Exploration of Urine and Plasma Biomarkers in Liver Fibrosis and Hepatocellular Carcinoma

by

Ayman Sharafeldin Bannaga
MB BS MRCP PGCert Med Ed

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

University of Warwick, Warwick Medical School
September 2021
بِسْمِِ ٱللَّهِِ ٱلرَّحْمَـٰنِ ٱلرَّحِيمِ

In the name of Allah, Most Merciful, Most Compassionate

الرُّسُخُوۡنَ فِي ٱلۡعِلَّمِ يُقُولُوۡنَ أَمِنًا بِهِ كُلُّ مِّن ٍّ عِنۡدِ رَبِّي وَمَا يِذَّكَرُ

إِلَّا أُولُو ٱلۡلَّبَـٰبِ

But those firm in knowledge say, "We believe in it. All [of it] is from our Lord." And no one will be reminded except those of understanding - Quran 3:7

وَقُولِْ رَبِّي ذِي عِلْمٍ عَلِيمٍ

And above every knowing one is a Knower - Quran 12:76

وَقُلِْ رَبِّ زَدْنِي عِلَّمًا

And say: My Lord! Increase me in knowledge - Quran 20:114
This thesis is dedicated to my daughter Farah and to my parents Sharafeldin and Sayda
# Table of Contents

I. List of Abbreviations 7

II. List of tables and figures 12

III. Acknowledgment 18

IV. Declaration 19

V. List of Publications 20

VI. Awards and educational bursaries 21

VII. Presentations 22

VIII. Abstract 23

## Chapter 1. Introduction 24

1.1 Overview 25

1.2 Origin of the liver 26

1.3 Macro-anatomy of the liver 27

1.4 Liver segments 27

1.5 Hepatic lobule 28

1.6 Perisinusoidal space 28

1.7 Extracellular skeleton of the liver tissue 28

1.8 Liver functions 31

1.9 The liver and functional immunity 33

1.9.1 Kupffer cells 33

1.9.2 Dendritic cells 34

1.9.3 Resident liver lymphocytes 34

1.9.4 Hepatic stellate cells 34

1.10 The liver and disease 35

1.11 Liver fibrosis 35

1.12 Pathogenesis of liver fibrosis 36

1.13 Logistic regression models for predicting liver fibrosis and cirrhosis 39

1.14 Enhanced liver fibrosis blood test 39

1.15 Imaging in liver fibrosis 40

1.16 Diagnostic pathways in liver fibrosis 42

1.17 Biopsy of the liver 44

1.18 Hepatocellular carcinoma 46

1.19 Pathogenesis of hepatocellular carcinoma 50

1.19.1 Telomere maintenance 52

1.19.2 Wnt and β catenin pathway 52

1.19.3 P53 inactivation 52

1.19.4 Chromatin remodeling 52
1.19.5 Activation of the reactive oxygen species (ROS) pathway  52
1.19.6 Mitogen activated protein kinase (MAPK) pathway  53
1.20 Diagnosis of hepatocellular carcinoma  55
1.21 Surveillance of hepatocellular carcinoma  59
1.22 Explored biomarkers in hepatocellular carcinoma  59
1.23 Treatment of hepatocellular carcinoma  60
1.24 Prognosis of hepatocellular carcinoma  62
1.25 Proteomic analysis  68
1.26 Volatile organic compounds (VOCs)  69
1.27 Liquid biopsy and DNA methylation  73
1.28 Methylated SEPTIN9 (mSEPT9)  73
1.29 Research justification  73
1.30 Research questions  74

Chapter 2. Methodology  75
2.1 Research design  76
2.2 Inclusion criteria  76
2.3 Exclusion criteria  77
2.4 Ethical considerations  77
2.5 Recruitment  78
2.6 Research procedures  78
2.7 Capillary electrophoresis coupled with mass spectrometry (CEMS)  81
2.7.1 Urine sample preparation for CEMS  81
2.7.2 CEMS urine analysis  81
2.7.3 CEMS data processing  82
2.7.4 Support vector machine model generation and classification  82
2.7.5 Peptide sequencing  83
2.7.6 In-silico protease assessment  83
2.7.7 Tissue transcriptomic data analysis  83
2.7.8 Statistical analysis for CEMS  84
2.9 Immunohistochemistry  86
2.10 Techniques used in volatile organic compounds analysis  89
2.11 Solid-phase microextraction (SPME)  91
2.11.1 Urine sample processing using SPME  92
2.11.2 SPME data processing  93
2.11.3 SPME data analysis  94
2.12 Gas chromatography coupled with ion mobility spectrometry (GC-IMS)  95
2.13 Gas chromatography time-of-flight mass spectrometry (GC-TOF-MS)  96
I. **List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBP</td>
<td>4-Methyl-2,4-bis(p-hydroxyphenyl)pent-1-ene, 2TMS derivative</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>ALT</td>
<td>Alanine aminotransferase</td>
</tr>
<tr>
<td>ALBI</td>
<td>Albumin-bilirubin grade</td>
</tr>
<tr>
<td>ALP</td>
<td>Alkaline phosphatase</td>
</tr>
<tr>
<td>AFP</td>
<td>Alpha fetoprotein</td>
</tr>
<tr>
<td>AA</td>
<td>Amino acid sequence</td>
</tr>
<tr>
<td>NH4OH</td>
<td>Ammonium hydroxide</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under the curve</td>
</tr>
<tr>
<td>APRI</td>
<td>Aspartate aminotransferase to platelet ratio</td>
</tr>
<tr>
<td>AST</td>
<td>Aspartate aminotransferase</td>
</tr>
<tr>
<td>ARID2</td>
<td>AT-rich interactive domain-containing protein 2</td>
</tr>
<tr>
<td>AIH</td>
<td>Autoimmune hepatitis</td>
</tr>
<tr>
<td>BCLC</td>
<td>Barcelona clinic liver cancer staging system</td>
</tr>
<tr>
<td>ACTB</td>
<td>Beta actin</td>
</tr>
<tr>
<td>BPA</td>
<td>Bisphenol A</td>
</tr>
<tr>
<td>BCA</td>
<td>Bladder carcinoma cases</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
</tr>
<tr>
<td>CDH</td>
<td>Cadherin protein</td>
</tr>
<tr>
<td>CLIP</td>
<td>Cancer of the liver Italian programme staging system</td>
</tr>
<tr>
<td>CEMS</td>
<td>Capillary electrophoresis mass spectrometry</td>
</tr>
<tr>
<td>CCl4</td>
<td>Carbon tetrachloride</td>
</tr>
<tr>
<td>CTNNB1</td>
<td>Catenin Beta 1 gene</td>
</tr>
<tr>
<td>CTS</td>
<td>Cathepsin</td>
</tr>
<tr>
<td>cfDNA</td>
<td>Cell free deoxyribonucleic acid</td>
</tr>
<tr>
<td>CTP</td>
<td>Child Turcotte Pugh score</td>
</tr>
<tr>
<td>CLD</td>
<td>Chronic liver disease</td>
</tr>
<tr>
<td>ctDNA</td>
<td>Circulating tumour deoxyribonucleic acid</td>
</tr>
<tr>
<td>CRN</td>
<td>Clinical research network</td>
</tr>
<tr>
<td>CLU</td>
<td>Clusterin</td>
</tr>
<tr>
<td>COL</td>
<td>Collagen</td>
</tr>
<tr>
<td>CT</td>
<td>Computed tomography</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence intervals</td>
</tr>
<tr>
<td>COVID19</td>
<td>Coronavirus disease-19</td>
</tr>
<tr>
<td>Ct</td>
<td>Cycle threshold</td>
</tr>
<tr>
<td>P450</td>
<td>Cytochrome</td>
</tr>
</tbody>
</table>
Cytosine preceding a guanine
Deoxyribonucleic acid
Des-gamma-carboxy-prothrombin
Dihydronicotinamide-adenine dinucleotide phosphate
Dynamic baseline correction
Enhanced liver fibrosis
Enzyme linked immunosorbent assay
Epidermal growth factor
Ethylendiaminetetraacetic acid
Extracellular matrix
False discovery rate
False positive rate
Female
Femtomole
Fibrinogen alpha chain
Fibrosis index 4
Gamma-glutamyl transferase
Gas chromatography ion mass spectrometry
Gas chromatography time of flight mass spectrometry
Gelsolin
Glypican 3
Haematoxylin and Eosin
Haemoglobin
Health research authority
Hepatic dendritic cells
Hepatic stellate cells
Hepatitis B virus
Hepatitis C virus
Hepatocellular carcinoma
Hepatocellular carcinoma-31 model
Hepatocyte growth factor
Hepato-pancreatico-biliary multidisciplinary team
High density lipoprotein
Human immunodeficiency virus
Human liver cancer cell line
Hyaluronic acid
Hydrogen peroxide

CpG
DNA
DCP
NADPH
DBS
ELF
ELISA
EGF
EDTA
ECM
FDR
FPR
F
fmol
FGA
FIB-4
GGT
GC-IMS
GC-TOF-MS
GSN
GPC3
H&E
Hb
HRA
HDCs
HSCs
HBV
HCV
HCC
HCC-31
HGF
HPB-MDT
HDL
HIV
HepG2
HA
H2O2
Immunohistochemistry (IHC)
Interferon gamma (IFγ)
Interleukin (IL)
International Liver Cancer Association (ILCA)
International normalisation ratio (INR)
Kallikrein related peptidase 6 (KLK6)
Kilo Dalton (kDa)
Kilopascal (kPa)
Lens culinaris agglutinin (LCA)
Linear regression (LR)
Lipopolysaccharides (LPS)
Liver cirrhosis (LC)
Liver fibrosis 50 model (LivFib-50)
Liver function tests (LFTs)
Liver imaging reporting and data system (LI-RADS)
Magnetic Resonance imaging (MRI)
Major histocompatibility complex (MHC)
Male (M)
Matrix metalloproteinases (MMPs)
Meprin A subunit alpha (MEP1A)
Metal oxide sensor (MOS)
Methylated SEPTIN9 gene (mSEPT9)
Micro ribonucleic acid (miRNAs)
Mitogen activated protein kinase (MAPK)
Model for end stage liver disease (MELD)
National institute for health care and clinical excellence (NICE)
National institute of health research (NIHR)
National institute of standards and technology (NIST)
Negative predictive value (NPV)
Neuroblast differentiation-associated protein (AHNAK)
Neutrophil to lymphocyte ratio (NLR)
Nitric oxide (NO)
Non-alcoholic fatty liver disease (NAFLD)
Non-alcoholic steatohepatitis (NASH)
Non-alcoholic steatohepatitis cirrhosis (NASH-C)
Oestrogen receptor (ER)
Osteopontin (OSN)
Patient information sheet
Percentage of methylated reference
Peroxisome proliferator activated receptor alpha
Peroxisome proliferator activated receptor gamma
Platelet derived growth factor
Platelets
Positive predictive value
Primary biliary cholangitis
Primary sclerosing cholangitis
Principal component analysis
Prostate cancer cases
Prostate specific antigen
Protein induced by vitamin K absence-II
Radial basis function network
Radiofrequency Ablation
Random Forrest
Reactive oxygen species
Real time polymerase chain reaction
Receiver operating characteristic
Selective internal radiation therapy
Serum separating tube
Septin 9 gene
Solid phase microextraction
Squamous cell carcinoma antigen
standard error
Support vector machine
Telomerase reverse transcriptase
The fatty liver and liver cancer study
Thioredoxins
Tissue inhibitor of metalloproteinase-1
Toll like receptors
Total cholesterol
Transarterial chemoembolisation
Transarterial radioembolisation
Transforming growth factor beta 1
Tricarboxylic acid
Triglycerides

PIS
PMR
PPARα
PPARγ
PDGF
PLT
PPV
PBC
PSC
PCA
PCA
PSA
PIVKA-II
RBFN
RFA
RF
ROS
rtPCR
ROC
SIRT
SST
SEPTIN9
SPME
SCA
SE
SVM
TERT
TENDENCY
TRXs
TIMP1
TLRs
TC
TACE
TARE
TGFβ1
TCA
TG
<table>
<thead>
<tr>
<th>Tris-buffered saline</th>
<th>TBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positive rate</td>
<td>TPR</td>
</tr>
<tr>
<td>Tumour necrosis factor alpha</td>
<td>TNFα</td>
</tr>
<tr>
<td>Tumour, node metastases staging system</td>
<td>TNM</td>
</tr>
<tr>
<td>Type III procollagen peptide</td>
<td>PIIINP</td>
</tr>
<tr>
<td>Ultrasound scan</td>
<td>USS</td>
</tr>
<tr>
<td>United Kingdom model for end stage liver disease</td>
<td>UKELD</td>
</tr>
<tr>
<td>United States food and drug administration</td>
<td>FDA</td>
</tr>
<tr>
<td>University Hospital Coventry and Warwickshire</td>
<td>UHCW</td>
</tr>
<tr>
<td>Vascular endothelial growth factor</td>
<td>VEGF</td>
</tr>
<tr>
<td>Vibration controlled transient elastography</td>
<td>VCTE</td>
</tr>
<tr>
<td>Volatile organic compounds</td>
<td>VOCs</td>
</tr>
<tr>
<td>White blood cells</td>
<td>WBCs</td>
</tr>
</tbody>
</table>
II. List of Tables and Figures

Chapter 1 Tables:

Table 1.1  Elastography modalities in different liver diseases
Table 1.2  Histopathological liver fibrosis grading systems
Table 1.3  Biomarker study designs and methodologies
Table 1.4  VOCs analysis instruments
Table 1.5  Commonly occurring VOCs in HCC

Chapter 1 Figures:

Figure 1.1  The ultrastructure of the normal human liver
Figure 1.2  Haematoxylin and Eosin (H&E) stain of normal liver tissue section
Figure 1.3  Features of quiescent and activated HSCs
Figure 1.4  Noninvasive approach for assessment of liver fibrosis
Figure 1.5  HCC observed and projected age-standardised incidence rates in the UK
Figure 1.6  HCC average number of deaths per year in the UK
Figure 1.7  Diagnosed patients with HCC in UHCW
Figure 1.8  Cell division cycle
Figure 1.9  Molecular pathway for HCC development.
Figure 1.10 Radiological features of HCC on triple phase CT scan
Figure 1.11 Haematoxylin and Eosin (H&E) stain of HCC tissue section
Figure 1.12 BCLC staging system
Figure 1.13 Survival curves comparing HCC patients with high and low NLR
Figure 1.14 Survival curves comparing HCC patients as per ALBI grades
Figure 1.15 ROC curves comparing CTP to ALBI in HCC patients

Chapter 2 Tables:

Table 2.1  Samples used for experimental work in this research project
Table 2.2  Observational comparison between SPME, GC-IMS and GC-TOF-MS

Chapter 2 Figures
Figure 2.1  Research procedures and hypotheses for the experimented work
Figure 2.2  CEMS steps for analyzing a urine sample
Figure 2.3  Steps followed in identifying HCC-31 biomarker (CEMS and Immunohistochemistry)
Figure 2.4a SPME Sensor tab
Figure 2.4b SensAm Instrument showing tab inserted for desorption on the sensor array
Figure 2.5 SPME preconcentration of urine volatile components on SensAm Tab.
Figure 2.6 Schematic flow chart showing steps for urine data acquisition and data analysis of SPME gas sensor signals.
Figure 2.7 FlavourSpec GC-IMS (G.A.S, Dortmund, Germany)
Figure 2.8 GC-TOF-MS system (TRACE 1300 GC Thermo Fisher Scientific, Loughborough, UK and BenchTOF-HD TOF-MS Markes Intl., Llantrisant, UK).
Figure 2.9 Schematic flow chart showing the process of acquiring data from urine samples and steps involved in statistical analyses in GC-TOF-MS and GC-IMS.
Figure 2.10 Schematic flow graphical summary showing hypothesis and all steps followed in analysing mSEPT9 in HCC.

Chapter 3 Tables:

Table 3.1 Characteristics of recruited participants investigated by CEMS to generate the LivFib-50 urinary peptide biomarker in liver fibrosis
Table 3.2 CEMS characteristics and descriptive statistics of the 50 urinary peptides included in the LivFib-50 peptide marker
Table 3.3 Amino acid alignment of all sequence-identified naturally occurring urinary peptides included in the LivFib-50 peptide marker
Table 3.4 List of protease cleavage sites for the sequence-identified LivFib-50 peptide markers
Table 3.5 Tissue transcriptomics data of 22 cirrhotic and 14 normal human livers retrieved from the NCBI GEO database (GSE77627) to compare it with the proteomic results reached in this CEMS LivFib-50
Chapter 3 Figures:

Figure 3.1 Schematic flow chart. 391 patients were included in the discovery and validation phases of the LivFib-50 peptide marker model for the detection of liver fibrosis.

Figure 3.2 Contour Plot showing the LivFib-50 peptide marker profiles for liver fibrosis and non-liver fibrosis cases.

Figure 3.3 ROC characteristics for the discovery set.

Figure 3.4 Box and Whisker distribution plots for classification of the different patient groups of the discovery set with the LivFib-50 biomarker.

Figure 3.5 Distribution of classification scores of the LivFib-50 marker pattern in normal liver and liver fibrosis groups of the first validation set of patients.

Figure 3.6 Classification of normal controls without clinical signs of liver disease (N=123) and those with clinical manifestations of liver cirrhosis (N=31) with the age-adjusted LivFib-50 peptide marker model in comparison to the proteomic model without age adjustment and age alone.

Figure 3.7 ROC characteristics of the age-adjusted LivFib-50 peptide marker model for patients with LC in the absence or presence of HCC (N=19) compared to non-fibrotic control patients (N=17) in independent validation.

Figure 3.8 Box and Whisker distribution plots for classification of the different patient groups of the validation set with the age-adjusted LivFib-50 classification model.

Chapter 4 Tables:

Table 4.1 Characteristics of HCC and non-HCC patients included in the discovery and validation phase.

Table 4.2 Characterization of the 31 urinary HCC peptide markers by amino acid sequencing and in silico protease prediction analysis together with their experimental spectrometry mass, retention time in capillary electrophoresis, and location in the protein's linear sequence.

Table 4.3 Differences in the activities of the seven in silico predicted proteases between the HCC and non-HCC liver disease patients.

Table 4.4 Allred Scoring for all retrieved tissue specimens.
Table 4.5  List of HCC-31 peptide markers that were identified in human tissue or body fluids other than urine.

Chapter 4 Figures:

Figure 4.1  Schematic flow chart for the discovery and validation phases of the HCC-31 peptide marker model for the detection of HCC.

Figure 4.2  ROC characteristics of the proteomic HCC classification model HCC-31 on the validation set of 39 HCC cases and 87 non-HCC liver diseased controls.

Figure 4.3  Overall survival of patients with a positive or negative HCC-31 test

Figure 4.4  IHC characteristics of kallikrein-6 (KLK6) in HCC, LC and normal liver

Figure 4.5  IHC characteristics of meprin A subunit a (MEP1A) in HCC, LC and normal liver

Chapter 5 Tables:

Table 5.1  Characteristics of recruited participants for SPME work

Table 5.2  Characteristics of recruited participants for the GC-IMS and GC-TOF-MS work

Table 5.3  Discrimination of cancers via urinary VOCs using an RBF neural network

Table 5.4  GC-IMS result for different liver disease groups

Table 5.5  GC-TOF-MS result for different liver disease groups.

Table 5.6  List of relevant chemicals identified by GC-TOF-MS for HCC vs non-HCC.

Table 5.7  List of relevant chemicals identified by GC-TOF-MS for fibrosis vs non fibrosis.

Chapter 5 Figures:

Figure 5.1  PCA plot showing SPME differentiation of urine volatile compounds captured from patients with HCC and controls

Figure 5.2  PCA plot showing SPME differentiation of urine volatile compounds captured from patients with prostate cancer and controls

Figure 5.3  PCA plot showing SPME differentiation of urine volatile compounds captured from patients with bladder cancer and controls
Figure 5.4 Urinary VOCs ROC curve for detection of HCC with normal AFP vs raised AFP levels.

Figure 5.5 Box plots to compare the relative response of the eight metal sensors in the SPME work (cancer groups and controls).

Figure 5.6 Example output of GC-IMS instrument when applied to urinalysis of VOCs.

Figure 5.7 ROC characteristics for GC-IMS Analysis (a) HCC vs Fibrosis, (b) HCC vs Non-Fibrosis and (c) Fibrosis vs Non-Fibrosis.

Figure 5.8 Example output of GC-TOF-MS instrument when applied VOCs urinalysis.

Figure 5.9 ROCs for (a) HCC and Fibrosis samples, (b) HCC and Non-Fibrosis and (c) Fibrosis and Non-Fibrosis using GC-TOF-MS.

Chapter 6 Tables:

Table 6.1 Characteristics of recruited participants for the mSEPT9 work at time of obtaining the samples.

Table 6.2 Diagnostic performance of AFP and mSEPT9 in identifying HCC.

Table 6.3 Combined AFP and mSEPT9 results in identifying or excluding HCC.

Table 6.4 Nonparametric comparison of AFP and mSEPT9 PMR values in patient with liver disease, LC without HCC, and LC with HCC.

Table 6.5 Logistic regression p-values for HCC, LC and mSEPT9 associations with other covariates.

Table 6.6 Time-to-Event Analysis for death from HCC and associations with other covariates.

Chapter 6 Figures:

Figure 6.1 The mSEPT9 amplification graphs obtained from rt-PCR output.

Figure 6.2 ROC characteristics for AFP and mSEPT9 in detection of HCC in all participants.

Figure 6.3 ROC characteristics for AFP and mSEPT9 PMR in detection of HCC with LC patients.

Figure 6.4 Simple box plot for mSEPT9 levels across patients with various stages of liver disease.

Figure 6.5 Simple box plot for AFP across patients with various stages of liver disease.
Figure 6.6  Kaplan-Meier curves showing HCC patients’ survival over time in relation to the mSEPT9 and AFP results.

Figure 6.7  Kaplan-Meier curves showing survival over time in all participants in relation to the AFP and mSEPT9 results.

Chapter 7 Tables

Table 7.1  Levels of evidence in biomarker studies
Acknowledgments

I would like to acknowledge many without whom this work would not have been possible. I am forever grateful to my primary supervisor Professor Ramesh Arasaradnam. Ramesh was there throughout my research journey providing me with constant unrestricted advice. Ramesh’s support, guidance and insightful academic acumen have made this work achievable.

I am also thankful to my second supervisor Professor Harpal Randeva for his advice. Dr Jochen Metzger for his scientific input and expertise on proteomics. The Gastroenterology and Hepatology team of Hannover Medical School (Germany) for their help in recruiting patients. Hypertension and Cardiovascular Epidemiology Unit, University of Leuven (Belgium) for their collaborative effort in providing further urine samples for validation of the urinary peptide biomarker in liver fibrosis. Dr Kishore Gopalakrishnan for supervising me in interpretation of the liver tissue slides. Sean James for teaching me immunohistochemistry. Professor Megan Hitchins for her help and expertise in methylation techniques and rt-PCR technology. Rocio Alvarez, Lisa Zhou and Dr Michael Petchey for analysing the plasma and serum samples. Dr Angela Noufaily for her guidance and supervision in biostatistics. Professor Krishna Persaud and Dr Frank Kvasnik for employing SPME on the urine samples. Professor James Covington for employing GC-IMS and GC-TOF-MS on the urine samples.

Jasmeet Bhambra for her help in securing NHS ethical approval and UHCW internal sponsorship. Sarah O’Toole, Emma Cooper and Yvonne Tatton for organising my research clinics. Parmjit Dahaley for her help with storing and transferring the urine and plasma samples. HPB-MDT coordinators, Performance and Information team, Research and Development department, and Arden Tissue Bank for their constant support and help in accessing data and recruiting patients.

Sami Abdelrahman, Hussam Ahmed, Ahmed Ibrahim, my late friend Rolland Iriarte and Masoud Shebani for their encouraging discussions about how to persevere as a PhD student and a clinician at the same time! My colleagues Subashini, Maria and Nick who have made my clinical job enjoyable and supported.

My parents Sharafeldin and Sayda and all my family in the Sudan for continuously inspiring and motivating me. Farah for being an amazing daughter, constantly cheering me up!

Crucially, I acknowledge the funding bodies who provided support for this work to be a reality, Medical and Life Sciences Research Fund, Midlands Gastroenterological Society, and the Aileen Lynn Bequest Fund of the Royal College of Physicians and Surgeons Glasgow. Most importantly, I acknowledge all the patients who provided me with access to their data and samples.
IV. Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. I, Ayman Bannaga, declare this thesis has been composed by myself and has not been submitted in any previous application for any degree. All the research has been undertaken in accordance with University of Warwick safety policy and guidelines on ethical practice. Most of this work has been successfully published and have also been presented nationally and internationally.

The presented work, including data generation and analysis, was carried out by the author except in the cases outlined below:

- Peptide urinary analysis using CEMS was completed by Mosaiques Diagnostics, Hannover, Germany.
- I participated in VOCs urinary analysis using SPME, this was further partly completed by Professor Krishna Persaud and Dr Frank Kvasnik in University of Manchester.
- VOCs urinary analysis using GC-IMS & GC-TOF-MS was completed by Professor James Covington research group in University of Warwick.
- Plasma methylated SEPTIN9 analysis was completed by Professor Megan Hitchins research group in Cedars Sinai, Los Angles, USA.
- I participated in immunohistochemistry of tissue samples, this was further partly completed by Sean James and interpretation was supervised by Dr Kishore Gopalakrishnan
- Dr Angela Noufaily biostatistician (University of Warwick) and Dr Jochen Metzger biomedical scientist (Hannover, Germany) were consulted in the biostatics of the CEMS and methylated SEPTIN9 data.
V. List of Publications


VI. Awards and educational bursaries


2. Medical and Life Sciences Research Fund Award (2018)


4. United European Gastroenterology Week (2020). National Scholar Award


6. Aileen Lynn Bequest Fund Award (2020). Royal College of Physicians and Surgeons Glasgow

7. European Association for the Study of Liver (EASL). Liver Cancer Summit Travel Bursary (2021)

VII. Presentations


2. Methylated SEPTIN9 (mSEPT9) as a biomarker for hepatocellular carcinoma. Digital Liver Cancer Summit 2021. 05/02/2021 (Poster)

3. Urinary volatile organic compounds characterization in hepatocellular carcinoma – pilot study. Digital Liver Cancer Summit 2021. 05/02/2021 (Poster)


8. Albumin-bilirubin grade (ALBI) suggests better prognostic model than Child Turcotte Pugh (CTP) score in predicting survival from Hepatocellular Carcinoma (HCC) 27th United European Gastroenterology Week. Barcelona, Spain (October 2019) (Poster)

9. Neutrophil to lymphocyte ratio (NLR) is better than alpha fetoprotein (AFP) as a predictor of survival in hepatocellular carcinoma (HCC). BASL Conference. Glasgow, UK (September 2019) (Poster)

10. Size of largest nodule in hepatocellular carcinoma (HCC) is better than neutrophil to platelets ratio (NPR) and alpha fetoprotein (AFP) as a predictor of overall survival. BASL Conference. Glasgow, UK (September 2019) (Poster)
VIII. Abstract

Liver fibrosis is a major risk factor for development of hepatocellular carcinoma. Both liver fibrosis and hepatocellular carcinoma are associated with molecular pathogenic mechanisms involving alterations in the hepatocellular proteome, metabolome, and genome. Both liver fibrosis and hepatocellular carcinoma lack suitable biological predictive biomarkers in clinical practice.

Therefore, to aid in identifying suitable biomarkers, three approaches were employed in patients with liver fibrosis and hepatocellular carcinoma. Firstly, proteomic analysis was applied to identify post-translational enzymatic protein modifications peripherally present in the urine. Secondly, metabolic profiling was applied to characterise small volatile organic compounds present in the urine. Thirdly, DNA methylation detection technology was applied, to identify methylated SEPTIN9 patterns among the circulating hepatocellular carcinoma DNA molecules within the cell-free DNA pool present peripherally in the plasma.

Urinary proteomic analysis identified novel specific peptides for liver fibrosis and hepatocellular carcinoma. Additionally, proteases potentially involved in liver fibrosis and hepatocellular carcinoma were predicted from the peptides sequence with further demonstration of these proteases by immunohistochemistry in human normal liver tissue, liver fibrosis and hepatocellular carcinoma. The identified urinary peptides showed good diagnostic and prognostic performance in liver fibrosis and hepatocellular carcinoma.

Urinary metabolic profiling technologies demonstrated that volatile organic compounds patterns can be used noninvasively to detect hepatocellular carcinoma and they also revealed chemical composition of novel volatile organic compounds related to liver fibrosis and hepatocellular carcinoma.

DNA methylation analysis showed that methylated SEPTIN9 has good sensitivity and specificity for hepatocellular carcinoma. It was also a prognostic indicator in patients with liver disease and hepatocellular carcinoma. The methylated SEPTIN9 was also associated with other surrogate biomarkers for liver function, liver fibrosis and inflammation. Additionally, methylated SEPTIN9 was noted to incrementally increase in various stages of liver disease.

The researched biomarkers in this work provided some insight into the pathogenic mechanisms of liver fibrosis and hepatocellular carcinoma. If further validated, the identified biomarkers in this work could offer cost-effective tools for screening, diagnosis, prognosis and/or surveillance, particularly in low resource settings where access to advanced imaging and invasive biopsy is not feasible.
Chapter 1. Introduction
Chapter 1. Introduction

This chapter summarizes the current literature and highlights the need for further work to investigate biomarkers related to liver fibrosis and hepatocellular carcinoma (HCC).

1.1 Overview

HCC represents around 90% of all primary liver cancer cases. HCC is the third leading cause of cancer related death in the world. There are around 850,000 cases identified every year. The main risk factor for developing HCC is chronic liver damage caused by inflammation and advanced fibrosis/liver cirrhosis (LC) (1,2).

The main aetiological factors for chronic liver damage are well known including alcohol abuse, non-alcoholic fatty liver disease (NAFLD), chronic hepatitis B virus (HBV) infection, chronic hepatitis C virus (HCV) infection, haemochromatosis, Wilson's disease, α1 antitrypsin deficiency, granulomatous diseases (e.g., sarcoidosis), drug induced liver diseases, glycogen storage disease, autoimmune liver disease and vascular liver disease. Ingestion of the fungal metabolite aflatoxin β1 has also been implicated in HCC (3,4).

Liver fibrosis therapies are targeted to treat the underlying causes. Meanwhile, the HCC therapeutic profile includes both curative and noncurative modalities. The modalities with curative intent include local radiofrequency ablation (RFA), surgical resection and liver transplantation. The noncurative modalities with palliative intent include transarterial chemoembolization (TACE), transarterial radioembolisation (TARE), chemotherapy and immunotherapy (3-5).

Both liver fibrosis and HCC do not have accurate noninvasive diagnostic biological biomarkers (3-5). There are no cost-effective population based screening programmes for liver fibrosis and HCC (3,4). The surveillance strategy for HCC development in patients with LC is ineffective as it relies on liver ultrasound (USS) for the detection of cancerous nodules. This strategy is dependent on the quality of training of the USS operator. The role of alpha fetoprotein (AFP) in surveillance for HCC in LC is also questionable due its poor sensitivity and is no longer recommended for routine use. The key role of AFP is as prognostic biomarker in HCCs that secrete it (6-9).

Once liver fibrosis is suspected clinicians usually apply surrogate clinico-biochemical markers based on liver function tests (LFTs), body mass index (BMI) and platelet count. These surrogate markers are poor at estimating mild to moderate stages of liver fibrosis. In addition, the present liver fibrosis serum markers in clinical practice lack accuracy. Elastography also has varied cut-offs for diagnosing liver fibrosis depending on the aetiology and has limitations for use in patients with a high BMI (10,11).
Currently diagnosing advanced stages of liver fibrosis or LC, relies on clinical suspicion and/or clinical features of decompensation of liver disease through various imaging techniques (12). For conclusive diagnosis of the various stages of liver fibrosis, invasive assessment by a liver biopsy is required which possesses its own complications risk. A liver biopsy is only representative of the sampled liver tissue and not necessarily comprehensive for all of the fibrotic stage in liver disease (13).

On the other hand, HCC diagnosis relies on the ability of advanced, high-resolution imaging techniques for the detection of liver lesion with early arterial enhancement followed by early washout. These scans are not easily accessible and can be less accurate in detecting lesions <1-2cm. The current modalities used are contrast-enhanced triphasic computed tomography (CT) and/or contrast-enhanced magnetic resonance imaging (MRI). If the scans are inconclusive, the diagnosis is then confirmed with a cytological or histopathological evaluation of the liver lesion via a tissue biopsy. Treatment and prognostication of patients with HCC consider the size and number of tumour nodules, their relation to the portal vein, and the degree of liver impairment (8,9).

It is important to note that liver fibrosis and HCC are connected, with the first progressing to the latter. Both can lead to significant morbidity and mortality and can be clinically silent till late stages. Given these factors there is a need for noninvasive methods to identify liver fibrosis and HCC. This is also called for by the research statements issued by international clinical bodies (8,9). A detailed literature review about liver fibrosis and HCC and their current diagnostic methods are explained in the remaining sections of this chapter.

1.2 Origin of the liver

In the embryonic life the liver appears at the 3rd gestational week as an outgrowth of the endodermal epithelium of the distal end of the foregut. The liver then continues to grow and joins the umbilical veins which then form the hepatic sinusoids (14). During embryonic life, the liver tissue undergoes differentiation into liver parenchyma including the lining of the biliary tract, hematopoietic cells, Kupffer cells, and connective tissue cells. This differentiation is under control by the surrounding tissues through growth factors (e.g., fibroblast growth factors and bone morphogenetic proteins). These growth factors have also been implicated in the pathophysiology of liver fibrosis (15,16).

The liver is an important site of shunting in embryonic life. A portion of the placental umbilical vein blood flows into left portal vein then into ductus venosus and then directly to the inferior vena cava. Thus, allowing oxygenated blood from the placenta to bypass the liver. After birth, the umbilical vein and ductus venosus are obliterated and form the ligamentum teres and ligamentum venosum (17).
1.3 Macro-anatomy of the liver

Following birth, the liver continues to grow forming a wedge-shaped organ. By adulthood, the liver is considered one of the largest organs in the human body, measuring around 1.5 kg. It occupies the right hypochondrium under the cover of the thoracic cage and extends across the epigastric region. The liver has a large right lobe and a small left lobe divided by the falciform ligament. The right lobe is further divided into a quadrate lobe and a caudate lobe by the presence of the gallbladder, the fissure for the ligamentum teres, the inferior vena cava, and the fissure for the ligamentum venosum (18-21).

The liver receives 30% oxygenated arterial blood supply and 70% portal venous blood supply. The arterial supply is from the hepatic artery, a branch of the coeliac artery, which divides into right and left terminal branches when it enters the porta hepatis. The portal venous supply is from the portal vein which originates from the superior mesenteric vein and splenic vein. The portal vein delivers venous blood rich in the products of digestion from the gastrointestinal tract. Both arterial and portal venous vessels then convey blood to the central vein of each hepatic lobule via the liver sinusoids. The central veins then drain into the right and left hepatic veins, and these leave the posterior surface of the liver and open directly into the inferior vena cava into the main systemic circulation (18-21).

The liver secretes bile via the biliary system into the intestine. When digestion is not taking place, the extra bile is stored in the gallbladder for a use at a later time. The liver forms a main part of the lymphoreticular system. The amount of lymph it represents is around one third of all of the body lymph. The lymph is secreted into vessels which leave the liver and enter several lymph nodes in the porta hepatis and then into the rest of the lymphatic system (18-21).

1.4 Liver segments

On the basis of blood supply and biliary drainage, the liver is divided into eight segments. Segment I is the caudate lobe of the liver. It is an autonomous segment receiving blood from right and left branches of the hepatic artery and portal vein, draining bile into right and left hepatic ducts and having independent venous drainage into the inferior vena cava. the caudate lobe (segment I) is also further subdivided into right and left parts and the caudate process. The left lateral aspect of the liver contains segment II posteriorly and segment III anteriorly, with the left hepatic vein separating them. Segment IV is found on the visceral surface as the quadrate lobe. Segments V and VI are the inferior segments of the right medial and right lateral aspect respectively. Segments VII and VIII are the superior segments of the right lateral and right medial aspect respectively. Further segmental subdivision partitions
segment IV into IVa (superior) and IVb (inferior) segments, the latter coinciding more accurately with the quadrate lobe (21,22).

1.5 Hepatic lobule

The liver is formed of many small functional units called hepatic lobules. These are collections of hepatocytes in a hexagonal shape with the centre being a central vein. Within the lobules, the hepatocytes are arranged in cords, and in between the cords is a vascular space with a thin fenestrated endothelium and a discontinuous membrane called a sinusoid. At the vertices of the hexagon is a triad of a bile duct branch, a portal vein branch, and a hepatic artery branch referred to as the portal triad. Blood flows from the portal vein branch and hepatic artery branch across the lobule and finally into the central vein which is a branch of the hepatic vein (18-22).

An alternative arrangement is called the portal acinus and is helpful in describing the functional zones of the liver. Zone 1 hepatocytes immediately surround the portal tracts and primarily are involved in oxidative energy metabolism. Zone 3 hepatocytes immediately surround the central veins and are the primary location for the biotransformation of drugs. Zone 2 hepatocytes are between Zone 1 and 3 and have mixed functionality (20).

1.6 Perisinusoidal space

The perisinusoidal space (or space of Disse) is located between a hepatocyte and a sinusoid. It contains the blood plasma. Microvilli of hepatocytes extend into this space to facilitate transfer of the proteins and other plasma components from the sinusoids to be absorbed by the hepatocytes. Hepatic stellate cells (HSCs) are also located in the space of Disse between the sinusoidal endothelial cells and hepatic epithelial cells, and account for 5%–8% of the cells in the liver. HSCs play a significant role in fibrogenesis (23,24).

1.7 Extracellular skeleton of the liver tissue

The hepatocytes are surrounded by extracellular matrix (ECM). Commonly well characterised liver ECM proteins are collagen and elastin. They serve as ultra-small units of the liver biomechanical structure. In the normal liver, the collagen and elastin are formed by the fibroblasts which are present in the interstitial space and then secreted into the ECM. Following this, enzymatic modification leads to maturation of these proteins to form stronger fibres. Most of the collagens are the fibrillar types I, III and V. Normal collagens are highly stable molecules, with half-lives as long as several years. However, the liver ECM undergoes dynamic remodelling, often in response to growth or injury of the tissue. The breakdown of collagen fibres is dependent on the proteolytic action of collagenases, which are part of a large family of matrix metalloproteinases (24,25).

Figure 1.1 and Figure 1.2 show schematic and histological microanatomy of the liver.
Figure 1.1 Schematic illustration for the ultrastructure of the normal human liver. This was acquired from Sheila Sherlock Textbook of Hepatology, Diseases of the Liver and Biliary System (20)
Figure 1.2 Haematoxylin and Eosin (H&E) stain of normal liver tissue section at low power magnification (X5). Central veins (CV) and hepatic lobule, building block of the liver tissue consisting of a portal triad (inset at high power field X40 within the right lower corner of the slide), hepatocytes arranged in linear cords between a capillary network, and a central vein. Zone 1 hepatocytes immediately surround the portal tracts. Zone 3 hepatocytes immediately surround the central veins. Zone 2 hepatocytes lie in between Zone 1 and 3. Courtesy of Dr Kishore Gopalakrishnan, Consultant Pathologist, UHCW
1.8 Liver functions

The liver filters the blood coming from the gastrointestinal tract via the portal vein. This blood reaches the sinusoids between plates of hepatic cells and eventually drains into the hepatic veins, which then enters the inferior vena cava to the systemic circulation. During the passage of blood through the hepatic tissue, it undergoes extensive chemical modification (26).

The liver produces the bile, which is a water based (97%) alkaline fluid. The rest of bile which is around 3% is formed of bile acids (Cholic acid and Chenodeoxycholic), bile salts (bile acids conjugated with the amino acids glycine and taurine), bile pigments (biliverdin and bilirubin), Cholesterol, inorganic salts, fatty acids, phosphatidylcholine, fat and ALP. Bile pigments are by-products of haem catabolism in the liver, they further undergo degradation in the intestine by the bacteria and secreted in the urine (Urobilinogen) and the stool (Stercobilinogen). The bile leading role is to digest and absorb fats. The liver produces around 500 mL of bile per day. It is important to note that the bile undergoes a circle of intestinal reabsorption and further re-secretion by the liver; the bile ends up secreted via the bile duct into the duodenum and then reabsorbed in the terminal ileum via superior mesenteric vein and into the portal vein then into the liver and then excreted again by the liver into the duodenum (enterohepatic circulation). It is important to note that the cholangiocytes also have a role in modifying the bile before it is being stored and further concentrated in the gallbladder (27-30).

Furthermore, the liver is a major site for lipid metabolism. Following degradation of dietary lipids in the gastrointestinal tract, they undergo further emulsification by the bile. The liver is then able to process one type of lipid into another and is also able to store excess protein and carbohydrate in the form of lipids for storage into the adipose tissue and other extrahepatic tissues. This process is under control by the insulin hormone in the fed state and is termed lipogenesis. The liver also participates in ketogenesis, degradation of fatty acids, providing adenosine triphosphate (ATP) via the tricarboxylic acid (TCA) cycle. This process is under control of Glucagon and catecholamines in the starvation state and is also partly regulated by the peroxisome proliferator activated receptor alpha (PPARα) (31-33).

The liver also plays an important role in protein synthesis. Using amino acids consumed in the diet and stimulated by insulin and the growth hormone, the liver is involved in production of many plasma proteins. It is responsible for albumin production and many coagulation factors in addition to α fetoprotein (AFP) which physiologically has a role in osmotic regulation, can carry proteins, hormones and amino acids. Pathologically AFP can be raised in HCC (34,35).

The liver is also involved in protein catabolism. Ammonia is a toxic by-product that is produced following the metabolism of amino acids/proteins. Up to 85% of ammonia in the body ends up in the liver via the portal circulation. The Ammonia then undergoes processing to urea via several coordinative steps in both the cytoplasm and the mitochondria of the hepatocytes.
Urea is water-soluble that then undergoes excretion by the kidneys in the urine. This is referred to as the urea cycle. Liver dysfunction in decompensated cirrhosis leads to an inability to convert ammonia into urea, which then builds up in the blood reaching the brain with subsequent hepatic encephalopathy (36). The liver has a leading role in maintaining the stability of blood glucose levels, removing excess glucose from the blood and returning it as needed (37). In the fed state, carbohydrates are stored as glycogen in the liver while in the fasting state, this glycogen can be broken down for systemic use (glycogenolysis) (38). Lactate, amino acids and glycerol can also be converted into glucose (gluconeogenesis) (39).

The liver is considered the centre of drug/toxin (xenobiotic) metabolism. It is usually the first organ that processes drugs, depending on the route of administration, prior to reaching the systemic circulation (first pass metabolism). Following metabolism of xenobiotics, the yield can either be active or inactive. It is also important to note that metabolic by-products of some drugs can have longer half-life than the original drug compound. Drug metabolism involves two phases (40). Phase I reactions include metabolic modification of the drug compound and phase II reactions include conjugation reactions. In this regard, liver disease can therefore result in overactivity or underactivity of different drugs. These reactions are completed partly by the hepatic endoplasmic reticulum and mitochondria via the cytochrome P450 (CYP450) enzyme superfamily (41).

Phase I reactions particularly involved in reduction of oxygen lead to production of hydrogen peroxide which is one of the reactive oxygen species (ROS). The ROS species are implicated in pathologic processes like cancer and inflammatory diseases. ROS can cause serious chemical damage to the DNA, proteins, and unsaturated lipids leading to cell death (42). This is important in the context of HCC pathogenesis through continuous exposure to environmental toxins. In this context it is important to note that, the ROS and CYP450 are responsible for production of various volatile organic compounds (VOCs) (43,44).

The liver is also involved in vitamin metabolism. Vitamins are important for many cellular functions as coenzymes in metabolism reactions. The liver stores up to 80% of vitamin A in the body. Functionally vitamin A is a component of Rhodopsin, a protein in the eye that facilitates vision in low light conditions. Vitamin A within the liver is stored in the HSCs cytoplasmic droplets. The liver is also responsible for hydroxylation of the Cholecalciferol changes into 25-hydroxyvitamin D. This is then converted in the kidney to calcitriol (1,25-dihydroxyvitamin D) and leads to subsequent increased calcium absorption from the intestine (45).

Vitamin K has significant role in modification of various blood clotting factors (II, VII, IX, X). This modification is required to facilitate interaction between clotting factors and the platelets. Vitamin K serves as a major coenzyme in the carboxylation of glutamic acid residues present
in these proteins (46). Carboxylation of glutamic acid residue occurs in the liver, and it represents site for inhibition by dicumarol, an anticoagulant occurring naturally in spoiled sweet clover (yellow blossom), and by warfarin, a synthetic analogue of vitamin K (47). Protein induced by vitamin K absence-II or PIVKA-II also known as des-γ-carboxy-prothrombin (DCP) was identified as a serum biomarker linked to HCC (48). It is an abnormal prothrombin molecule produced as a consequence of an acquired defect in the carboxylation of the prothrombin precursor in malignant cells (49).

Vitamin E is an antioxidant, preventing the nonenzymic oxidation of cell components (50). Vitamin E was found to have modest roles in improving liver biochemistry and histology of NASH related fibrosis (51).

**1.9 The liver and functional immunity**

The liver is continuously exposed to the portal blood coming from the gastrointestinal tract. Bearing in mind that portal blood flow contains various pathogens and intestinal by-products (antigens), this will require the liver to deal with these antigens. The liver is therefore considered, an integral preventative filter, it fights infections and detoxifies the portal blood before it reaches the rest of the body sparing further excessive immune activations. The immunity centre within the liver lies in the space of Disse. The endothelial cells here form a barrier between the sinusoidal contents and the hepatocytes. In the endoluminal space there are further important cells; Kupffer cells, dendritic cells, lymphocytes and the hepatic stellate cells. It is important to note that the full spectrum of immune cells resident within the liver is still unclear (52).

**1.9.1 Kupffer cells**

Kupffer cells, are the resident liver macrophages. They are the first defence against microorganisms. They are involved in phagocytosis and breakdown of the red blood cells. Kupffer cells are also involved in clearance of endotoxins, or lipopolysaccharides (LPS), from the portal circulation through binding to the expressed toll like receptors (TLRs). Once Kupffer cells are activated, they release Interleukins (ILs) and tumour necrosis factor alpha (TNFα) involved in the regulation of the hepatic immune response. The cascade of released cytokines can activate fibrogenic responses through production of transforming growth factor beta 1 (TGF-β1), matrix metalloproteinases (MMPs), platelet-derived growth factor (PDGF), and reactive oxygen species (ROS). These cytokines lead to activation of the hepatic stellate cells (HSCs) with subsequent transdifferentiation into myofibroblasts. Here, extracellular matrix components are produced, including actin, laminin and collagens (53,54).
1.9.2 Dendritic cells

Hepatic dendritic cells (HDC) are derived from the bone marrow myeloid progenitor cells. HDC has a stellate morphology which can efficiently present antigens on major histocompatibility complex (MHC) molecules to activate naive T lymphocytes response. This T lymphocytes response is adaptive, it takes days to weeks to form and is responsible for specific pathogens and retains memory capabilities. The adaptive immune response usually orchestrates the chronicity of inflammation and liver damage (55,56).

1.9.3 Resident liver lymphocytes

The resident liver lymphocytes are partly formed of innate hepatic lymphocytes. These lack antigen specific receptors (57). They participate in the development and repair of lymphoid tissue. Hepatic lymphocytes include the natural killer (NK) cells, these have a key role against hepatic viral infection and neoplastic cells transformation (58). Interferon gamma (IFNγ), tumour necrosis factor alpha (TNFα) and granzyme are effector mediators used by the NK cells to destroy infected hepatic cells and malignant cells in order to stop the spread of the infection and to stop hepatocellular neoplastic formation (58). Other lymphocytes present in the liver are CD8 liver resident memory T lymphocytes. These act by directly lysing target cells and confer protective functions through recruitment of circulating T lymphocytes and other immune cells by chemokine production. They have a role in resisting parasites infection (59). Resident liver lymphocytes also include the gamma delta lymphocytes which participate in protecting against bacterial infection and also have a role in dealing with lipid antigens and therefore have a role in pathogenesis of NAFLD (60).

1.9.4 Hepatic stellate cells

Hepatic stellate cells (HSCs) are fat-storing cells located in the space of Disse. In fetal life they originate from the septum transversum as it forms from cardiac mesenchyme during invagination of the hepatic bud. Following birth HSCs are usually derived from bone marrow cells as shown by many studies. Stellate cells in other organs were also described. HSCs comprise around 15% of the total number of resident cells in a normal liver. HSCs have dendritic cytoplasmic processes and are major storage site for vitamin A and metabolism. HSCs are important to normal healthy liver as they have a role in hepatic regeneration and immune regulation (61,62).

Activation of HSCs following chronic liver injury/inflammation is considered a key step in development of fibrosis and cirrhosis. Here, HSCs undergo transformation to contractile myofibroblasts that produce ECM collagens and laminin. HSCs activation usually occurs under signalling from Kupffer cells, endothelial cells, hepatocytes, lymphocytes and platelets in response to liver injury (61,62).
1.10 The liver and disease

There are many aetiological factors involved in development of liver disease including alcohol, drugs, viral infections (e.g., HBV and HCV), ischaemia (e.g., Budd-Chiari syndrome), metabolic (e.g., NAFLD), genetic conditions (e.g., Wilson’s disease and Haemochromatosis) and autoimmune liver diseases (e.g., PBC, PSC and AIH). Generally, depending on the duration, liver diseases can be classed as either acute or chronic.

Acute liver failure refers to a rare syndrome of raised liver transaminases with features of impaired liver function namely jaundice, coagulopathy and encephalopathy. Acute liver failure can be divided into hyper-acute (<7 days onset), acute (8-28 days onset) and sub-acute (29 days to 12 weeks onset) (63). Acute liver failure is a life threatening condition with poor prognosis. The King’s College criteria for acute liver failure are accepted as indicators of poor prognosis requiring the need for urgent referral to a liver centre for further treatment including multiorgan support and liver transplantation (64).

Chronic liver disease (CLD) is defined as progressive deterioration of liver function for more than six months. CLD is a major cause of death worldwide and is among the top 20 causes of disability-adjusted life years and years of life lost (65). Even though most causes of CLD are preventable, there is still low public awareness about CLD as a significant cause morbidity and mortality (66). The severity of CLD can be assessed using several systems, including Child-Turcotte-Pugh score (CTP) (67,68), the model for end stage liver disease (MELD) score (69) and the United Kingdom Model for End-Stage Liver Disease (UKELD) (70). Each of these predicts mortality with CLD. UKELD and MELD scores are also used for prioritising liver transplantation. Over the progressive course of CLD, the liver undergoes significant morphological changes ranging from early to late liver fibrosis, or LC.

1.11 Liver fibrosis

This is a complex morpho-pathological process where normal liver tissue is slowly transformed into fibrous tissue over years in response to chronic liver injury of various aetiologies. Liver fibrosis has numerous stages, starting with early, then moderate and finally advanced liver fibrosis with regenerative hepatic nodules or LC (71). Histologically LC can be categorised to either micronodular LC, macronodular or both mixed micro and macronodular type (72). LC may be compensated when it is silent clinically and discovered incidentally or decompensated when patients start to have clinical features (jaundice, ascites, variceal bleeding and hepatic encephalopathy). Decompensated LC is usually driven by certain stressful factors like bleeding, infection, surgery, drugs or malignant transformation (73).

A recent large population-based UK study including 4021 patients estimated that 2.7% had moderate to advanced liver fibrosis when participants were assessed by elastography. The
study identified leading risk factors for this to be related to NAFLD and alcohol abuse (74). This was similar to an Italian study that identified liver fibrosis to be around 3.1% in the general population (75).

No matter what the cause of liver disease is, liver-related mortality increases exponentially with the increase in fibrosis stage. Higher fibrosis stage is associated with higher all-cause mortality when compared to patients without fibrosis. In alcoholic liver disease, severe fibrosis without frank LC had a major impact on 10-year mortality (76). In NAFLD, the higher stage of fibrosis was the most important predictor of mortality (77).

1.12 Pathogenesis of liver fibrosis

Complete understanding of liver fibrogenesis is still lacking. The literature describes this to start first at the molecular level with activation of HSCs. Chronic liver injury secondary to various aetiologies is the initiator with further influence from molecular signalling by platelets, lymphocytes, endothelial cells, Kupffer cells and hepatocytes (Figure 1.3).

The HSCs activation usually occurs via growth factors like the platelet-derived growth factor (PDGF) and transforming growth factor-β (TGF-β), which are released in response to chronic hepatocellular necrosis by the platelets and Kupffer cells. In addition, chronic injury to HSCs also reduces the expression of peroxisome proliferator activated receptor gamma (PPARγ) which is important for cellular differentiation, leading the HSCs to lose their lipid and retinoid droplet components (78,79).

Activation of HSCs will eventually lead to their transformation into myofibroblasts. These myofibroblasts move to the site of the hepatocellular necrosis and secrete high density collagens (I and III) and other ECM components. Myofibroblasts also continue to release vascular endothelial growth factor (VEGF) which directly promotes HSCs proliferation further. Following this there is continuous process of regeneration and degeneration. The hepatocytes start to regenerate under influence of the hepatocyte growth factor (HGF), epidermal growth factor (EGF) and interleukin 1 (IL1) and this is further inhibited by TGF-β (80,81).

Structurally fibrous septa or bridges are usually located between central vein and portal tracts (centro-portal) or between two portal tracts (portal-portal) because fibrogenesis by the myofibroblasts usually follows hepatocellular necrosis patterns in the liver. Subsequent fibrous septa then prevent metabolite exchange between the blood and the hepatocytes leading to chronic hepatocellular failure with associated morbidity and mortality (82).

It is important to note that studies showed that activated HSCs can also have a role in directly influencing HCC development through continued secretion of PDGF, TGF-β, VEGF and HGF.
These growth factors were found to promote the HCC microenvironment and angiogenesis (83).

Studies performed in patients and experimental fibrosis models have shown that fibrosis can be reversed when the injury is removed in initial stages. In particular, myofibroblasts can undergo apoptosis or inactivation when the aetiological factor is cleared, leading to fibrosis regression (84-86). At present time there are no treatments available for liver fibrosis, therapeutic interventions are targeted to the underlying aetiological factor of the liver disease (87).
Figure 1.3 Features of quiescent and activated HSCs. Quiescent HSCs store retinoid droplets and proliferate slowly. Upon liver injury, increase in molecular signalling transforms HSCs into myofibroblasts secreting rich collagens and other ECM components.
Currently, noninvasive methods used to identify patients with liver fibrosis have limitations (88). These include surrogate biochemical and clinical scores and various imaging modalities. The gold standard tool for diagnosis and staging of liver fibrosis remains the liver biopsy. It is important to detect clinically silent liver fibrosis because it triggers close monitoring and surveillance for complications of end stage liver disease like HCC and gastroesophageal varices. Methods aimed to monitor liver fibrosis could also be used to assess the need for antiviral therapy in chronic viral hepatitis and could prompt the need for liver transplantation (89).

1.13 Logistic regression models for predicting liver fibrosis and cirrhosis

Many models were described in the literature to assess for liver fibrosis and LC. These were developed based on the observation that AST levels usually exceed ALT levels in advanced fibrosis or LC, due to decreased clearance of AST and decreased synthesis of ALT. The commonly used models are AST to ALT ratio, AST to platelet ratio (APRI), NAFLD fibrosis score (age, body mass index, impaired fasting glycaemia or diabetes, AST:ALT, platelet count and albumin) and FIB-4 index (Age, AST, Platelets and ALT). These models were developed from large cohorts and were validated by different research groups against liver fibrosis histologically. However, these models have limitations, they were developed in patients with HCV, HIV and NAFLD, therefore extending their use into to other patients with liver fibrosis secondary to different aetiologies may be misleading. Use of these models in patients <35 or >65 years old should be done cautiously, as they have been shown to be less reliable in these age groups. Liver fibrosis often progresses non-linearly, it is not sequential or straightforward and can affect various parts of the liver, so it is important to re-assess fibrosis scores in individual patients over time. These models could also have a high rate of intermediate scores and invasive assessments to assess for severe fibrosis/LC scores may still be required (90-101).

1.14 Enhanced liver fibrosis blood test

Enhanced liver fibrosis (ELF) is a blood test that measures 3 biomarkers, type III procollagen peptide (PIIINP), hyaluronic acid (HA), and tissue inhibitor of metalloproteinase-1 (TIMP1). These molecules participate in liver ECM metabolism (102). ELF is recommended by national institute for health and care excellence (NICE) to assess NAFLD patients for presence of advanced liver fibrosis in primary care and need for specialist care referral (103). For advanced fibrosis, higher cut-offs of ELF could have up to 80% sensitivity and 90% specificity. However, a meta-analysis in 2020 found that ELF test has different used cut-offs in the guidelines and by the manufacturer. In settings with low prevalence of NAFLD, ELF was found to have a low positive predictive value, suggesting that additional strategies may be needed to assess for fibrosis in such settings (104).
It is also notable that the formula and components of ELF test that they were primarily included in the research work in 2004 (102) have been altered sequentially. Initially, the combination of these biomarkers included age as a factor. Age was then removed, and the ELF test was simplified to a newer algorithm which was published in 2008 (105). Therefore, not all studies on ELF reflect the performance of the currently available test. The ELF test was also found to have false positive results in patients with other chronic inflammatory conditions, chronic renal disease, and/or other extrahepatic causes of abnormal fibrogenesis (106-108).

1.15 Imaging in liver fibrosis

Morphological features of LC can be identified by USS, CT and MRI. An inhomogeneous hepatic texture or surface, shrunken size, enlarged caudate lobe, nodular liver with other associated features of portal hypertension (splenomegaly and/or collateral vessels) could aid in identifying LC with high accuracy via these imaging modalities. USS is the most widely used test and can provide valuable information about gross hepatic architecture, atrophy and hypertrophy of hepatic lobes, and thrombosis in the portal system (109). However, these conventional imaging modalities cannot detect early to moderate liver fibrosis.

Over the last decade, elastography research into liver fibrosis identified the technology as feasible in assessment. Elastography is a noninvasive method that uses low frequency vibrations during an ultrasound or MRI to measure the stiffness or elasticity of an organ. This is based on the measurement of the return velocity of a vibration wave travelling from an organ to a probe, this is then converted into kilopascals to assess the stiffness (110-113).

USS vibration controlled transient elastography is directed to the skin overlying the surface anatomy of the liver. It can be limited by raised BMI with large subcutaneous tissue, narrow intercostal spaces overlying the liver, ascites, hepatic congestion, cholestasis and the fasting state of the patient. MRI based elastography could assess the whole liver but is not widely accessible and results could be affected by increased iron stores in the liver and may be contraindicated in claustrophobic patients. The cost of MR elastography is also another limitation for its wide use (114,115).

The specific calculations of kilopascals differ between MRI and USS based vibration controlled transient elastography, and these measures vary between different liver diseases and have also been found to have different cut-offs as demonstrated in Table 1.1 from studies that have researched this in the literature (116).
Table 1.1 Studies showing variability in elastography modalities in different liver diseases, modified from reference (116)

<table>
<thead>
<tr>
<th>Test</th>
<th>Study</th>
<th>kPa Cut-off*</th>
<th>Sens%</th>
<th>Spec%</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCV</td>
<td>Vibration controlled transient elastography</td>
<td>Castéra et al (117)</td>
<td>&gt;9.5</td>
<td>73</td>
</tr>
<tr>
<td>HBV</td>
<td>Vibration controlled transient elastography</td>
<td>Marcellin et al (118)</td>
<td>&lt;8.1</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;10.5</td>
<td>72</td>
</tr>
<tr>
<td>NAFLD</td>
<td>Vibration controlled transient elastography</td>
<td>Tapper et al (119)</td>
<td>&gt;9.9</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Magnetic resonance elastography</td>
<td>Loomba et al (120)</td>
<td>&gt;3.64</td>
<td>86</td>
</tr>
<tr>
<td>Primary sclerosing cholangitis</td>
<td>Vibration controlled transient elastography</td>
<td>Corpechot et al (121)</td>
<td>&gt;10.7</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Vibration controlled transient elastography</td>
<td>Corpechot et al (122)</td>
<td>&gt;9.6</td>
<td>93</td>
</tr>
<tr>
<td>All liver diseases</td>
<td>Magnetic resonance elastography</td>
<td>Singh et al (123)</td>
<td>&gt;4.11</td>
<td>85</td>
</tr>
</tbody>
</table>

kPa, kilopascals. *cut-off are for advanced fibrosis. Sens, sensitivity. Spec, specificity
1.16 Diagnostic pathways in liver fibrosis

The current clinical guidelines in Europe and USA advise on assessing for liver fibrosis in considerable risk patients by performing serum fibrosis markers (e.g., NAFLD fibrosis score, FIB-4, and/or ELF test) or vibration controlled transient elastography. If the results of any of these suggest advanced fibrosis, further specialist referral to liver services is warranted for further assessment and counselling. However, if serum markers are negative, the advice is to repeat serum fibrosis markers every 2-3 years for monitoring (11,124).

This advice could miss some patients with fibrosis due to variable ability of these markers in detecting fibrosis as mentioned earlier. Therefore, a combined approach of using serum fibrosis markers with elastography techniques to increase the yield for detection of liver fibrosis could be another sensible approach as shown on Figure 1.4.

Despite all of these diagnostic approaches a liver biopsy would be still required to give accurate assessment of the fibrotic stage in a patient.
Figure 1.4. A noninvasive approach suggested by Tapper et al for assessment of liver fibrosis. This approach suggested combination of noninvasive tools for liver fibrosis, seeking concordance between serologic and elastography tests. This approach has been evaluated most extensively in patients with HCV and NAFLD. (MRE, magnetic resonance elastography). Acquired from (116)
1.17 Biopsy of the liver

Liver biopsy is usually required for diagnosis and prognosis or to stratify need for treatment when safe noninvasive methods are not informative. It is also needed as gold standard comparator in liver disease research. A liver biopsy can be commonly obtained transcutaneously, transvenously via a transjugular approach or via open/laparoscopic surgical approach. A core step to perform liver biopsy is informed consent and patient counselling prior to the procedure explaining benefits, risks, alternatives diagnostic tools and follow-up arrangements (125).

A common complication of liver biopsy is bleeding, it can occur between 1-10% with major bleeding occurring in <2%. Risk factors for bleeding from a liver biopsy include older age, bleeding disorders, anti-platelets/coagulants and presence of other comorbidities. A transvenous/transjugular biopsy is usually performed in patients who have coagulopathies or where there is a risk of bleeding, as it is considered safer than the percutaneous approach in these scenarios. Other liver biopsy complications include organ perforation, sepsis and death. Mortality associated with a liver biopsy is <1 in 1000 (126,127).

In the UK most liver biopsies are obtained by radiologists directly under Imaging (USS, CT or x-ray), due to evidence that this is approach is associated with reduced complications and increase in adequate samples (126).

Limitations to note in liver biopsy is the patient discomfort from the procedure as it is invasive. Another limitation is the cost, particularly if there is associated complication and hospitalisation and this could reach up to £2500 (128). Sampling error is also a technical aspect, there is an agreement that the minimum biopsy sample for research purposes should be around 20mm obtained with a 16G needle (129).

Histopathological considerations are also important to note due to intraobserver and interobserver variability which may lead to underestimation of underlying liver pathology or fibrosis. Due to variability in histopathological reporting, many histopathological grading systems for liver fibrosis have been developed and (130) some of these are demonstrated in Table 1.2. Furthermore, there is a consensus on how to approach a formal reporting for liver biopsy of a focal lesion in the UK as suggested by the Royal College of Pathologists (131).
Table 1.2 Knodell, Batts-Ludwig, Scheuer and Ishak histopathological liver fibrosis grading systems (132-135).

<table>
<thead>
<tr>
<th>Fibrosis stage</th>
<th>Knodell (132)</th>
<th>Batts-Ludwig (133)</th>
<th>Scheuer (134)</th>
<th>Ishak (135)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No fibrosis</td>
<td>0-No fibrosis</td>
<td>0-Normal connective tissue</td>
<td>0-No fibrosis</td>
<td>0-No fibrosis of some portal areas with or without short fibrous septa</td>
</tr>
<tr>
<td></td>
<td>Fibrous portal expansion</td>
<td>1-Fibrous portal expansion</td>
<td>1-Enlarged, fibrotic portal tracts</td>
<td></td>
</tr>
<tr>
<td>Early fibrosis</td>
<td></td>
<td></td>
<td></td>
<td>2-Fibrous expansion of most portal areas with or without short fibrous septa</td>
</tr>
<tr>
<td>Periportal Fibrosis</td>
<td>Bridging fibrosis (Portal-portal or portal-central)</td>
<td>2-Periportal fibrosis ± portal–portal septa</td>
<td>2-Periportal fibrosis or portal–portal septa, intact architecture</td>
<td>3-Fibrous expansion of most portal areas with occasional portal to portal bridging</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4-Fibrous expansion of portal areas with marked portal-portal and portal-central bridging</td>
</tr>
<tr>
<td>Late Bridging fibrosis</td>
<td></td>
<td>3-Bridging fibrosis but no obvious cirrhosis</td>
<td>3-Fibrosis with architectural distortion</td>
<td>5-Marked bridging with occasional nodules (incomplete cirrhosis)</td>
</tr>
<tr>
<td>Cirrhosis</td>
<td>Cirrhosis</td>
<td>4-Regenerative nodules encircled by fibrous septa</td>
<td>4-Probably or definite cirrhosis</td>
<td>6-Cirrhosis, probably or definite</td>
</tr>
</tbody>
</table>
1.18 Hepatocellular carcinoma

Liver cancer is the second most common cause of cancer death in the world. Around 90% of primary liver cancer cases are caused by hepatocellular carcinoma (HCC). It is estimated that 72% of cases occur in Asia (>50% in China), 10% in Europe, 7.8% in Africa, 5.1% in North America, 4.6% in Latin America and 0.5% in Oceania. The HCC prevalence in Asia follows HBV prevalence which is the major cause of HCC in this part of the world (1-4,136,137). Neonatal vaccination against HBV reduced the number of children who could become chronically infected. Vaccination programmes in China and sub-Saharan Africa have already produced a decreased risk of HCC at young ages (138,139).

In Europe and USA, HCV is considered to be the leading cause of HCC followed by excessive alcohol consumption. Although, it is expected that NAFLD would be the leading HCC cause in the coming decades, given successful clinical implementation of new HCV direct antiviral treatment with projected decreased HCV-HCC related incidence (140-143).

In the UK HCC incidence rate increased by 160% over the last three decades. HCC incidence rate is projected to further rise by 38% (15 cases/100,000) in the UK by the year 2035. Increased incidence will carry a male predominance, with an increased mortality rate in the older population. (144). HCC Incidence rate locally in UHCW also increased during the last decade (unpublished data). (Figure 1.5, Figure 1.6 and Figure 1.7)

HCC is more frequent in males usually occurring between the 5th and 7th decade of life. Patients who develop HCC usually have no symptoms other than those related to the underlying chronic liver disease. However, in patients with sudden hepatic decompensation such as ascites, jaundice, hepatic encephalopathy or variceal bleeding often caused by portal vein thrombosis there is an increased likelihood of HCC. Other uncommon features in HCC patient presentation include the triad of pain in the right upper quadrant, hepatomegaly, and weight loss and/or very rarely a sudden abdominal pain and swelling associated with haemoperitoneum due to HCC tumour rupture. Occasionally patients develop paraneoplastic syndromes (pyrexia of unknown origin, hypoglycaemia, erythrocytosis, hypercalcaemia, severe watery diarrhoea, dermatomyositis and several types of skin lesions) (3,4).
Figure 1.5 HCC observed and projected age-standardised incidence rates, by sex, UK, 1979-2035 (144).
Figure 1.6 HCC average number of deaths per year and age-specific mortality rates per 100,000 population, UK, 2015-2017 (144).
Figure 1.7 Diagnosed patients with HCC between October 2013 to October 2018 in UHCW (n=148). Note the trend line of increasing cases year on year. Courtesy of Performance and Information team at UHCW.
1.19 Pathogenesis of hepatocellular carcinoma

The majority of HCC cases occur in the setting of chronic liver disease, with advanced fibrosis/LC being the major risk factor, independent of liver disease aetiology. Around 80% of patients with HCC have background fibrosis/LC, while 20% of HCCs can occur on a normal background liver tissue. It is estimated that 1 in 3 LC patients develop HCC during their lifetime with a 1 to 8% annual incidence reported in long-term follow-up studies. Incidence of HCC appears lower in alcohol and NASH related liver diseases in comparison to active chronic viral hepatitis (HBV and HCV). HCC also occurs in patients with haemochromatosis, a1-antitrypsin deficiency, Wilson’s disease, autoimmune hepatitis, primary biliary cholangitis, glycogen storage disease, porphyria disorders. Alagille syndrome and hepatic vascular disease (Budd Chiari syndrome) (3,4,8,9).

Aflatoxin β1, direct hepatocarcinogen produced by Aspergillus flavus and present in stored foods like peanuts, has been implicated in susceptibility to HCC via direct disruption of the P53 tumour-suppressor gene in both cirrhotic and noncirrhotic livers (145).

Industrial chemicals were also implicated in HCC development e.g., nitrosamines, azo dyes, aromatic amines, vinyl chloride, organic solvents, pesticides and arsenic compounds. Carbon tetrachloride (CCl4) can also lead to HCC and usually used for research purpose in experimental animal models (146,147). Excessive androgens have also been implicated in HCC development. This was described first in patients with Fanconi’s anaemia on anabolic androgenic steroids. Therefore, it is important to counsel patients who take these treatments (148).

Many patients with HCC usually have more than one aetiological risk factor. Diabetes is considered an independent risk factor for HCC and treatment with Metformin was shown to reduce HCC development (149,150). In addition to these interventions there are data that showed Atorvastatin and Fluvastatin are associated with lower risk for LC and HCC in HCV patients (151).

The link between advancing liver fibrosis and HCC is not fully understood. HCC pathogenesis is described as a multifactorial event involving genetic and epigenetic changes, ongoing fibro-inflammation, regeneration and involvement of signalling molecules (152,153).

Biologically there are three main stages in hepatocarcinogenesis, these are genetic instability, changes in cellular proliferation and death and finally invasion and metastasis. Genetic instability occurs during the cell cycle division. The parental cell transfers the full and exact component of genetic information to the daughter cell (Figure 1.8). Mistakes in this process lead to genetic instability and mutations. For cancer to develop mutations occur in genes whose role is to maintain integrity and repair of the deoxyribonucleic acid (DNA). Following
cell division, the cellular proliferation and death is under control by growth factors and cell-cycle "checkpoints" at which the cell checks the process of DNA replication is occurring normally and if not; the cycle is stopped, and the cell may pass into apoptosis if the problem is not corrected. Impairment in this process leads to carcinogenesis. If a hepatic cancer cell escapes the check points, it then starts to detach and acquire motility. This is achieved by acquiring the ability to digest structural matrix proteins, to travel through the ECM. This is usually associated by angiogenesis and lymphangiogenesis which allow enlargement and spread of the tumour. This is brought about by cytokines and growth factors, principally vascular endothelial growth factor (VEGF) (154).

**Figure 1.8 Cell division cycle**
Furthermore, the literature described many molecular mechanisms involved in HCC development. These mechanisms occur in presence of complex immune signalling from the resident liver lymphocytes, natural killer cells, dendritic cells and HSCs. Below are the main molecular pathways described. (Figure 1.9)

1.19.1 Telomere maintenance

Telomere is a region of repetitive nucleotide sequences TTAGGG at each end of a chromosome. It protects the end of the chromosome from fusion with neighbouring/surrounding chromosomes. Shortening of telomeres occurs in proliferating phase, triggering the cell to become senescent or apoptotic. In cancer cells the enzyme telomerase reverse transcriptase (hTERT) is expressed to maintain the telomere, adding tandem repeats of TTAGGG to the end of the chromosome. This reactivation is found in most (>95%) of cancers including HCC. To date there is no strong telomerase inhibitor for use in HCC (155,156).

1.19.2 Wnt and β catenin pathway

Wnt and β catenin proteins are responsible for regulating hepatocellular proliferation and adhesion. The β catenin acts as an intracellular signal transducer in the Wnt signalling pathway. β catenin is encoded by the CTNNB1 (Catenin β1) gene. Mutations of this gene leading to overexpression of this protein in the hepatocellular tissue was found to be strongly associated with HCC development and its poor prognosis. Overexpression of β-catenin was also found in many other cancers. Inhibition of this pathway as a target therapy is under active research to come up with therapeutic options for HCC (157-159).

1.19.3 P53 inactivation

Tumour protein p53 is responsible for DNA repair and activates apoptosis. In HCC the p53 was found to be inactivated by chronic liver inflammation, nitric oxide (NO) telomere shortening, aflatoxin B1 and HBV infection (160).

1.19.4. Chromatin remodeling

This is the process by which enzymes help condense the DNA inside the nucleus. The AT-rich interactive domain-containing protein 2 (ARID2) gene is partly responsible for chromatin remodeling. This gene was found to be deactivated in HCV associated HCC suggesting its vital role in hepatocarcinogenesis (161,162).

1.19.5 Activation of the reactive oxygen species (ROS) pathway

This is a metabolic pathway occurring in the intracellular level leading to production of oxygen and nitrogen containing chemicals (e.g., H₂O₂ and NO) which cause damage to the lipids, proteins and DNA leading to formation of hydrocarbon based compounds like aldehydes or
other volatile organic compounds (VOCs). This pathway is linked to many diseases. Altered blood flow, hypoxic hepatocytes and chronic liver inflammation induce this pathway with resultant formation of HCC. Lipopolysaccharides and microbiota accumulation from the gut via the portal circulation are also contributing factors through this pathway (163-166).

1.19.6 Mitogen activated protein kinase (MAPK) pathway

MAPK is partly responsible for cell proliferation, metabolism, motility and death. It is a signalling pathway between the cell surface and the nucleus. It starts when an extracellular stimulus binds a cell surface receptor leading to cascade of events that are transmitted through the cell to the DNA with resultant protein expression. MAPK pathway is mediated by phosphorylation of different molecules by protein kinases (167). In HCC it was found that HCV core proteins are able to directly activate MAPK pathway leading to mitotic changes and HCC development (168).

MAPK pathway is of importance in HCC because one of the subclasses of protein kinases, tyrosine kinase, is a target for HCC chemotherapy treatment. Sorafenib is a small molecule that was shown to reduce HCC proliferation and angiogenesis. It also increases the rate of apoptosis. It acts by inhibiting tyrosine kinase activity, supressing the production of vascular endothelial growth factor receptors (VEGFRs) 1, 2, and 3 and platelet-derived growth factor receptor β (PDGFR-β) (169,170). Sorafenib was found to increase median survival time by three months in advanced HCC. The MAPK pathway is also a target for Regorafenib which is a multi-protein kinase inhibitor that have shown HCC survival benefit for patients who have progressed on Sorafenib treatment (171,172).
Figure 1.9 Molecular pathway for HCC development in cirrhotic and non-cirrhotic liver.

Main signaling pathways altered in HCC

- Telomere maintenance
- P53
- Oxidative stress pathway
- Wnt/β-catenin pathway
- Map kinase pathway
1.20 Diagnosis of hepatocellular carcinoma

HCC early detection is rare. There are no accurate diagnostic biological biomarkers for HCC nor population-based screening. Clinically it is usually suspected in patients with who have chronic liver disease with features of hepatic decompensation or when picked up on surveillance of fibrotic patients for HCC nodules. In cases of clinical suspicion, HCC diagnosis relies on the ability of advanced, high-resolution imaging techniques for the detection of liver lesion early arterial enhancement followed by early washout (8,9). (Figure 1.10)

These scans are not easily accessible and can be less accurate in detecting lesions <1-2cm. The current modalities used are contrast-enhanced triphasic computed tomography (CT) and/or contrast-enhanced magnetic resonance imaging (MRI) (8,9). These have sensitivity and specificity range between 90 to 95%. The Liver Imaging Reporting and Data System (LI-RADS) is now an adopted classification system for liver lesions which is used in patients at risk of HCC. The LI-RADS category reflects the probability of HCC and is based on the typical CT and MR-findings in HCC (173).

Therefore, HCC diagnosis is considered radiological however if the scans are inconclusive, the diagnosis is then confirmed with a cytological or histopathological evaluation of the liver lesion from tissue biopsy. (Figure 1.11). If the biopsy is inconclusive repeat imaging with or without biopsy is repeated then every 3-6 months (8,9).

HCC staging is usually undertaken in a multidisciplinary approach involving various specialities contributing to the HCC patient care. Many systems have been developed for staging and prognosis (174) like the tumour, node and metastasis (TNM) staging system, the Hong Kong liver cancer staging system, the Japanese integrated system and the cancer of the liver Italian program (CLIP). The Barcelona clinic liver cancer (BCLC) staging system is the most validated and widely used to stratify HCC patients. BCLC staging system is implemented by the HCC management guidelines. BCLC is linked to liver function, overall functional status of the patient, treatment modality and overall prognosis (175). (Figure 1.12)
Figure 1.10 Hepatocellular carcinoma (arrow) in a cirrhotic liver demonstrating typical arterial enhancement (a) and portal phase washout (b) on CT (176).
**Figure 1.11** Haematoxylin and Eosin (H&E) stain of tissue section at low power magnification (X2). The whole slide demonstrates a well-differentiated HCC showing trabecular pattern (Inset in the left lower corner at high porefield magnification X40). The trabecular pattern of HCC comprises thickened trabeculae of large, polygonal, eosinophilic tumour cells, which are separated by a network of vascular channels lined by the endothelium. Courtesy of Dr Kishore Gopalakrishnan, Consultant Pathologist, UHCW.
Figure 1.12 HCC BCLC staging system (175). (BSC, best supportive care. PS, performance status)

HCC in cirrhotic liver

Prognostic stage

- Very early stage (0)
  - Single <2 cm
  - Preserved liver function
  - PS 0

- Early stage (A)
  - Solitary or
  - 2–3 nodules <3 cm
  - Preserved liver function
  - PS 0

- Intermediate stage (B)
  - Multinodular, unresectable
  - Preserved liver function
  - PS 0

- Advanced stage (C)
  - Portal invasion/extrahepatic spread
  - Preserved liver function
  - PS 1–2

- Terminal stage (D)
  - Not transferable HCC
  - End-stage liver function
  - PS 3–4

Treatment

- Ablation
- Resection
- Transplant
- Chemoembolization
- Systemic therapy
- BSC

Survival

- >5 years
- >2.5 years
- ≥10 months
- 3 months

Optimal surgical candidate

- Yes
- No

Transplant candidate

- Yes
- No
1.21 Surveillance of hepatocellular carcinoma

This refers to active six monthly monitoring of high risk patients with established LC for early HCC lesions in the liver. According to international guidelines other at risk patients for HCC are African and Asian HBV carriers without cirrhosis and those with a family history of HCC (8,9). This strategy has shown there are benefits (177). Surveillance leads to early detection and brings higher chances for receiving prompt treatment and subsequent increased survival (177). There is a debate in the HCC literature on the choice of suitable biomarkers for surveillance in LC (178). Surveillance is usually done by performing 6 monthly liver ultrasound scan (USS) on cirrhotic patients to detect early suspicious nodules (179). However, USS is an operator dependent test and could miss early diagnosis of HCC in non-experienced hands (180-182). Nonetheless, it was shown to be cost-effective (8,9). In addition to USS, a blood test called alpha fetoprotein (AFP) is still used by clinicians to help in detection of HCC.

AFP was discovered in HCC patients in 1964. In the human, AFP is produced by the liver. Its plasma concentration reaches the maximum in the first trimester. It declines after birth to adult levels, usually between 0 to 10 ng/ml. In cirrhotic patients with a suspicious liver mass on USS, an AFP level above 200 ng/ml can be considered clinically diagnostic. Serum AFP usually increases with tumour growth. Serial measurements showing a rising trend of AFP are indicative of HCC. However, this is not the case for all HCCs (176).

AFP has a low sensitivity for HCC detection, it has a sensitivity between 40–65%. Due to this poor sensitivity, it was advised not to be used routinely by the international guidelines in 2018 (8,9). This was also demonstrated in the work done to look at outcomes of HCC in UHCW, AFP had sensitivity of 50% in patients diagnosed between 2013 and 2018 (unpublished data). Elevated AFP level can also be associated with non-seminomatous germ cell tumours, hepatoblastoma (in infants), and hepatic metastases (183,184).

AFP was further investigated to improve its ability for HCC detection. Past research showed that AFP has three isoforms that display affinity to lens culinaris agglutinin (LCA). These three isoforms are AFP-L1, AFP-L2, and AFP-L3. When investigated they displayed poor sensitivity of 60% for early HCC detection (185,186).

1.22 Explored biomarkers in hepatocellular carcinoma

There are many biomarkers that were explored in HCC. Below is a summary of the most commonly evaluated markers:

Des gamma carboxyprothrombin (DCP), also known as protein induced by vitamin K absence (PIVKA II). It is produced by malignant liver cells following inhibition of the vitamin K carboxylase system. Numerous studies used different cut off levels for measurement in HCC. DCP role in early HCC detection is not clear because data are lacking from studies around
cost effectiveness for surveillance. DCP was also found to be raised in portal vein invasion and advanced HCC tumour stage hence its use in early detection has been questionable (187-190).

Glypican 3 (GPC3), is a proteoglycan that is attached to cellular membrane by a glycosyl-phosphatidylinositol (GPI) anchor. GPC3 was found to be down regulated in in breast, ovarian and lung cancer. It was also found to be upregulated in HCC and thought to be involved in activation of the Wnt pathway. Serum GPC3 was assessed in the diagnosis of early stage HCC, the observed sensitivity was around 55% (191-195).

Micro ribonucleic acids (miRNAs) are small, non-coding RNAs (20-25 nucleotides). The miRNAs participate in the development of cancers. They have a role in proliferation and apoptosis of cancers cells. The miRNAs can be detected in urine and plasma of HCC patients. Studies have identified varying panels of miRNAs in HCC. This is an of ongoing and active area in HCC research. Many miRNAs were found to have a role in both the Wnt/β-catenin and MAPK pathways (196-198).

Further literature search also identified many other biomarkers found in relation to HCC like Osteopontin (OPN), Golgi protein-73 (GP73), Squamous cell carcinoma antigen (SCCA), Annexin A2, Soluble urokinase plasminogen activator receptor, Midkine (MDK) and Thioredoxins (TRXs). All of these markers are not implemented in clinical practice due to low sensitivity and also because they lack validation by large longitudinal clinical studies (199).

1.23 Treatment of hepatocellular carcinoma

The treatment approach is decided on the stage of HCC. Curative options include surgical resection, ablation and liver transplantation. Non-curative options include transarterial chemotherapeutic embolisation (TACE), systemic chemotherapy and general palliative supportive care (8,9).

Ablation used in HCC can vary between radiofrequency, microwave, cryoablation or laser techniques. Ablation can be delivered using image guidance like USS scan or via open or laparoscopic surgical approach. The choice of the technique usually depends on the availability of ablative method in the local health service and also on the training of interventional radiologist/specialist surgeon delivering this treatment (200,201).

Radiofrequency ablation (RFA) is the most commonly used method. HCC ablation is usually undertaken in patients who have BCLC stage 0-A with ≤ 3 HCC nodules of ≤ 3 cm in size. Ablation therapy has also been used as a bridging therapy to facilitate further surgical resection/transplantation. Median survival for ablation is between 1 to 3 years (202,203).
Resection is the treatment of choice for patients who have non-cirrhotic HCC. Patients with compensated cirrhosis should be assessed carefully for risk of decompensation and complications of portal hypertension; these two remain major contraindications for surgery. Liver resection has demonstrated good overall 5 year survival post-surgery between 40-75%. Resection can be complicated by 70% rate of recurrence at 5 years (204,205).

The publication of the Milan criteria in 1996 introduced liver transplantation as an option for cirrhosis related HCC treatment for patients with small unresectable early stage disease. The original study showed that with either one tumour ≤5 cm, or patients with 2-3 tumours each with a diameter ≤3 cm have post-liver transplantation survival similar to that of patients without HCC. When these criteria are followed, 5-year survival rate is around 70-80% and tumour recurrence rate is in the range of 10%. (206,207)

Transarterial chemoembolisation (TACE) is a non-curative option for patients who have BCLC stage B HCC. TACE involves embolisation of the HCC tumours with delivering local chemotherapy. Current chemotherapeutic agents that are used include Doxorubicin or Cisplatin frequently embedded in Lipiodol as a vehicle for increasing tumour exposure to the drug, followed by embolisation of the blood vessel with embolic agents such as gelatin sponge particles, metallic coils, starch microspheres, or autologous blood clots. Median survival rate for TACE is between 20 to 40 months. TACE could be repeated depending on the tumour nodule number and size (208).

Trans arterial embolisation of HCC is also explored by using yttrium-90 (⁹⁰Y) radioactive beads. This procedure is called selective internal radiation therapy (SIRT). However, a recent study found that when SIRT was employed in patients with locally advanced or intermediate-stage hepatocellular carcinoma after unsuccessful TACE, overall survival did not significantly differ to patients receiving Sorafenib (209).

There were limited options for chemotherapy till the use of Sorafenib for HCC patients started in 2007. Sorafenib is a tyrosine kinase inhibitor acting on the MAPK pathway. HCC patients with advanced disease can benefit of median survival of 3 months when receiving this therapy in comparison to placebo. A new drug called Regorafenib, a multi kinase inhibitor, has shown some survival benefit on patients who progressed on Sorafenib and in addition to this Lenvatinib, which is another multikinase inhibitor, was found to be non-inferior to Sorafenib in advanced HCC (169-172).

Nivolumab is check point inhibitor and was approved for HCC treatment in 2017 by the United States Food and Drug Administration (FDA). It is a human IgG4 monoclonal antibody that blocks programmed cell death protein-1 (PD-1). The approval was based on the results of CHECKMATE-040 clinical trial (NCT 01658878) which showed that median overall survival was 16.4 months with Nivolumab compared to 14.7 months with sorafenib. In 2018 the FDA
granted Pembrolizumab, another PD-1 inhibitor, an accelerated approval for the treatment of patients with HCC who had received Sorafenib. This was approved as per the KEYNOTE-240 (NCT02702401) clinical trial which showed improvement in the overall survival and progression free survival in HCC patients. Additionally, Durvalumab and Tremelimumab are other checkpoint inhibitors currently being investigated by the HIMALAYA clinical trial (NCT03298451) as first-line treatment in patients with advanced HCC (210-213).

### 1.24 Prognosis of hepatocellular carcinoma

Symptomatic presentation carries a poor prognosis. Less than 10% of HCC patients survive 3 years. Estimating prognosis depends on tumour stage, liver function, and overall health. Several staging systems have been proposed and are in clinical use. The Barcelona Clinic Liver Cancer (BCLC) system links tumour stage and Child–Turcotte-Pugh (CTP) status with treatment strategy to optimize prognosis. These two systems are complicated and require imaging and clinical assessment to formulate the exact prognosis (8,9).

The literature recently described an additional two prognostic systems that have a role in HCC prognosis. These are the neutrophil to lymphocyte ratio (NLR) and albumin bilirubin grade (ALBI). Both only use numerical values from routine blood tests done on HCC patients. NLR considered as a generic biomarker for the inflammatory status and ALBI described in 2015 by Johnson et al as a simpler and more objective way to estimate liver function and mortality in HCC (214-217).

Evaluation of the performance of these prognostic outcomes in patients between 2013 and 2018 locally in UHCW showed comparable results to those in the literature (218). Neutrophil to lymphocyte ratio (NLR) had a significant (p=0.011) negative relationship to survival. Mean survival for Albumin-bilirubin (ALBI) grade 1 was 37.7 months, ALBI grade 2 was 13.4 months and for ALBI grade 3 was 4.5 months. ALBI was better than CTP in detecting HCC death. Their potential relies in their ease to apply on patients from routine blood tests, particularly in low resource settings when access to imaging is not available or limited. (Figure 1.13, Figure 1.14 and Figure 1.15)
Figure 1.13 Survival curves comparing HCC patients with high neutrophil to lymphocyte ratio (NLR) > 3 and low NLR < 3 in UHCW between December 2013 and December 2018 (218)
Figure 1.14 survival curves comparing HCC patients according to the different albumin to bilirubin (ALBI) grades (1, 2 and 3) in UHCW between December 2013 and December 2018. ALBI = (log10 bilirubin × 0.66) + (albumin × -0.085), where bilirubin is in μmol/L and albumin in g/L. (218)
Figure 1.15 Receiver operating characteristics (ROC) curves comparing CTP to ALBI in HCC patients in UHCW between December 2013 and December 2018. Area under the curve (AUC) represents death from HCC. (218)
For the rest of this chapter, the focus will continue into the basic scientific background of the experimental work completed in this thesis exploring for biomarkers of liver fibrosis and HCC.

A biomarker is a biological molecule found in blood, other body fluids (e.g., urine), or tissues. A biomarker could be a sign of a normal or abnormal process, or of a condition or disease (219). A biomarker can also allow classification of patients based on common features and facilitate risk stratification, early detection, diagnosis, and prediction of prognosis or treatment response (219). Therefore, there is considerable research activity toward the development of more and better biomarkers in diseases that do not have accurate and available biomarkers (e.g., HCC and liver fibrosis).

The discovery and implementation of a biomarker involve various phases. These phases extend from biomarker discovery (phase I) to evaluation of biomarker performance (phases II-III) and clinical benefits and harms (phase IV-V), as demonstrated in Table 1.3. (220,221) Furthermore, biomarker research methodology concerning implementation in clinical practice should have key considerations; the study design should clearly define the target population for which the experimented biomarker is designed for, the biomarker assay should not be affected by using archived versus new specimens, and finally the data generated from the study should inform about the biomarker performance without bias (220,221).

The work herein this thesis used designs related to phase I-II biomarker studies. Phase I study aim to identify biomarkers and determine how well they distinguish HCC and non-HCC controls, that is, the true positive rate (TPR) and false positive rate (FPR). These studies can include genes, proteins and organic molecule panels and may start with measurement of the biomarker at the tissue level, with or without correlation with plasma or urine. Phase II is a clinical assay development based on a specimen that can be obtained noninvasively (220,221). Outcomes at this phase are estimation of TPR and FPR or the receiver operating characteristics curve (ROC) for the biomarker to distinguish subjects with HCC from those with liver fibrosis/LC but without HCC.
<table>
<thead>
<tr>
<th>Phase</th>
<th>Outcomes</th>
<th>Study design</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Exploratory TPR and FPR</td>
<td>Preclinical case-control study</td>
<td>TPR, FPR, ROC curves to identify potential biomarkers</td>
</tr>
<tr>
<td>II</td>
<td>Clinical assay TPR and FPR</td>
<td>Case-control study</td>
<td>TPR, FPR, ROC curves -Comparison with established disease biomarkers -Combinations of markers -Impact of covariates on biomarker performance</td>
</tr>
<tr>
<td>III</td>
<td>Detection of preclinical disease</td>
<td>PRoBE (prospective specimen collection, retrospective blinded evaluation) design</td>
<td>TPR, FPR, ROC curves -Comparison with established disease biomarkers -Time between outcome and biomarker measurement -Combination of markers</td>
</tr>
<tr>
<td>IV</td>
<td>Disease detection rates</td>
<td>Prospective cohort</td>
<td>-PPV -Disease stage and distribution -TPR, FPR, ROC curves -Benefits (early disease detection) and harms (false positive rate) -Comparison with established disease biomarkers</td>
</tr>
<tr>
<td>V</td>
<td>Decrease in disease mortality</td>
<td>Randomized study</td>
<td>Survival analysis and Cox regression -Cost and quality of life -Overdiagnosis and overtreatment</td>
</tr>
</tbody>
</table>

FPR, false positive rate; PPV, positive predictive value; ROC, receiver operating characteristics curve; TPR, true positive rate.
1.25 Proteomic analysis

Proteins are essential molecules in any living organism. They are formed from amino acids which are joined together by peptide bonds. This bond is formed from C1 (Carbon) part of an amino acid and N2 (Nitrogen) part of another amino acid (C-N terminals). The proteins are originally translated from the genetic material in the cell. Any alteration in the nucleic acid sequence may result in an incorrect amino acid being inserted into the polypeptide chain, potentially causing disease or even death of the organism. Proteins functionally have many roles in the human body. Proteins participate in catalysis, protection, regulation, signal transduction, storage and transport. They can also be enzymes and hormones. Proteins have various structures (primary, secondary, tertiary and quaternary) which can change according to the function or disease process. Proteins structure usually change in response to chemical hydrophobic and electrostatic interactions. The scientific discipline concerned with studying structure and function of proteins, is referred to as proteomics (222-224).

Proteomic analysis methods have gained attention over the last few decades as suitable tools for biomarker discovery with diagnostic or prognostic value. Proteomic analysis facilitate measurement of protein levels in accessible biofluids (e.g., plasma, urine or cerebrospinal fluid) and tissues. Proteomic analysis could also identify mechanistic insights into disease biology to prioritise targets for therapeutic development (225-227).

In liver fibrosis and HCC, there are many protein alterations as previously mentioned, this makes proteomic analysis an attractive tool to study these changes. The work herein this thesis targeted the urine as it is easy to collect and acceptable by many individuals as a way to investigate for disease when compared to invasive methods like tissue biopsy and blood tests. Also, urine has a potential to be easily scalable to mass population if there is an accurate test for liver fibrosis and HCC for screening purposes for example. This is an advantage in comparison to imaging modalities which are not easily accessed in low resource settings (228).

There are many methods used for proteomic analysis. Conventional methods include chromatography, enzyme-linked immunosorbet assay (ELISA), western blot and capillary electrophoresis (CE) (229-231), while advanced proteomic analytical methods include mass spectrometry (MS) detection technologies. The MS technology certainly advanced molecular research specially when further connected to other conventional methods like liquid chromatography or CE (233,234).

The work herein this thesis applied capillary electrophoresis mass spectrometry (CEMS) as the method of choice in urinary proteomic analysis due to its high resolution capacity for proteomic analysis in various complex biofluids (235-237).
CEMS in medical research facilitates proteomic characterization, biomarker identification and diagnostic evaluation. Several proteomics studies involving CEMS demonstrated good diagnostic potential of urinary peptide biomarkers. These peptide biomarkers have been identified in the context of a single type of cancer, chronic diseases and severe infections (239-243).

Furthermore, the origin of peptides in proteomic studies could be investigated by resolving their amino acid sequence and by searching for proteases involved in their generation through the in silico mapping software tools like Proteasix. This tool enables the linking of peptide fragments to active proteases and is therefore the bridge between the phenotype depicted in the low molecular weight proteome (consisting of naturally occurring peptides as a result of proteolysis) and the protease activity as a result of molecular pathophysiological mechanisms that are altered in diseases. This online open-source tool uses an input peptide list and allows for automatic cleavage site reconstruction and protease associations based on C-N terminal mapping (244,245).

1.26 Volatile organic compounds

Volatile organic compounds (VOCs) are small molecules that have the ability to transform from solid and liquid phases to gaseous compounds in room temperature. In the human body, VOCs, represent different metabolites. For example, in the liver, the reactive oxygen species (ROS) and cytochrome P450 (CYP450) pathways are responsible for generation of various VOCs. After the tissues produce VOCs they permeate through the cell membranes to the systemic circulation reaching different biological fluids like urine, blood, stool, sweat and breath. In diseased cells and tissues the metabolites are different and therefore investigations of the pattern and composition of these VOCs was shown to be promising for studying pathophysiology of disease and also in reaching various biomarker profiles that could be used for diagnosis and prognosis (44,163-166,246,247).

The attractiveness of VOCs analysis for disease biomarker discovery relies on the ability of the metabolic profiling techniques in measuring VOCs noninvasively. Currently, urine and breath are commonly used for this approach. In particular the urine was found to have good potential as a tool for VOCs measurement. A recent publication by Ronald Becker (249) showed the pros and cons of breath versus urine sampling. The author of this thesis opinion is that urinalysis is better than breath analysis for VOCs detection noninvasively. At present breath sampling has limitations, particularly due to the Coronavirus (COVID-19) pandemic. Adding to this, breath sampling is also limited by the presence of both endogenous and exogenous VOCs in exhaled air. In contrast, urine as a medium for VOCs analysis has strong characteristics that favour its application in future diagnostic longitudinal studies. Urinary VOCs were already able to differentiate between controls and many diseases. Urine samples
have also been shown to remain stable for VOC analysis at room temperature for up to 12 hours. In addition, leaving urine samples stored at 4°C for up to 24 hours demonstrated no significant difference in the chemical composition of VOCs. The optimum temperature for storing urine for VOCs analysis was found to be -80°C, with storage time of up to one year. Two studies showed two to nine cycles of -80°C freezing and thawing revealed stable VOC results. Another characteristic that distinguish urine as a biological sample in VOCs analysis is the ability to undergo freeze-drying as an alternative to freezing. This was shown to have potential as an alternative method for VOCs analysis. Furthermore, in comparison to breath sampling urine collection from patients may be more acceptable (248-256).

Various instruments have been explored successfully in the literature that could assess for VOCs in different biological media. A summary of methods for VOC detection in biological fluids was discussed by Arasaradnam et al (257) and is shown in Table 1.4. The University Hospital Coventry and Warwickshire (UHCW), Faculty of Engineering, University of Warwick and Department of Analytical Chemistry, University of Manchester research groups have extensive experience in analysis of the bio-signature of VOCs in many gastrointestinal diseases. The group conducted a pilot study that provides proof of concept for further research project into the role of VOCs in NAFLD. The study showed urinary VOCs could distinguish between various stages of NAFLD including non-alcoholic steatohepatitis (NASH) and NASH cirrhosis (NASH-C) and controls (258). The group also looked into the pattern of VOCs in colon and pancreatic cancers (259,260).

VOCs were shown to have a significant role in liver disease and HCC. VOCs can be both released and metabolised by Hepatocytes. Understanding the VOCs and the metabolome is important in understanding the pathogenesis of liver fibrosis and HCC. One in vitro study on cell lines of HCC (261) has shown that total of nine compounds were found to be metabolised by HCC cells, these comprised of 6 aldehydes, n-propyl propionate, n-butyl acetate, and isoprene. Same cell lines also were found to release 12 VOCs, there were five ketones, five sulphur compounds, n-propyl acetate, and 2-heptene. Another in vitro study in HCC cells identified methane-sulfonyl chloride and di-hydro-benzofuran to have significant statistical difference when compared to normal cells. In addition to these one in vivo study involving breath analysis of 30 HCC patients identified that 3-hydroxy-2-butanone, styrene, and decane to have the best diagnostic values (262). Few other studies demonstrated that VOCs could have potential diagnostic performance in various liver diseases (263-268). Commonly described HCC VOCs are summarised in Table 1.5 as per current literature.
Table 1.4 Gas analysis instruments describing its advantages and limitations including practical applicability for VOCs analysis. Table adapted from (245). GC-MS, gas chromatography mass spectrometry; MS, mass spectrometry; SIFT-MS, selected ion flow tube mass spectrometry; FAIMS, field asymmetric ion mobility spectroscopy. Depends upon type of IMS technology deployed. *Drift tubes are medium, FAIMS high. † Pre-concentration required.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Breadth of analysis</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>Speed</th>
<th>User skill level</th>
<th>Consumable cost per item</th>
<th>Maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas analyzer (e.g., nitric)</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Real-time</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Electronic Nose</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Real-time</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Ion mobility spectrometer</td>
<td>Medium</td>
<td>Medium/High*</td>
<td>Medium/High</td>
<td>High</td>
<td>Real-time</td>
<td>Medium/High</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>GC-MS</td>
<td>High</td>
<td>Very high†</td>
<td>Very High</td>
<td>High</td>
<td>Off-line</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>SIFT-MS</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Real-time</td>
<td>Medium/High</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>FAIMS</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Real-time</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
</tr>
</tbody>
</table>

*Electronic Nose subgroups*

| Metal oxide                | Medium              | High        | Medium      | Medium   | Real-time   | Low               | Low                     | Low         |
| Optical                    | Medium              | Medium      | Medium      | High     | Real-time   | Low               | Low                     | Low         |
| Polymer                    | Medium              | Medium      | Low         | Low      | Real-time   | Low               | Low                     | Low         |
| GC-Based                   | High/Medium         | Medium/High | High        | Medium   | Real-time   | Medium            | Medium                  | Medium      |
| Electrochemical            | Low                 | High        | High        | High     | Real-time   | Low               | Low                     | Low         |
| Quartz Crystal             | Medium              | Medium      | Medium      | High     | Real-time   | Medium            | Low                     | Low         |
Table 1.5 Commonly occurring VOCs in HCC as per current literature, table adapted from (263-268)

<table>
<thead>
<tr>
<th>2-methyl 2-propanal</th>
<th>2-nonanone</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-methyl propanal</td>
<td>dimethyl sulfide</td>
</tr>
<tr>
<td>2-ethylacrolein</td>
<td>ethyl methyl sulphide</td>
</tr>
<tr>
<td>3-methyl butanal</td>
<td>3-methyl thiophene</td>
</tr>
<tr>
<td>n-hexanal</td>
<td>2-methyl-1-(methylthio)- propane</td>
</tr>
<tr>
<td>Benzaldehyde</td>
<td>2-methyl-5-(methylthio) furan</td>
</tr>
<tr>
<td>n-propyl propionate</td>
<td>n-propyl acetate</td>
</tr>
<tr>
<td>n-butyl acetate</td>
<td>2-heptene</td>
</tr>
<tr>
<td>Isoprene</td>
<td>methane-sulfonyl chloride</td>
</tr>
<tr>
<td>2-pentanone</td>
<td>di-hydro-benzofuran</td>
</tr>
<tr>
<td>3-heptanone</td>
<td>3-hydroxy-2-butanoine</td>
</tr>
<tr>
<td>2-heptanone</td>
<td>Styrene</td>
</tr>
<tr>
<td>3-octanone</td>
<td>Decane</td>
</tr>
</tbody>
</table>

1.27 Liquid biopsy and DNA methylation

Short fragments of cell free (cf) DNA are released into the plasma when cells die. This was particularly noted in patients with cancer, here, DNA fragments originating from the dead tumour cells circulate as part of the cfDNA. Circulating (ct) tumour DNA was then researched using advancing technologies (real-time PCR) to detect genomic abnormalities and mutations, this is referred to with the term liquid biopsy (269). This approach was also noted to be particularly useful in pregnant women where DNA derived from the foetus can be detected peripherally, and this has helped in process of prenatal screening (270-272). Liquid biopsy was further investigated in other diseases associated with cell death. Analysis of ctDNA can provide useful clinical information for detection or monitoring of cancer (273,274). With further improvement in analytical technologies, new genomic biomarkers markers will lead to the application of liquid biopsy in routine diagnostics in the future.

To put this in perspective not only the sequence of the ctDNA can be detected but also epigenomic changes associated with it can also be detected. Nowadays, with advanced technologies methylation changes can be detected peripherally too in ctDNA. The DNA methylation is an epigenetic mechanism, involving the addition of a methyl group (-CH3) to cytosines preceding a guanine (referred to as CpG sites) in genomic DNA without changing its sequence. The typical consequence of aberrant DNA methylation at CpG-rich promoter regions of genes is repression of transcription, and thereby, loss of protein expression. All cell types of the body have the same DNA sequence. However, the methylation patterns are different. For example, many genes expressed specifically in the liver, are unmethylated in hepatocytes and methylated in all other tissues, and genes that are preferentially read in the heart are unmethylated exclusively in cardiomyocytes. DNA methylation is an important
process to maintain genomic stability. DNA methylation has been reported in organ development and function and it was also studied in development of disease (275,278).

DNA methylation has been described in patients with viral hepatitis, cirrhosis and HCC (277-281). There are many markers that have been identified in HCC, however advancement of methylation sequencing techniques will require standardization between researchers. Further examples of methylation biomarkers include the p16 and RASSF1A genes which are hypermethylated in multiple cancer types (282,283). Another example is the hypermethylation of the SEPTIN9 gene that has been used for the screening of colorectal carcinoma (284-290).

1.28 Methylated SEPTIN9

The first methylated SEPTIN9 DNA (mSEPT9) test was developed by Lofton-Day et al in 2008 as a research kit (284). The second-generation test that is commercially available, Epi proColon® V2.0 CE (Epigenomics AG, Berlin, Germany). This test was approved in 2016 by the United States Food and Drug Administration (FDA) as screening test for the detection of colorectal cancers in people who are at average-risk (≥50 years, who refuse colonoscopy and other high sensitivity faecal occult blood tests) (291,292).

There is evidence originating from experimental studies completed at the molecular level about the SEPTIN9 gene in pathogenesis of both liver fibrosis and HCC. The Septin 9 expression was found to be abundant in healthy tissue but was frequently decreased by aberrant promoter hypermethylation in HCC. In particular, in the liver Septin 9 protein was found to have a vital role in the growth of lipid droplets. These droplets have an important intracellular role in interacting with other cell organelles and their accumulation is an important pathogenic step in the development of non-alcoholic steatohepatitis (NASH) and viral-related cirrhosis. In addition, methylation of SEPTIN9 resulted in loss of apoptotic cellular function and was directly involved in the activation of hepatic stellate cells with subsequent liver fibrogenesis. These are two important steps in inducing hepatocarcinogenesis (293-294).

Given that the mSEPT9 assay is available commercially and showing promise, the work herein this thesis aimed to evaluate this biomarker further in liver fibrosis and HCC.

1.29 Research justification

Current methods used to detect liver fibrosis and HCC have limitations. Liver fibrosis and HCC both have poor survival due to late diagnosis and late therapeutic intervention. Therefore, new biomarkers that could aid in screening, diagnosis and/or prognosis are needed. New biomarkers should be easy to obtain to facilitate cost-effective approach for future implementation in clinical practice. Biologically, urine collection and blood (plasma) sampling are common and easy to obtain. Therefore, they are acceptable to patients in comparison to invasive liver biopsy. In addition, expensive advanced imaging scans are not scalable to mass
population. Scans are only requested on case by case basis when there is a clinical suspicion or clear risk factor. Scans are also limited in poor resource settings and the developing world. This thesis aimed to address this unmet need by exploring the urine and plasma of patients with liver fibrosis and HCC. Work here looked into urinary peptides, urinary volatile organic compounds (VOCs) and plasma methylated SEPTIN9 (mSEPT9).

1.30 Research questions

1. Identify urinary peptides and proteases in patients with liver fibrosis and HCC

2. Identify the pattern and chemical composition of urinary VOCs in patient with liver fibrosis and HCC

3. Evaluate the diagnostic and prognostic performance of plasma mSEPT9 in patient with liver fibrosis and HCC
Chapter 2. Methodology
Chapter 2. Methodology

This chapter summarises the research design, ethical considerations and methodological applications used in relation to the exploratory experimental work completed herein this thesis to answer the research questions in section 1.30.

2.1. Research design

This research aimed to explore new novel biomarkers for detection of both liver fibrosis and HCC. The design of the research project was therefore descriptive in nature. It involved cohort of patients with HCC, healthy controls and patients with various chronic liver diseases including NAFLD, NASH and NASH cirrhosis. The choice of chronic liver disease group was made to include patients who are at risk of developing varying degrees of liver fibrosis and HCC (296).

Definitive sample size calculations were difficult as the technologies used were novel with little available literature for comparison. This was discussed at the Warwick Medical School statistics clinic at the start of the research project. The advice was that a small sample size was justifiable, given the exploratory nature of project. In addition to this, there were constraints of low case load of HCC and NAFLD meeting inclusion criteria in UHCW. The main aim of the research is to identify new biomarkers, with emerging results naturally needing further validation in bigger cohort of patients over a longitudinal prospective course. Ultimately, a target sample of 150 participants was made.

For further review of the research design, the UHCW research and development department sought informal peer review from Hepatologists in Queen Elizabeth Hospital Birmingham and Leicester Royal Infirmary. These reviews helped in refining the inclusion criteria for the project.

2.1 Inclusion criteria:

For the HCC, patients had to be aged ≥ 18 years. Participant had to be willing and able to provide written informed consent. HCC case had to be diagnosed radiologically or histologically by the hepato-pancreatico-biliary multidisciplinary team meeting (HPB-MDT) prior to inclusion to research study. These are in line with the international criteria for HCC diagnosis (8,9).

For the liver disease group, patients had to be aged ≥ 18 years. Participant had to be willing and able to provide written informed consent. Given clinical diagnosis of NAFLD and/or other chronic liver disease by a gastroenterologist. In addition, NAFLD spectrum participants had to have histologically proven disease (NAFLD, NASH or NASH-cirrhosis).

For the healthy control group, participants had to be aged ≥ 18 years. Participant had to be willing and able to provide written informed consent. To fulfil this further, participants had to
have normal liver functions tests, normal liver on ultrasound scan or other abdominal imaging and without clinical features of liver disease.

2.3 Exclusion criteria

The following were excluded from the research study; patients aged < 18 years, participants who are unable to provide written informed consent, pregnant women and any patient who already received anti-HCC treatment.

2.4 Ethical considerations

For this experimental work, a small research study, the fatty liver and liver cancer study (TENDENCY) was developed. Health Research Authority (HRA) and North-East Yorkshire NHS research committee ethical approvals were obtained (Ref 19/NE/0213). On 02/07/2021 TENDENCY study was successfully adopted on the national institute of health research (NIHR) portfolio to receive clinical research network (CRN) support nationally for further recruitment to validate the results of this work in the near future (CPMS ID 42438, pending CRN ID). This is the future direction which the author of this thesis is aiming for as the next step in this research.

The work in this thesis also had access to archival samples fulfilling the recruitment criteria from the FAMISHED study (Food and Fermentation using Metagenomics in Health and Disease). This study was approved by the Coventry and Warwickshire NHS research ethics committee (Ref 09/H1211/38).

Part of experimental work in this thesis also lead to forming a collaborative link with Hannover Medical School to obtain further samples for the experimental work. In Germany, the study was approved by the Ethics Committee of the Hannover Medical School (Ref 901). To access stored tissue samples for the purpose of immunohistochemistry an ethical approval from the Arden Tissue Bank, UK was obtained (Ref ATB19-013). For work on outcomes of HCC patients in UHCW, Internal approval from UHCW research and development department was obtained (Ref GF0336).

Using the UHCW informatics systems and notifications of the hepato-pancreatico-biliary multidisciplinary team (HPB-MDT) meetings the potential participants were identified. Patient information sheet (PIS) was posted to eligible research participants, this was followed by conducting a phone call. The participant was given an option of either attending a dedicated research clinic or meeting after a routine clinic appointment in UHCW. Before taking consent, the patient was given adequate time to read the PIS and was given the opportunity to ask any questions. The right of a participant to refuse participation without giving any reason was respected. All experimental research in this thesis conformed to the World Medical Association Declaration of Helsinki, with all participants included provided written informed consent.
2.5 Recruitment

Participants were recruited from UHCW (UK), Arden Tissue Bank (UK) and Hannover Medical School (Germany). During the course of recruitment, a further collaborative link with the Hypertension and Cardiovascular Epidemiology Unit, University of Leuven (Belgium) was formed. The unit provided urine samples to aid in discovery and validation of the urinary peptide markers in liver fibrosis (chapter 3). The biomarker analysis of samples took place over two years, between 01/06/2019 to 01/06/2021. There were total of 474 participants involved in this research project. Further characterisation of the participants is demonstrated in each result chapter (chapter 3 to chapter 6). The samples used herein this thesis is summarised in Table 2.1.

Table 2.1. Samples used in experimental work in this research project. UHCW and Hannover medical school had participants without liver disease as well as participants with varying degrees of liver disease including liver fibrosis and HCC. University of Leuven participants were selected from epidemiological study populations ensuring absence of diabetes, heart disease, hypertension, hyperlipidaemia and obesity.

<table>
<thead>
<tr>
<th>Centre</th>
<th>Type of samples</th>
<th>Experiment</th>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>UHCW</td>
<td>141 Urine samples</td>
<td>CEMS</td>
<td>Chapters 3 &amp; 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SPME</td>
<td>Chapter 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GC-IMS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GC-TOF-MS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>141 plasma samples</td>
<td>rt-PCR</td>
<td>Chapter 6</td>
</tr>
<tr>
<td>Arden tissue bank</td>
<td>14 experiment tissue sections</td>
<td>IHC</td>
<td>Chapter 4</td>
</tr>
<tr>
<td></td>
<td>40 test tissue sections</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hannover Medical School</td>
<td>77 urine samples</td>
<td>CEMS</td>
<td>Chapter 3 &amp; 4</td>
</tr>
<tr>
<td>University of Leuven</td>
<td>244 urine samples</td>
<td>CEMS</td>
<td>Chapter 3</td>
</tr>
</tbody>
</table>

2.6 Research procedures

Urine and blood samples were collected in one visit. After collection, urine specimens were frozen at -80 °C (within 2 hours) and stored at the Arden Tissue Bank for later analysis. Participants then underwent venepuncture for two blood samples (5 ml each). Samples were collected in a serum separating tube (SST) which contain spray-coated silica and ethylenediaminetetraacetic acid (EDTA) tube. Blood samples were then immediately spun in a centrifuge for 15 minutes at 5000 rpm. Serum, plasma and whole blood were then aliquoted into 1 ml tubes and labelled S-serum, P-plasma, WB-whole blood. Samples were then frozen
immediately at -80°C till further batch analysis at end of recruitment (297-299). All samples were anonymised, only identified by unique study number. Figure 2.1 shows the hypotheses behind the experimental work allowing for exploring urine and plasma for liver disease biomarkers.

All participants had basic clinical data collected including age, sex, medical history, drug history, body mass index (BMI), results of various imaging scans, blood test results, liver histology and endoscopy. All of these data were done as part of clinical care and were collected at time of sampling. In NAFLD patients, vibration controlled transient elastography (VCTE) was used to ensure presence or absence of liver fibrosis. VCTE is a non-invasive technique that uses both ultrasound and low-frequency waves to quantify liver fibrosis. Analysis of aspartate aminotransferase (AST) was further analysed in all recruited participants as this is not done routinely in UHCW, this facilitated calculating surrogate liver fibrosis scores. Additionally, AFP was also analysed separately as it was used as a comparator in the research work and was completed in participants without HCC.
Figure 2.1 Research procedures and hypotheses for the experimented work
To answer research questions in section 1.30 experimental research methods described below were chosen dependent on the accuracy of these methods for detecting urinary peptides, urinary VOCs and plasma mSEPT9 from previous similar studies done in other patients for detection of cancer and other diseases.

2.7 Capillary electrophoresis coupled with mass spectrometry (CEMS)

Capillary electrophoresis mass spectrometry (CEMS) has emerged in recent years as a hybrid advanced technology using capillary electrophoresis (CE) instead of liquid chromatography for sensitive (up to 1 fmol) and high-resolution low molecular weight protein and peptide separation before mass spectrometry (MS). CEMS aims to achieve molecular separation through the charge and then the size of the molecule (Figure 2.2). Notably, this method enables profiling of the proteomic content of body fluids, such as urine, plasma or bile, in a mass range of 0.8 to 20 kilodalton (kDa). So far, it is one of the most applicable methods for monitoring of systemic catabolic processes and post translation enzymatic peptide changes caused by differences in the proteolytical environment at tissue and organ sites. In this work, CEMS was used to answer question 1 in section 1.30.

2.7.1 Urine sample preparation for CEMS

This work was completed commercially by Mosaiques Diagnostics.

After participants had their urine collected in standard universal specimen containers (Newport, UK) and frozen to -80°C. For proteomic analysis, samples had to be prepared. In brief, a 0.7 mL aliquot was thawed immediately before use and diluted with 0.7 mL 2 M urea, 10 mM NH₄OH containing 0.02% Sodium Dodecyl Sulfate. To remove proteins of higher molecular mass (e.g., albumin and immunoglobulin G) the sample was filtered using a Centrisart ultracentrifugation filter device (20 kDa molecular weight cut-off; Sartorius, Goettingen, Germany) at 3,000 rcf until 1.1 ml filtrate was obtained. Subsequently, the filtrate was loaded onto a PD-10 desalting column (GE Healthcare, München, Germany), and equilibrated in 0.01% NH₄OH in HPLC-grade H₂O (Roth, Karlsruhe, Germany) in order to decrease matrix effects by removing urea, electrolytes, and salts, and also to enrich polypeptides. Finally, all samples were lyophilized, stored at 4°C, and resuspended in HPLC-grade H₂O shortly before CEMS analysis (300).

2.7.2 CEMS urine analysis

CEMS analysis was performed using a P/ACE MDQ capillary electrophoresis system (Beckman Coulter, Fullerton, USA) on-line coupled to a Micro Time-of-Flight MS (Bruker Daltonic, Bremen, Germany). The ESI sprayer (Agilent Technologies, Palo Alto, USA) was grounded, and the ion spray interface potential was set between -4.0 and -4.5 kV. Data acquisition and MS acquisition methods were automatically controlled by the CE via contact-
close-relays. Spectra were accumulated every 3 s over a range of m/z 350 to 3000. Details on accuracy, precision, selectivity, sensitivity, stability, and reproducibility of the CEMS method have been established (300).

2.7.3 CEMS data processing

Mass spectral ion peaks, representing identical molecules at different charge states, were deconvoluted into single masses using MosaiquesVisu software. For noise filtering, signals with z>1 observed in a minimum of 3 consecutive spectra with a signal-to-noise ratio of at least 4 were considered. MosaiquesVisu employs a probabilistic clustering algorithm and uses both isotopic distribution (for z ≤ 6) and conjugated masses for charge-state determination of peptides/proteins (300).

The resulting peak list characterizes each polypeptide by its mass and its migration time. Time-of-flight-MS data were calibrated utilizing 150 reference mass data points and 452 reference migration time data points by applying global and local linear regression, respectively. Ion signal intensity (amplitude) showed variability, mostly due to different amounts of salt and peptides in the sample and were normalized. Reference signals of 29 highly abundant peptides were used as “internal standard” peptides for calibration using local linear regression. This procedure was shown to be an easy and reliable method to address both analytical and dilution variances in a single calibration step. The obtained peak list characterizes each polypeptide by its calibrated molecular mass [Da], calibrated CE migration time [min] and normalized signal intensity (300).

All detected peptides were deposited, matched, and annotated in a Microsoft SQL database allowing further statistical analysis. For clustering, peptides in different samples were considered identical, if mass deviation was <50 parts per million for small (<4,000 Da) or 75 parts per million for larger peptides. Due to analyte diffusion effects, CE peak widths increase with CE migration time. For data clustering, this effect was considered by linearly increasing cluster widths over the entire electropherogram (19 min to 45 min) from 2 to 5% (298).

2.7.4 Support vector machine model generation and classification

For the integration of a set of peptides to a support vector machine (SVM) classification model the MosaCluster v.1.7.5 software was applied (Biomosaiques Software GmbH, Hannover, Germany). MosaCluster constructs a high dimensional parameter space based on the amplitudes of the selected peptides and defines a separation hyper-plane between two groups defined as case or control during the supervised learning phase. After establishment, such a SVM peptide marker model can be used for diagnosis by assigning to each patient’s CEMS profile a membership value according to the level of similarity to either the case or control group used for training. To compensate for imbalanced data, MosaCluster includes a class-
weighting function based on the ratio of the two classes which is used for assigning higher misclassification penalties to the larger group.

2.7.5 Peptide sequencing

Peptide sequencing was carried out both on a Dionex Ultimate 3000 RSLS nanoflow system (Dionex, Camberley, UK) and a Beckman CE/Orbitrap Q Exactive plus combination (Thermo Scientific, Waltham, MA). Spectra files were analyzed with Proteome Discoverer 2.4 (Thermo Scientific) allowing a precursor mass tolerance of 5 ppm and a fragment mass tolerance of 0.05 Da. This was followed by a search using the SEQUEST search engine against the UniProt human non-redundant database (https://www.uniprot.org/) without any protease specificity or fixed modification. Oxidation of methionine and proline were considered as variable modifications. Only sequences with high confidence (Xcorr ≥1.9) and without unmodified cysteine were accepted (due to the application of non-reducing conditions). A strong correlation between peptide charge at the CE operating pH of 2 deduced from the number of basic amino acids in the annotated peptide sequence and the migration time was use as another criterion to prevent false sequence assignments (233,302).

2.7.6 In-silico protease assessment

In silico protease assessment was performed using Proteasix (www.proteasix.org), the web-based tool for investigation of proteolytic events involved in naturally occurring peptide generation (303). Observed specific proteases responsible for cleavage of N- or C- terminus of a peptide were retrieved from CutDB proteolytic event database available at www.cutdb.burnham.org (304).

Protease activity was assessed in the patient’s CEMS peptide profiles to gain fold-changes between HCC cases, liver fibrosis, cirrhotic and non-cirrhotic liver diseases based on the average of associated peptide intensities. This was carried to similar work completed by Voigtländer et al (305).

2.7.7 Tissue transcriptomic data analysis

This was completed by Dr Agnieszka Latosinska of Mosaiques diagnostics.

For assessing liver fibrosis peptides further tissue transcriptomics data were retrieved from the NCBI Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) and from the EMBL-EBI Expression Atlas (https://www.ebi.ac.uk/gxa/home) (306,307). This includes expression profiling data by array from 1) 22 patients with cirrhosis (LC) and 14 histologically normal livers (GEO accession: GSE77627), 2) four biological replicates of histologically normal kidney and three biological replicates of histologically normal liver of human origin (E-MTAB-2836) (308), as well as 3) four biological replicates of histologically
normal kidney and four biological replicates of histologically normal liver of human origin (E-MTAB-5782) (309).

The GEO data sets were analysed with GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/). Analysis was conducted using default settings, with false discovery rate adjustment of p-values based on the method by Benjamini & Hochberg (310).

2.7.8 Statistical analysis for CEMS

This was completed by Dr Jochen Metzger of Mosaiques diagnostics.

Clinical and demographic data, as well as peptide distributions in the samples of the patients included in the different patient subgroups were presented either as mean ± standard deviation or as percentage. For intergroup comparison, continuous data were compared by Student's t-test or Mann Whitney rank-sum test, and categorical data by Chi square or Fisher's exact test after testing for normal distribution. A two-sided p value < 0.05 was considered to be statistically significant (311).

For urinary peptide markers identification in the discovery phase, the distribution differences of the urinary peptides between the patients with liver disease and normal controls were based on natural logarithm transformed intensities and a non-parametric Wilcoxon rank sum test. For correction of p values due to multiple testing the method by Benjamini and Hochberg was used (310).

Receiver-operating-characteristic (ROC) curves were generated for the classification of the patient samples with the urinary peptide markers for liver fibrosis and HCC. The ROC curve was obtained by plotting all sensitivity values (true positive fraction) on the y axis against their equivalent (1-specificity) values (false positive fraction) on the x axis for all available thresholds. The area under the ROC curve (AUC) was evaluated as it provides a single measure of overall accuracy independent of any threshold. Calculation of 95% confidence intervals (CI) was based on exact binomial calculations and the optimal balance of sensitivity and specificity was determined on the basis of the Youden index (312). ROC analysis and the determination of AUC values thereof were used as these are accepted descriptors to determine diagnostic test accuracy. A major characteristic of ROC analysis is that it describes the classifiers performance over the entire range of criterion values and therefore provides the advantage to be independent of any particular threshold (312).

Overall survival was analyzed by Kaplan-Meier methodology and a log-rank test to compare patients with a positive test result versus those with a negative test result by the HCC urinary peptide marker test (313). All statistical analyses were carried out using MedCalc version 12.7.5.0 (MedCalc Software; Mariakerke, Belgium).
Figure 2.2 CEMS steps for analyzing a urine sample.

CEMS

Capillary Electrophoresis (CE)

Ionization

Mass Spectrometry (MS)

Data Storage And Evaluation

CE
- Fast
- Robust
- Reproducible

MS
- Resolution
- Scan speed

Urine Sample

Diagnostic

Disease specific Biomarker pattern

Report
2.9 Immunohistochemistry

This was work was completed by author of the thesis and Mr. Sean James.

To further evaluate the presence of predicted proteases from the discovered CEMS peptides at the liver tissue site, immunohistochemistry (IHC) was performed. In this work, IHC was completed for the two most significant predicted proteases (KLK6 and MEP1A) from the urinary HCC-31 peptide model. (Figure 2.3). This is further demonstrated in chapter 4.

The Arden tissue bank was interrogated for liver tissue sections/biopsies. 14 cases (5 with HCC, 4 with benign liver disease including cirrhosis but no HCC and 5 cases with normal liver tissue without disease).

For both KLK6 and MEP1A detection commercial polyclonal goat IgG antibodies known to react to human tissue (R&D systems, Abingdon, UK) was used. The Human protein atlas was interrogated (https://www.proteinatlas.org/) (314), to investigate the current guidance on antibody dilutions and anticipated staining patterns for both KLK6 and MEP1A. However, no suitable strengths were found on human protein atlas data at the time of the experiment. Therefore, the stains were optimized first on test cases (for positive controls, used liver and small bowel tissue for MEP1A and tonsil and liver for KLK6 and negative controls were also used) (314). Following this, KLK6 specific primary antibody at a dilution of 1:200 and MEP1A specific primary antibody at a dilution of 1:1400 were chosen for all IHC experiments. Each tissue section/biopsy was reviewed by a Gastrointestinal Pathologist to ensure adequate tissue was present prior to staining.

The tissue sections/biopsies were formalin-fixed and paraffin-embedded using a standardized protocol in UHCW (unpublished). The tissue sections/biopsies were cut at 3-4 μm and dried overnight in a 56°C oven. A Leica polymer detection kit (Leica, UK) was used throughout. Pre-labelled slides were washed with distilled water. A hydrophobic pen was used to mark borders around the tissue sections, they were then placed in a metal moist chamber. Endogenous peroxidase activity was blocked using 100 µl/2 drops of Leica peroxidase block and left for 5 minutes. Two washes were carried out using Tris-buffered saline (TBS) for five minutes each. Sections were then incubated in Leica protein block 100 μl for 20 mins, this reduces non-specific binding of primary and polymer. The protein block was drained off and a further 2 TBS washes were completed at five minutes each.

This was followed by incubation in KLK6 at 1:200 and MEP1A at 1:1400 dilution as determined by prior serial dilutions. Incubation with antibody was performed overnight in the fridge at 4°C to help reduce the level of background staining.

After overnight incubation, sections were washed in TBS for 10 minutes, twice. Leica post primary block was then applied to slides using 100µl and incubated for 30 minutes, this
solution recognizes the goat immunoglobulins, this step is followed by two washes with TBS for 5 minutes each. Sections were then incubated in Leica Novolink polymer solution using 100μl to ensure adequate coverage and incubated for 30 minutes followed by two washes with TBS for 5 minutes each time. Sections were incubated in Leica diaminobenzidine (DAB) working solution. This was constituted using 50μl of DAB chromogen in one ml of Novolink DAB substrate.

Each slide was covered with two-three drops and left for five minutes reaction with the peroxidase produces a visible brown precipitate at the antigen site. This step was followed by a further two washes with TBS for five minutes each time. Slides were then rinsed in distilled water and drained off. A counterstain of haematoxylin (brown colour) was applied for 30 seconds and slides were again washed in distilled water, followed by incubation in TBS for 2-3 minutes. Once left to dehydrate they were then ready for viewing. Results were interpreted using a light microscope and Panoramic viewer (3DHISTECH Ltd, Hungary), where digitally converted slides were displayed.

For IHC assessment of the detected protease the Allred Scoring system for stain intensity was used; 0 for Negative (no staining of any nuclei at high magnification), 1 for weak (only visible at high magnification), 2 for Moderate (readily visible at low magnification) and 3 for strong (strikingly positive at low magnification) (315).
Figure 2.3 All steps followed in identifying HCC-31 urinary peptide markers. (CEMS and Immunohistochemistry)

- Hepatitis B and C, Alcohol, Fat, Autoimmune disease, and Inherited conditions
- Extracellular matrix
- Low molecular weight peptide fragments in the systemic circulation
- Kidney filtration
- Databases and proteases prediction
- Testing various proteases in normal liver, cirrhosis and HCC
- Capillary electrophoresis Mass spectrometry
- Urine samples
- Statistical analysis and discovery of HCC-specific diagnostic peptide marker model
- Histopathological interpretation

*Hepatocellular carcinoma (HCC), Horseradish peroxidase (HRP), Immunohistochemistry (IHC)*
2.10 Techniques used in volatile organic compounds analysis

Volatile organic compounds (VOCs) have the ability to emit as gases from solid or liquid substances. They have characteristic smell patterns (316). To put this in perspective, the smell has been a valuable tool used by clinicians in detection of disease. For example, diabetes mellitus, is derived from the Greek meaning of sweet urination. Clinicians used to diagnose it from the musty sweet urine smell and acetone breath (317). Furthermore, through the ages, clinicians identified patients with severe liver disease through the fishy smell, or foetor hepaticus. In 1949, Davidson demonstrated the presence of mercaptans in the breath of patients with severe liver disease and postulated that these VOCs were the origins of the odour known as ‘foetor hepaticus’ (318). In 1971, Nobel laureate Linus Pauling took this further by describing the complex mixture of volatile compounds (~250 compounds) present in breath and urine in response to change in diet (319). There are also reports that canines can be trained to detect initial stages of disease, through smell sensing of VOCs (320).

The VOCs arise from the human tissue and then excreted in the human biological waste including breath, urine, blood, faeces and sweat. Techniques used to identify VOCs have gained momentum over the last few decades, particularly because they are noninvasive. There are several available analytical methods that can detect VOCs.

Experimented urinary work in this thesis, started by applying an electronic nose that utilizes solid phase microextraction (SPME) technique, SensAm®. This device is broad, referring to a method rather than specific sensor technology. It doesn't detect every single chemical component but, like the human olfactory system, it attempts to identify patterns in an array of 8 nonspecific sensors.

Further experimentation on the urine in this thesis moved on to use advanced technologies namely gas chromatography coupled with ion mass spectrometry (GC-IMS) and gas chromatography coupled with time-of-flight mass spectrometry (GC-TOF-MS). These methods are described in sections 2.11, 2.12 and 2.13. During the research time The author of the thesis with help of the Warwick university group has reached to an observational comparison between each of these methods as demonstrated in Table 2.2.
Table 2.2 Observational comparison for the methods used herein this thesis for urinary VOCs pattern recognition and identification.

<table>
<thead>
<tr>
<th>Observations</th>
<th>SPME</th>
<th>GC-IMS</th>
<th>GC-TOF-MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytical features</td>
<td>This depends on the sensors array chosen in the device. Molecular separation usually give group of volatile compound groups rather than an exact identified chemical.</td>
<td>This depends on the proton Affinity. Molecular separation here is based on the specific drift times, that ionized compounds need to pass a fixed distance (drift tube) in an electric field</td>
<td>This depends on the choice of the separation tube, or column in the gas chromatography device. It is important to note if there is no preconcentration prior to use, small volatile compounds, particularly sulphur containing compounds are difficult to identify.</td>
</tr>
<tr>
<td>Analytical time per sample</td>
<td>5 minutes per sample&lt;br&gt;10 minutes between samples</td>
<td>&lt;10 minutes&lt;br&gt;No flush cycle in between samples</td>
<td>Up to 60 minutes per sample</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>Moderate</td>
<td>High</td>
<td>Moderate to high</td>
</tr>
<tr>
<td>Specificity</td>
<td>Moderate</td>
<td>Moderate to high</td>
<td>High</td>
</tr>
<tr>
<td>Operative skill level</td>
<td>Low</td>
<td>Moderate</td>
<td>High</td>
</tr>
<tr>
<td>Maintenance cost</td>
<td>Low cost</td>
<td>Medium cost</td>
<td>High</td>
</tr>
<tr>
<td>Device cost</td>
<td>£10,000-15,000</td>
<td>£40,000-50,000</td>
<td>£120,000-150,000</td>
</tr>
<tr>
<td>Future potential</td>
<td>Point of care device&lt;br&gt;Rapid decision making e.g., infections</td>
<td>Lab based</td>
<td>Potential for pathology lab Mainly targeting specificity</td>
</tr>
<tr>
<td>Data analysis</td>
<td>Quick straightforward process</td>
<td>Time consuming, least developed device it usually takes more effort compared to GC-TOF-MS</td>
<td>Established analytical algorithms</td>
</tr>
</tbody>
</table>
2.11 Solid-phase microextraction (SPME)

This is a technique where special absorptive or adsorptive ‘sorptive’ materials are used to extract analytes of interest by capturing them in their matrices. The sorptive materials are then heated to a suitable temperature to cause the analyte to desorb from the coating to be analysed. This is established as a means for preconcentrating samples for analysis by gas chromatography where the SPME matrix is in the form of a coating on a fibre that is introduced into the injection port. This yields information about the identification and concentration of volatile compounds and is applicable for both gas and liquid samples (321).

Here, a new variation of the technique was used, it was developed by SensAm Ltd, UK. Unlike a gas chromatograph, which is used for separating chemical components from a mixture and then applying them to an appropriate detector, the strategy was to capture and preconcentrate volatiles from urine headspace on to a flat tab that contains a proprietary semiconducting polymer (Figure 2.4a and Figure 2.4b). This is then desorbed on to detector that comprises an array of metal oxide gas sensors that generates a pattern of data that can be used to discriminate complex mixtures without separation into individual chemical components, which forms the basis of “electronic nose” technology (322). These polymer tabs are single use devices that have been pre-activated by heating during manufacture, do not require regeneration and thus do not suffer from recovery and carry-over problems that are found with SPME fibres used for gas chromatography.

The commercial metal oxide sensors (MOS), used as detectors in the system, are based on tin oxide that are doped with various metals that confer some selectivity to various chemical species. When heated to a temperature of about 300°C, they function as semiconductors. Atmospheric oxygen residing on the MOS surface is reduced by the target gases, allowing more electrons in the conduction band of the metal oxide material. They exhibit a change in electrical resistance of the metal oxide due to adsorption of gases or vapours, and this is proportional to the concentration adsorbed. Using a range of sensors with different chemical selectivities, incorporated into an array, allows a pattern of information to be generated that can be processed by multivariate techniques to allow discrimination of complex mixtures of volatile chemicals.
2.11.1 Urine sample processing using SPME

Following collection of 5 mL of urine from the participants, the samples were left to freeze within 2 hours to -80°C. Prior to analysis, samples were left to thaw in a water bath at 23°C for 1 hour. Urine was then placed into a 50 mL Falcon conical centrifuge tube with a modified cap which had two slots to allow two sensor tabs to be inserted. These tabs absorbed VOCs from the sample head space for a period of 300 seconds. (Figure 2.5)
2.11.2 SPME data processing

When the sensor tabs were inserted into the SensAm device (Figure 2.4b) for VOCs analysis, they were then automatically heated in the device to 120°C. Responses from the array of gas sensors to the desorbed vapours were captured over a period of 180 seconds, digitized and stored. A schematic diagram of the data processing steps adopted is shown in (Figure 2.6).

The portion of the gas sensor response, corresponding to the detection of desorbed volatiles, was extracted and averaged to produce n=5 samples straddling the response profile. This was normalised to minimise concentration effects. The autoscaling method was according to equation 1.

\[
A_{ij} = \frac{A_{ij} - \text{mean}(A)}{\text{std}(A)}
\]

Where \( A \) is the feature matrix for \( n \) samples over \( p \) sensors, \( A_{ij} \) is the \( i \)th sample of the \( j \)th sensor, \( A_j \) contains all \( n \) samples for sensor \( j \), and \( A_i \) contains all \( p \) samples for the sensors at the \( i \)th sample. This constituted a pattern of relative responses of the entire sensor array to a particular analyte sample.
2.11.3 SPME data analysis

This analysis was completed by Professor Krishna Persaud, University of Manchester.

For each patient sample, databases of the patterns were created, and these formed the basis for further data processing. To visualise the data from all the samples that were processed, the method of principal components analysis (PCA) was used (323).

PCA visualises the data while retaining as much of the variance as possible. This is by mathematical orthogonal data linear transformation to principal components. The greatest variance is explained by first principal component, the second greatest variance is explained by the second principal component and so on. PCA makes no assumptions to which classes the data belong and to discriminate between different classes of data a neural network based on radial basis functions was employed because of speed and robustness of performance in many practical applications (324,325).

Radial basis function networks (RBFN) are a variant of three-layer feedforward neural networks. They contain an input layer, a hidden layer and an output layer where the transfer function in the hidden layer is called a radial basis function (RBF). To each individual pattern in the databases a class name was assigned – these were Normal and HCC. The neural network was trained against 50% of the samples and the rest were used for evaluating the prediction accuracy of the neural network against previously unseen samples. The output nodes of the neural network were scaled to provide an output scaled between 0 to 1 representing the probability of an input pattern belonging to a certain class. To generate receiver operating characteristics (ROC) curves, the neural network was trained on a series of test cases versus controls, and ROC curves were generated for each of the test cases from the outputs of the neural network when assessed against previously unseen patterns, using established algorithms (326,327).

These produce reports on Sensitivity (the ability of the test to correctly identify those patients with the disease), Specificity (the ability of the test to correctly identify those patients without the disease), the Positive Predictive Value – how likely it is that the patient has the disease if the test is positive (this is also known as Precision), and Negative Predictive Value - how likely is it that the patient does not have the disease if the test result is negative (also known as Recall). These are used to produce a F-score which is a measure of a test's accuracy calculated from the precision and recall of the test, where the precision is the number of correctly identified positive results divided by the number of all positive results, including those not identified correctly, and the recall is the number of correctly identified positive results divided by the number of all samples that should have been identified as positive (328). The neural network was also tested to discriminate all cancers individually against controls as well as to discriminate each type of cancer.
2.12 Gas chromatography coupled with ion mobility spectrometry (GC-IMS)

Samples were shipped from UHCW in universal sample containers on dry ice to the School of Engineering, University of Warwick where they were stored at -20°C until use. Prior to testing the samples were thawed overnight in a laboratory fridge at 4°C. Once thawed, 5mL of each urine sample was aliquoted into 20mL glass vials (Thames Restek, UK), and sealed with a PTFE crimp cap (Thames Restek, UK). Samples were then analysed using a FlavourSpec GC-IMS (G.A.S, Dortmund, Germany) (Figure 2.7). The FlavourSpec is fitted with a combi PAL autosampler allowing for high-throughput automatic analysis of samples. The samples were loaded into a cooled autosampler tray keeping the samples at 4°C. Each sample was heated to 40°C and agitated for 10 minutes prior to analysis. A 0.5mL sample of the headspace was then taken using the autosampler syringe and injected directly into the GC-IMS for sampling. The GC-IMS settings were as follows: drift gas flow of 150mL/m, and a carrier gas flow rate of 20mL/min. The drift gas used was 99.99% Nitrogen. The IMS was heated to 45°C (T1), the GC to 40°C (T2), the injector to 80°C (T3), the T4 transfer line to 80°C and the T5 transfer line to 45°C. Sample analysis was for 10 minutes. Once completed, the data acquired was viewed using LAV software (G.A.S, Dortmund, Germany) and then exported for further analysis. This method has been developed over several urinary VOC studies and is designed
to maximize information content and chemical separation (329,330). This includes the volume of urine, agitation period and temperature. For quality control, blank samples were added at the beginning and end of each run, with the instrument having regular calibration standards run. Furthermore, the information content of each sample was checked, which included a visual inspection of each sample file.

**Figure 2.7** FlavourSpec GC-IMS (G.A.S, Dortmund, Germany)

---

### 2.13 Gas chromatography coupled to time-of-flight mass spectrometry (GC-TOF-MS)

A subset of samples was also analyzed on GC-TOF-MS (Markes International, UK), with a Unity thermal desorber and Ultra autosampler (Markes International, UK) (**Figure 2.8**). Urine samples for GC-TOF-MS were aliquoted as outlined about with 5mL of each sample in a 20mL vial, which was sealed with a crimp camp. The headspace from each urine sample was then adsorbed onto a Markes bio-monitoring tube (C2-AAXX-5149). The septa of the vial was pierced, and the sorbent tube pushed through into the headspace in the vial. The samples were then heated to 40°C for 20 minutes, before a pump was attached to the sorbent tube and the sample was pulled through onto the sorbent bed of the tube for 20 minutes whilst still being heated to 40°C. Once complete, the tube was removed from the vial and placed into the Markes Ultra autosampler. The Ultra autosampler is set to run with a standby split of 150°C, and the GC temperature ramp of 20°C per minute heating from 40°C to 280°C with a GC run
time of 25 minutes. The samples are each pre-purged for 1 minute, following which the sorbent tube is desorbed onto the trap for 10 minutes at 250°C. Once complete the trap was purged for a further minute, the trap was cooled to 30°C before being heated to 300°C for 3 minutes. Post analysis, a dynamic baseline correction (DBS) is applied using the native TOF-DS software, and the chromatogram was integrated and deconvoluted with the following settings: global height reject of 10000, global width reject 0.01, baseline threshold 3, and global area reject 10000. The peaks identified were then compared with the NIST list, with a match (forward and reverse) factor of 450, to identify compounds present. As with the GC-IMS, this method has been used on a number of VOC studies, including those associated with cancer and has been previously reported (330).

Figure 2.8 GC-TOF-MS system (TRACE 1300 GC Thermo Fisher Scientific, Loughborough, UK and BenchTOF-HD TOF-MS Markes Intl., Llantrisant, UK).
2.14 GC-IMS and GC-TOF-MS statistical Analysis

Urinalysis using GC-IMS and GC-TOF-MS was completed by Professor Covington research group in University of Warwick.

The analysis of the data was undertaken using previously reported data analysis pipeline for GC-IMS and GC-TOF-MS data, using “R” (version 3.6.3) (330-332). In brief, for GC–IMS data, a two-stage pre-processing step was applied. This was under-taken because the dataset has high dimensionality (typically 11 million data points), but low chemical information. The first step was to crop the central section of the out-put data, where all of the chemical information is located. This was followed by the application of a threshold, below which all values were given a value of zero. This was undertaken to remove the background, leaving just the chemical information. The crop parameters were manually selected, and the same values were applied to all of the data. The threshold was defined by the value of the background noise. The data were then processed using a 10-fold cross validation. Here, the data were split into a 90% training set and a 10% test set. Within each fold, a Wilcoxon rank sum test was undertaken, and the 100 features with the lowest p-value were extracted. Classification models were constructed using two classifiers (eXtreme Gradient Boosting (XGBoost), and logistic regression). This process was repeated until all of the samples had been in the test group. The results were then collated, and from the resultant probabilities, statistical parameters, including sensitivity and specificity, were calculated.

For GC-TOF-MS, a similar process was undertaken. However, in this case, chemical identification to create features was used and due to the much lower dimensionality, these were used directly by the classifier with no additional feature reduction. A further step used here was to undertake the statistical analysis of each chemical. A non-parametric t-test was undertaken in order to calculate the p-value of each chemical, comparing the samples in the two groups. Those chemicals had a p-value of <0.05 were considered statistically important. Figure 2.9 shows the process of acquiring data from urine samples and steps involved in statistical analyses in GC-TOF-MS and GC-IMS.
**Figure 2.9** schematic flow chart showing the process of acquiring data from urine samples and steps involved in statistical analyses in GC-TOF-MS and GC-IMS.
2.15 Testing for mSEPT9 in plasma using real time polymerase chain reaction (rt-PCR)

This work was completed by Professor Megan Hitchins research group, Cedars Sinai, Los Angeles, USA. Using modified single reaction PCR for mSEPT9 analysis. Schematic flow graphical summary showing hypothesis and all steps followed in analysing mSEPT9 is shown in Figure 2.10.

Testing for the presence of mSEPT9 ctDNA was performed using the Epi proColon V2.0 kits (Epigenomics, Inc.). For this work, up to 1 mL of plasma was available per subject for mSEPT9 testing. However, clinical testing using the Epi proColon 2.0 CE assay, as stipulated by the manufacturer and as approved by the FDA uses 3.5 mL plasma per subject. Available at www.accessdata.fda.gov/cdrh_docs/pdf13/p130001c.pdf (333).

From this, cfDNA was extracted and bisulfite-converted, then divided equally for input into three duplex real-time PCR amplification of mSEPT9 and ACTB (Actin, Beta). The ACTB serves as the internal control for cfDNA input and integrity to indicate the validity of each reaction. When the ACTB signal is detected before it reaches a designated cycle threshold (Ct) value ≤32.1 the reaction is considered valid. The FDA has approved the test scoring “1/3 algorithm” for ACTB-validated PCR reactions, whereby a positive mSEPT9 test result is designated if the mSEPT9 signal is detected at a Ct threshold value <45 in one or more of the three reactions. A negative test result is given if all three reactions are negative for mSEPT9.

By extrapolation, each PCR reaction contained bisulfite-converted cfDNA template derived from an initial plasma was calculated to be a volume of 0.875 mL.

Accordingly, a modified Epi proColon 2.0 CE cfDNA extraction and bisulfite-conversion component of the protocol for a starting plasma volume of 1 mL and the derived template was input into a single duplex PCR reaction. The amount of template input was derived from the equivalent of 0.875 mL plasma volume. Thus, the single real-time PCR reaction performed in this study used the same amount of bisulfite-converted cfDNA input, per starting volume of plasma, as each reaction performed in triplicate according to clinical testing conditions.

This was achieved by proportionally adjusting the volumes of reagents from the Epi proColon Plasma Quick Kit (Epigenomics, catalogue# M5-02-001) (333) used for cfDNA extraction from plasma and subsequent bisulfite conversion. Briefly, 1 mL plasma was thawed then treated with an equal volume (1 mL) of Epi proColon lysis buffer, vortexed and incubated at room temperature for 10 minutes. The cfDNA was bound to magnetic beads by adding 25.7 µL of the Epi proColon magnetic beads followed by 714 µL of absolute ethanol. The tube was gently inverted and mounted to on a rotating platform to gyrate at 20 rpm for 45 minutes at a 450 angle. The magnetic beads with bound cfDNA were then captured on a DynaMag-15 magnetic rack for 5 minutes and the supernatant removed. The beads were resuspended with 500 µL
of Epi proColon Wash A Buffer and transferred into a 2.0 mL tube that is subsequently inserted onto a DynaMag-2 magnetic rack. Once captured, the wash buffer was completely removed from the beads and 50 µL of Epi proColon Elution Buffer added. The reaction tube was shaken on a thermoshaker at 1000 rpm at 80° C for 10 minutes to elute the cfDNA from the magnetic beads. The beads were captured on the DynaMag-2 magnetic rack, and the 50 µl eluate containing the cfDNA transferred to a new 2.0 mL tube. For bisulfite conversion, 75 µL of Epi proColon bisulfite solution followed by 12.5 µL protection buffer were added to the cfDNA eluate (334).

The reaction was mixed thoroughly and placed on a thermoshaker at 80° C for 45 minutes. After bisulfite conversion, 500 µL of Epi proColon Wash A Buffer, followed by 10 µL of magnetic beads were added, and placed on a thermoshaker at 23° C and 1000 rpm for 45 minutes to bind the bisulfite-converted cfDNA to the magnetic beads. The beads were then captured on a DynaMag-2 magnetic rack and all buffer removed. The bead-bound bisulfite-converted cfDNA was washed with three washes as follows: 1) 500 µL Wash A Buffer, 2) 400 µL Wash B Buffer and 3) 200 µL Wash B Buffer, with capture of the beads on the DynaMag-2 magnetic rack after each wash. After the final wash, the buffer was completely removed, and the beads allowed to dry slightly by placing the tube at 23° C for 10 minutes. To elute the bisulfite-converted cfDNA from the magnetic beads, 17 µL of elution buffer was added and incubated on a thermoshaker set at 23° C, shaking at 1000 rpm for 10 minutes. The bisulfite-converted cfDNA eluate was collected following capture of the beads on the DynaMag-2 magnetic rack (334).

The real-time PCR reaction was performed according to the manufacturer’s protocol with 15 µl input volume of the bisulfite-converted cfDNA. With each batch of patient samples, 1 mL of the Epi proColon Positive and Negative Controls was processed alongside. Real-time PCR was conducted on the Applied Biosystems 7500 Fast, with thermal cycling conditions and results interpretation strictly according to the standard operating procedure. The tests were considered valid when the designated threshold for ACTB was reached in each reaction, and valid for each batch of samples when the Positive Control was positive for mSEPT9, and the Negative Control were negative for mSEPT9, according to the prespecified Ct (cycle threshold) values. Sensitivity and specificity to detect HCC were calculated based on the “1/1 algorithm”, whereby a mSEPT9 positive test result was designated if the plasma sample produced a signal under the specified Ct value <45 threshold in the single reaction, and negative if mSEPT9 was undetected, and the reaction was ACTB-validated (334).

Sample processing and mSEPT9 PCR analysis was completed for all participants in two batches of mixed samples (i.e., HCC and non-HCC plasma were included in both batches), and the operators were blinded to the disease status of the samples.
2.15.1 Quantification of mSEPT9 levels

To obtain semi-quantitative levels of plasma SEPT9 methylation, the percentage of methylated reference (PMR) was calculated, which has previously been applied to mSEPT9 qPCR results. Herein, use of the Ct values obtained for the Epi proColon® Positive Control sample (included during cfDNA extraction, bisulfite conversion, and qPCR for each batch of patient samples) as the reference sample for calibration of the Ct values for each patient sample included in the respective batch. The ΔΔCt value for each patient sample was calculated as ΔCtSample(ACTB-mSEPT9) - ΔCtReference(ACTB-mSEPT9). The PMR was calculated as $2^{\Delta\Delta C_t} \times 100$. Following PMR calculation, 0 represented a negative mSEPT9 test and >0 represented a positive mSEPT9 test (335).

2.15.2 mSEPT9 data analysis

This work was completed by author of the thesis and supervised by Dr Angela Noufaily from Warwick Medical School. Statistical analyses were undertaken using the R4.0.3 and IBM SPSS Statistics 26 software.

Descriptive statistics for each of the HCC and non-HCC subgroups were computed. Sensitivity and specificity for each of mSEPT9, AFP and their combination in identifying HCC were computed. Using the mSEPT9 PMR numerical values, receiver operating characteristic (ROC) analyses was performed to assess mSEPT9 performance amongst HCC patients and amongst HCC patients that also have LC. Mean AFP and PMR at different liver disease stages (liver disease alone, LC without HCC, and LC with HCC) was compared using non-parametric statistical tests. These results were demonstrated visually using box plots. Logistic regression analyses were performed, with subsequent odds ratios and p-values to identify the associations between each of mSEPT9, LC and HCC with each other and with the potential influential covariates age, gender, BMI, history of diabetes and blood count levels for platelets, NLR, albumin, bilirubin, ALBI and FIB-4 index. Cox-proportional hazards survival models to investigate the associations between participants' time-to-death from blood sampling with LC, HCC, mSEPT9, AFP and the other influential covariates were also completed. Kaplan-Meier survival curves for both AFP and mSEPT9 in patients with HCC and in all study participants were further performed (336-339).
Figure 2.10 schematic flow graphical summary showing hypothesis and all steps followed in analysing mSEPT9 in HCC.

Vascularized HCC nodules

Dying HCC cell releases circulatory tumour (ct) DNA molecules

Systemic circulation

Venipuncture followed by centrifugation

DNA methylation changes can be detected among the ctDNA molecules within the cell-free DNA (cfDNA) pool in peripheral blood plasma

DNA extraction and RT-PCR using Epi proColon 2.0 CE assay adapted for 1mL plasma using 1:1 algorithm

Data statistical analyses and clinical interpretation

Qualitative result with positive test for mSEPT9 if the signal reached the specified threshold in the single (ACTB - validated) reaction, and negative test if it did not
Chapter 3. LivFib-50 urinary peptide biomarker in liver fibrosis
Chapter 3. LivFib-50 urinary peptide biomarker in liver fibrosis

This work was done to explore urinary peptides and proteases in patients with liver fibrosis. This is to answer the first part of question 1 in section 1.30. The author declares that work described herein this chapter has been published (340).

3.1 Introduction

Liver fibrosis develops slowly as a consequence of progressive chronic inflammation, and it has clinical and life threatening implications to the affected patients. Liver fibrosis is associated with proteins changes in the liver. Liver fibrosis can be silent in patients and is often diagnosed late. Noninvasive diagnosis of liver fibrosis in practice is still lacking with the definitive tool being an invasive histopathological assessment of the liver by a biopsy (340).

3.2 Methods

Here, CEMS was applied to the urine of patients to search in the low molecular weight proteome of urine for peptide markers sensitive for liver fibrosis (LivFib-50). The CEMS steps used were mentioned in sections 2.7.1 to 2.7.8. The aim was to explore value of diagnostic utility of urinary peptides in those with liver fibrosis (discovery cohort) and then to validate this within two different populations.

Participants were recruited between 2014 to 2019 from both UHCW and Hannover Medical School, Germany. Work was approved by the relevant Ethics and R&D Committees (Ref 18717, Ref 260179 & Ref 901). Work was conducted according to the World Medical Association Declaration of Helsinki, with all participants providing written informed consent (341). The steps used during this experiment to generate the urinary LivFib 50 are shown in Figure 3.1.

In the discovery phase, the following were prospectively recruited: (i) 19 patients with NAFLD; (ii) 9 patients with NASH, but without LC manifestations; (iii) 13 patients with NASH and well established LC; (iv) 19 patients with histologically confirmed cirrhosis not related to NASH, mainly of viral or alcoholic origin; and (v) 19 patients with HCC and LC.

Differential diagnosis of these patients was established by a combination of liver ultrasound, Fibroscan, laboratory markers (e.g., AST/PLT-ratio), and histology. In order to reduce confounding factors, only stable cirrhotic patients from the liver outpatient clinics were included; none of whom had recently or were currently receiving a course of antibiotics. Patients with type 2 diabetes or other gastrointestinal conditions (e.g., inflammatory bowel disease or coeliac disease) were excluded from the work.
Eighty-one normal individuals adjusted for age, gender and renal function and without clinical or biochemical evidence of liver disease served as controls. These were selected from epidemiological study populations ensuring absence of diabetes, heart disease, hypertension, hyperlipidaemia and obesity. In the validation phase the following were recruited; 123 normal control subjects without evidence of liver disease and 19 patients with non-fibrotic liver diseases of various aetiologies including two with NAFLD, three with NASH and three with HCC as non-LC control groups and 50 patients with biopsy-proven LC of whom 11 had concomitant HCC, 13 benign biliary disorders, 9 chronic pancreatitis, 4 cholangiocarcinoma and 4 pancreatic cancer as LC case group. A set of 41 patients with no liver, but kidney fibrosis was included to evaluate potential interference of the liver fibrosis-specific urinary peptide profile with expressions of kidney fibrosis. Composition of the liver disease case and normal control groups together with the applied exclusion criteria and adjusted demographic variables are presented in Table 3.1.
Biomarker selection & model establishment

Discovery set of 79 liver diseased patients and 81 age- and gender-matched normal controls (NC).

1. **1st sequential biomarker selection step:** Statistical comparison of peptide amplitudes by a group-wise non-parametric Wilcoxon rank sum test and correction for multiple testing by the Benjamini & Hochberg procedure.

2. **2nd sequential biomarker selection step:** Rank sum correlation of peptide amplitudes to a liver disease severity score (0 = NC, 1 = NAFLD, 2 = NASH, 3 = LC).

3. **3rd sequential biomarker selection step:** Kruskal-Wallis rank sum test for gradual decrease or increase of peptide amplitudes from NC over NAFLD and LC-negative NASH patients to LC-positive patient groups.

4. Integration of the remaining peptide markers from the peptide marker selection steps by support vector machine modelling with subsequent multivariate parameter optimization.

Model evaluation & confounder analysis

1. **1st validation set (cross-sectional) of 31 liver diseased patients with LC and 123 NC.**

5. Evaluation of the peptide marker model’s classification characteristics on an independent set of patient samples.

6. Confounder analysis of the peptide marker model’s classification values to the relevant clinical and demographic patient characteristics and age-adjustment of the peptide marker model by logistic regression.

Model validation

2. **2nd validation set of 19 liver diseased patients with LC, 17 non-fibrotic liver diseased patients (both groups prospective) and 41 patients with kidney, but no liver fibrosis.**

7. Evaluation of the age-adjusted peptide marker model’s classification characteristics on an independent set of liver diseased patients in a clinical setting of intended use.

8. Evaluation of the model’s potential interference with kidney fibrosis.

**Figure 3.1 Schematic flow chart.** A total of 391 patients was included in the discovery and validation phases of the LivFib-50 peptide marker model for the detection of liver fibrosis. Biomarker selection was carried out in three sequential steps resulting in 50 urinary peptides with differential and graded expression ranging from disease-free normal individuals over NAFLD and NASH patients without LC to patients with well-established LC. Combining the peptide markers to the LivFib-50 model was followed by a first evaluation of the model’s classification performance and confounder analysis in patients with LC and normal controls. Since the classification factors significantly correlate with the age of the patients, the LivFib-50 classification model was adjusted for age on these patient groups by logistic regression. In a final validation phase the age-adjusted LivFib-50 classification model was validated in an independent group of patients with liver disease, with or without LC and interference of classification was tested in a set of patients with renal fibrosis, but no liver fibrosis.
Table 3.1 Selected clinical and demographic characteristics of recruited participants investigated by CEMS to generate the LivFib 50 urinary peptide biomarker in liver fibrosis

<table>
<thead>
<tr>
<th>Patient group</th>
<th>Liver diseased patients</th>
<th>Normal controls</th>
<th>Other validation control groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Discovery</td>
<td>Validation</td>
<td>P*</td>
</tr>
<tr>
<td>No patients, n</td>
<td>79</td>
<td>50</td>
<td>---</td>
</tr>
<tr>
<td>Age mean (range), y</td>
<td>56</td>
<td>59</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>(18-79)</td>
<td>(31-86)</td>
<td></td>
</tr>
<tr>
<td>Gender female/male, n</td>
<td>34/45</td>
<td>10/40</td>
<td>0.08</td>
</tr>
<tr>
<td>No peptides mean (range), 1 x 10^3</td>
<td>2.3</td>
<td>2.4</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>(0.6-4.0)</td>
<td>(0.8-4.6)</td>
<td></td>
</tr>
<tr>
<td>WBC mean (range), 1 x 10^3/μL</td>
<td>5</td>
<td>7</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(1-13)</td>
<td>(2-18)</td>
<td></td>
</tr>
<tr>
<td>Hb mean (range), g/dL</td>
<td>13</td>
<td>13</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>(8-18)</td>
<td>(9-41)</td>
<td></td>
</tr>
<tr>
<td>Pit mean (range), 1 x 10^9/L</td>
<td>145</td>
<td>182</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>(25-402)</td>
<td>(51-568)</td>
<td></td>
</tr>
<tr>
<td>INR mean (range)</td>
<td>1.2</td>
<td>1.2</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>(0.9-2.0)</td>
<td>(1.0-1.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC mean (range), mg/dL</td>
<td>329</td>
<td>314</td>
<td>0.85</td>
</tr>
<tr>
<td>HDL mean (range), mg/dL</td>
<td>121</td>
<td>114</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>(9-340)</td>
<td>(27-350)</td>
<td></td>
</tr>
<tr>
<td>TG mean (range), mg/dL</td>
<td>130</td>
<td>139</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>(46-400)</td>
<td>(48-610)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AST</td>
<td>ALT</td>
<td>ALP</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>mean</td>
<td>(range), U/L</td>
<td>mean (range), U/L</td>
<td>mean (range), U/L</td>
</tr>
<tr>
<td></td>
<td>50 (13-284)</td>
<td>48 (11-288)</td>
<td>128 (44-797)</td>
</tr>
<tr>
<td></td>
<td>68 (20-180)</td>
<td>76 (15-1,038)</td>
<td>230 (44-693)</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.40</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>21 (13-65)</td>
<td>68 (12-46)</td>
<td>68 (19-174)</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>---</td>
<td>88 (45-358)</td>
</tr>
<tr>
<td></td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>44 (5-246)</td>
<td>45 (5-246)</td>
<td>88 (45-358)</td>
</tr>
</tbody>
</table>

* Difference between discovery and validation cohort by two-tailed probability for continuous data and significance level by Fisher exact test for categorical data. Diagnosis was established by a combination of liver ultrasound, Fibroscan, laboratory markers, e.g., AST/PLT-ratio, and histology. Patients with type 2 diabetes or other gastrointestinal conditions (e.g., inflammatory bowel disease or coeliac disease) were excluded. Controls were adjusted for age, gender and renal function and without clinical or biochemical evidence of liver disease served as controls. Absence of diabetes, heart disease, hypertension, hyperlipidemia and obesity were ensured.

**Abbreviations:** AFP, alpha-fetoprotein; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; Hb, hemoglobin; HDL, high density lipoprotein; INR, International Normalized Ratio; Plt, platelets; TC, total cholesterol; TG, triglycerides; WBC, white blood cells.
3.3 Results

For the identification of urinary peptide markers associated with fibrosis, the low molecular weight urine proteome was investigated from 79 patients with chronic liver diseases and 81 age and gender-matched normal subjects by CEMS. Age- and gender-adjustment of controls was implemented in the exploratory work design of the discovery phase to ensure that only those peptides that are unaffected by age or gender were selected as marker for liver fibrosis.

For age and gender adjustment, 81 normal controls were selected from a pool of 204 normal controls by using the R-script “MatchIt”, which conducts nearest neighbour matching with logistic regression used to estimate the propensity score (342). The age of the 81 selected normal controls was subsequently compared with that of the 79 liver fibrosis case patients by a Mann-Whitney rank sum test and resulted in a p-value of 0.18. The difference in gender was proven to be statistically insignificant (p = 0.63) by a Chi-squared test. For renal function it was verified that all patients have no renal dysfunction at the date of urine collection as indicated by an eGFR-value above 60 mL/min/1.73m².

Fibrosis markers showing a gradual increase or decrease in their CEMS-detected amplitude signals from normal over non-fibrotic NAFLD and NASH to LC were specifically searched for. For this purpose, three-step strategy which is schematically presented in Figure 3.1 was followed.

In the first selection step, the CEMS peptide profiles of the patients with liver diseases to those of the normal control subjects by a group-wise non-parametric Wilcoxon rank sum test were statistically compared. From a total of 21,559 annotated peptides in the mass range of 0.8 to 20 kDa, 854 peptides were identified with a frequency threshold of 30% and p-values below 0.05. Out of these, 732 remained significant after adjustment of the false discovery rate for multiple testing by the Benjamini and Hochberg procedure (310).

In the second selection step, the abundance of the 732 preselected peptides with a severity score of liver disease were investigated for correlation, assigning 0 for normal controls, 1 for patients with NAFLD, 2 for patients with NASH (both without signs of liver fibrosis), and 3 for patients with LC to select peptides that appear associated with progression of liver fibrosis.

By selecting a threshold above 0.2 or below -0.2 for Spearman rho correlation (FDR-adjusted p-values < 0.05), 500 out of the 732 peptides selected in step 1 fulfilled also the selection criteria from step 2 (343). Out of the selected peptides, 198 showed negative correlation and 302 were positively correlated to the severity scores of the liver insult. In this way, selection step 2 was used as a first step to identify those liver disease-associated peptides that show a progression in their abundance depending on the presence of liver fibrosis.
In the final biomarker selection step, the list of the preselected 500 peptides for those fibrosis associated markers were searched for, showing a strictly linear decrease or increase from normal controls over the NAFLD and NASH patients to the LC-positive patients when applying a Kruskal-Wallis rank sum test. This last, very stringent selection step reduced the number of fibrosis grading markers to 50 peptides. The final list of fibrosis markers together with their statistical characteristics of all three selection steps is presented in Table 3.2.

These 50 liver fibrosis peptide markers were combined into a support vector machine (SVM)-based peptide multi-marker panel for non-invasive detection of liver fibrosis. The peptide marker model was trained by SVM using a Radial Basis Function (RBF) kernel using a cost C value of 9.7 and a gamma value of 0.00073 as penalty settings. In combination, these two parameters control the trade-off between allowing training errors and forcing rigid margins. After optimization of the SVM parameters, the peptide marker pattern designated as LivFib-50 including the 50 peptides resulted in an AUC of 0.95 and 95% (CI) in the range of 0.90 to 0.98 (p<0.0001) on the 19 NAFLD, 9 NASH and 51 LC (out of which 12 are NASH-LC) patients and 81 normal control subjects in receiver operating characteristics analysis (ROC) after take-one-out total cross-validation. Figure 3.2 shows contour plot for the LivFib-50 peptide marker profiles for liver fibrosis cases and non-liver fibrosis controls. The ROC curve and ROC characteristics are presented in Figure 3.3. As presented in the group-specific Box-and-Whisker distribution plot of Figure 3.4, the established peptide marker pattern not only allowed differentiation of the normal controls from NAFLD, NASH and LC, but also the classification scores for NAFLD and NASH-LC are significantly different, with higher values in the case of LC.

Adjustment of the LivFib-50 peptide marker model for patient age. In a first validation phase, LivFib-50 was evaluated for generalizability on independent cross-sectional datasets of 31 patients with LC and 123 normal controls. ROC analysis revealed an AUC of 0.94 (95% CI: 0.89-0.97, p<0.0001) in this diagnostic phase II testing. Besides performance evaluation, correlation of LivFib-50 classification values to relevant clinical and demographic patient characteristics was done in this first validation cohort and not in the training cohort since the former represents an independent patient group free of any overfitting bias and thus without the need for any bootstrapping (344).

Confounder analysis revealed correlation of LivFib-50 classification to age (r = 0.43, p < 0.0001) as the only confounder with impact on the LivFib-50 classification result. Most importantly and with the results shown in Figure 3.5, a potential impact of diabetes mellitus (DM) was excluded by dividing the 31 patients with LC into those with (N=9) and without (N=22) DM and by performing a global Kruskal Wallis analysis also including the 123 patients without signs of liver disease. As a result of this exploratory confounder analysis on this first
validation set, an age-adjusted LivFib-50 model was constructed by logistic regression of the LivFib-50 classification values and age on the basis of their odds ratios of 27.54 (95% CI: 6.87-110.33) and 1.16 (95% CI: 1.06-1.27) resulting in a logit function for an age-adjusted LivFib-50 score of -7.37 + 3.32 × LivFib-50 classification factor + 0.15 × patient age. As presented in Figure 3.6, the age-adjusted LivFib-50 classification model significantly improved classification performance relative to the LivFib-50 peptide marker model alone by an AUC increase of 0.027 (AUC = 0.96, p = 0.044) on the 31 LC cases in comparison to the 123 normal controls.

Validation of the age-adjusted LivFib-50 classification model. For validation in a clinical setting of intended use, the age-adjusted LivFib-50 model was evaluated in an independent set of 36 prospectively collected chronic liver disease patients, 19 with and 17 without LC. This second validation set is essential for the diagnostic test validation process in order to compensate for potential classification bias by the fitted logistic regression model algorithm. As demonstrated by an AUC of 0.91 and a 95% CI range of 0.76 to 0.98 at a significance level of p<0.0001 (Figure 3.7), patients with LC are clearly distinguishable in their LivFib-50 model scores from patients without LC. Without age-adjustment the LivFib-50 model resulted in an AUC of 0.87 (95% CI: 0.71-0.96, p<0.0001) on this second validation set. The positive (PPV) and negative (NPV) predictive values were determined to be 67% (95% CI: 42-85 %) and 92% (95% CI: 81-97 %), respectively.

Of note, the PPV and NPV were corrected for a 30% prevalence of liver fibrosis in chronic liver disease patients since the prospective validation cohort does not fully reflect the reported value for this patient collective.

Applying the optimal classification threshold based on the Youden index at 0.41, classification of this validation cohort resulted in a sensitivity of 84.2 % (95% CI: 60.4-96.6) and a specificity of 82.4 % (95% CI: 56.6-96.2). As revealed by the Box-and-Whisker distribution plot of LibFib-50 classification values in Figure 3.8, there is a clear separation between the fibrotic and non-fibrotic patient groups irrespective of concomitant HCC. A clear distinction from the case group is also observed for patients with no liver, but kidney fibrosis as clearly demonstrated by a specificity of 73 % for LivFib-50 positivity in this interference risk group.

Sequencing of liver fibrosis peptide markers reveal differential urinary excretion of collagen chain fragments. This work was able to obtain high confidence amino acid sequence information for 22 of these 50 peptides. All liver fibrosis peptide markers that could be successfully resolved for their amino acid sequence are presented in Table 3.3. These peptides are fragments of collagen α-1(I) (n=14) and five other collagen fragments. Of these, 10 showed increased and nine showed decreased excretion during fibrosis. Other peptides
are derived from the renal homeostasis proteins such as uromodulin (n=2) and Na/K-transporting ATPase subunit γ (n=1), with decreased urinary levels in the former and increased urinary levels in the latter case.

Following the hypothesis that naturally occurring peptides arise from proteolytical degradation of their parent proteins, observed protease cleavage sites in the CutDB database were searched for. This is on the basis of the octamer amino acid sequence motifs around the N- and C-terminal ends of the 22 sequence-identified fibrosis peptide markers. As presented in Table 3.4, this search identified ten putative proteolytic cleavage sites, eight with specificity for matrix metallopeptidases and two for cathepsins. Protease activity was subsequently assessed in the patient’s CEMS peptide profiles based on the average of associated peptide intensities in the 31 LC cases compared to the 123 normal controls of the cross-sectional validation set. This analysis revealed a fold change (FC) in the activity between LC cases and normal controls of 0.86 (p=0.0005) for MMP2, 0.89 (p=0.0016) for MMP3, 0.56 (p<0.0001) for MMP13, 0.90 (p=0.0034) for MMP14 and 0.74 (p<0.0001) for CTSB.

Further analysis in this work included tissue transcriptomics data of 22 cirrhotic and 14 normal human livers retrieved from the NCBI GEO database (GSE77627) to compare it with the proteomic results reached in this CEMS exploratory experimental work. As presented in Table 3.5, this analysis identified significant differences in gene expression levels between cirrhotic and normal tissue for the collagen chains COL1A2 (FC: 0.66, p = 0.04) and COL3A1 (FC: 0.28, p = 5.26E-06), both with a decreased fold change ratio in LC, whereas COL18A1 (FC: 1.49, p = 0.026) showed increased expression in LC. For the protease CTSB a significant decrease in its expression levels was found, with the fold change in LC being 0.45 (p = 4.31E-04). The relevant MMP’s demonstrate an increase in expression, as in the case of MMP2 (FC: 3.08, p = 0.002), or no significant expression differences between cirrhotic and normal livers, as in the case of the MMP’s 3, 13 and 14.

Tissue transcriptomics data from 22 cirrhotic and 14 normal human livers (GSE77627) were retrieved from the NCBI Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). The GEO data sets were analyzed with GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/). Analysis was conducted using default settings, with false discovery rate (FDR) adjustment of p-values based on the method by Benjamini & Hochberg (308-310).
### Table 3.2. CEMS characteristics and descriptive statistics of the 50 urinary peptides included in the LivFib-50 peptide marker model for diagnosis and grading of progressive liver fibrosis.

<table>
<thead>
<tr>
<th>Peptide ID*</th>
<th>CEMS characteristics</th>
<th>Group-wise statistical comparison of peptide distributions between liver disease patients and NC</th>
<th>Rank sum correlation of peptide amplitudes to a liver disease grading score</th>
<th>Kruskal-Wallis rank sum test for gradual changes of peptide amplitudes between patient subgroups from NC over non-fibrotic NAFLD and NASH to LC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2136</td>
<td>1,137.55</td>
<td>23.85</td>
<td>7.18E-04</td>
<td>0.62</td>
</tr>
<tr>
<td>2740</td>
<td>1,203.57</td>
<td>26.90</td>
<td>2.14E-03</td>
<td>0.62</td>
</tr>
<tr>
<td>2844</td>
<td>1,215.44</td>
<td>26.94</td>
<td>7.37E-08</td>
<td>0.74</td>
</tr>
<tr>
<td>3079</td>
<td>1,241.54</td>
<td>26.61</td>
<td>2.30E-06</td>
<td>0.69</td>
</tr>
<tr>
<td>3793</td>
<td>1,312.55</td>
<td>29.69</td>
<td>1.36E-07</td>
<td>0.74</td>
</tr>
<tr>
<td>4419</td>
<td>1,378.61</td>
<td>28.88</td>
<td>1.33E-12</td>
<td>0.82</td>
</tr>
<tr>
<td>5241</td>
<td>1,467.81</td>
<td>24.78</td>
<td>2.02E-02</td>
<td>0.61</td>
</tr>
<tr>
<td>5787</td>
<td>1,536.68</td>
<td>22.29</td>
<td>6.93E-04</td>
<td>0.65</td>
</tr>
<tr>
<td>5810</td>
<td>1,538.69</td>
<td>29.69</td>
<td>1.76E-05</td>
<td>0.69</td>
</tr>
<tr>
<td>5833</td>
<td>1,540.77</td>
<td>29.95</td>
<td>1.29E-05</td>
<td>0.69</td>
</tr>
<tr>
<td>6546</td>
<td>1,620.70</td>
<td>22.69</td>
<td>4.58E-03</td>
<td>0.63</td>
</tr>
<tr>
<td>7414</td>
<td>1,711.63</td>
<td>22.87</td>
<td>1.14E-02</td>
<td>0.59</td>
</tr>
<tr>
<td>7550</td>
<td>1,725.58</td>
<td>38.53</td>
<td>4.17E-05</td>
<td>0.69</td>
</tr>
<tr>
<td>7762</td>
<td>1,746.73</td>
<td>30.23</td>
<td>7.16E-03</td>
<td>0.59</td>
</tr>
<tr>
<td>8397</td>
<td>1,817.70</td>
<td>20.25</td>
<td>1.90E-09</td>
<td>0.77</td>
</tr>
<tr>
<td>8462</td>
<td>1,825.79</td>
<td>31.80</td>
<td>4.82E-04</td>
<td>0.65</td>
</tr>
<tr>
<td>8534</td>
<td>1,835.71</td>
<td>19.87</td>
<td>4.60E-07</td>
<td>0.73</td>
</tr>
<tr>
<td>9061</td>
<td>1,892.88</td>
<td>24.45</td>
<td>2.29E-07</td>
<td>0.72</td>
</tr>
<tr>
<td>9253</td>
<td>1,915.00</td>
<td>33.62</td>
<td>1.82E-04</td>
<td>0.63</td>
</tr>
<tr>
<td>9627</td>
<td>1,962.81</td>
<td>21.88</td>
<td>2.76E-03</td>
<td>0.62</td>
</tr>
<tr>
<td>Peptide ID*</td>
<td>CEMS characteristics</td>
<td>Group-wise statistical comparison of peptide distributions between liver disease patients and NC</td>
<td>Rank sum correlation of peptide amplitudes to a liver disease grading score</td>
<td>Kruskal-Wallis rank sum test for gradual changes of peptide amplitudes between patient subgroups from NC over non-fibrotic NAFLD and NASH to LC</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>9911</td>
<td>1,996.78</td>
<td>20.96</td>
<td>2.98E-09 0.77 1.72E-07 -0.51 1.01E-09</td>
<td>43.63 &lt;1.00E-06 101.66 80.71 78.83 47.11</td>
</tr>
<tr>
<td>10170</td>
<td>2,035.01</td>
<td>40.03</td>
<td>1.17E-02 0.60 2.98E-02 -0.27 4.57E-03</td>
<td>17.85 4.72E-04 87.86 95.79 87.94 61.79</td>
</tr>
<tr>
<td>10424</td>
<td>2,067.82</td>
<td>20.60</td>
<td>7.90E-10 0.78 8.15E-08 -0.51 1.01E-09</td>
<td>42.92 &lt;1.00E-06 102.43 80.26 59.22 49.51</td>
</tr>
<tr>
<td>10953</td>
<td>2,149.95</td>
<td>27.60</td>
<td>2.28E-06 0.72 3.24E-05 0.42 1.47E-06</td>
<td>27.50 5.00E-06 63.67 75.79 98.11 105.88</td>
</tr>
<tr>
<td>11209</td>
<td>2,187.96</td>
<td>39.54</td>
<td>2.06E-08 0.75 7.09E-07 -0.52 7.09E-10</td>
<td>52.84 &lt;1.00E-06 100.94 94.74 83.78 42.16</td>
</tr>
<tr>
<td>11744</td>
<td>2,261.00</td>
<td>27.14</td>
<td>6.86E-03 0.60 1.92E-02 -0.26 6.88E-03</td>
<td>12.32 6.38E-03 88.48 87.74 83.00 64.69</td>
</tr>
<tr>
<td>11753</td>
<td>2,261.99</td>
<td>27.14</td>
<td>1.89E-03 0.64 6.70E-03 0.27 4.98E-03</td>
<td>10.02 1.84E-02 69.53 85.47 85.22 95.24</td>
</tr>
<tr>
<td>12124</td>
<td>2,310.05</td>
<td>41.01</td>
<td>1.47E-02 0.61 3.53E-02 -0.29 2.86E-03</td>
<td>20.98 1.06E-04 89.15 98.74 94.17 57.55</td>
</tr>
<tr>
<td>13021</td>
<td>2,439.09</td>
<td>27.84</td>
<td>8.91E-09 0.75 4.00E-07 0.43 7.19E-07</td>
<td>34.17 &lt;1.00E-06 60.85 93.58 95.33 104.22</td>
</tr>
<tr>
<td>13342</td>
<td>2,503.13</td>
<td>28.32</td>
<td>1.38E-04 0.66 8.12E-04 0.31 8.46E-04</td>
<td>19.95 1.73E-04 67.54 84.26 71.00 101.35</td>
</tr>
<tr>
<td>13730</td>
<td>2,567.20</td>
<td>28.23</td>
<td>3.76E-03 0.62 1.16E-02 0.25 9.40E-03</td>
<td>10.09 1.78E-02 71.47 80.47 83.00 94.41</td>
</tr>
<tr>
<td>13746</td>
<td>2,570.19</td>
<td>42.35</td>
<td>1.44E-06 0.72 2.23E-05 -0.48 2.47E-08</td>
<td>43.01 &lt;1.00E-06 97.85 100.05 78.22 46.07</td>
</tr>
<tr>
<td>13779</td>
<td>2,577.25</td>
<td>24.62</td>
<td>2.02E-06 0.72 2.94E-05 0.37 5.28E-05</td>
<td>21.96 6.70E-05 63.91 88.63 101.00 100.20</td>
</tr>
<tr>
<td>14801</td>
<td>2,754.28</td>
<td>29.59</td>
<td>2.10E-02 0.60 4.70E-02 -0.23 2.30E-02</td>
<td>11.62 8.79E-03 89.33 88.32 87.78 62.27</td>
</tr>
<tr>
<td>15053</td>
<td>2,796.24</td>
<td>28.66</td>
<td>4.87E-03 0.61 1.45E-02 0.24 1.60E-02</td>
<td>9.36 2.48E-02 72.30 78.79 83.33 93.67</td>
</tr>
<tr>
<td>15811</td>
<td>2,933.37</td>
<td>27.64</td>
<td>6.78E-04 0.65 2.90E-03 -0.33 4.74E-04</td>
<td>21.87 6.90E-05 92.86 91.08 80.06 57.00</td>
</tr>
<tr>
<td>16149</td>
<td>3,005.44</td>
<td>24.59</td>
<td>5.17E-06 0.64 5.69E-05 0.38 2.12E-05</td>
<td>27.49 5.00E-06 69.41 81.92 73.89 98.75</td>
</tr>
<tr>
<td>16149</td>
<td>3,064.34</td>
<td>20.64</td>
<td>1.04E-02 0.61 2.72E-02 -0.27 6.45E-03</td>
<td>16.58 8.63E-04 89.35 99.05 74.56 60.59</td>
</tr>
<tr>
<td>16427</td>
<td>3,067.31</td>
<td>48.89</td>
<td>1.05E-04 0.64 6.54E-04 -0.32 5.50E-04</td>
<td>16.97 7.17E-04 91.14 79.68 74.61 64.95</td>
</tr>
<tr>
<td>16632</td>
<td>3,114.42</td>
<td>30.34</td>
<td>7.32E-05 0.67 4.91E-04 0.33 3.28E-04</td>
<td>17.78 4.84E-04 67.60 79.68 93.11 99.06</td>
</tr>
<tr>
<td>16811</td>
<td>3,158.42</td>
<td>29.80</td>
<td>1.08E-04 0.68 6.69E-04 0.33 4.49E-04</td>
<td>15.60 1.37E-03 67.07 82.08 95.11 98.67</td>
</tr>
<tr>
<td>16969</td>
<td>3,194.43</td>
<td>30.28</td>
<td>5.33E-08 0.72 1.57E-06 0.45 2.11E-07</td>
<td>30.59 1.00E-06 63.75 84.87 95.11 102.89</td>
</tr>
<tr>
<td>17021</td>
<td>3,205.28</td>
<td>19.51</td>
<td>6.75E-09 0.76 3.22E-07 -0.48 1.95E-08</td>
<td>39.65 &lt;1.00E-06 101.26 82.74 47.28 52.56</td>
</tr>
<tr>
<td>18250</td>
<td>3,521.61</td>
<td>30.50</td>
<td>8.93E-04 0.63 3.59E-03 -0.30 1.68E-03</td>
<td>16.05 1.11E-03 90.83 84.84 79.56 62.65</td>
</tr>
<tr>
<td>Peptide ID*</td>
<td>MS exptl. mass [Da]</td>
<td>CE migr. time [Min]</td>
<td>CEMS characteristics</td>
<td>Group-wise statistical comparison of peptide distributions between liver disease patients and NC</td>
</tr>
<tr>
<td>------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>19314</td>
<td>3,886.83</td>
<td>33.49</td>
<td>unadj. p-value</td>
<td>AUC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.96E-03</td>
<td>0.62</td>
</tr>
<tr>
<td>19533</td>
<td>4,002.63</td>
<td>20.69</td>
<td>9.59E-09</td>
<td>0.76</td>
</tr>
<tr>
<td>19706</td>
<td>4,078.81</td>
<td>33.06</td>
<td>1.54E-04</td>
<td>0.67</td>
</tr>
<tr>
<td>20181</td>
<td>4,368.87</td>
<td>20.21</td>
<td>1.07E-05</td>
<td>0.69</td>
</tr>
<tr>
<td>20959</td>
<td>5,526.35</td>
<td>27.28</td>
<td>6.63E-03</td>
<td>0.59</td>
</tr>
<tr>
<td>21405</td>
<td>10,341.86</td>
<td>22.96</td>
<td>3.96E-05</td>
<td>0.64</td>
</tr>
</tbody>
</table>

* Peptide identification number.

**Abbreviations:** CE, capillary electrophoresis; coef., coefficient; exptl., experimental; FDR-adj., false discovery rate-adjusted; LC, liver cirrhosis. migr., migration; MS, mass spectrometry; NAFLD, non-alcoholic fatty liver disease; NASH, Non-alcoholic steatohepatitis; NC, normal control; unadj., unadjusted.
Figure 3.2 Contour plot showing the LivFib-50 peptide marker profiles for liver fibrosis cases and non-liver fibrosis controls with the CE migration time on the x-axis and the molecular weight in log scale on the y-axis.
Figure 3.3 ROC curve and ROC characteristics for the discovery set. Patients with NAFLD, NASH, LC or HCC were treated as case group (N=79) and were compared to non-diseased age- and gender-matched normal controls (N=81). Bootstrapping of the classification results was performed by leave one out total cross-validation. Dotted lines represent the upper and lower bounds of the confidence interval.
Figure 3.4 Box and Whisker distribution plots for classification of the different patient groups of the discovery set with the LivFib-50 model. A Kruskal-Wallis test was performed for rank sum differences in the LivFib-50 classification scores and revealed significant differences in post-hoc analysis between normal controls to all liver diseased patient groups (p < 0.0001) as well as between patients with combined NASH and LC (NASH-LC) compared to NASH without concomitant LC (p = 0.04).
Figure 3.5 Distribution of classification scores of the LivFib-50 marker pattern in normal liver and liver fibrosis groups of the first validation set of patients. The liver fibrosis group (N=31) was further divided into those with (N=9) or without (N=22) concomitant diabetes mellitus in order to evaluate the impact of diabetes mellitus on the LivFib-50 classification results. A post hoc test was performed for average rank differences between the three different subgroups (each with p< 0.05) after a significant result in the global Kruskal-Wallis test. The abbreviation n.s. indicates a non-significant result.
**Figure 3.6** Classification of normal controls without clinical signs of liver disease (NC, N=123) and those with clinical manifestations of liver cirrhosis (LC, N=31) with the age-adjusted LivFib-50 peptide marker model in comparison to the proteomic model without age adjustment and age alone. Age adjustment of the LivFib-50 peptide marker model was performed using logistic regression. Significance P values for each ROC curve were determined to be <0.0001.
Figure 3.7 ROC curve and ROC characteristics of the age-adjusted LivFib-50 peptide marker model for patients with LC in the absence or presence of HCC (N=19) compared to non-fibrotic control patients (N=17) in independent validation. Of note, the three HCC patients without LC manifestations were treated as controls.
**Figure 3.8** Box and Whisker distribution plots for classification of the different patient groups of the validation set with the age-adjusted LivFib-50 classification model. The validation set consists of patients without clinical signs of liver fibrosis (N=17), patients with kidney fibrosis (N=41) and patients with LC or fibrosis (N=19).
Table 3.3 Amino acid alignment of all sequence-identified naturally occurring urinary peptides included in the LivFib-50 peptide marker model due to their graded association with progressive liver fibrosis. Peptide markers are overlapping fragments derived from the triple helical region of the collagen α1(I) chain. Opposite regulation of overlapping fragments might be attributed to changes in the activity of extracellular matrix degrading proteases during fibrosis progression.

<table>
<thead>
<tr>
<th>Peptide ID&lt;sup&gt;†&lt;/sup&gt;</th>
<th>Sequence‡</th>
<th>AA§</th>
<th>Regulation in liver fibrosisǁ</th>
<th>Protein name</th>
</tr>
</thead>
<tbody>
<tr>
<td>6546 PpGPpGKNGDDGEAGKP</td>
<td>222-238</td>
<td>↓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9627 PpGPpGKNGDDGEAGKpGRp</td>
<td>222-241</td>
<td>↓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13021 LDGAKGDAGpAGpKGEpGSpGENApG</td>
<td>273-299</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5810 PpGEAGKpGEQGVpGD</td>
<td>651-666</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3793 GEAKKPGEQGVPGD</td>
<td>653-666</td>
<td>↓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8462 GANGApGNDGAKGDAGApGApG</td>
<td>698-719</td>
<td>↓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4419 ApGDRGEpGPpGPAG</td>
<td>798-812</td>
<td>↓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2136 GDRGEpGPpGPA</td>
<td>800-811</td>
<td>↓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14801 GPpGADQpGAKGEpGDAGAKGDAGpPGPAGP</td>
<td>815-846</td>
<td>↓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11753 GADQpGAKGEpGDAGAKGDAGpGP</td>
<td>818-843</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13342 GADQpGAKGEpGDAGAKGDAGpPGPAGP</td>
<td>818-846</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10953 DGQpGAKGEpGDAGAKGDAGpGP</td>
<td>820-843</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3079 EKGSpGADGpAGAP</td>
<td>933-946</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13730 AGPpGAPGApGAPGPVPGAPKSGDRGETGP</td>
<td>1042-1071</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11744 LQGLpGTGGpGENKpGEpGPKG</td>
<td>640-663</td>
<td>↓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peptide ID</td>
<td>Sequence</td>
<td>Position</td>
<td>Regulation</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>----------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>9061</td>
<td>GApGApGGKGDAGApGERGPpG</td>
<td>666-687</td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>13779</td>
<td>DDILASPPRLPEPQYPGAPHHSS</td>
<td>1534-1557</td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>16419</td>
<td>EAGRDGNpGNDGPpGRDGQpGHkGERGYPG</td>
<td>923-952</td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>2740</td>
<td>YLGSPKGDVDP</td>
<td>5-16</td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>5833</td>
<td>SGSVIDQSRVLNLGP</td>
<td>589-603</td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>5241</td>
<td>DQSRVNLGPITR</td>
<td>594-606</td>
<td>↓</td>
<td></td>
</tr>
</tbody>
</table>

**Collagen α-1(III) chain (COL3A1)**

**Collagen α-1(XVIII) chain (COL18A1)**

**Collagen α-2(I) chain (COL1A2)**

**Na/K-transporting ATPase subunit γ (FXYD2)**

**Uromodulin (UMOD)**

†Peptide identification numbers; ‡Lower case p and k indicates hydroxyproline and hydroxylysine; §Amino acid positions according to UniProt Knowledge Base numbering; ¶Regulation determined sequentially from normal controls over NAFLD and NASH (all without clinical signs of liver cirrhosis) to clinically well-documented liver cirrhosis by the Kruskal-Wallis rank sum test.

**Abbreviations:** AA, amino acid sequence; Da, Dalton; MS, mass spectrometry; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis.
Table 3.4 List of protease cleavage sites for the sequence-identified LivFib-50 peptide markers derived from the CutDB proteolytic event database based on the octamer amino acid sequence motifs around the peptides N- and C-terminal ends.

<table>
<thead>
<tr>
<th>Peptide ID</th>
<th>Peptide AA sequence &amp; cleavage site†</th>
<th>Protease Human</th>
<th>CutDB-ID or reference‡</th>
<th>Plausible mode of proteolysis§</th>
</tr>
</thead>
<tbody>
<tr>
<td>2136</td>
<td>GDRGEpGPpGPA</td>
<td>GFAGP</td>
<td>MMP9</td>
<td>18544, 18503</td>
</tr>
<tr>
<td>4419</td>
<td>ApGDRGEpGPpGPAG</td>
<td>FAGP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6546</td>
<td>MGPRG</td>
<td>PpGPpGKNGDDGEAGKP</td>
<td>CTSK</td>
<td>26955</td>
</tr>
<tr>
<td>9627</td>
<td>MGPRG</td>
<td>PPpGpGKNGDDGEAGKpGrp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11744</td>
<td>GPQG</td>
<td>LQGLpGTGGpGENGpGsGEpGPKG</td>
<td>MMP9, MMP13, MMP20</td>
<td>22880, 22857, 20126</td>
</tr>
<tr>
<td>13730</td>
<td>AGpGAPGApGAPGVPAGKSGDRGETGP</td>
<td>AGPTGPV</td>
<td>MMP25</td>
<td>(29)</td>
</tr>
<tr>
<td>13779</td>
<td>TARPWRA</td>
<td>DdILASPPRLPEPQPGAPPHSS</td>
<td>MMP9, MMP13</td>
<td>18955, 18957</td>
</tr>
<tr>
<td></td>
<td>DDdILASPPRLPEPQPGAPPHSS</td>
<td>YVHL</td>
<td>MMP3, MMP7, MMP13, MMP14, MMP20, CTS, CTSB, CTSK</td>
<td>18950, 18954, 18959, 18960, (30)</td>
</tr>
<tr>
<td>14801</td>
<td>GPAGFA</td>
<td>GPpGADGQPGAKGEpGDAGAGKGDAGpPGPAGP</td>
<td>MMP9</td>
<td>18544, 18503</td>
</tr>
<tr>
<td>16811</td>
<td>GPQGVK</td>
<td>GERGspGpGAAGpGARGIpGpPSNNGPpGp</td>
<td>MMP8, MMP9, MMP12, MMP13</td>
<td>22882, 22875, 22859, 22845</td>
</tr>
</tbody>
</table>

† peptide sequence together with the N- or C-terminal cleavage sequence motif marked in bold. The cleavage site is indicated by a vertical bar.
‡ The observed cleavage site according to the 5-digit CutDB entry number or the literature reference, as given in the manuscript.
§ The mode of proteolysis indicates if a cleavage site arise directly from cleavage of the endopeptidase or is further processed by sequential cleavage of single amino acids specified by the number of AA.

Abbreviations: AA, amino acid; CTS, cathepsin; MMP, matrix metalloproteinase.
Table 3.5 Changes in gene expression levels between cirrhotic and normal livers for the collagen chains COL1A1, COL1A2, COL3A1 and COL18A1, from which the LivFiv-50 peptide markers are derived, and for the proteases MMP 2, 3, 13, 14 and CTSB, which were predicted to be responsible for the release of the LivFib-50 collagen peptide markers.

<table>
<thead>
<tr>
<th>Protein symbol</th>
<th>Protein name</th>
<th>GSE77627</th>
<th></th>
<th>Fold change Liver cirrhosis vs Normal liver</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ID</td>
<td>FDR-adj. P</td>
<td></td>
</tr>
<tr>
<td>COL1A1</td>
<td>Collagen α-1(I) chain</td>
<td>ILMN_1701308</td>
<td>1.84E-01</td>
<td>1.61</td>
</tr>
<tr>
<td>COL1A2</td>
<td>Collagen α-2(I) chain</td>
<td>ILMN_1785272</td>
<td>4.22E-02</td>
<td>0.66</td>
</tr>
<tr>
<td>COL3A1</td>
<td>Collagen α-1(III) chain</td>
<td>ILMN_1773079</td>
<td>5.26E-06</td>
<td>0.28</td>
</tr>
<tr>
<td>COL18A1</td>
<td>Collagen α-1(XVIII) chain</td>
<td>ILMN_1806733</td>
<td>2.62E-02</td>
<td>1.49</td>
</tr>
<tr>
<td>MMP2</td>
<td>Matrix metallopeptidase 2</td>
<td>ILMN_1762106</td>
<td>2.10E-03</td>
<td>3.08</td>
</tr>
<tr>
<td>MMP3</td>
<td>Matrix metallopeptidase 3</td>
<td>ILMN_1784459</td>
<td>8.39E-01</td>
<td>0.95</td>
</tr>
<tr>
<td>MMP13</td>
<td>Matrix metallopeptidase 13</td>
<td>ILMN_1663873</td>
<td>3.66E-01</td>
<td>1.13</td>
</tr>
<tr>
<td>MMP14</td>
<td>Matrix metallopeptidase 14</td>
<td>ILMN_1774739</td>
<td>8.04E-01</td>
<td>0.95</td>
</tr>
<tr>
<td>CTSB</td>
<td>Cathepsin B</td>
<td>ILMN_1696360</td>
<td>4.31E-04</td>
<td>0.45</td>
</tr>
</tbody>
</table>
3.4 Discussion

This is the first exploratory work using CEMS to identify urinary peptides as biomarkers for liver fibrosis. The findings reported here demonstrate feasibility to identify specific peptide markers significantly associated with liver fibrosis, which, can be combined into a multi-marker panel for accurate detection of liver fibrosis, irrespective of other comorbidities. The initial hypothesis that the transition from a healthy liver state to NAFLD and further to NASH on the molecular level is characterized by fibrosis which displays in urinary collagen fragments could be verified in this exploratory work (345,346). To select specific fibrosis marker(s), the search strategy for urinary biomarkers was not restricted to a statistical group comparison of peptide distributions in patients with NAFLD and NASH compared to normal individuals, but also extended to patients with LC (both alcohol and non-alcohol aetiology). Aside from this group-wise comparison, the fibrosis marker search strategy was further adapted to peptides that showed gradual differences in their excretion profiles between normal liver states, NAFLD, NASH and LC, representing groups with none, mild, moderate to advanced stages of fibrosis. Validation of the findings was performed in a separate set of samples, further strengthening the validity of these findings.

Even after applying highly stringent criteria, CEMS-based urinary proteome analysis, enabled the definition of a set of 50 naturally occurring peptides in the molecular mass range of 0.8 to 20 kDa that showed significant differences in their excretion profiles with high confidence (due to adjustment of p-values by the Benjamini and Hochberg procedure (310) between normal subjects and patients with NAFLD, NASH or LC. Moreover, these peptides are correlated with the severity of liver injury from normal liver state over NAFLD and NASH to LC, as demonstrated by Spearman Rho correlation factors above 0.2 or below -0.2.

The amino acid sequence could be resolved from 22 of these 50 peptides. Reasons for not retrieving the amino acid sequence for all peptides include the inability to identify all possible post-translational modifications, e.g., due to high glycan heterogeneity, and resistance of the peptide to its total fragmentation even at high collision energies in tandem mass spectrometry (347,348). This is of no relevance for the diagnostic use of the peptide marker model, as indicated for example by clinical application of a combined bile and urine proteome test for cholangiocarcinoma diagnosis based on similar peptide marker models (349-350) and is based on the fact that the peptide markers are clearly defined in the CEMS as the device for operating the diagnostic test by their molecular mass and by their capillary electrophoresis migration time.

In liver fibrosis, normal hepatic tissue is replaced by connective tissue. Morphologically liver fibrosis is characterized by an excessive deposition of collagen-rich ECM components in the
liver, which is thought to be caused by the transdifferentiation of hepatic stellate cells into collagen producing cells and reduced degradation of collagen fibrils. Since urine in its low molecular weight protein composition consists to a large degree of collagen fragments, partially arising from glomerular filtration and endosomal-exosomal passage of blood circulating collagen fragments, the hypothesis was that fibrosis can be detected by the profiling of the urinary low molecular weight proteome.

A further advantage of the approach chosen is the ability to adjust for confounders based on additional data available from the CEMS database. As such, specificity of the age-adjusted LivFib-50 peptide model for liver fibrosis was evaluated. When investigating data from patients with chronic kidney disease and kidney fibrosis, the later defined as ≥ 2.5 % of fibrosis per total tissue area, 30 out of 41 (73%) of these patients scored negative for liver fibrosis. This is in line with the finding, that no impact of diabetes mellitus on LivFib-50 classification was observed in a global Kruskal-Wallis rank sum test when subdividing the LC patients of this cross-sectional cohort into those with or without diabetes mellitus.

Not unexpectedly, due to the impact of dysregulated ECM remodelling on the progression of liver fibrosis, sequencing of the peptide markers included in the LivFib-50 classification model revealed a predominant abundance of peptides derived from fibrillar collagen chains, and particularly from collagen α-1(I) (COL1A1) and collagen α-1(III) (COL3A1). Amino acid alignment of the 14 COL1A1 peptides to the linear protein sequence revealed that some of the peptides are overlapping fragments of certain areas within the triple helical COL1A1 region. In respect to the differences in peptide distributions between patients with liver disease and normal controls, even overlapping peptides demonstrate opposite regulation. The restriction of the peptide markers to certain protein regions and the non-directional type of regulation of the COL1A1 peptide markers might be explained by sites of increased proteolytical accessibility and changes in the activity of ECM degrading proteases during fibrosis progression. In this respect, the comparison of the octamer amino acid motifs surrounding the N- and C-terminal ends of the 50 fibrosis peptide markers with already reported proteolytic cleavage sites reported in the CutDB, proteolytic event database was performed and resulted in the identification of ten putative proteolytic cleavage sites, eight specific for matrix metallopeptidases and two for cathepsins (304,352,353).

The results on proteases predicted to be responsible for cleavage of the collagen markers are in line with the role of matrix metallopeptidases and cathepsins in liver fibrotic disease processes. Both are essential together with their tissue inhibitors of metalloproteinases (TIMP) counterparts in maintaining the structure of the hepatic ECM. Their normal function becomes impaired in the presence of chronic liver inflammation leading to fibrogenesis (354,355).
To provide further evidence on the findings of this work, comparison was made with tissue transcriptomics data of human cirrhotic and normal livers and found evidence for differential regulation of the targeted collagen chains and predicted proteases also in liver tissue. It must be stated, however, that the regulation of protease activity and the degradation of the ECM leading to the release of endogenous collagen peptides into circulation and then into urine is much more complex than simply explained by changes in tissue gene expression levels (308,309).

To determine the origin of the collagen peptide markers, gene expression levels of their collagen precursors in normal livers were compared to normal kidneys and found higher expression levels of COL1A1, COL3A1 and COL18A1 in the liver than in the kidney. Together with the observation that a significant part of the urinary collagen peptides is derived from extrarenal origin, this provides evidence that the identified collagen peptide markers are derived from fibrotic processes in the liver.

When the activity of the predicted proteases was assessed based on the associated peptide marker expressions in the patient’s CEMS peptide profiles (305), significance could be reached for the MMP’s 2, 3, 13 and 14, as well as CTSB. Since for all of these proteases a decrease in activity was predicted in the case of LC, this leads us to the hypothesis based on literature findings that TIMP-2 might be the inducer and/or hepatorenal connector leading to the observed changes associated with liver fibrosis in the urinary peptidome (356,357).

Recently, diagnostic tests for cardiac fibrosis were proposed on the immunological detection of COL1A1 fragments in human serum. The tests are based on assessment of the carboxy-terminal propeptide (PICP) and the carboxy-terminal telopeptide (CITP) of COL1A1. Similarly, the LITMUS consortium reported the development of the two diagnostic panels FIBC3 and ABC3D both based on the detection of plasma levels of a PRO-C3 collagen neo-epitope for diagnosis and quantification of advanced liver fibrosis in patients with NAFLD (358,359).

These recent developments underpin the significance of circulating collagen fragments in non-invasive fibrosis testing. The LivFib-50 peptide marker model reported here is based on simultaneous detection of 50 urinary peptides by mass spectrometry. Such a multiplexed analysis has the advantage over single biomarker detection and it better compensates for biological and analytical variances making the diagnostic test more robust against interferences and variances caused by a complex biological matrix. Mass spectrometry in addition allows direct detection of the analyte in the sample and thus eliminates the need for antigen binding reagents with the potential to introduce bias by antigen cross reactivity.
In addition, the MS-based approach profits from the higher selectivity: each peptide is detected based on its exact mass, while antibodies detect an epitope that may be present in multiple similar peptides, thereby severely compromising selectivity and accuracy. Due to its wide and adjustable measurement range and its independence from a sieving matrix and the requirement for continuous adaptation of electrospray conditions, CEMS combined with stringent calibration and proteomic data normalization procedures is highly adapted for the use in routine clinical practice.

Over the last years, this was demonstrated in technical reports and large-scale prospective and/or longitudinal clinical studies. Recently, the value and applicability of CEMS (300) in the early detection of chronic kidney disease in a large multicentric randomized controlled trial has been demonstrated. In this respect, the results demonstrate that with further validation in larger prospective studies, urinary peptide detection can extend to screening in the primary care for liver fibrosis, which is lacking at present. This approach also offers insights into the pathophysiology of liver fibrosis and relevant peptidases involved in the hepatic ECM remodelling as well as potential antifibrotic therapeutic interventions in liver fibrosis (300,360-363).

Collagen fragments were reported as biomarkers for a variety of other pathologies, especially for association with fibrosis in kidney disease and heart failure. To estimate a potential interference, the degree of overlap with the chronic kidney disease-specific classifier CKD273 and the heart failure-specific classifier HF1 was investigated. As a result, an overlap of six peptides (12%) was detected for CKD273. Three peptides (6%) thereof had the same direction of regulation compared to controls. In the case of HF1, there was an overlap of five peptides (10%) of which four (8%) showed the same trend. Based on these results, interference of LivFib-50 with CKD273 and HF1 is estimated to be low, this will however need further comparative cohorts’ studies (364,365).

The experimental work for the LivFib-50 was limited with the variable degrees of liver fibrosis in patients recruited, hence these markers are indicative of presence of liver fibrosis rather than the stage of it. Development of classifiers for staging requires further investigation and validation against different histological grades of liver fibrosis. Nonetheless, these markers relate to changes within an individual patient rather than being a surrogate indication of liver function.

Another limitation is that not all peptides included in the LivFib-50 model could be sequenced. In CEMS, a peptide is identified simply by its physicochemical characteristics, namely number of positive charges at pH 2 and molecular mass. This is sufficient for its definite detection in a patient sample. Nevertheless, identification of endogenous peptide markers is desirable to
understand the pathophysiological context behind their altered expression. Sequence identification is however challenging, due to unknown post-translational modifications or resistance of larger peptides to fragmentation during the MS/MS. As such, sequence information can frequently not be assigned with the high confidence level. In the case where only the 21 sequenced peptides of the LivFib-50 model were used for classification of patients from the first validation set, an AUC of 0.84 (95% CI: 0.78-0.90, p<0.0001) was received under identical SVM settings, which is statistically inferior (p = 0.014 for the difference in areas) to that of 0.94 (95% CI: 0.89-0.97, p<0.0001) received with all LivFib-50 peptide markers. Thus, also the peptides for which the sequence could not be resolved until to date should be kept in the classification model.

The findings indicate that specific biomarkers of liver fibrosis in the urine of patients with different liver diseases and different grades of liver injury/fibrosis exist. These urinary peptide markers can be combined into a multivariate model with high discriminatory potential. Further validation that qualifies the selected peptide markers for liver fibrosis and thorough evaluation of the established peptide multi-marker model from this initial exploratory experimental work appear warranted to assess their exact value in predicting and staging liver fibrosis.
Chapter 4. HCC-31 urinary peptide biomarker in hepatocellular carcinoma
Chapter 4. HCC-31 urinary peptide biomarker in hepatocellular carcinoma

This work was done to explore urinary peptides and proteases in patients with HCC. This is to answer the second part of question 1 in section 1.30. The author declares that work described herein this chapter has been published (366).

3.2.1 Introduction

HCC major risk factors is liver fibrosis. It develops due to genomic instability with alteration to the microenvironment, particularly in the scaffolding and structural proteins that allow HCC cells to grow and spread further in the body. HCC diagnosis is usually radiological with advanced CT/MRI scans with or without a liver biopsy. Noninvasive diagnosis is still lacking without suitable diagnostic or screening biomarkers (366).

3.2.2 Methods

In this work, CEMS technology was applied to investigate the low molecular weight proteome of urine from patients with HCC and non-HCC but with various liver diseases including NAFLD, NASH and LC. This is to identify peptide biomarkers specific for HCC. The CEMS steps used were mentioned in sections 2.7.1 to 2.7.6 and 2.7.8. Proteases in normal liver tissue, LC and HCC were further investigated by IHC following the steps in section 2.9. Participants were recruited between 2014 to 2019 from both UHCW and Hannover Medical School, Germany. This combined approach of CEMS and IHC is demonstrated visually in the graphical abstract shown in Figure 2.3. A follow-up period to note death outcomes was closed on 15 November 2020. Work was approved by the relevant Ethics and R&D Committees (Ref 18717, Ref 260179 & Ref 901). Work was conducted according to the World Medical Association Declaration of Helsinki, with all participants providing written informed consent.

In the discovery phase, 18 HCC cases and 51 non-HCC cases were prospectively recruited. In the validation phase, 39 HCC and 87 non-HCC cases were prospectively recruited. Diagnosis of these patients was established by a combination of liver ultrasound, laboratory markers, Fibrosis 4 index (FIB-4), CT/MRI scans, and histology. The HCC diagnosis was in line with international diagnostic criteria used in Europe. HCC patients were recruited before receiving anti-cancer treatment. Schematic flow chart showing the separate phases of this experimental exploratory work are shown in Figure 4.1. Clinical and characteristics of recruited patients are shown in Table 4.1.
Figure 4.1 Schematic flow chart for the discovery and validation phases of the HCC-31 peptide marker model for the detection of HCC. List of Abbreviations: AUROC, area under the ROC curve; CI, confidence interval; HCC, hepatocellular carcinoma; IHC, immunohistochemistry; KLK6, Kallikrein-6; LC, liver cirrhosis; MEP1A, Meprin A subunit α; NAFLD, Non-alcoholic fatty liver disease; NASH, Non-Alcoholic Steatohepatitis; NC, normal controls; ROC, receiver operating characteristics; SVM, support vector machine.
**Table 4.1** Clinical and demographic data of HCC case and non-HCC control patients included in the discovery and validation phase of the study. Parameters, demonstrating significant differences between the HCC and non-HCC groups.

<table>
<thead>
<tr>
<th>Study Phase</th>
<th>Discovery</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient Group</strong></td>
<td>HCC</td>
<td>Non-HCC</td>
</tr>
<tr>
<td>Patients/samples, n</td>
<td>18/18</td>
<td>51/51</td>
</tr>
<tr>
<td>Age, years, mean/range</td>
<td>58/28–76</td>
<td>52/18–82</td>
</tr>
<tr>
<td>Female/male, n</td>
<td>3/15</td>
<td>20/31</td>
</tr>
<tr>
<td>No. detected peptides, mean/range</td>
<td>1,753/623–2,965</td>
<td>2,329/920–4,488</td>
</tr>
<tr>
<td>Liver cirrhosis, n (%)</td>
<td>18 (100)</td>
<td>25 (49)</td>
</tr>
<tr>
<td>Diabetes mellitus, n (%)</td>
<td>3 (17)</td>
<td>13 (25)</td>
</tr>
<tr>
<td>Body mass index, mean/range</td>
<td>26.8/18.3–32.1</td>
<td>27.8/20.3–41.7</td>
</tr>
<tr>
<td>Platelet count, ×10^9/L, mean/range</td>
<td>99/26–310</td>
<td>162/25–391</td>
</tr>
<tr>
<td>a-Fetoprotein (AFP), µg/L, mean/range</td>
<td>1,821/3–22,826</td>
<td>4/1–50</td>
</tr>
<tr>
<td>Alkaline phosphatase, U/L, mean/range</td>
<td>151/60–380</td>
<td>125/45–797</td>
</tr>
<tr>
<td>Aspartate aminotransferase (AST), U/L, mean/range</td>
<td>74/24–284</td>
<td>49/13–120</td>
</tr>
<tr>
<td>Alanine aminotransferase (ALT), U/L, mean/range</td>
<td>61/14–288</td>
<td>53/10–304</td>
</tr>
<tr>
<td>AST:ALT ratio, mean/range</td>
<td>1.48/0.89–3.41</td>
<td>1.17/0.20–3.20</td>
</tr>
<tr>
<td>Approx. Ishak Fibrosis Score as per FIB-4 index, n, 0–1/2–3/4–6</td>
<td>0/1/17</td>
<td>14/12/17</td>
</tr>
<tr>
<td>Albumin, g/L, mean/range</td>
<td>32/19–48</td>
<td>40/26–51</td>
</tr>
<tr>
<td>Bilirubin, µmol/L, mean/range</td>
<td>27/9–86</td>
<td>26/3–163</td>
</tr>
<tr>
<td>ALBI stage, n, 1/2/3</td>
<td>2/11/5</td>
<td>26/15/4</td>
</tr>
</tbody>
</table>

**Liver Disease Aetiology, n**

|  | Primary HCC |  |  |  |  |  |
|  | 0 | 0 | 1 | 0 |  |  |

<p>|  | Alcoholic cirrhosis (C2) |  |  |  |  |  |
|  | 7 | 4 | 7 | 6 |  |  |</p>
<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Mann–Whitney Test</th>
<th>Fisher’s Exact or Chi-Squared Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virus-related cirrhosis (HBV/HCV/HDV)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cryptogenic/Biliary cirrhosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hereditary (Mucoviscidosis/Hemochromatosis/AATD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cholestasis (PBC/PSC/SSC/PFIC)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autoimmune hepatitis (AIH)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAFLD/NASH/NASH-LC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GI cancer (CCA/PCA) with LC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other benign liver diseases (CHP/CDL)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No liver disease</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Abbreviations: AATD, a-1 antitrypsin deficiency; CCA, cholangiocarcinoma; CDL, choledocholithiasis; CHP, chronic pancreatitis; HCC, hepatocellular carcinoma; LC, liver cirrhosis; NAFLD, non-alcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; PBC, primary biliary cholangitis; PCA, pancreatic cancer; PFIC, progressive familial intrahepatic cholestasis; PSC, primary sclerosing cholangitis; SSC, secondary sclerosing cholangitis.
3.2.3 Results

The urine samples of the HCC cohort were analysed by CEMS resulting in a list of 7259 peptides in the molecular mass range between 800 and 20,000 Dalton and with a frequency of occurrence in at least 20% of samples. A threshold of 20% was chosen to have on the one side a sufficient high parameter space and on the other side can still manage the zero-inflated data matrix of peptide signal amplitudes in the CEMS peptide profiles. The latter criterion is of particular relevance for differential analysis of single peptide distributions.

In order to identify urinary peptides with differential regulation between HCC case and normal or liver fibrotic control groups, a two-step selection approach was followed. Firstly, a group-wise comparison of 18 urine samples from HCC patients and 51 non-HCC controls (25 LC, 8 NASH w/o LC, 9 NAFLD, 9 center-matched healthy individuals) by a parametric Wilcoxon rank sum test. This resulted in the identification of 123 peptides with a p-value below 0.05 after false discovery rate (FDR) adjustment by the method of Benjamini and Hochberg. Using this extended set of markers, a search was subsequently performed for those peptide markers that showed a gradual increase or decrease in their CEMS-detected amplitude signals from normal or non-cirrhotic liver disease including NAFLD and NASH without cirrhosis over LC of different aetiology to HCC. This selection procedure resulted in a list of 31 out of the set of 123 peptides with significant Spearman Rho correlation coefficients after FDR adjustment either above 0.3 or below −0.3 defining a source of HCC progression markers.

The 31 selected peptides were combined to a support vector machine (SVM)-based peptide model. This SVM model, named HCC-31, was trained during the supervised learning phase using the 18 HCC and 51 non-HCC control patients of the discovery cohort to differentiate between HCC and non-HCC specific peptide marker patterns. Concerning SVM characteristics, HCC-31 is based on a radial basis function (RBF) kernel of C-SVC type with C = 2.2691, g = 0.0764 and eps = 0.001 as fixed kernel parameter settings. Selection of an RBF kernel of C-SVC was found to be the best option for data matrices with frequent occurrence of zero intensity values as represented by CEMS peptide profiles. After optimization of the SVM parameters and total cross-validation on the original training data, the peptide marker pattern resulted in an AUC of 0.92 (95% confidence interval (CI): 0.87 to 0.96, p < 0.0001) in receiver operating characteristics analysis (ROC). The optimal threshold for an HCC positive test result was determined based on the Youden index to be −0.25 resulting in a sensitivity of 86.8% (95% CI: 74.7–94.5) and a specificity of 89.0% (95% CI: 81.2–94.4).

In order to determine the model’s accuracy without overfitting bias, the HCC-31 model was subsequently tested on an independent cross-sectional cohort of patients of whom 39 had a clinical diagnosis for HCC and 87 for other liver diseases. As presented in Figure 4.2,
independent validation of the HCC-31 model resulted in an AUC of 0.88 (95% CI: 0.81–0.93, \( p < 0.0001 \)), and 79.5% sensitivity and 85.1% specificity at the predetermined threshold at −0.25. When classification by the HCC-31 model was adjusted for age and gender of the patients in the validation cohort, the AUC in ROC analysis was significantly increased from 0.88 to 0.94 \( (p = 0.008) \). Based on the age- and gender-matched HCC-31 model, only two out of the 39 HCC cases were missed, as they were classified as controls (false negatives).

Subsequently, the prognostic value of a positive HCC-31 test was investigated, by assessing whether the classification result is a significant predictor of overall mortality during a follow-up period of 500 days starting from the date of sample collection. As revealed by the Kaplan–Meier survival curves in Figure 4.3, patients with a positive HCC-31 test had a 4.1-fold increased risk of death (95% CI: 1.7–9.8, \( p = 0.0005 \)) compared to patients with a negative test during the 500-days follow-up.

For the 31 peptides that were identified as differentially excreted in the urine between HCC cases and disease matched controls, amino acid sequences were assigned based on mapping of the CEMS characteristics (CE migration time and MS-detected molecular mass) to the urinary peptide sequence database. The CEMS characteristics for the 31 peptides together and the sequence information for all sequence identified peptides \( (n = 27) \) are presented in Table 4.2.

Following the hypothesis that peptides emerge from proteolytical processing of proteins and that peptides serve as substrates of disease-specific changes to the proteolytic environment, in silico mapping was performed on the 27 sequence identified HCC peptide marker of the HCC-31 model. In total, 18 protease candidates were found to be associated with the sequence motifs at the N- and C-terminal ends of the 27 peptides. Out of these, seven showed significant differences in the ion signal intensities of their mapped peptide substrates between the HCC case and other liver disease control groups after adjustment for multiple testing. As presented in Table 4.3, kallikrein-6 (KLK6), the matrix metallopeptidase (MMP) 3 and 13 and the cathepsins (CTS) D and E were predicted to be significantly increased, whereas meprin A subunit α (MEP1A) and CTSB were found to be decreased in their activities in HCC compared to non-HCC liver diseases \( (p < 0.05) \) in the Mann–Whitney U test.

The proteases KLK6 and MEP1A were selected for immunohistochemical (IHC) staining of liver tissue sections since they showed the highest difference in activity between HCC cases and controls in Table 4.3. For the investigation of KLK6 and MEP1A tissue expression, liver biopsy sections from five HCC, four benign liver disease including cirrhosis and five cases with normal liver tissue without disease were selected from the Arden tissue bank. As shown in Figure 4.4 incremental gradient staining ranging from mild staining in normal liver tissue, to
moderate staining in liver cirrhosis and then to strong diffuse staining in HCC was observed for KLK6. For MEP1A, there was an absence of staining in cirrhosis and HCC, whereas it was mildly present in normal liver tissue (Figures 4.5). The IHC suggests that KLK6 increase with cirrhosis and HCC while MEP1A decrease in cirrhosis and HCC. For gradient staining the Allred score was used (315), and this is demonstrated in all tested histopathological sections in Table 4.4.

The 31-HCC model consists of peptides derived from different protein sources including cell-derived and structural proteins. As revealed by a literature review, several of the HCC-31 peptide markers were also identified in other human body fluids, like serum, plasma, cerebrospinal fluid, or as HLA-associated immunopeptides in tissue and are therefore proven not to be restricted to urine (for details, see Table 4.5).
Figure 4.2 ROC characteristics of the proteomic HCC classification model HCC-31 on the validation set of 39 HCC cases and 87 non-HCC liver diseased controls.

<table>
<thead>
<tr>
<th>ROC characteristics HCC-31</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCC case / other liver diseased controls</td>
</tr>
<tr>
<td>Area under curve (AUC)</td>
</tr>
<tr>
<td>95% Confidence Interval (CI)</td>
</tr>
<tr>
<td>Significance P</td>
</tr>
<tr>
<td>Sensitivity / Specificity [%] (at -0.25 cut-off)</td>
</tr>
</tbody>
</table>
Figure 4.3 Overall survival of patients with a positive or negative HCC-31 test result during a follow-up of 500 days starting from the date of urine sample collections. Patients lost to follow-up were censored.

<table>
<thead>
<tr>
<th>Time after sample collection</th>
<th>Survival probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>90</td>
</tr>
<tr>
<td>200</td>
<td>80</td>
</tr>
<tr>
<td>300</td>
<td>70</td>
</tr>
<tr>
<td>400</td>
<td>60</td>
</tr>
<tr>
<td>500</td>
<td>50</td>
</tr>
</tbody>
</table>

**HCC-31 negative**

<table>
<thead>
<tr>
<th>Time after sample collection</th>
<th>Number at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>82</td>
</tr>
<tr>
<td>100</td>
<td>73</td>
</tr>
<tr>
<td>200</td>
<td>67</td>
</tr>
<tr>
<td>300</td>
<td>65</td>
</tr>
<tr>
<td>400</td>
<td>64</td>
</tr>
<tr>
<td>500</td>
<td>0</td>
</tr>
</tbody>
</table>

**HCC-31 positive**

<table>
<thead>
<tr>
<th>Time after sample collection</th>
<th>Number at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>44</td>
</tr>
<tr>
<td>100</td>
<td>35</td>
</tr>
<tr>
<td>200</td>
<td>31</td>
</tr>
<tr>
<td>300</td>
<td>26</td>
</tr>
<tr>
<td>400</td>
<td>24</td>
</tr>
<tr>
<td>500</td>
<td>0</td>
</tr>
</tbody>
</table>

HR: 4.05
95% CI: 1.7-9.8
p = 0.0005
Table 4.2 Characterization of the 31 urinary HCC peptide markers by amino acid sequencing and in silico protease prediction analysis together with their experimental spectrometry mass, retention time in capillary electrophoresis, and location in the proteins linear sequence.

<table>
<thead>
<tr>
<th>Peptide ID</th>
<th>Exp. Mass [Da]</th>
<th>CE Time [min]</th>
<th>Protein</th>
<th>AA</th>
<th>Proteases for N-Terminal Cleavage</th>
<th>Peptide Sequence (Black) with Flanking Regions (Grey)</th>
<th>Proteases for C-Terminal Cleavage</th>
</tr>
</thead>
<tbody>
<tr>
<td>54</td>
<td>807.39</td>
<td>23.2</td>
<td>CLU</td>
<td>366–371</td>
<td>---</td>
<td>LNEQ</td>
<td>FNWVSR</td>
</tr>
<tr>
<td>1059</td>
<td>920.34</td>
<td>21.2</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1160</td>
<td>928.51</td>
<td>24.6</td>
<td>UMOD</td>
<td>592–599</td>
<td>MEP1A, MMP3</td>
<td>RSGS</td>
<td>VIDQSRV</td>
</tr>
<tr>
<td>1778</td>
<td>981.49</td>
<td>24.4</td>
<td>ACTB</td>
<td>107–115</td>
<td>MEP1A, MMP[3,13], CTS[B,D,E]</td>
<td>HPVL</td>
<td>LTEAPLNPK</td>
</tr>
<tr>
<td>2314</td>
<td>1,032.45</td>
<td>25.9</td>
<td>ACTB</td>
<td>95–103</td>
<td>---</td>
<td>YNEL</td>
<td>RVAPEEH</td>
</tr>
<tr>
<td>3559</td>
<td>1,134.59</td>
<td>24.0</td>
<td>COL3A1</td>
<td>755–766</td>
<td>---</td>
<td>ADGV</td>
<td>PGKDPRG</td>
</tr>
<tr>
<td>3662</td>
<td>1,142.55</td>
<td>22.0</td>
<td>ADGRF3</td>
<td>56–64</td>
<td>---</td>
<td>DKAW</td>
<td>NERIDRP</td>
</tr>
<tr>
<td>4564</td>
<td>1,199.55</td>
<td>20.8</td>
<td>TKT</td>
<td>343–352</td>
<td>---</td>
<td>DGDT</td>
<td>KNSTF</td>
</tr>
<tr>
<td>4610</td>
<td>1,201.53</td>
<td>24.9</td>
<td>GAGE12H</td>
<td>58–70</td>
<td>MEP1A, MMP3, CTSB</td>
<td>AAAQ</td>
<td>KGEDEGASAGQGP</td>
</tr>
<tr>
<td>4829</td>
<td>1,217.48</td>
<td>27.7</td>
<td>COL3A1</td>
<td>179–191</td>
<td>---</td>
<td>PAGP</td>
<td>pGPpGpGp</td>
</tr>
<tr>
<td>5284</td>
<td>1,250.64</td>
<td>20.6</td>
<td>HBB</td>
<td>136–147</td>
<td>---</td>
<td>QKVV</td>
<td>AGVANALA</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>p-value</td>
<td>Fold Change</td>
<td>Description</td>
<td>Protein Sequence</td>
<td>Protein Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>---------</td>
<td>-------------</td>
<td>-------------</td>
<td>------------------</td>
<td>---------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6601</td>
<td>1,352.78</td>
<td>22.0</td>
<td>AHNAK</td>
<td>EVDVNLPKADVDISGPKIDVT</td>
<td>MEP1A, KLK6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6607</td>
<td>1,353.53</td>
<td>23.8</td>
<td>FGA</td>
<td>YKMADEAGSEADHEGHTSTKR</td>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8720</td>
<td>1,513.62</td>
<td>29.5</td>
<td>CDH1</td>
<td>ITTLKVTDAANPNTAPEAVYT</td>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9080</td>
<td>1,539.74</td>
<td>40.4</td>
<td>COL18A1</td>
<td>EGROGPpGPpGPpPGPPSFPGP</td>
<td>HRQT</td>
<td>MMP3</td>
<td></td>
</tr>
<tr>
<td>9510</td>
<td>1,576.68</td>
<td>44.9</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9728</td>
<td>1,594.77</td>
<td>40.3</td>
<td>COL1A1</td>
<td>DAGPPGPpGPpGPpGPPSA</td>
<td>GFDF</td>
<td>MMP[3,13], CTSB</td>
<td></td>
</tr>
<tr>
<td>10177</td>
<td>1,624.73</td>
<td>25.1</td>
<td>COL2A1</td>
<td>GPSQDQGASGpAgpSpRGpGP</td>
<td>PVGP</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>11725</td>
<td>1,733.73</td>
<td>29.8</td>
<td>GSN</td>
<td>AAYLWVTGASEAEKTGAQELLRVL</td>
<td>MEP1A, MMP3, CTS[D,E]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12459</td>
<td>1,782.85</td>
<td>26.0</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13134</td>
<td>1,836.79</td>
<td>31.1</td>
<td>COL1A2</td>
<td>SPGVNGApGEAGRDGNPGNDGPPp</td>
<td>RDGQ</td>
<td>MEP1A, MMP [3,13], CTSB</td>
<td></td>
</tr>
<tr>
<td>13176</td>
<td>1,840.83</td>
<td>41.9</td>
<td>COL2A1</td>
<td>PRGRSGETGPAGppGNPGPPGPpGP</td>
<td>PGPG</td>
<td>KLK6</td>
<td></td>
</tr>
<tr>
<td>14389</td>
<td>1,930.89</td>
<td>31.6</td>
<td>COL3A1</td>
<td>TGPQPpGPTGPpGDGGDGTGpGPpGQQ</td>
<td>LQGL</td>
<td>MEP1A, MMP [3,13], CTSB</td>
<td></td>
</tr>
<tr>
<td>14925</td>
<td>1,972.96</td>
<td>25.0</td>
<td>PCSK1N</td>
<td>RRAADHDVGGSELPPPEGVLGALLR</td>
<td>VKRL</td>
<td>MEP1A, MMP [3,13]</td>
<td></td>
</tr>
<tr>
<td>PRT</td>
<td>Exp. mass</td>
<td>PRT</td>
<td>PRT</td>
<td>PEPTIDE</td>
<td>PROTEIN</td>
<td>EXP. mass</td>
<td>PRT</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>-------</td>
<td>-------</td>
<td>--------------------------</td>
<td>----------</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td>15342</td>
<td>2,009.88</td>
<td>32.4</td>
<td>COL5A2</td>
<td>136–157</td>
<td>MEP1A, MMP [3,13], CTSB</td>
<td>GAPG</td>
<td>SKGEAGpTGPMGDpGTVP</td>
</tr>
<tr>
<td>17066</td>
<td>2,169.98</td>
<td>33.7</td>
<td>COL16A1</td>
<td>1,145–1,166</td>
<td>MEP1A, MMP [3,13], CTSB</td>
<td>GPQG</td>
<td>NGEKGDGQFGQFGQFGpGPPGP</td>
</tr>
<tr>
<td>17805</td>
<td>2,232.00</td>
<td>33.6</td>
<td>COL5A1</td>
<td>999–1,021</td>
<td>---</td>
<td>PPQVG</td>
<td>GQpGpTGETGpMEP1A, CTSB</td>
</tr>
<tr>
<td>19681</td>
<td>2,380.08</td>
<td>35.8</td>
<td>COL3A1</td>
<td>201–228</td>
<td>MEP1A, CTSB</td>
<td>PGYQ</td>
<td>GPPGEPQAGpSGpPGppGAI</td>
</tr>
<tr>
<td>20237</td>
<td>2,430.60</td>
<td>35.7</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>GETG</td>
<td>PQPpGPTGpGDpGDPpGp</td>
</tr>
<tr>
<td>24328</td>
<td>2,854.37</td>
<td>34.6</td>
<td>COL3A1</td>
<td>616–646</td>
<td>CTSB</td>
<td>VTTV</td>
<td>ASHTDSDVPSGVEVVKLF</td>
</tr>
<tr>
<td>29919</td>
<td>3,524.75</td>
<td>32.4</td>
<td>CLU</td>
<td>390–423</td>
<td>---</td>
<td>VTTV</td>
<td>ASHTDSDVPSGVEVVKLF</td>
</tr>
</tbody>
</table>

ACTB, Actin, cytoplasmic; ADGRF3, Adhesion G-protein coupled receptor F3; AHNAK, Neuroblast differentiation-associated protein; CDH1, Cadherin-1; CLU, Clusterin; COL1A1, Collagen α-1(I) chain; COL1A2, Collagen α-2(I) chain; COL2A1, Collagen α-1(II) chain; COL3A1, Collagen α-1(III) chain; COL5A1, Collagen α-1(V) chain COL5A2, Collagen α-2(V) chain; COL16A1, Collagen α-1(XVI) chain; COL18A1, Collagen α-1(XVIII) chain; CTS, Cathepsin; FGA, Fibrinogen α chain; GAGE12H, G antigen 12H; GSN, Gelsolin; HBB, Haemoglobin subunit β; KLK6, Kallikrein-6; MEP1A, Meprin A subunit α; MMP, Matrix metallopeptidase; PCSK1N, ProSAAS; TKT, Transketolase; UMOD, Uromodulin.

Abbreviation: AA, amino acid sequence; CE, capillary electrophoresis; Exp. mass, experimental mass. † Peptide identification number. ‡ Lower case p expresses hydroxyproline.
Table 4.3. Differences in the activities of the seven in silico predicted proteases meprin A subunit α (MEP1A), matrix metallopeptidase (MMP) 3 and 13, kallikrein-6 (KLK6) and cathepsin (CTS) B, D and E based on the fold change of the protease associated peptide substrate’s ion signals between the HCC case and non-HCC liver disease control groups. *p*-values were calculated by the Mann–Whitney U test.

<table>
<thead>
<tr>
<th>Protease</th>
<th>Peptide Substrate Distribution [Avg. Ion Counts ± SD]</th>
<th>Fold-Change Case/Control</th>
<th><em>p</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HCC Case Group (n = 18)</td>
<td>Non-HCC Liver Disease Control Group (n = 51)</td>
<td></td>
</tr>
<tr>
<td>MEP1A</td>
<td>196.23 ± 93.19</td>
<td>365.65 ± 231.12</td>
<td>0.54</td>
</tr>
<tr>
<td>MMP3</td>
<td>632.99 ± 317.56</td>
<td>393.63 ± 331.29</td>
<td>1.60</td>
</tr>
<tr>
<td>MMP13</td>
<td>729.36 ± 402.76</td>
<td>495.03 ± 539.78</td>
<td>1.47</td>
</tr>
<tr>
<td>KLK6</td>
<td>166.40 ± 79.87</td>
<td>67.09 ± 88.51</td>
<td>2.47</td>
</tr>
<tr>
<td>CTSB</td>
<td>347.63 ± 173.67</td>
<td>643.77 ± 399.10</td>
<td>0.53</td>
</tr>
<tr>
<td>CTSD</td>
<td>32.90 ± 33.44</td>
<td>17.35 ± 31.21</td>
<td>1.89</td>
</tr>
<tr>
<td>CTSE</td>
<td>34.41 ± 31.99</td>
<td>22.56 ± 41.25</td>
<td>1.52</td>
</tr>
</tbody>
</table>
**Figure 4.4** The top row shows IHC characteristics of kallikrein-6 (KLK6) at low power magnification (X10) in HCC, liver cirrhosis and normal liver tissue sections. The brown chromogen indicates protein expression of KLK6 in the nucleus and cytoplasm of the hepatic and stromal cells. The HCC (left) demonstrated high intensity of nuclear and cytoplasmic staining for KLK 6 (Allred score 3). The liver cirrhosis (middle) demonstrated moderate intensity of nuclear and cytoplasmic staining for KLK 6 (Allred score 2). The normal liver right demonstrated mild intensity of nuclear and cytoplasmic staining for KLK6 (Allred score 1). The bottom row shows IHC characteristics of kallikrein-6 (KLK6) at high power magnification highlighting cellular features of interest. Left slide at X40 showing irregular and infiltrative cords and nests of atypical hepatocytes with moderate amount of cytoplasm and pleomorphic nuclei. Here KLK6 showing both high nuclear and cytoplasmic staining in the tumour cells (blue arrow). There is associated desmoplastic reaction in the stroma (yellow arrow). Middle slide at X20 showing nodules of regenerating hepatocytes with moderate KLK6 nuclear and cytoplasmic staining (blue arrow) and expanded portal tract by dense fibrosis (yellow arrow). Right slide at X40 showing mild nuclear and cytoplasmic staining of KLK6.
Figure 4.5 The top row shows IHC characteristics of meprin A subunit a (MEP1A) at low power magnification (X10) in HCC, liver cirrhosis and normal liver tissue sections. The brown chromogen indicates protein expression of MEP1A in the nucleus and cytoplasm of the hepatic and stromal cells. The HCC (left) and liver cirrhosis (middle) demonstrated negative nuclear and cytoplasmic staining for MEP1A (Allred score 0). The normal liver tissue (right) demonstrated mild intensity of nuclear and cytoplasmic staining for MEP1A (Allred score 1). The middle row shows IHC characteristics of MEP1A at high power magnification (X40). Left slide at X40 showing HCC comprising solid sheets of atypical hepatocytes with mild nuclear pleomorphism, loss of portal tracts within the tumour and complete loss of MEP1A staining. Middle slide at X20 showing nodules of regenerating hepatocytes separated by fibrous septum (yellow arrow) and complete loss of MEP1A staining. Right slide at X40 showing normal liver parenchyma with mild nuclear and cytoplasmic staining of MEP1A in hepatocytes. The bottom row demonstrates a positive control of MEP1A in enterocytes of the villi of small intestine showing mild nuclear and cytoplasmic staining. MEP1A is normally expressed in small intestine (314).
Table 4.4 Allred Scoring (315) for all retrieved tissue specimens.

<table>
<thead>
<tr>
<th>Tissue Type</th>
<th>Allred IHC Score</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KLK6</td>
<td>MEP1A</td>
<td></td>
</tr>
<tr>
<td>HCC 1</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>HCC 2</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>HCC 3</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>HCC 4</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>HCC 5</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Cirrhosis 1</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Cirrhosis 2</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Cirrhosis 3</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Cirrhosis 4</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Normal liver 1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Normal liver 2</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Normal liver 3</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Normal liver 4</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Normal liver 5</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.5 List of HCC-31 peptide markers that were identified in human tissue or body fluids other than urine.

<table>
<thead>
<tr>
<th>HCC-31 Peptide Marker ID</th>
<th>Amino Acid Sequence</th>
<th>Protein Name</th>
<th>Protein Symbol</th>
<th>Biological Source of Identification (Other Than Urine)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>11354</td>
<td>107-LTEAPLNPK-115</td>
<td>Actin; α skeletal muscle</td>
<td>ACTA1</td>
<td>Cerebellum tissue‡</td>
<td>Marcu et al. (367)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HCC tumor tissue‡</td>
<td>Dong et al. (368)</td>
</tr>
<tr>
<td>14071</td>
<td>95-RVAPEEHPV-103</td>
<td>Actin, cytoplasmic 1</td>
<td>ACTB</td>
<td>Lung tissue‡</td>
<td>Marcu et al. (367)</td>
</tr>
<tr>
<td>25411</td>
<td>179-PGPPGPPTSGHP-191</td>
<td>Collagen α-1(III) chain</td>
<td>COL3A1</td>
<td>Plasma</td>
<td>Zakharova et al. (369)</td>
</tr>
<tr>
<td>33901</td>
<td>605-DEAGSEADHEGTH-617</td>
<td>Fibrinogen α chain</td>
<td>FGA</td>
<td>Serum</td>
<td>Ueda et al. (370)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Plasma</td>
<td>Koomen et al. (371)</td>
</tr>
<tr>
<td>135817</td>
<td>390-ASHTSDSDVPSGVTEV VVKLFDSDPITVTPVEV-423</td>
<td>Clusterin</td>
<td>CLU</td>
<td>Cerebrospinal fluid</td>
<td>Belogurov et al. (372)</td>
</tr>
<tr>
<td>57312</td>
<td>605-WVGTGASEAEK TQAQEL-621</td>
<td>Gelsolin</td>
<td>GSN</td>
<td>Plasma</td>
<td>Modzdiak et al. (373)</td>
</tr>
</tbody>
</table>

‡ Identified as HLA-associated immunopeptide.
3.2.4 Discussion

The CEMS experimental work herein this chapter focused on the identification of HCC-specific peptides in urine to test their diagnostic utility by integration into a peptide marker model and to trace back these peptides to systemic alterations of HCC pathophysiological processes at the tumor site.

The HCC-31 classifier adds to the current modalities for non- or minimal-invasive HCC diagnosis. To put this in a clinical perspective, the HCC-31 performed better in comparison to AFP. HCC-31 showed sensitivity of 79.5% while the quoted literature showed that AFP usually has low sensitivity for HCC detection between 40–65% (184). Therefore, potential use of HCC-31 is promising if further validated as substitute to AFP in aiding HCC diagnosis or as a prognostic marker.

HCC-31 utilizes a molecular pattern of 31 peptides, which are surrogate markers for differential proteolytic activity at the HCC tumor site in comparison to other cirrhotic and non-cirrhotic liver diseases. Validation of HCC-31 on an independent cross-sectional cohort of 39 HCC and 87 highly heterogeneous non-HCC liver disease patients from two clinical populations, one in England and the other in Germany, resulted in an accuracy of 83.3% of the pure classification model and 91.3% when the model was adjusted for gender and age. Moreover, HCC-31 positivity was associated with a 4-fold increased risk of death during a 500-day observational period providing further evidence for its clinical applicability.

Some of the proteins from which the peptide markers are derived are already described in the context of HCC, such as CDH1/E-cadherin and AHNAK (374,375). Peptides as disease markers add another level of complexity since their expression differences might not be caused by differential regulation of their parent protein but changes in the activity of the proteases leading to their generation. Therefore, the common features qualifying the 31 peptides as HCC markers are most likely their cleavage by proteases with changes in activity during the course of HCC and the same route of clearance by incorporation into exosomes, release into the circulation and final excretion into the urine. Most of the peptides included in HCC-31 are fragments of collagen chains, which were also identified as source of peptide markers for other diseases (376,377).

Collagen chains are the main components of the extracellular matrix, and their fragments are predominant in the low molecular weight fraction of the urinary proteome. Various proteases are able to cleave collagen chains, most prominent are matrix metalloproteinases and cathepsins (378). In this respect, more than 600 different partially overlapping peptides derived from collagen α-1(I) chain in urine were detected. As already described in other studies, the
composition of collagen peptides in urine is strongly associated with changes in specific protease activities at the site of disease, particularly in progressing tumours and surrounding microenvironment (350,379).

The exact mechanisms leading to carcinogenesis are yet to be identified. However, cancer cells’ metabolism involves extracellular proteolytic degradation. This plays a role in cell migration, tumour growth and distant spreading in the body (380). Therefore, investigations at the protein level (proteomics) are advantageous particularly in the case of in-depth characterization of cancer progression and invasiveness. CEMS has demonstrated in this context a good diagnostic potential of urinary peptide biomarkers even for non-renal diseases with exosomes as the potential trans-renal carriers. These biomarkers have been identified in the context of a single type of cancer (e.g., bladder, prostate, pancreatic, renal cell carcinoma and cholangiocarcinoma) (350,381-383).

The results herein this chapter have demonstrated that proteolytically processed peptides in the urine can be used in diagnosis and prognosis of HCC, and this is actually a promising non-invasive tool for precision medicine in the future. This CEMS exploratory experimental work also demonstrated that these urinary peptides are related to proteolytic activities at the tumour site.

The predicted proteases in various stages of liver tissue ranging from normal to LC and HCC were demonstrated to identify firstly if these proteases are present and if their staining differs between normal and disease groups, namely LC and HCC. KLK6 and MEP1A were shortlisted as per the lowest p-value. The gradient staining confirmed the predicted activity, showing that KLK6 increases with cirrhosis and HCC, and MEP1A decreases in cirrhosis and HCC.

KLK6 is a protease that belongs to the kallikrein family of fifteen members located on chromosome 19. KLK6 was shown to be involved in many cancers’ formation and progression (384-387). In the liver, KLK6 was shown to catalyze ubiquitin, an important cellular regulatory protein involved in protein synthesis. KLK6 also was shown to induce de novo cirrhosis and was increased in HCC tissues (388). Additionally, a study designed to check the activity of KLK6 on ECM peptides in HCC revealed that KLK6 has an upregulated activity (389).

MEP1A is a metalloproteinase that belongs to the metzincin family with the main function in intracellular transport of proteins (390). MEP1A has been implicated in kidney, colorectal and pancreatic cancers (381,391,392). In HCC, MEP1A was shown to promote cell proliferation, migration and invasion (393,394). In the present work, MEP1A related peptides in the urine are present in HCC at decreased levels than in cirrhosis and non-cirrhosis control group. However, both the staining in cirrhosis and HCC tissues were negative but present in normal livers. This was also noted by OuYang et al (394) on HCC tissues, where
immunohistochemical MEP1A expression levels in the tumour cell cytoplasm varied widely among different HCC specimens. However, the same group showed that MEP1A was found to be elevated following analysis of the HCC tissues using quantitative real-time polymerase chain reaction compared with matched adjacent nonneoplastic tissues and non-malignant liver disease tissues. Differential regulation in this respect might occur on the protein level, e.g., by secretion of soluble MEP1A, rather than forming a membrane-bound complex within the cell or on the cell surface (395). In addition, the presence of MEP1A in HCC tissues also demonstrated poor prognosis (394).

The predicted seven proteases in this CEMS exploratory experimental work could also be potential sites for anti-protease treatment in HCC. An example was demonstrated in a study by Tran et al (396). They showed that injection of metalloproteinases (MMPs) inhibitors to HCC cell lines resulted in delaying HCC growth without treatment related toxicity. MMP inhibitors also lead to inhibition of angiogenesis and tumour necrosis. Furthermore, anti-cathepsins were found to promote cell death in a study completed on HepG2 cell lines (397). These anti-proteases could be used through an immunotherapy approach in combination with conventional chemotherapy and/or nanoparticle based intervention.

This work herein in this chapter was limited by the small number of patients, small number of human liver tissue samples and its exploratory nature; nonetheless, it was multicentre and validated across two populations. In addition, the presentation of the predicted proteases was verified at the tissue level demonstrating that these urinary peptides are related to the HCC disease formation in the liver.

In conclusion, CEMS technology has identified an important sequence of urinary peptides related to proteolytic activity in HCC. The technology paves the way for future work on these peptides to develop a noninvasive test that could be applied early for purpose of screening, prognosis, surveillance and/or diagnosis.
Chapter 5. Exploration of patterns and composition of the urinary volatile organic compounds (VOCs) in liver fibrosis and hepatocellular carcinoma
Chapter 5. Exploration of patterns and composition of the urinary volatile organic compounds (VOCs) in liver fibrosis and hepatocellular carcinoma

This work was done to explore urinary VOCs in patients with liver fibrosis and HCC. This is to answer question 2 in section 1.30. The aim here was to identify the urine for pattern and chemical composition of VOCs. The author declares that work described herein this chapter has been published. (398,399)

5.1 Introduction

VOCs have a role in the pathogenesis of liver disease. In the liver, VOCs could be derived from metabolic pathways like the ROS and CYP450. They can also be derived from the gut microbiome metabolic by-products reaching the liver via the portal circulation. Following metabolic processing in the liver they reach the kidney via the systemic circulation with subsequent filtration and excretion in the urine. These VOCs have the ability to emit from solid or liquid forms into gases. Further methodologies were applied to identify these VOCs in the headspace of the urine samples of HCC and liver fibrosis patients.

5.2 Methods

Three methods were applied here. SPME, GC-IMS and GC-TOF-MS. This is to evaluate the hypothesis that liver fibrosis and HCC urinary VOCs could be detected in the urine and to also test if there are any differences between these methodologies.

SPME analysis was completed following steps explained in section 2.11, 2.11.1 and 2.11.2. GC-IMS and GC-TOF-MS analyses were completed following the steps explained in 2.12, 2.13 and 2.14.

This exploratory work was ethically approved by the Coventry and Warwickshire and Northeast Yorkshire NHS Ethics Committees (Ref 18717 and Ref 260179). The exploratory work conformed to the ethical principles of the Declaration of Helsinki. Participants were recruited from UHCW. All participants provided written informed consent. Five mL urine samples were collected into universal bottles from each participant. These samples were then immediately frozen at -80C. The samples were then stored until further sample analysis at the end of the recruitment process. Work done previously assessed the stability of urine samples in storage, and all methods used here were in line with this.

A prospective recruitment was completed between 2018 and 2019. HCC and other cancer patients were recruited before receiving any form of cancer treatment for example chemo/radiotherapy or surgery. Patients who are less than 18 years old or pregnant were excluded.
For SPME work, other cancer patients (bladder and prostate cancers) were additionally recruited, as disease group with a cancer not related to the liver. This choice was made as these cancers originate in the urinary tract and are in contact with urine and so would form basic comparison if VOCs can differentiate HCC from urinary tract cancers and other healthy participants without disease.

For SPME work, there were 31 HCC cases, 62 prostate carcinoma cases (PCA), 29 urinary bladder carcinoma cases (BCA) and 18 non cancer cases. PCA and BCA represented other cancers controls. Male to female ratio was 5:1 and mean age was 72 years (range 42–94). All cancer patients were diagnosed by conventional clinical methods, including various imaging techniques, to locate the tumour and/or biopsy for final diagnosis. Non-cancer controls included patients that were suspected of cancer but had negative investigations. Table 5.1 shows the characteristics of patients involved in the SPME work.

Table 5.1 Characteristics of recruited participants at time of obtaining the urine samples for the SPME work

<table>
<thead>
<tr>
<th>Group</th>
<th>Gender</th>
<th>Mean age (range)</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCC (n=31)</td>
<td>F: 5 M: 26</td>
<td>71 years (46 to 86)</td>
<td>7 non-cirrhotic&lt;br&gt;24 cirrhotic&lt;br&gt;14 &lt;10 ku/L (normal AFP)&lt;br&gt;17 &gt;10 ku/L (raised AFP)&lt;br&gt;12 cases confirmed by histology and 19 by CT/MRI</td>
</tr>
<tr>
<td>PCA (n=62)</td>
<td>M: 62</td>
<td>71.2 years (51 to 88)</td>
<td>62 confirmed by histology&lt;br&gt;Mean PSA 21.3 (range 4.4-153)&lt;br&gt;All cases had raised PSA &gt; 4 ng/mL&lt;br&gt;PSA: prostate specific antigen</td>
</tr>
<tr>
<td>BCA (n=29)</td>
<td>F: 5 M: 24</td>
<td>74.2 years (50 to 92)</td>
<td>29 confirmed by histology</td>
</tr>
<tr>
<td>Controls (n=18)</td>
<td>F: 7 M: 11</td>
<td>70.3 years (52 to 86)</td>
<td>All patients were suspected to have cancer and referred for investigations of abdominal symptoms and all patients had normal CT scan and colonoscopy</td>
</tr>
</tbody>
</table>

HCC: hepatocellular carcinoma, F: Female, M: male, PCA: prostate cancer, PSA: prostate specific antigen, BCA: bladder cancer

For GC-IMS and GC-TOF-MS work, there were a total of 58 participants. These included 20 HCC cases and 38 non-HCC cases. The non-HCC cases were recruited from two sources to decrease bias. The first source was healthy individuals without liver disease. The second source was patients with various stages of NAFLD, the advantage here is that these patients represent those at risk of becoming HCC cases in the future. The non-HCC cases were then further divided into 31 non-fibrotic and 7 fibrotic/cirrhotic cases. The characteristics of participants in the GC-IMS and GC-TOF-MS work are shown in Table 5.2.
Table 5.2 Clinical and biochemical characteristics of recruited participants at time of obtaining the urine samples for the GC-IMS and GC-TOF-MS work

<table>
<thead>
<tr>
<th>Covariate</th>
<th>HCC cases</th>
<th>Non-HCC cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Patients</td>
<td>20</td>
<td>38</td>
</tr>
<tr>
<td>Age, mean (range)</td>
<td>73 (53-84)</td>
<td>58.08 (29-89)</td>
</tr>
<tr>
<td>Gender Female/male</td>
<td>2/18</td>
<td>11/27</td>
</tr>
<tr>
<td>Cause of liver disease</td>
<td>3 Alcohol</td>
<td>1 HBV cirrhosis</td>
</tr>
<tr>
<td></td>
<td>1 HBV</td>
<td>9 NAFLD</td>
</tr>
<tr>
<td></td>
<td>1 HCV</td>
<td>10 NASH</td>
</tr>
<tr>
<td></td>
<td>13 NASH</td>
<td>6 NASH cirrhosis</td>
</tr>
<tr>
<td></td>
<td>2 primary/idiopathic</td>
<td>12 without liver disease</td>
</tr>
<tr>
<td>Histological/Radiological Features</td>
<td>Liver cirrhosis</td>
<td></td>
</tr>
<tr>
<td>of</td>
<td>Present/absent</td>
<td>16/4</td>
</tr>
<tr>
<td>Liver cirrhosis</td>
<td></td>
<td>7/31</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Present/absent</td>
<td>11/9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7/31</td>
</tr>
<tr>
<td>AFP mean (range), KU/L</td>
<td>1380.60 (1-9400)</td>
<td>---</td>
</tr>
<tr>
<td>ALT mean (range), U/L</td>
<td>44.60 (13-149)</td>
<td>50.74 (5-304)</td>
</tr>
<tr>
<td>ALP mean (range), U/L</td>
<td>150.90 (83-326)</td>
<td>89.76 (53-279)</td>
</tr>
<tr>
<td>Albumin mean (range), g/L</td>
<td>39 (24-44)</td>
<td>43.87 (28-50)</td>
</tr>
<tr>
<td>Bilirubin mean (range), µmol/L</td>
<td>24.30 (5-84)</td>
<td>7.97 (5-21)</td>
</tr>
<tr>
<td>Stage of the HCC</td>
<td>Hepatic/Extra-hepatic</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>13/7</td>
<td>---</td>
</tr>
</tbody>
</table>

5.3 Results of the SPME urinary VOCs detection

SPME urine test was able to clearly differentiate between healthy controls and each cancer group as shown in the PCA plots (Figure 5.1, Figure 5.2 & Figure 5.3). For HCC, the sensitivity of the conventional alpha fetoprotein (AFP) alone in this work was 54.8%. AFP (>10 ku/L) was found in only 17 cases. When comparing the urine test to AFP, it showed good discrimination in diagnosis of HCC (Figure 5.4). The SPME urine test showed AUC for detection of HCC patients with normal AFP of 0.68 (ROC Curve Area was 0.68, SE 0.06, 95% CI 0.54 to 0.81 and P <0.005). The AUC increased in patients with raised AFP to 0.83. (ROC Curve Area was 0.83, SE 0.05, 95% CI 0.73 to 0.93 and P<0.0001). Visually, sensor 4 and sensor 8 were the two sensors with the least relative responses in all groups. For HCC, sensor 5 was predominantly responsive. However, no unique relative response between different cancers or controls was identified. Relative responses from the eight sensors are shown in (Figure 5.5).

Table 5.3 shows the results obtained from training a radial basis function neural network against the imbalanced data sets available to investigate whether these cancers could be discriminated from each other on the basis of the urine VOCs data that were captured. Unlike an ROC curve which is a binary classifier, this method allowed addressing a multi-class
problem to discriminate between the different cancers assessed in this work. Evaluating the trained network against previously unseen samples, showed that the average potential accuracy for this could be as great as 93.7%, based on the limited number of samples available. These results should be taken as cautiously indicative at this time as the neural network could be overtrained and would require a population study with a large cohort of samples for verification.

**Figure 5.1**

Principal component analysis (PCA) measuring degree of variance of data captured by the SPME device. PCA plot showing captured data of urine VOCs from patients with HCC and non-cancer controls.
**Figure 5.2** Principal component analysis (PCA) measuring degree of variance of data captured by the SPME device. PCA plot showing captured data of urine VOCs from patients with prostate cancer and non-cancer controls.

**Figure 5.3** Principal component analysis (PCA) measuring degree of variance of data captured by the SPME device. PCA plot showing captured data of urine VOCs from patients with bladder cancer and non-cancer controls.
Figure 5.4 Urinary VOCs ROC curve for detection of HCC with normal AFP vs raised AFP levels
Figure 5.5 Box plots to compare the relative response of the eight metal sensors in the SPME work (cancer groups and controls)
Table 5.3 Discrimination of cancers via urinary VOCs using an RBF neural network

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Sensitivity</th>
<th>Selectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCC</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>0.86</td>
<td>0.96</td>
</tr>
<tr>
<td>Bladder cancer</td>
<td>1.0</td>
<td>0.97</td>
</tr>
<tr>
<td>Controls</td>
<td>0.72</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Statistics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PPV %</td>
<td>89.23</td>
</tr>
<tr>
<td>NPV %</td>
<td>98.02</td>
</tr>
<tr>
<td>F Score %</td>
<td>93.42</td>
</tr>
<tr>
<td>Average System Accuracy %</td>
<td>93.7</td>
</tr>
<tr>
<td>System Error %</td>
<td>6.3</td>
</tr>
</tbody>
</table>

Sensitivity (range 0-1.0) (the ability of the test to correctly identify those patients with the disease). Unlike the ROC curve which evaluates a binary problem, a neural network addresses a multi-class problem to discriminate between the different cancers investigated. Selectivity (range 0-1.0) reflects the ability to discriminate between the different classes in a multiclass problem. Specificity as displayed in an ROC curve is used to show the method’s ability responding to one single analyte only, while Selectivity is used when the method is able to respond to several different analytes in the sample. *PPV = positive predictive value (Precision), NPV = negative predictive value (Recall), F Score = the weighted harmonic mean of the test’s PPV and Sensitivity, Average System Accuracy = the accuracy of the neural network to classify correctly and System Error – neural network misclassification error.
5.4 Results of GC-IMS urinary VOCs detection

The GC-IMS produces an analysis output, this is shown in Figure 5.6. The background is defined in blue, with the red peaks showing areas of high intensity. The long red line is the output of the instrument to the carrier gas (in this case Nitrogen). The results show that GC-IMS was able to separate different chemicals within the urine sample without saturating the machine and without chemical overlap. Table 5.4 shows results of the separation of those with HCC from non-HCC (liver fibrosis). The AUC (area under the curve), sensitivity and specificity were 0.97, 0.43 and 0.95, respectively. Conversely, the separation of those with HCC compared to non-HCC (without liver fibrosis) shows modest separation with an AUC, sensitivity and specificity of 0.62, 0.60 and 0.74, respectively. Comparison of both fibrosis and non-fibrosis revealed an AUC, sensitivity and specificity of 0.63, 0.29 and 0.90. The ROC curves for different liver groups, using GC-IMS, are presented in Figure 5.7. The optimal threshold value was applied for the comparison of HCC and fibrosis samples, HCC and Non-fibrosis samples and fibrosis and non-fibrosis samples and were 0.39, 0.35 and 0.52, respectively. The results showed that the diagnostic test gave 4 false positive for comparison between HCC and fibrosis samples, 8 false positive tests for HCC and non-fibrosis samples, with only 3 false positive tests for fibrosis and non-fibrosis samples. Also, the number of false negative tests for HCC and fibrosis samples was only 1 whereas the number of false negative tests for HCC and non-fibrosis samples and fibrosis and non-fibrosis samples were 12 and 5, respectively.
Figure 5.6. Example output of GC-IMS instrument when applied to urinalysis of VOCs. The background is defined in blue, with the red peaks shows areas of high intensity. The long red line is the output of the instrument to the carrier gas (in this case Nitrogen).
Table 5.4 GC-IMS result for different liver disease groups (95% confidence intervals are in brackets). compares HCC with non-HCC using GC-IMS analysis, providing area under the curve (AUC), sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV). LR: linear regression, RF: Random Forest.

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>Classifiers</th>
<th>Threshold</th>
<th>AUC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCC vs Fibrosis</td>
<td>RF</td>
<td>0.39</td>
<td>0.97</td>
<td>0.43</td>
<td>0.95</td>
<td>0.75</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.91–1)</td>
<td>(0.13–75)</td>
<td>(0.86–1)</td>
<td>(0.53–1)</td>
<td>(0.68–0.96)</td>
</tr>
<tr>
<td>HCC vs Fibrosis</td>
<td>LR</td>
<td>0.39</td>
<td>0.92</td>
<td>0.71</td>
<td>0.9</td>
<td>0.71</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.84–1)</td>
<td>(0.40–1)</td>
<td>(0.77–1)</td>
<td>(0.4–1)</td>
<td>(0.78–1)</td>
</tr>
<tr>
<td>HCC vs Non-Fibrosis</td>
<td>RF</td>
<td>0.35</td>
<td>0.62</td>
<td>0.6</td>
<td>0.74</td>
<td>0.6</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.48–0.76)</td>
<td>(0.41–0.78)</td>
<td>(0.61–0.87)</td>
<td>(0.42–0.78)</td>
<td>(0.61–0.88)</td>
</tr>
<tr>
<td>HCC vs Non-Fibrosis</td>
<td>LR</td>
<td>0.34</td>
<td>0.62</td>
<td>0.71</td>
<td>0.9</td>
<td>0.48</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.48–0.75)</td>
<td>(0.4–0.68)</td>
<td>(0.77–0.78)</td>
<td>(0.29–0.65)</td>
<td>(0.52–0.8)</td>
</tr>
<tr>
<td>Fibrosis vs Non-Fibrosis</td>
<td>RF</td>
<td>0.41</td>
<td>0.5</td>
<td>0.14</td>
<td>0.97</td>
<td>0.5</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.25–0.74)</td>
<td>(0–0.4)</td>
<td>(0.91–1)</td>
<td>(0.5–1)</td>
<td>(0.73–0.92)</td>
</tr>
<tr>
<td>Fibrosis vs Non-Fibrosis</td>
<td>LR</td>
<td>0.46</td>
<td>0.63</td>
<td>0.29</td>
<td>0.90</td>
<td>0.4</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.36–0.89)</td>
<td>(0–0.6)</td>
<td>(0.81–0.97)</td>
<td>(0–0.83)</td>
<td>(0.74–0.94)</td>
</tr>
</tbody>
</table>
Figure 5.7 ROC for GC-IMS Analysis: (a) HCC vs Fibrosis; (b) HCC vs Non-Fibrosis; (c) Fibrosis vs Non-Fibrosis
5.5 Results of GC-TOF-MS urinary VOCs detection

The GC-TOF-MS produces an analysis output, this is shown in Figure 5.8. The GC-TOF-MS produces a broad range of chemical peaks throughout the spectra with good separation. On average, the total number of peaks detected by GC-TOF-MS after analyzing HCC and fibrosis samples were 112 and the total number of peaks detected for HCC and Non-Fibrosis samples were 74. Similarly, for fibrosis and Non-Fibrosis samples, on average 79 peaks were detected. Test accuracies for HCC and non-HCC cases using GC-TOF-MS are provided in Table 5.5 and Figure 5.9.

GC-TOF-MS identified 200 chemicals for different liver diseases groups. Then, using the National Institute of Standards and Technology (NIST) software, statistically significant chemicals between the disease groups was reached, with a p-value <0.05. These chemicals are listed in Table 5.6 and Table 5.7. In HCC cases the abundance of whether a chemical increased or decreased was looked for. This has not been attempted in the liver fibrosis group due to the small sample size (n=7).

Figure 5.8 Example output of GC-TOF-MS instrument when applied to urinalysis of VOCs
Table 5.5 GC-TOF-MS result for different liver disease groups (95% confidence intervals are in brackets). compares HCC with non-HCC using GC-TOF-MS analysis, providing AUC, sensitivity, specificity, negative predictive value and positive predictive value. LR: Linear Regression, RF: Random Forest.

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>Classifiers</th>
<th>Threshold</th>
<th>AUC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCC vs Fibrosis</td>
<td>LR</td>
<td>0.32</td>
<td>0.79</td>
<td>0.71</td>
<td>0.66</td>
<td>0.66</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.50 – 1)</td>
<td>(0.40 – 1)</td>
<td>(0.33 – 1)</td>
<td>(0.23 – 1)</td>
<td>(0.40 -1)</td>
</tr>
<tr>
<td>HCC vs Fibrosis</td>
<td>RF</td>
<td>0.33</td>
<td>0.71</td>
<td>0.71</td>
<td>0.5</td>
<td>0.60</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.40 – 0.91)</td>
<td>(0.40 – 1)</td>
<td>(0.17 – 0.86)</td>
<td>(0.25 – 1)</td>
<td>(0.40 – 0.89)</td>
</tr>
<tr>
<td>HCC vs Non-Fibrosis</td>
<td>LR</td>
<td>0.32</td>
<td>0.83</td>
<td>0.66</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.62 – 1)</td>
<td>(0.33 – 1)</td>
<td>(0.43 – 1)</td>
<td>(0.40 – 1)</td>
<td>(0.33 – 1)</td>
</tr>
<tr>
<td>HCC vs Non-Fibrosis</td>
<td>RF</td>
<td>0.31</td>
<td>0.71</td>
<td>0.66</td>
<td>0.43</td>
<td>0.60</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.43 – 0.95)</td>
<td>(0.30 – 1)</td>
<td>(0.14 – 0.75)</td>
<td>(0.20 – 1)</td>
<td>(0.20 – 0.80)</td>
</tr>
<tr>
<td>Fibrosis vs Non-Fibrosis</td>
<td>LR</td>
<td>0.34</td>
<td>0.84</td>
<td>0.86</td>
<td>0.57</td>
<td>0.80</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.58 – 1)</td>
<td>(0.60 – 1)</td>
<td>(0.25 – 0.88)</td>
<td>(1 – 0.5)</td>
<td>(0.38 – 0.90)</td>
</tr>
<tr>
<td>Fibrosis vs Non-Fibrosis</td>
<td>RF</td>
<td>0.47</td>
<td>0.60</td>
<td>0.71</td>
<td>0.57</td>
<td>0.66</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.33 – 0.88)</td>
<td>(0.40 – 1)</td>
<td>(0.25 – 0.87)</td>
<td>(0.33 – 1)</td>
<td>(0.22 – 0.88)</td>
</tr>
</tbody>
</table>
Figure 5.9 ROCs for (a) HCC and Fibrosis samples, (b) HCC and Non-Fibrosis and (c) Fibrosis and Non-Fibrosis using GC-TOF-MS.
<table>
<thead>
<tr>
<th>No.</th>
<th>Retention Time (min)</th>
<th>Chemical</th>
<th>p-value</th>
<th>Abundance change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15.25</td>
<td>4-Methyl-2,4-bis(p-hydroxyphenyl)pent-1-ene, 2TMS derivative</td>
<td>&lt;0.01</td>
<td>Lower for HCC</td>
</tr>
<tr>
<td>2</td>
<td>2.5998</td>
<td>2-Butanone</td>
<td>0.03637</td>
<td>Higher for HCC</td>
</tr>
<tr>
<td>3</td>
<td>4.5684</td>
<td>2-Hexanone</td>
<td>0.04309</td>
<td>Lower for HCC</td>
</tr>
<tr>
<td>4</td>
<td>6.3215</td>
<td>Benzene, 1-ethyl-2-methyl-</td>
<td>0.04183</td>
<td>Lower for HCC</td>
</tr>
<tr>
<td>5</td>
<td>12.1318</td>
<td>3-Butene-1,2-diol, 1-(2-furanyl)-</td>
<td>0.03247</td>
<td>Lower for HCC</td>
</tr>
<tr>
<td>6</td>
<td>8.2054</td>
<td>Bicyclo[4.1.0]heptane, 3,7,7-trimethyl-, [1S-(1a,3β,6a)]-</td>
<td>0.03553</td>
<td>Lower for HCC</td>
</tr>
<tr>
<td>7</td>
<td>13.861</td>
<td>Sulpiride</td>
<td>0.04369</td>
<td>Lower for HCC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No.</th>
<th>Retention time (min)</th>
<th>Chemical</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.13</td>
<td>3-Butene-1,2-diol, 1-(2-furanyl)-</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>2</td>
<td>12.26</td>
<td>Isopropyl myristate</td>
<td>0.00361</td>
</tr>
<tr>
<td>3</td>
<td>5.59</td>
<td>Octane, 2,2,6-trimethyl-</td>
<td>0.00457</td>
</tr>
<tr>
<td>4</td>
<td>6.27</td>
<td>Benzene, 1-ethyl-2-methyl-</td>
<td>0.04183</td>
</tr>
<tr>
<td>5</td>
<td>9.83</td>
<td>Biphenyl</td>
<td>0.04714</td>
</tr>
</tbody>
</table>
SPME was able to detect relative responses of signatures of the metabolic by-products present in urine of different cancer patients. SPME identified that use of VOCs from urine in detection of HCC has potential and that it can be cancer specific as it also detected urinary tract cancers. UHCW uses a common immunoassay for AFP detection with the lower range of detection being up to 10 ku/L. It is accepted that values > 10 ku/L are considered raised and suspicious of HCC. In the SPME work, HCC group had 17 patients with raised AFP, leaving the sensitivity of AFP in this group for HCC detection to be 54.84% (95%CI: 36.03%-72.68%). Urinary VOCs detection for HCC in patients with normal AFP alone was better at 68% and in those with a raised AFP was 83%. This indicates that the patterns of urinary VOCs hold promise for HCC detection even in combination with AFP.

Using GC-IMS and GC-TOF-MS further consolidated this in addition to the existing published studies utilizing urine VOCs for cancer detection. The work identified high specificity of 0.95 (95%CI: 0.86-1.00) in separating HCC from those with liver fibrosis offering important insights into role of urine VOC as a screening modality. Two studies showed that VOCs can be detected in the headspace of incubated in vitro HCC cells, supporting VOCs analysis for the assessment of hepatic enzymes function and also in predicting of HCC progression and metastasis (261,263). Qin et al utilized breath VOCs to identify HCC, independent of AFP levels or clinical stage (262). A recent study by Miller-Atkins et al showed that using 22 breath VOCs, detected HCC with 0.73 sensitivity compared with 0.53 for AFP in the same cohort (267).

The SPME AUC for HCC with negative AFP was 0.68, and it rose to 0.83 when combined with raised AFP and this was comparable to GC-IMS and GC-TOF-MS. For HCC detection the AUCs were 0.62 for GC-IMS and 0.79 for GC-TOF-MS (394,395). GC-IMS and GC-TOF-MS herein this work were advanced in proving the feasibility of urinary VOCs in differentiating between non-fibrotic, fibrotic and HCC stages.

Using GC-TOF-MS additionally tentatively identified 7 VOCs related to HCC. Literature search of these VOCs in relation to the development of HCC was completed, this concluded that the most described VOC in HCC was 2-butanone. In experimental models, exposure to 2-butanone lead to hepatotoxicity through potentiating the dihydronicotinamide-adenine dinucleotide phosphate (NADPH)‐cytochrome c reductase activity and the concentrations of cytochromes P450 enzymes. In addition, 2-butanone exposure in these models concomitantly with the known hepatocarcinogenic agent Carbon tetrachloride (CCI4) enhanced worse accelerated hepatotoxic metabolites and HCC formation. The 2-butanone was also found to inhibit the activity of membrane-bound monoamine oxidase, this is important because
monoamine oxidase was found to suppress HCC metastasis and progression by inhibiting the adrenergic system and its transactivation of epidermal growth factor receptor (EGFR) signaling (400-409).

In human studies, 2-Butanone was found in breath of HCC patients and was found to have the best diagnostic value among other organic compounds. In NAFLD paediatric patients 2-Butanone had significantly higher levels in the faeces and was related to faecal Lachnospiraceaee, family of anaerobic, spore-forming bacteria. Additionally, the study identified that Oscillospira family decrease in comparison to 2-butanol upregulation. The 2-butanol was found to be elevated in cirrhotic patients who underwent liver transplantation. The blood levels of 2-butanol were found to be discrimiant of liver cancer patients in comparison healthy individuals. Serum bilirubin also correlated with 2-butanol. The breath 2-butanol also discriminated different classes of liver cirrhosis demonstrated by child turcotte pugh (CTP) scores A, B and C (262,268,410-413).

The GC-TOF-MS further identified 4-Methyl-2,4-bis(p-hydroxyphenyl)pent-1-ene (MBP) as a urinary HCC VOC, it is a derivative of bisphenol A (BPA), a major pollutant. In the liver MBP metabolic activation from BPA occurs via the cytochrome P450 system. MBP can induce the function of oestrogen in experimental models via activating the oestrogen receptor (ER). In patients with HCC ERs are present and functional in around 50% of the cases but the role in promoting carcinogenesis is still not fully clear. The presence of urinary MBP in HCC patients in this work consolidated that MBP has a role in HCC perhaps by activation of ERs but this needs further research (414-416).

Another VOC found in this work related to HCC is 2-Hexanone. It was found to have a potentiating effect of the hepatotoxic agent chloroform and subsequent liver injury in experimental animal models. The mechanism for this was found due to induction of the CYP450 system. Chronic inhalation of an isomer of 2-Hexanone (methyl isobutyl ketone, MiBK) was found to cause hepatocellular adenomas and HCC in mice. This was shown to be in part through the activation of Pregnane X and constitutive androstane nuclear receptors, these receptors are responsible for regulation of CYP450 activity (417-423).

Further literature search identified urinary Benzene, 1-ethyl-2-methyl- has in HCC, this has been identified as blood biomarker of HCC in a study using SPME-GC/MS. Sulpiride is another chemical found in this work which is closely related to many chronic liver diseases. In particular, Sulpiride was found to be related biliary liver cirrhosis, NAFLD and cholestatic hepatits. Though it has not been identified as biomarker for HCC, the presence of Sulpiride may indicate that it may be a significant chemical for HCC. A study has suggested 3-Butene-1,2-diol, 1-(2-furanyl)- as an important VOC for lung cancer, but it has not been verified as
HCC biomarker. Whereas Bicyclo[4.1.0]heptane, 3,7,7-trimethyl-, [1S-(1a,3ß,6a)]- found in this work have not been identified as HCC or liver disease biomarkers requiring further investigation to confirm the presence of these VOCs in a larger cohort (425-428).

This work herein this chapter involving SPME, GC-IMS and GC-TOF-MS, points to the role of different VOCs in fibrogenesis and HCC pathogenesis. However, the work was limited by not accounting for other factors that can be involved in production of VOCs, like occupational environmental factors, diet, smoking and drugs. The other limitation was the small number of study participants. Nevertheless, the experimental work has answered the hypothetical questions that VOCs related to the function of CYP450 and ROS pathways in HCC and can be detected in the urine. In particular, as discussed earlier, the tentative identification of urinary VOCs in this study has been identified in various experimental and clinical studies. The strong literature around 2-butanone is encouraging to do more work to identify the exact biochemical pathway for this compound in liver fibrosis and HCC pathogenesis. However, these chemicals were not validated, nor quantified. In addition, the data from the GC-IMS system was analyzed using a pattern recognition approach and no attempt to identify chemical components.

The potential of this work lies in the noninvasive urinary detection for HCC and liver fibrosis and that it could provide real time check of metabolic activity and perhaps these tools could be beneficial in cancer diagnostics in the future and/or even monitoring response to treatment. The future for VOCs work in HCC and liver fibrosis will require further validation in a larger cohort prospectively obtaining longitudinal data.

In conclusion, urinary VOCs analysis in liver fibrosis and HCC has future potential as a detection method, it appears sensitive and specific. Characterisation of the precise metabolic VOCs will shine light on the biology of liver fibrosis and HCC but unlikely to improve the diagnostic potential.
Chapter 6. Exploration of plasma methylated SEPTIN9 in liver fibrosis and hepatocellular carcinoma
Chapter 6. Exploration of plasma methylated \textit{SEPTIN9} in liver fibrosis and hepatocellular carcinoma

This work was done to evaluate the diagnostic value and prognostic performance of plasma mSEPT9 in patient with liver fibrosis and HCC. This is to answer question 3 in section 1.30. The author declares that work described herein this chapter has been published (429).

6.1 Introduction

Emerging literature suggested that methylated \textit{SEPTIN9} (mSEPT9) DNA, has a role in HSCs activation, liver fibrogenesis and hepatocarcinogenesis. Loss of Septin 9 protein through epigenetic methylation leads to interfering with cellular signalling, proliferation and apoptosis (295-295).

6.2 Methods

This work was completed by following steps in sections 2.15, 2.15.1 and 2.15.2. This work was ethically approved by relevant NHS Ethics Committees in the UK (Ref 18717 and Ref 260179). It conformed with the ethical principles of the declaration of Helsinki. All participants provided written informed consent.

HCC diagnosis was made according to the current international guidelines, all inconclusive cases being confirmed by a liver biopsy. Liver disease patients were diagnosed following confirmation by clinical history and examination by liver physician with associated liver biochemical and/or radiological abnormalities continuing for more than 6 months. Advanced liver fibrosis/liver cirrhosis (LC) was further confirmed by stigmata of chronic liver disease and different radiological tests and/or Fibroscan. In case of ambiguity about the clinical diagnosis a liver biopsy was performed to ascertain the cause of liver disease and to look for presence or absence of LC.

Collected clinical parameters of interest included age at sampling, age at death, gender, survival outcomes, any history of diabetes, and body mass index (BMI). Further variables included were biochemical laboratory tests including platelets count, neutrophils, lymphocytes, aspartate aminotransferase (AST) and alanine aminotransferase (ALT), alkaline phosphatase (ALP), albumin, bilirubin and albumin-bilirubin (ALBI). NLR and ALBI were further mathematically calculated for each participant according to standard methods. In addition, surrogate FIB-4 Index was calculated in all participants to study its relationship to mSEPT9. The AFP was completed in line with UHCW guidelines using the AFP Roche Cobas® immunoassay with results ≤ 10KU/L as negative test and >10KU/L as positive test for HCC detection.
Participants provided 5 to 10 mL of peripheral whole blood samples by standard venepuncture in K2EDTA tubes and serum separating (SST) tubes. Samples were centrifuged for 15 minutes at 1500 g. Plasma and serum were then aliquoted (1mL each) into small universal vials. Samples were then immediately frozen at -80°C until analysis at the end of the recruitment. Plasma samples were then further analysed to obtain results for mSEPT9. This was completed by rt-PCR. This was completed in two batches and operators were blinded to the clinical data of the participants, the rt-PCR amplification are shown in Figure 6.1.

The recruitment took place between 1 July 2013 and 30 November 2019. A follow-up period to measure survival outcomes was closed on 1 August 2020. 141 participants were recruited from the liver ward and clinics at UHCW. All participants were recruited prior to any anticancer therapy received and patients were excluded if they were pregnant or <18 years. There were 38 HCC cases and 103 non-HCC cases. All participants, except five, had liver disease.

Liver disease aetiologies were varied, including non-alcoholic fatty liver disease (NAFLD), metabolic liver disease, autoimmune liver disease, viral liver disease, alcohol related liver disease and cryptogenic causes. The advantage of having patients with liver disease in the non-HCC group was to have cases that are at risk of developing HCC in the future. Both HCC and non-HCC groups had features of LC. There were 60 cirrhotic and 81 non-cirrhotic cases. Among the cirrhotics there were (29/60) with HCC and (31/60) without HCC. The participants’ characteristics are shown in Table 6.1.
Figure 6.1 The mSEPT9 amplification graphs obtained from rt-PCR output. The two graphs show each batch of plasma samples analysed to obtain the mSEPT9 results. X-axis represents copies per reaction and Y-axis represents the fluorescence. Red curves are positive tests above the threshold and green curves are negative tests below the threshold. These graphs were obtained in real time using Applied Biosystems 7500 Fast Real-time PCR system. Post runs were analyzed with software version SDS V1.4 from fluorescence data of 45 cycles. Reporter dyes were FAM (red colour) for methylated SEPTIN9 and JOE (green colour) for the housekeeping gene ACTB. Data was analyzed post run in the Delta Rn vs Cycle mode with manual threshold levels set at 50000 for mSept9 and 25000 for ACTB, both with consecutive baselines of 10 and end cycles set at 22. Ct cycles for both detectors were then automatically calculated and graphs generated by the software. Valid limits for Epi proColon controls are as follows: Positive controls should generate Ct ≤ 41.1 for a positive mSept9 and Ct ≤ 29.8 for ACTB; Negative controls should yield no Ct values (“undetermined”) for mSept9 and Ct ≤ 37.2 for ACTB. Limits for test specimens are positive for mSept9 if obtain a Ct < 45 for mSept9 in conjunction with Ct ≤ 32.1 for ACTB; the test is considered negative for mSept9 if obtain no CT (“undetermined”) for mSept9 with Ct ≤ 32.1 for ACTB. The test is invalid if the Ct for ACTB is > 32.1 or “undetermined” irrespective of the mSept9 Ct value.
Table 6.1 Clinical and biochemical characteristics of recruited participants for the mSEPT9 work at time of obtaining the samples.*

<table>
<thead>
<tr>
<th>Covariate</th>
<th>HCC</th>
<th>Non-HCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Patients</td>
<td>38</td>
<td>103</td>
</tr>
<tr>
<td>Age, mean (range)</td>
<td>71.24 (46-87)</td>
<td>56.80 (26-88)</td>
</tr>
<tr>
<td>Gender Female/male</td>
<td>7/31</td>
<td>53/50</td>
</tr>
<tr>
<td>Survival outcome Dead/Alive</td>
<td>23/15</td>
<td>18/85</td>
</tr>
<tr>
<td>Days to death from diagnosis mean (range), days</td>
<td>237.70 (36-557)</td>
<td>1086.50 (349-2099)</td>
</tr>
<tr>
<td>Cause of liver disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>HBV</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>HCV</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>miscellaneous</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>NASH</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>primary/idiopathic</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>without liver disease</td>
<td>20</td>
<td>NAFLD</td>
</tr>
<tr>
<td>Histological/Radiological Features of Liver cirrhosis Present/absent</td>
<td>29/9</td>
<td>31/72</td>
</tr>
<tr>
<td>Diabetes Present/absent</td>
<td>17/21</td>
<td>29/74</td>
</tr>
<tr>
<td>BMI mean (range) kg/m²</td>
<td>27.508 (20.8-41)</td>
<td>30.6746 (18-48.8)</td>
</tr>
<tr>
<td>Platelets mean (range), 1 x 10⁹/L</td>
<td>153.37 (50-595)</td>
<td>204.15 (37-370)</td>
</tr>
<tr>
<td>Neutrophils mean (range), 1 x 10⁹/µL</td>
<td>4.61 (1.41-11.81)</td>
<td>6.32 (1.02-26.5)</td>
</tr>
<tr>
<td>Lymphocytes mean (range), 1 x 10⁹/µL</td>
<td>1.34 (0.26-3.74)</td>
<td>2.05 (0.48-7.12)</td>
</tr>
<tr>
<td>NLR Mean (range)</td>
<td>4.96 (0.79-39.23)</td>
<td>4.44 (0.37-23.4)</td>
</tr>
<tr>
<td>AFP mean (range), KU/L</td>
<td>3568.03 (1-46289)</td>
<td>3.5 (1-37)</td>
</tr>
<tr>
<td>AST mean (range), U/L</td>
<td>89.24 (20-355)</td>
<td>38.01 (14-236)</td>
</tr>
<tr>
<td>ALT mean (range), U/L</td>
<td>44.34 (12-149)</td>
<td>34.40 (5-304)</td>
</tr>
<tr>
<td>AST : ALT mean (range)</td>
<td>2.10 (0.28-5.3)</td>
<td>1.52 (0.07-5.81)</td>
</tr>
<tr>
<td>FIB-4 Mean (range)</td>
<td>8.18 (0.87-35.13)</td>
<td>3.14 (0.19-33.88)</td>
</tr>
<tr>
<td>FIB-4 stage (n)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 0-1 (1)</td>
<td></td>
<td>Stage 0-1 (44)</td>
</tr>
<tr>
<td>Stage 2-3 (11)</td>
<td></td>
<td>Stage 2-3 (36)</td>
</tr>
<tr>
<td>Stage 4-6 (26)</td>
<td></td>
<td>Stage 4-6 (23)</td>
</tr>
<tr>
<td>ALP mean (range), U/L</td>
<td>183.37 (83-551)</td>
<td>93.22 (37-279)</td>
</tr>
<tr>
<td>Albumin mean (range), g/L</td>
<td>37.29 (23-47)</td>
<td>42.63 (27-52)</td>
</tr>
<tr>
<td>Bilirubin mean (range), µmol/L</td>
<td>23.28 (5-84)</td>
<td>13.44 (2-123)</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>--------------</td>
<td>---------------</td>
</tr>
<tr>
<td>ALBI mean (range)</td>
<td>-2.34 (-3.53 to -0.77)</td>
<td>-2.96 (-4.14 to -1.22)</td>
</tr>
<tr>
<td>ALBI grade (n)</td>
<td>ALBI grade 1 (18)</td>
<td>ALBI grade 2 (16)</td>
</tr>
<tr>
<td></td>
<td>ALBI grade 3 (4)</td>
<td>ALBI grade 1 (87)</td>
</tr>
<tr>
<td></td>
<td>ALBI grade 2 (16)</td>
<td>ALBI grade 3 (2)</td>
</tr>
<tr>
<td>Stage of the HCC Hepatic/Extra-hepatic</td>
<td>27/11</td>
<td>--</td>
</tr>
<tr>
<td>Methylated SEPTIN9 positive</td>
<td>34</td>
<td>19</td>
</tr>
<tr>
<td>Methylated SEPTIN9 negative</td>
<td>4</td>
<td>83</td>
</tr>
<tr>
<td>Methylated SEPTIN9 failed test</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>HCC treatment modality received by patient (n)</td>
<td>Palliative chemotherapy (8)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>RFA (1)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>Surgical resection (2)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>TACE (17)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>Best supportive care (9)</td>
<td>--</td>
</tr>
</tbody>
</table>

*Characteristics of the HCC and non-HCC groups. HCC diagnosis was made in line with international guidelines. Liver disease was established by a combination of radiological scans, Fibroscan, laboratory markers, and histology. All covariates were collected at the time of plasma sampling for mSEPT9. Survival outcome was obtained on 01/08/2020, time of closing data collection and follow up period for recruited participants. Abbreviations: AFP, alpha-fetoprotein; ALBI, albumin-bilirubin grade calculated using the following formula ALBI = (log10 bilirubin × 0.66) + (albumin × -0.085), where bilirubin is in µmol/L and albumin in g/L; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; FIB-4 Score, fibrosis 4 score calculated using the following equation (Age x AST) / (Platelets x √(ALT); HBV, hepatitis B virus; HCV, hepatitis C virus; mSEPT9, methylated SEPTIN9; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; NLR, neutrophil to lymphocyte ratio.
6.3 Results

Descriptive statistics for both HCC and non-HCC patients (displayed in Table 6.1) show that among the 141 participants, there were 38 HCC patients (mean age 71 years) and 103 non-HCC patients (mean age 56.8 years). Out of the 60 females, 7 had HCC whilst out of the 81 males 31 had HCC. 41 patients died (of which, 23 were HCC cases) by the end of the follow-up period with the average survival time for the HCC patients being 237.7 days in comparison with 1086.5 days for the non-HCC patients. On average, BMI, platelets, albumin, lymphocytes and neutrophils were lower for the HCC patients, whereas NLR, AFP, mSEPT9, AST, ALT, bilirubin, ALBI and FIB-4 index were higher for the HCC patients. The most common causes for liver disease amongst the participants were alcohol, NASH, and NAFLD.

The diagnostic accuracy results of both AFP and mSEPT9 in identifying HCC are given in Table 6.2. AFP sensitivity and specificity were 50% (CI: 33.8%-66.62%) and 97.09% (CI: 91.72%-99.40%) respectively. mSEPT9 sensitivity was 89.47% (CI: 75.20%-97.06%) and specificity was 81.55% (CI:72.70%-88.51%). Thus, mSEPT9 showed higher sensitivity but lower specificity than AFP. Negative predictive value (NPV) for mSEPT9 was 95.45% (CI: 89.22% to 98.16%) compared to AFP NPV of 84.03% (CI: 79.27% to 87.87%). This suggests that a negative mSEPT9 was more accurate than AFP for excluding HCC when the disease was absent. In addition, combinations of AFP and mSEPT9 in identifying HCC were further assessed. The combined AFP and mSEPT9 sensitivity (both true positives) were 81% whilst the combined specificity (both true negatives) was 80% (Table 6.3). Combination of mSEPT9 and AFP has not shown superiority to detect HCC disease, indicating that mSEPT9 alone performed better for detecting HCC.

The ROC curves in Figure 6.2 evaluated the performance of mSEPT9 and AFP, separately, in identifying HCC amongst all participants. For AFP, the area under the ROC curve (AUROC) was 0.781 (SE 0.054, P=0.000, CI: 0.674-0.888) in comparison with mSEPT9 PMR AUROC of 0.888 (SE 0.034, P=0.000, CI: 0.821-0.955). ROC curves (Figure 6.3) for LC cases with HCC (29/60) obtained comparable results. The AFP AUROC is 0.710 (SE .071, P=0.005, CI 0.571-0.849) whilst the mSEPT9 PMR AUROC is 0.835 (SE 0.052, P=0.000, CI: 0.734-0.937). These ROC curves showed that mSEPT9 had good ability to identify HCC in patients with and without LC.

Little is known about the diagnostic performance of mSEPT9 at the various stages of liver disease. Therefore, the work here assessed mSEPT9 performance in comparison to AFP to identify patients at three different stages of liver disease from less severe to more severe, as follows: liver disease without LC or HCC (67/127); LC without HCC (31/27); LC with HCC
The 5 patients without liver disease and those with primary HCC but without LC were excluded. PMR mSEPT9 values and AFP levels among patients in these three stages were compared using nonparametric statistical tests, based on the assumption that AFP and PMR values were not normally distributed (Table 6.4). The levels of PMR mSEPT9 and AFP in all three stages were compared using the Kruskal-Wallis H test, which showed a significant difference between the three stages in PMR mSEPT9 values and in AFP levels (p=0.000).

Levels of PMR mSEPT9 and AFP for every two stages at a time using the Mann-Whitney U test were further compared. PMR mSEPT9 values were significantly different in patients with liver disease alone in comparison to those with LC (p=0.003), however there was no similar comparable difference with AFP levels (p=0.067), indicating that mSEPT9 is potentially better at distinguishing those two stages. Both AFP levels and PMR mSEPT9 values were significantly associated with the two more severe stages of the disease, with PMR mSEPT9 performing better (p<0.05). These results are further demonstrated visually in the box plots (Figure 6.4 & Figure 6.5), which show that AFP levels and PMR mSEPT9 values increase with liver disease severity, but this increase is more evident for mSEPT9 PMR values from liver disease alone to LC as compared to AFP.

Further analysis the data using logistic regression to investigate the effects of several conventional parameters and clinical covariates on the binary outcome (positive vs negative) in relation to mSEPT9, HCC, and LC is shown in Table 6.5. Starting with age as a covariate, logistic regression showed that mSEPT9, LC, and HCC were significantly associated with age, with odds ratios of 1.06(1.03,1.10)), 1.04(1.01,1.06)) and 1.10(1.06,1.15) respectively. Therefore, every year older increased the odds of a positive mSEPT9 result by 6%, having LC by 4% and having HCC by 1%. Age-adjusted logistic regression showed that there is no significant association between either LC or HCC with BMI or NLR. Age-adjusted logistic regression also showed that recruited HCC patients had similar risk factors to the ones described in the literature since HCC was significantly associated with age, male gender and presence of LC (p<0.05). The odds of patients with LC to have HCC were 7.95 times more than patients without LC. Patients with AFP >10KU/L had 59.6 more odds of having HCC, this was particularly noted due to the high AFP levels in secretory HCCs. Patients with positive mSEPT9 had 29.6 times more odds of having HCC than patients who had negative mSEPT9.

Logistic regression showed that mSEPT9 results were significantly associated with all covariates (p<0.05) apart from history of diabetes when adjusted for age (OR=1.51(0.69,3.28)). The age-adjusted odds ratio for being female was 0.37(0.16,0.80). Hence, females were 63% less likely to have a positive mSEPT9 result. This was particularly noted due to the small number of females compared to males (60:81).
The odds for patients with LC and HCC to have a positive mSEPT9 were respectively 6.74 and 27.4 times more than for patients with a negative result. This is important from a clinical perspective, as though there were positive mSEPT9 results in LC cases without HCC, the odds were higher in HCC cases. Logistic regression also identified that AFP >10KU/L, bilirubin, ALBI and FIB-4 index have higher odds of a positive mSEPT9 whereas lower platelets and albumin had higher odds of a positive mSEPT9. This suggests that surrogate markers related to liver fibrosis and liver function were significantly associated with mSEPT9.

For survival analysis, Cox-proportional hazards models were calculated from the time-to-death to blood sampling dependent on the conventional parameters and clinical covariates of interest (Table 6.6). The analyses showed that time-to-death was significantly associated with age (p=0.001) and so it is a potential confounder. When adjusted for age, time-to-death from blood sampling was significantly associated with gender, LC, HCC, AFP, mSEPT9, platelets count, albumin, bilirubin, ALBI and FIB-4 index (p<0.05). The only covariate that showed no significant association with time-to-death was history of diabetes (p=0.085). Further plotting of Kaplan-Meier (KM) survival curves, showed that mSEPT9 was more accurate survival marker than AFP in HCC cases at 20 months and in liver disease cases at 80 months (Figure 6.6 & Figure 6.7), therefore, mSEPT9 was a more accurate prognostic biomarker.
<table>
<thead>
<tr>
<th>mSEPT9 in HCC detection</th>
<th>Value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>89.47%</td>
<td>75.20% to 97.06%</td>
</tr>
<tr>
<td>Specificity</td>
<td>81.55%</td>
<td>72.70% to 88.51%</td>
</tr>
<tr>
<td>Positive Likelihood Ratio</td>
<td>4.85</td>
<td>3.19 to 7.39</td>
</tr>
<tr>
<td>Negative Likelihood Ratio</td>
<td>0.13</td>
<td>0.05 to 0.33</td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>64.15%</td>
<td>54.03% to 73.15%</td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>95.45%</td>
<td>89.22% to 98.16%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>83.69%</td>
<td>76.54% to 89.37%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AFP in HCC detection</th>
<th>Value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>50.00%</td>
<td>33.38% to 66.62%</td>
</tr>
<tr>
<td>Specificity</td>
<td>97.09%</td>
<td>91.72% to 99.40%</td>
</tr>
<tr>
<td>Positive Likelihood Ratio</td>
<td>17.17</td>
<td>5.38 to 54.73</td>
</tr>
<tr>
<td>Negative Likelihood Ratio</td>
<td>0.52</td>
<td>0.37 to 0.71</td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>86.36%</td>
<td>66.52% to 95.28%</td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>84.03%</td>
<td>79.27% to 87.87%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>84.40%</td>
<td>77.34% to 89.96%</td>
</tr>
</tbody>
</table>

*Sensitivity: probability that a test result will be positive when the disease is present (true positive rate). Specificity: probability that a test result will be negative when the disease is not present (true negative rate). Positive likelihood ratio: ratio between the probability of a positive test result given the presence of the disease and the probability of a positive test result given the absence of the disease. Negative likelihood ratio: ratio between the probability of a negative test result given the presence of the disease and the probability of a negative test result given the absence of the disease. Positive predictive value: probability that the disease is present when the test is positive. Negative predictive value: probability that the disease is not present when the test is negative. Accuracy: overall probability that a patient is correctly classified.
Table 6.3 Combined AFP and methylated *SEPTIN9* (mSEPT9) results in identifying or excluding HCC

<table>
<thead>
<tr>
<th>True Diagnosis</th>
<th>% Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HCC</strong></td>
<td></td>
</tr>
<tr>
<td>mSEPT9 Positive and AFP &gt;10</td>
<td>0.81</td>
</tr>
<tr>
<td>mSEPT9 Negative and AFP ≤ 10</td>
<td>0.08</td>
</tr>
<tr>
<td>mSEPT9 Positive and AFP ≤ 10</td>
<td>0.42</td>
</tr>
<tr>
<td>mSEPT9 Negative and AFP &gt;10</td>
<td>0.03</td>
</tr>
<tr>
<td>mSEPT9 and AFP discordant results</td>
<td>0.45</td>
</tr>
<tr>
<td>mSEPT9 Positive or AFP &gt;10</td>
<td>0.92</td>
</tr>
<tr>
<td>mSEPT9 Negative or AFP ≤ 10</td>
<td>0.53</td>
</tr>
<tr>
<td><strong>No HCC</strong></td>
<td></td>
</tr>
<tr>
<td>mSEPT9 Positive and AFP &gt;10</td>
<td>0.02</td>
</tr>
<tr>
<td>mSEPT9 Negative and AFP ≤ 10</td>
<td>0.80</td>
</tr>
<tr>
<td>mSEPT9 Positive and AFP ≤ 10</td>
<td>0.17</td>
</tr>
<tr>
<td>mSEPT9 Negative and AFP &gt;10</td>
<td>0.01</td>
</tr>
<tr>
<td>mSEPT9 and AFP discordant results</td>
<td>0.18</td>
</tr>
<tr>
<td>mSEPT9 Positive or AFP &gt;10</td>
<td>0.19</td>
</tr>
<tr>
<td>mSEPT9 Negative or AFP ≤ 10</td>
<td>0.98</td>
</tr>
</tbody>
</table>
Table 6.4 Nonparametric comparison of AFP and methylated SEPTIN9 (mSEPT9) PMR values in patient with liver disease, LC without HCC, and LC with HCC

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Group</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
<th>Mann-Whitney U</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFP</td>
<td>Liver disease without LC</td>
<td>67</td>
<td>46.04</td>
<td>3085</td>
<td>807</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>LC without HCC</td>
<td>31</td>
<td>56.97</td>
<td>1766</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mSEPT9 PMR</td>
<td>Liver disease without LC</td>
<td>67</td>
<td>45.58</td>
<td>3054</td>
<td>776</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>LC without HCC</td>
<td>31</td>
<td>57.97</td>
<td>1797</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFP</td>
<td>LC without HCC</td>
<td>31</td>
<td>24.42</td>
<td>757</td>
<td>261</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>LC with HCC</td>
<td>29</td>
<td>37</td>
<td>1073</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mSEPT9 PMR</td>
<td>LC without HCC</td>
<td>31</td>
<td>20.77</td>
<td>644</td>
<td>148</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>LC with HCC</td>
<td>29</td>
<td>40.90</td>
<td>1186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFP</td>
<td>Liver disease without LC</td>
<td>67</td>
<td>40.60</td>
<td>2720.5</td>
<td>442.50</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>LC with HCC</td>
<td>29</td>
<td>66.74</td>
<td>1935.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mSEPT9 PMR</td>
<td>Liver disease without LC</td>
<td>67</td>
<td>35.96</td>
<td>2409</td>
<td>131</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>LC with HCC</td>
<td>29</td>
<td>77.48</td>
<td>2247</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Kruskal-Wallis H</td>
<td>P value</td>
</tr>
<tr>
<td>AFP</td>
<td>Liver disease without LC</td>
<td>67</td>
<td>52.65</td>
<td></td>
<td>20.33</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>LC without HCC</td>
<td>31</td>
<td>65.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LC with HCC</td>
<td>29</td>
<td>88.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mSEPT9 PMR</td>
<td>Liver disease without LC</td>
<td>67</td>
<td>47.54</td>
<td></td>
<td>63.813</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>LC without HCC</td>
<td>31</td>
<td>62.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LC with HCC</td>
<td>29</td>
<td>103.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR (CI) p-value</td>
<td>Age Adj OR (CI) p-value</td>
<td>OR (CI) p-value</td>
<td>Age Adj OR (CI) p-value</td>
<td>OR (CI) p-value</td>
<td>Age Adj OR (CI) p-value</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------</td>
<td>------------------------</td>
<td>----------------</td>
<td>------------------------</td>
<td>----------------</td>
<td>------------------------</td>
</tr>
<tr>
<td><strong>HCC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (continuous)</td>
<td>1.10(1.06,1.15)</td>
<td>0.000</td>
<td>1.04(1.01,1.06)</td>
<td>0.004</td>
<td>1.06(1.03,1.10)</td>
<td>0.000</td>
</tr>
<tr>
<td>Age (&gt;60)</td>
<td>6.61(2.70,18.8)</td>
<td>0.000</td>
<td>2.0(1.01,4.02)</td>
<td>0.048</td>
<td>2.98(1.45,6.32)</td>
<td>0.003</td>
</tr>
<tr>
<td>Gender (Female)</td>
<td>0.22(0.08,0.52)</td>
<td>0.001</td>
<td>0.15(0.05,0.40)</td>
<td>0.000</td>
<td>0.41(0.20,0.82)</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>0.013</td>
<td></td>
<td>0.37(0.17,0.76)</td>
<td>0.008</td>
<td>0.45(0.22,0.93)</td>
<td>0.033</td>
</tr>
<tr>
<td>BMI</td>
<td>0.96(0.89,1.01)</td>
<td>0.202</td>
<td>0.95(0.88,1.01)</td>
<td>0.177</td>
<td>1.01(0.99,1.06)</td>
<td>0.396</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.01(0.99,1.06)</td>
<td>0.427</td>
</tr>
<tr>
<td></td>
<td>0.436</td>
<td></td>
<td></td>
<td></td>
<td>0.91(0.85,0.98)</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td>0.91(0.84,0.97)</td>
<td>0.008</td>
</tr>
<tr>
<td><strong>Cirrhosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCC</td>
<td>7.48(3.28,18.5)</td>
<td>0.000</td>
<td>7.48(3.28,18.5)</td>
<td>0.000</td>
<td>7.19(3.54,16.6)</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td></td>
<td>7.95(2.09,22.9)</td>
<td>0.000</td>
<td>7.48(3.54,16.6)</td>
<td>0.000</td>
</tr>
<tr>
<td>Cirrhosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFP (&gt;10)</td>
<td>33.3(10.2,152)</td>
<td>0.000</td>
<td>33.3(10.2,152)</td>
<td>0.000</td>
<td>33.3(10.2,152)</td>
<td>0.000</td>
</tr>
<tr>
<td>mSEPT9 (Pos,Neg)</td>
<td>37.1(13.0,136)</td>
<td>0.000</td>
<td>29.6(9.66,116)</td>
<td>0.000</td>
<td>37.1(13.0,136)</td>
<td>0.000</td>
</tr>
<tr>
<td>Diabetes</td>
<td>2.07(0.95,4.47)</td>
<td>0.065</td>
<td>2.07(0.95,4.47)</td>
<td>0.065</td>
<td>2.07(0.95,4.47)</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>1.33(0.56,3.13)</td>
<td>0.514</td>
<td>1.33(0.56,3.13)</td>
<td>0.514</td>
<td>1.33(0.56,3.13)</td>
<td>0.514</td>
</tr>
<tr>
<td></td>
<td>2.33(1.14,4.83)</td>
<td>0.021</td>
<td>2.33(1.14,4.83)</td>
<td>0.021</td>
<td>2.33(1.14,4.83)</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>1.87(0.89,3.97)</td>
<td>0.101</td>
<td>1.87(0.89,3.97)</td>
<td>0.101</td>
<td>1.87(0.89,3.97)</td>
<td>0.101</td>
</tr>
<tr>
<td></td>
<td>2.13(1.04,4.42)</td>
<td>0.040</td>
<td>2.13(1.04,4.42)</td>
<td>0.040</td>
<td>2.13(1.04,4.42)</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>1.51(0.69,3.28)</td>
<td>0.301</td>
<td>1.51(0.69,3.28)</td>
<td>0.301</td>
<td>1.51(0.69,3.28)</td>
<td>0.301</td>
</tr>
<tr>
<td></td>
<td>Platelets</td>
<td>NLR</td>
<td>Albumin</td>
<td>Bilirubin</td>
<td>ALBI</td>
<td>FIB-4</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>----------</td>
<td>----------</td>
<td>-----------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>191</td>
<td>0.992(0.987, 0.997)</td>
<td>1.001(0.973, 1.020)</td>
<td>0.83(0.76,0.90)</td>
<td>1.03(1.01,1.06)</td>
<td>5.62(2.82,12.3)</td>
<td>1.14(1.07,1.24)</td>
</tr>
<tr>
<td></td>
<td>0.994(0.988, 0.999)</td>
<td>1.004(0.971, 1.023)</td>
<td>0.81(0.74,0.89)</td>
<td>1.04(1.01,1.06)</td>
<td>7.51(3.28,19.7)</td>
<td>1.10(1.03,1.20)</td>
</tr>
<tr>
<td></td>
<td>0.980(0.973, 0.986)</td>
<td>0.99(0.94,1.01)</td>
<td>0.80(0.72,0.87)</td>
<td>1.12(1.07,1.18)</td>
<td>13.3(5.52,38.1)</td>
<td>1.56(1.31,1.95)</td>
</tr>
<tr>
<td></td>
<td>0.981(0.974, 0.987)</td>
<td>0.99(0.928,1.012)</td>
<td>0.81(0.73,0.88)</td>
<td>1.12(1.07,1.19)</td>
<td>12.0(4.99,34.3)</td>
<td>1.54(1.27,1.95)</td>
</tr>
<tr>
<td></td>
<td>0.992(0.987, 0.996)</td>
<td>1.51(1.22,1.94)</td>
<td>0.73(0.64,0.81)</td>
<td>1.04(1.02,1.08)</td>
<td>14.0(5.79,40.6)</td>
<td>1.30(1.16,1.50)</td>
</tr>
<tr>
<td></td>
<td>0.993(0.988, 0.998)</td>
<td>1.52(1.20,1.99)</td>
<td>0.73(0.63,0.81)</td>
<td>1.04(1.01,1.07)</td>
<td>14.9(5.86,46.0)</td>
<td>1.23(1.11,1.41)</td>
</tr>
<tr>
<td></td>
<td>0.003</td>
<td>0.891</td>
<td>0.000</td>
<td>0.007</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>0.021</td>
<td>0.682</td>
<td>0.000</td>
<td>0.007</td>
<td>0.000</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.563</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.582</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>0.001</td>
<td>0.000</td>
<td>0.006</td>
<td>0.000</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Table 6.6 Time-to-Event Analysis for death from HCC and associations with other covariates

<table>
<thead>
<tr>
<th></th>
<th>HCC</th>
<th>Age Adj p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age continuous</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Age (&gt;60)</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>Gender (F/M)</td>
<td>0.047</td>
<td>0.023</td>
</tr>
<tr>
<td>BMI</td>
<td>0.561</td>
<td>0.484</td>
</tr>
<tr>
<td>HCC</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Cirrhosis</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>AFP (&gt;10, ≤10)</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>mSEPT9 (Pos,Neg)</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Diabetes</td>
<td>0.006</td>
<td>0.085</td>
</tr>
<tr>
<td>Platelets</td>
<td>0.002</td>
<td>0.007</td>
</tr>
<tr>
<td>NLR</td>
<td>0.140</td>
<td>0.074</td>
</tr>
<tr>
<td>Albumin</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Bilirubin</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>ALBI</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>FIB-4</td>
<td>0.002</td>
<td>0.026</td>
</tr>
</tbody>
</table>
**Figure 6.2.** Receiver operating characteristics (ROC) curves for AFP and mSEPT9 in detection of HCC in all participants. AFP area under the receiver operating characteristics curve (AUROC) is 0.781 (SE 0.054, P=0.000, CI 0.674-0.888) while methylated SEPTIN9 (mSEPT9) PMR AUROC is 0.888 (SE 0.034, P=0.000, CI 0.821-0.955).

**Figure 6.3.** Receiver operating characteristics (ROC) curves for AFP and mSEPT9 PMR in detection of HCC with LC patients. AFP area under the receiver operating characteristics curve (AUROC) is 0.710 (SE 0.071, P=0.005, CI 0.571-0.849) while methylated SEPTIN9 (mSEPT9) PMR AUROC is 0.835 (SE 0.052, P=0.000, CI 0.734-0.937).
Figure 6.4 Simple box plot for methylated SEPTIN9 (mSEPT9) levels across patients with various stages of liver disease (n=127). To plot this PMR values were used, they are semi-quantitative values of mSEPT9 methylation (0 represent a negative test/absent methylation, >0 is a positive test). This plot demonstrates that, in the presence of methylation, the PMR values increase between liver disease alone, LC, and LC with HCC. The mSEPT9 has high NPV for HCC and Mann-Whitney U test for PMR mean values between liver disease, LC, and LC with HCC were significant (p<0.05).

Figure 6.5 Simple box plot for AFP across patients with various stages of liver disease (n=127). To plot this AFP quantitative values were used (≤ 10 represent a negative test and >10 is a positive test). This plot demonstrates that AFP values increase between LC and HCC but remain the same between liver disease alone and LC. Mann-Whitney U test for the mean AFP values between liver disease without LC and LC without HCC was not significant (p=0.067) but was significant between LC without HCC and LC with HCC (p<0.05).
Figure 6.6 Kaplan-Meier curves showing HCC patients’ survival over time in relation to the methylated \textit{SEPTIN9} (mSEPT9) and AFP results. The upper curves represent negative mSEPT9 and AFP \( \leq 10 \text{ Ku/L} \) and lower curve represents positive mSEPT9 and AFP \( > 10\text{Ku/L} \) (x axis represents time in months and y axis represents cumulative survival). The mSEPT9 was a better survival marker at 20 months follow-up.
Figure 6.7 Kaplan-Meier curves showing survival over time in all participants in relation to the AFP and methylated SEPTIN9 (mSEPT9) results. The upper curves represent negative mSEPT9 and AFP ≤ 10 Ku/L and lower curves represent positive mSEPT9 and AFP > 10Ku/L (x axis represents time in months and y axis represents cumulative survival). The mSEPT9 was a better survival marker at 80 months follow-up.
6.4 Discussion

Solid organ malignancy is associated with epigenetic changes including promoter hypermethylation of certain genes (275). Growing evidence indicates that epigenetic DNA markers could be used as diagnostic and prognostic tools (275). HCC currently has no accurate, scalable, non-invasive, simple diagnostic test. Not all patients have access to advanced scans, especially in areas of low resource settings. Given that the mSEPT9 assay is available commercially and showing promise in HCC, the work herein aimed to evaluate this biomarker in the patient cohort. The results comprised head-to-head comparisons with all conventional parameters and clinical covariates, as these were obtained at the time of blood sampling for the mSEPT9 and other blood-based tests.

The first mSEPT9 test was developed by Lofton-Day, et al in 2008 as a research kit (284). The second-generation test that is commercially available, Epi proColon® V2.0 CE (Epigenomics AG, Berlin, Germany) was approved in 2016 by the United States Food and Drug Administration (FDA) as a CRC screening test for members of the population at average-risk aged 50 years and above, who have refused colonoscopy and high-sensitivity faecal-based tests. The clinical testing assay requires large amount of plasma of (3.5 mL), on which three qPCR reactions are performed on cfDNA template derived from 0.875 mL plasma per qPCR reaction. If any one of the three qPCR reactions produce a positive test result, the clinical test is considered positive, according to the “1/3 algorithm”. Only 1 mL plasma available for testing. Therefore, a protocol that was previously validated prior to testing for CRC in Lynch syndrome patients was followed (334). The key difference was a reduced plasma volume to allow for a single PCR reaction from cfDNA template derived from 0.875 mL with result scoring based on the “1/1 algorithm”. This difference can result in a slight under-estimation of test sensitivity to detect disease and a slight over-estimation of test specificity, as compared to the “1/3 algorithm” (334,335,430).

This work herein this chapter is the first mSEPT9 analytical work in the UK to assess the diagnostic performance of mSEPT9 for HCC detection and survival. The work herein found mSEPT9 sensitivity and specificity to be 89% and 81%, respectively, for HCC detection. The AUROC was 0.854 (SE 0.037, P=0.000, CI 0.782-0.926). These results are comparable to prior findings. A study by Oussalah et al in German and French patients demonstrated that positive mSEPT9 was associated with HCC diagnosis with sensitivity of 80.8% and specificity of 95.8%. Oussalah et al methodology followed triplicate PCR reaction for detecting mSEPT9 (431). The comparable results suggest that the 1:1 algorithm protocol followed in this work for methylated SEPTIN9 (mSEPT9) detection is applicable for identifying HCC patients on a research basis.
Similarly, He et al demonstrated sensitivity and specificity of mSEPT9 to detect HCC among Chinese patients were 76.7% and 95.9%, respectively. However, their study used the “SensiColon” assay, which was modified from the US FDA-approved Epi proColon® V2.0 test and also used a triplicate qPCR assay to detect mSEPT9 (432).

A further study in a Japanese population used a newly developed assay; the combined restriction digital PCR (CORD) assay, which circumvents the need for bisulfite treatment and methylated DNA immunoprecipitation (without the need for DNA bisulfite treatment). Using this methodology, mSEPT9 had low sensitivity at 63.2% but good specificity at 90% for detecting HCC. This study also assessed HCC tumour tissues, which had a significantly higher copy number of mSEPT9 than non-tumour tissues; The median copy number was 2,360 in the tumour tissues and 34 in adjacent non-tumour tissues in patients with liver disease (433). Another recent study by Lewin et al showed that The mSEPT9 in addition to a DNA-methylation panel established by next generation sequencing to have 76.7% sensitivity and 64.1% specificity, In a post-hoc analysis, a combination with AFP (at 20 ng/mL) achieved 68% sensitivity at 97% specificity, highlighting potential combined use with AFP in HCC detection (434).

The findings by Oussalah et al and He et al showed that LC patients have abnormal plasma mSEPT9 levels (431,432). This in particular was comparable to the findings herein this work. Significant differences were found (incremental rise) in the means of the PMR values in patients with liver disease, LC (without HCC), and LC with HCC. This suggests a role for mSEPT9 in HCC development on the background of progression from liver disease to LC, due to the accumulation of aberrancies which aid in the malignant transformation of liver cells. These findings raise the potential clinical use for mSEPT9 in chronic liver disease and cirrhosis surveillance, because the PMR mSEPT9 changes were shown in this exploratory work to be present peripherally in plasma, making it an alternative blood test to AFP use in liver disease and LC. Further prospective studies that longitudinally follow cases with liver disease as they may develop into HCC will be required to confirm these findings.

HCC was noted to be more common in males and in older age which is consistent with existing literature. The mSEPT9 biomarker was significantly associated with age. For every year older, increased the odds of positive mSEPT9 by 6%. mSEPT9 had no association with diabetes and females were 63% less likely to have a positive mSEPT9 result (though there were lower number of females (60/141)). Similarly, the analysis by He, et al showed mSEPT9 to be independent of gender, however it also identified that patients >50 years of age exhibited higher sensitivity compared with those <50 years of age for mSEPT9.
The work herein showed that age is significantly associated with mSEPT9 results, LC, HCC, and time-to-death from blood sampling which is an accepted confounder in methylation epigenetic processes. This is important for future work to determine exact cut-off levels needed for a positive test in older HCC patients (430).

Through survival analysis, the work herein also demonstrated that mSEPT9 had more accurate survival outcomes in liver disease and HCC patients in comparison to AFP. Both were significantly associated with time-to-death (p<0.000) with higher survival rates for mSEPT9, and results were comparable to the discussed published literature (431,432).

It is important to note that mSEPT9 was significantly associated with BMI, LC, platelet count, NLR, ALBI and FIB-4. This suggests that mSEPT9 is in turn associated with risk factors for HCC and the surrogate markers for liver fibrosis, liver function and the inflammatory status of the patient.

It is known that AFP secreting HCC nodules can secrete at very high levels (>100,000 KU/L). This explains the high specificity for AFP in HCC. However more than 50% of HCCs are non-AFP secreting and hence the low sensitivity and false negative rates for the AFP test. This pattern was similar within the recruited cohort in herein this exploratory work with half of HCC patients with negative AFP ≤ 10 KU/L. Given AFP has low sensitivity and higher false negative rate, diagnosis usually requires further evaluation with a liver biopsy in absence of diagnostic scans. To improve AFP performance, many cut off levels for AFP in HCC has been suggested. These AFP characteristics necessitate an additional endeavour to find better biomarkers in HCC detection, this can be found in the commercially available mSEPT9 as demonstrated by the results herein this chapter with both high sensitivity and specificity and (435-437).

This exploratory work was limited in part by the small cohort, single measurement of mSEPT9 at the recruitment stage and short follow up period. This in turn limited further evaluation of the true significance of plasma mSEPT9 as an independent prognostic marker in relation to the conventional parameters and clinical covariates explored in this work. Nevertheless, the findings are the first reported in a UK population and consistent with those of others.

In summary, the results herein this chapter show the feasibility of mSEPT9 as an additional biomarker for HCC detection, prognosis and surveillance. It appears a superior alternative to AFP with better negative predictive value.
Chapter 7. Discussion and future directions
Liver fibrosis and hepatocellular carcinoma (HCC) continue to cause significant morbidity and mortality. They both lack simple noninvasive biological biomarkers. From the start, this project aimed to address the knowledge gaps in liver fibrosis and HCC.

The project investigated whether peptides could be demonstrated peripherally in the urine and if these could be linked to liver fibrogenesis and HCC microenvironment. This resulted in identifying urinary peptides for liver fibrosis and HCC with good diagnostic and prognostic performance as demonstrated in chapter 3 and chapter 4. Furthermore, the identified peptides formed basis for predicting the proteases involved in their production, which were further compared with transcriptomics of human liver genome and tested in human liver tissue specimens including liver fibrosis and HCC. This approach proved these peptides are directly related to the pathophysiology in the liver, rather than changes occurring at a different site before reaching the urine. The envisioned use of these urinary peptides as biomarkers could be in the primary care setting to identify at risk patients. Unfortunately, the work did not provide answers in relation to comparison with histological grade of liver fibrosis or the stage of HCC and further work will be required to establish this.

Nonetheless, the idea of a suitable urinary test to check for liver fibrosis and HCC is promising. Urine is easy to obtain and could be more acceptable by patients to provide in comparison to a blood test. Urine test could easily be scalable to use in the wider population. Further assays could also be developed to assess the discovered proteases individually whether in the tissue or in the blood. For example, staining for Kallikrein-6 (KLK6) and Meprin A Subunit α (MEP1A) could determine prognostic course or choice of therapy if further tested against stage of HCC.

The project also aimed to address the feasibility of using urinary volatile organic compounds (VOCs) as diagnostic tool and if they could be linked to liver fibrosis and HCC pathophysiology. Using urinary VOCs technologies, the project showed the pattern of these compounds could identify liver fibrosis and HCC patients. When comparing, solid phase microextraction (SPME) to more advanced analytical methods like gas chromatography coupled with ion mobility spectrometry (GC-IMS) and gas chromatography time-of-flight mass spectrometry (GC-TOF-MS), the area under the curve (AUC) of the analytical method did not improve significantly as demonstrated in Chapter 5, however this part of the work included smaller cohorts. GC-TOF-MS was able to tentatively identify the nature of the VOCs linked to both HCC and liver fibrosis. These VOCs have been linked to liver fibrosis, HCC and other liver diseases when researched in the literature. Each compound will require further molecular analysis in the liver cells to understand its importance in liver disease development.

Additionally, this project aimed to also evaluate a commercially available methylation DNA assay, methylated SEPTIN9 (mSEPT9), in liver disease patients including liver fibrosis and
HCC. As demonstrated in Chapter 6, the mSEPT9 showed promise as it increases incrementally between stages of liver disease and it is also associated with liver function surrogate marker (ALBI grade) and another surrogate marker for liver fibrosis (FIB-4 Score), these characteristics make it suitable for this blood test to be further evaluated on mass scale for future use. Future studies should compare mSEPT9 to radiology tests to assess feasibility as a detection modality for liver fibrosis and HCC. The envisioned example for use of mSEPT9 could be as an adjunct with or without USS as well as with or without AFP in LC surveillance for HCC.

The data analyses applied to identify these biomarkers were appropriate. The project used the receiver operating characteristic (ROC) curve, the area under the curve (AUC), sensitivity and specificity in identifying the accuracies of these biomarkers against set international criteria for HCC diagnosis and against liver biopsy and other surrogate biomarkers in liver fibrosis. The ROC analysis is an acceptable method for evaluating a biomarker accuracy. ROC analysis has been used extensively in assessment of biomarkers to separate between the diseased and healthy subjects. ROC as a predictive model is also commonly used to estimate the risk of adverse outcome based on the patient’s risk profile in medical researches. ROC analysis has several advantages, it is independent of prevalence of disease since it is based on sensitivity and specificity. ROC analysis can also assess several diagnostic tasks on the same subjects and can consider the covariance between two correlated ROC curves. The ROC analysis can also easily obtain the sensitivity by visualizing the curve and can be used in determining the optimal cut-off value through the Youden Index. Herein this thesis, this index been used in the urinary peptide work. Youden index maximizes the difference between sensitivity and false positive rate (1-specificity). Therefore, maximizing sensitivity and specificity across various cut-off points, the optimal cut-off point is then calculated.

The project aimed to decrease bias through comparing diseased patients (HCC) and patients at risk of disease (liver fibrosis). Thus, the clinical spectrum of disease was included with varying degrees of severity. Also, samples analyses were completed blindly and operators were not aware of the data of the cases till after the analyses were completed to decrease bias. The work also aimed to eliminate confounders in the discovery of these biomarkers by adjusting for a common confounding factor which is age. It is well known that age could change the shape of the ROC curve (337).

Naturally, given the exploratory approach of the work herein this thesis, the identified biomarkers will require further validation before implementation in clinical practice. To complete this in the near in the future the proposed projects should be in line with the recent published vision of the International Liver Cancer Association (ILCA) on biomarker development (438). Studies should have longitudinal study designs that follow patients before
developing liver fibrosis, during the liver fibrosis and HCC disease and also covering the progression of the HCC course, while comparing the performance of these biomarkers to conventional diagnostic and prognostic methods. Case selection should also include patients on treatment and without treatment, with varying stages of HCC and varying degrees of liver fibrosis based on histology. The last point is understandably difficult to achieve given the difficulty of obtaining liver biopsy for research purposes due to the associated complication risks. Also, it is important to have larger cohort of patients.

Further work will also be required to simplify the methodologies used in this project. Capillary electrophoresis coupled with mass spectrometry (CEMS), GC-IMS and GC-TOF-MS lead to discovery of the urinary biomarkers, however, they are labour intensive and are used as research tools. Therefore, further work is still required to develop simpler assays for the discovered urinary biomarkers.

The results demonstrated herein this thesis can further be taken forward to provide clinical impact provided that are validated through study designs that implement repeated experimental biomarker validation to gather high quality clinical evidence about their use. The clinical evidence in this scenario relates to collection of summaries of the best available results about what works and what doesn't work in identifying liver fibrosis and HCC (439-442). Table 7.1 further illustrates the quality of clinical evidence and the associated study design linked to future biomarker development. The author envision that the identified novel biomarkers herein this thesis could have clinical role in diagnosis, screening, surveillance and prognosis and assessment of treatment in both liver fibrosis and HCC conditions.

In conclusion, this work has addressed the research questions. It lead to a) discovery of new urinary peptides and proteases in patients with liver fibrosis and HCC, b) identified the pattern and chemical composition of urinary VOCs in patient with liver fibrosis and HCC and c) evaluated the diagnostic and prognostic performance of plasma methylated SEPTIN9 (mSEPT9) in patient with liver fibrosis and HCC.
Table 7.1 Levels of evidence in biomarker studies. The table demonstrates key biomarker characteristics required prior to implementation in practice. Adapted from Ref (438)

<table>
<thead>
<tr>
<th>Biomarker characteristics</th>
<th>Level 1 RCT</th>
<th>Level 2a - Archived samples from RCT</th>
<th>Level 2b- Prospective cohort</th>
<th>Level 2b- Retrospective cohort</th>
<th>Level 3-Convenience or case-control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk stratification</td>
<td>Could be incorporated in RCT evaluating early detection biomarker</td>
<td>N/A</td>
<td>Useful for late stage model validation</td>
<td>Used for model derivation and optimization</td>
<td>Limited utility as can confound early detection and risk stratification</td>
</tr>
<tr>
<td>Early detection</td>
<td>Determine if biomarker-based strategy reduces mortality</td>
<td>Evaluate ability of biomarker to detect preclinical disease</td>
<td>Enumerates detection and false referral rates after acting on biomarker for diagnostic work up</td>
<td>Determines ability of biomarker to detect preclinical disease</td>
<td>Distinguishes early HCC from liver fibrosis/LC</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>N/A</td>
<td>N/A</td>
<td>Used model for validation</td>
<td>Determines performance metrics of the biomarker</td>
<td>Discriminates HCC from other focal hepatic lesions</td>
</tr>
<tr>
<td>Prognosis</td>
<td>N/A</td>
<td>Useful to derive novel prognostic stratification factors</td>
<td>Used model for validation</td>
<td>Used for optimization of prognostic model performance</td>
<td>Classifies patients based on outcomes</td>
</tr>
<tr>
<td>Treatment response</td>
<td>Determine if biomarker-based strategy improves treatment response and survival</td>
<td>Allows identification of subset of responders in negative trials</td>
<td>Limited utility due confound prognosis and treatment response</td>
<td>Used to estimate effect size and optimize prediction model</td>
<td>Categorize patients based on odds to respond to therapy</td>
</tr>
</tbody>
</table>

Randomised controlled trial (RCT), not applicable (N/A), hepatocellular carcinoma (HCC), liver cirrhosis (LC)
IX. Appendix

Arden Tissue Bank Ethical Approval
UHCW Internal Ethical Approval
NHS Health Research Authority (HRA) Approval
NHS Research Ethics Committee (REC) Approval
Example of Patient Information Sheet (PIS)
Medical and Life Sciences Research Fund Award
Midlands Gastroenterological Society Research Grant Award
United European Gastroenterology National Schola Award
United European Gastroenterology Best submission in clinical Hepatology
Aileen Lynn Bequest Fund. Royal College of Physicians and Surgeons Glasgow
Turner-Warwick RCP London Prize
11.09.2019

Ayman Bannaga  
Gastroenterology Department  
University Hospital Coventry and Warwickshire  
Clifford Bridge Road  
CV2 2DX

Dear Ayman,

Project Title: The Fatty Liver and Liver Cancer Study (TENDENCY)

Applicant No: ATB19-013

Thank you for submitting your project to the Arden Tissue Bank Management Committee for consideration on 09.08.2019. I am pleased to inform you that it was the decision of the committee to support your project, and it was approved on 11.09.2019 by the Arden Tissue Bank Management Committee with the following comments:

- This approval grants you ethical approval for your project under the Arden Tissue Bank NRES generic ethical approval. Should you need to quote an ethics number this is 18/SC/0180.

- Should you wish to make any changes to your project, you must gain approval from our Committee prior to implementation.

- An Annual Progress Report (APR) should be submitted to us once a year throughout the study. The first report is due on: 11.09.2020.

- All publications resulting from the use of this tissue must acknowledge Arden Tissue Bank as the provider of the tissues.

- Notification of all publications resulting from the use of this tissue must be provided to Arden Tissue Bank at the time of publication. Citations will be used by Arden Tissue Bank on its website as evidence of its ongoing commitment to place quality tissues in the hands of researchers.

- Notifications of any serious breaches of the trial protocol, or incidents that affect the
integration of the tissue must be reported to the Tissue Bank Manager within 24 hours of any suspected breach being identified and confirmed.

May I take this opportunity to remind you that, as a researcher, you must ensure that your research is conducted in a way that protects the dignity, rights, safety and well-being of participants. This approval assumes that you have read and understood the Research Governance Framework for Health and Social Care and accept that your responsibilities as a researcher are to comply with it, the Human Tissue, Data Protection and Health & Safety Acts.

We wish you every success with your project.

Yours Sincerely

Sean James
Arden Tissue Bank Manager
Arden Tissue Bank Management Committee.
05 July 2019

Dr Ayman Bannaga
University Hospitals Coventry and Warwickshire NHS Trust
Clifford Bridge Road
Coventry
CV2 2DX

Dear Ayman,

Project Title: Predictors of survival and prognosis of hepatocellular carcinoma patients

R&D Ref: GF0336

Thank you for sending in the required documents and completing the GafREC form for the above study. Having reviewed the details of your proposed project, research involving previously collected, non-identifiable information including research undertaken by staff within a care team using previously collected information during the course of care of their own patients or clients, are excluded from NHS Research Ethics Committee (REC) review therefore; I can confirm that we are happy for you to carry out this project within UHCW NHS Trust.

Please be aware that should you wish to change the project in anyway, you must notify our office using the above reference.

I have logged your study on behalf of the Trust, which means you can proceed. I wish you every success with your project.

Yours Sincerely,

Jasmeet Bhambra
Research Administration Specialist.
Please note: This is the favourable opinion of the REC only and does not allow you to start your study at NHS sites in England until you receive HRA Approval.

14 June 2019

Dr Ayman Bannaga
Research Fellow
University Hospitals Coventry and Warwickshire NHS Trust
Clifford Bridge Road
Coventry
CV2 2DX

Dear Dr Bannaga

Study title: The Fatty Liver and Liver Cancer Study
REC reference: 19/NE/0213
IRAS project ID: 260179

The Proportionate Review Sub-committee of the North East - York Research Ethics Committee reviewed the above application on via correspondence.

Ethical opinion

On behalf of the Committee, the sub-committee gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or NHS management permission (in Scotland) must be obtained from each host organisation prior to the start of the study at the site concerned.

Confirmation of Capacity and Capability (in England, Northern Ireland and Wales) or NHS

A Research Ethics Committee established by the Health Research Authority
Professor Ramesh Arasaradnam
University Hospitals Coventry and Warwickshire NHS
Trust
Clifford Bridge Road
Coventry
CV2 2DX

02 July 2019

Dear Professor Arasaradnam,

Study title: The Fatty Liver and Liver Cancer Study
IRAS project ID: 260179
REC reference: 19/NE/0213
Sponsor University Hospitals Coventry and Warwickshire NHS Trust

I am pleased to confirm that HRA and Health and Care Research Wales (HCRW) Approval has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

Please now work with participating NHS organisations to confirm capacity and capability, in line with the instructions provided in the “information to support study set up” section towards the end of this letter.

How should I work with participating NHS/HSC organisations in Northern Ireland and
LETTER OF ACCEPTANCE AND COMMITMENT
(to be signed by the named researcher)

To: The Medical and Life Sciences Research Fund
email: charityadmin@mlsrf.co.uk

From: Medical and Life Sciences Research Fund bursary holder:
Dr Ayman Sharafeldin Bannaga
Gastroenterology
UHCV
Clifford Bridge Road
Coventry CV2 2DX

January 2019

To: Medical and Life Sciences Research Fund trustees,

I accept your offer of the awarded of a Medical and Life Sciences Research Fund bursary.

As the named researcher, I confirm that a report of the progress of the subject of the research programme - Understanding the Metabolome and Microbiome in Hepatocellular Carcinoma

- will be forwarded to the Medical and Life Sciences Research Fund Scientific committee in one year’s time.

The Medical and Life Sciences Research Fund will be acknowledged in any publication/presentation arising from the bursary.

Yours faithfully,

(Signed)

Dr Ayman Sharafeldin Bannaga
Gastroenterology
UHCV
Dear Participant,

You are invited to take part in a research study. Before you decide, it is important for you to understand why the study is being done and what it will involve for you. Please take time to read the following information carefully. Discuss it with friends, relatives or your GP if you wish. It is up to you to decide whether or not to take part in this study. Whatever you decide, the standard of care you receive will not be affected. If there is anything that is unclear, or you would like more information, please do not hesitate to ask.

**What is the purpose of the study?**
To study the blood and urine to identify new diagnostic tests in liver disease.

**Where is this study being carried out?**
This study will take place at University Hospitals Coventry and Warwickshire NHS Trust.

**Why have I been invited to take part?**
You have been invited to take part because you have been diagnosed with Hepatocellular Carcinoma (HCC).

**What does taking part in this study involve?**
You will first be sent the study documents via mail to inform you about the study – the documents will include a copy of the Participant Information Sheet and Invitation letter. You will then be contacted via telephone by a member of the research team to discuss the study further and to confirm if you would like to take part. This way, you will have sufficient time to decide whether you would like to participate or not.

The study will require one visit – if you decide to take part, you will be given the choice to meet after your routine appointment or attend a separate research clinic for the purpose of the study. You will be asked to provide blood and urine samples in one visit to the hospital. If you are unable to provide a urine sample at the appointment, you will be given the option to send the sample via post. Your samples will then be compared with the two other patient groups: Non-alcoholic fatty liver disease (NAFLD) and Healthy volunteers. This may potentially help find new methods of diagnosing liver disease. This will be of potential benefits to future patients and not necessarily you.

If you agree to participate,

Sign CONSENT form

↓

Clinic visit – sample (urine and blood) collection

↓
End of study procedures

Do I have to take part?

It is up to you to decide whether or not to take part. We will go through this information sheet with you, which we will then give to you to keep. If you do agree to take part, you will be asked to sign a consent form. You are free to withdraw at any time without giving a reason. If you decide not to take part, or to withdraw at any time, the standard of care you receive at this hospital will not be affected in any way.

Will I receive reimbursement for taking part?

If you attend a research specific appointment (i.e., not a planned hospital appointment), you will be reimbursed for car parking and travel costs. Please keep your receipts and give them to the research team on arrival.

What if something goes wrong?

In the very unlikely event of you being harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone’s negligence, then you may have grounds for legal action but you may have to pay for it. If you wish to complain or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal National Health Service complaints mechanism is available to you. Please contact:

Complaints Manager
University Hospitals of Coventry & Warwickshire NHS Trust
Clifford Bridge Road
CV2 2DX
Telephone no: 02476 965 198

For independent advice on research, you can contact PALS (Patient Advice and Liaison Service) on freephone 0800 028 4203, Email: feedback@uhcw.nhs.uk

In the unlikely event of you losing your capacity to consent during the course of the study, you would be withdrawn from the study. Identifiable data or tissue already collected with consent would be retained and used in the study. No further data or tissue would be collected or any other research procedures carried out on or in relation to the participant.

Will my taking part in this study be kept confidential?

If you consent to take part in the study, all information which is collected about you during the course of the research will be kept strictly confidential. Your name and address will be removed from all study data (anonymised) and you will have a study identification number assigned.

University Hospitals Coventry & Warwickshire (UHCW NHS Trust) is the sponsor for this study based in the United Kingdom. We will be using information from you and your medical records in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. UHCW NHS Trust will keep identifiable information about you for 25 years after the study has finished.

Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you
withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible. You can find out more about how we use your information at http://www.uhcw.nhs.uk/privacy/.

The Trust has a Data Protection Officer who ensures that individual rights are respected and that we comply with the law. If you have any concerns or questions about how we look after your personal information, please contact the Data Protection Officer, at information.governance@uhcw.nhs.uk. Alternatively, you could write to:

data Protection Officer
Information Governance Unit – FM Building
University Hospital Coventry
Clifford Bridge Road
Coventry
CV2 2DX

UHCW NHS Trust will use your name and contact details to contact you about the research study, and make sure that relevant information about the study is recorded for your care, and to oversee the quality of the study. Individuals from UHCW NHS Trust and regulatory organisations may look at your medical and research records to check the accuracy of the research study. The only people in UHCW NHS Trust who will have access to information that identifies you will be people who need to contact you about the study. The people who analyse the information will not be able to identify you and will not be able to find out your name or contact details.


What will happen to the results of the research study?
Results of the study will be shared with the public via peer reviewed scientific journals, conferences and meetings.

What are the possible risks of taking part?
With regards to taking a blood sample, it can be a little uncomfortable and occasionally there is a little bruising. However, you can opt out from this and provide only urine samples.

What are the benefits of taking part?
You will not receive any direct benefit from taking part in this study. However, your participation may help inform us more about the development of liver disease and also help to identify biomarkers that could be related to liver disease.

What happens to the samples taken?
The blood and urine samples will be stored anonymously in a freezer at the Arden Tissue bank which can only be accessed by the research team via the Tissue Bank team. Urine and blood samples will be analysed using techniques to identify proteins, nuclear markers and organic
molecules. After the analysis of samples, any remaining blood and urine samples will be disposed of in accordance with the Human Tissue Act (2004).

**Who is organising and funding this study?**

This study is funded by the Medical Life Sciences Research Fund.

This study has been organised by UHCW NHS Trust in conjunction with the University of Warwick. None of the researchers will be paid in person nor will they receive any financial gain for doing this study.

**Who has reviewed this project?**

This study has been reviewed and approved by <Insert REC Name and Ref>. Health Research Authority (HRA) Approval was granted on <Insert Date>. **Contact for further information:**

If you require further information or have any questions, please contact:

Study Investigator:
<insert contact details>
14/01/2020

Dr Ayman Bannaga
University Hospital Coventry
Clifford Bridge Road
CV2 2DX

Dear Dr Bannaga

Study grant application – November 2019

The Fatty Liver and Liver Cancer Study (TENDENCY)

Thank you for submitting your application for the small study grant for consideration by the Midland Gastroenterological Society Executive Committee. I am pleased to inform you your application has been successful.

Please be aware of the following:

Your grant of £1000 will be presented to you at the prize presentation at the forthcoming Midland Gastroenterological Society conference on 10 November 2019. I have attached the programme and application forms for your perusal. If you wish to receive a certificate of attendance you will need to register for this event.

- You will be required to submit an abstract, with a summary of the outcomes of your project, for presentation at a Midland Gastroenterological Society Conference within the next 18 months.

- We would ask that any academic output from this project (publications, conference papers etc) should acknowledge the grant and the society would be grateful to receive a copy of such outcomes for our records.

Should you have any queries please do not hesitate to contact me.

Yours sincerely

Dr Neeraj Bhal placing the name as a reference to the letter
Honorary Treasurer of the Midlands Gastroenterological Society
Notification National Scholar Award UEG Week Virtual 2020

Dear Ayman Bannaga

To stimulate the career of young investigators in the field of GI research, UEG will present you with National Scholar Award.

National Scholar Awardees will be highlighted in the online programme.

We are looking forward to meeting you soon and remain with kind regards

Sent by the UEG secretariat on behalf of
Herbert Tilg
Chairman UEG Scientific Committee

Sophie Gouché
Association Management
United European Gastroenterology (UEG)
Wickenburggasse 1
A-1080 Vienna, Austria

17/08/2020
Dear Dr. Ayman Bannaga,

Congratulations again that your abstract ‘P0978 - PROFILING OF URINARY PEPTIDES IN HEPATOCELLULAR CARCINOMA: AN EXPLORATORY STUDY’ has been selected for poster presentation at UEG Week Virtual 2020 and will be displayed in the poster exhibition!

I am happy to inform you that your poster is one of the best submissions we received this year.

Subsequently, your poster has been selected to take part in our ‘Moderated Posters’ sessions and we would like to congratulate you for this!

With best regards,

Prof Herbert Tilg
UEG Scientific Committee Chair

Sent by the UEG Secretariat
Maresa Wiener
Congress Management
United European Gastroenterology (UEG)
Wickenburggasse 1
A-1080 Vienna, Austria

31/07/2020
219

7 October 2020

Dear Dr. Bannaga

Scholarship Committee 7 October 2020

It is with pleasure that I write to inform you that you have been successful in your application for the Aileen Lynn Bequest Fund for funding towards your project entitled: The Fatty Liver and Liver Cancer Study.

The amount of £5,000 has been awarded to fund your project and arrangements for transfer of funds will be made. Please find attached bank details form for completion and return to scholarships@rcpsg.ac.uk

Your name, institution and project may be used in College literature outlining details of recipients of College awards and we would welcome any reports of progress in your research project.

Yours sincerely

Professor Matthew Walters FRCP(Glasg)
Chair - Scholarship Committee

scholarships@rcpsg.ac.uk
Dear Ayman

Lecture: ‘Noninvasive urinary detection of hepatocellular carcinoma. Multicentre study’

Thank you for submitting your application for the West Midlands Turner-Warwick Lecture 2021. The standard of entries was extremely high and I am pleased to advise you that have the highest overall score and that you are our winner - Congratulations!

This year, Turner Warwick Lecture award winners will be invited to record their lecture and it will be shared on the RCP Player platform which is viewed by doctors both in the UK and internationally. More details will be sent over in due course but for now please can you just confirm that you are happy to present.

If you could also send over your home address I will arrange for your award letter to be sent in the post. Please also provide a head and shoulders shot of yourself for inclusion in the TW year book.

You will also receive the following:

1. All lecturers receive either one year’s free membership or a free membership on renewal.
2. TW yearbook entry
3. Plaque
4. Virtual reception with PRCP and registrar

Please do let me know if you have any questions in the meantime.

Kindly acknowledge this email as soon as possible and provide the information required, after which one of my colleagues will be in touch with you to arrange the recording of your lecture.

Kind regards and congratulations!

Jenny

Jenny Ward | Regional Manager

West Midlands Region

Oxford and Thames Valley Region

Royal College of Physicians Regional Office

Birmingham Research Park, Vincent Drive, Edgbaston, Birmingham, B15 2SQ

Tel: 0121 803 1061 | 07458 084548 www.rcplondon.ac.uk | facebook
X. References


41. Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: regulation of
gene expression, enzyme activities, and impact of genetic variation. Pharmacol Ther.

42. Ray PD, Huang BW, Tsuji Y. Reactive oxygen species (ROS) homeostasis and redox

43. Dalle-Donne I, Rossi R, Colombo R, Giustarini D, Milzani A. Biomarkers of oxidative

44. Mathews JM, Raymer JH, Etheridge AS, Velez GR, Bucher JR. Do endogenous volatile
organic chemicals measured in breath reflect and maintain CYP2E1 levels in vivo? Toxicol


46. Shearer MJ, Fu X, Booth SL. Vitamin K nutrition, metabolism, and requirements: current

47. Lurie Y, Loebstein R, Kurnik D, Almog S, Halkin H. Warfarin and vitamin K intake in the

48. Zakhary NI, Khodeer SM, Shafik HE, Abdel Malak CA. Impact of PIVKA-II in diagnosis of

49. Okuda H, Obata H, Nakanishi T, Furukawa R, Hashimoto E. Production of abnormal
prothrombin (des-gamma-carboxy prothrombin) by hepatocellular carcinoma. A clinical and


51. El Hadi H, Vettor R, Rossato M. Vitamin E as a Treatment for Nonalcoholic Fatty Liver

52. Robinson MW, Harmon C, O'Farrelly C. Liver immunology and its role in inflammation


213. Study of Durvalumab and Tremelimumab as First-line Treatment in Patients With Advanced Hepatocellular Carcinoma (HIMALAYA).


314. https://www.proteinatlas.org/ Accessed 01/08/2021


423. National Toxicology Program. Toxicology and carcinogenesis studies of methyl isobutyl ketone (Cas No. 108-10-1) in F344/N rats and B6C3F1 mice (inhalation studies). Natl Toxicol Program Tech Rep Ser. 2007;538:1-236.


