blue indicates upregulated, and red indicates downregulated. (B) The pathways in which the dysregulated proteins are involved. The top downregulated pathways are the smooth muscle contraction and cell-ECM interactions. The metabolism and ECM pathways are enriched by both upregulated and downregulated proteins.

A

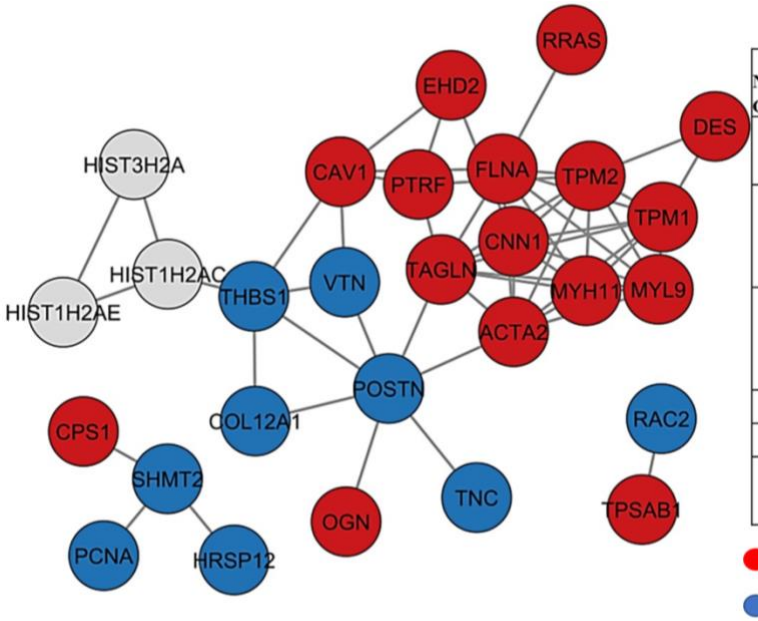


B

No. of Genes	Pathway	FDR value	Genes
32	Metabolism	3.22E-07	GSTM3|ARF3|DLAT|PRSS1|NQO1|GSTA1|MAOA|CYB5A|PPA2|RPL22|AKR1B10|CES1|AKR7A3|MGST3|GOT1|ALDOB|AKR1C3|AKR1C1|GPX2|BDH1|SUCLG1|PTGR1|TST|RPL35A|SUCLG2|HSD17B4|CA1|ARF1|BPNT1|RPL28|COX4I1|EPHX1
18	Developmental Biology	0.0014	KRT14|MYL12A|KRT1|KRT15|ANK1|RHOC|HIST1H2AE|KRT7|CTNNB1|RPL22|KRT19|HIST1H2AC|KRT6A|RPL35A|KRT8|SPTB|RPL28|MYL12B
7	Formation of the cornified envelope	0.0016	KRT14|KRT1|KRT15|KRT7|KRT19|KRT6A|KRT8
8	Biological oxidations	0.0034	GSTM3|GSTA1|MAOA|CES1|AKR7A3|MGST3|BPNT1|EPHX1
9	Asparagine N-linked glycosylation	0.0034	ARF5|TFG|ARF3|ANK1|ARF4|GFPT1|STT3A|ARF1|SPTB
8	RHO GTPase Effectors	0.0099	YWHAH|IQGAP2|RHOC|HIST1H2AE|CTNNB1|HIST1H2AC|DIAPH1|MYL12B
10	Neutrophil degranulation	0.0099	EPX|CTSD|CAT|KRT1|IQGAP2|ANPEP|RNASE3|HP|FTL|DIAPH1
14	Innate Immune System	0.0258	EPX|VTN|CTSD|CAT|KRT1|IQGAP2|ANPEP|RNASE3|FGA|CTNNB1|HP|C4BPA|FTL|DIAPH1
11	Vesicle-mediated transport	0.0258	ARF5|TFG|YWHAH|IGJ|ARF3|ANK1|ARF4|HP|FTL|ARF1|SPTB
3	Synthesis of bile acids and bile salts via 7alpha-hydroxycholesterol	0.029	AKR1C3|AKR1C1|HSD17B4
12	Infectious disease	0.0377	SLC25A4|HIST1H2AE|CTNNB1|RPL22|HIST3H2A|ATP1B1|HIST1H2AC|STT3A|RPL35A|ARF1|ATP1A1|RPL28

Figure 3.6: Network and pathway analysis of dysregulated proteins in GD compared to normal tissue samples. (A) Network analysis of the dysregulated proteins indicating interactions between the proteins in GD compared to normal. The colour change is determined by the average Log_2 fold change of the protein; blue indicates upregulated, and red indicates downregulated. (B) Enriched pathways by the dysregulated proteins; neutrophil degranulation was the top upregulated pathway, metabolism was the top downregulated pathway, and the innate immune system pathway was enriched by both upregulated and downregulated proteins.

Network and pathway analysis was then performed on the 38 proteins commonly dysregulated in GBC compared to GD and normal controls. This may determine pathway aberrations specific to GBC irrespective of precursor condition (Figure 3.7). The smooth muscle contraction pathway was shown to be downregulated in GBC tissues. Moreover, gene ontology enrichment analysis for the dysregulated proteins for the tissue groups determined that the molecular functions of most of the dysregulated proteins were related to binding and catalytic activity (Figure 3.8).



No. of Genes	Pathway	FDR Value	Genes Involved
5	Smooth Muscle Contraction	2.68E-05	MYL9|TPM1|TPM2|MYH 11|ACTA2
4	RUNX1 Regulates Genes Involved in Megakaryocyte Differentiation and Platelet Function	0.0044	THBS1|MYL9|HIST1H2A E|HIST1H2AC
6	RHO GTPase Effectors	0.0045	RAC2|MYL9|HIST1H2AE |FLNA|HIST1H2AC|MYH 11
3	Sema4D In Semaphorin Signalling	0.0046	RRAS|MYL9|MYH11
3	Syndecan Interactions	0.0056	VTN|THBS1|TNC
3	Metalloprotease Dubs	0.0073	HIST1H2AE|HIST3H2A| HIST1H2AC

Downregulated in both groups

Upregulated in both groups

Upregulated in GBC/Normal but downregulated in GBC/GD

Figure 3.7: Pathway and network analyses for the commonly dysregulated proteins between GBC/Normal and GBC/GD tissue groups. Smooth muscle contraction was the most downregulated pathway. Red indicates downregulated proteins, blue indicates upregulated proteins, and grey indicates upregulated in GBC/Normal but downregulated in GBC/GD.

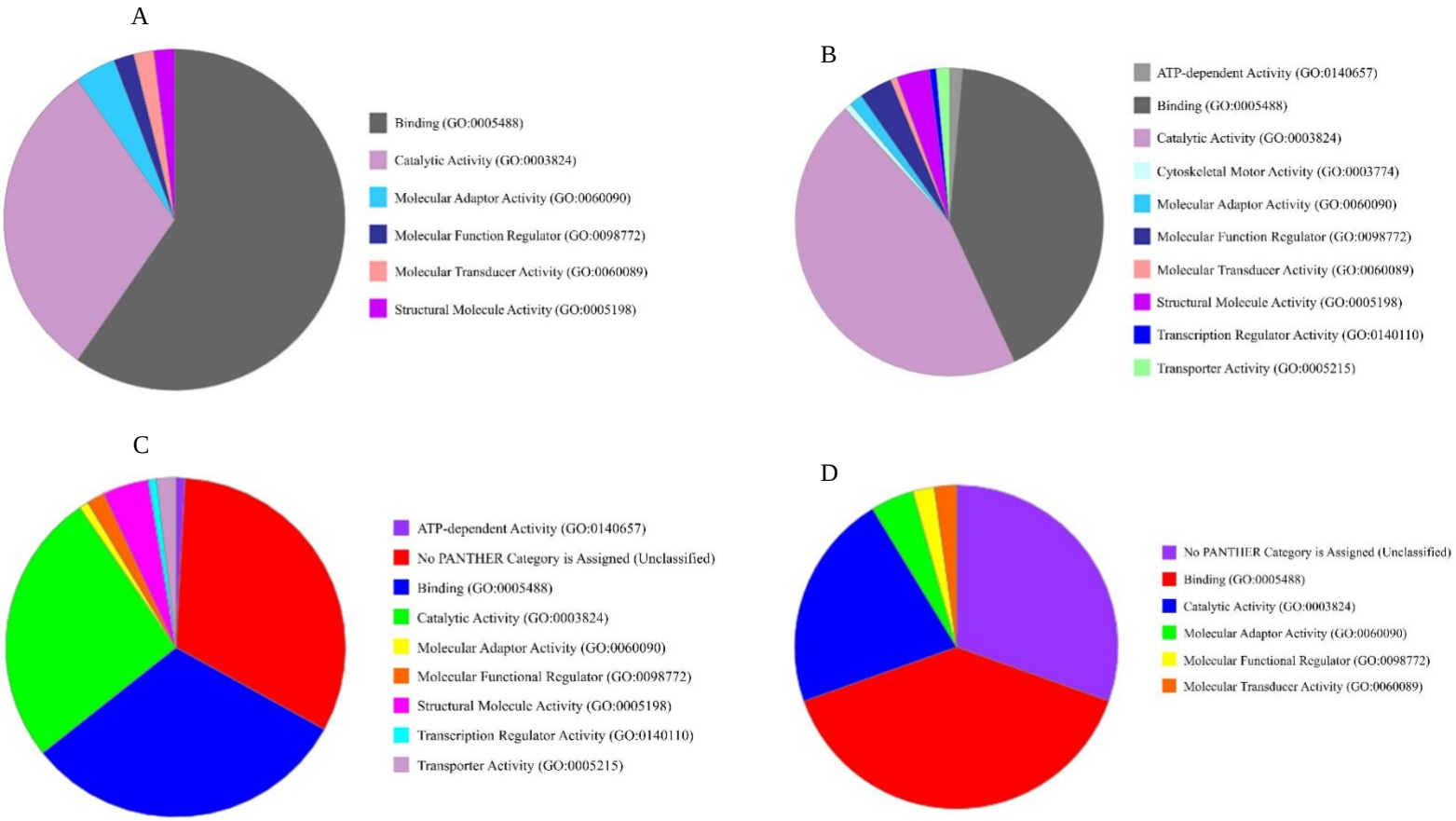


Figure 3.8: Annotated molecular functions for dysregulated proteins in tissue groups. Pie charts representing the molecular functions of dysregulated proteins. (A) GBC tumours compared to normal tissues. (B) GBC tumours compared to GD tissues (C) GD compared to normal tissues. (D) Thirty-eight proteins commonly dysregulated in GBC compared to GD and normal tissues. The annotation was conducted using PANTHER v17.0.

3.4.3 Altered proteins identified in the plasma sample sample group

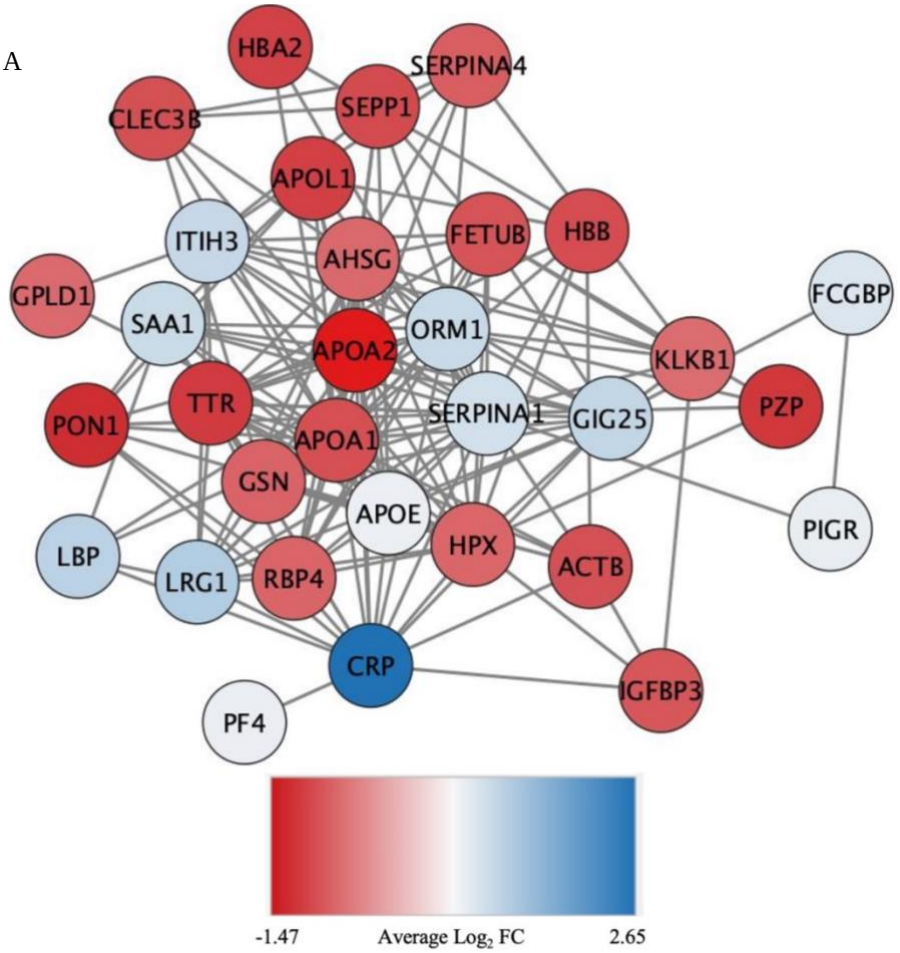
Plasma protein levels were also investigated in a separate sample group of GBC and BBP patients. SWATH-MS identified 3310 peptides which matched 260 protein groups in the custom-built spectral library. There were 2577 peptides matching 226 protein groups identified in 50% of the total runs. A total of 149 proteins were detected in full profiles and 292 in sparse profiles. The details of precursors, peptides, and protein groups for each of the GBC and BBP plasma patients are described in Table S3.9. There was an average of 27.1% missed cleavages and 68.6% total data completeness. There were 5.0 data points per peak and median peak widths of approximately 0.11 min for both the GBC and BBP groups. The retrospective power analysis indicated that proteins exhibiting a Log_2 FC $\geq$ 1.6 were considered significantly differentially expressed. Consequently, 33 proteins were identified as significantly dysregulated (Table S3.10). Of these 33 significantly dysregulated proteins, 12 were upregulated, and 21 were downregulated. The top upregulated protein identified in the plasma sample group was CRP, and the top downregulated protein was apolipoprotein A-II (APOA2).

A subset analysis was conducted to determine proteins significantly altered in GBC patients with a history of GD to those without. At a Log_2 FC $\geq$ 2.45, the levels of two proteins, S100A8 (FC = -2.651) and S100A9 (FC = -1.397), were identified to be significantly lowered in GBC patients with a history of GD. Furthermore, proteins whose levels were associated with tumour staging were investigated, and GBC patients were categorised as non-metastatic (Stage I, II, and III) and metastatic (Stage IV). Two proteins were elevated in non-metastatic patients compared to metastatic patients: apolipoprotein E and inter-alpha-trypsin inhibitory heavy chain H3 (ITIH3) (Table S3.11).

3.4.4 Pathway enrichment by the dysregulated proteins in the plasma sample group

Network and pathway analyses were performed to elucidate interactions and pathway involvement of the differentially expressed proteins identified in the plasma patient sample group (Figure 3.9). The dysregulated pathways include platelet degranulation, haemostasis, innate immune system, and neutrophil degranulation (Figure 3.9B). There were similar dysregulated pathways in the plasma and tissue sample group. These pathways are enriched by proteins that are both upregulated and downregulated.

A



B

No. of Genes	Pathway	FDR Value	Genes Involved
10	Platelet degranulation	3.50E-11	APOA1|ORM1|PF4|CLEC3B|GIG25|AHSG|ITIH3|SERPINA1|SEPP1|SERPINA4
13	Haemostasis	6.25E-09	APOA1|ORM1|KLKB1|PF4|CLEC3B|HBB|ACTB|GIG25|AHSG|ITIH3|SERPINA1|SEPP1|SERPINA4
6	Binding and Uptake of Ligands by Scavenger Receptors	6.26E-08	APOA1|APOE|HPX|APOL1|HBB|SAA1
14	Innate Immune System	1.46E-07	LBP|TTR|CRP|ORM1|LRG1|HBB|PGLYRP2|PIGR|ACTB|GSN|GIG25|SAA1|AHSG|SERPINA1
7	Post-translational protein phosphorylation	2.19E-07	APOA1|APOE|APOL1|APOA2|IGFBP3|AHSG|SERPINA1
7	Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	5.09E-07	APOA1|APOE|APOL1|APOA2|IGFBP3|AHSG|SERPINA1
9	Neutrophil degranulation	1.26E-05	TTR|ORM1|LRG1|HBB|PIGR|GSN|GIG25|AHSG|SERPINA1
12	Metabolism of proteins	0.005	GPLD1|APOA1|TTR|APOE|APOL1|ACTB|APOA2|GSN|IGFBP3|SAA1|AHSG|SERPINA1

Figure 3.9: Pathway and network analysis of the dysregulated proteins in the plasma group. (A) Network analysis of the dysregulated proteins indicating interactions between the proteins in GBC compared to BBP plasma. The colour change is determined by the average Log_2 fold change of the protein; blue indicates upregulated, and red indicates downregulated. (B) Enriched pathways by the dysregulated proteins; platelet degranulation, haemostasis, innate immune system, and neutrophil degranulation pathways were enriched by both upregulated and downregulated proteins.

Additionally, gene ontology enrichment analysis for the molecular functions of the altered plasma proteins determined that most of the proteins were involved in binding and catalytic activity. The other molecular functions of these proteins were molecular function regulators and molecular transducer activity (Figure 3.10).

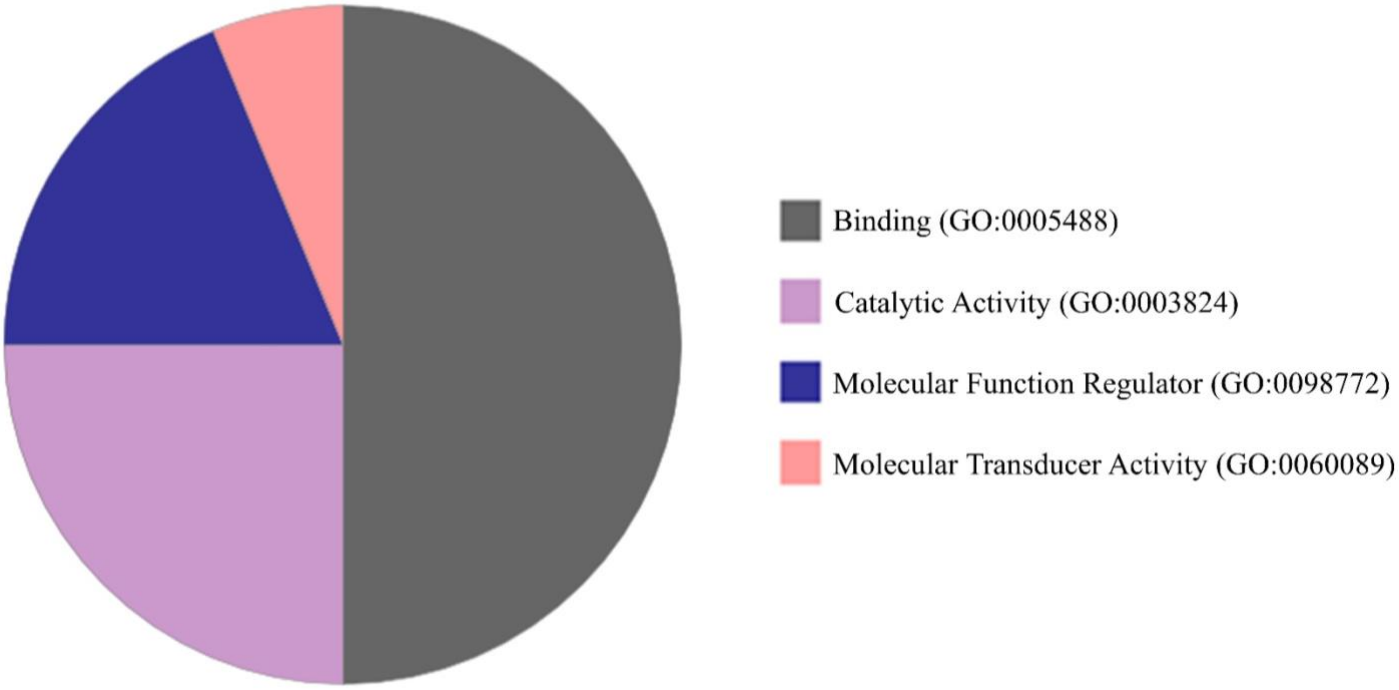
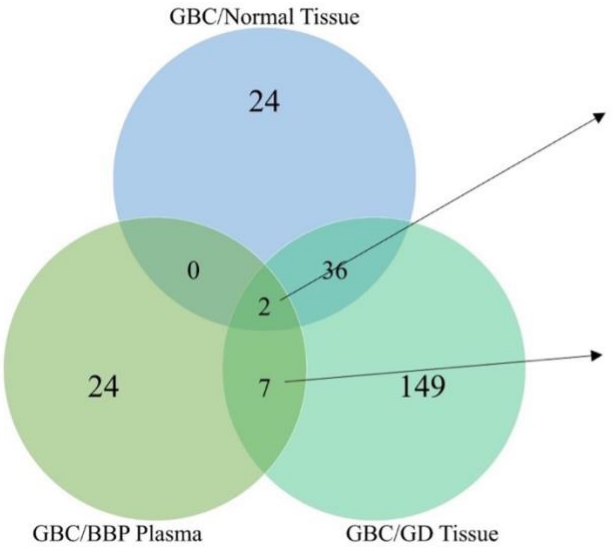


Figure 3.10: Annotated molecular functions for plasma dysregulated proteins. Pie charts representing the molecular functions of dysregulated proteins in the GBC compared to the BBP plasma sample group. The annotation was done using PANTHER v17.0.

3.5 Commonly dysregulated proteins in tissue and plasma sample sample groups

Additional analysis was performed to determine whether any dysregulated proteins in the tissue sample group were present in the plasma sample group (Figure 3.11). There were 24, 149, and 24 dysregulated proteins unique to GBC/Normal, GBC/GD, and GBC/BBP sample group comparisons, respectively. There were 36 common proteins between GBC/Normal and GBC/GD tissue comparison groups. Seven proteins were common between GBC/GD tissue and GBC/BBP plasma: apolipoprotein A-I (APOA1), apolipoprotein A-II (APOA2), retinol-binding protein 4 (RET4), transthyretin (TTR), hemopexin (HEMO), haemoglobin subunit alpha (HBA), haemoglobin subunit beta (HBB). Two proteins were common among all three groups; these are polymeric immunoglobulin receptor (PIGR) and apolipoprotein E (APOE) (Table S3.12). These nine proteins are are subsequently referred to as ‘commonly dysregulated

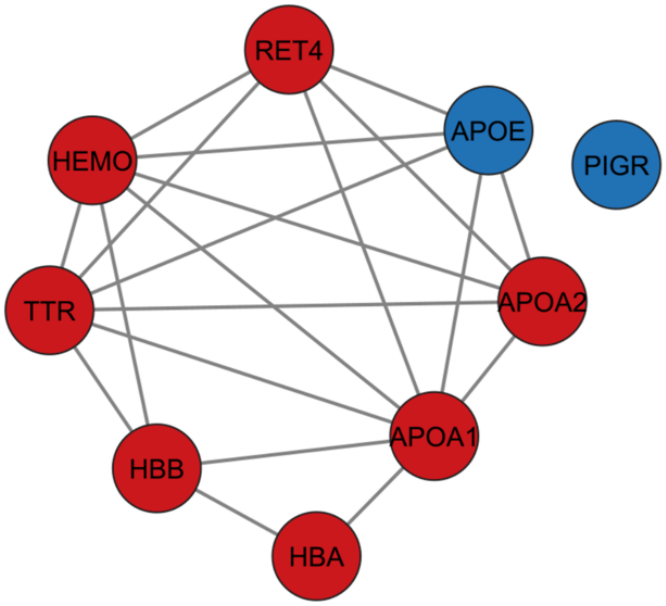
proteins (CDPs)'. Interestingly, the CDPs are dysregulated in the same direction (either up- or downregulated) across the sample types (tissue and plasma groups). Additionally, a Kruskal-Wallis H test demonstrated that age and sex did not significantly affect the CDP expression for any sample group. The Log_2 quantities for each of the CDPs for each patient are shown in Figures S3.1 and S3.2.



Potential Protein Markers (PPMs)	Dysregulation
Polymeric immunoglobulin receptor (PIGR)	Upregulated
Apolipoprotein E (APOE)	Upregulated
Apolipoprotein A-I (APOA1)	Downregulated
Apolipoprotein A-II (APOA2)	Downregulated
Retinol-binding protein 4 (RET4)	Downregulated
Transthyretin (TTR)	Downregulated
Hemopexin (HEMO)	Downregulated
Haemoglobin subunit alpha (HBA)	Downregulated
Haemoglobin subunit beta (HBB)	Downregulated

Figure 3.11: Venn diagram for common dysregulated proteins among sample group comparisons. There are 62, 194, and 33 identified dysregulated proteins across GBC/Normal tissue (blue), GBC/GD tissue (turquoise), and GBC/BBP (green) plasma groups. There are commonly dysregulated proteins, as indicated by the overlaps. The Venn diagram was generated using Venny (v2.1.0).

Network analysis was then performed on the CDPs to determine interactions and pathway enrichment (Figure 3.12). The network analysis indicated various interactions between the CDPs; only PIGR was not involved in the network and has not been shown to interact with the other CDPs. The pathways associated with metabolism were demonstrated to be the most enriched by the CDPs.



No. of Genes	Pathway	FDR value	Genes Involved
5	Retinoid metabolism and transport	2.16E-08	APOA1|TTR|APOE|APOA2 |RET4
4	Binding and Uptake of Ligands by Scavenger Receptors	1.65E-06	APOA1|APOE|HEMO|HBB
3	Chylomicron assembly and remodeling	6.94E-06	APOA1|APOE|APOA2
3	Scavenging of heme from plasma	1.02E-05	APOA1|HEMO|HBB
2	HDL remodeling	0.0021	APOA1|APOE
3	Post-translational protein phosphorylation	0.0022	APOA1|APOE|APOA2
2	Retinoid cycle disease events	0.0025	TTR|RET4
3	Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	0.0026	APOA1|APOE|APOA2
2	Scavenging by Class A Receptors	0.0041	APOA1|APOE
2	The canonical retinoid cycle in rods (twilight vision)	0.0061	TTR|RET4
2	Plasma lipoprotein clearance	0.0105	APOA1|APOE
4	Transport of small molecules	0.0176	APOA1|APOE|HBB|APOA2
2	Amyloid fibre formation	0.0487	APOA1|TTR

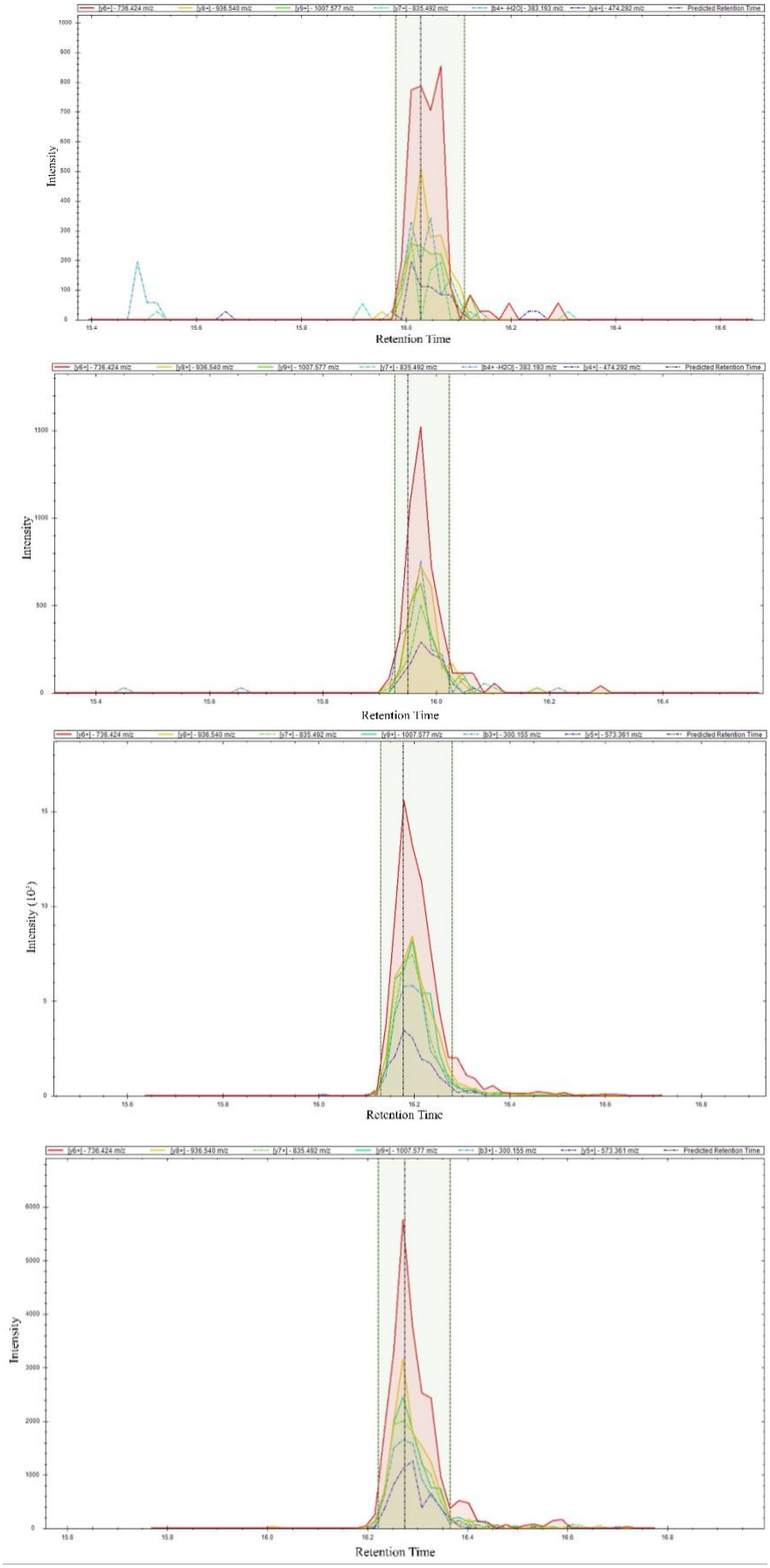
Figure 3.12: Network and pathway analysis for the identified Commonly Dysregulated Proteins between the tissue and plasma groups. Network analysis shows the intricate relationships between the CDPs. Blue and red indicate up- and downregulated proteins in GBC tissue and plasma compared to GD tissue and BBP plasma, respectively.

3.6 Quality control of the commonly dysregulated proteins

Quality control of the CDPs was done to determine the suitability of these proteins to distinguish GBC from benign pathologies correctly. The number of samples in which CDPs were detected and their average sequence coverage was investigated (Table S3.13). APOA1, APOA2, TTR, HEMO, HBA, and HBB were detected in 100% of samples across the tissue and plasma sample populations. RET4 was detected in 100% of samples in the GBC/BBP comparison and 97.5% in the GBC/GD comparison. APOA1 had 73% sequence coverage for GBC/GD and 72% for GBC/BBP. APOA2 had a coverage of 64% for GBC/GD and 42% for GBC/BBP. RET4 had 41% and 59% coverage for GBC/GD and GBC/BBP, respectively. The coverage for TTR was 69% for GBC/GD and GBC/BBP. HEMO had a coverage of 51% and 71% for GBC/GD and GBC/BBP, respectively. The coverage for HBA was 96% for GBC/GD and 65% for GBC/BBP. The coverage for HBB was 96% for both GBC/GD and GBC/BBP.

The two proteins common (PIGR and APOE) between all three comparisons (GBC/Normal, GBC/GD, and GBC/BBP) were also evaluated. PIGR was detected in 81.2%, 75%, and 93.7% of samples for GBC/Normal, GBC/GD, and GBC/BBP, respectively. The sequence coverage was 35%, 36%, and 21% for GBC/Normal, GBC/GD, and GBC/BBP, respectively. APOE was detected in 100% of samples for all three comparisons and had a coverage of 64% for GBC/Normal and GBC/GD and 59% for GBC/BBP (Table S3.13).

The MS2 fragment ion chromatograms for the nine CDPs were investigated (Figure 3.13 and S3.3 – S3.10). The CDPs fragment ions indicated overlapped retention times, which suggests correct matching to the MS1 spectra. This shows the fragments are eluting in the predicted retention time. Most of the precursor and fragment ion peaks show a high signal-to-noise ratio, which indicates high-quality raw data and high-quality identification of the CDPs. Also, thin, resolved peaks of the fragment ions indicate good chromatographic retention and elution. Moreover, the CV for each CDP were investigated in the plasma sample group (Table S3.13). Large CVs for many of the CDPs indicate a wide range of detected abundance values. In general, the CV for each CDP was lower in the BBP group compared to GBC except for HBA, HBB and PIGR.



GBC Tissue

GD Tissue

GBC Plasma

BBP Plasma

Figure 3.13: Representative MS2 _DLATVYVDVLK_.2 precursor chromatogram overview for APOA1 in the tissue and plasma sample group. The MS2 precursor chromatogram overview for representative samples for GBC tissue, GD tissue, GBC plasma, and BBP plasma sample group.

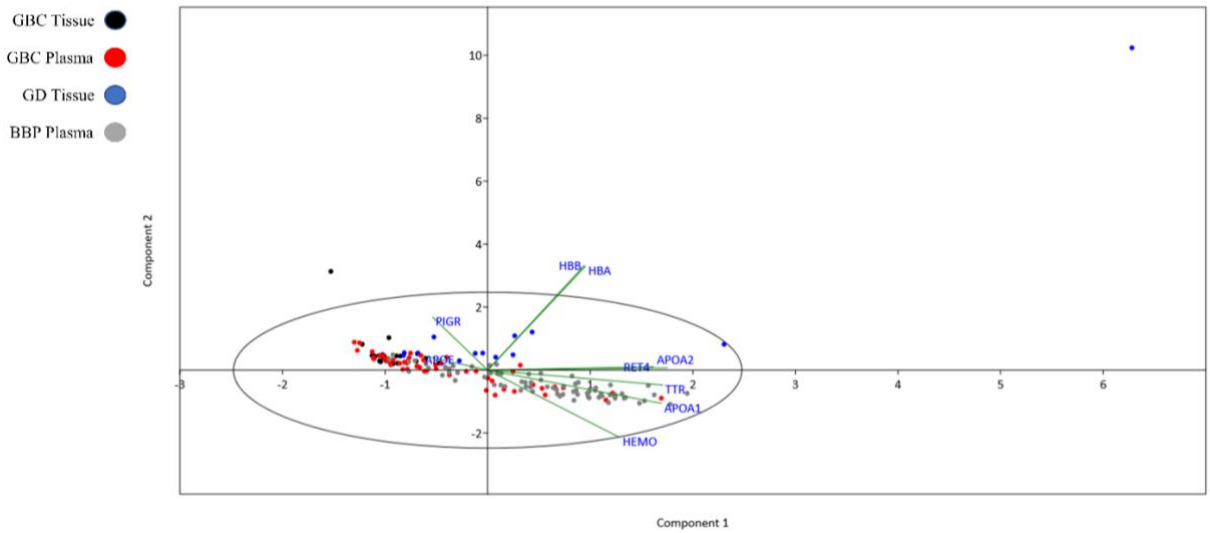
3.7 Hierarchical clustering, principal component analysis, and Spearman's rank correlation test of the commonly dysregulated proteins

Hierarchical clustering was performed on all the quantified proteins for GBC/Normal, GBC/GD, and GBC/BBP (Figure S3.11). The tissue groups showed distinct clustering by condition (tumour and control). However, the conditions for the plasma dataset showed some overlap.

Principal component analysis (PCA) was performed to determine the ability of the CDPs to distinguish between GBC and controls and reduce data dimensionality. Visually, the GBC tissue group (black) showed considerable overlap with GBC plasma (red) samples following the reduction of dimensionality (Figure 3.14A). GBC tissue indicated tight clustering, and GBC plasma was more dispersed. Specifically, the generated principal component 1 (PC1) and PC2 accounted for 67.37% of the CDP variation. Both GBC tissue and plasma clustering were influenced by PC1 (44.76% variation) with strong positive contributions from APOA1, APOA2, RET4, and TTR. The relationship between GBC tissue and plasma was delineated from GD and BBP in the vertical PC2 axis (22.61% variation) by positive influences from HBB and HBA with a negative influence from HEMO (Figure 3.14B).

The Spearman's rank correlation test was performed to determine any correlated pattern of expression of the CDPs in the context of proteins, that is, protein expression only, regardless of sample source (Figure 3.14C) with the specific Rho and p-values noted in Figure S3.12. APOA1, APOA2, RET4, TTR, and HEMO were all significantly positively correlated with each other and are closely related in an intricate network (Figure 3.12). Additionally, these proteins share similar pathways, such as retinoid metabolism. Although these seven proteins common between the tissue and plasma groups are all downregulated, HBB and HBA are negatively correlated with the other proteins, such as APOA1. Conversely, while PIGR and APOE are commonly dysregulated across all groups, these two proteins are not significantly correlated with any of the other CDPs.

A



B

Positive Loading
Negative Loading

Loadings	PC 1	PC 2
APOA1	0.42856	-0.18997
APOA2	0.44345	0.01261
RET4	0.4093	0.018916
TTR	0.43151	-0.08514
HEMO	0.32619	-0.3854
HBB	0.23835	0.59783
HBA	0.24084	0.59286
PIGR	-0.13562	0.30266
APOE	-0.16234	0.084351
Eigenvalue	4.02841	2.03502
Percentage Variance	44.76%	22.611%
Total Variance	67.371%	

C

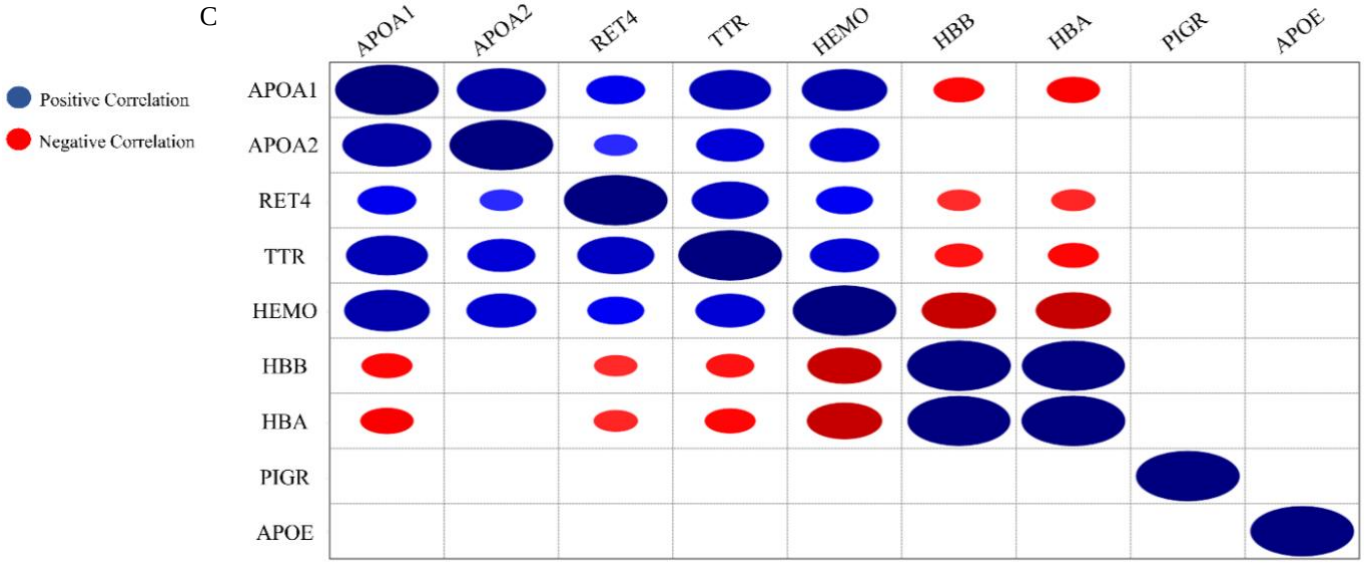


Figure 3.14: Principal component analysis and Spearman's rank correlation test for commonly dysregulated proteins (CDPs). (A) The scatterplot indicates PC1 vs PC2 of the CDPs. Black dots indicate GBC tissue patients, red dots indicate GBC plasma patients, blue dots indicate GD patients and grey dots indicate BBP patients. (B) The loadings for the CDPs for each PC. Green highlighted values are significantly positive, and red highlighted values are significantly negative loadings. (C) The Spearman's rank correlation test indicates significant positive (blue ellipses) and negative (red ellipses) correlation between the CDPs. The presence of ellipses indicate a significant correlation, and the size represents the strength of the correlation.

3.8 Summary of results

Clinical and demographic data for the tissue and plasma sample groups were recorded during sample collection. This included standardised blood tests routinely performed on GBC patients at CHBAH. There were significant alterations in age and sex (GBC plasma patients compared to BBP patients). Inflammatory markers (WCC and CRP) and LFTs were also significantly elevated in GBC compared to GD and BBP patients. GBC tissue patients presented with jaundice more frequently than GD patients. CA 19-9 was elevated in some GBC patients, and adenocarcinoma was identified as the prevalent histological type of GBC. A project-specific spectral library was built using Spectronaut v16 and the Pulsar search algorithm to detect relatively abundant proteins in GBC compared to normal, GD, and BBP patients. There were 62 (GBC/Normal), 194 (GBC/GD), 105 (GD/Normal), and 33 (GBC/BBP plasma) relatively dysregulated proteins observed in the tissue and plasma sample groups.

Various pathways were dysregulated in the tissue and plasma sample groups, especially those involved in metabolism, ECM organisation, integrin and non-integrin cell surface interactions, cell-ECM interactions, platelet and neutrophil degranulation, innate immune system, haemostasis, and smooth muscle contraction. Additionally, 38 proteins in the tissue sample group were commonly dysregulated in GBC/Normal and GBC/GD comparisons. The top pathway enriched by these 38 proteins was smooth muscle contraction. Gene ontology indicates that the molecular functions of the dysregulated proteins in the tissue and plasma sample groups are involved in binding and catalytic activity. Furthermore, nine proteins (APOA1, APOA2, RET4, TTR, HEMO, HBA, HBB, PIGR, and APOE) were discovered to be commonly dysregulated in GBC compared to normal, GD, and BBP sample groups (tissue and plasma sample groups). These proteins were termed commonly dysregulated proteins (CDPs). The CDPs interacted closely in an interaction network; however, PIGR was observed not to interact directly with any of the other CDPs.

Hierarchical cluster analysis indicated complete separation between GBC and GD and normal controls (tissue group). However, the plasma group had various overlaps between GBC and BBP groups. The PCA analysis showed that the CDPs clustered the GBC tissue and plasma groups separately from the GD and BPP groups, with minimal overlap between the disease states. The Spearman's rank test showed that APOA1, APOA2, RET4, TTR, and HEMO were significantly correlated, and HBB/HBA was inversely correlated with APOA1, RET4, TTR, and HEMO.

4 CHAPTER 4: DISCUSSION

Gallbladder cancer (GBC) has a poor prognosis with an ever-growing incidence and mortality, particularly in South Africa (50,180). In most cases, GBC is detected at advanced stages leading to a poorer prognosis and survival rate (181). This calls for a better understanding of the disease progression and identifying suitable biomarkers. Notably, the variations observed in incidence and mortality across different geographical regions reinforce the possible involvement of molecular, clinicopathological, and environmental factors. Various studies have been published which have determined the molecular profiles highlighting potential mechanisms of progression and biomarkers for the disease. However, this information is lacking in African patients (17,182). This study has demonstrated potential novel molecular mechanisms and markers linked to GBC in an African sample group by conducting proteomics profiling of tissue and plasma samples.

4.1 Elevated clinical parameters in gallbladder cancer patients

Clinical and demographic information for each patient was captured and compared to determine any differences in various parameters such as inflammatory markers, liver function tests (LFTs), age, and sex. This study showed that both GBC patient sample group presented with a significantly advanced age compared to control patients (GD and BBP). This finding is similar to others that observed GBC patients present at an advanced age (183,184). In general, more females presented with diseased gallbladders more frequently than men. It is well-documented that females develop a diseased gallbladder at a rate of two to six-fold more frequently than males (9).

The results showed that most of the clinical parameters for GD and BPP groups were within normal parameters. In contrast, GBC patients had significantly altered parameters (Tables 3.1 and 3.2). For example, clinical inflammatory markers (WCC and CRP) were significantly elevated in GBC patients compared to the control groups, even though GD and BBP are both inflammatory diseases, further suggesting an increased inflammatory environment in GBC patients. WCC is a non-specific inflammatory marker that is often increased in acute or chronic infections (185). CRP is primarily expressed in hepatocytes, and its expression is regulated by interleukin-6 (IL6), a well-known pro-inflammatory cytokine (186).

CRP as a marker increases in inflammatory diseases, including GD and BBP, and in tumour progression. Gallbladder inflammation is a known risk factor for GBC onset and links to an

ongoing inflammatory environment to induce tumour progression. (187–191). Haemoglobin (Hb) was significantly reduced in GBC patients, suggesting an anaemic condition. Reduced Hb could be attributed to several reasons, including excessive bleeding induced by the tumour or inflammation, which can lead to iron deficiency (192). The tumour marker CA 19-9 was elevated in GBC patients, although this was not significant in the plasma sample group, likely due to many missing values (Table 3.2). However, on its own, CA 19-9 is unsuitable for detecting GBC and should instead be used with other known markers, such as CEA, to increase sensitivity and specificity (124).

Most LFTs, including bilirubin, ALP, GGT, and AST, were elevated in GBC patients. The tumour invades the liver during GBC progression and, consequently, alters the functioning of these enzymes produced by the liver. Bilirubin, produced and excreted by the liver via haem degradation, can accumulate due to biliary obstruction, which causes jaundice (193,194). This study also showed that jaundice is significantly more prevalent in GBC patients (Table 3.1). GGT functions to metabolise glutathione, an antioxidant found in plants, fungi, and animals. Glutathione can prevent cellular damage by removing reactive oxygen species and free radicals, and serves as an important antioxidant and facilitates important detoxification of the liver (195,196). Increased GGT levels will reduce the amount of glutathione present, inhibiting the function of glutathione to minimise cellular damage. Therefore, elevated GGT has been associated with high cancer risk and poor prognosis due to the reduced protective role of glutathione (196–198).

Conversely, albumin and total protein are reduced in GBC patients. Albumin is synthesised by hepatocytes and is rapidly secreted into the blood as only a small amount is stored in the liver. Albumin is a plasma oncotic pressure modulator (osmotic pressure induced by proteins) and transporter of endogenous and exogenous ligands (199). Hypoalbuminaemia indicates malnutrition, liver dysfunction, and systemic inflammation (197). ALT, another liver protein, shows no significant changes in GBC patients compared to GD and BBP patients. ALT is predominantly produced by the liver, and elevated levels indicate hepatic and biliary injury. In GBC, GD, and BBP patients, ALT is elevated, indicating hepatic injury in both GBC and control sample group despite no significant differences (200). Although, in the present study, several of these clinical parameters were significantly altered in GBC patients, these alterations are not specific to the disease and are known to occur in diverse conditions (201).

4.2 Pathways dysregulated in the tissue sample group

This study identified dysregulated proteins in the tumours of a sample group of South African GBC patients compared to GD and normal gallbladder (ie, non-diseased gallbladder) tissues. There were 62 proteins dysregulated in the GBC/Normal comparison, 194 proteins in the GBC/GD comparison, and 105 proteins in the GD/Normal comparison (Tables S3.3, S3.5, and S3.7). Many of these proteins are uniquely dysregulated in the comparisons, but 38 proteins were commonly dysregulated in GBC/GD and GBC/Normal comparisons. Pathway analysis identified pathways enriched by the dysregulated proteins in the GBC/Normal, GBC/GD, and GD/Normal comparisons.

4.2.1 Enriched signalling pathways by the dysregulated proteins in the GBC/Normal comparison

Comparing GBC to normal tissue samples, the top pathways enriched by the upregulated proteins in the GBC/Normal comparison were ECM organisation, integrin cell surface interactions, syndecan interactions, and metabolism. The smooth muscle contraction pathway was downregulated (Figure ***3.4***).

The ECM is a complex non-cellular structure involved in biological processes such as tissue homeostasis and development regulation. Normal functioning of the ECM includes cellular differentiation, adhesion, and cellular communication. Significantly, the ECM interacts with surrounding cells and induces signal transduction cascades which mediate these processes (202,203). The ECM mainly comprises proteoglycans and fibrous proteins. However, even though the ECM regulates normal tissue functioning, it is also responsible for regulating several critical hallmarks of cancer, such as proliferation, evasion of the immune response, and cellular death. It is, therefore, crucial in promoting tumour progression and metastasis (169,203,204). Re-organisation of the ECM has been documented as playing an important role in tumourigenesis. This process involves hundreds of proteins having an altered expression, which changes the overall biochemical and biophysical properties of the ECM. It is a highly regulated process involved in tissue homeostasis and wound healing but becomes dysregulated in inflammatory diseases and cancer (Figure 4.1A) (203). Tissue-associated aberrations in ECM re-organisation have been highlighted as crucial for metastasis (203). One protein, periostin (POSTN), which was found to be the most upregulated in this study, plays a significant role in ECM organisation (205). POSTN aids in forming collagen cross-links which provides stability in collagen-rich connective tissues (205). POSTN has been identified to be

dysregulated in inflammation and malignancy, and is highly expressed in cancers of the prostate, lung, and pancreas (205). Although POSTN is involved in ECM organisation, it was not observed as part of the ECM organisation network in this study (Figure 3.4).

Some of the dysregulated proteins involved in the ECM pathway and enriched in this study are integrin alpha-M (ITAM), type I and type XII collagen (CO1A1 and COCA1), vitronectin (VTNC), and fibronectin (FINC). These are known ECM proteins involved in binding to the cell surface to induce downstream signal transduction responses (206). Integrins bind ECM macromolecules, such as collagen and laminin, to induce a signal transduction cascade. Collagen is the predominant macromolecule of the ECM, with 28 types and CO1A1 and COCA1 were upregulated in this study. Increased expression of these collagen types has been associated with cancer metastasis, poor survival, and proliferation (207,208). VTNC and FINC are glycoproteins with adhesive properties that link cells and the ECM by binding integrins to facilitate cellular adhesion (209,210).

ECM reorganisation involves several key pathways, including integrin cell surface interaction. This pathway was identified as significantly dysregulated in GBC tumours. Integrins are glycoproteins which are heterodimeric receptors for the ECM. The main functions of these proteins are adhesion and cellular motility (211). Some of the ECM components that integrins bind to are laminins, collagen, VTNC, and FINC which are the main scaffolding molecules of the ECM. In carcinogenesis, integrins regulate tumourigenic processes such as survival, proliferation, and migration via integrin activation after binding to ECM components (212,213). Also, integrins, and their regulated processes such as survival, proliferation, and migration, are associated with cancer stem-like properties, metastasis, and drug resistance (214). The close interactions between integrins and ECM could allow for the remodelling of the ECM to promote a metastatic environment.

Another crucial component of the ECM is the syndecans. In this study, pathways linked to syndecan interactions were upregulated in GBC tumours compared to normal tissues. Syndecans are transmembrane proteins comprising heparan sulphate proteoglycans. They interact with various components of the ECM, which contain heparan-binding domains, to induce signal transductions in response to ECM changes (215). Syndecans regulate internal to external cellular signal transduction, and vice versa, to effect changes in the ECM components that directly link to the ECM organisation pathway (216).

The metabolism pathway was another pathway dysregulated in GBC patients; this is a well-known, fundamental process in cancer. The Warburg effect describes cancer upregulating glucose usage to produce more lactic acid than normal tissues (217). Various substrates, such as lipids and glucose, are metabolised for energy production. The altered metabolic pathways allow tumours to have a higher proliferation rate, even with limited nutrients and oxygen contributing to tumourigenesis and malignancy (218). Metabolic pathways are under intense scrutiny to identify potential therapeutic intervention strategies using large-scale profiling methods (219–221). The gallbladder's function is to store and concentrate bile to digest fats into fatty acids. Proteins which enrich the metabolic pathway have been shown to play a role in lipid transport, uptake, and oxidation. A vital protein which enriches this pathway is APOE, a protein identified as "CDP", and plays a key role in lipid metabolism by acting as a lipid transporter. The specific function of APOE is discussed in detail in Section 4.4. Other important dysregulated proteins involved in this pathway are caveolin-1 (CAV1), peroxisomal bifunctional enzyme (ECHP), and fatty acid-binding protein (FABPL). These proteins are involved in either the direct oxidation or the transport of lipids for metabolism (222–224). CAV1 is essential in lipid metabolism as it is a fatty acid- and cholesterol-binding protein, alluding to a lipid transport role for metabolism (222). ECHP is also involved in lipid metabolism by catalysing the oxidation of long-chain fatty acids (223). Furthermore, FABPL regulates lipoprotein-mediated uptake of cholesterol and binds free fatty acids (224). The functions of these proteins support the possible involvement of lipid metabolism in GBC tumourigenesis.

Fatty acids, triglycerides, and cholesterol are metabolised in the lipid metabolism pathway (225). Various lipids, such as glycolipids, phospholipids, and cholesterol, represent significant components of membranes (226). In the cancer tumour microenvironment, nutrient availability is variable, so lipids become an alternative energy source for the proliferation, migration, and metastasis of cancer cells (227). Lipid metabolism may not occur solely for energy production but can generate messenger molecules after hydrolysing membrane lipids (phospholipids) by phospholipases. These messenger molecules activate other enzymes, such as Ras and PI3K. These enzymes activate the MAPK and PI3K/Akt pathways, which are known to increase GBC progression (227). In addition, cholesterols can covalently bind to proteins Hedgehog and Smoothened, which allows for activation of the Hedgehog pathway. The Hedgehog pathway has also been described to be aberrant in GBC which promotes the notion of lipid metabolism being imperative for GBC advancement (Figure 4.1B) (77,228).

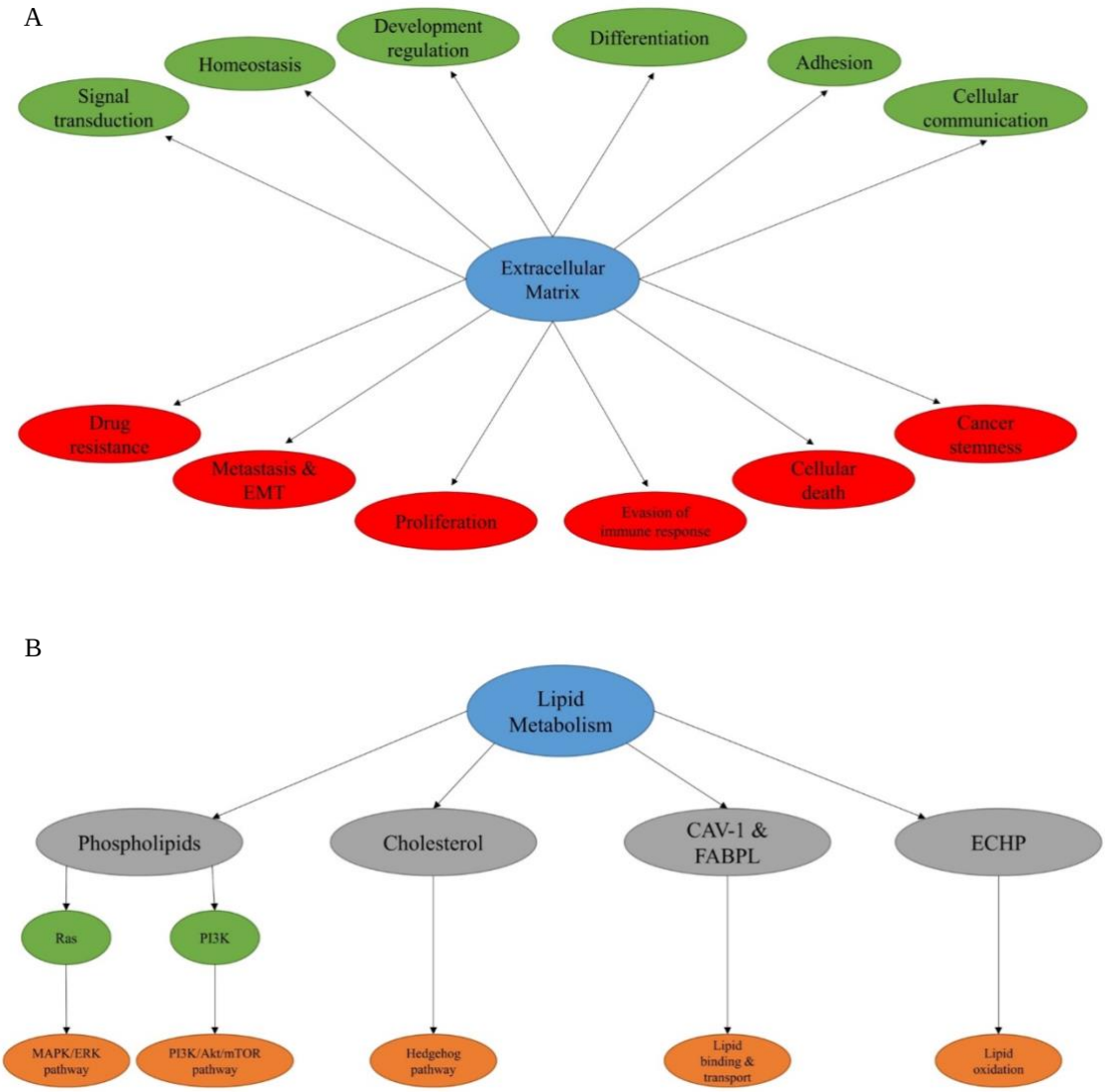


Figure 4.1: Summary of the roles of the ECM and lipid metabolism pathways. (A) The ECM microenvironment under normal (green) and tumourigenic (red) conditions. (B) The lipid metabolism pathway constitutes several other pathways and targets. Grey ellipses are effector biomolecules involved in lipid metabolism. Green ellipses are proteins which become active after phospholipid hydrolysation. Orange ellipses indicate the downstream involvement.

The smooth muscle contraction pathway was the top dysregulated pathway and was enriched by several proteins downregulated in GBC tumours. It is important to note that the downregulation of the smooth muscle contraction pathway may be due to the microenvironment of GBC tumours which consists of predominantly stroma and epithelial cells, compared to normal tissues composed of muscular and epithelial cells (229,230). Moreover, smooth muscle contractility regulates the motility and emptying of the gallbladder (231). The motility of the gallbladder depends on the periodic contractions of the muscular layer, which allows bile to empty from the gallbladder. Decreased gallbladder motility is associated with an increased risk of gallstone disease (10). This process is regulated by the

hormone cholecystokinin (CCK), which binds to the membrane receptor (CCKR), inducing gallbladder contraction and emptying (23,232,233). A pilot study examined the expression of CCKRs in GBC, GD, and normal gallbladder tissues and observed a decrease in tumours, although not significant (234). Although CCK and CCKR regulate the contractility of gallbladder smooth muscle, neither of these proteins has been observed to be dysregulated in the GBC/Normal comparison. Additionally, oxidative stress resulting from gallbladder contraction dysfunction can damage CCKRs, lead to altered lipid metabolism and induce inflammation. This study found alteration of lipid metabolism and an inflammatory state to be characteristics of GBC patients (Tables 3.1 and 3.2; Figure 3.4) (235–237). Moreover, the individual proteins involved in the smooth muscle contraction pathway are actin (ACTA), myosin (MYH11, ML12A, ML12B, and MYL9), and tropomyosin (TPM1 and TPM2). These proteins have been identified to play a role in the contractility and motility of muscle (238).

4.2.2 Pathways enriched by proteins dysregulated in gallbladder cancer patients compared to gallstone disease patients

The signalling pathways enriched by the dysregulated proteins in the GBC/GD comparison were similar to that of the GBC/Normal comparison. However, there were differences in some of the key enriching proteins.

In the GBC/Normal comparison, there were 19 proteins involved in the metabolism pathway while the GBC/GD comparison had 70 proteins (Figure 3.4B and Figure 3.5B). Like the GBC/Normal comparison, lipid metabolism was dysregulated in the GBC/GD comparison. Lipid metabolism in a cancer context was described in Section 4.2.1. The top upregulated protein in this comparison was AK1C2 (aldo-keto reductase family 1 member C2), which metabolises critical steroid hormones such as dihydrotestosterone (239). AK1C2 is also involved in the metabolism of bile acid and salts, consolidating this protein's role in the lipid metabolism pathway. Other aldo-keto reductase proteins involved in lipid metabolism in the GBC/GD comparison are AK1C1 and AK1C3 (both upregulated). Other proteins which enrich the lipid metabolism pathway are fatty acid-binding protein 5 (FABP5) and 3-ketoacyl-CoA thiolase (THIM). FABP5 functions to transport long-chain fatty acids, and THIM catalyses the oxidation of fatty acids (240,241).

Similar to the GBC/Normal group, the ECM organisation pathway was also enriched in the GBC/GD group with some overlapping enriched proteins such as TSP1, VTNC, and TENA. Unique ECM-related pathways to the GBC/GD group are non-integrin membrane-ECM and

laminin interactions. The proteins involved in the non-integrin interactions are laminins (LAMA5, LAMB2, and LAMC1) and vitronectin (VTNC). Laminins are glycoproteins which bind to collagen to increase basement membrane stability and cellular adhesion (242). LAMA5, LAMB2, and LAMC1 are also involved in the laminin interaction pathway. Like the GBC/Normal group, the smooth muscle contraction pathway was also downregulated in the GBC/GD group (Figure 3.5B). Interestingly, the top downregulated protein in the GBC/GD comparison was DESM (Desmin), a prominent intermediate filament of cardiac, smooth, and skeletal muscles (243). It maintains muscle integrity and is involved in muscle contraction. However, this protein is not observed to be directly involved in the smooth muscle contraction pathway.

4.2.3 Enriched signalling pathways by the dysregulated proteins in the GD/Normal group

The signalling pathways enriched by the dysregulated proteins in the GD/Normal comparison were metabolism, innate immune system, and neutrophil degranulation (Figure 3.6B). The metabolism pathway was similarly dysregulated in the GBC/Normal and GBC/GD comparisons. Although the metabolic pathway is upregulated in those comparisons, it is downregulated in GD compared to normal tissue groups. The biological oxidation pathway was also reduced in the GD/Normal comparison; oxidation is the transport of electrons in various metabolic cycles, such as the Krebs cycle or glycolysis, for energy production. ARK73 (aflatoxin B1 aldehyde reductase member 3), the top downregulated protein in the GD/Normal comparison, was enriched in the metabolism and biological oxidation pathways. ARK73 detoxifies aflatoxin B1, a known potent hepatocarcinogen, indicating a vital role in drug metabolism (244). In this study, liver carboxylesterase 1 (EST1) and glutathione S-transferase mu 3 (GSTM3) are also dysregulated and enriched in the metabolic pathway, and play a role in drug metabolism (245,246). Proteins involved in the metabolism of drugs are usually synthesised by the liver (247) and GD may cause aberrant expression of these proteins due to bile buildup from blockage of the biliary tree.

Another pathway enriched by the dysregulated proteins in GD tissues compared to normal gallbladders is the innate immune system pathway. Some innate immune cells are macrophages, fibroblasts, mast cells, and circulating leukocytes, including monocytes and neutrophils. These cells respond to various signals: pathogen-associated molecular patterns (PAMPs) in response to infection and damage-associated molecular patterns (DAMPs) in

response to tissue injury. PAMPs and DAMPs are detected using pattern recognition receptors (PRRs) (248). Inflammation can be classified as the cascade of events responding to infection or tissue damage (249). In the context of GD, inflammation is in response to tissue damage caused by the formation of gallstones. H2A variant histones H2A1B, H2A3, and H2A1C, were the top upregulated proteins in the GD/Normal group. While these proteins usually function to organise DNA structure, they may be released from the nucleus during cellular injury and inflammation to act as DAMPs. When histones function as DAMPs, the innate immune system is activated (250). However, although these proteins play a role in activating the immune response, they do not appear to be involved in the innate immune system network and pathway (Figure 3.6). The proteins enriched by the innate immunity pathway were also found in the neutrophil degranulation pathway (Figure 3.6B), corroborating a well-established link between these pathways.

In a normal state, inactivated neutrophils circulate in the body to not release toxic contents and cause cellular damage. Neutrophils become activated after stimulation from various signalling factors, such as bacterial products, cytokines and chemokines. Once neutrophils become activated, they secrete pro-inflammatory cytokines to activate the adaptive immune system. Neutrophils also contribute to various inflammatory diseases, such as inflammatory arthritis and chronic obstructive pulmonary disease (251). The degranulation of neutrophils occurs at sites of infection or tissue injury. As such, the release of the granules may further induce tissue injury and prolong the inflammatory state. It has been observed that histones may also activate neutrophils to form neutrophil extracellular traps (NETs), which promote the activation of the innate immune system (250).

4.2.4 Proteins similarly dysregulated in gallbladder cancer compared to gallstone disease and normal gallbladders

Thirty-eight proteins were identified to be similarly dysregulated in the tissue sample group comparisons: GBC/Normal and GBC/GD (Tables S3.8 and S3.12). Most of these proteins are dysregulated in the same direction (up- or downregulated in GBC compared to the controls), indicating that these alterations are specific to GBC irrespective of the precursor state. However, H2A1B, H2A3, and H2A1C do not follow this pattern. These proteins are upregulated in GBC compared to normal gallbladder but downregulated in GBC compared to GD. As previously described in Section 4.2.3, extracellular histones play a role in innate immunity and may be upregulated in response to cellular injury and inflammation. Thus, a

combination of increased cellular injury and inflammation may increase the levels of these histones in GD compared to in GBC (252).

The top dysregulated pathway enriched by the 38 proteins was the smooth muscle contraction pathway. Other dysregulated pathways include syndecan interactions, Sema4D in semaphorin signalling, and Rho GTPase effectors (Figure 3.7). The smooth muscle contraction pathway was dysregulated in the GBC/Normal and GBC/GD groups, with actin, myosin, and tropomyosin proteins enriching the pathway. As discussed in Section 4.2.1, the smooth muscle contraction pathway may be characteristic of the morphology of GBC and some of the proteins involved have been implicated in tumourigenesis. Another pathway dysregulated by the common proteins was semaphorin signalling. Semaphorin signalling plays a role in the immune response by activating dendritic cell activation and migration. Dendritic cells migrate by the action of myosin (253). Myosin has been observed to be enriched in this pathway. Additionally, plexins are the primary receptors for semaphorin signalling, which regulate R-RAS (Ras-related protein R-RAS) (254), and R-RAS has been observed to be involved in this pathway. In addition, plexins have been observed to regulate Rho GTPase activity. The Rho GTPase pathway was enriched by the common dysregulated GBC proteins. Rho (Ras homologous) GTPases are a family of small GTPases. RAC2 (Ras-related C3 botulinum toxin substrate 2) is a Rho GTPase enriched in this pathway and has been shown to regulate cytoskeleton remodelling (255). This pathway supports cytoskeleton remodelling by associating with myosin proteins, also enriched in the Rho GTPase pathway. This would suggest a role in cellular migration due to cytoskeleton remodelling.

4.3 Enriched pathways by the dysregulated proteins in the plasma sample group

Network and pathway analysis was performed on the relatively abundant proteins in the GBC/BBP plasma group comparison. Platelet degranulation, innate immune system, neutrophil degranulation, and haemostasis were among the pathways dysregulated (Figure 3.9B).

Platelets are discoid-shaped cells lacking a nucleus circulating in the blood. The primary function of platelets is haemostasis and coagulation. Platelets are filled with α-granules (the most abundant), dense granules, and lysosomes (256). There is evidence that platelets aid in initiating and propagating the inflammatory process (257). The α-granules influence inflammation by secreting pro-inflammatory chemokines. These chemokines influence a feedback loop to stimulate chemokine receptors on the platelet surface to induce further inflammatory responses (258). Platelet factor 4 (PLF4), a chemokine which aids in coagulation

by neutralising the anticoagulant heparin (259), was enriched in this pathway. Elevated PLF4 levels indicate the presence of activated platelets, which have been observed to accelerate tumour growth by modulating the tumour microenvironment and inducing angiogenesis (260,261). Another protein enriched in the platelet degranulation pathway is A1AT (alpha-1-antitrypsin). This is an important glycoprotein synthesised mainly by the liver and it maintains homeostasis by improving tissue repair. Elevated A1AT has been associated with increased tumour stage and poor prognosis in breast and prostate cancer (262). Haemostasis is the process of inducing clot formation, which links back to the platelet degranulation pathway as activated platelets induce coagulation (263,264). Furthermore, the present study found that proteins such as PLF4 were associated with platelet degranulation and haemostasis pathways (259). The pathogenesis of coagulation activation is multifactorial, although malignancies exhibit a unique feature by expressing clot-promoting molecules inducing a clotting cascade. It has been observed that elevated biomarkers of coagulation activation are associated with disease progression and poor outcomes in cancer (265). For example, fibrinogen a marker for haemostasis and an independent prognostic marker for GBC was demonstrated to be elevated in GBC patients (266).

The specific mechanisms of action for neutrophil degranulation and the innate immune system pathways in inflammation were described previously in Section 4.2.3. An increase in circulating neutrophils is associated with tumour metastasis and poor outcomes (267). This study observed the involvement of A1AT and AACT, which were upregulated and downregulated, respectively, in the neutrophil degranulation pathway. A1AT is present in neutrophil granules and may act as a neutralising factor against the proteases released by neutrophil granules. Neutrophil proteases cause cellular damage and A1AT protects against cellular damage by inactivating them (268). This points to a potential protective role of A1AT to protect the host against further cellular damage. AACT also functions as a protease inhibitor against neutrophil degranulation proteases. Its downregulation may indicate a shift to an increase of neutrophil protease enzymatic activity (269). The opposing roles of A1AT and AACT may indicate a complex mechanism in an attempt to protect the host from protease cellular damage (A1AT) and induce further cellular damage (AACT) for tumour progression. In addition, it was observed that the dysregulated proteins which enrich the neutrophil degranulation pathway were also involved in the innate immune system pathway, suggesting close functioning and cross-talk. One of the proteins, CRP, was involved in the innate immunity

pathway. CRP was also shown to be elevated in clinical tests in GBC patients compared to BBP patients (Table 3.2).

4.4 Commonly dysregulated proteins may be linked to gallbladder cancer in the patient sample group

4.4.1 Functions and pathways enriched by the identified commonly dysregulated proteins

Comparing the dysregulated proteins in the tissue and plasma sample group showed nine commonly dysregulated proteins. Seven of these proteins were common between the GBC/GD and GBC/BBP groups, and two between GBC/Normal, GBC/GD, and GBC/BBP groups (Figure 3.11; Table S3.12). The similarity of expression of these proteins at both the tissue and plasma levels hints towards secretion into the bloodstream and their involvement in tumour expression. (270). The nine common proteins are APOA1, APOA2, RET4, TTR, HEMO, HBB, and HBA which were downregulated while PIGR and APOE were upregulated (Table 4.1).

Quality control (QC) was performed on the mass spectrometry raw data of the CDPs to evaluate their suitability as markers of GBC. The MS2 fragment ion chromatogram was investigated for each CDP in tissue and plasma samples. Each of the MS2 fragment ion chromatograms for the CDPs indicated the correct predicted retention time for the fragment ions and alignment with the corresponding MS1 precursor detection (Figure 3.13 and S3.3 – S3.10). In addition, the intensity of the peaks indicated good signal-to-noise ratios with minimal interfering signals from contaminants. The peak widths showed good chromatographic protein elution. All these factors for the CDPs, together with their biological function, further suggest their possible role as potential biomarkers. The CDP subset showed a wide CV variation in the plasma sample group (Table S3.13). However, RET4 (76%), HEMO (67.2%), and PIGR (56.4%) showed the least variance while HBB (165.4%), and HBA (168.2%) showed the most. This suggests a wide range of detection abundance, although the CVs are generally lower in the GBC group compared to the BBP. This may indicate more consistent abundance detection in GBC compared to benign gallbladder diseases. The biological functions of the CDPs are discussed subsequently.

Table 4.1: Commonly Dysregulated Proteins (CDPs) and their functions

Protein	Function	Pathway	Dysregulation
Apolipoprotein A-I	Lipid transport	Lipid metabolism	Downregulated
Apolipoprotein A-II			Downregulated
Apolipoprotein E			Upregulated
Retinol-binding protein 4	Retinol transport	JAK2-STAT3 pathway	Downregulated
Transthyretin	Thyroid hormone transport		Downregulated
Hemopexin	Haem transport	Haem detoxification	Downregulated
Haemoglobin subunit beta	Oxygen binding	Oxygen transport	Downregulated
Haemoglobin subunit alpha			Downregulated
Polymeric Immunoglobulin receptor	Transcytosis of immunoglobulin IgA and IgM	Innate immune system	Upregulated

Apolipoproteins A-I (APOA1), A-II (APOA2), and E (APOE) are all involved in the transportation of lipids where they become metabolised; however, there have been recent discoveries which have noted diverse functions of these proteins in diseases. In a USA sample group, *APOA1* was underexpressed in GBC tumours compared to normal tissues (271). Similarly, a reduction of APOA1 was demonstrated in both tumours and serum samples obtained from liver cancer patients (272). APOA1 is the major structural and functional component of high-density lipoprotein (HDL), approximately 70%. It is a critical mediator in lipid transport by binding various receptors and transporters (273). APOA1 may function as a tumour suppressor in various cancer models and points towards an anti-inflammatory function which aligns with anti-tumourigenic processes (274). APOA2 was the most downregulated protein in plasma. Conversely, APOA2 has been detected to be reduced in other cancer types; these include gastric cancer (275), multiple myeloma (276), and pancreatic cancer (277). APOA2 is the second most abundant component of HDL, comprising approximately 20% and is routinely detected in the plasma of human patients due to its function in lipid transport. The exact mechanistic role of APOA2 in lipid transport is unclear (278).

APOE was one of two dysregulated proteins across all three comparisons; it transports metabolism-related lipids by binding low-density lipoprotein receptors. It is speculated that APOE does not function solely to transport lipid metabolism but also plays a role in immune response and cellular proliferation. In one study, APOE activation restricted the response of the innate immune system to suppress cancer proliferation, promoting tumour growth and metastasis (279). APOE has been linked to cancer development in ovarian and prostate epithelial malignancies (280,281). An elevated mRNA expression of APOE was detected in gastric cancer, associated with deep tissue invasion, lymph node metastasis, and reduced survival time (282).

RET4 (Retinol-binding protein 4) is a retinol (vitamin A) transporter protein in the blood, delivering retinol from the liver to peripheral tissues. RET4 can be transported across the plasma membrane of cells via the cell surface protein Retinoic Acid 6 (STRA6) leading to a signalling cascade via the JAK2-STAT3 pathway (283). A decrease in RET4 in sera of epithelial ovarian cancer patients has been observed (284). Transthyretin (TTR) is a protein involved in thyroid hormone transport and plays a role in retinol metabolism. TTR and RET4 bind in a complex with retinol to prevent the loss of small proteins after kidney filtration (285). The network analysis supports close interactions between TTR and RET4 (Figure ***3.7***). TTR expression is decreased in severe cases of liver disease and acute inflammation. TTR serum levels were decreased in ovarian cancer and advanced cervical and endometrial carcinomas (286). It is important to note that TTR is not a complete requirement for RET4 functioning. TTR knockout mice still indicated detectable circulating RET4, albeit at reduced levels. TTR-deficient mice did not indicate reduced retinoid levels in tissues. This suggests that TTR increases RET4 half-life and functional time by reducing renal filtration (285,287,288). Additionally, it has been observed that TTR is decreased in the serum of lung cancer patients compared to those with benign lung disease and healthy individuals (289). Although RET4 has been extensively studied in various cancer types, the specific mechanisms of how this protein affects tumour initiation are still unclear (283).

Hemopexin (HEMO) is expressed predominantly by the liver and has a high affinity for binding haem and labile (metabolically active) haem. This allows HEMO to transport the haem in a non-toxic form to promote liver-mediated haem detoxification (290). The current literature on HEMO and cancer is limited, although there are emerging studies which indicate a potential role in cancer. It was noted that plasma levels of HEMO in prostate cancer patients were lower

than that of healthy individuals (291). Similar findings were seen in HEMO-null mice, where the tumours observed were larger compared to HEMO-expressing mice (290). Haemoglobin subunits beta (HBB) and alpha (HBA) are significantly downregulated in GBC tissues and plasma. This downregulation is corroborated by the significant decrease in clinical haemoglobin in GBC tissue and plasma patient sample populations (Tables 3.1 and 3.2). Haemoglobin transports oxygen throughout the body and comprises two alpha and beta subunits each (292). This study also indicated that clinical Hb levels were significantly reduced in GBC patients compared to GD and BBP patients (Tables 3.1 and 3.2). A study conducted in India showed that clinical haemoglobin was significantly decreased in GBC patients compared to GD patients, as observed in this study (293). In non-small cell lung carcinoma (NSCLC), it was shown that HBB and HBA were downregulated, and the expression of these proteins correlated with tumour stage and lymph node spread. This indicates a protective role by HBB and HBA in NSCLC (292).

Polymeric immunoglobulin receptor (PIGR) was the other dysregulated protein across all three comparisons. This protein is responsible for the transcytosis of immunoglobulins IgA and IgM across mucosal surfaces; these are the first-line antibodies against infection which points to a role in innate and adaptive immunity (294). This is corroborated by the fact that PIGR is enriched in the innate immune system pathway dysregulated in the GBC/BBP group (Figure ***3.9***B). Proinflammatory cytokines upregulate PIGR in response to infections. However, it has been observed that this protein is also highly expressed in various cancers, including colon, hepatocellular, and ovarian. In colon carcinoma, increased expression of PIGR was linked to poor prognosis and hepatic metastasis (295). Elevated PIGR levels were detected in the sera of pancreatic, lung, and colon cancer patients (296–298). While new findings have indicated aberrant expression of PIGR, the mechanism behind the function of this protein in tumours remains uncertain, although there are some indications of the ribosome pathway being activated (294,299).

4.4.2 Commonly dysregulated proteins significantly correlate with each other and can distinguish gallbladder cancer from control groups

Principal component analysis (PCA) was performed to reduce the dimensionality of data to interpret it visually. Although there may be some overlap between GBC tissue/plasma and GD/BBP groups, it can be said that there is a separation between the GBC and control groups (Figure 3.14A). The Spearman's rank correlation test was performed to determine the

correlation between the CDPs in a protein-only context, i.e, the protein level abundance data regardless of sample source (tissue and plasma) (Figure 3.14C). APOA1, APOA2, RET4, TTR and HEMO are all significantly positively correlated. Strong positive contributions to PC1 by APOA1, APOA2, RET4, and TTR may indicate a similar ability of these proteins to distinguish between GBC and controls. HBB and HBA are positively correlated with each other but are both negatively correlated with APOA1, RET4, TTR, and HEMO, despite these proteins being all downregulated. There are significant positive contributions to PC2 by HBB and HBA, with a significant negative contribution by HEMO. The opposing contributions to PC2 by these proteins are supported by the significant negative correlation between HBB/HBA and HEO (Figure 3.14B and 3.14C). The opposing PCA contributions and negative correlations between HBB/HBA and HEMO link back to their diverse biological functions, as described in Section 4.4.1.

Some negative correlations were observed between HBB/HBA and the other CDPs. The expression patterns of these proteins seem to be independent of each other and therefore, their specific roles may be diverse and even antagonistic within similar processes. For example, HBB/HBA and APOA1 were negatively correlated but showed strong interactions demonstrated by the network analysis (Figure 3.12) (300). The main functions of APOA1 and HBB/HBA are cholesterol and oxygen transport, respectively. Elevated blood cholesterol transport leads to diminished blood oxygen levels. Notably, a decrease in clinical haemoglobin was observed in GBC patients (Tables 3.1 and 3.2), and raised cholesterol increases the risk of GBC (301). It is important to note that another study indicated that identifying significant dysregulation of blood proteins, such as HBB and HBA, may be due to erythrocyte contamination from plasma samples (302). However, in this study, HBB and HBA expression was observed in plasma and tissues, strongly suggesting that these proteins may be linked to the disease.

4.5 Proteins associated with a history of gallstones and tumour staging

A subset analysis was performed on Spectronaut, using an unpaired Student's t-test, to identify significantly dysregulated proteins in GBC plasma patients with a history of GD versus those without. Gallstone disease is a well-documented precursor linked to an increased risk of GBC. Approximately 70 – 80% of GBC cases progressed from GD (17,303). Two proteins (S100A8 and S100A9) were downregulated in patients with a history of GD compared to those without. These proteins are expressed by neutrophils and monocytes and function as calcium ion sensors

(304). Although these proteins are routinely expressed at high levels during inflammation, it has been demonstrated that at reduced levels they promote cell growth and proliferation (304,305). Therefore, the reduction of S100A8/9 in GBC patients with GD history may suggest this group of patients may experience an increase in tumour aggressiveness compared to thouse without GD history (182,303). However, this would need to be verified in future functional studies.

An additional subset analysis was performed to identify proteins associated with tumour staging (Table S3.11). From the 33 proteins identified to be dysregulated in the GBC plasma sample group, only two (APOE and ITIH3) were found to be significantly elevated in non-metastatic (Stage I-III) compared to metastatic (Stage IV) patients. APOE is a protein involved in cholesterol homeostasis, lipid metabolism, and immune suppression, similar to APOA1 and APOA2, as described in Section 4.4.1 (301,306). One study demonstrated that APOE overexpression in Stage II colorectal tumours is an independent prognostic factor for overall survival. Also, elevated serum APOE has been associated with early-stage hepatocellular carcinoma (307). Taken together, the increased expression of APOE in non-metastatic GBC may suggest a role in progressing tumourigenesis (308). ITIH3 (inter-alpha-trypsin inhibitor heavy chain 3) covalently links to hyaluronic acid, a major component of the ECM. Since linkage to hyaluronic acid strongly depends on ITIH3, dysregulation of this protein influences tumour development and progression (309). It can be speculated that loss of expression of ITIH3, as GBC progresses, can be attributed to reduced ECM interactions to drive the metastatic process.

4.6 Dysregulated proteins observed in a sample group of South African gallbladder cancer patients compares to other populations

GBC has been studied in various populations, predominantly in China and India. Several proteomic research studies across these populations suggest similarities and differences. For example, proteins such as cytokeratin 19 (CYFRA 21-1) (310) and TYMP (311) found to be dysregulated in the present study, have been identified in other patient sample group. CYFRA 21-1 has been identified as a potential marker in a Chinese population, and TYMP in a South Korean population. A recent Indian study using plasma-derived extracellular vesicles showed 86 proteins abundant in GBC patients (312). Of the 86 dysregulated proteins identified, 16 were also dysregulated in our study. In another study, GBC cell line-derived extracellular vesicles were investigated using LC-MS and iTRAQ labelling for biomarker discovery (313).

There were 29 proteins identified as enriched in the cell line-derived extracellular vesicles. Four of these (FETUA, TETN, ALBU, and G6PD) were also dysregulated in GBC samples in the present study. A separate study performed on an Indian patient cohort used 2D gel electrophoresis and LC-MS-based immunoproteomics to identify dysregulated proteins in GBC patients compared to GD and normal gallbladders due to an immune response. The study found 27 upregulated proteins in GBC patients. Of those, six (CAH2, HBA, HBB, THIL, ALDOB, and CAH1) were altered in the patients used in the present study (314). Taken together, there are similarities observed in the expression of certain proteins in this study and others. However, in some cases, although the proteins were significantly altered they were dysregulated in the opposite direction. For example, ALDOB was found to be upregulated across Stage I, II, and IIIA GBC patients and was observed to be upregulated in patients in this study. Notably, some significantly altered proteins in our patient sample group were not identified in other groups. However, these observations may be attributed to differences in sample cohorts or types but may also be a result of alternative proteomic technologies or strategies used for investigation.

5 CHAPTER 5: CONCLUSION AND FUTURE DIRECTION

This study provides novel insights into the dysregulated proteins and pathways associated with GBC onset and progression in South African patients. The clinical inflammatory markers WCC and CRP were identified as significantly elevated in GBC patients in the tissue and plasma patient sample group. Additionally, liver function tests, including total and direct bilirubin, albumin, ALP, GGT, and AST, were also significantly increased in GBC patients. The significant changes in the liver function tests are because the gallbladder and liver function closely together, so gallbladder diseases can cause aberrant expression of these enzymes. CA 19-9, a widely used marker that is considered for GBC diagnosis, was significantly elevated in the GBC tissue sample group but not in the plasma. Although some significant differences were observed in clinical parameters (such as the liver function tests and CA 19-9), these tests are also aberrant in various other diseases, including other malignancies.

SWATH-MS was performed in a South African GBC patient cohort for the first time providing much-needed molecular data on this group of patients. There were 62 (GBC/Normal), 194 (GBC/GD), 105 (GD/Normal), and 33 (GBC/BBP plasma) proteins dysregulated in each of the relevant comparisons. These differentially expressed proteins enrich several signalling pathways including lipid metabolism, smooth muscle interactions, innate immunity, ECM reorganisation, platelet and neutrophil degranulation, and integrin and syndecan interactions. The discovery of these signalling pathways aids in the understanding of GBC in this study's population and may be useful targets in future therapeutic studies. While some of the different comparisons share similar pathways, the differences observed especially in the enrichment of specific proteins may help further elucidate the mechanisms of GBC initiation and progression. Furthermore, it was observed that S100A8 and S100A9 were downregulated in patients with GD history compared to those without. Reduced levels of S100A8/9 induced by a pro-inflammatory environment may promote tumour growth and proliferation. Follow-up work is required to further demonstrate the potential role of these proteins and their significance in the group of patients and their implications in GBC prognosis. Two proteins, APOE and ITIH3, from the differentially expressed plasma proteins were shown to be elevated in non-metastatic (Stages I-III) compared to metastatic (Stage IV) GBC patients and may be further investigated as markers. APOE and ITIH3 may also have prognostic functions in GBC patients which would need to be investigated further.

The study also identified a group of commonly dysregulated proteins (CDPs) expressed at both the tissue and plasma levels. Statistical tests demonstrated their ability to distinguish between GBC and control groups. The expression of these proteins and their ability to distinguish between the different groups suggest their significance in GBC as potential biomarkers in this patient sample group, however, this needs to be further validated. Future studies would need to be conducted to verify the expression of the identified dysregulated proteins in a larger, independent patient sample group and evaluate the plasma of healthy participants. Moreover, the suitability of the identified dysregulated proteins compared to the well-known clinical biomarkers CA 19-9 and CEA could also be further investigated.

6 CHAPTER 6: STUDY LIMITATIONS

The main limitation of this study was the low number of tissue samples, especially in the normal gallbladder tissue group. Specifically, the low number of GBC tumours could be attributed to patients not being biopsied because of presenting at an advanced stage of the disease. Some missing clinicopathological information, such as tumour differentiation, CA 19-9 and CEA levels also limited the analyses that could be conducted. For example, comparison of the utility of clinical tumour markers such as CA 19-9 and CEA to identified proteins.

Additionally, the use of unmatched tissue and plasma samples may have limited the identification of more potential markers of GBC in this patient cohort. Lastly, circulating levels of the target proteins were not evaluated in healthy individuals preventing their assessment in an additional and important control group.

REFERENCES

1. Diamond JM. The reabsorptive function of the gall-bladder. J Physiol. 1962;161(3):442–73.

2. Nagral S. Anatomy relevant to cholecystectomy. J Minimal Access Surg. 2005;1(2):53.

3. Housset C, Chrétien Y, Debray D, Chignard N. Functions of the Gallbladder. Compr Physiol. 2016 Jun 13;6(3):1549–77.

4. Gray H, Standring S. Gray's Anatomy: the anatomical basis of clinical practice. Churchill Livingstone; 2008. 1181–1187 p.

5. Reid KM, Ramos-De la Medina A, Donohue JH. Diagnosis and Surgical Management of Gallbladder Cancer: A Review. J Gastrointest Surg. 2007 Apr 28;11(5):671–81.

6. Schnelldorfer T. Porcelain Gallbladder: A Benign Process or Concern for Malignancy? J Gastrointest Surg. 2013 Jun;17(6):1161–8.

7. Morimoto M, Matsuo T, Mori N. Management of Porcelain Gallbladder, Its Risk Factors, and Complications: A Review. Diagnostics. 2021 Jun 10;11(6):1073.

8. DesJardins H, Duy L, Scheirey C, Schnelldorfer T. Porcelain Gallbladder: Is Observation a Safe Option in Select Populations? J Am Coll Surg. 2018;226(6):1064–9.

9. Halaseh SA, Halaseh S, Shakman R. A Review of the Etiology and Epidemiology of Gallbladder Cancer: What You Need to Know. Cureus. 2022 Aug 22;14(8):e28260.

10. Paumgartner G, Sauerbruch T. Gallstones: pathogenesis. The Lancet. 1991 Nov;338(8775):1117–21.

11. Ansaloni L, Pisano M, Coccolini F, Peitzmann AB, Fingerhut A, Catena F, et al. 2016 WSES guidelines on acute calculous cholecystitis. World J Emerg Surg. 2016 Dec;11(1):25.

12. Knab LM, Boller AM, Mahvi DM. Cholecystitis. Surg Clin North Am. 2014 Apr;94(2):455–70.

13. Gutt C, Schläfer S, Lammert F. The Treatment of Gallstone Disease. Dtsch Ärztebl Int. 2020 Feb 28;117(9):148.

14. Rebholz C, Krawczyk M, Lammert F. Genetics of gallstone disease. Eur J Clin Invest. 2018 Jul;48(7):e12935.

15. Huang J, Patel HK, Boakye D, Chandrasekar VT, Koulaouzidis A, Lucero-Prisno III DE, et al. Worldwide distribution, associated factors, and trends of gallbladder cancer: A global country-level analysis. Cancer Lett. 2021 Nov;521:238–51.

16. Marschall HU, Einarsson C. Gallstone disease. J Intern Med. 2007 Jun;261(6):529–42.

17. Schmidt MA, Marcano-Bonilla L, Roberts LR. Gallbladder cancer: epidemiology and genetic risk associations. Chin Clin Oncol. 2019 Aug;8(4):31.

18. Valle JW, Kelley RK, Nervi B, Oh DY, Zhu AX. Biliary tract cancer. The Lancet. 2021 Jan;397(10272):428–44.

19. Stinton LM, Shaffer EA. Epidemiology of Gallbladder Disease: Cholelithiasis and Cancer. Gut Liver. 2012 Apr 15;6(2):172–87.

20. Parekh D, Lawson HH, Kuyl JM. Gallstone disease among black South Africans: A review of the Baragwanath Hospital experience. S Afr Med J. 1987;72(1):23–6.

21. Strasberg SM. Acute Calculous Cholecystitis. N Engl J Med. 2008;358(26):2804–11.

22. Savoca PE, Longo WE, Zucker KA, Mcmillen MM, Modlin IM. The Increasing Prevalence of Acalculous Cholecystitis in Outpatients: Results of a 7-Year Study. Ann Surg. 1990 Apr;211(4):433–7.

23. Upp JR, Nealon WH, Singh PO, Fagan CJ, Jonas AS, Greeley Jr GH, et al. Correlation of Cholecystokinin Receptors with Gallbladder Contractility in Patients with Gallstones. Ann Surg. 1987 Jun;205(6):641–8.

24. Sun H, Warren J, Yip J, Ji Y, Hao S, Han W, et al. Factors Influencing Gallstone Formation: A Review of the Literature. Biomolecules. 2022 Apr 6;12(4):550.

25. Wang TY, Portincasa P, Liu M, Tso P, Wang DQH. Mouse models of gallstone disease. Curr Opin Gastroenterol. 2018 Mar;34(2):59–70.

26. Wang HH, Liu M, Portincasa P, Wang DQH. Recent Advances in the Critical Role of the Sterol Efflux Transporters ABCG5/G8 in Health and Disease. Adv Exp Med Biol. 2020;1276:105–36.

27. Lee MH, Lu K, Hazard S, Yu H, Shulenin S, Hidaka H, et al. Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption. Nat Genet. 2001;27(1):79–83.

28. Wang HH, Portincasa P, Liu M, Tso P, Wang DQH. An Update on the Lithogenic Mechanisms of Cholecystokinin a Receptor (CCKAR), an Important Gallstone Gene for Lith13. Genes. 2020 Nov 29;11(12):1438.

29. Wang DQH, Schmitz F, Kopin AS, Carey MC. Targeted disruption of the murine cholecystokinin-1 receptor promotes intestinal cholesterol absorption and susceptibility to cholesterol cholelithiasis. J Clin Invest. 2004 Aug 16;114(4):521–8.

30. Figueiredo JC, Haiman C, Porcel J, Buxbaum J, Stram D, Tambe N, et al. Sex and ethnic/racial-specific risk factors for gallbladder disease. BMC Gastroenterol. 2017 Dec;17(1):153.

31. Katzenellenbogen BS, Choi I, Delage-Mourroux R, Ediger TR, Martini PGV, Montano M, et al. Molecular mechanisms of estrogen action: selective ligands and receptor pharmacology. J Steroid Biochem Mol Biol. 2000 Nov;74(5):279–85.

32. Berge KE, Tian H, Graf GA, Yu L, Grishin NV, Schultz J, et al. Accumulation of Dietary Cholesterol in Sitosterolemia Caused by Mutations in Adjacent ABC Transporters. Science. 2000 Dec;290(5497):1771–5.

33. Rege RV, Prystowsky JB. Inflammation and a Thickened Mucus Layer in Mice with Cholesterol Gallstones. J Surg Res. 1998 Jan;74(1):81–5.

34. Shrikhande SV, Barreto SG, Singh S, Udwadia TE, Agarwal AK. Cholelithiasis in gallbladder cancer: Coincidence, cofactor, or cause! Eur J Surg Oncol EJSO. 2010 Jun;36(6):514–9.

35. Landskron G, De la Fuente M, Thuwajit P, Thuwajit C, Hermoso MA. Chronic Inflammation and Cytokines in the Tumor Microenvironment. J Immunol Res. 2014;2014:1–19.

36. Perwez Hussain S, Harris CC. Inflammation and cancer: An ancient link with novel potentials. Int J Cancer. 2007 Dec 1;121(11):2373–80.

37. Karin M. Nuclear factor-κB in cancer development and progression. Nature. 2006 May;441(7092):431–6.

38. Biswas R, Bagchi A. NFkB pathway and inhibition: an overview. Comput Mol Biol. 2016;6(1):1–20.

39. Mittal M, Siddiqui MR, Tran K, Reddy SP, Malik AB. Reactive Oxygen Species in Inflammation and Tissue Injury. Antioxid Redox Signal. 2014 Mar;20(7):1126–67.

40. Mauro C, Zazzeroni F, Papa S, Bubici C, Franzoso G. The NF-κB Transcription Factor Pathway as a Therapeutic Target in Cancer: Methods for Detection of NF-κB Activity. In: Inflammation and Cancer [Internet]. Humana Press; 2009 [cited 2023 Jan 15]. p. 169–207. (Methods in Molecular Biology™; vol. 512). Available from: http://link.springer.com/10.1007/978-1-60327-530-9_10

41. Fan Y, Mao R, Yang J. NF-κB and STAT3 signaling pathways collaboratively link inflammation to cancer. Protein Cell. 2013 Mar;4(3):176–85.

42. Bowman T, Broome MA, Sinibaldi D, Wharton W, Pledger WJ, Sedivy JM, et al. Stat3-mediated Myc expression is required for Src transformation and PDGF-induced mitogenesis. Proc Natl Acad Sci. 2001 Jun 19;98(13):7319–24.

43. Gonda TA, Tu S, Wang TC. Chronic inflammation, the tumor microenvironment and carcinogenesis. Cell Cycle. 2009 Jul;8(13):2005–13.

44. Tu S, Bhagat G, Cui G, Takaishi S, Kurt-Jones EA, Rickman B, et al. Overexpression of Interleukin-1β Induces Gastric Inflammation and Cancer and Mobilizes Myeloid-Derived Suppressor Cells in Mice. Cancer Cell. 2008 Nov;14(5):408–19.

45. Bromberg J, Wang TC. Inflammation and Cancer: IL-6 and STAT3 Complete the Link. Cancer Cell. 2009 Feb;15(2):79–80.

46. Hundal R, Shaffer EA. Gallbladder cancer: epidemiology and outcome. Clin Epidemiol. 2014;6(99):99–109.

47. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209–49.

48. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin. 2023 Jan;73(1):17–48.

49. Arroyo GF, Gentile A, Parada LA. Gallbladder cancer: South American experience. Chin Clin Oncol. 2016 Oct;5(5):67.

50. Khan ZA, Khan MU, Brand M. Gallbladder cancer in Africa: A higher than expected rate in a "low-risk" population. Surgery. 2022 Apr;171(4):855–8.

51. Bruni L, Albero G, Serrano B, Mena M, Gómez D, Muñoz J, et al. ICO/IARC information centre on HPV and cancer (HPV information centre). Human papillomavirus and related diseases in the world. Summ Rep. 2019;30(17).

52. Halder A, Swain JR, Mohanty D, Majumder R. Gallbladder Carcinoma: Epidemiology, Risk Factors and Modes of Presentation in Eastern Part of India. Ann Clin Case Rep. 2020;5(1818):1–5.

53. Pilgrim CHC, Groeschl RT, Christians KK, Gamblin TC. Modern perspectives on factors predisposing to the development of gallbladder cancer. HPB. 2013 Nov;15(11):839–44.

54. Pandey M, Shukla VK. Lifestyle, parity, menstrual and reproductive factors and risk of gallbladder cancer: Eur J Cancer Prev. 2003 Aug;12(4):269–72.

55. Tsuchiya Y, Terao M, Okano K, Nakamura K, Oyama M, Ikegami K, et al. Mutagenicity and Mutagens of the Red Chili Pepper as Gallbladder Cancer Risk Factor in Chilean Women. Asian Pac J Cancer Prev. 2011;12(2):471–6.

56. Ikoma T, Tsuchiya Y, Asai T, Okano K, Ito N, Endoh K, et al. Ochratoxin A Contamination of Red Chili Peppers from Chile, Bolivia and Peru, Countries with a High Incidence of Gallbladder Cancer. Asian Pac J Cancer Prev. 2015 Sep 2;16(14):5987–91.

57. Andrén-Sandberg Å. Diagnosis and management of gallbladder cancer. North Am J Med Sci. 2012;4(7):293–9.

58. Koshiol J, Wozniak A, Cook P, Adaniel C, Acevedo J, Azócar L, et al. *Salmonella enterica* serovar Typhi and gallbladder cancer: a case–control study and meta-analysis. Cancer Med. 2016 Nov;5(11):3310–3235.

59. Murphy G, Michel A, Taylor PR, Albanes D, Weinstein SJ, Virtamo J, et al. Association of seropositivity to *Helicobacter* species and biliary tract cancer in the ATBC study. Hepatology. 2014 Dec;60(6):1963–71.

60. Kamisawa T, Kuruma S, Tabata T, Chiba K, Iwasaki S, Koizumi S, et al. Pancreaticobiliary maljunction and biliary cancer. J Gastroenterol. 2015 Mar;50(3):273–9.

61. Kamisawa T, Kuruma S, Chiba K, Tabata T, Koizumi S, Kikuyama M. Biliary carcinogenesis in pancreaticobiliary maljunction. J Gastroenterol. 2017 Feb;52(2):158–63.

62. Kimura W. Congenital dilatation of the common bile duct and pancreaticobiliary maljunction—clinical implications. Langenbecks Arch Surg. 2009 Mar;394(2):209–13.

63. Ramachandran A, Srivastava DN, Madhusudhan KS. Gallbladder cancer revisited: the evolving role of a radiologist. Br J Radiol. 2021 Jan 1;94(1117):20200726.

64. Cariati A, Piromalli E, Cetta F. Gallbladder cancers: associated conditions, histological types, prognosis, and prevention. Eur J Gastroenterol Hepatol. 2014 May;26(5):562–9.

65. Sánchez-Danés A, Blanpain C. Deciphering the cells of origin of squamous cell carcinomas. Nat Rev Cancer. 2018 Sep;18(9):549–61.

66. White VA, Hyrcza MD, Lennerz JK, Thierauf J, Lokuhetty D, Cree IA, et al. Mucoepidermoid carcinoma (MEC) and adenosquamous carcinoma (ASC), the same or different entities? Mod Pathol. 2022 Oct;35(10):1484–93.

67. Hofland J, Zandee WT, de Herder WW. Role of biomarker tests for diagnosis of neuroendocrine tumours. Nat Rev Endocrinol. 2018 Nov;14(11):656–69.

68. Moskal TL, Zhang PJ, Nava HR. Small cell carcinoma of the gallbladder. J Surg Oncol. 1999 Jan;70(1):54–9.

69. Mukkamalla SKR, Kashyap S, Recio-Boiles A, Babiker HM. Gallbladder Cancer. [Updated 2022 Jul 10] [Internet]. Treasure Island (FL): StatPearls Publishing; 2022. Available from: https://www.ncbi.nlm.nih.gov/books/NBK442002/?report=classic

70. Benson AB, D'Angelica MI, Abbot DE, Anaya DA, Anders R, Are C, et al. Hepatobiliary cancers, version 2.2021, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2021;19(5):541–65.

71. Koimtzis GD, Chalklin CG, Carrington-Windo E, Ramsden M, Stefanopoulos L, Kosmidis CS. The Role of Fine Needle Aspiration Cytology in the Diagnosis of Gallbladder Cancer: A Systematic Review. Diagnostics. 2021 Aug 6;11(8):1427.

72. Moreels TG. Endoscopic retrograde cholangiopancreatography in patients with altered anatomy: How to deal with the challenges? World J Gastrointest Endosc. 2014;6(8):345–51.

73. Kumar KS, Rao PS. The Role of MRCP in Evaluation of Pancreatic and Biliary Pathologies in Correlation with ERCP. J Contemp Med Dent. 2021;9(2):24–8.

74. Gupta P, Meghashyam K, Marodia Y, Gupta V, Basher R, Das CK, et al. Locally advanced gallbladder cancer: a review of the criteria and role of imaging. Abdom Radiol. 2021 Mar;46(3):998–1007.

75. Amin MB, Edge SB, Greene FL, Byrd DR, Brookland RK, Washington MK, et al. AJCC Cancer Staging Manual. 8th ed. New York: Springer Cham; 2017.

76. Li M, Liu F, Zhang F, Zhou W, Jiang X, Yang Y, et al. Genomic *ERBB2* / *ERBB3* mutations promote PD-L1-mediated immune escape in gallbladder cancer: a whole-exome sequencing analysis. Gut. 2018 Jun 28;68(6):1024–33.

77. Baichan P, Naicker P, Devar JWS, Smith M, Candy GP, Nweke E. Targeting gallbladder cancer: a pathway based perspective. Mol Biol Rep. 2020;47(3):2361–9.

78. Buchegger K, Silva R, López J, Ili C, Araya JC, Leal P, et al. The ERK/MAPK pathway is overexpressed and activated in gallbladder cancer. Pathol - Res Pract. 2017 May;213(5):476–82.

79. King D, Yeomanson D, Bryant HE. PI3King the Lock: Targeting the PI3K/Akt/mTOR Pathway as a Novel Therapeutic Strategy in Neuroblastoma. J Pediatr Hematol Oncol. 2015 May;37(4):245–51.

80. Liu L, Yang Z lin, Wang C, Miao X, Liu Z, Li D, et al. The Expression of Notch 1 and Notch 3 in Gallbladder Cancer and Their Clinicopathological Significance. Pathol Oncol Res. 2016 Jul;22(3):483–92.

81. Tanno S, Obara T, Fujii T, Mizukami Y, Shudo R, Nishino N, et al. Proliferative potential and K-ras mutation in epithelial hyperplasia of the gallbladder in patients with anomalous pancreaticobiliary ductal union. Cancer. 1998 Jul 15;83(2):267–75.

82. Song X, Hu Y, Li Y, Shao R, Liu F, Liu Y. Overview of current targeted therapy in gallbladder cancer. Signal Transduct Target Ther. 2020;5(1):1–9.

83. Sharma U, Tuli HS, Uttam V, Choudhary R, Sharma B, Sharma U, et al. Role of Hedgehog and Hippo signaling pathways in cancer: A special focus on non-coding RNAs. Pharmacol Res. 2022 Dec;186:106523.

84. Goldin RD, Roa JC. Gallbladder cancer: a morphological and molecular update. Histopathology. 2009 Aug;55(2):218–29.

85. Zhu G, Pan C, Bei JX, Li B, Liang C, Xu Y, et al. Mutant p53 in Cancer Progression and Targeted Therapies. Front Oncol. 2020 Nov 6;10:595187.

86. Mishra SK, Kumari N, Krishnani N. Molecular pathogenesis of gallbladder cancer: An update. Mutat Res Mol Mech Mutagen. 2019 Nov;816–818:111674.

87. Fan K, Wang J, Shen S, Ni X, Gong Z, Zheng B, et al. ERBB2 S310F mutation independently activates PI3K/AKT and MAPK pathways through homodimers to contribute gallbladder carcinoma growth. Med Oncol. 2022;39(5):1–10.

88. Kim YW, Huh SH, Park YK, Yoon TY, Lee SM, Hong SH. Expression of the c-erb-B2 and p53 protein in gallbladder carcinomas. Oncol Rep. 2001;8(5):1127–32.

89. Gianni L. Is there room for another HER2-targeting drug? Lancet Oncol. 2018;19(7):847–9.

90. Hashemi Goradel N, Najafi M, Salehi E, Farhood B, Mortezaee K. Cyclooxygenase-2 in cancer: A review. J Cell Physiol. 2019 May;234(5):5683–99.

91. Sharma P, Caldwell TS, Rivera MN, Gullapalli RR. Cadmium exposure activates Akt/ERK Signaling and pro-inflammatory COX-2 expression in human gallbladder epithelial cells via a ROS dependent mechanism. Toxicol In Vitro. 2020 Sep;67:104912.

92. Li Y, Zhang J, Ma H. Chronic inflammation and gallbladder cancer. Cancer Lett. 2014 Apr;345(2):242–8.

93. Bizama C, García P, Espinoza JA, Weber H, Leal P, Nervi B, et al. Targeting specific molecular pathways holds promise for advanced gallbladder cancer therapy. Cancer Treat Rev. 2015 Mar;41(3):222–34.

94. Sheng W, Tang J, Cao R, Shi X, Ma Y, Dong M. Numb-PRRL promotes TGF-β1- and EGF-induced epithelial-to-mesenchymal transition in pancreatic cancer. Cell Death Dis. 2022 Feb 23;13(2):173.

95. Du P, Liang H, Fu X, Wu P, Wang C, Chen H, et al. SLC25A22 promotes proliferation and metastasis by activating MAPK/ERK pathway in gallbladder cancer. Cancer Cell Int [Internet]. 2019 Dec [cited 2019 Mar 7];19(1). Available from: https://cancerci.biomedcentral.com/articles/10.1186/s12935-019-0746-9

96. Wu XS, Wang XA, Wu WG, Hu YP, Li ML, Ding Q, et al. MALAT1 promotes the proliferation and metastasis of gallbladder cancer cells by activating the ERK/MAPK pathway. Cancer Biol Ther. 2014 Jun;15(6):806–14.

97. Canale M, Monti M, Rapposelli IG, Ulivi P, Sullo FG, Bartolini G, et al. Molecular Targets and Emerging Therapies for Advanced Gallbladder Cancer. Cancers. 2021 Nov 12;13(22):5671.

98. Yun J, Rago C, Cheong I, Pagliarini R, Angenendt P, Rajagopalan H, et al. Glucose Deprivation Contributes to the Development of KRAS Pathway Mutations in Tumor Cells. Science. 2009 Sep 18;325(5947):1555–9.

99. Kemp SB, Cheng N, Markosyan N, Sor R, Kim IK, Hallin J, et al. Efficacy of a Small-Molecule Inhibitor of KrasG12D in Immunocompetent Models of Pancreatic Cancer. Cancer Discov. 2023 Feb 6;13(2):298–311.

100. Yang D, Chen T, Zhan M, Xu S, Yin X, Liu Q, et al. Modulation of mTOR and epigenetic pathways as therapeutics in gallbladder cancer. Mol Ther - Oncolytics. 2021 Mar;20:59–70.

101. Lunardi A, Webster KA, Papa A, Padmani B, Clohessy JG, Bronson RT, et al. Role of aberrant PI3K pathway activation in gallbladder tumorigenesis. Oncotarget. 2014 Feb 28;5(4):894–900.

102. Tewari D, Patni P, Bishayee A, Sah AN, Bishayee A. Natural products targeting the PI3K-Akt-mTOR signaling pathway in cancer: A novel therapeutic strategy. Semin Cancer Biol. 2022 May;80:1–17.

103. Cao Y, Liu X, Lu W, Chen Y, Wu X, Li M, et al. Fibronectin promotes cell proliferation and invasion through mTOR signaling pathway activation in gallbladder cancer. Cancer Lett. 2015 May;360(2):141–50.

104. Sun L rui, Li S yu, Guo Q shi, Zhou W, Zhang H mei. SPOCK1 Involvement in Epithelial-to-Mesenchymal Transition: A New Target in Cancer Therapy? Cancer Manag Res. 2020 May;12:3561–9.

105. Shu YJ, Weng H, Ye YY, Hu YP, Bao RF, Cao Y, et al. SPOCK1 as a potential cancer prognostic marker promotes the proliferation and metastasis of gallbladder cancer cells by activating the PI3K/AKT pathway. Mol Cancer. 2015 Dec;14(1):p.12.

106. Ye Z, Chen J, Hu X, Yang S, Xuan Z, Lu X, et al. SPOCK1: a multi-domain proteoglycan at the crossroads of extracellular matrix remodeling and cancer development. Am J Cancer Res. 2020;10(10):3127.

107. Iwasaki H, Nakano K, Shinkai K, Kunisawa Y, Hirahashi M, Oda Y, et al. Hedgehog Gli3 activator signal augments tumorigenicity of colorectal cancer via upregulation of adherence-related genes. Cancer Sci. 2013;104(3):328–36.

108. Matsushita S, Onishi H, Nakano K, Nagamatsu I, Imaizumi A, Hattori M, et al. Hedgehog signaling pathway is a potential therapeutic target for gallbladder cancer. Cancer Sci. 2014 Mar;105(3):272–80.

109. Azevedo Martins JM, Rabelo-Santos SH, do Amaral Westin MC, Zeferino LC. Tumoral and stromal expression of MMP-2, MMP-9, MMP-14, TIMP-1, TIMP-2, and VEGF-A in cervical cancer patient survival: a competing risk analysis. BMC Cancer. 2020 Dec;20(1):1–11.

110. Jeng KS, Chang CF, Lin SS. Sonic Hedgehog Signaling in Organogenesis, Tumors, and Tumor Microenvironments. Int J Mol Sci. 2020 Jan 23;21(3):758.

111. Xie F, Xu X, Xu A, Liu C, Liang F, Xue M, et al. Aberrant activation of Sonic hedgehog signaling in chronic cholecystitis and gallbladder carcinoma. Hum Pathol. 2014 Mar;45(3):513–21.

112. Wee P, Wang Z. Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. Cancers. 2017 May 17;9(5):52.

113. Rajaram P, Chandra P, Ticku S, Pallavi BK, Rudresh KB, Mansabdar P. Epidermal growth factor receptor: Role in human cancer. Indian J Dent Res. 2017;28(6):687–94.

114. Barreto SG, Dutt A, Chaudhary A. A genetic model for gallbladder carcinogenesis and its dissemination. Ann Oncol. 2014 Jun;25(6):1086–97.

115. Sigismund S, Avanzato D, Lanzetti L. Emerging functions of the EGFR in cancer. Mol Oncol. 2018 Jan;12(1):3–20.

116. Das C, Mukhopadhyay M, Subba S, Saha AK, Mukhopadhyay B. Role of EGFR and HER-2/NEU Expression in Gall Bladder Carcinoma (GBC). J Lab Physicians. 2021 Mar;13(01):029–35.

117. Zhao S, Cao Y, Liu S bo, Wang X an, Bao R fa, Shu Y jun, et al. The E545K mutation of PIK3CA promotes gallbladder carcinoma progression through enhanced binding to EGFR. J Exp Clin Cancer Res. 2016 Dec;35(1):97.

118. Ariyama H, Qin B, Baba E, Tanaka R, Mitsugi K, Harada M, et al. Gefitinib, a selective EGFR tyrosine kinase inhibitor, induces apoptosis through activation of Bax in human gallbladder adenocarcinoma cells. J Cell Biochem. 2006 Mar 1;97(4):724–34.

119. García P, Lamarca A, Díaz J, Carrera E, Roa J, on behalf of the European-Latin American ESCALON Consortium. Current and New Biomarkers for Early Detection, Prognostic Stratification, and Management of Gallbladder Cancer Patients. Cancers. 2020 Dec 7;12(12):3670.

120. Težak Ž, Kondratovich MV, Mansfield E. US FDA and personalized medicine: *in vitro* diagnostic regulatory perspective. Pers Med. 2010 Sep;7(5):517–30.

121. Lin HZ, Zhang T, Chen MY, Shen JL. Novel biomarkers for the diagnosis and prognosis of gallbladder cancer. J Dig Dis. 2021 Feb;22(2):62–71.

122. Bind MK, Mishra RR, Kumar V, Misra V, Singh PA. Serum CA 19-9 and CA 125 as a diagnostic marker in carcinoma of gallbladder. J Pathol Microbiol. 2021;64(1):65.

123. Agrawal S, Saxena R. Raised carbohydrate antigen 19-9 levels detect recurrences and impacts overall and disease-free survival in radically resected gallbladder cancer: A simple surveillance marker? J Cancer Res Ther. 2023;

124. Lee T, Teng TZJ, Shelat VG. Carbohydrate antigen 19-9 — tumor marker: Past, present, and future. World J Gastrointest Surg. 2020 Dec 27;12(12):468–90.

125. Shukla VK, Sharma D, Dixit VK. Diagnostic value of serum CA242, CA 19-9, CA 15-3 and CA 125 in patients with carcinoma of the gallbladder. Trop Gastroenterol Off J Dig Dis Found. 2006;27(4):160–5.

126. Lakemeyer L, Sander S, Wittau M, Henne-Bruns D, Kornmann M, Lemke J. Diagnostic and Prognostic Value of CEA and CA19-9 in Colorectal Cancer. Diseases. 2021 Mar 17;9(1):21.

127. Ince AL, Yildiz K, Baysal B, Danalioglu A, Kocaman O, Tozlu M, et al. Roles of serum and biliary CEA, CA19-9, VEGFR3, and TAC in differentiating between malignant and benign biliary obstructions. Turk J Gastroenterol. 2014 Jul 6;25(2):162–9.

128. Baj J, Bryliński Ł, Woliński F, Granat M, Kostelecka K, Duda P, et al. Biomarkers and Genetic Markers of Hepatocellular Carcinoma and Cholangiocarcinoma—What Do We Already Know. Cancers. 2022 Mar 15;14(6):1493.

129. Okamura K, Takayama K, Izumi M, Harada T, Furuyama K, Nakanishi Y. Diagnostic value of CEA and CYFRA 21-1 tumor markers in primary lung cancer. Lung Cancer. 2013 Apr;80(1):45–9.

130. Barak V, Meirovitz A, Leibovici V, Rachmut J, Peretz T, Eliashar R, et al. The Diagnostic and Prognostic Value of Tumor Markers (CEA, SCC, CYFRA 21-1, TPS) in Head and Neck Cancer Patients. Anticancer Res. 2015;35(10):5519–24.

131. Anderson NL, Anderson NG. Proteome and proteomics: New technologies, new concepts, and new words. Electrophoresis. 1998 Aug;19(11):1853–61.

132. Yan Y, Yeon SY, Qian C, You S, Yang W. On the Road to Accurate Protein Biomarkers in Prostate Cancer Diagnosis and Prognosis: Current Status and Future Advances. Int J Mol Sci. 2021 Dec 17;22(24):13537.

133. Dupree EJ, Jayathirtha M, Yorkey H, Mihasan M, Petre BA, Darie CC. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. Proteomes. 2020 Jul 6;8(3):14.

134. Schirle M, Bantscheff M, Kuster B. Mass Spectrometry-Based Proteomics in Preclinical Drug Discovery. Chem Biol. 2012 Jan;19(1):72–84.

135. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. J Biol Chem. 1975 May;250(10):4007–21.

136. Ünlü M, Morgan ME, Minden JS. Difference gel electrophoresis. A single gel method for detecting changes in protein extracts. Electrophoresis. 1997;18(11):2071–7.

137. Marouga R, David S, Hawkins E. The development of the DIGE system: 2D fluorescence difference gel analysis technology. Anal Bioanal Chem. 2005 Jun;382(3):669–78.

138. Van den Bergh G, Clerens S, Cnops L, Vandesande F, Arckens L. Fluorescent two-dimensional difference gel electrophoresis and mass spectrometry identify age-related protein expression differences for the primary visual cortex of kitten and adult cat: 2D-DIGE and visual cortex development. J Neurochem. 2003 Mar 18;85(1):193–205.

139. Tannu NS, Hemby SE. Two-dimensional fluorescence difference gel electrophoresis for comparative proteomics profiling. Nat Protoc. 2006 Nov;1(4):1732–42.

140. Sutandy FXR, Qian J, Chen C, Zhu H. Overview of Protein Microarrays. Curr Protoc Protein Sci. 2013 Apr;72(1):27.1.1-27.1.16.

141. Huang Y, Zhu H. Protein Array-based Approaches for Biomarker Discovery in Cancer. Genomics Proteomics Bioinformatics. 2017 Apr;15(2):73–81.

142. Moore CD, Ajala OZ, Zhu H. Applications in high-content functional protein microarrays. Curr Opin Chem Biol. 2016 Feb;30:21–7.

143. Benvenuto M, Mattera R, Masuelli L, Tresoldi I, Giganti MG, Frajese GV, et al. The crossroads between cancer immunity and autoimmunity: antibodies to self antigens. Front Biosci-Landmark. 2017;22(8):1289–329.

144. Paweletz CP, Charboneau L, Bichsel VE, Simone NL, Chen T, Gillespie JW, et al. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene. 2001 Apr 12;20(16):1981–9.

145. Nishizuka SS, Mills GB. New era of integrated cancer biomarker discovery using reverse-phase protein arrays. Drug Metab Pharmacokinet. 2016 Feb;31(1):35–45.

146. Rapkiewicz A, Espina V, Zujewski JA, Lebowitz PF, Filie A, Wulfkuhle J, et al. The needle in the haystack: Application of breast fine-needle aspirate samples to quantitative protein microarray technology. Cancer Cytopathol Interdiscip Int J Am Cancer Soc. 2007 May 8;111(3):173–84.

147. Yang L, Wang J, Li J, Zhang H, Guo S, Yan M, et al. Identification of Serum Biomarkers for Gastric Cancer Diagnosis Using a Human Proteome Microarray. Mol Cell Proteomics. 2016 Feb;15(2):614–23.

148. Maurya P, Meleady P, Dowling P, Clynes M. Proteomic Approaches for Serum Biomarker Discovery in Cancer. Anticancer Res. 2007;27(3A):1247–55.

149. Vreeker GCM, Wuhrer M. Reversed-phase separation methods for glycan analysis. Anal Bioanal Chem. 2017 Jan;409(2):359–78.

150. Zhao Q, Fang F, Shan Y, Sui Z, Zhao B, Liang Z, et al. In-Depth Proteome Coverage by Improving Efficiency for Membrane Proteome Analysis. Anal Chem. 2017 May 16;89(10):5179–85.

151. Coskun O. Separation Tecniques: Chromatography. North Clin Istanb. 2016;3(2):156–60.

152. Nesvizhskii AI, editor. Protein Identification by Tandem Mass Spectrometry and Sequence Database Searching. Mass Spectrom Data Anal Proteomics. 2007;87–119.

153. Dunn WB. Mass spectrometry in systems biology: An introduction. Methods Enzymol. 2011;500:15–35.

154. Karpievitch YV, Polpitiya AD, Anderson GA, Smith RD, Dabney AR. Liquid chromatography mass spectrometry-based proteomics: Biological and technological aspects. Ann Appl Stat [Internet]. 2010 Dec 1 [cited 2023 Jan 31];4(4). Available from: https://projecteuclid.org/journals/annals-of-applied-statistics/volume-4/issue-4/Liquid-chromatography-mass-spectrometry-based-proteomics--Biological-and-technological/10.1214/10-AOAS341.full

155. Qian WJ, Jacobs JM, Liu T, Camp II DG, Smith RD. Advances and Challenges in Liquid Chromatography-Mass Spectrometry-based Proteomics Profiling for Clinical Applications. Mol Cell Proteomics. 2006 Oct;5(10):1727–44.

156. Bauer A, Kuster B. Affinity purification-mass spectrometry: Powerful tools for the characterization of protein complexes. Eur J Biochem. 2003 Jan 24;270(4):570–8.

157. Kinoshita Y, Uo T, Jayadev S, Garden GA, Conrads TP, Veenstra TD, et al. Potential Applications and Limitations of Proteomics in the Study of Neurological Disease. Arch Neurol. 2006 Dec 1;63(12):1692–6.

158. Aslam B, Basit M, Nisar MA, Khurshid M, Rasool MH. Proteomics: Technologies and Their Applications. J Chromatogr Sci. 2017 Feb;55(2):182–96.

159. Chait BT. Mass spectrometry: bottom-up or top-down? Science. 2006;314(5796):65–6.

160. Cristobal A, Marino F, Post H, van den Toorn HWP, Mohammed S, Heck AJR. Toward an Optimized Workflow for Middle-Down Proteomics. Anal Chem. 2017 Mar 21;89(6):3318–25.

161. Tsiatsiani L, Heck AJR. Proteomics beyond trypsin. FEBS J. 2015 Jul;282(14):2612–26.

162. Dau T, Gupta K, Berger I, Rappsilber J. Sequential Digestion with Trypsin and Elastase in Cross-Linking Mass Spectrometry. Anal Chem. 2019 Apr 2;91(7):4472–8.

163. Huesgen PF, Lange PF, Rogers LD, Solis N, Eckhard U, Kleifeld O, et al. LysargiNase mirrors trypsin for protein C-termini and methylation site identification. Nat Methods. 2015;12(1):55–8.

164. Glatter T, Ludwig C, Ahrné E, Aebersold R, Heck AJR, Schmidt A. Large-Scale Quantitative Assessment of Different In-Solution Protein Digestion Protocols Reveals Superior Cleavage Efficiency of Tandem Lys-C/Trypsin Proteolysis over Trypsin Digestion. J Proteome Res. 2012 Nov 2;11(11):5145–56.

165. Klammer AA, MacCoss MJ. Effects of Modified Digestion Schemes on the Identification of Proteins from Complex Mixtures. J Proteome Res. 2006 Mar 1;5(3):695–700.

166. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, et al. 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. 2012;11(6):016717-1-016717–17.

167. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Appendix to: Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol Syst Biol. 2018 Aug;14(8):e8126.

168. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol Syst Biol. 2018 Aug;14(8):e8126.

169. Nweke EE, Naicker P, Aron S, Stoychev S, Devar J, Tabb DL, et al. SWATH-MS based proteomic profiling of pancreatic ductal adenocarcinoma tumours reveals the interplay between the extracellular matrix and related intracellular pathways. PloS ONE. 2020 Oct 13;15(10):e0240453.

170. Sajic T, Liu Y, Arvaniti E, Surinova S, Williams EG, Schiess R, et al. Similarities and Differences of Blood N-Glycoproteins in Five Solid Carcinomas at Localized Clinical Stage Analyzed by SWATH-MS. Cell Rep. 2018 May;23(9):2819-2831.e5.

171. Milighetti M, Krasny L, Lee ATJ, Morani G, Szecsei C, Chen Y, et al. Proteomic profiling of soft tissue sarcomas with SWATH mass spectrometry. J Proteomics. 2021 Jun;241:104236.

172. Janacova L, Faktor J, Capkova L, Paralova V, Pospisilova A, Podhorec J, et al. SWATH-MS Analysis of FFPE Tissues Identifies Stathmin as a Potential Marker of Endometrial Cancer in Patients Exposed to Tamoxifen. J Proteome Res. 2020 Jul 2;19(7):2617–30.

173. Campbell MC, Tishkoff SA. African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. Annu Rev Genomics Hum Genet. 2008 Sep 1;9(1):403–33.

174. Batth TS, Tollenaere MAX, Rüther P, Gonzalez-Franquesa A, Prabhakar BS, Bekker-Jensen S, et al. Protein Aggregation Capture on Microparticles Enables Multipurpose Proteomics Sample Preparation*. Mol Cell Proteomics. 2019 May;18(5):1027–35.

175. Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol. 2014 Mar;32(3):223–6.

176. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003 Nov 1;13(11):2498–504.

177. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. J Proteome Res. 2019 Feb 1;18(2):623–32.

178. Fabregat A, Sidiropoulos K, Viteri G, Forner O, Marin-Garcia P, Arnau V, et al. Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinformatics. 2017 Dec;18(1):1–9.

179. Mi H, Ebert D, Muruganujan A, Mills C, Albou LP, Mushayamaha T, et al. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res. 2021 Jan 8;49(D1):D394–403.

180. Ouyang G, Liu Q, Wu Y, Liu Z, Lu W, Li S, et al. The global, regional, and national burden of gallbladder and biliary tract cancer and its attributable risk factors in 195 countries and territories, 1990 to 2017: A systematic analysis for the Global Burden of Disease Study 2017. Cancer. 2021 Jul;127(13):2238–50.

181. Lazcano-Ponce EC, Miquel JF, Munoz N, Herrero R, Ferrecio C, Wistuba II, et al. Epidemiology and Molecular Pathology of Gallbladder Cancer. CA Cancer J Clin. 2001 Nov 1;51(6):349–64.

182. Rawla P, Sunkara T, Thandra KC, Barsouk A. Epidemiology of gallbladder cancer. Clin Exp Hepatol. 2019;5(2):93–102.

183. Jackson SS, Adami HO, Andreotti G, Beane-Freeman LE, Berrington de González A, Buring JE, et al. Associations between reproductive factors and biliary tract cancers in women from the Biliary Tract Cancers Pooling Project. J Hepatol. 2020 Oct;73(4):863–72.

184. Van Dyke AL, Shiels MS, Jones GS, Pfeiffer RM, Petrick JL, Beebe-Dimmer JL, et al. Biliary tract cancer incidence and trends in the United States by demographic group, 1999-2013. Cancer. 2019 May;125(9):1489–98.

185. Erlinger TP, Muntner P, Helzlsouer KJ. WBC Count and the Risk of Cancer Mortality in a National Sample of U.S. Adults: Results from the Second National Health and Nutrition Examination Survey Mortality Study. Cancer Epidemiol Biomark Prev. 2004;13(6):1052–6.

186. Heikkilä K, Ebrahim S, Lawlor DA. A systematic review of the association between circulating concentrations of C reactive protein and cancer. J Epidemiol Community Health. 2007 Sep 1;61(9):824–33.

187. Espinoza JA, Bizama C, García P, Ferreccio C, Javle M, Miquel JF, et al. The inflammatory inception of gallbladder cancer. Biochim Biophys Acta. 2016;1865(2):245–54.

188. Ji M, Du L, Ma Z, Xie J, Huang Y, Wei X, et al. Circulating C-reactive protein increases lung cancer risk: Results from a prospective cohort of UK Biobank. Int J Cancer. 2021;150(1):47–55.

189. Kemp TJ, Castro FA, Gao YT, Hildesheim A, Nogueira L, Wang BS, et al. Application of multiplex arrays for cytokine and chemokine profiling of bile. Cytokine. 2015 May;73(1):84–90.

190. Koper-Lenkiewicz OM, Dymicka-Piekarska V, Milewska AJ, Zińczuk J, Kamińska J. The Relationship between Inflammation Markers (CRP, IL-6, sCD40L) and Colorectal Cancer Stage, Grade, Size and Location. Diagnostics. 2021 Jul 31;11(8):1382.

191. Patel SH, Derweesh IH, Saito K, Patil D, Meagher MF, Bindayi A, et al. Preoperative Elevation of C-Reactive Protein Is a Predictor for Adverse Oncologic Survival Outcomes for Renal Cell Carcinoma: Analysis from the International Marker Consortium Renal Cancer (INMARC). Clin Genitourin Cancer. 2021 Aug;19(4):e206–15.

192. Gilreath JA, Stenehjem DD, Rodgers GM. Diagnosis and treatment of cancer-related anemia: Diagnosis and treatment of cancer-related anemia. Am J Hematol. 2014 Feb;89(2):203–12.

193. Saluja SS, Gulati M, Garg PK, Pal H, Pal S, Sahni P, et al. Endoscopic or Percutaneous Biliary Drainage for Gallbladder Cancer: A Randomized Trial and Quality of Life Assessment. Clin Gastroenterol Hepatol. 2008 Aug;6(8):944–50.

194. Stevenson DK, Vreman HJ, Wong RJ. Bilirubin Production and the Risk of Bilirubin Neurotoxicity. Semin Perinatol. 2011 Jun;35(3):121–6.

195. Pompella A, Visvikis A, Paolicchi A, Tata VD, Casini AF. The changing faces of glutathione, a cellular protagonist. Biochem Pharmacol. 2003 Oct;66(8):1499–503.

196. Kennedy L, Sandhu JK, Harper ME, Cuperlovic-Culf M. Role of Glutathione in Cancer: From Mechanisms to Therapies. Biomolecules. 2020 Oct 9;10(10):1429.

197. Sun L, Ke X, Wang D, Yin H, Jin B, Xu H, et al. Prognostic Value of the Albumin-to-γ-glutamyltransferase Ratio for Gallbladder Cancer Patients and Establishing a Nomogram for Overall Survival. J Cancer. 2021;12(14):4172–82.

198. Kunutsor SK, Apekey TA, Van Hemelrijck M, Calori G, Perseghin G. Gamma glutamyltransferase, alanine aminotransferase and risk of cancer: Systematic review and meta-analysis: GGT, ALT and cancer risk. Int J Cancer. 2015 Mar 1;136(5):1162–70.

199. Yuwen P, Chen W, Lv H, Feng C, Li Y, Zhang T, et al. Albumin and surgical site infection risk in orthopaedics: a meta-analysis. BMC Surg. 2017 Dec;17(1):7.

200. Giboney PT. Mildly Elevated Liver Transaminase Levels in the Asymptomatic Patient. Am Fam Physician. 2005;71(6):1105–10.

201. Næser E, Møller H, Fredberg U, Frystyk J, Vedsted P. Routine blood tests and probability of cancer in patients referred with non-specific serious symptoms: a cohort study. BMC Cancer. 2017 Dec;17(1):1–11.

202. Lv K, Yang J, Sun J, Guan J. Identification of key candidate genes for pancreatic cancer by bioinformatics analysis. Exp Ther Med. 2019;18(1):8.

203. Winkler J, Abisoye-Ogunniyan A, Metcalf KJ, Werb Z. Concepts of extracellular matrix remodelling in tumour progression and metastasis. Nat Commun. 2020 Dec;11(1):5120.

204. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. EMBO Rep. 2014 Dec;15(12):1243–53.

205. González-González L, Alonso J. Periostin: A Matricellular Protein With Multiple Functions in Cancer Development and Progression. Front Oncol. 2018 Jun 12;8:225.

206. Sainio A, Järveläinen H. Extracellular matrix-cell interactions: Focus on therapeutic applications. Cell Signal. 2020 Feb;66:109487.

207. Song Y, Wang L, Wang K, Lu Y, Zhou P. COL12A1 Acts as a Novel Prognosis Biomarker and Activates Cancer-Associated Fibroblasts in Pancreatic Cancer through Bioinformatics and Experimental Validation. Cancers. 2023 Feb 26;15(5):1480.

208. Li X, Sun X, Kan C, Chen B, Qu N, Hou N, et al. COL1A1: A novel oncogenic gene and therapeutic target in malignancies. Pathol - Res Pract. 2022 Aug;236:154013.

209. Efthymiou G, Saint A, Ruff M, Rekad Z, Ciais D, Van Obberghen-Schilling E. Shaping Up the Tumor Microenvironment With Cellular Fibronectin. Front Oncol. 2020 Apr 30;10:641.

210. Burgos-Panadero R, Noguera I, Cañete A, Navarro S, Noguera R. Vitronectin as a molecular player of the tumor microenvironment in neuroblastoma. BMC Cancer. 2019 Dec;19(1):479.

211. Rathinam R, Alahari SK. Important role of integrins in the cancer biology. Cancer Metastasis Rev. 2010 Mar;29(1):223–37.

212. Hamidi H, Ivaska J. Every step of the way: integrins in cancer progression and metastasis. Nat Rev Cancer. 2018 Sep;18(9):533–48.

213. Marsico G, Russo L, Quondamatteo F, Pandit A. Glycosylation and Integrin Regulation in Cancer. Trends Cancer. 2018 Aug;4(8):537–52.

214. Hamidi H, Pietilä M, Ivaska J. The complexity of integrins in cancer and new scopes for therapeutic targeting. Br J Cancer. 2016 Oct;115(9):1017–23.

215. Xian X, Gopal S, Couchman JR. Syndecans as receptors and organizers of the extracellular matrix. Cell Tissue Res. 2010 Jan;339(1):31–46.

216. Jang B, Kim A, Hwang J, Song HK, Kim Y, Oh ES. Emerging Role of Syndecans in Extracellular Matrix Remodeling in Cancer. J Histochem Cytochem. 2020 Dec;68(12):863–70.

217. Warburg O, Wind F, Negelein E. The Metabolism of Tumors in the Body. J Gen Physiol. 1927;8(6):519–30.

218. Fakhri S, Moradi SZ, Farzaei MH, Bishayee A. Modulation of dysregulated cancer metabolism by plant secondary metabolites: A mechanistic review. Semin Cancer Biol. 2022 May;80:276–305.

219. Benjamin DI, Cravatt BF, Nomura DK. Global Profiling Strategies for Mapping Dysregulated Metabolic Pathways in Cancer. Cell Metab. 2012 Nov;16(5):565–77.

220. Robey RB, Weisz J, Kuemmerle NB, Salzberg AC, Berg A, Brown DG, et al. Metabolic reprogramming and dysregulated metabolism: cause, consequence and/or enabler of environmental carcinogenesis? Carcinogenesis. 2015 Jun;36(Suppl 1):S203–31.

221. Sreedhar A, Zhao Y. Dysregulated metabolic enzymes and metabolic reprogramming in cancer cells (Review). Biomed Rep. 2018;8(1):3–10.

222. Fernandez-Rojo MA, Ramm GA. Caveolin-1 Function in Liver Physiology and Disease. Trends Mol Med. 2016 Oct;22(10):889–904.

223. Houten SM, Denis S, Argmann CA, Jia Y, Ferdinandusse S, Reddy JK, et al. Peroxisomal L-bifunctional enzyme (Ehhadh) is essential for the production of medium-chain dicarboxylic acids. J Lipid Res. 2012 Jul;53(7):1296–303.

224. Liu RZ, Godbout R. An Amplified Fatty Acid-Binding Protein Gene Cluster in Prostate Cancer: Emerging Roles in Lipid Metabolism and Metastasis. Cancers. 2020 Dec 18;12(12):3823.

225. Cortés VA, Barrera F, Nervi F. Pathophysiological connections between gallstone disease, insulin resistance, and obesity. Obes Rev. 2020 Apr;21(4):e12983.

226. Luo J, Yang H, Song BL. Mechanisms and regulation of cholesterol homeostasis. Nat Rev Mol Cell Biol. 2020 Apr;21(4):225–45.

227. Bian X, Liu R, Meng Y, Xing D, Xu D, Lu Z. Lipid metabolism and cancer. J Exp Med. 2021 Jan 4;218(1):e20201606.

228. Xiao X, Tang JJ, Peng C, Wang Y, Fu L, Qiu ZP, et al. Cholesterol Modification of Smoothened Is Required for Hedgehog Signaling. Mol Cell. 2017 Apr;66(1):154-162.e10.

229. Chen P, Wang Y, Li J, Bo X, Wang J, Nan L, et al. Diversity and intratumoral heterogeneity in human gallbladder cancer progression revealed by single-cell RNA sequencing. Clin Transl Med. 2021 Jun;11(6):e462.

230. Ebata N, Fujita M, Sasagawa S, Maejima K, Okawa Y, Hatanaka Y, et al. Molecular Classification and Tumor Microenvironment Characterization of Gallbladder Cancer by Comprehensive Genomic and Transcriptomic Analysis. Cancers. 2021 Feb 10;13(4):733.

231. Gether IM, Nexøe-Larsen C, Knop FK. New Avenues in the Regulation of Gallbladder Motility—Implications for the Use of Glucagon-Like Peptide–Derived Drugs. J Clin Endocrinol Metab. 2019 Jul 1;104(7):2463–72.

232. Portincasa P, Di Ciaula A, vanBerge-Henegouwen GP. Smooth muscle function and dysfunction in gallbladder disease. Curr Gastroenterol Rep. 2004 Apr;6(2):151–62.

233. Xu Q, Shaffer E. The potential site of impaired gallbladder contractility in an animal model of cholesterol gallstone disease. Gastroenterology. 1996 Jan;110(1):251–7.

234. Faridi MS, Jaiswal MSD, Goel SK. Expression of CCK Receptors in Carcinoma Gallbladder and Cholelithiasis: A Pilot Study. J Clin Diagn Res. 2015;9(7):PC04.

235. Chatterjee S. Chapter Two - Oxidative Stress, Inflammation, and Disease. In: Oxidative Stress and Biomaterials. Academic Press; 2016. p. 35–58.

236. Frijhoff J, Winyard PG, Zarkovic N, Davies SS, Stocker R, Cheng D, et al. Clinical Relevance of Biomarkers of Oxidative Stress. Antioxid Redox Signal. 2015 Nov 10;23(14):1144–70.

237. Zhang X, Saarinen AM, Hitosugi T, Wang Z, Wang L, Ho TH, et al. Inhibition of intracellular lipolysis promotes human cancer cell adaptation to hypoxia. eLife. 2017 Dec 19;6:e31132.

238. Doran MH, Pavadai E, Rynkiewicz MJ, Walklate J, Bullitt E, Moore JR, et al. Cryo-EM and Molecular Docking Shows Myosin Loop 4 Contacts Actin and Tropomyosin on Thin Filaments. Biophys J. 2020 Aug;119(4):821–30.

239. Ji Q, Chang L, VanDenBerg D, Stanczyk FZ, Stolz A. Selective reduction of AKR1C2 in prostate cancer and its role in DHT metabolism. The Prostate. 2003 Mar 1;54(4):275–89.

240. Kaczocha M, Vivieca S, Sun J, Glaser ST, Deutsch DG. Fatty Acid-binding Proteins Transport N-Acylethanolamines to Nuclear Receptors and Are Targets of Endocannabinoid Transport Inhibitors. J Biol Chem. 2012 Jan;287(5):3415–24.

241. Cao W, Liu N, Tang S, Bao L, Shen L, Yuan H, et al. Acetyl-Coenzyme A acyltransferase 2 attenuates the apoptotic effects of BNIP3 in two human cell lines. Biochim Biophys Acta BBA - Gen Subj. 2008 Jun;1780(6):873–80.

242. Aleman JD, Young CD, Karam SD, Wang X. Revisiting laminin and extracellular matrix remodeling in metastatic squamous cell carcinoma: What have we learned after more than four decades of research? Mol Carcinog. 2023 Jan;62(1):5–23.

243. Smolina N, Khudiakov A, Knyazeva A, Zlotina A, Sukhareva K, Kondratov K, et al. Desmin mutations result in mitochondrial dysfunction regardless of their aggregation properties. Biochim Biophys Acta BBA - Mol Basis Dis. 2020 Jun;1866(6):165745.

244. Chow RKK, Tsz-Kwan Sin S, Liu M, Li Y, Man Chan TH, Song Y, et al. AKR7A3 suppresses tumorigenicity and chemoresistance in hepatocellular carcinoma through attenuation of ERK, c-Jun and NF-κB signaling pathways. Oncotarget. 2017 Oct 13;8(48):83469–79.

245. Allocati N, Masulli M, Di Ilio C, Federici L. Glutathione transferases: substrates, inihibitors and pro-drugs in cancer and neurodegenerative diseases. Oncogenesis. 2018 Jan 24;7(1):8.

246. Qian Y, Markowitz JS. Natural Products as Modulators of CES1 Activity. Drug Metab Dispos. 2020 Oct;48(10):993–1007.

247. Almazroo OA, Miah MK, Venkataramanan R. Drug Metabolism in the Liver. Clin Liver Dis. 2017 Feb;21(1):1–20.

248. Newton K, Dixit VM. Signaling in Innate Immunity and Inflammation. Cold Spring Harb Perspect Biol. 2012 Mar 1;4(3):a006049.

249. Barton GM. A calculated response: control of inflammation by the innate immune system. J Clin Invest. 2008 Feb 1;118(2):413–20.

250. Li X, Ye Y, Peng K, Zeng Z, Chen L, Zeng Y. Histones: The critical players in innate immunity. Front Immunol [Internet]. 2022 Nov 21 [cited 2023 Mar 10];13. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2022.1030610/full

251. Wright HL, Moots RJ, Bucknall RC, Edwards SW. Neutrophil function in inflammation and inflammatory diseases. Rheumatology. 2010 Sep 1;49(9):1618–31.

252. Chang CW, Chang WH, Lin CC, Chu CH, Wang TE, Shih SC. Acute transient hepatocellular injury in cholelithiasis and cholecystitis without evidence of choledocholithiasis. World J Gastroenterol. 2009;15(30):3788–92.

253. Takamatsu H, Takegahara N, Nakagawa Y, Tomura M, Taniguchi M, Friedel RH, et al. Semaphorins guide the entry of dendritic cells into the lymphatics by activating myosin II. Nat Immunol. 2010 Jul;11(7):594–600.

254. Liu WN, Yan M, Chan AM. A thirty-year quest for a role of R-Ras in cancer: from an oncogene to a multitasking GTPase. Cancer Lett. 2017 Sep;403:59–65.

255. Liu Y, Cheng G, Song Z, Xu T, Ruan H, Cao Q, et al. RAC2 acts as a prognostic biomarker and promotes the progression of clear cell renal cell carcinoma. Int J Oncol. 2019 Jul 26;55(3):645–56.

256. Blair P, Flaumenhaft R. Platelet α-granules: Basic biology and clinical correlates. Blood Rev. 2009 Jul;23(4):177–89.

257. May AE, Seizer P, Gawaz M. Platelets: Inflammatory Firebugs of Vascular Walls. Arterioscler Thromb Vasc Biol. 2008 Mar;28(3):s5–10.

258. Gleissner CA, von Hundelshausen P, Ley K. Platelet Chemokines in Vascular Disease. Arterioscler Thromb Vasc Biol. 2008 Nov;28(11):1920–7.

259. Goodarzi A, Yari F, Mohammadipour M, Deyhim MR, Naghadeh HT. Capability of Platelet Factor 4 to Induce Apoptosis in the Cancerous Cell Lines in Vitro. Int J Med Lab. 2018;5(3):195–207.

260. Sabrkhany S, Kuijpers MJE, van Kuijk SMJ, Sanders L, Pineda S, Olde Damink SWM, et al. A combination of platelet features allows detection of early-stage cancer. Eur J Cancer. 2017 Jul;80:5–13.

261. Pucci F, Rickelt S, Newton AP, Garris C, Nunes E, Evavold C, et al. PF4 Promotes Platelet Production and Lung Cancer Growth. Cell Rep. 2016 Nov;17(7):1764–72.

262. Ercetin E, Richtmann S, Delgado BM, Gomez-Mariano G, Wrenger S, Korenbaum E, et al. Clinical Significance of SERPINA1 Gene and Its Encoded Alpha1-antitrypsin Protein in NSCLC. Cancers. 2019 Sep 4;11(9):1306.

263. Falanga A, Marchetti M, Vignoli A. Coagulation and cancer: biological and clinical aspects. J Thromb Haemost. 2013 Feb;11(2):223–33.

264. Sang Y, Roest M, de Laat B, de Groot PG, Huskens D. Interplay between platelets and coagulation. Blood Rev. 2021 Mar;46:100733.

265. Garnier D, Magnus N, D'Asti E, Hashemi M, Meehan B, Milsom C, et al. Genetic pathways linking hemostasis and cancer. Thromb Res. 2012 Apr;129:S22–9.

266. Xu WY, Zhang HH, Yang XB, Bai Y, Lin JZ, Long JY, et al. Prognostic significance of combined preoperative fibrinogen and CA199 in gallbladder cancer patients. World J Gastroenterol. 2018 Apr 7;24(13):1451–63.

267. Mishalian I, Granot Z, Fridlender ZG. The diversity of circulating neutrophils in cancer. Immunobiology. 2017 Jan;222(1):82–8.

268. Janciauskiene S, Wrenger S, Immenschuh S, Olejnicka B, Greulich T, Welte T, et al. The Multifaceted Effects of Alpha1-Antitrypsin on Neutrophil Functions. Front Pharmacol. 2018 Apr 17;9:341.

269. Sehgal R, Kaur N, Maiwall R, Ramakrishna G, Maras JS, Trehanpati N. Plasma Proteomic Analysis Identified Proteins Associated with Faulty Neutrophils Functionality in Decompensated Cirrhosis Patients with Sepsis. Cells. 2022 May 25;11(11):1745.

270. Nolen BM, Lokshin AE. Protein biomarkers of ovarian cancer: the forest and the trees. Future Oncol. 2012 Jan;8(1):55–71.

271. Zuo M, Rashid A, Wang Y, Jain A, Li D, Behari A, et al. RNA sequencing-based analysis of gallbladder cancer reveals the importance of the liver X receptor and lipid metabolism in gallbladder cancer. Oncotarget. 2016 Jun 7;7(23):35302–12.

272. Georgila K, Vyrla D, Drakos. Apolipoprotein A-I (ApoA-I), Immunity, Inflammation and Cancer. Cancers. 2019 Aug 1;11(8):1097.

273. Mangaraj M, Nanda R, Panda S. Apolipoprotein A-I: A Molecule of Diverse Function. Indian J Clin Biochem. 2016 Jul;31(3):253–9.

274. Cochran BJ, Ong KL, Manandhar B, Rye KA. APOA1: a Protein with Multiple Therapeutic Functions. Curr Atheroscler Rep. 2021 Mar;23(3):1–10.

275. Humphries JM, Penno MAS, Weiland F, Klingler-Hoffmann M, Zuber A, Boussioutas A, et al. Identification and validation of novel candidate protein biomarkers for the detection of human gastric cancer. Biochim Biophys Acta BBA - Proteins Proteomics. 2014 May;1844(5):1051–8.

276. Hachem H, Favre G, Ghalim N, Puchois P, Fruchart J, Soula G. Quantitative abnormalities of lipoprotein particles in multiple myeloma. J Clin Chem Clin Biochem. 1987;25:675–9.

277. Honda K, Kobayashi M, Okusaka T, Rinaudo JA, Huang Y, Marsh T, et al. Plasma biomarker for detection of early stage pancreatic cancer and risk factors for pancreatic malignancy using antibodies for apolipoprotein-AII isoforms. Sci Rep. 2015 Dec;5(1):15921.

278. Feingold KR. Lipid and Lipoprotein Metabolism. Endocrinol Metab. 2022;51(3):437–58.

279. Tavazoie MF, Pollack I, Tanqueco R, Ostendorf BN, Reis BS, Gonsalves FC, et al. LXR/ApoE Activation Restricts Innate Immune Suppression in Cancer. Cell. 2018 Feb;172(4):825-840.e18.

280. Chen YC, Pohl G, Wang TL, Morin PJ, Risberg B, Kristensen GB, et al. Apolipoprotein E Is Required for Cell Proliferation and Survival in Ovarian Cancer. Cancer Res. 2005 Jan 1;65(1):331–7.

281. Venanzoni MC, Giunta S, Muraro GB, Storari L, Crescini C, Mazzucchelli R, et al. Apolipoprotein E expression in localized prostate cancers. Int J Oncol. 2003;22(4):779–86.

282. Sakashita K, Tanaka F, Zhang X, Mimori K, Kamohara Y, Inoue H, et al. Clinical significance of ApoE expression in human gastric cancer. Oncol Rep. 2008;20(6):1313–9.

283. Karunanithi S, Levi L, DeVecchio J, Karagkounis G, Reizes O, Lathia JD, et al. RBP4-STRA6 Pathway Drives Cancer Stem Cell Maintenance and Mediates High-Fat Diet-Induced Colon Carcinogenesis. Stem Cell Rep. 2017 Aug;9(2):438–50.

284. Lorkova L, Pospisilova J, Lacheta J, Leahomschi S, Zivny J, Cibula D, et al. Decreased concentrations of retinol-binding protein 4 in sera of epithelial ovarian cancer patients: A potential biomarker identified by proteomics. Oncol Rep. 2012;27(2):318–24.

285. Steinhoff JS, Lass A, Schupp M. Biological Functions of RBP4 and Its Relevance for Human Diseases. Front Physiol. 2021 Mar 11;12:659977.

286. Kozak KR, Su F, Whitelegge JP, Faull K, Reddy S, Farias-Eisner R. Characterization of serum biomarkers for detection of early stage ovarian cancer. PROTEOMICS. 2005 Oct 20;5(17):4589–96.

287. van Bennekum AM, Wei S, Gamble MV, Vogel S, Piantedosi R, Gottesman M, et al. Biochemical Basis for Depressed Serum Retinol Levels in Transthyretin-deficient Mice. J Biol Chem. 2001 Jan;276(2):1107–13.

288. Wei S, Episkopou V, Piantedosi R, Maeda S, Shimada K, Gottesman ME, et al. Studies on the Metabolism of Retinol and Retinol-binding Protein in Transthyretin-deficient Mice Produced by Homologous Recombination. J Biol Chem. 1995 Jan;270(2):866–70.

289. Liu L, Liu J, Dai S, Wang X, Wu S, Wang J, et al. Reduced transthyretin expression in sera of lung cancer. Cancer Sci. 2007 Oct;98(10):1617–24.

290. Fiorito V, Tolosano E. Hemopexin and Cancer. Int J Mol Sci. 2022 Jan 17;23(2):997.

291. Canesin G, Di Ruscio A, Li M, Ummarino S, Hedblom A, Choudhury R, et al. Scavenging of Labile Heme by Hemopexin Is a Key Checkpoint in Cancer Growth and Metastases. Cell Rep. 2020 Sep;32(12):108181.

292. Kang N, Qiu W, Wang B, Tang D, Shen X. Role of hemoglobin alpha and hemoglobin beta in non-small-cell lung cancer based on bioinformatics analysis. Mol Carcinog. 2022 Jun;61(6):587–602.

293. Rai A, Tewari M, Mohapatra SC, Shukla HS. Correlation of nutritional parameters of gallbladder cancer patients. J Surg Oncol. 2006 Jun 15;93(8):705–8.

294. Tey SK, Wong SWK, Chan JYT, Mao X, Ng TH, Yeung CLS, et al. Patient pIgR-enriched extracellular vesicles drive cancer stemness, tumorigenesis and metastasis in hepatocellular carcinoma. J Hepatol. 2022 Apr;76(4):883–95.

295. Liu Y, Hu Y, Deng L. The Underlying Roles of Exosome-Associated PIGR in Fatty Acid Metabolism and Immune Signaling in Colorectal Cancer. J Oncol. 2022 Sep 15;2022:1–12.

296. Berntsson J, Lundgren S, Nodin B, Uhlén M, Gaber A, Jirström K. Expression and prognostic significance of the polymeric immunoglobulin receptor in epithelial ovarian cancer. J Ovarian Res. 2014 Dec;7(1):1–8.

297. Ohkuma R, Yada E, Ishikawa S, Komura D, Kubota Y, Hamada K, et al. High expression levels of polymeric immunoglobulin receptor are correlated with chemoresistance and poor prognosis in pancreatic cancer. Oncol Rep. 2020 May 11;44(1):252–62.

298. Yue X, Ai J, Xu Y, Chen Y, Huang M, Yang X, et al. Polymeric immunoglobulin receptor promotes tumor growth in hepatocellular carcinoma. Hepatology. 2017 Jun;65(6):1948–62.

299. Zhang Y, Zhang J, Chen X, Yang Z. Polymeric immunoglobulin receptor (PIGR) exerts oncogenic functions via activating ribosome pathway in hepatocellular carcinoma. Int J Med Sci. 2021;18(2):364–71.

300. Buchwald H, O'Dea TJ, Menchaca HJ, Michalek VN, Rohde TD. Effect Of Plasma Cholesterol On Red Blood Cell Oxygen Transport. Clin Exp Pharmacol Physiol. 2000 Dec;27(12):951–5.

301. Andreotti G, Chen J, Gao YT, Rashid A, Chang SC, Shen MC, et al. Serum lipid levels and the risk of biliary tract cancers and biliary stones: A population-based study in China. Int J Cancer. 2008;122(10):2322–9.

302. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M. Plasma Proteome Profiling to Assess Human Health and Disease. Cell Syst. 2016 Mar;2(3):185–95.

303. Hsing AW, Gao YT, Han TQ, Rashid A, Sakoda LC, Wang BS, et al. Gallstones and the risk of biliary tract cancer: a population-based study in China. Br J Cancer. 2007 Dec;97(11):1577–82.

304. Wang S, Song R, Wang Z, Jing Z, Wang S, Ma J. S100A8/A9 in Inflammation. Front Immunol. 2018 Jun 11;9:1298.

305. Khorrami S, Tavakoli M, Safari E. Clinical Value of Serum S100A8/A9 and CA15-3 in the Diagnosis of Breast Cancer. Iran J Pathol. 2019 Apr 1;14(2):104–12.

306. Lumsden AL, Mulugeta A, Zhou A, Hyppönen E. Apolipoprotein E (APOE) genotype-associated disease risks: a phenome-wide, registry-based, case-control study utilising the UK Biobank. eBioMedicine. 2020 Sep;59:102954.

307. Ritorto MS, Borlak J. Combined Serum and Tissue Proteomic Study Applied to a c-Myc Transgenic Mouse Model of Hepatocellular Carcinoma Identified Novel Disease Regulated Proteins Suitable for Diagnosis and Therapeutic Intervention Strategies. J Proteome Res. 2011 Jul 1;10(7):3012–30.

308. Zhao Z, Zou S, Guan X, Wang M, Jiang Z, Liu Z, et al. Apolipoprotein E Overexpression Is Associated With Tumor Progression and Poor Survival in Colorectal Cancer. Front Genet. 2018 Dec 13;9:650.

309. Jiang X, Bai XY, Li B, Li Y, Xia K, Wang M, et al. Plasma Inter-Alpha-Trypsin Inhibitor Heavy Chains H3 and H4 Serve as Novel Diagnostic Biomarkers in Human Colorectal Cancer. Dis Markers. 2019 Jul 31;2019:1–10.

310. Huang L, Chen W, Liang P, Hu W, Zhang K, Shen S, et al. Serum CYFRA 21-1 in Biliary Tract Cancers: A Reliable Biomarker for Gallbladder Carcinoma and Intrahepatic Cholangiocarcinoma. Dig Dis Sci. 2015 May;60(5):1273–83.

311. Won HS, Lee MA, Chung ES, Kim DG, You YK, Hong TH, et al. Comparison of thymidine phosphorylase expression and prognostic factors in gallbladder and bile duct cancer. BMC Cancer. 2010 Dec;10(1):1–8.

312. Priya R, Jain V, Akhtar J, Chauhan G, Sakhuja P, Goyal S, et al. Plasma-derived candidate biomarkers for detection of gallbladder carcinoma. Sci Rep. 2021 Dec;11(1):23554.

313. Priya R, Jain V, Akhtar J, Saklani N, Sakhuja P, Agarwal AK, et al. Proteomic profiling of cell line-derived extracellular vesicles to identify candidate circulatory markers for detection of gallbladder cancer. Front Oncol. 2022 Nov 23;12:1–12.

314. Akhtar J, Priya R, Jain V, Sakhuja P, Agarwal AK, Goyal S, et al. Immunoproteomics approach revealed elevated autoantibody levels against ANXA1 in early stage gallbladder carcinoma. BMC Cancer. 2020 Dec;20(1):1–13.

SUPPLEMENTARY DATA

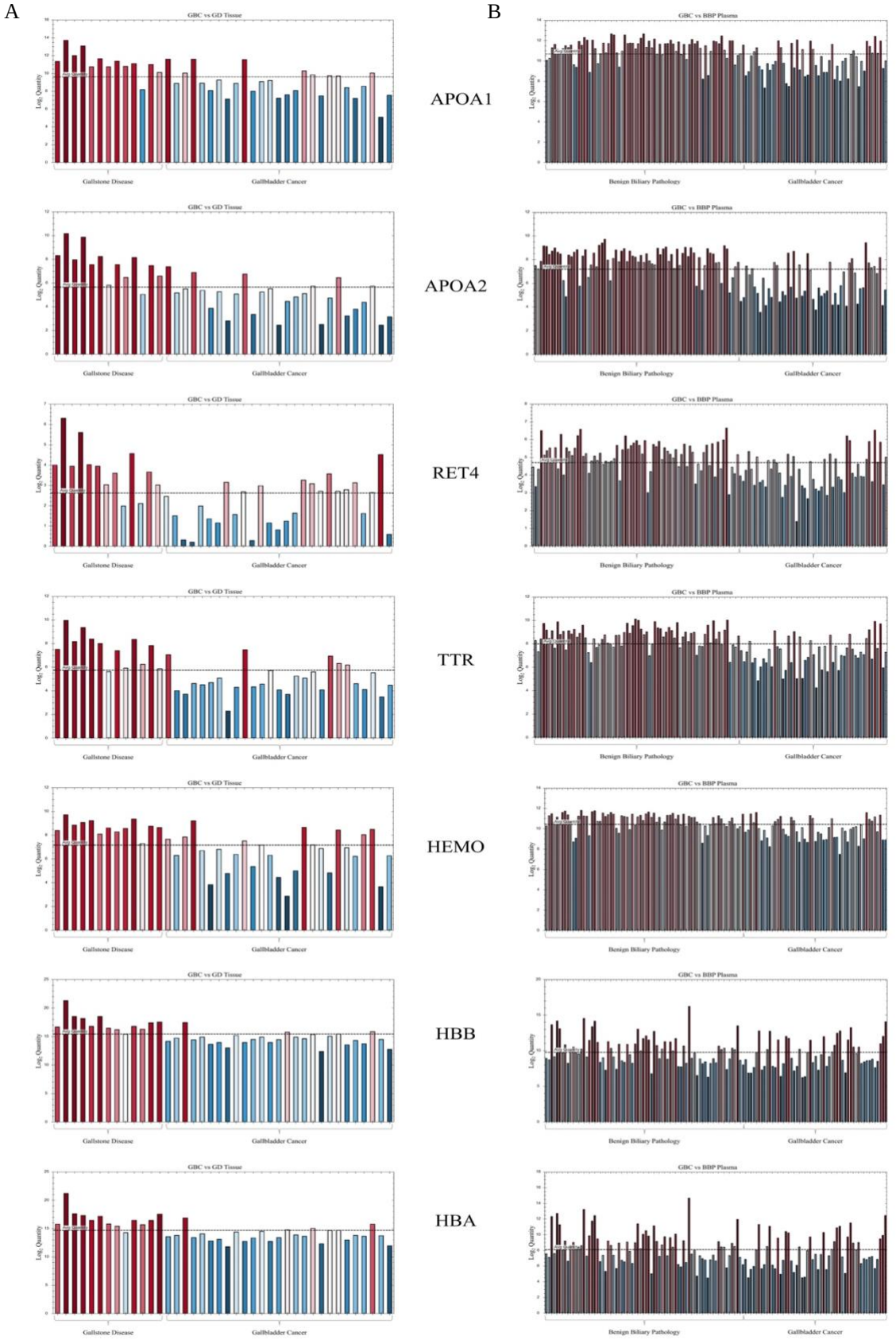


Figure S3.1: Log_2 quantities for the commonly dysregulated proteins identified in GBC/GD and GBC/BBP comparisons per patient. (A) The Log_2 quantities per patient for the GBC/GD comparison for each CDP. (B) The Log_2 quantities for each patient for the GBC/BBP plasma comparison.

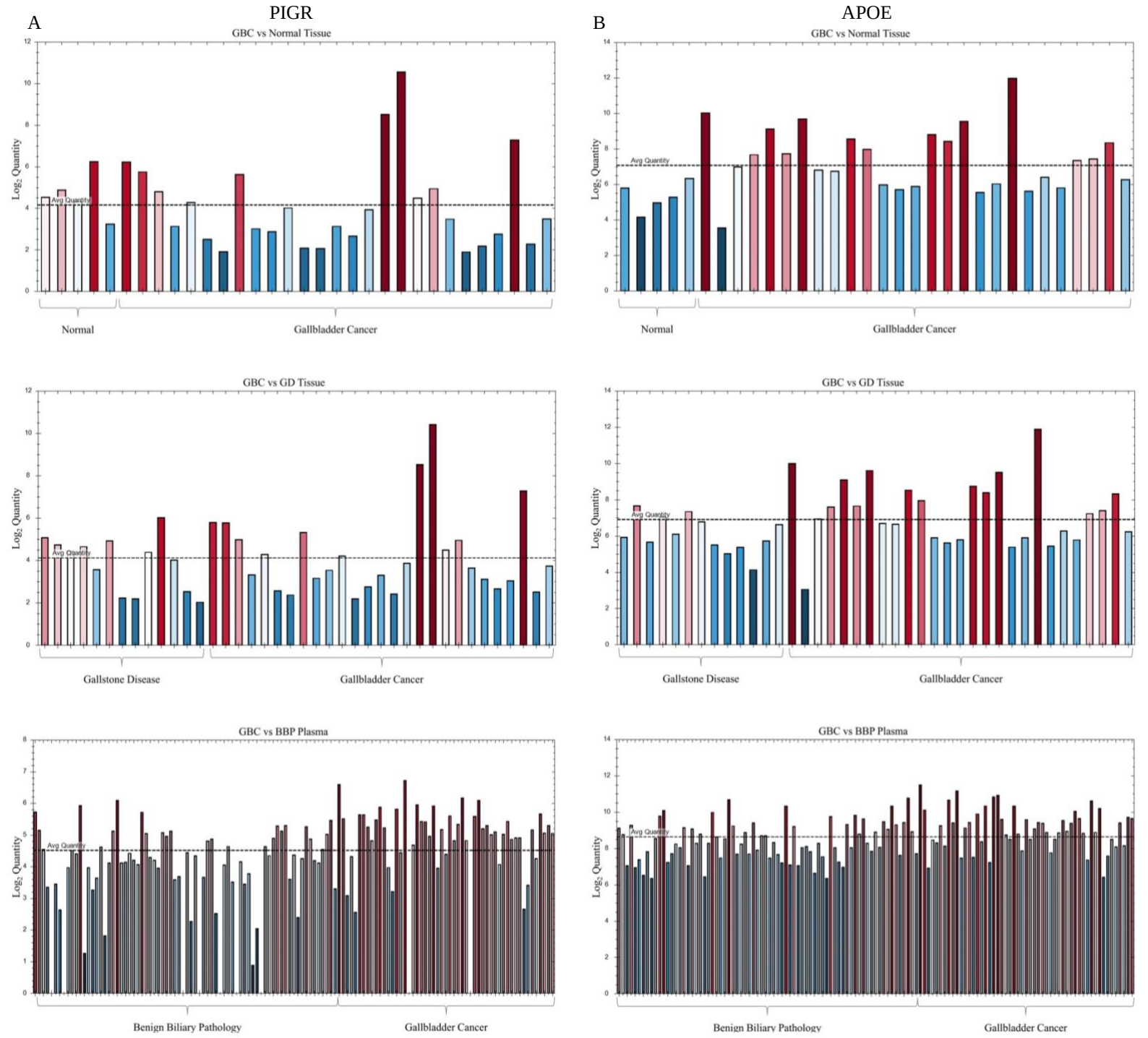
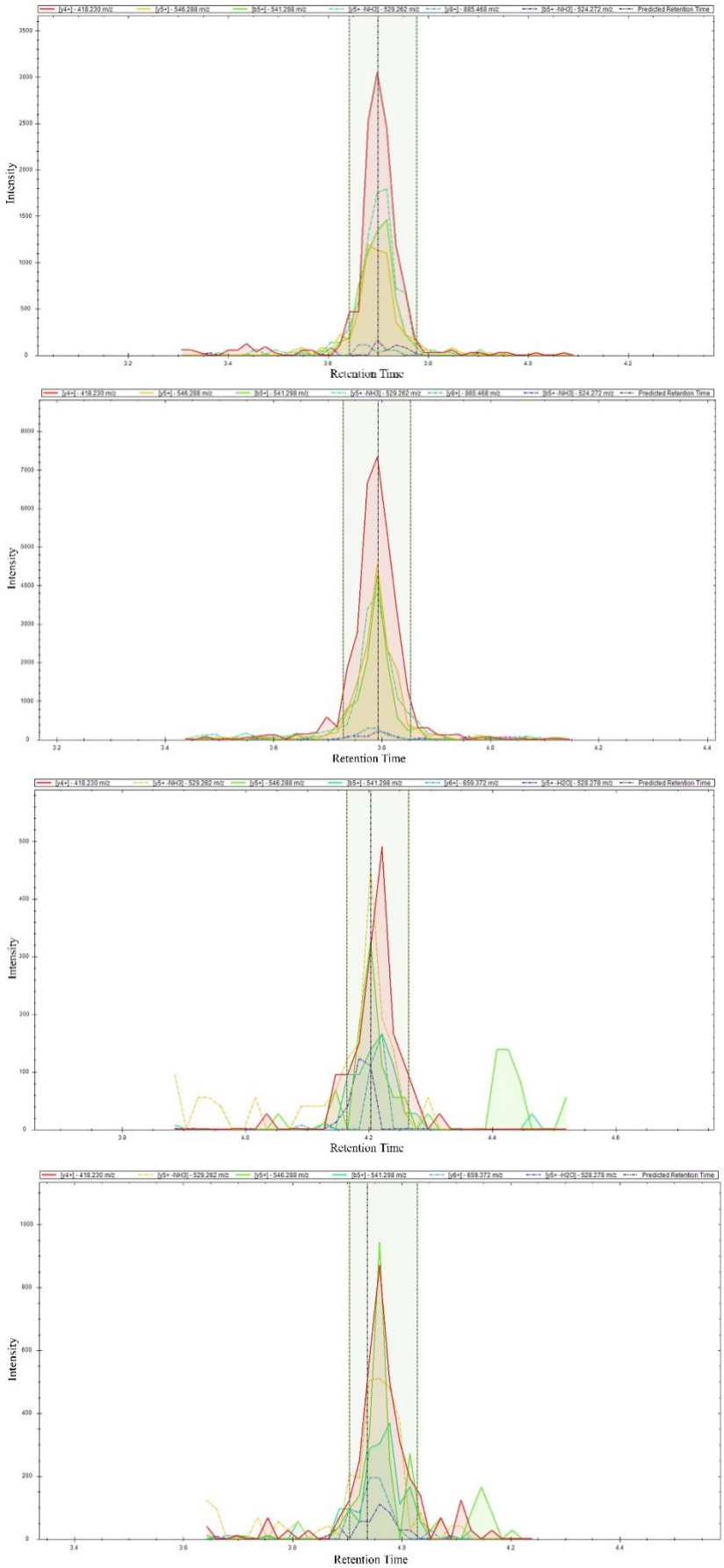


Figure S3.2: The Log_2 quantities for the commonly dysregulated proteins identified in GBC/Normal, GBC/GD, and GBC/BBP Comparisons. (A) The individual patient Log_2 quantities for PIGR in GBC/Normal, GBC/GD, and GBC/BBP comparisons. (B) The individual patient Log_2 quantities for APOE in GBC/Normal, GBC/GD, and GBC/BBP comparisons.



GBC Tissue

GD Tissue

GBC Plasma

BBP Plasma

Figure S3.3: Representative MS2 _VKSPELQAEAK_.3 precursor chromatogram overview for apolipoprotein A-II in tissue and plasma sample populations. The MS2 precursor chromatogram overview for representative samples for GBC tissue, GD tissue, GBC plasma, and BBP plasma sample populations.

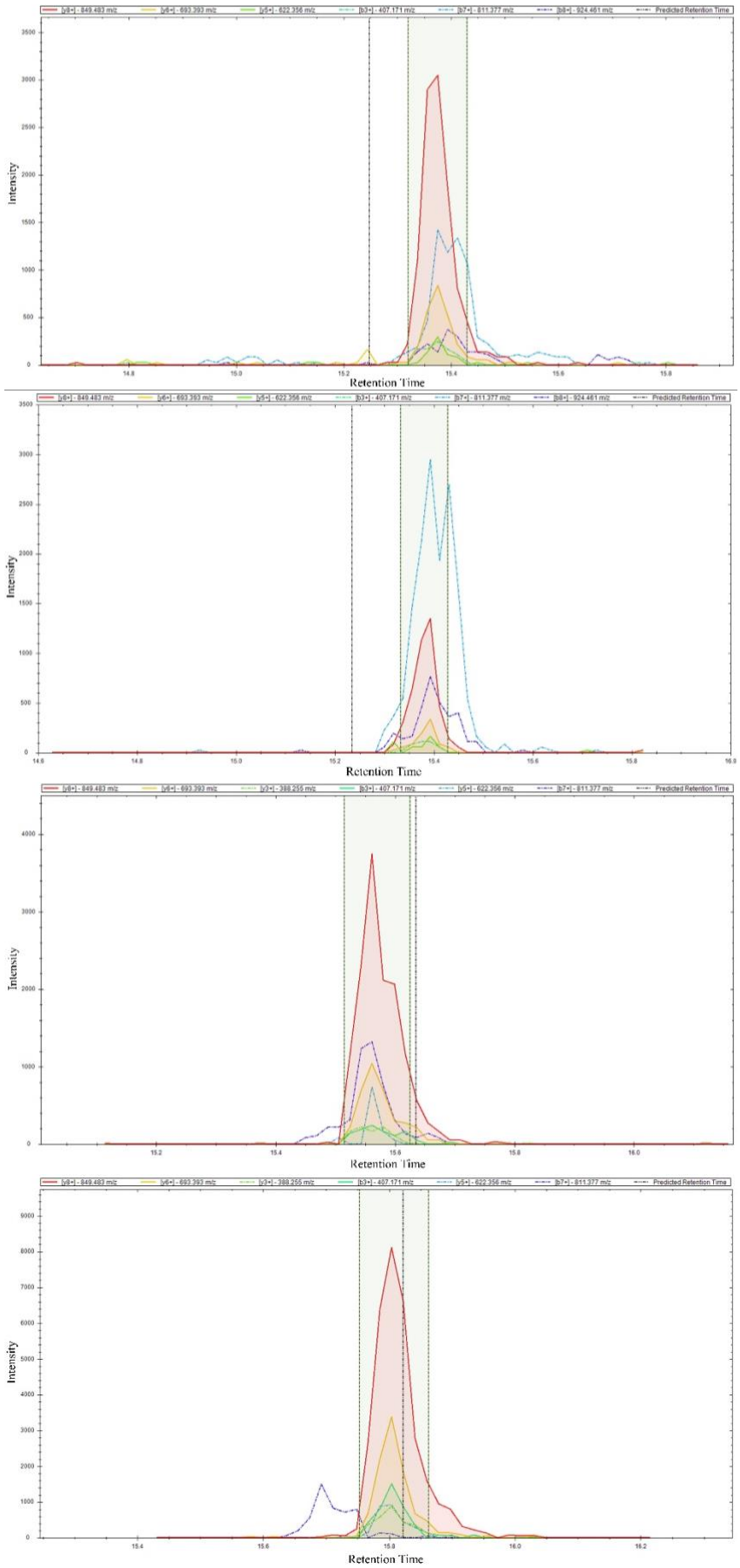
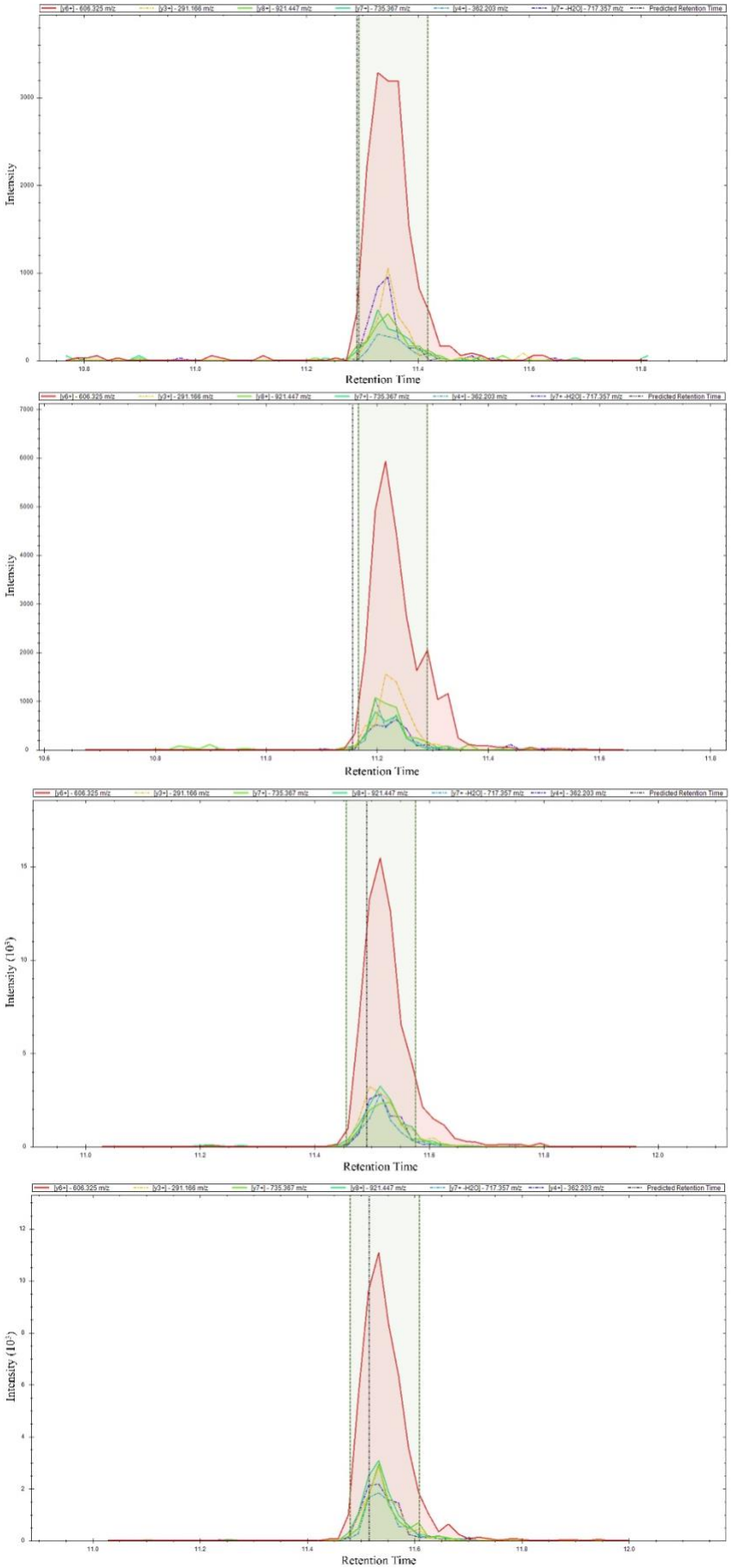


Figure S3.4: Representative MS2 _YWGVASFLQK_.2 precursor chromatogram overview for retinol-binding protein 4 in tissue and plasma sample populations. The MS2 precursor chromatogram overview for representative samples for GBC tissue, GD tissue, GBC plasma, and BBP plasma sample populations.



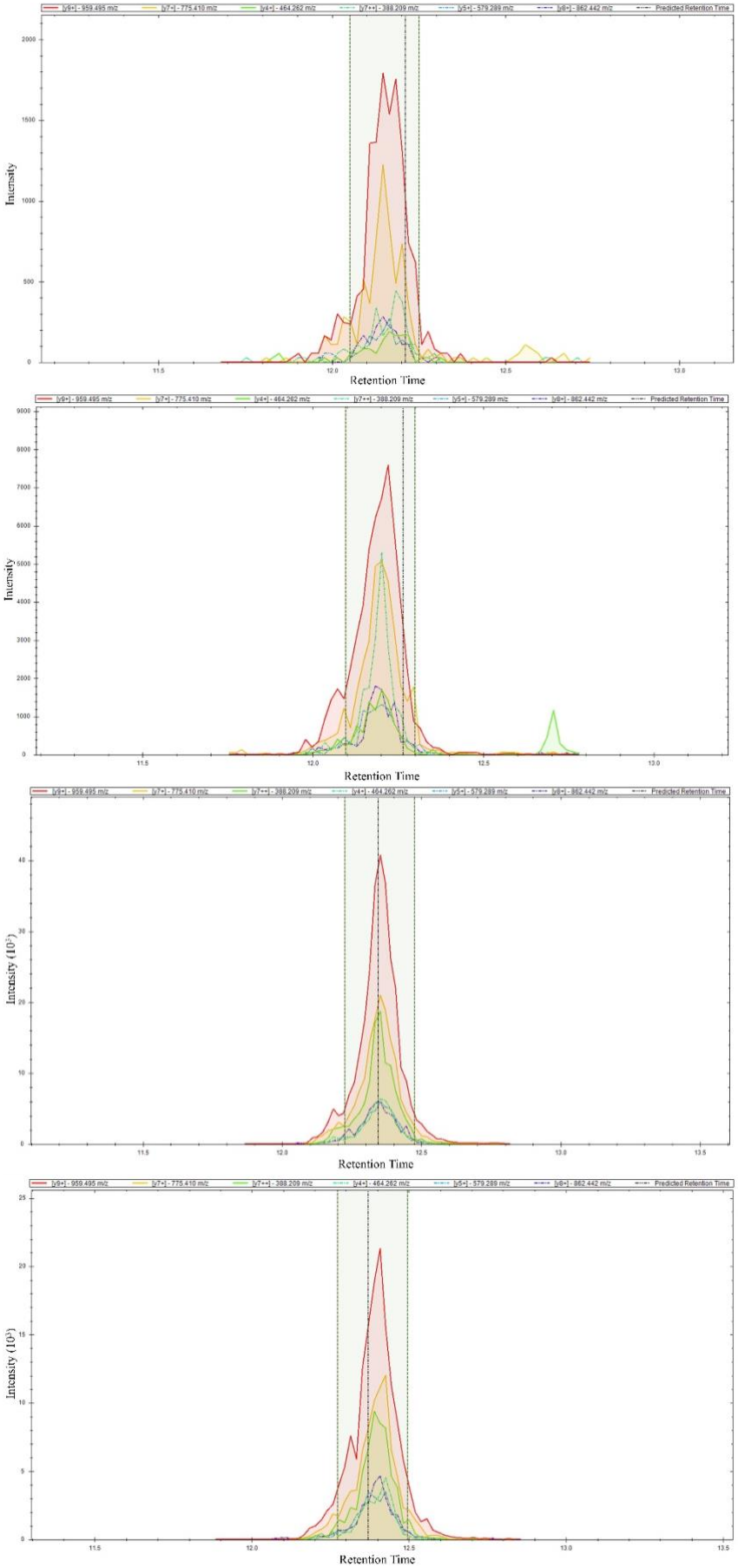
GBC Tissue

GD Tissue

GBC Plasma

BBP Plasma

Figure S3.5: Representative MS2 _AADDTWEPFASGK_.2 precursor chromatogram overview for transthyretin in tissue and plasma sample populations. The MS2 precursor chromatogram overview for representative samples for GBC tissue, GD tissue, GBC plasma, and BBP plasma sample populations.



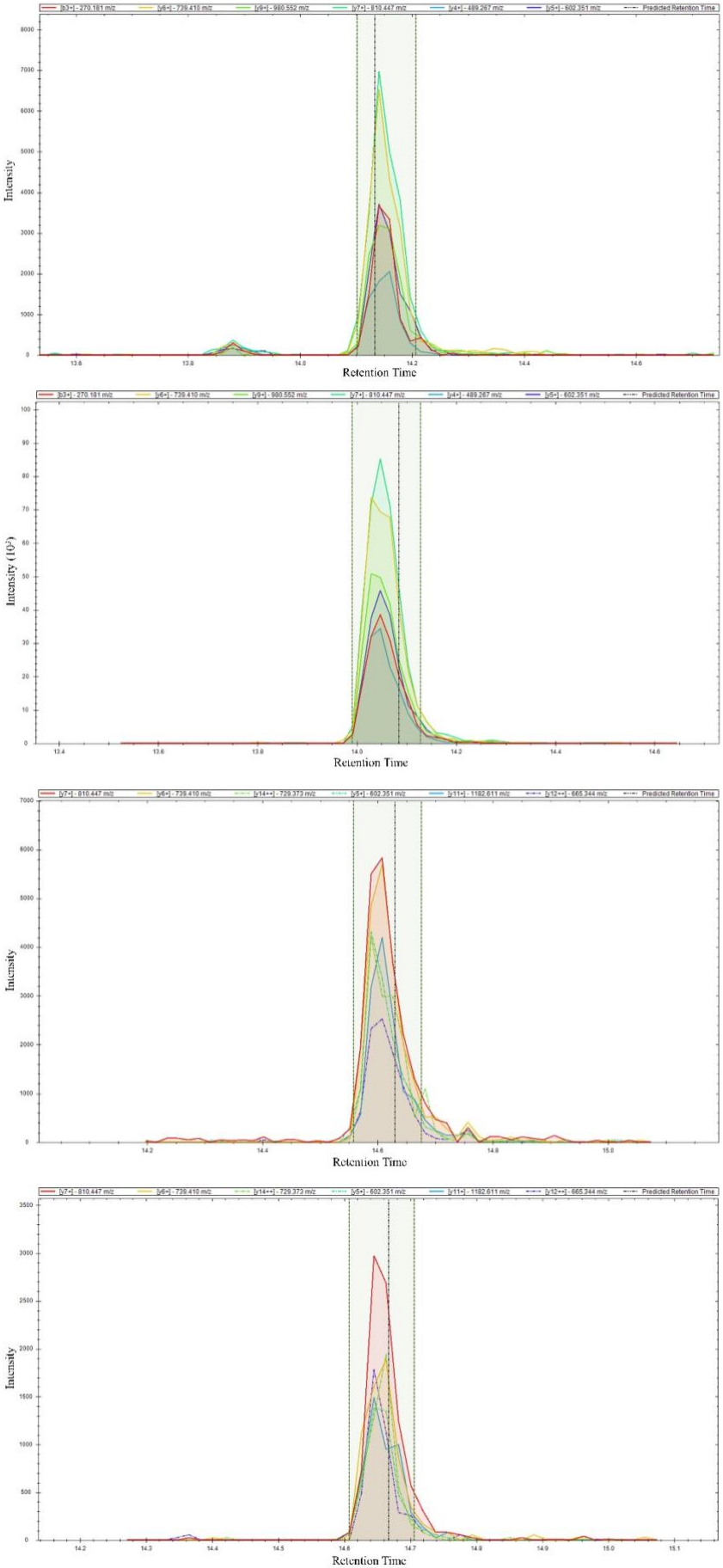
GBC Tissue

GD Tissue

GBC Plasma

BBP Plasma

Figure S3.6: Representative MS2 _NFPSPVDAAFR_.2 precursor chromatogram overview for hemopexin in tissue and plasma sample populations. The MS2 precursor chromatogram overview for representative samples for GBC tissue, GD tissue, GBC plasma, and BBP plasma sample populations.



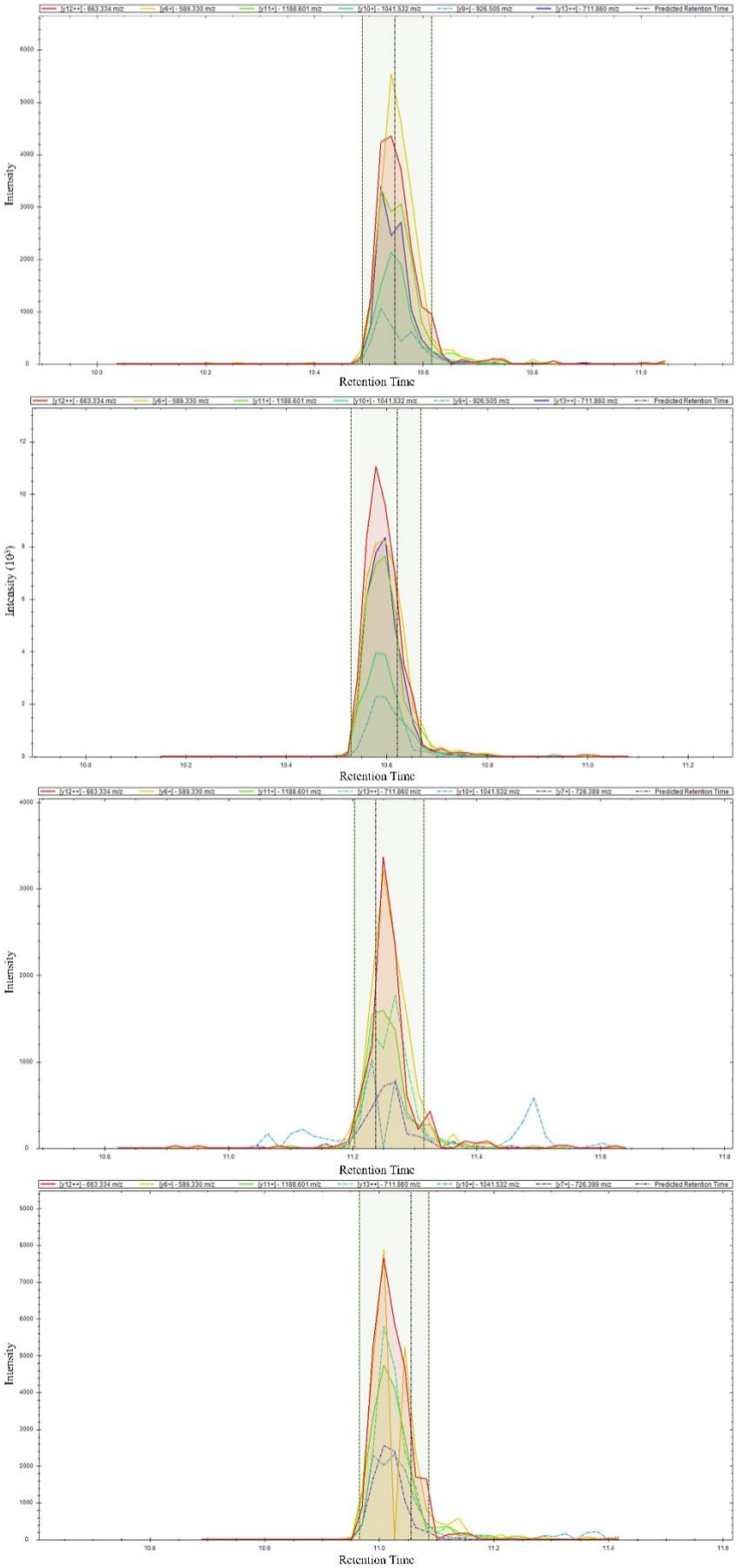
GBC Tissue

GD Tissue

GBC Plasma

BBP Plasma

Figure S3.7: Representative MS2 _VLGAFSDGLAHLDNLK_.3 precursor chromatogram overview for haemoglobin subunit B in tissue and plasma sample populations. The MS2 precursor chromatogram overview for representative samples for GBC tissue, GD tissue, GBC plasma, and BBP plasma sample populations.



GBC Tissue

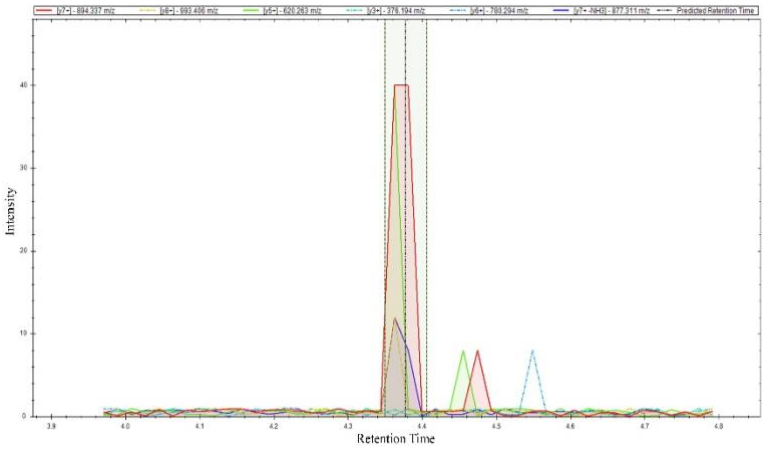
GD Tissue

GBC Plasma

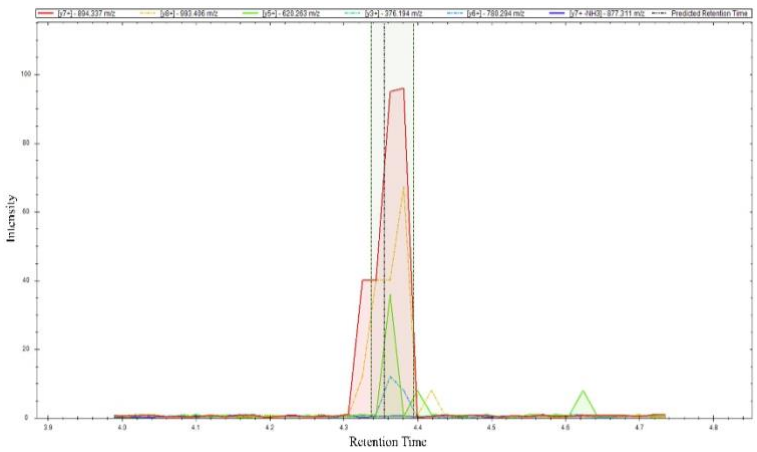
BBP Plasma

Figure S3.8: Representative MS2 _TYFPHFDLSHGSAQVK_.3 precursor chromatogram overview for haemoglobin subunit A in tissue and plasma sample populations. The MS2 precursor chromatogram overview for representative samples for GBC tissue, GD tissue, GBC plasma, and BBP plasma sample populations.

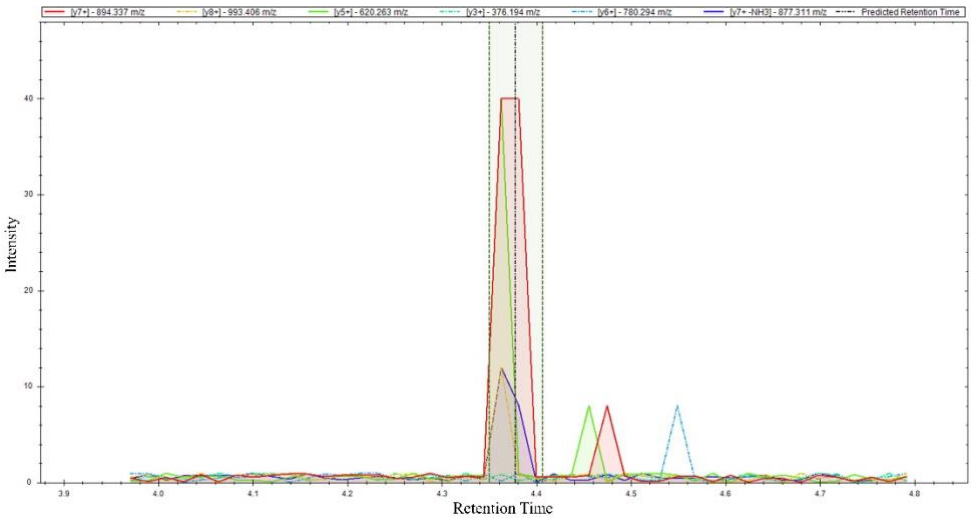
GBC Tissue



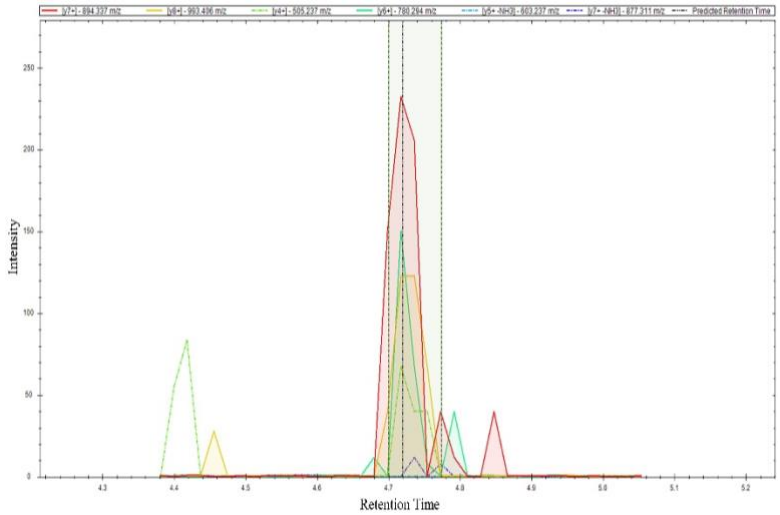
GD Tissue



Normal Tissue



GBC Plasma



BBP Plasma

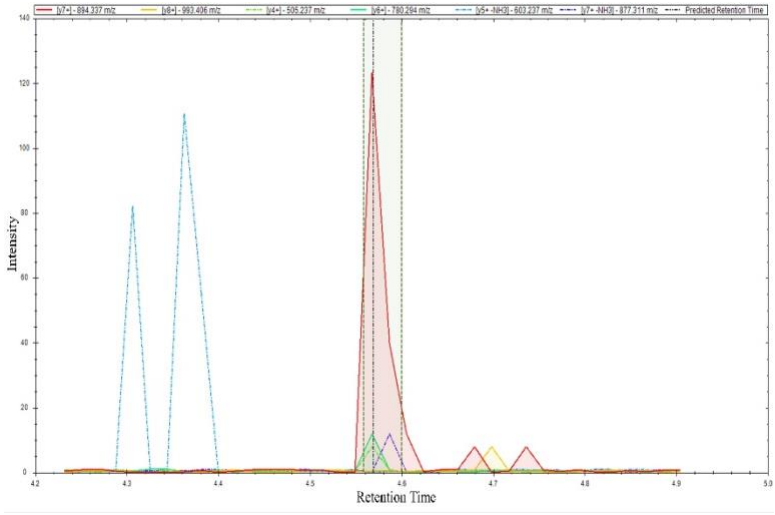


Figure S3.9: Representative MS2 _AFVNC[Carbamidomethyl (C)]DENSR_.2 precursor chromatogram overview for polymeric immunoglobulin receptor in tissue and plasma sample populations. The MS2 precursor chromatogram overview for representative samples for GBC tissue, GD tissue, GBC plasma, and BBP plasma sample populations.

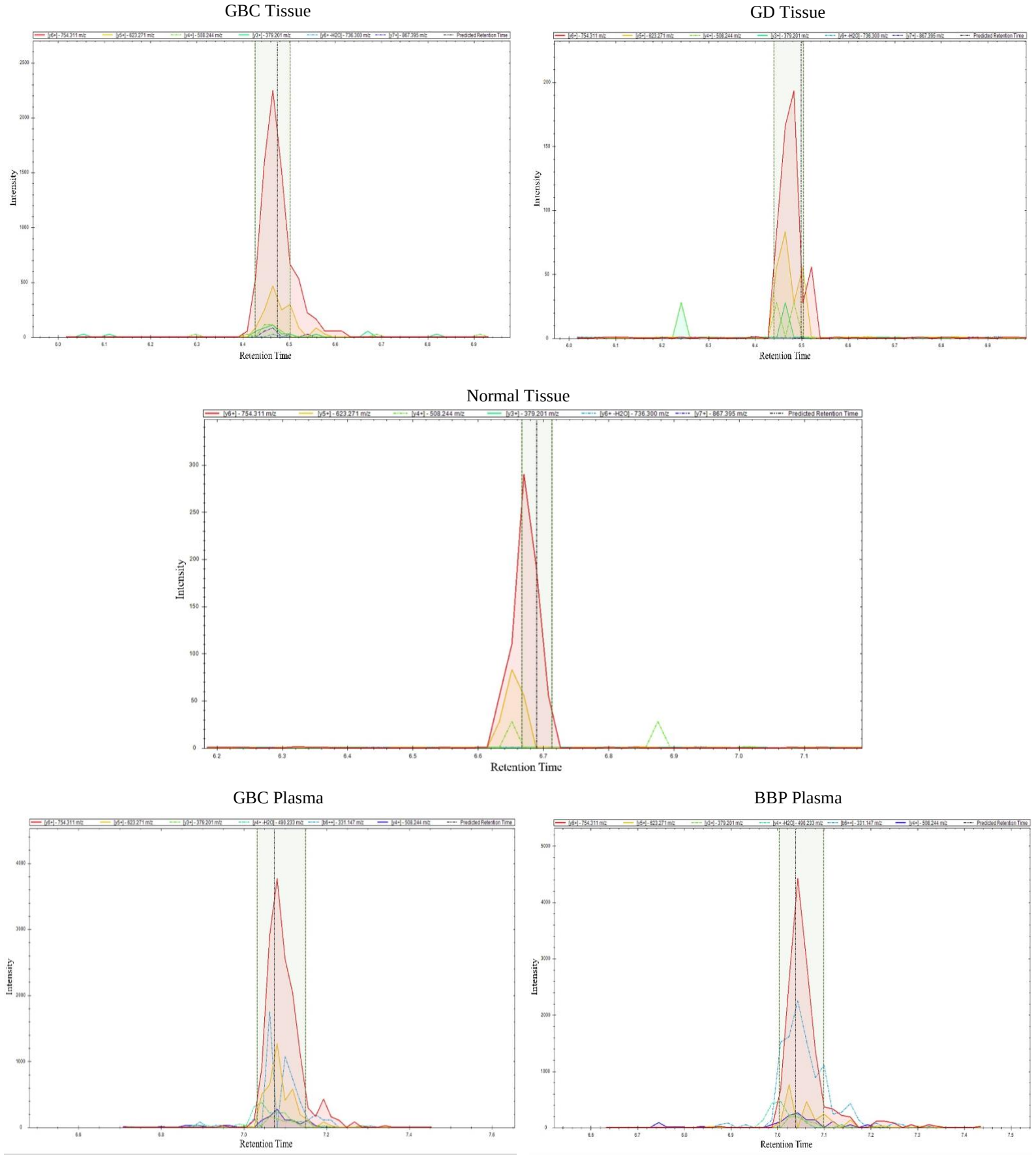


Figure S3.10: Representative MS2 _ALMDETMK_.2 precursor chromatogram overview for apolipoprotein E in tissue and plasma sample populations. The MS2 precursor chromatogram overview for representative samples for GBC tissue, GD tissue, GBC plasma, and BBP plasma sample populations.

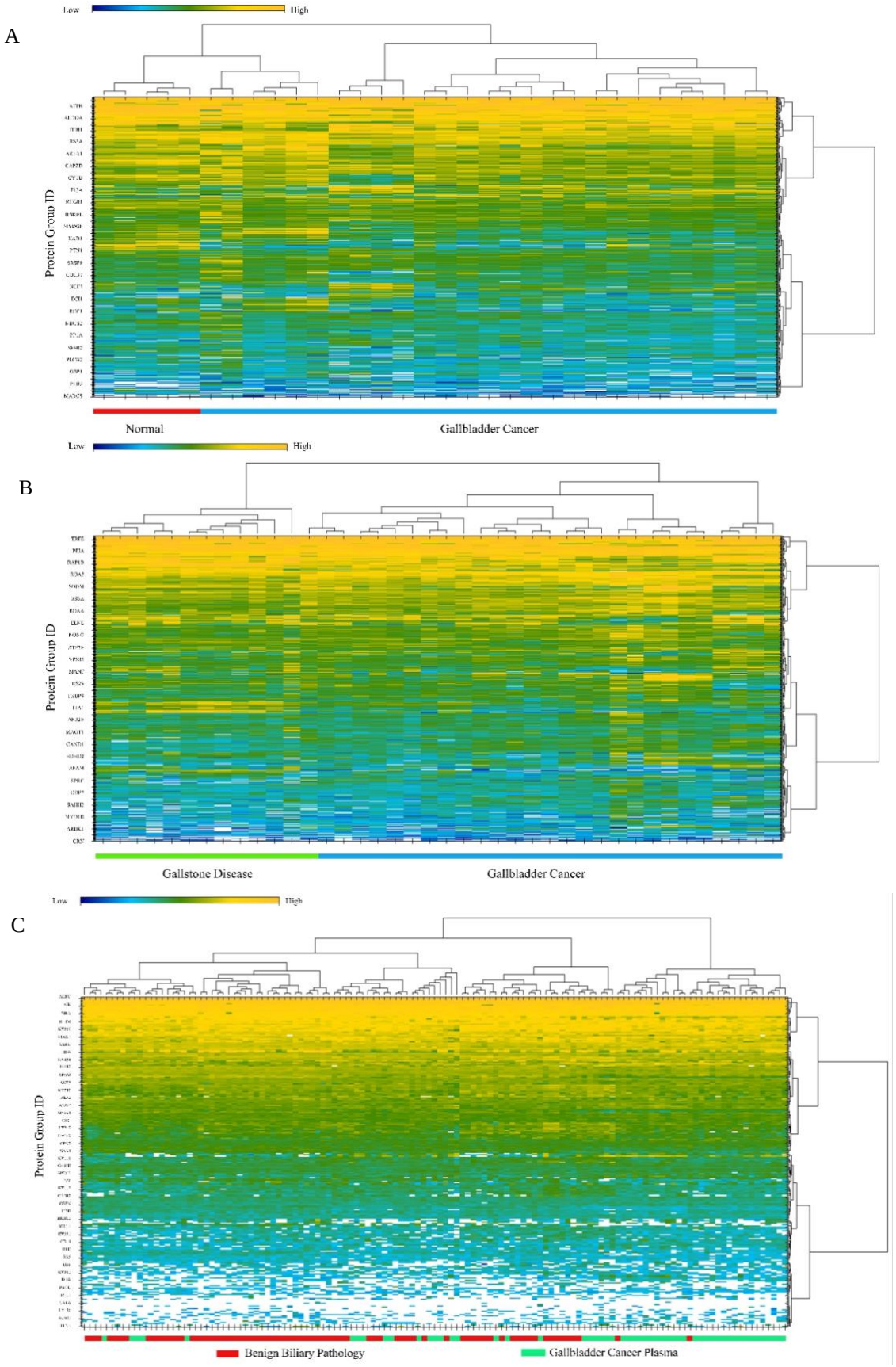


Figure S3.11: Hierarchical cluster analysis for the differentially expressed proteins. Hierarchical cluster analyses are shown in the heatmap and dendrograms for all quantified proteins in the GBC/Normal (A), GBC/GD (B), and GBC/BBP plasma (C) comparisons. The cluster brackets on the right side of the heatmaps indicate proteins clustered together based on detection intensity. The clustering brackets on the top indicate clustering based on similarity across the individual samples. The larger brackets indicate a low similarity, and the small brackets indicate a close similarity. Blue to yellow colouring indicates low to high expression of the proteins. The heatmaps were generated in Spectronaut v16.

	APOA1	APOA2	RET4	TTR	HEMO	HBB	HBA	PIGR	APOE
APOA1		3.58E-22	3.05E-06	4.11E-16	6.55E-19	0.000198	5.87E-05	1	1
APOA2	0.85834		0.005758	3.46E-09	6.20E-10	0.95427	0.59297	1	1
RET4	0.56533	0.41724		4.52E-13	9.41E-06	0.006011	0.004293	1	1
TTR	0.78791	0.65252	0.74134		9.58E-10	0.000698	0.000201	1	0.82194
HEMO	0.82435	0.67142	0.54719	0.66676		6.26E-12	2.45E-12	1	1
HBB	-0.48914	-0.25123	-0.41616	-0.46335	-0.71595		2.74E-70	1	0.95145
HBA	-0.51229	-0.27083	-0.4245	-0.48894	-0.72412	0.99282		1	0.50729
PIGR	-0.019218	-0.17453	0.039569	0.011728	-0.037669	-0.1224	-0.15513		0.084261
APOE	0.11367	-0.081096	0.082602	0.25753	0.09708	-0.25136	-0.27698	0.36054	

Figure S3.12: Spearman's Rho values and p-values for correlation of commonly dysregulated proteins. The Rho correlation values (below) and the corresponding p-values (above) for the CDPs. The yellow-coloured cells indicate significant Rho and p-values.

Table S3.1: TNM staging and gallstone disease history for gallbladder cancer plasma patients

Patient Number	TNM Staging	Gallstone History
1	T4 N2 M1	No
2	T4 N2 M1	Yes
3	Unknown	Unknown
4	T4 N2 M1	Unknown
5	Unknown	Yes
6	Unknown	Yes
7	T2a N1 M1	Yes
8	Unknown	No
9	T1 N0 M0	No
10	T3 N0 M0	Unknown
11	T3 N0 M0	Unknown
12	T3 N1 M1	No
13	T2b N1 M0	Unknown
14	T2b N1 M1	Yes
15	T3 N0 M0	Unknown
16	T3 N0 M0	No
17	T2b N1 M0	Unknown
18	T1	No
19	Unknown	Yes
20	T2b	Unknown
21	Unknown	No
22	Unknown	No
23	Unknown	Yes
24	Unknown	Unknown
25	T3 N1 M1	No
26	Unknown	Unknown
27	Unknown	Yes
28	Unknown	Yes
29	Unknown	No
30	Unknown	No
31	T4	No
32	T3 N1 Mx	No
33	T4	No
34	Unknown	Yes
35	T4	No
36	T4	No
37	Unknown	Yes

38	T4 N2 M1	No
39	T4	No
40	T1	No
42	T1	No
43	T1	No
44	T4 N2 Mx	Yes
45	T1	No
46	Unknown	No
47	T1	No
48	T4	No
49	T4	Yes
50	T4 N2 M0	No
51	T1	Yes
52	Unknown	Unknown
53	Unknown	Unknown
55	Unknown	Unknown
56	Unknown	Unknown

Table S3.2: The total precursors, peptides, and protein groups identified for each individual patient for the gallbladder cancer versus normal comparison

Sample number	Precursors	Peptides	Protein Groups
Normal 1	13 787	11 276	1 813
Normal 2	14 787	12 108	1 869
Normal 3	18 102	14 760	2 116
Normal 4	12 269	10 042	1 680
Normal 5	12 968	1 043	1 749
Gallbladder cancer (GBC) 7	10 697	8 692	1 605
GBC 8	22 169	18 111	2 349
GBC 11	12 654	10 528	1 801
GBC 12	8 768	7 202	1 320
GBC 13	16 236	12 998	1 957
GBC 14	24 683	19 538	2 413
GBC 15	22 469	17 795	2 415
GBC 16	24 670	19 762	2 457
GBC 17	18 541	14 956	2 161
GBC 18	11 367	9 482	1 674
GBC 19	19 298	15 580	2 159
GBC 21	22 882	17 873	2 258
GBC 22	23 240	18 334	2 338
GBC 24	20 129	16 169	2 145
GBC 25	21 035	16 673	2 259
GBC 26	21 206	16 754	2 192
GBC 27	12 779	10 327	1 701

GBC 28	21 113	16 615	2 284
GBC 29	21 030	16 689	2 245
GBC 30	10 382	8 452	1 583
GBC 31	18 048	14 368	2 001
GBC 32	19 162	15 315	2 214
GBC 33	13 757	11 461	1 943
GBC 34	11 388	9 301	1 688
GBC 35	13 199	10 534	1 629
GBC 36	24 206	19 161	2 430
GBC 37	21 250	17 136	2 339

Table S3.3: All dysregulated proteins in gallbladder cancer compared to normal

Protein Group	**Protein Name**	**Protein Name**	**No. of Unique Peptides**	**Average Log_2 Ratio**	**p-value**	**q-value**
Q15063	POSTN – Periostin	POSTN	59	5.02066359	<0.001	<0.001
P04732	MT1E – Metallothionein 1E	MT1E	3	4.712206662	<0.001	<0.001
P31327	CPSM – Carbamoyl-phosphate synthase	CPSM	103	4.483369762	<0.001	<0.001
P07148	FABPL – Fatty acid binding protein	FABPL	14	4.024563036	<0.001	<0.001
Q99715	COCA1 – Collagen alpha-1(XII) chain	COCA1	143	4.004558467	<0.001	<0.001
P02792	FRIL – Ferritin light chain	FRIL	11	3.964146323	<0.001	<0.001
P04004	VTNC – Vitronectin	VTNC	18	3.915829431	<0.001	<0.001
P54868	HMCS2 – Hydroxymethylglutaryl-CoA synthase	HMCS2	26	3.915381366	<0.001	<0.001
P52895	AK1C2 – Aldo-keto reductase family 1 member C2	AK1C2	5	3.848081936	<0.001	<0.001
P52758	RIDA – 2-iminobutanoate/2-iminopropanoate deaminase	RIDA	6	3.70928337	<0.001	<0.001
P04908	H2A1B – Histone H2A type 1-B/E	H2A1B	2	3.601816369	<0.001	<0.001
Q7L7L0	H2A3 – Histone H2A type 3	H2A3	2	3.601816369	<0.001	<0.001
Q93077	H2A1C – Histone H2A type 1-C	H2A1C	2	3.601816369	<0.001	<0.001
P07996	TSP – Thrombospondin-1	TSP	54	3.251710872	<0.001	<0.001

P04264	KSC1 – Keratin, type II cytoskeletal 1	KSC1	54	3.25004815	<0.001	<0.001
P12004	PCNA – Proliferating cell nuclear antigen	PCNA	10	3.172973883	<0.001	<0.001
P02751	FINC – Fibronectin	FINC	117	3.0986923	<0.001	<0.001
Q7Z4W1	DCXR – L-xylulose reductase	DCXR	13	2.983416149	<0.001	<0.001
P02649	APOE – Apolipoprotein E	APOE	26	2.968525542	<0.001	<0.001
Q06278	AOXA – Aldehyde oxidase 1	AOXA	43	2.942234669	<0.001	<0.001
P07477	TRY1 – Serine protease 1	TRY1	2	2.931102462	<0.001	<0.001
Q02338	BDH – 3-hydroxybutyrate dehydrogenase 1	BDH	15	2.900769273	<0.001	<0.001
P01833	PIGR – Polymeric immunoglobulin receptor	PIGR	29	2.898754081	<0.001	<0.001
P01911	DRB1 – HLA class II histocompatibility antigen, DRB1 beta chain	DRB1	10	2.740000646	<0.001	<0.001
P15153	RAC2 – Rac family small GTPase 2	RAC2	5	2.737986073	<0.001	<0.001
P05413	FABPH – Fatty acid binding protein	FABPH	11	2.731079529	<0.01	<0.01
P24821	TENA – Tenascin C	TENA	84	2.712643782	<0.001	<0.001
P13611	CSPG2 – Versican core protein	CSPG2	31	2.635654171	<0.001	<0.001
P35555	FBN1 – Fibrillin-1	FBN1	108	2.619872137	<0.001	<0.001
Q08426	ECHP – Peroxisomal bifunctional enzyme	ECHP	39	2.608263782	<0.001	<0.001
P00747	PLMN – Plasminogen	PLMN	54	2.587929691	<0.001	<0.001
P19971	TYPH – Thymidine phosphorylase	TYPH	25	2.561671217	<0.001	<0.001
P34897	GLYM – Serine hydroxymethyltransferase 2	GLYM	21	2.51078038	<0.001	<0.001
P02452	CO1A1 – Collagen alpha-1(I) chain	CO1A1	29	2.507681268	<0.001	<0.001
P02748	CO9 – Complement C9	CO9	26	2.454653675	<0.001	<0.001
P11215	ITAM – Integrin subunit alpha M	ITAM	31	2.441464589	<0.001	<0.001
Q00688	FKBP3 – Integrin alpha-M	FKBP3	11	2.427845643	<0.001	<0.001
P09110	THIK - Acetyl-CoA acyltransferase	THIK	17	2.383977059	<0.001	<0.001

P21333	FLNA – Filamin-A	FLNA	169	-2.489748582	<0.001	<0.001
P62736	ACTA – Actin alpha 2	ACTA	21	-2.492619937	<0.001	<0.001
P09493	TPM1 – Tropomyosin alpha-1 chain	TPM1	25	-2.60007637	<0.01	<0.01
Q03135	CAV1 – Caveolin 1	CAV1	9	-2.651299406	<0.001	<0.001
P51911	CNN1 – Calponin 1	CNN1	27	-2.741320166	<0.001	<0.001
P18085	ARF4 – ADP ribosylation factor 4	ARF4	6	-2.756697069	<0.001	<0.001
P09525	ANXA4 – Annexin A4	ANXA4	36	-2.787707074	<0.01	<0.01
P10301	RRAS – Ras-related protein R-Ras	RRAS	14	-2.788038724	<0.001	<0.001
P20231	TRYB2 – Tryptase beta 2	TRYB2	16	-2.800721074	<0.001	<0.001
Q15661	TRYB1 – Tryptase alpha/beta 1	TRYB1	16	-2.800721074	<0.001	<0.001
P20774	MIME – Mimecan	MIME	23	-2.895872701	<0.001	<0.001
P07951	TPM2 – Tropomyosin beta chain	TPM2	26	-2.980807556	<0.001	<0.001
Q9HBI1	PARVB – Beta-parvin	PARVB	3	-3.106309788	<0.001	<0.001
P00918	CAH2 – Carbonic anhydrase 2	CAH2	18	-3.359399413	<0.001	<0.001
P06703	S10A6 – Protein S100-A6	S10A6	9	-3.362875006	<0.001	<0.001
Q01995	TAGL – Transgelin	TAGL	31	-3.372011937	<0.01	<0.01
P17661	DESM – Desmin	DESM	53	-3.441316866	<0.001	<0.001
Q9NZN4	EHD2 – EH domain-containing protein 2	EHD2	36	-3.475631992	<0.001	<0.001
O14950	ML12B – Myosin regulatory light chain 12B	ML12B	10	-3.549636751	<0.001	<0.001
P19105	ML12A – Myosin regulatory light chain 12A	ML12A	10	-3.549636751	<0.001	<0.001
P24844	MYL9 – Myosin regulatory light polypeptide 9	MYL9	6	-3.628772097	<0.001	<0.001
P35749	MYH11 – Myosin-11	MYH11	231	-3.652336209	<0.001	<0.001
Q6NZI2	CAVN1 – Caveolae-associated protein 1	CAVN1	19	-3.822693564	<0.001	<0.001
O60271	JIP4 – C-Jun-amino-terminal kinase-interacting protein 4	JIP4	24	-4.856406351	<0.01	<0.01

Table S3.4: The total precursors, peptides, and protein groups identified for each individual patient for the gallbladder cancer versus gallstone disease comparison

Sample number	Precursors	Peptides	Protein Groups
Gallstone disease (GD) 1	12 737	10 125	1 623
GD 2	6 375	5 128	958
GD 3	11 297	9 099	1 472
GD 4	8 701	7 031	1 203
GD 5	10 836	8 758	138
GD 6	14 447	11 514	1 802
GD 7	13 195	10 688	1 758
GD 9	18 144	14 182	1 993
GD 10	11 417	9 157	1 446
GD 11	13 248	10 413	1 581
GD 12	16 728	13 401	2 013
GD 14	15 737	12 323	1 793
GD 15	12 518	9 927	1 612
Gallbladder cancer (GBC) 7	10 699	8 693	1 545
GBC 8	22 129	18 073	2 338
GBC 11	12 663	10 536	1 809
GBC 12	8 768	7 201	1 319
GBC 13	16 239	13 000	1 957
GBC 14	24 647	19 502	2 400
GBC 15	22 470	17 795	2 414
GBC 16	24 653	19 744	2 450
GBC 17	18 551	14 966	2 163
GBC 18	11 360	9 474	1 671
GBC 19	19 301	15 581	2 159
GBC 21	22 882	17 872	2 245
GBC 22	23 220	18 313	2 332
GBC 24	20 115	16 154	2 142
GBC 25	21 040	16 677	2 263
GBC 26	21 099	16 745	2 191
GBC 27	12 785	10 333	1 707
GBC 28	21 105	16 605	2 281
GBC 29	21 035	16 693	2 244
GBC 30	10 390	8 459	1 589
GBC 31	18 053	14 372	2 006
GBC 32	19 163	15 314	2 210
GBC 33	13 755	11 459	1 945
GBC 34	11 387	9 300	1 687
GBC 35	13 203	10 537	1 631
GBC 36	21 187	19 141	2 423
GBC 37	21 223	17 107	2 326

Table S3.5: All differentially abundant proteins in gallbladder cancer compared to gallstone disease

Protein Group	Protein Name	No. of Unique Peptides	Average Log_2 Ratio	p-value	q-value
P52895	AK1C2 – Aldo-keto reductase family 1 member C2	5	4.953323214	<0.001	<0.001
P31327	CPSM – Carbamoyl-phosphate synthase	103	4.323381454	<0.001	<0.001
P11498	PYC – Pyruvate carboxylase	56	4.065746017	<0.001	<0.001
Q04828	AK1C1 – Aldo-keto reductase family 1 member C1	23	3.67721375	<0.001	<0.001
Q02338	BDH – D-beta-hydroxybutyrate dehydrogenase	15	3.638123672	<0.001	<0.001
P52758	RIDA – 2-iminobutanoate/2-iminopropanoate deaminase	6	3.587142384	<0.001	<0.001
Q99715	COCA1 – Collagen alpha-1(XII) chain	143	3.374782172	<0.001	<0.001
O60701	UGDH – UDP-glucose 6-dehydrogenase	31	3.088888829	<0.001	<0.001
P07996	TSP1 – Thrombospondin-1	54	3.003882122	<0.001	<0.001
Q7Z4W1	DCXR – L-xylulose reductase	13	2.844273258	<0.001	<0.001
P35442	TSP2 – Thrombospondin-2	37	2.780811447	<0.001	<0.001
P42765	THIM – 3-ketoacyl-CoA thiolase	31	2.691117194	<0.001	<0.001
P05062	ALDOB – Fructose-bisphosphate aldolase B	26	2.64136246	<0.001	<0.001
P42330	AK1C3 – Aldo-keto reductase family 1 member C3	12	2.626561093	<0.001	<0.001
Q8IUX7	AEBP1 – Adipocyte enhancer-binding protein 1	24	2.608124891	<0.001	<0.001
Q14956	GPNMB – Transmembrane glycoprotein NMB	2	2.592714206	<0.001	<0.001
P16219	ACADS – Short-chain specific acyl-CoA dehydrogenase	17	2.557209494	<0.001	<0.001
P12004	PCNA – Proliferating cell nuclear antigen	10	2.479580402	<0.001	<0.001
P06731	CEAM5 – Carcinoembryonic antigen-related cell adhesion molecule 5	5	2.468455016	<0.001	<0.001
Q96HE7	ERO1A – ERO1-like protein alpha	23	2.418763344	<0.001	<0.001
Q08380	LG3BP – Galectin-3-binding protein	17	2.371437513	<0.001	<0.001
O60218	AK1BA – Aldo-keto reductase family 1 member B10	26	2.356581673	<0.001	<0.001
Q15063	POSTN – Periostin	59	2.216168479	<0.001	<0.001
P01833	PIGR – Polymeric immunoglobulin receptor	29	2.196173212	<0.001	<0.001
P00966	ASSY – Argininosuccinate synthase	21	2.176407723	<0.001	<0.001
P11413	G6PD – Glucose-6-phosphate 1-dehydrogenase	40	2.17522764	<0.001	<0.001
P08263	GSTA1 – Glutathione S-transferase A1	31	2.172464657	<0.001	<0.001
P16930	FAAA – Fumarylacetoacetase	15	2.127056816	<0.001	<0.001
P23141	EST1 – Liver carboxylesterase 1	36	2.109240352	<0.001	<0.001
P02649	APOE – Apolipoprotein E	26	2.088429222	<0.001	<0.001

P51659	DHB4 – Peroxisomal multifunctional enzyme type 2	36	2.081996564	<0.001	<0.001
P00167	CYB5 – Cytochrome b5	8	2.076020282	<0.001	<0.001
Q30154	DRB5 – HLA class II histocompatibility antigen	3	2.07583031	<0.01	<0.01
P09467	F16P1 – Fructose-1,6-bisphosphatase 1	17	2.063445957	<0.001	<0.001
P08729	K2C7 – Keratin, type II cytoskeletal 7	48	2.054256483	<0.001	<0.001
P17174	AATC – Aspartate aminotransferase	26	2.046466287	<0.001	<0.001
Q16698	DECR – 2,4-dienoyl-CoA reductase	18	2.016535491	<0.001	<0.001
Q16881	TRXR1 – Thioredoxin reductase 1	19	1.980805794	<0.001	<0.001
P05783	K1C18 – Keratin, type I cytoskeletal 18	37	1.975434726	<0.001	<0.001
P10809	CH60 – 60 kDa heat shock protein	54	1.920904009	<0.001	<0.001
P24752	THIL – Acetyl-CoA acetyltransferase	33	1.914611807	<0.001	<0.001
P16949	STMN1 – Stathmin	6	1.895356424	<0.001	<0.001
P34897	GLYM – Serine hydroxymethyltransferase	21	1.862362934	<0.001	<0.001
P16435	NCPR – NADPH--cytochrome P450 reductase	27	1.848717579	<0.001	<0.001
P59665	DEF1 – Neutrophil defensin 1	4	1.838381823	<0.001	<0.001
P59666	DEF3 – Neutrophil defensin 3	4	1.838381823	<0.001	<0.001
P00367	DHE3 – Glutamate dehydrogenase 1	33	1.830040693	<0.001	<0.001
O95831	AIFM1 – Apoptosis-inducing factor 1	27	1.805722015	<0.001	<0.001
P19971	TYPH – Thymidine phosphorylase	25	1.795493292	<0.001	<0.001
P38117	ETFB – Electron transfer flavoprotein subunit beta	21	1.791739209	<0.001	<0.001
P30084	ECHM – Enoyl-CoA hydratase	22	1.764578278	<0.001	<0.001
P24821	TENA – Tenascin	84	1.763552536	<0.001	<0.001
P05787	K2C8 – Keratin, type II cytoskeletal 8	47	1.751221966	<0.001	<0.001
P07858	CATB – Cathepsin B	14	1.745793698	<0.001	<0.001
P38646	GRP75 – Stress-70 protein	43	1.744767062	<0.001	<0.001
P11586	C1TC – C-1-tetrahydrofolate synthase	55	1.742940969	<0.001	<0.001
Q01813	PFKAP – ATP-dependent 6-phosphofructokinase	31	1.742191304	<0.001	<0.001
Q01469	FABP5 – Fatty acid-binding protein 5	21	1.736882963	<0.05	<0.01
P08727	K1C19 – Keratin, type I cytoskeletal 19	30	1.703228728	<0.001	<0.001
P05141	ADT2 – ADP/ATP translocase 2	10	1.692862602	<0.001	<0.001
Q9Y2Q3	GSTK1 – Glutathione S-transferase kappa 1	12	1.680971617	<0.001	<0.001
P15153	RAC2 – Ras-related C3 botulinum toxin substrate 2	5	1.675119445	<0.001	<0.001
P08195	4F2 – 4F2 cell-surface antigen heavy chain	15	1.665178905	<0.001	<0.001
P31949	S10AB – Protein S100-A11	6	1.662021978	<0.001	<0.001
P07339	CATD – Cathepsin D	28	1.649144014	<0.001	<0.001
Q9BPW8	NIPS1 – NipSnap homolog 1	9	1.626564893	<0.01	<0.01
P21796	VDAC1 – Voltage-dependent anion-selective channel protein 1	21	1.614328869	<0.001	<0.001

P43304	GPDM – Glycerol-3-phosphate dehydrogenase	24	1.597357909	<0.001	<0.001
O75874	IDHC – Isocitrate dehydrogenase	26	1.595013547	<0.001	<0.001
P07099	HYEP – Epoxide hydrolase 1	33	1.586795424	<0.001	<0.001
P08238	HS90B – Heat shock protein HSP 90-beta	53	1.576348982	<0.001	<0.001
P13804	ETFA – Electron transfer flavoprotein subunit alpha	17	1.570324464	<0.001	<0.001
Q9H0W9	CK054 – Ester hydrolase C11orf54	11	1.569835702	<0.01	<0.01
P22307	SCP2 – Sterol carrier protein 2	36	1.565157933	<0.001	<0.001
P05107	ITB2 – Integrin beta-2	32	1.563240053	<0.001	<0.001
P48735	IDHP – Isocitrate dehydrogenase	30	1.558495224	<0.001	<0.001
O15533	TPSN – Tapasin	6	1.558452978	<0.001	<0.001
P01903	DRA – HLA class II histocompatibility antigen, DR alpha chain	8	1.552950249	<0.01	<0.01
P40121	CAPG – Macrophage-capping protein	17	1.551524099	<0.001	<0.001
P50225	ST1A1 – Sulfotransferase 1A1	11	1.546602828	<0.01	<0.01
P20292	AL5AP – Arachidonate 5-lipoxygenase-activating protein	2	1.54098875	<0.001	<0.001
P61604	CH10 – 10 kDa heat shock protein	14	1.533125541	<0.001	<0.001
P23229	ITA6 – Integrin alpha-6	23	1.528409047	<0.001	<0.001
P04004	VTNC – Vitronectin	18	1.508175969	<0.01	<0.01
Q6YN16	HSDL2 – Hydroxysteroid dehydrogenase-like protein 2	19	1.477875786	<0.001	<0.001
Q13283	G3BP1 – Ras GTPase-activating protein-binding protein 1	15	1.476789382	<0.001	<0.001
Q15185	TEBP – Prostaglandin E synthase 3	6	1.476785049	<0.001	<0.001
P28161	GSTM2 – Glutathione S-transferase Mu 2	6	-1.479098041	<0.001	<0.001
O43294	TGFI1 – Transforming growth factor beta-1-induced transcript 1 protein	13	-1.481158888	<0.001	<0.001
P05090	APOD – Apolipoprotein D	12	-1.497688594	<0.01	<0.01
P01619	KV320 – Immunoglobulin kappa variable 3-20	4	-1.508669451	<0.001	<0.001
P14543	NID1 – Nidogen-1	34	-1.512686623	<0.001	<0.001
P04433	KV311 – Immunoglobulin kappa variable 3-11	2	-1.531428044	<0.01	<0.01
Q7Z5L7	PODN – Podocan	20	-1.535722608	<0.001	<0.001
P12109	CO6A1 – Collagen alpha-1(VI) chain	51	-1.53845217	<0.001	<0.001
P08185	CBG – Corticosteroid-binding globulin	5	-1.541950634	<0.001	<0.001
P04196	HRG – Histidine-rich glycoprotein	18	-1.546626855	<0.01	<0.01
Q96AC1	FERM2 – Fermitin family homolog 2	32	-1.553301196	<0.001	<0.001
Q9HBI1	PARVB – Beta-parvin	3	-1.560212729	<0.001	<0.001
Q15404	RSU1 – Ras suppressor protein 1	11	-1.562357193	<0.001	<0.001
Q06828	FMOD – Fibromodulin	11	-1.578801641	<0.001	<0.001
P02787	TRFE – Serotransferrin	74	-1.583597892	<0.001	<0.001
P11047	LAMC1 – Laminin subunit gamma-1	67	-1.59617644	<0.001	<0.001
P04908	H2A1B – Histone H2A type 1-B/E	2	-1.597290839	<0.001	<0.001

Q7L7L0	H2A3 – Histone H2A type 3	2	-1.597290839	<0.001	<0.001
Q93077	H2A1C – Histone H2A type 1-C	2	-1.597290839	<0.001	<0.001
Q9BZQ8	NIBA1 – Niban 1	17	-1.598688541	<0.001	<0.001
P43121	MUC18 – Cell surface glycoprotein MUC18	17	-1.598813306	<0.001	<0.001
P12111	CO6A3 – Collagen alpha-3(VI) chain	184	-1.61580628	<0.001	<0.001
P02790	HEMO – Hemopexin	31	-1.618488319	<0.001	<0.001
Q13418	ILK – Integrin-linked protein kinase	23	-1.636966035	<0.001	<0.001
P43652	AFAM – Afamin	23	-1.650870889	<0.001	<0.001
Q9UBX5	FBLN5 – Fibulin-5	15	-1.674528913	<0.001	<0.001
P55083	MFAP4 – Microfibril-associated glycoprotein 4	6	-1.674834435	<0.001	<0.001
P01614	KVD40 – Immunoglobulin kappa variable 2D-40	4	-1.687001286	<0.001	<0.001
P02753	RET4 – Retinol-binding protein 4	11	-1.68830487	<0.001	<0.001
P01871	IGHM – Immunoglobulin heavy constant mu	7	-1.709143093	<0.05	<0.01
P00568	KAD1 – Adenylate kinase isoenzyme 1	12	-1.735247741	<0.01	<0.01
Q13228	SBP1 – Methanethiol oxidase	30	-1.73889701	<0.01	<0.01
A0A075B6H7	KV37 – Probable non-functional immunoglobulin kappa variable 3-7	2	-1.744984995	<0.001	<0.001
P60033	CD81 – CD81 antigen	4	-1.749183973	<0.001	<0.001
P60660	MYL6 – Myosin light polypeptide 6	10	-1.766841827	<0.001	<0.001
P49961	ENTP1 – Ectonucleoside triphosphate diphosphohydrolase 1	9	-1.770423909	<0.001	<0.001
P14649	MYL6B – Myosin light chain 6B	9	-1.777243042	<0.001	<0.001
P05546	HEP2 – Heparin cofactor 2	16	-1.778561033	<0.001	<0.001
P01591	IGJ – Immunoglobulin J chain	8	-1.784962119	<0.01	<0.01
Q93052	LPP – Lipoma-preferred partner	29	-1.792208928	<0.001	<0.001
O00159	MYO1C – Unconventional myosin-Ic	49	-1.79914756	<0.001	<0.001
P00739	HPTR – Haptoglobin-related protein	7	-1.807881632	<0.01	<0.01
Q6UWY5	OLFL1 – Olfactomedin-like protein 1	13	-1.813201967	<0.001	<0.001
Q53TN4	CYBR1 – Cytochrome b reductase 1	3	-1.82094853	<0.001	<0.001
P09455	RET1 – Retinol-binding protein 1	9	-1.828235131	<0.001	<0.001
P02671	FIBA – Fibrinogen alpha chain	60	-1.879968132	<0.05	<0.01
P68133	ACTS – Alpha-actin-1	2	-1.911461545	<0.001	<0.001
Q14315	FLNC – Filamin-C	107	-1.92873136	<0.001	<0.001
P02743	SAMP – Serum amyloid P-component	11	-1.943145428	<0.001	<0.001
P60981	DEST – Destrin	12	-1.967869924	<0.001	<0.001
P10301	RRAS – Ras-related protein R-Ras	14	-1.986053122	<0.001	<0.001
P01023	A2MG – Alpha-2-macroglobulin	75	-2.00585659	<0.001	<0.001
Q68CZ2	TENS3 – Tensin-3	20	-2.008561926	<0.001	<0.001
P07738	PMGE – Bisphosphoglycerate mutase	11	-2.021951202	<0.001	<0.001
O75368	SH3L1 – SH3 domain-binding glutamic acid-rich-like protein 1	8	-2.022416573	<0.001	<0.001
P02768	ALBU – Albumin	100	-2.027978309	<0.001	<0.001

P18206	VINC – Vinculin	82	-2.036145639	<0.001	<0.001
P06753	TPM3 – Tropomyosin alpha-3 chain	36	-2.048250457	<0.001	<0.001
P69891	HBG1 – Haemoglobin subunit gamma-1	7	-2.075686751	<0.001	<0.001
Q8WU39	MZB1 – Marginal zone B- and B1-cell-specific protein	8	-2.080293581	<0.01	<0.01
Q05707	COEA1 – Collagen alpha-1(XIV) chain	68	-2.09873457	<0.001	<0.001
P21291	CSRP1 – Cysteine and glycine-rich protein 1	13	-2.1514253	<0.001	<0.001
P13716	HEM2 – Delta-aminolevulinic acid dehydratase	15	-2.180358021	<0.01	<0.01
P21333	FLNA – Filamin-A	169	-2.187590377	<0.001	<0.001
P02647	APOA1 – Apolipoprotein A-I	35	-2.18857284	<0.001	<0.001
P07585	PGS2 – Decorin	26	-2.206845035	<0.001	<0.001
P08294	SODE – Extracellular superoxide dismutase	12	-2.218963629	<0.001	<0.001
Q86Y46	K2C73 – Keratin, type II cytoskeletal 73	3	-2.230272282	<0.001	<0.001
P55268	LAMB2 – Laminin subunit beta-2	58	-2.231997226	<0.001	<0.001
P39059	COFA1 – Collagen alpha-1(XV) chain	15	-2.26081612	<0.001	<0.001
O15230	LAMA5 – Laminin subunit alpha-5	84	-2.279449202	<0.001	<0.001
Q9HBL0	TENS1 – Tensin-1	46	-2.299559932	<0.001	<0.001
Q13425	SNTB2 – Beta-2-syntrophin	13	-2.41437426	<0.001	<0.001
Q15746	MYLK – Myosin light chain kinase	48	-2.437637008	<0.001	<0.001
P02654	APOC1 – Apolipoprotein C-I	4	-2.463745607	<0.05	<0.05
Q9NR12	PDLI7 – PDZ and LIM domain protein 7	21	-2.514497229	<0.001	<0.001
P62736	ACTA – Actin alpha 2	21	-2.608621477	<0.001	<0.001
P12277	KCRB – Creatine kinase B-type	21	-2.779624442	<0.001	<0.001
P02766	TTR – Transthyretin	11	-2.796960584	<0.001	<0.001
P09493	TPM1 – Tropomyosin alpha-1 chain	25	-2.832043183	<0.001	<0.001
P07951	TPM2 – Tropomyosin beta chain	26	-2.841326704	<0.001	<0.001
Q16853	AOC3 – Membrane primary amine oxidase	21	-2.882510929	<0.001	<0.001
P02652	APOA2 – Apolipoprotein A-II	7	-2.910543359	<0.001	<0.001
P34949	MPI – Mannose-6-phosphate isomerase	7	-3.016282551	<0.001	<0.001
Q9BX66	SRBS1 – Sorbin and SH3 domain-containing protein 1	35	-3.046542825	<0.001	<0.001
Q01995	TAGL – Transgelin	31	-3.075365695	<0.001	<0.001
Q99459	CDC5L – Cell division cycle 5-like protein	17	-3.092589754	<0.01	<0.01
P35749	MYH11 – Myosin-11	231	-3.238178757	<0.001	<0.001
P32119	PRDX2 – Peroxiredoxin-2	17	-3.29406503	<0.01	<0.01
P20231	TRYB2 – Tryptase beta-2	16	-3.371269387	<0.001	<0.001
Q15661	TRYB1 – Tryptase alpha/beta-1	16	-3.371269387	<0.001	<0.001
Q03135	CAV1 – Caveolin-1	9	-3.408389459	<0.001	<0.001
P68871	HBB – Haemoglobin subunit beta	22	-3.465642881	<0.01	<0.01
Q6NZI2	CAVN1 – Caveolae-associated protein 1	19	-3.521064707	<0.001	<0.001
P11277	SPTB1 – Spectrin beta chain	86	-3.671171173	<0.01	<0.01

P20774	MIME – Mimecan	23	-3.70936941	<0.001	<0.001
Q9NZN4	EHD2 – EH domain-containing protein 2	36	-3.733255327	<0.001	<0.001
P69905	HBA – Haemoglobin subunit alpha	20	-3.769500318	<0.01	<0.01
P51911	CNN1 – Calponin-1	27	-3.789798779	<0.001	<0.001
P24844	MYL9 – Myosin regulatory light polypeptide 9	6	-3.939013593	<0.001	<0.001
P02042	HBD – Haemoglobin subunit delta	10	-4.021552107	<0.01	<0.01
P00918	CAH2 – Carbonic anhydrase 2	18	-4.035396298	<0.01	<0.01
O60271	JIP4 – C-Jun-amino-terminal kinase-interacting protein 4	24	-4.346847515	<0.001	<0.001
P02730	B3AT – Band 3 anion transport protein	23	-4.674344703	<0.01	<0.01
P00915	CAH1 – Carbonic anhydrase 1	19	-4.723264666	<0.01	<0.01
P17661	DESM – Desmin	53	-4.845020339	<0.001	<0.001

Table S3.6: The total precursors, peptides, and protein groups identified for each individual patient for the gallstone disease versus normal comparison

Sample number	Precursors	Peptides	Protein Groups
Normal 1	13 662	11 157	1706
Normal 2	14 615	11 940	1723
Normal 3	17 719	14 386	1870
Normal 4	12 162	9940	1650
Normal 5	12 849	10 426	1652
Gallstone disease (GD) 1	12 635	10 030	1540
GD 2	6356	5107	931
GD 3	11 225	9029	1410
GD 4	8658	6990	1163
GD 5	10 718	8645	1443
GD 6	14 238	11 310	1647
GD 7	13 001	10 498	1614
GD 9	17 922	13 967	1828
GD 10	11 333	9076	1379
GD 11	13 161	10 329	1511
GD 12	16 474	13 159	1823
GD 14	15 571	12 161	1658
GD 15	12 409	9822	1523

Table S3.7: All differentially expressed proteins in gallstone disease compared to normal

Protein Groups	Protein Name	No. of Unique Peptides	Average Log2 Ratio	p-value	q-value
P04908	H2A1B – Histone H2A type 1-B/E A	2	5.026480401	<0.001	<0.001
Q7L7L0	H2A3 – Histone H2A type 3	2	5.026480401	<0.001	<0.001
Q93077	H2A1C – Histone H2A type 1-C	2	5.026480401	<0.001	<0.001
A0A075B6H7	KV37 – Probable non-functional immunoglobulin kappa variable 3-7	2	3.512519933	<0.001	<0.001
P07477	TRY1 – Serine protease 1	2	3.471879053	<0.001	<0.001
P00747	PLMN – Plasminogen	54	3.378687829	<0.001	<0.001
P62316	SMD2 – Small nuclear ribonucleoprotein Sm D2	12	3.366337484	<0.05	<0.01
P01834	IGKC – Immunoglobulin kappa constant	2	3.331233775	<0.001	<0.001
P01871	IGHM – Immunoglobulin heavy constant mu	7	3.146650537	<0.001	<0.001
Q15063	POSTN – Periostin	59	3.104959339	<0.001	<0.001
P01714	LV319 – Immunoglobulin lambda variable 3-19	2	3.013745266	<0.001	<0.001
A0A0A0MS15	HV349 – Immunoglobulin heavy variable 3-49	4	2.947311103	<0.001	<0.001
P16157	ANK1 – Ankyrin-1	56	2.852903234	<0.01	<0.001
P01860	IGHG3 – Immunoglobulin heavy constant gamma 3	9	2.839500671	<0.001	<0.001
P00739	HPTR – Haptoglobin-related protein	7	2.830130286	<0.001	<0.001
P04003	C4BPA – C4b-binding protein alpha chain	27	2.828428874	<0.001	<0.001
P04264	KSC1 – Keratin, type II cytoskeletal 1	33	2.776773217	<0.001	<0.001
P02730	B3AT – Band 3 anion transport protein	23	2.770443432	<0.01	<0.001
P00915	CAH1 – Carbonic anhydrase 1	19	2.750828743	<0.001	<0.001
P11277	SPTB1 – Spectrin beta chain	86	2.721829734	<0.01	<0.01
A0A0B4J1X5	HV374 – Immunoglobulin heavy variable 3-74	3	2.687754081	<0.001	<0.001
P02671	FIBA – Fibrinogen alpha chain	60	2.681953918	<0.001	<0.001
P02792	FRIL – Ferritin light chain	11	2.681790254	<0.001	<0.001
P06331	HV434 – Immunoglobulin heavy variable 4-34	4	2.634933097	<0.001	<0.001
P04732	MT1E – Metallothionein-1E	3	2.611219085	<0.05	<0.01
P04004	VTNC – vitronectin	18	2.583194465	<0.001	<0.001
P01859	IGHG2 – Immunoglobulin heavy constant gamma 2	16	2.576554825	<0.001	<0.001
P01701	LV151 – Immunoglobulin lambda variable 1-51	4	2.549826612	<0.001	<0.001
P0DOX7	IGK – Immunoglobulin kappa light chain	14	2.521000794	<0.001	<0.001
Q99459	CDC5L – Cell division 5-like protein	17	2.51908653	<0.001	<0.001
A0A087WSY6	KVD15 – Immunoglobulin kappa variable 3D-15	2	2.516676165	<0.001	<0.001
P04433	KV311 – Immunoglobulin kappa variable 3-11	2	2.507837652	<0.001	<0.001
A0A0B4J1V0	HV315 – Immunoglobulin heavy variable 3-15	9	2.433286309	<0.001	<0.001

P01619	KV320 – Immunoglobulin kappa variable 3-20	4	2.412259567	<0.001	<0.001
P11678	PERE – Eosinophil peroxidase	30	2.374181688	<0.001	<0.001
P12724	ECP – Eosinophil cationic protein	8	2.351447295	<0.001	<0.001
A0A0C4DH67	KV108 – Immunoglobulin kappa variable 1-8	3	2.325851945	<0.001	<0.001
P01591	IGJ – Immunoglobulin J chain	8	2.274713422	<0.001	<0.001
P0DOX5	IGG1 – Immunoglobulin gamma-1 heavy chain	30	2.261475057	<0.001	<0.001
P18077	RL35A – 60S ribosomal protein L35a	2	2.259865522	<0.001	<0.001
P04040	CATA – Catalase	35	2.236763293	<0.01	<0.001
P0DOX6	IGM – Immunoglobulin mu heavy chain	20	2.222687712	<0.001	<0.001
P51659	DHB4 – Peroxisomal multifunctional enzyme type 2	36	-2.234912265	<0.001	<0.001
O14950	ML12B – Myosin regulatory light chain 12B	10	-2.238596873	<0.001	<0.001
P19105	ML12A – Myosin regulatory light chain 12A	10	-2.238596873	<0.001	<0.001
P46779	RL28 – 60S ribosomal protein L28	6	-2.248073806	<0.001	<0.001
Q92734	TFG – Protein TFG	9	-2.258641507	<0.001	<0.001
O95861	BPNT1 – 3’(2’),5’-biphosphate nucleotidase 1	15	-2.264004031	<0.01	<0.01
P02533	K1C14 – Keratin 1, cytoskeletal 14	27	-2.264244572	<0.001	<0.001
P21397	AOFA – Amine oxidase [flavin-containing] A	20	-2.27090425	<0.01	<0.001
P07339	CATD – Cathepsin D	28	-2.281231719	<0.001	<0.001
Q96I99	SUCB2 – Succinate--CoA ligase subunit beta	20	-2.284271329	<0.01	<0.01
Q16762	THTR – thiosulphate sulphurtransferase	14	-2.296616563	<0.01	<0.01
Q9NX63	MIC19 – MICOS complex subunit	13	-2.300496493	<0.01	<0.01
P53597	SUCA – Succinate--CoA ligase subunit alpha	9	-2.32923445	<0.01	<0.01
P15559	NQO1 – NAD(P)H dehydrogenase 1	14	-2.34160178	<0.001	<0.001
Q06210	GFPT1 – Glutamine—fructose-6-phosphate aminotransferase	31	-2.373925504	<0.01	<0.001
P12235	ADT1 – ADP/ATP translocase 1	5	-2.379055509	<0.001	<0.001
P02538	K2C6A – Keratin, type II cytoskeletal 6A	38	-2.41007921	<0.001	<0.001
P46977	STT3A – Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit	8	-2.4514888	<0.001	<0.001
P06703	S10A6 – Protein S100-A6	9	-2.468148229	<0.001	<0.001
P21266	GSTM3 – Glutathione S-transferase mu	19	-2.477088168	<0.001	<0.001
P17174	AATC – Aspartate aminotransferase	26	-2.499709041	<0.001	<0.001
Q04917	1433F – 14-3-3 protein eta	15	-2.512646513	<0.001	<0.001
P10515	ODP2 – Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	15	-2.513441998	<0.01	<0.01
P35268	RL22 – 60S ribosomal protein L22	5	-2.527634306	<0.001	<0.001
Q96C19	EFHD2 – EF-hand domain-containing D2	18	-2.535237465	<0.001	<0.001
P08134	RHOC – Rho-related GTP-binding protein RhoC	2	-2.555907738	<0.05	<0.01

P42330	AK1C3 – Aldo-keto reductase family 1 member C3	12	-2.558817553	<0.01	<0.001
O94788	AL1A2 – Retinal dehydrogenase 2	2	-2.568421381	<0.001	<0.001
P61204	ARF3 – ADP-ribosylation factor 3	6	-2.584212523	<0.001	<0.001
P84077	ARF1 – ADP-ribosylation factor 1	6	-2.584212523	<0.001	<0.001
P35222	CTNB1 – Catenin beta-1	24	-2.586923641	<0.001	<0.001
P05787	K2C8 – Keratin, type II cytoskeletal 8	47	-2.673299682	<0.001	<0.001
P07099	HYEP – Epoxide hydrolase 1	33	-2.703562532	<0.001	<0.001
P05023	AT1A1 – Sodium/potassium-transporting ATPase subunit alpha-1	41	-2.74670505	<0.001	<0.001
P08727	K1C19 – Keratin, type I cytoskeletal 19	30	-2.752332751	<0.001	<0.001
P13073	COX41 – Cytochrome c oxidase subunit 4 isoform 1	11	-2.81690633	<0.001	<0.001
P09525	ANXA4 – Annexin A4	36	-2.869026412	<0.001	<0.001
P05062	ALDOB – Fructose-biphosphate aldolase B	2	-2.877305513	<0.01	<0.01
P08729	K2C7 – Keratin, type II cytoskeletal 7	48	-2.926438029	<0.001	<0.001
P14406	CX7A2 – Cytochrome c oxidase subunit 7A2	3	-2.934592568	<0.001	<0.001
P15144	AMPN – Aminopeptidase N	36	-2.944084687	<0.01	<0.01
Q9Y646	CBPQ – Carboxypeptidase	8	-2.949875634	<0.01	<0.01
Q14914	PTGR1 – Prostaglandin reductase 1	16	-2.991988947	<0.01	<0.01
O60610	DIAP1 – Protein diaphanous homolog 1	2	-2.994977469	<0.001	<0.001
P84085	ARF5 – ADP-ribosylation factor 5	11	-3.01626769	<0.001	<0.001
O14880	MGST3 – Microsomal glutathione S-transferase 3	4	-3.035528926	<0.001	<0.001
Q04828	AK1C1 – Aldo-keto reductase family 1 member C1	23	-3.047828204	<0.001	<0.001
P05026	AT1B1 - Sodium/potassium-transporting ATPase subunit beta-1	12	-3.064770063	<0.001	<0.001
P19012	K1C15 – Keratin, type I cytoskeletal 15	2	-3.115263073	<0.001	<0.001
P23141	EST1 – Liver carboxylesterase 1	36	-3.14804674	<0.01	<0.01
A5A3E0	POTEF – POTE ankyrin domain family member F	2	-3.244731193	<0.01	<0.001
Q9H2U2	IPYR2 – Inorganic pyrophosphatase 2	13	-3.362175046	<0.001	<0.001
Q02338	BDH – D-beta-hydroxybutyrate dehydrogenase	2	-3.382846245	<0.001	<0.001
P27216	ANX13 – Annexin A13	20	-3.549757494	<0.001	<0.001
P56470	LEG4 – Galectin-4	15	-3.61095977	<0.001	<0.001
P18085	ARF4 – ADP-ribosylation factor 4	6	-3.67162195	<0.001	<0.001
P18283	GPX2 – Glutathione peroxidase 2	9	-3.702170816	<0.001	<0.001
P00167	CYB5 – Cytochrome b5	8	-3.74134519	<0.001	<0.001
P08263	GSTA1 – Glutathione S-transferase A1	2	-3.777706713	<0.01	<0.01
Q58FF6	H90B4 – Putative heat shock protein HSP 90-beta 4	2	-3.788347121	<0.001	<0.001
Q13576	IQGA2 – Ras GTPase-activating-like protein	2	-3.808015946	<0.001	<0.001
O60218	AK1BA – Aldo-keto reductase family 1 member B10	26	-3.832896149	<0.001	<0.001
O95154	ARK73 – Aflatoxin B1 aldehyde reductase member 3	2	-4.998339249	<0.001	<0.001

Table S3.8: Proteins commonly dysregulated in gallbladder cancer compared to gallstone disease and normal with relative average Log_2 fold changes

Common Proteins in GBC/Normal & GBC/GD	Average Log_2 Fold Change	
Protein Name	**GBC/Normal**	**GBC/GD**
CPSM – Carbamoyl-phosphate synthase	4.483369762	4.323381454
COCA1 – Collagen alpha-1(XII) chain	4.004558467	3.374782172
AK1C2 – Aldo-keto reductase family 1 member C2	3.848081936	4.953323214
RIDA – 2-iminobutanoate/2-iminopropanoate deaminase	3.70928337	3.587142384
TSP1 – Thrombospondin-1	3.251710872	3.003882122
PCNA – Proliferating cell nuclear antigen	3.172973883	2.479580402
DCXR – L-xylulose reductase	2.983416149	2.844273258
APOE – Apolipoprotein E	2.968525542	2.088429222
PIGR – Polymeric immunoglobulin receptor	2.898754081	2.196173212
RAC2 – Ras-related C3 botulinum toxin substrate 2	2.737986073	1.675119445
TENA – Tenascin	2.712643782	1.763552536
TYPH – Thymidine phosphorylase	2.561671217	1.795493292
GLYM – Serine hydroxymethyltransferase 2	2.51078038	1.862362934
FLNA – Filamin-A	-2.489748582	-2.187590377
ACTA – Actin alpha 2	-2.492619937	-2.608621477
TPM1 – Tropomyosin alpha-1 chain	-2.60007637	-2.832043183
CAV1 – Caveolin 1	-2.651299406	-3.408389459
CNN1 – Calponin 1	-2.741320166	-3.789798779
RRAS – Ras-related protein R-Ras	-2.788038724	-1.986053122
TRYB2 – Tryptase beta 2	-2.800721074	-3.371269387
TRYB1 – Tryptase alpha/beta 1	-2.800721074	-3.371269387
MIME – Mimecan	-2.895872701	-3.70936941
TPM2 – Tropomyosin beta chain	-2.980807556	-2.841326704
PARVB – Beta-parvin	-3.106309788	-1.560212729
CAH2 – Carbonic anhydrase 2	-3.359399413	-4.035396298
TAGL – Transgelin	-3.372011937	-3.075365695
DESM – Desmin	-3.441316866	-4.845020339
EHD2 – EH domain-containing protein 2	-3.475631992	-3.733255327
MYL9 – Myosin regulatory light polypeptide 9	-3.628772097	-3.939013593
MYH11 – Myosin-11	-3.652336209	-3.238178757
CAVN1 – Caveolae-associated protein 1	-3.822693564	-3.521064707
JIP4 – C-Jun-amino-terminal kinase-interacting protein 4	-4.856406351	-4.346847515

Common Proteins in GBC/GD & GD/Normal	Average Log_2 Fold Change	
Protein Name	**GBC/GD**	**GD/Normal**
AK1C1 – Aldo-keto reductase family 1 member C1	3.67721375	-3.047828204
ALDOB – Fructose-bisphosphate aldolase B	2.64136246	-2.877305513
AK1C3 – Aldo-keto reductase family 1 member C3	2.626561093	-2.558817553
AK1BA – Aldo-keto reductase family 1 member B10	2.356581673	-3.832896149
GSTA1 – Glutathione S-transferase A1	2.172464657	-3.777706713
EST1 – Liver carboxylesterase 1	2.109240352	-3.14804674
DHB4 – Peroxisomal multifunctional enzyme type 2	2.081996564	-2.234912265
CYB5 – Cytochrome b5	2.076020282	-3.74134519
K2C7 – Keratin, type II cytoskeletal 7	2.054256483	-2.926438029

AATC – Aspartate aminotransferase	2.046466287	-2.499709041
K2C8 – Keratin, type II cytoskeletal 8	1.751221966	-2.673299682
K1C19 – Keratin, type I cytoskeletal 19	1.703228728	-2.752332751
CATD – Cathepsin D	1.649144014	-2.281231719
HYEP – Epoxide hydrolase 1	1.586795424	-2.703562532
KV320 – Immunoglobulin kappa variable 3-20	-1.508669451	2.412259567
KV311 – Immunoglobulin kappa variable 3-11	-1.531428044	2.507837652
IGHM – Immunoglobulin heavy constant mu	-1.709143093	3.146650537
KV37 – Probable non-functional immunoglobulin kappa variable 3-7	-1.744984995	3.512519933
IGJ – Immunoglobulin J chain	-1.784962119	2.274713422
HPTR – Haptoglobin-related protein	-1.807881632	2.830130286
FIBA – Fibrinogen alpha chain	-1.879968132	2.681953918
CDC5L – Cell division cycle 5-like protein	-3.092589754	2.51908653
SPTB1 – Spectrin beta chain	-3.671171173	2.721829734
B3AT – Band 3 anion transport protein	-4.674344703	2.770443432
CAH1 – Carbonic anhydrase 1	-4.723264666	2.750828743

Common Proteins in GBC/Normal & GD/Normal	**Average Log_2 Fold Change**	
Protein Name	**GBC/Normal**	**GD/Normal**
MT1E – Metallothionein 1E	4.712206662	2.611219085
FRIL – Ferritin light chain	3.964146323	2.681790254
KSC1 – Keratin, type II cytoskeletal 1	3.25004815	2.776773217
TRY1 – Serine protease 1	2.931102462	3.471879053
PLMN – Plasminogen	2.587929691	3.378687829
ARF4 – ADP ribosylation factor 4	-2.756697069	-3.67162195
ANXA4 – Annexin A4	-2.787707074	-2.869026412
S10A6 – Protein S100-A6	-3.362875006	-2.468148229
ML12B – Myosin regulatory light chain 12B	-3.549636751	-2.238596873
ML12A – Myosin regulatory light chain 12A	-3.549636751	-2.238596873

Common Proteins in GBC/Normal, GBC/GD, & GD/Normal	**Average Log_2 Fold Change**		
Protein Name	**GBC/Normal**	**GBC/GD**	**GD/Normal**
POSTN – Periostin	5.02066359	2.216168479	3.104959339
VTNC – Vitronectin	3.915829431	1.508175969	2.583194465
H2A1B – Histone H2A type 1-B/E	3.601816369	-1.597290839	5.026480401
H2A3 – Histone H2A type 3;	3.601816369	-1.597290839	5.026480401
H2A1C – Histone H2A type 1-C	3.601816369	-1.597290839	5.026480401
BDH – D-beta-hydroxybutyrate dehydrogenase	2.900769273	3.638123672	-3.382846245

Table S3.9: The total precursors, peptides, and protein groups identified for each individual patient for the gallbladder cancer versus benign biliary pathology plasma comparison

Sample number	Precursor	Peptides	Protein Groups
Benign biliary pathology (BBP) 1	3 915	2 618	232
BBP 2	3 714	2 501	225
BBP 3	3 516	2 381	220
BBP 4	3 992	2 654	237
BBP 5	3 673	2 461	228
BBP 6	3 854	2 544	230
BBP 9	3 666	2 479	224
BBP 10	3 754	2 504	229
BBP 11	3 809	2 552	226
BBP 12	3 981	2 612	230
BBP 13	3 826	2 554	235
BBP 14	4 082	2 704	244
BBP 15	3 695	2 488	230
BBP 16	3 639	2 442	226
BBP 17	3 836	2 567	232
BBP 18	3 807	2 519	229
BBP 19	4 099	2 686	229
BBP 23	3 789	2 536	236
BBP 24	3 746	2 509	231
BBP 25	3 890	2 594	237
BBP 26	3 717	2 497	221
BBP 27	3 816	2 544	232
BBP 28	3 863	2 587	234
BBP 30	3 629	2 451	240
BBP 31	3 807	2 515	234
BBP 32	3 613	2 438	222
BBP 33	3 925	2 575	227
BBP 34	3 969	2 615	234
BBP 35	3 747	2 484	225
BBP 36	3 958	2 594	232
BBP 37	3 737	2 509	238
BBP 38	3 569	2 402	225
BBP 39	3 845	2 552	234
BBP 40	4 122	2 681	229
BBP 41	3 957	2 582	232
BBP 42	4 088	2 671	246
BBP 43	4 070	2 638	227
BBP 44	4 062	2 661	239
BBP 45	4 012	2 642	245
BBP 46	4 230	2 753	238
BBP 47	3 999	2 626	236
BBP 48	4 034	2 649	238
BBP 49	4 043	2 648	230
BBP 50	3 910	2 574	233
BBP 51	3 950	2 610	239
BBP 52	3 790	2 503	227
BBP 53	3 859	2 541	236

BBP 54	3 696	2 460	232
BBP 55	3 864	2 546	231
BBP 56	3 853	2 534	230
BBP 57	4 123	2 681	235
BBP 58	3 879	2 567	229
BBP 59	4 002	2 633	233
BBP 60	3 874	2 590	237
BBP 61	3 883	2 560	227
BBP 62	3 829	2 530	227
BBP 63	4 062	2 647	235
BBP 64	4 049	2 617	229
BBP 65	3 739	2 463	224
BBP 67	3 664	2 429	216
BBP 68	3 809	2 513	227
BBP 70	3 766	2 479	224
BBP 71	3 551	2 372	225
BBP 72	3 539	2 323	224
BBP 73	3 785	2 485	222
BBP 74	3 395	2 250	215
BBP 75	3 704	2 454	228
BBP 76	3 746	2 460	231
BBP 77	3 499	2 313	221
BBP 78	3 182	2 167	216
BBP 79	3 897	2 562	228
BBP 80	4 030	2 641	240
BBP 83	3 727	2 454	216
Gallbladder cancer (GBC) 1	3 850	2 569	227
GBC 2	4 016	2 656	229
GBC 3	3 963	2 639	240
GBC 4	3 713	2 500	229
GBC 5	3 932	2 592	231
GBC 6	3 740	2 488	224
GBC 7	3 895	2 590	224
GBC 8	3 741	2 530	234
GBC 9	3 854	2 561	226
GBC 10	3 970	2 635	229
GBC 11	3 991	2 638	232
GBC 12	3 964	2 618	227
GBC 13	3 833	2 566	227
GBC 14	4 035	2 666	233
GBC 15	3 931	2 585	215
GBC 16	3 826	2 545	224
GBC 17	3 732	2 520	230
GBC 18	3 813	2 569	238
GBC 19	3 722	2 467	225
GBC 20	3 783	2 538	238
GBC 21	3 893	2 585	221
GBC 22	3 783	2 514	224
GBC 23	3 994	2 616	221
GBC 24	3 884	2 562	236
GBC 25	3 934	2 593	230

GBC 26	3 627	2 419	227
GBC 27	3 976	2 605	228
GBC 28	4 155	2 679	238
GBC 29	3 675	2 441	220
GBC 30	4 026	2 659	231
GBC 31	4 192	2 715	242
GBC 32	4 060	2 642	231
GBC 33	3 890	2 555	228
GBC 34	3 958	2 623	246
GBC 35	4 141	2 716	234
GBC 36	4 161	2 693	242
GBC 37	4 101	2 676	229
GBC 38	4 171	2 714	240
GBC 39	4 184	2 734	240
GBC 40	4 094	2 700	240
GBC 42	4 007	2 631	235
GBC 43	4 005	2 630	233
GBC 44	3 996	2 615	237
GBC 45	3 692	2 430	223
GBC 46	3 734	2 468	234
GBC 47	3 760	2 453	232
GBC 48	3 620	2 388	216
GBC 49	3 786	2 488	220
GBC 50	3 728	2 451	225
GBC 51	3 933	2 542	227
GBC 52	3 946	2 566	227
GBC 53	3 593	2 403	226
GBC 55	3 952	2 577	230
GBC 56	3 914	2 588	236

Table S3.10: All differentially abundant proteins in gallbladder cancer compared to benign biliary pathology plasma

Protein Groups	Protein Name	No. of Unique Peptides	Average Log_2 Ratio	p-value	q-value
P02741	CRP – C-reactive protein	6	2.650327304	<0.001	<0.001
P02750	A2GL – Leucine-rich alpha-2-glycoprotein	12	1.239250891	<0.001	<0.001
P18428	LBP – Lipopolysaccharide-binding protein	10	1.15874307	<0.001	<0.001
P01011	AACT – Alpha-1-antichymotrypsin	24	1.07655644	<0.001	<0.001
Q06033	ITIH3 – Inter-alpha-trypsin inhibitor heavy chain H3	18	1.05262098	<0.001	<0.001
P02763	A1AG1 – Alpha-1-acid glycoprotein 1	13	1.029932151	<0.001	<0.001
P0DJI8	SAA1 – Serum amyloid A-1 protein	5	1.010432073	<0.001	<0.001
P01009	A1AT – Alpha-1-antitrypsin	32	0.933125857	<0.001	<0.001
Q9Y6R7	FCGBP – IgGFc-binding protein	12	0.843403272	<0.001	<0.001

P01833	PIGR – Polymeric immunoglobulin receptor	16	0.727327957	<0.001	<0.001
P02776	PLF4 – Platelet factor 4	4	0.704884661	<0.05	<0.01
P02649	APOE – Apolipoprotein E	18	0.691933152	<0.001	<0.001
P03952	KLKB1 – Plasma kallikrein	23	-0.681121022	<0.001	<0.001
P01700	LV147 – Immunoglobulin lambda variable 1-47	2	-0.70219583	<0.001	<0.001
P02765	FETUA – Fetuin-A	10	-0.711729896	<0.001	<0.001
P02790	HEMO – Hemopexin	26	-0.739818383	<0.001	<0.001
P02753	RET4 – Retinol-binding protein 4	9	-0.748467682	<0.001	<0.001
P06396	GELS – Gelsolin	26	-0.773859647	<0.001	<0.001
P29622	KAIN – Kallistatin	12	-0.815267411	<0.001	<0.001
P17936	IBP3 – Insulin-like growth factor-binding protein 3	2	-0.889184722	<0.001	<0.001
Q9UGM5	FETUB – Fetuin-B	6	-0.919334394	<0.001	<0.001
P60709	ACTB – Actin, cytoplasmic 1	16	-0.94638678	<0.001	<0.001
P05452	TETN – Tetranectin	7	-0.951549061	<0.001	<0.001
P02647	APOA1 – Apolipoprotein A-I	30	-0.957554639	<0.001	<0.001
P68871	HBB – Haemoglobin subunit beta	15	-0.977033982	<0.001	<0.001
P49908	SEPP1 – Selenoprotein P	3	-0.990550309	<0.001	<0.001
P69905	HBA – Haemoglobin subunit alpha	8	-1.088464667	<0.001	<0.001
O14791	APOL1 – Apolipoprotein L1	10	-1.094391403	<0.001	<0.001
Q96PD5	PGRP2 – Peptidoglycan recognition protein 2	13	-1.125975222	<0.001	<0.001
P02766	TTR – Transthyretin	10	-1.147028578	<0.001	<0.001
P20742	PZP – Pregnancy zone protein	35	-1.170088816	<0.001	<0.001
P27169	PON1 – Serum paraoxonase/arylesterase 1	9	-1.263906253	<0.001	<0.001
P02652	APOA2 – Apolipoprotein A-II	7	-1.474317396	<0.001	<0.001

Table S3.11: Proteins associated with non-metastatic vs metastatic staging for gallbladder cancer plasma patients

Protein Name	**Non-metastatic Raw Value Mean**	**Metastatic Raw Value Mean**	**p-value**
APOA1 – Apolipoprotein A-I	1180.14	1497.55	0.1275
APOA2 – Apolipoprotein A-II	80.07	146.11	0.2348
RET4 – Retinol-binding protein 4	26.12	20.39	0.5730
TTR - Transthyretin	205.75	197.60	0.7853
HEMO - Hemopexin	1033.80	1457.41	0.1474
HBB – Haemoglobin subunit beta	2196.72	1183.42	0.6761
HBA – Haemoglobin subunit alpha	697.44	379.40	0.7029
PIGR – Polymeric immunoglobulin receptor	39.21	35.96	0.7029
APOE* – Apolipoprotein E	926.01	552.41	0.0143*
CRP – C-reactive protein	461.85	274.26	0.8990

GELS – Gelsolin	31.96	34.29	0.9855
APOL1 – Apolipoprotein L1	11.40	11.52	0.6497
A1AT – Alpha-1-antitrypsin	40335.51	30706.32	0.1098
AACT – Alpha-1-antichymotrypsin	3379.44	2353.12	0.3526
LV147 – Immunoglobulin lambda variable 1-47	84.55	53.27	0.5009
A2GL – Leucine-rich alpha-2-glycoprotein	705.06	498.34	0.2348
A1AG1 – Alpha-1-acid glycoprotein 1	1982.38	1477.00	0.5244
FETUA – Fetuin-A	430.90	476.94	0.2348
PLF4 – Platelet factor 4	11.28	3.52	0.1653
KLKB1 – Plasma kallikrein	27.03	26.95	0.7029
TETN – Tetranectin	4.28	5.17	0.4779
SAA1 – Serum amyloid A-1 protein	68.78	62.65	0.6995
IBP3 – Insulin-like growth factor-binding protein 3	7.82	9.94	0.5358
LBP – Lipopolysaccharide-binding protein	49.98	24.60	0.3919
PZP – Pregnancy zone protein	17.63	23.37	0.3426
PON1 – Serum paraoxonase/arylesterase 1	38.29	52.04	0.1184
KAIN – Kallistatin	6.14	5.91	0.8079
SEPP1 – Selenoprotein P	2.83	3.36	0.5855
ACTB – Actin, cytoplasmic 1	10.81	22.42	0.0940
ITIH3* – Inter-alpha-trypsin inhibitor heavy chain H3	94.64	60.16	0.0173*
PGRP2 – Peptidoglycan recognition protein 2	7.39	7.16	0.4337
FETUB – Fetuin-B	24.49	27.70	0.3460
FCGBP – IgGFc-binding protein	3.68	3.31	0.7606

Table S3.12: Commonly dysregulated proteins across group comparisons with relative average Log_2 fold changes

Common Proteins in GBC/Normal & GBC/GD	**Average Log_2 Fold Change**		**Common Proteins in GBC/GD & GBC/BBP**	**Average Log_2 Fold Change**	
	GBC/Norm	**GBC/GD**		**GBC/GD**	**GBC/BBP**
CAH2 – Carbonic anhydrase 2	-3.359399413	-4.035396298	**APOA1 – Apolipoprotein A-I**	-2.18857284	-0.957554639
JIP4 – C-Jun-amino-terminal kinase-interacting protein 4	-4.856406351	-4.346847515	**APOA2 – Apolipoprotein A-II**	-2.910543359	-1.474317396
VTNC – Vitronectin	3.915829431	1.508175969	**RET4 – Retinol-binding protein 4**	-1.68830487	-0.748467682
H2A1B; H2A3; H2A1C – Histone H2A type 1-B/E; Histone H2A type 3; Histone H2A type 1-C	3.601816369	-1.597290839	**TTR – Transthyretin**	-2.796960584	-1.147028578
TPM2 – Tropomyosin beta chain	-2.980807556	-2.841326704	**HEMO – Hemopexin**	-1.618488319	-0.739818383

TSP1 – Thrombospondin-1	3.251710872	3.003882122	**HBB – Haemoglobin subunit beta**	-3.465642881	-0.977033982	
TPM1 – Tropomyosin alpha-1 chain	-2.60007637	-2.832043183	**HBA – Haemoglobin subunit alpha**	-3.769500318	-1.088464667	
RRAS – Ras-related protein R-Ras	-2.788038724	-1.986053122				
PCNA – Proliferating cell nuclear antigen	3.172973883	2.479580402	**Common Proteins in GBC/Normal,**	**Average Log_2 Fold Change**		
RAC2 – Ras-related C3 botulinum toxin substrate 2	2.737986073	1.675119445	**GBC/GD, & GBC/BBP**	**GBC/Normal**	**GBC/GD**	**GBC/BBP**
DESM – Desmin	-3.441316866	-4.845020339	**PIGR – Polymeric immunoglobulin receptor**	2.898754081	2.196173212	0.727327957
TYPH – Thymidine phosphorylase	2.561671217	1.795493292	**APOE – Apolipoprotein E**	2.968525542	2.088429222	0.691933152
TRYB2; TRYB1 – Tryptase beta-2; Tryptase alpha/beta-1	-2.800721074	-3.371269387				
MIME – Mimecan	-2.895872701	-3.70936941				
FLNA – Filamin-A	-2.489748582	-2.187590377				
TENA – Tenascin	2.712643782	1.763552536				
MYL9 – Myosin regulatory light polypeptide 9	-3.628772097	-3.939013593				
GLYM – Serine hydroxymethyltransferase	2.51078038	1.862362934				
MYH11 – Myosin-11	-3.652336209	-3.238178757				
CNN1 – Calponin-1	-2.741320166	-3.789798779				
RIDA – 2-iminobutanoate/2-iminopropanoate deaminase	3.70928337	3.587142384				
AK1C2 – Aldo-keto reductase family 1 member C2	3.848081936	4.953323214				
ACTA – Actin, aortic smooth muscle	-2.492619937	-2.608621477				
TAGL – Transgelin	-3.372011937	-3.075365695				
BDH – D-beta-hydroxybutyrate dehydrogenase	2.900769273	3.638123672				
CAV1 – Caveolin-1	-2.651299406	-3.408389459				
POSTN – Periostin	5.02066359	2.216168479				
CAVN1 – **Caveolae-associated protein 1**	-3.822693564	-3.521064707				
DCXR – L-xylulose reductase	2.983416149	2.844273258				
COCA1 – Collagen alpha-1(XII) chain	4.004558467	3.374782172				
PARVB – Beta-parvin	-3.106309788	-1.560212729				
EHD2 – EH domain-containing protein 2	-3.475631992	-3.733255327				
CPSM – Carbamoyl-phosphate synthase	4.483369762	4.323381454				
Norm = Normal						
GD = Gallstone Disease						

GBC = Gallbladder Cancer
BBP = Benign Biliary Pathology

Table S3.13: Protein coverage and number of samples identified in for the commonly dysregulated proteins

Protein	Protein Sequence Coverage			Number of Samples Identified In			CVs	
	GBC/GD	GBC/BBP	GBC/Normal	GBC/GD	GBC/BBP	GBC/Normal	GBC Plasma	BBP Plasma
APOA1 – Apolipoprotein A-I	73%	72%		100%	100%		89%	51.2%
APOA2 – Apolipoprotein A-II	64%	42%		100%	100%		110.6%	53.8%
RET4 – Retinol-binding protein 4	41%	59%		97.5%	100%		76%	51.3%
TTR – Transthyretin	69%	69%		100%	100%		93.7%	54.8%
HEMO – Hemopexin	51%	71%		100%	100%		67.2%	39.6%
HBB – Haemoglobin subunit beta	96%	96%		100%	100%		165.4%	273.6%
HBA – Haemoglobin subunit alpha	96%	65%		100%	100%		168.2%	273.4%
PIGR – Polymeric immunoglobulin receptor	36%	24%	35%	75%	93.7%	81.2%	56.4%	73.6%
APOE – Apolipoprotein E	64%	59%	64%	100%	100%	100%	83.2%	81.1%

APPENDIX A: TURNITIN REPORT

Thesis_v6-7.docx

ORIGINALITY REPORT

26%	23%	23%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet Source	13%
2	link.springer.com Internet Source	2%
3	clinicalproteomicsjournal.biomedcentral.com Internet Source	1%
4	researchonline.ljmu.ac.uk Internet Source	<1%
5	irep.ntu.ac.uk Internet Source	<1%
6	storage.googleapis.com Internet Source	<1%
7	cco.amegroups.com Internet Source	<1%
8	www.omicsdi.org Internet Source	<1%
9	Submitted to University of Salford Student Paper	<1%

Ekene Nweke 12:08 (30 minutes ago)
to me

Dear Pavan,
Your overall similarity index is 26% with a maximum single match of 13% followed by 2%. These individual matches are from your 2 published articles from your PhD. All other matches are ≤ 1%. Therefore, your similarity index is **ACCEPTABLE** within our Department.
For PG submissions - Please only submit the first page of your turnitin report, as well as this email stating that your plagiarism report is **ACCEPTABLE** within the Department of Surgery.

Kind regards
Emmanuel

Dr. Ekene Emmanuel Nweke (PhD), *Pri Sci Nat*
Ass. Research Scientist and MMED Research Coordinator
Department of Surgery, Room 9M07

WITS UNIVERSITY

Tel: +27(0)11 717 2801
Email: Ekene.Nweke@wits.ac.za
Website: www.wits.ac.za

WITS SCHOOL OF CLINICAL MEDICINE
FACULTY OF HEALTH SCIENCES 100

APPENDIX B: ETHICS CLEARANCE CERTIFICATE

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG	HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

Office of the Deputy Vice-Chancellor (Research & Post Graduate Affairs)

TO: Mr P Baichan
School of Clinical Medicine
Department of Surgery
Chris Hani Baragwanath Academic Hospital

E-mail: 677795@students.wits.ac.za

CC: Supervisor: Professor G Candy, et al <Geoffrey.Candy@wits.ac.za>
and <HREC-Medical.ResearchOffice@wits.ac.za>

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

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

DATE: 2019/10/09

REF: *R14/49*

PROTOCOL NO: **M190555** *(This is your ethics application study reference number. Please quote this reference number in all correspondence relating to this study)*

PROJECT TITLE: *Potential protein biomarker discovery for gallbladder cancer in a South African cohort*

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 the Government funding of the University.

MSWorks2000/Iain0007/Clearscan.wps

APPENDIX C: STUDY INFORMATION



STUDY INFORMATION DOCUMENT

Study title: <u>Potential protein biomarker discovery for gallbladder cancer in a South African cohort</u>

Greetings Sir/Madam:

Introduction:
Good day, we, Prof. Geoffrey Candy and Mr. Pavan Baichan, are doing research on gallbladder carcinoma. Research is a process used in seeking new knowledge. In this study we want to learn the protein expression profile of gallbladder carcinoma. This will be done so by drawing blood samples as well as tumour biopsies. This will allow us to study to protein expression for gallbladder carcinoma and potentially develop a new, simpler diagnostic technique in order to diagnose Gallbladder carcinomas sooner to provide better quality care.

This is a research study and is not part of your normal and routine care which you will receive while in hospital. The study hopes to find out why the cancer starts, what makes it different, and can we find ways to best kill it by using drugs or other ways.

Invitation to Participate: In order for us to study the cancer, we are asking you to take part in a research study.

What is involved in the study: You have been diagnosed with gallbladder carcinoma, cholecystitis, or have been invited to volunteer as you have no gallbladder disease. The doctors would have told you that they need to biopsy your gallbladder in order to determine a way forward in treating your illness. When they perform this, we are asking your permission to take one additional small (5 mm X 5 mm) piece of the cancer or gallbladder (normal or inflamed). We will then study the differences in the proteins between the normal, inflamed, and cancer tissues. We would do this by separating and identifying the proteins using laboratory methods and instruments.
In this way we hope to find differences between the proteins. If we are fortunate and successful, these proteins could be targeted by removing, destroying, or inactivating them using drugs. If so we may find a marker for the cancer or even slow or stop the cancer from growing.

Risks of being involved in the study: We are taking pieces of the cancer, inflamed, and normal tissue from the gallbladder. There is no risk to you directly. Furthermore, we will only do this if the treating surgeon doing the procedure says we may safely do this. If the surgeon decides it is not safe to take the samples, we will not take the sample or include you in the study.

Benefits of being in the study: This is a research study. We do not know at this time what we will find, if anything. Research studies like this can take years to complete and we need samples from a number of patients to make sure we are seeing the same differences in all the patients. So at this time you will not benefit from taking part in this study. However, if you do take part, we may find better ways to treat or diagnose other patients in the future. However, should there

be a discovery which may benefit you we would contact your treating doctor who would contact you through your clinic.

Participation is voluntary: If you decide to not participate in the study, this is your right and this will not influence the way you are treated while in hospital. If you decline to take part, we simply will not take the samples indicated. If you change your mind later, we would destroy the samples we have taken. Again there will be no penalty or loss of benefits or care that you receive. Either way you will need to contact Prof. Geoffrey Candy to ensure that this happens – the contact details are below. Finally, your participation in this study will not cost you anything.

Reimbursements: Participation in the study is entirely voluntary and you would not be paid for taking part in the study. None of the scientists or doctors are being paid if you decide to participate in the study.

Confidentiality: As with any medical records, every effort will be made to keep all your personal information confidential. Your name will not be disclosed to any external third parties. In order to keep your identify confidential, a specialised identity code will be assigned to you; only the investigators of this study will have access to the files which links your name to the specialised identity code. However, absolute confidentiality cannot be guaranteed and personal information may be disclosed if required by law. Organisations such as the Research Ethics Committee of the University of the Witwatersrand may inspect and/or copy your research records for quality assurance and data analysis where this is required (this includes groups such as the Medicines Control Council.
For this study we are only interested in the cancer and diagnosis, not in names and personal information. It is the intention to publish the data, showing proteins which may be of interest for all the patients in the group – not that of individuals participating in the study.

Contact details of researcher/s: If you have further queries or questions, Prof. Geoffrey Candy can be contacted on 011-717 2574 or 083-226 2436

Outputs: As previously mentioned, it is the intention to publish the data, showing proteins which may be of interest for all the patients in the group – not that of individuals participating in the study.

Contact details of HREC administrator and chair:
This study has been approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand, Johannesburg ("Committee"). A principal function of this Committee is to safeguard the rights and dignity of all human subjects who agree to participate in a research project and the integrity of the research.
If you have any concern over the way the study is being conducted, please contact the Chairperson of this Committee who is Professor Clement Penny, who may be contacted on telephone number 011 717 2301, or by e-mail on Clement.Penny@wits.ac.za. The telephone numbers for the Committee secretariat are 011 717 2700/1234 and the e-mail addresses are Zanele.Ndlovu@wits.ac.za and Rhulani.Mukansi@wits.ac.za

Thank you for reading this Study Information Sheet.

Date: February 2019

APPENDIX D: INFORMED CONSENT (GALLBLADDER CANCER PATIENTS)

Consent Form: Protein expression profiling of Gallbladder Carcinoma in a South African Cohort

Dear Patient,
You are currently admitted to our hospital and have been diagnosed with gallbladder cancer. The Hospital not only renders treatment but is also actively involved in conducting research aimed at improving the quality of care that we deliver. From time to time such research involves the use of patient records from which information is extracted. We will also in this study be using a tissue and blood sample taken via percutaneous biopsy and venously for protein expression profiling of the gallbladder. This sampling will not jeopardize your safety and will be part of the standard processes for diagnosing your tumour. The use of such information and sample is subject to the following:

1. Approval from the Human Research Ethics Committee (Medical) of the University of the Witwatersrand.
2. As with any medical records, every effort will be made to keep all your personal information confidential. Your name will not be disclosed to any external third parties. In order to keep your identify confidential, a specialised identity code will be assigned to you; only the investigators of this study will have access to the files which links your name to the specialised identity code. However, absolute confidentiality cannot be guaranteed and personal information may be disclosed if required by law. Organisations such as the Research Ethics Committee of the University of the Witwatersrand may inspect and/or copy your research records for quality assurance and data analysis where this is required for this study we are only interested in the cancer and diagnosis, not in names and personal information. It is the intention to publish the data, showing proteins which may be of interest for all the patients in the group – not that of individuals participating in the study.

We would like to obtain your consent to use information from your file for the purpose of research, subject to the aforementioned conditions. If you choose not to give consent, this will not compromise your treatment in any way. If at any time you choose to withdraw consent you are free to do so and will not be prejudiced in any way.

Should you wish to contact us at any stage regarding consent, contact Dr J Devar at (011) 9339267.

A. Consent Given

I_____ hereby give consent for my records to be used as per the above mentioned conditions for the purposes of research:

PATIENT: _____
Name Signature Date

RESEARCHER: _____
Name Signature Date

WITNESS: _____
Name Signature Date

APPENDIX E: INFORMED CONSENT (GALLSTONE DISEASE PATIENTS)

Consent Form: Protein expression profiling of Gallstone disease in a South African Cohort

Dear Patient,
You are currently admitted to our hospital and have been diagnosed with cholelithiasis (gallstone disease). The Hospital not only renders treatment but is also actively involved in conducting research aimed at improving the quality of care that we deliver. From time to time such research involves the use of patient records from which information is extracted. We will also in this study be using a tissue and blood sample taken via percutaneous biopsy and venously for protein expression profiling of the gallbladder. This sampling will not jeopardize your safety and will be part of the standard processes for diagnosing your tumour. The use of such information and sample is subject to the following:

3. Approval from the Human Research Ethics Committee (Medical) of the University of the Witwatersrand.
4. As with any medical records, every effort will be made to keep all your personal information confidential. Your name will not be disclosed to any external third parties. In order to keep your identify confidential, a specialised identity code will be assigned to you; only the investigators of this study will have access to the files which links your name to the specialised identity code. However, absolute confidentiality cannot be guaranteed and personal information may be disclosed if required by law. Organisations such as the Research Ethics Committee of the University of the Witwatersrand may inspect and/or copy your research records for quality assurance and data analysis where this is required For this study we are only interested in the cancer and diagnosis, not in names and personal information. It is the intention to publish the data, showing proteins which may be of interest for all the patients in the group – not that of individuals participating in the study.

We would like to obtain your consent to use information from your file for the purpose of research, subject to the aforementioned conditions. If you choose not to give consent, this will not compromise your treatment in any way. If at any time you choose to withdraw consent you are free to do so and will not be prejudiced in any way.

Should you wish to contact us at any stage regarding consent, contact Dr J Devar at (011) 9339267.

B. Consent Given

I_____ hereby give consent for my records to be used as per the above mentioned conditions for the purposes of research:

PATIENT: _____
Name Signature Date

RESEARCHER: _____
Name Signature Date

WITNESS: _____
Name Signature Date

APPENDIX F: INFORMED CONSENT (NORMAL GALLBLADDER PATIENTS)

Consent Form: Protein expression profiling of Gallbladder Carcinoma in a South African Cohort

Dear Patient,
You are currently admitted to our hospital for treatment of problems you are currently experiencing. The Hospital not only renders treatment but is also actively involved in conducting research aimed at improving the quality of care that we deliver. From time to time such research involves the use of patient records from which information is extracted. We will also in this study be using a tissue and blood sample taken via percutaneous biopsy and venously for protein expression profiling of the gallbladder. You have not been diagnosed with any gallbladder disease, so your sampling would serve as a normal sample group in which to compare against. This sampling will not jeopardize your safety and will be part of the standard processes for diagnosing your tumour. The use of such information and sample is subject to the following:

5. Approval from the Human Research Ethics Committee (Medical) of the University of the Witwatersrand.
6. As with any medical records, every effort will be made to keep all your personal information confidential. Your name will not be disclosed to any external third parties. In order to keep your identify confidential, a specialised identity code will be assigned to you; only the investigators of this study will have access to the files which links your name to the specialised identity code. However, absolute confidentiality cannot be guaranteed and personal information may be disclosed if required by law. Organisations such as the Research Ethics Committee of the University of the Witwatersrand may inspect and/or copy your research records for quality assurance and data analysis where this is required For this study we are only interested in the cancer and diagnosis, not in names and personal information. It is the intention to publish the data, showing proteins which may be of interest for all the patients in the group – not that of individuals participating in the study.

We would like to obtain your consent to use information from your file for the purpose of research, subject to the aforementioned conditions. If you choose not to give consent, this will not compromise your treatment in any way. If at any time you choose to withdraw consent you are free to do so and will not be prejudiced in any way.

Should you wish to contact us at any stage regarding consent, contact Dr J Devar at (011) 9339267.

C. Consent Given

I_____ hereby give consent for my records to be used as per the above mentioned conditions for the purposes of research:

PATIENT: _____
Name Signature Date

RESEARCHER: _____
Name Signature Date

WITNESS: _____
Name Signature Date

APPENDIX G: PERMISSION TO CONDUCT RESEARCH

GAUTENG PROVINCE
HEALTH
REPUBLIC OF SOUTH AFRICA

MEDICAL ADVISORY COMMITTEE

CHRIS HANI BARAGWANATH ACADEMIC HOSPITAL

PERMISSION TO CONDUCT RESEARCH

Date: 13th February 2019

TITLE OF PROJECT:

Potential Protein Biomarker Discovery for Gallbladder Cancer in a South African Cohort.

UNIVERSITY: Witswatersrand

Principal Investigator: Dr Baichan

Department: Surgery

Supervisor : Prof G Candy

Permission Head Department (where research conducted): Yes

The Medical Advisory Committee recommends that the said research be conducted at Chris Hani Baragwanath Academic Hospital. The CEO / management of Chris Hani Baragwanath Academic Hospital is accordingly informed and the study is subject to:-

- **Permission having been granted by the Committee for Research on Human Subjects of the University of the Witwatersrand.**
- The Hospital will not incur extra costs as a result of the research being conducted on its patients within the hospital
- The MAC will be informed of any serious adverse events as soon as they occur
- Permission is granted for the duration of the Ethics Committee Approval.

.....
Recommended
(On behalf of the MAC)
Date: 13/02/2019.

.....
Approved/Not Approved
Hospital Management
Date: 14/02/2019