VALIDATION OF VOLATILE ORGANIC COMPOUNDS FOR THE ASSESSMENT OF LIVER DISEASE

Dr Michael Jonathan Hewitt MBBS BSc

Department of Surgery and Cancer, Imperial College London,
Hammersmith Hospital, London, UK

2023

A thesis submitted to Imperial College London for the degree of Doctor of Philosophy

Supervisors

Professor George B. Hanna PhD FRCS

Department of Surgery and Cancer, Imperial College London,
Hammersmith Hospital, London, UK

Professor Mark Thursz PhD FRCP

Digestive Diseases Unit, Imperial College London
St Mary's Hospital, London, UK

DECLARATION OF ORIGINALITY

I hereby declare that I am the sole author of this thesis and that all the work within it is my own. Any individuals who carried out collaborative work are appropriately credited.

Signed electronically.

Michael Hewitt
21st June 2023

COPYRIGHT DECLARATION

ABSTRACT

Chronic liver disease is one of the few conditions with increasing morbidity and mortality rates. Up to 75% of individuals with cirrhosis are diagnosed during a decompensation episode, at which point the prognosis is poor. Cirrhotic patients also have an annual risk of 2 to 4% of developing hepatocellular carcinoma (HCC). HCC is currently the fourth leading cause of cancer-related mortality worldwide, which is at least in part due to late diagnosis and inadequate screening. Gas chromatography-mass spectrometry (GC-MS) analysis of volatile organic compounds (VOCs) in breath has the potential to form the basis of a non-invasive diagnostic test for chronic liver disease and HCC. However, exhaled VOCs can be influenced by multiple confounding factors and the equipment used to collect and analyse breath can be cost prohibitive.

The aims of my PhD were four-fold. Firstly, to develop and validate a novel, cost-effective breath collection device and to formulate a standard operating procedure for its use in clinical studies. Secondly, to analyse the VOC profile of background room air within common clinical sampling locations and to assess their potential impact upon the collection of breath samples. Thirdly, to investigate a methodology for sample splitting using GC-MS as a way of facilitating sample analysis across multiple mass spectrometry platforms. With the information garnered from this methodology work, my final aim was to perform a clinical study to profile the VOCs in the exhaled breath of patients with cirrhosis, HCC, and normal liver parenchyma. Prior to this, I also performed a critical analysis of the pre-existing literature on VOCs for assessment of liver disease to help guide my study design.

Analysis of the novel breath collection device revealed acceptable repeatability for a wide range of VOCs and optimum settings for flow rates and volumes of breath were determined and included within a standard operating procedure. Profiling the background air volatiles in sampling locations identified specific VOC signatures for each location. Breath samples did not separate by location but monitoring of background volatiles in parallel to breath sampling remains important for identification of contaminant VOCs. Splitting of desorbed breath samples via GC-MS and recollection of two samples back on to one thermal desorption tube provides the best discrimination between samples. For my main clinical study, breath samples of 149 patients were analysed using GC-MS. Elevated levels of limonene and 2-pentanone were identified in those with hepatopathology, validating the results of previous studies. Additional VOCs were also discovered as candidate biomarkers and further studies are required to validate these findings.

The results of my clinical study have added to the existing literature that specific VOCs in exhaled breath have the potential to form a non-invasive diagnostic test for hepatopathology that could potentially help enhance earlier diagnosis of liver disease and reverse the trend in mortality rates.

PEER REVIEWED PUBLICATIONS AND PRESENTATIONS

PEER REVIEWED PUBLICATIONS

Hewitt MJ, Belluomo I, Zuffa S, Boshier PR, Myridakis A. Variation of volatile organic compound levels within ambient room air and its impact upon the standardisation of breath sampling. *Scientific Reports.* 2022; 12 (1): 15887. 10.1038/s41598-022-20365-7.

PRESENTATIONS

International:
Volatile Organic Compound (VOC) Profiling in Liver Disease via Selected-Ion Flow-Tube Mass Spectrometry (SIFT-MS), Mass Spectrometry & Advances in the Clinical Lab (MSACL), Salzburg, Austria, September 2019.

National:
Volatile Organic Compound Profiling in Liver Disease, Anatune SIFT-MS Interest Group Meeting, Cambridge, UK, July 2019.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors **Professor George Hanna** and **Professor Mark Thursz** for their guidance, support, and encouragement. Their belief in me got me through the many challenges I faced throughout my PhD journey. I am also hugely indebted to **Mr Piers Boshier**, whose guidance and never-ending enthusiasm for my research has been truly invaluable. I am also forever grateful for the day-to-day support, supervision, wisdom, and friendship provided by **Dr Ilaria Belluomo** and **Dr Antonis Myridakis**. I would not have made it to this point without you both. A special thanks also to **Professor Patrik Španěl** for your endless generosity with your mass spectrometry and VOC expertise. I must also mention the constant support and encouragement from my partner, friends, and family throughout.

I would also like to specifically acknowledge the help of the following individuals and their contributions to my research:

Dr Ilaria Belluomo – for day-to-day supervision, guidance and support with study design and instrument use.

Dr Antonis Myridakis – for guidance and support with mass spectrometry instruments, data analysis, statistical support, and general day-to-day supervision.

Dr Aaron Parker – for invaluable assistance with mass spectrometry instrument use, assistance with quality control and data processing and analysis.

Mr Oscar Ayrton – for support with the principles of mass spectrometry and laboratory workflow.

Mr Bibek Das – for the support provided with data analysis.

Miss Sara Jamel – for the support provided with data analysis.

Dr Simone Zuffa – for the support with statistical analysis of my method development work.

Dr Ameet Dhar, Mr Gary Hanh, Miss Celia Diaz Moore, Miss Rebecca Coimbra, Miss Grace Wilson, and Miss Nicola Cook – for assistance with the recruitment of patients from St. Mary's Hospital.

Dr Rohini Sharma - for assistance with the recruitment of patients from Hammersmith Hospital.

Dr Nikhil Vergis and Professor Paul Monks – for the invaluable feedback provided in my both my early and late-stage assessments.

Professor Shahid Khan - for assistance with the recruitment of patients from Hammersmith Hospital and for your time and considered feedback during my viva.

Professor Chris Mayhew – for your time and considered feedback during my viva.

Finally, many thanks to all the **patients** who kindly volunteered their time and breath for my studies.

FUNDING

The research within my PhD was supported with funding from HCA International Ltd. and by infrastructure funding by NIHR London In Vitro Diagnostic Co-operative Programme (grant reference P67522).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AFLD	Alcohol Related Fatty Liver Disease
AH	Alcoholic Hepatitis
ALP	Alkaline Phosphatase
ALT	Alanine Transaminase
ANOVA	Analysis of Variance
APRI	AST to Platelet Ratio Index
AST	Aspartate Aminotransferase
AUC	Area Under Curve
AUROC	Area Under Receiver Operating Characteristic Curve
BCLC	Barcelona Clinic Liver Cancer
BMI	Body Mass Index
CAS	Chemical Abstracts Service
CI	Confidence Interval
CLD	Chronic Liver Disease
cm	Centimetre
CRF	Case Report Form
CSLR	Calibration Solution Loading Rig
CT	Computerised Tomography
CV%	Coefficient of Variation
CV-ANOVA	Cross-Validated Analysis of Variance
DBC	Dynamic Baseline Correction
EASL	European Association for Study of the Liver
EEG	Electroencephalogram
ELF	Enhanced Liver Fibrosis Test
ETOH	Ethanol Alcohol
FDR	False Discovery Rate
FIB-4	Fibrosis-4 Index
FID	Flame Ionisation Detector
GC	Gas Chromatography
GGT	Gamma-Glutamyl Transferase
HBV	Hepatitis B Virus
HCC	Hepatocellular Carcinoma
HCV	Hepatitis C Virus
HDV	Hepatitis D Virus

HE	Hepatic Encephalopathy
HH	Hammersmith Hospital
HIV	Human Immunodeficiency Virus
HRA	Health Research Authority
HSC	Hepatic Stellate Cell
Hz	Hertz
iCCA	Intrahepatic Cholangiocarcinoma
kg	Kilogram
IMR-MS	Ion Molecule Reaction Mass Spectrometry
INR	International Normalised Ratio
LCD	Liquid-Crystal Display
LFT	Liver Function Test
Ltd.	Limited
mAh	Milliampere Hour
MELD	Model for End Stage Liver Disease
MeSH	Map Term to Subject Heading
METAVIR	Meta-analysis of Histological Data in Viral Hepatitis
MRE	Magnetic Resonance Elastography
MRI	Magnetic Resonance Imaging
MS	Mass Spectrometry
NAFLD	Non-Alcoholic Fatty Liver Disease
NASH	Non-Alcoholic Steatohepatitis
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NIST	National Institute of Standards and Technology
OPLS-DA	Orthogonal Projections to Latent Structures Discriminant Analysis
PCA	Principal Component Analysis
PET	Polyethylene Terephthalate
PHES	Psychometric Hepatic Encephalopathy Score
PPB	Parts Per Billion
PPM	Parts Per Million
PPT	Parts Per Trillion
PQN	Probabilistic Quotient Normalisation
PTFE	Polytetrafluoroethylene

PTR	Proton Transfer Reaction
PVF	Polyvinyl Fluoride
REC	Research Ethics Committee
RFA	Radio-frequency Ablation
RH	Relative Humidity
ROC	Receiver Operating Characteristic
RT	Retention Time
SACT	Systemic Anti-Cancer Therapy
SBRT	Stereotactic Body Radiation Therapy
SESI-MS	Secondary Electro-spray Ionisation Mass Spectrometry
SIBO	Small Intestinal Bacterial Overgrowth
SIFT	Selected Ion Flow Tube
SMH	St. Mary's Hospital
SOP	Standard Operating Procedure
SPME	Solid-Phase Microextraction
STARD	Standards for the Reporting of Diagnostic Accuracy
SVM	Support Vector Machine
TACE	Trans-Arterial Chemo-Embolisation
TD	Thermal Desorption
TE	Transient Elastography
TIPSS	Trans-jugular Intrahepatic Portosystemic Shunt
TNM	Tumour Node Metastasis
ToF	Time of Flight
UV	Unit Variance
VOC	Volatile Organic Compound
VOCAL	Volatile Organic Compounds for Assessment of Liver Disease
VS	Versus

SECTION 1: INTRODUCTION

1 CHRONIC LIVER DISEASE

1.1 Introduction

Chronic liver disease (CLD) is broadly defined as the deterioration of liver function over a period of at least 6 months (2). Chronic inflammation of the liver can be driven by multiple pathologies including, but not limited to, steatosis (e.g., fatty liver disease and non-alcoholic steatohepatitis), alcohol consumption (i.e., alcohol related liver disease), autoimmune conditions (e.g., autoimmune hepatitis, primary biliary cholangitis, and primary sclerosing cholangitis) and chronic viral hepatitides (e.g., hepatitis B, hepatitis C, hepatitis D, and hepatitis E). Persistent insult to the liver results in chronic inflammation which in turn leads to activation of the usually quiescent hepatic stellate cells (HSCs). Activation of HSCs results in extracellular matrix deposition and development of fibrosis (3). Early stages of fibrosis have been shown to be reversible once the insult is removed or treated due to the regenerative capacity of the liver, but ongoing insult and inflammation can result in progression of fibrosis to irreversible scarring (i.e., cirrhosis, see figure 1.1) and its associated complications including the development of portal hypertension, hepatic encephalopathy (HE), hepatocellular carcinoma (HCC) (4), intrahepatic cholangiocarcinoma (iCCA) (5) and fulminant liver failure.

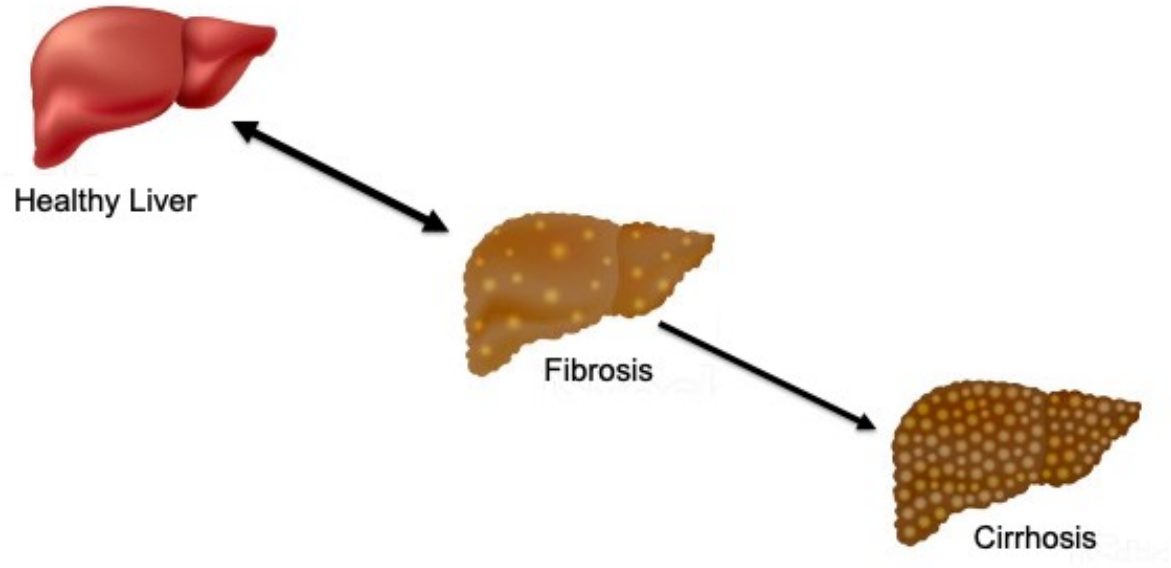


Figure 1.1 Visual representation of the progression of chronic liver disease from healthy liver parenchyma to cirrhosis.

Not all patients with the aforementioned conditions will go on to develop cirrhosis. There is an increased risk of progression to cirrhosis with advancing age, male sex, and additional

co-morbidities (e.g., HIV and hepatitis C co-infection) (6). Of those who develop cirrhosis, many will remain in an asymptomatic, "compensated" state, whereby the liver has retained sufficient function to prevent clinical manifestations of liver disease. Depending on the aetiology, 4-10% of those with cirrhosis will flip into a "decompensated" state (7) broadly characterised by the presence of ascites, hepatic encephalopathy, the development of varices, and often catastrophic variceal bleeding.

1.2 Epidemiology

While morbidity and mortality rates have been improving for many diseases in recent years, the opposite is true for CLD (figure 1.2). A rise in the incidence of obesity and alcohol consumption amongst the general population has led to a 400% increase in mortality from CLD and cirrhosis since 1970 in the UK (8) putting increasing strain on the National Health Service (NHS). Inadequacies in existing diagnostic pathways for CLD have contributed to this trend. Up to 75% of patients with cirrhosis are diagnosed during a decompensation episode, at which point the prognosis is poor (8). While established alcohol consumption will often trigger screening for liver disease in primary care, screening for hepatic steatosis is often overlooked. Those with the metabolic syndrome are often screened for cardiovascular disease but not liver disease. Earlier detection of liver disease at the fibrosis stage would allow for potential intervention (e.g., tighter diabetic control in hepatic steatosis and counselling for reduction of alcohol consumption in alcohol related liver disease) to prevent disease progression and allow possible regression.

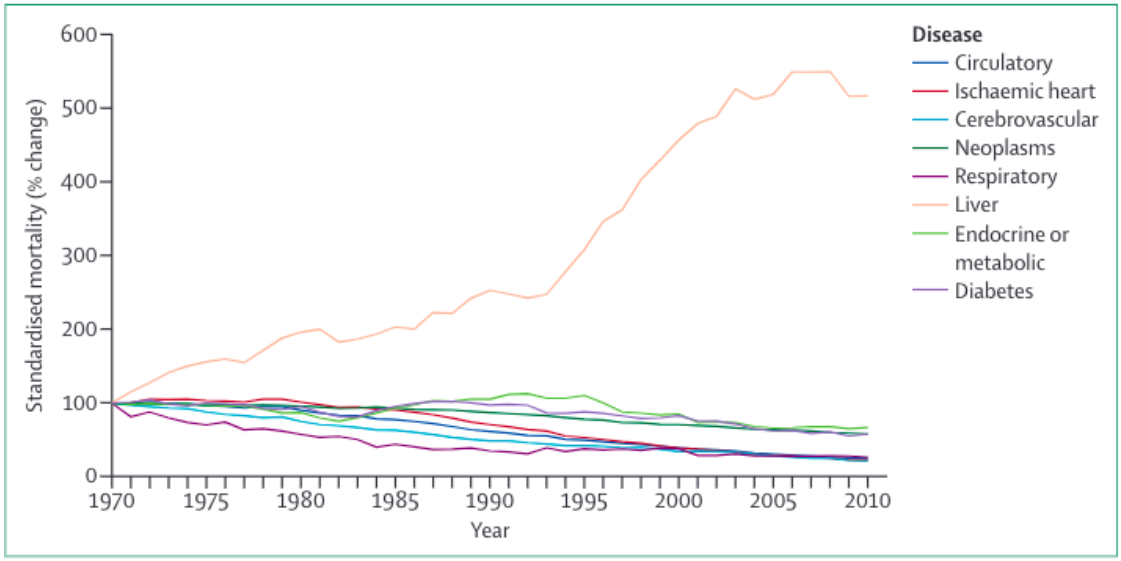


Figure 1.2 Graph of UK standardised mortality rate (SMR) data for CLD, demonstrating a significant increase in SMR since 1970 for liver disease. Reproduced with permission from Williams et al. (8). Copyright 2014 The Lancet.

1.3 Diagnosis of Chronic Liver Disease

The diagnosis of chronic liver disease is not without its challenges. There are several approaches currently used.

1.3.1 Clinical Examination

Early stage, compensated liver disease is rarely detectable on clinical assessment, but advanced or "decompensated" liver disease can be picked up by the presence of clinical signs including jaundice, ascites, skin changes (e.g., spider naevi, palmar erythema, and caput medusa), gynaecomastia, splenomegaly, and hepatic encephalopathy (manifested in its later stages as asterixis, confusion and reduced conscious level). Hepatic encephalopathy (HE) can often be detected at earlier stages using psychometric testing including the psychometric hepatic encephalopathy scoring (PHES) system. It can be used to diagnose covert HE but can take up to 20 minutes per patient to complete accurately (9) and can thus be difficult to perform within the current confines of clinical practice. A diagnosis of HE can also be achieved via an electroencephalogram (EEG), but this is not readily available in all hospitals (10) and is often reserved for individuals being considered for a trans-jugular intrahepatic portosystemic shunt (TIPSS), a procedure which aims to reduce hepatic portal pressures but as a consequence can exacerbate any underlying HE (11).

1.3.2 Histology

While histological assessment via liver biopsy is the gold standard for diagnosis and staging of fibrosis and cirrhosis, it is also invasive and flawed. Histology samples are commonly scored via the "meta-analysis of histological data in viral hepatitis" (METAVIR) staging system which gives a score from F0 (no fibrosis) to F4 (cirrhosis) (12). Liver biopsies can be performed by percutaneous or trans-jugular approaches. The latter is usually undertaken when there is concern about increased bleeding risk which is frequently present due to coagulopathy arising from liver synthetic dysfunction. Although considered the gold standard, the sample of tissue taken represents only 0.002% of the liver parenchyma (13) and as some aetiologies of chronic liver disease can cause patchy distribution of fibrosis and inflammation, it may under- or overestimate the true disease stage. It is also an invasive procedure and as such may be unpleasant for patients. Up to 80% of patients undergoing a liver biopsy reported pain during the procedure and in 20% of cases it was classified as "severe" (14). There is also a risk of infection, bleeding and a small but significant 1/10,000 mortality rate (15). As well as being invasive, liver biopsies are expensive, costing over £600 per procedure (16).

1.3.3 Radiology

While ultrasound imaging can detect changes in the liver parenchyma via its macroscopic appearance, it is not sensitive for detecting fibrosis. Transient elastography (e.g., a FibroScan®), which measures the stiffness of the liver, has been demonstrated to be a sensitive and specific diagnostic tool for the diagnosis of fibrosis but it is also not without its limitations. There can be significant inter-user variability (17) and there is reduced accuracy with increased body habitus, especially relevant for individuals with steatosis-related liver disease. It is also less sensitive at differentiating between advanced fibrosis and cirrhosis (18). Magnetic resonance elastography (MRE) has been demonstrated as having a sensitivity of up to 98% and a specificity of 100% (19) in the diagnosis of fibrosis, but it is time consuming, expensive, and not readily available to the majority of clinicians at present.

1.3.4 Blood Tests

"Liver function tests" (LFTs) are a misnomer. A standard LFT panel consists of serum levels of bilirubin and the enzymes alanine transaminase (ALT) and alkaline phosphatase (ALP). These do not provide any direct measurement of the function of the liver itself and are also non-specific for liver disease e.g., bilirubin can be elevated due to haemolysis, ALP can be elevated due to vitamin D deficiency and ALT can be raised by rhabdomyolysis amongst other causes. While these blood tests can be deranged in chronic liver disease, they can also frequently remain within normal parameters until the point of decompensation (20). They can also be elevated due to extra-hepatic disease (e.g., obstruction of the biliary tree) and acute liver injury.

Other blood tests used to assess liver function include an individual's platelet count and blood clotting profile. Those with portal hypertension (often, but not always, secondary to chronic liver disease) can develop thrombocytopenia. The mechanism of this is multifactorial including hypersplenism and reduced production of the cytokine thrombopoietin by the damaged liver (21) but once again this occurs at an advanced stage of liver disease. As the liver plays an integral role in producing clotting factors, measurements of blood clotting including prothrombin time (PT) can be used as an indicator of liver synthetic function. Elevations in a patient's PT are often found in severe liver disease but once again, it will not detect early-stage liver disease (22). Hypoalbuminaemia is common in patients with cirrhosis due to reduced production by the damaged hepatocytes and increased escape from capillaries (23). However, along with thrombocytopenia and deranged clotting, hypoalbuminaemia is non-specific and can be caused by a multitude of alternative non-hepatic pathologies (e.g., poor nutritional status).

In the absence of a single reliable blood test for measuring liver function, biomarker panels combining multiple blood tests have been developed to try and screen for fibrosis. A summary of them can be seen in table 1.1. However, they can be expensive, not readily available, may only be validated for limited aetiologies of liver disease and can be inaccurate in discriminating between early and late-stage fibrosis (24).

Table 1.1 Commercially available biomarker panels for liver fibrosis.

Test	Calculation / Components	AUROC for Fibrosis (F1-3) vs Cirrhosis (F4)
AST/ALT Ratio	AST/ALT	HCV: 0.66 (25) HBV: 0.61 (26)
AST to Platelet Ratio Index (APRI)	[(AST/upper limit of the normal AST range) X 100]/Platelet Count	HCV: 0.89 (27) HBV: 0.75 (28) NASH: 0.75 (29)
Fibrosis-4 Index (FIB-4)	(Age x AST) / (Platelets x square root of ALT)	HCV: 0.89 (30) HBV 0.93 (31) NASH 0.85 (29)
Enhanced Liver Fibrosis (ELF) Test	Hyaluronic acid, procollagen III amino-terminal peptide, tissue inhibitor of matrix metalloproteinase 1	HCV: 0.82 (30) HBV: 0.83 (32) NASH: 0.76 (33)
FibroTest™	Alpha-2-macroglobulin, Haptoglobin, Apolipoprotein A1, GGT, Bilirubin, ALT	HCV: 0.90 (34) HBV: 0.87 (35) NASH: 0.76 (36)
WFA(+)-M2BP	Wisteria floribunda agglutinin-positive human Mac-2-binding protein	HCV: 0.87 (37) HBV: 0.81 (37) NASH: 0.85 (37)

1.4 Assessing Severity of Chronic Liver Disease

There are several scores currently used in clinical practice to assess the severity of chronic liver disease.

1.4.1 Child-Pugh Score

The Child-Pugh score is the most common scoring system currently in use for chronic liver disease. It is a modified version of a score first published by Child and Turcotte in 1964 (38) which assigns a score based on several attributes as listed in table 1.2. While it was initially developed and validated to predict mortality following abdominal surgery in patients with

cirrhosis, it is now commonly used for all cirrhotic patients as a general assessment of disease state (39).

Table 1.2 Components and calculation of the Child-Pugh score.

Parameter	1 Point	2 Points	3 Points
Total Bilirubin (μmol/L)	<34	34-50	>50
Serum Albumin (g/L)	>35	28-35	<28
INR	<1.7	1.7-2.3	>2.3
Ascites	None	Mild	Moderate to severe
Hepatic Encephalopathy	None	Grade I – II	Grade III - IV

	A	B	C
Total Points	5-6	7-9	10-15
1 Year Survival (%)	100	80	45
2 Year Survival (%)	85	60	35

Limitations of the Child-Pugh score include its use of discrete rather than continuous variables, that all variables are given the same weighting and the subjectivity involved in the assessment of the presence of ascites and hepatic encephalopathy.

1.4.2 Model for End-stage Liver Disease Score

While the model for end-stage liver disease (MELD) score was developed as a tool for predicting the 3-month mortality rate for cirrhotics following a TIPSS procedure (40), it has also been shown to predict the 3-month mortality (table 1.3) for patients awaiting a liver transplant (41). It uses total bilirubin, serum creatinine and international normalised ratio (INR) in the following equation:

MELD = 3.78×ln[serum bilirubin (mg/dL)] + 11.2×ln[INR] + 9.57×ln[serum creatinine (mg/dL)] + 6.43

Table 1.3 3-month mortality based on MELD score, adapted from Wiesner et al. with permission (41).

MELD Score	3 Month Mortality (%)
<9	1.9
10-19	6.0
20-29	19.6
30-39	52.6
>40	71.3

1.4.3 United Kingdom Model for End-stage Liver Disease Score

The United Kingdom model for end-stage liver disease (UKELD) score is a modified form of the MELD score which incorporates serum sodium levels. It is used to prioritise the most unwell patients with liver disease for liver transplantation and is calculated using the following equation:

UKELD = 5×[1.5×ln(INR)+0.3×ln[creatinine (μmol/L)]+0.6×ln[bilirubin (μmol/L)]−13×ln[serum Na(mmol/L)] + 70]

A score of greater than 49 correlates with a 9% 1-year mortality rate which is the minimum requirement for transplant listing (42). While this is useful for risk-stratifying those being considered for a liver transplant, it is not a tool that has been validated for diagnosis or screening of CLD.

1.5 Hepatocellular Carcinoma

1.5.1 Epidemiology

Patients with cirrhosis, depending on the underlying aetiology, have a 2-4% annual risk of developing hepatocellular carcinoma (HCC). Those who smoke, consume alcohol, or have diabetes mellitus are at a higher risk (43). All type liver cancer is a major public health issue, ranking sixth worldwide in cancer incidence and fourth for cancer mortality (44). Patients with HCC have a mean survival of 6 to 20 months (45) and earlier diagnosis and enhanced screening programmes is vital for improving this. HCCs rarely arise from non-cirrhotic livers (less than 1% of all cases) (46). Cirrhosis can therefore be considered a pre-malignant state.

1.5.2 Diagnosis of Hepatocellular Carcinoma

Histology

Unlike the vast majority of cancers, the diagnosis of HCC does not hinge upon histology. While histological confirmation remains a gold standard for diagnosis of HCC, hallmark radiological features are frequently deemed sufficient for a diagnosis. Due to concerns regarding tumour seeding along the needle tract, biopsies are rarely done for suspected HCCs and are reserved for cases where there is diagnostic uncertainty (47) e.g., lesions arising from non-cirrhotic livers where the differential diagnosis remains broader.

Radiology

HCCs can be detected successfully on ultrasound, computerised tomography (CT) and MRI imaging. Ultrasound is less effective at picking up early-stage HCC as these lesions can appear isoechoic until greater than 1 cm (48). The sensitivity of ultrasound imaging can be increased with the use of doppler flow and contrast to further characterise the vasculature of any suspicious lesions. Any concerning lesions identified on ultrasound should go on to have cross sectional imaging for further characterisation and staging. CT imaging with contrast is a commonly used cross sectional modality for diagnosis of HCC. HCCs are hyper-vascular tumours and therefore the hallmark features expected for HCC are non-rim arterial hyperenhancement on the late arterial phase of scanning and non-peripheral wash-out appearance on the portal-venous and delayed phases. The same features can be found on contrast-enhanced MRI. MRI has an enhanced sensitivity compared to CT, especially for lesions less than 20 mm however it is more time consuming and less readily available compared to CT (49).

A summary of the HCC diagnostic algorithm produced by European Association for the Study of the Liver (EASL) can be seen in figure 1.3.

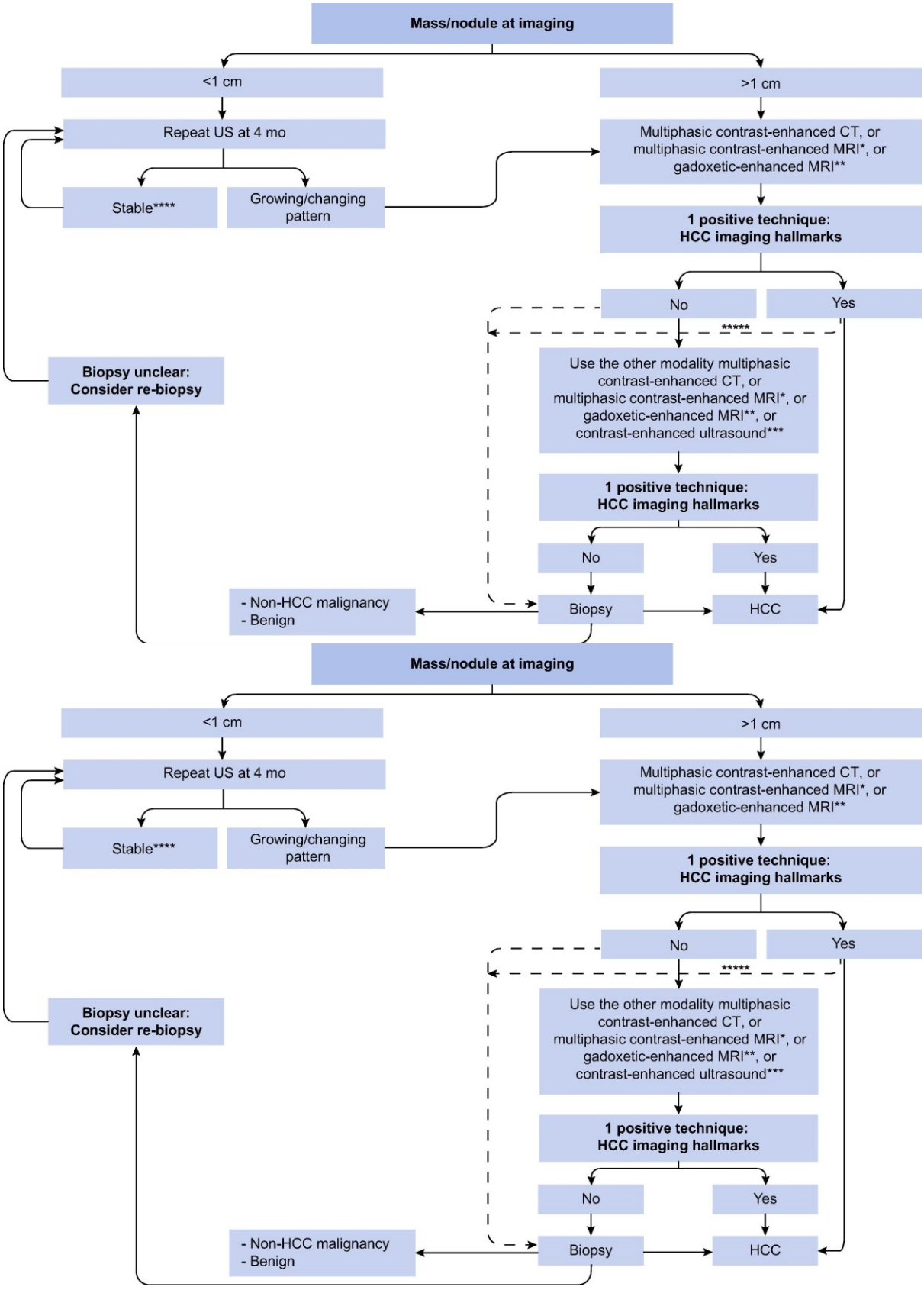


Figure 1.3 EASL's diagnostic algorithm and recall policy in cirrhotic livers. Reproduced with permission from European Association for the Study of the Liver. Copyright 2018 Journal of Hepatology (49).

1.5.3 Staging of Hepatocellular Carcinoma

HCC can be staged according to several staging systems as detailed below.

Tumour-Node-Metastasis (TNM) Staging

Like all cancers, HCC can be scored according to a TNM staging algorithm which considers the size of the tumour (T), the presence or absence of pathological lymph nodes (N) and the presence or absence of distant metastases (M). A summary of the TNM classification according to the American Joint Committee on Cancer Staging (AJCC) can be seen in table 1.4. Based on the TNM staging, the AJCC have provided prognostic groupings under their own staging system which can also be seen in table 1.4. The major limitation of the TNM staging algorithm for HCC is that it does not consider the patient's underlying liver function which can have significant impact on the treatments available and the prognosis.

Table 1.4 Classification of HCC via TNM staging algorithm and AJCC classification. Adapted with permission from College of American Pathologists (50).

T Category (pT)	
T1	Solitary, <2 cm, or >2 cm without vascular invasion
T1a	Solitary, ≤2 cm
T1b	Solitary, >2 cm without vascular invasion
T2	Solitary, >2 cm with vascular invasion; or multiple, none >5 cm
T3	Multiple, at least one >5 cm
T4	Any tumour involving a major branch of the portal vein or hepatic vein, or with direction invasion of adjacent organs other than the gallbladder or with perforation of visceral peritoneum
N Category (pN)	
N0	No regional lymph node metastasis
N1	Regional lymph node metastasis
Distant Metastasis (pM)	
M0	No distant metastasis
M1	Distant metastasis
AJCC Stage Groupings	
IA	T1a N0 M0
IB	T1B N0 M0
II	T2 N0 M0
IIIA	T3 N0 M0
IIIB	T4 N0 M0
IVA	Any T N1 M0
IVB	Any T Any N M1

Barcelona Clinic Liver Cancer (BCLC) Classification

The BCLC staging classification (table 1.5) has been accepted as the standard staging system for HCC by EASL since its publication in 1999. The BCLC classification takes a global assessment of the patient, including their performance status and underlying liver function alongside the degree of tumour burden. The main criticism of the BCLC staging system is that the intermediate 'B' stage has broad inclusion parameters, resulting in a large and heterogenous cohort.

Table 1.5 Barcelona Clinic Liver Cancer (BCLC) Classification of HCC. Reproduced with permission (51).

Stage		Performance Status	Tumour Stage	Liver Function
Early	**A1**	0	Single	No portal hypertension and normal bilirubin
	A2	0	Single	Portal hypertension and normal bilirubin
	A3	0	Single	Portal hypertension and abnormal bilirubin
	A4	0	3 Tumours <3cm	Child-Pugh A/B
Intermediate	**B**	0	Large multifocal	Child-Pugh A/B
Advanced	**C**	1-2	Vascular invasion or extrahepatic spread	Child-Pugh A/B
End Stage	**D**	3-4	Any	Child-Pugh C

1.5.4 Treatment of Hepatocellular Carcinoma

Gold standard treatment of HCC includes surgical resection or liver transplantation for those fit for surgery and with early-stage disease. Single HCCs which are less than 2 cm in size can also be successfully ablated. However, HCCs are frequently diagnosed at an unresectable stage either due to tumour size or severity of underlying liver disease. Non-operative locoregional treatment for HCC includes radiofrequency ablation (RFA), trans-arterial chemo-embolisation (TACE), and stereotactic body radiation therapy (SBRT). Systemic anti-cancer therapy (SACT) options for HCC include molecular targeted therapy

such as tyrosine kinase inhibitors (e.g., sorafenib, lenvatinib and regorafenib) and anti-vascular endothelial growth factor receptor therapies (e.g., ramucirumab) as well as immunotherapy (e.g., atezolizumab, bevacizumab, ipilimumab, nivolumab and pembrolizumab) and more traditional cytotoxic chemotherapy (e.g., doxorubicin, fluorouracil with leucovorin and oxaliplatin). Often, multiple treatments are used in combination (e.g., atezolizumab and bevacizumab) (52). For those with end-stage disease, best supportive care is the recommended course of action. EASL's treatment algorithm for HCC can be seen in figure 1.4.

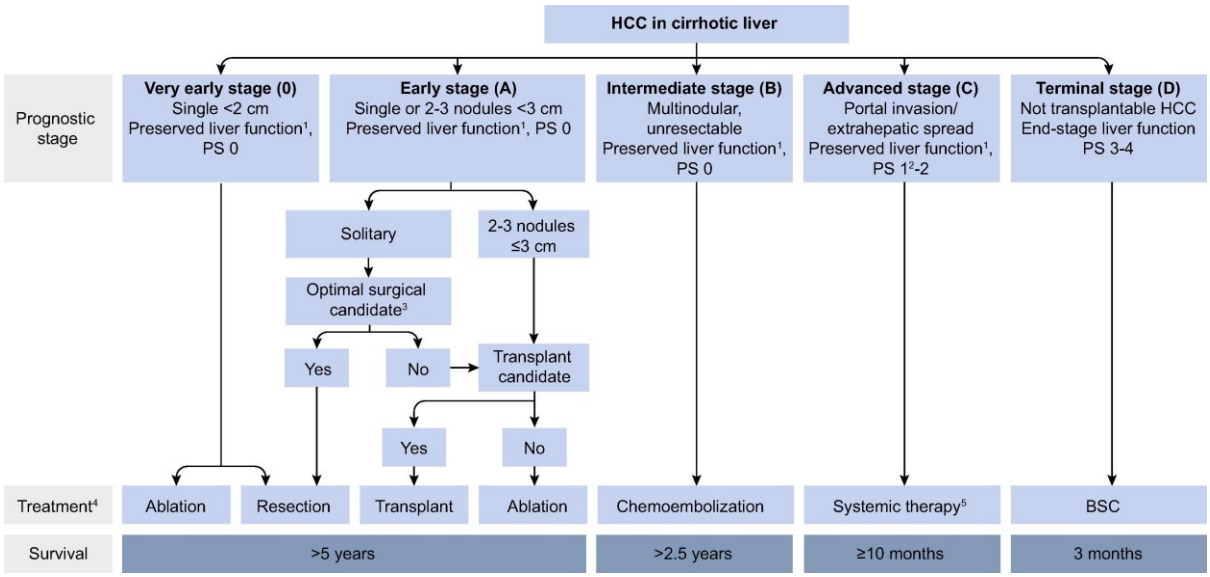


Figure 1.4 EASL HCC treatment algorithm. Reproduced with permission from European Association for the Study of the Liver. Copyright 2018 Journal of Hepatology (49).

1.5.5 Screening for Hepatocellular Carcinoma

While there is currently no national screening programme for HCC, the latest National Institute for Health and Care Excellence (NICE) guidelines recommend ultrasound imaging with or without serum alpha-fetoprotein (AFP) testing every 6 months to screen for HCC (53) in at risk patients (those with established cirrhosis or significant fibrosis in the context of chronic hepatitis B infection). However, as stated previously, ultrasound is poor at detecting at lesions less than 1 cm with a recent meta-analysis demonstrating only 47% sensitivity in detecting early-stage HCC (54). A recent nationwide survey also identified significant variation in the provision of NICE's recommendation, with many centres having no local screening programme implemented as well as difficulty in accessing radiology services to provide what was often perceived as a service that was "not a priority" (55). The same survey identified that 60% of HCC cases discussed at MDT were inoperable at the time of diagnosis.

AFP is the most widely available and commonly used serum tumour marker for HCC. It is a glycoprotein produced by the foetal liver during the early stages of pregnancy. It is present at low levels in adults but 60-80% of HCCs produce AFP and this can be detected in serum and used as a tumour marker and screening tool (56). However, as not all HCCs are AFP producers, and as other hepatic pathology can cause rises in serum AFP levels (e.g., acute hepatitides) (57), it is not a perfect biomarker. Using the well-established cut-off of 400 ng/mL, a recent meta-analysis gave a pooled summary sensitivity of 32% and specificity of 99% (57). Combining AFP with ultrasound imaging can enhance the sensitivity. Individuals with an elevated AFP but no identifiable lesion on imaging are at higher risk of going on to develop HCC and thus an elevated AFP in this context can trigger enhanced monitoring (49).

Recently, several alternative biomarkers have also been investigated for use in HCC screening and these are detailed in table 1.6.

Table 1.6 Alternative candidate serum biomarkers for HCC screening.

Biomarker	**Rationale**	**Sensitivity**	**Specificity**	**Ref.**
Des-γ-carboxy-prothrombin (DCP)	Prothrombin precursor that can be overexpressed in HCC.	0.74	0.70	(58)
AFP Lectin 3 Fraction (AFP L3%)	An isoform of AFP which is more specific to HCC based on its affinity to lectin lens culinaris agglutinin.	0.42	0.97	(59)
Alpha-L-Fucosidase	A lysosomal enzyme noted to be higher in some patients with HCC.	0.73	0.76	(60)
Glypican-3 (GPC3)	A heparan sulfate proteoglycan overexpressed in HCC.	0.55	0.58	(61,62)

None of these biomarkers have yet been incorporated into routine use for HCC screening. There thus remains a need for more robust and widely available screening tools for HCC.

2 VOLATILE ORGANIC COMPOUNDS

2.1 Introduction

Volatile organic compounds (VOCs) are carbon-based compounds that have substantial saturated vapour pressure at room temperature and can thus be detected as trace gases in human breath, stool, bodily fluids, and tissue headspace amongst other media (63). They have been of interest to researchers for several decades. In 1971, Pauling et al. (64), first reported that breath and urine contained approximately 250 and 280 VOCs respectively in normal human subjects. Since then, we now know that breath can contain thousands of VOCs (65) and the detection of VOCs in breath has become a focus of research in the hope of developing novel and non-invasive diagnostic tests.

2.2 Origins of VOCs

VOCs can broadly be divided into metabolites (either human, microbial or present in food) or xenobiotics (i.e., extrinsic to normal human metabolism). VOCs can also have endogenous and exogenous origins. Endogenous VOCs are a product of metabolic processes within the human body. For example, oxidative stress on cells (commonly as a consequence of disease and inflammation) can lead to lipid peroxidation. The process of lipid peroxidation then results in the generation of aldehydes. Aldehydes can then be metabolised to alkanes and alcohols. Trace amounts of aldehyde, alkane, and alcohol vapours can then be detected within breath (66). This is just one example of the numerous metabolic pathways within the body that can generate endogenous VOCs which can then potentially be exploited in the hunt for disease specific breath signatures.

Exogenous and xenobiotic VOCs generally result from external environments and substances including, food, drugs, and atmospheric conditions. While some exogenous VOCs may be solely contaminants from the ambient air that is inhaled, many exogenous VOCs can be absorbed by the body, metabolised, and then excreted. Thus, merely being an exogenous VOC does not exclude it from being utilised in development of breath testing for disease states. Exogenous VOCs can be metabolised differently by those with different disease states, resulting in varying concentrations of VOCs and their corresponding metabolites depending on clinical status (67-71). However, levels of exogenous VOCs can also be influenced by environmental contaminants, potentially skewing results (72,73).

To further confound the situation, some VOCs can originate from both endogenous and exogenous sources, e.g., aldehydes. As discussed previously, aldehydes can be generated

endogenously by lipid peroxidation, but they are also abundant in food and even in the inhaled air as industrial pollutants (e.g., petrol combustion) (74). This can complicate the analysis of exhaled breath VOCs as it can be challenging to accurately determine the specific origin of VOCs.

The total VOCs emitted from the human body can be referred to as the human volatilome.

2.3 Physiology of Exhaled VOCs

While VOCs can be detected from all human bodily fluids, VOCs in exhaled breath are of particular appeal. Breath is an attractive matrix as it can be collected quickly, non-invasively and in large volumes without detriment to patients. Technological developments have also facilitated real time measurement of exhaled volatiles, thus potentially giving a real time assessment of the physiological status of an individual (75).

VOCs are present in exhaled breath due to ventilation and perfusion within the lungs. VOCs circulate within the blood stream to the lungs and diffuse across the alveolar wall into the alveolar air space. They are then exhaled from the alveolar air space, through the lower and upper airways and out of the oral cavity where they detected in breath. VOCs have to pass through multiple body compartments consisting of different matrixes (e.g., tissue and blood), before reaching breath (76). The utilisation of exhaled VOCs for assessing the physiological status of the body assumes that levels of VOCs within exhaled breath correlate with the levels of VOCs within the other compartments within the body and can therefore provide a snapshot of the body's metabolism.

Historical work by Farhi in 1971 (77) on exhaled inert gases has also provided guidance on how alveolar concentrations of VOCs correlate with their equivalent mixed venous blood concentration. Based on a two-compartment model (i.e., the lungs and the rest of the body), Farhi's work dictates that the levels of gases within alveolar air depend upon the cardiac output, ventilation rate, the arterial concentration, the mixed venous concentration, the concentration of the gases in the inhaled air and the partition coefficient of the gases. The partition coefficient is the ratio of VOC concentration in one compartment (e.g., blood) to another (e.g., alveolar air space) and according to Henry's law (i.e., that solubility of a gas in a liquid is directly proportional to the partial pressure of the gas above the liquid) (78), the ratio is constant (79). To apply Farhi's work to VOCs, several assumptions have to be made: the level of the VOC in question is negligible in the inhaled air, the partition coefficients are constant according to Henry's law, the VOC does not undergo reactions with

other breath constituents, a steady state is reached within the lungs, there is uniform behaviour of the lungs, the VOC does not interact strongly with the airways and there is sampling of only end-tidal breath so that the exhaled VOC concentration mirrors the alveolar concentration (76).

Farhi's equation is defined as: $C_A = \frac{C_{\bar{V}}}{\lambda_{b:air} + \frac{\dot{V}}{\dot{Q}}}$.

Where C_A = alveolar concentration, $C_{\bar{V}}$ = mixed venous concentration, $\lambda_{b:air}$ = blood: air partition coefficient and $\dot{V}/\dot{Q}$= ventilation / perfusion (80). For VOCs with a low blood: air partition coefficient, the alveolar concentration will be close to the end tidal concentration.

As Farhi's equation demonstrates, changes in physiological status (e.g. ventilation rate and perfusion which is in turn influenced by cardiac output) can alter the alveolar levels of VOCs (and thus the concentration within exhaled breath) and an awareness of this is vital in understanding how VOCs can be influenced by disease state in order to avoid drawing false conclusions from changes in VOC levels driven by concurrent changes in other physiological parameters. However, it should also be noted that many VOCs do not fulfil all assumptions (e.g., water soluble VOCs).

A diagrammatic representation of the compartment-based model for distribution of exogenous VOCs, reproduced with permission from Amann et al. (81), can be seen in figure 2.1.

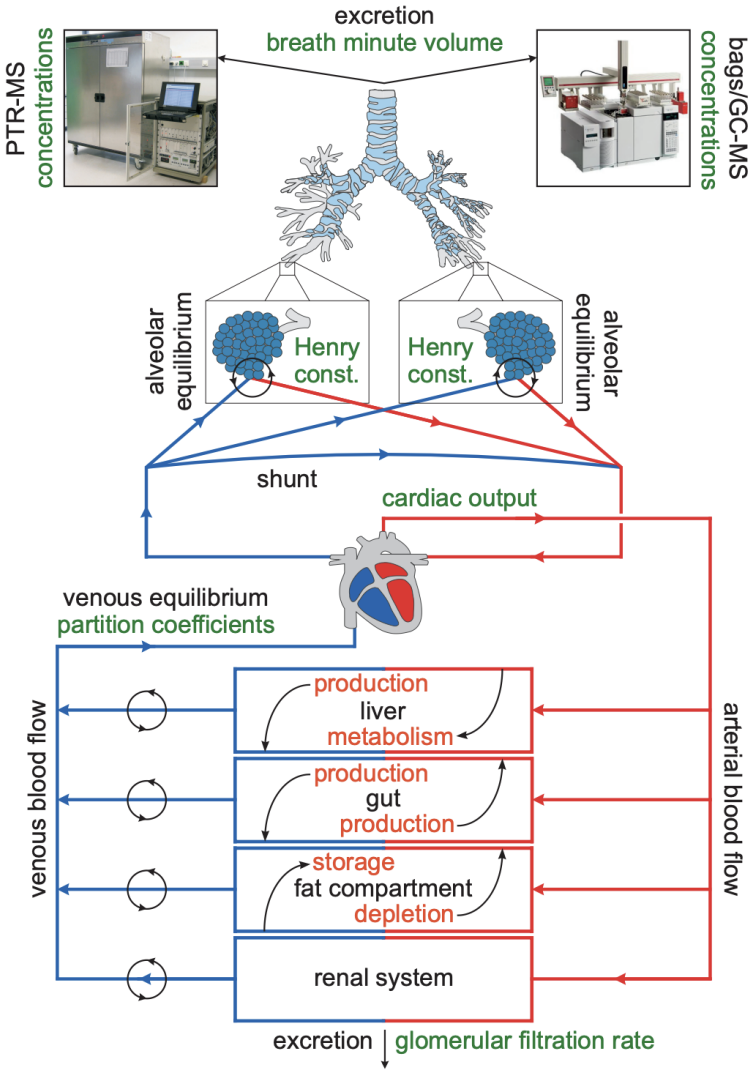


Figure 2.1 A diagram of the compartment-based model for distribution of endogenous VOCs, as published in Volatile Biomarkers - Non-Invasive Diagnosis in Physiology and Medicine by Amann et al. (81), copyright Elsevier (2013) and reproduced with permission.

2.4 VOCs and Disease State

There has been considerable research into the VOCs in exhaled breath in various disease states, often with a focus on diseases with the highest mortality rates such as cancers and respiratory diseases. A snapshot of some of the work performed to date follows.

2.4.1 Cancer

There has been extensive work to date looking at the VOC profiles in cancer patients. A review by Sutaria et al. (82) in 2022 of 44 papers published on lung cancer analysis via VOCs identified a common theme of aldehydes being elevated in those with lung cancer. They theorised that this could be accounted for by elevated levels of reactive oxygen

species within lung cancer patients, in turn reacting with unsaturated lipids to generate aldehydes. They did however note that there was significant variation between the types of aldehydes identified depending on the methods for breath collection and analysis used and that aldehydes can be associated with smoking behaviours. A review on breast cancer and VOCs by Leemans et al. (83) reviewed 32 studies, 10 of which focused on exhaled VOCs. Only one study underwent validation and once again significant heterogeneity between methodology was observed. While alcohols, ketones and alkanes were recurrent themes, only two specific VOCs (cyclohexanone and n-heptanal) were reproduced across several studies. Sixteen studies reviewed by Dima et al. (84) focusing on VOC analysis for gastrointestinal (GI) cancer found elevated levels of fatty acids along with alkanes and alcohols. Once again there was significant variation in methodology and minimal overlap for specific VOCs amongst papers.

Work by our group to date has included multi-centre studies on VOCs for diagnosis of GI cancers. Woodfield et al.'s COBRA study (85) identified elevated levels of alkanes, alcohols, esters, and sulfur-containing VOCs in those with colorectal cancer compared to those without and generated a model of 14 VOCs with an area under the receiver operating characteristic curve (AUC) of 0.87. Mina et al. (86) investigated both the exhaled breath and endoluminal air of those with oeosophago-gastric cancer and found elevated levels of volatile fatty acids in those with cancer compared to those without. Markar et al.'s (87) study on exhaled VOCs for diagnosis of pancreatic disease revealed elevated levels of formaldehyde, acetone, acetoin, undecane, isopropyl alcohol and a model generated gave an AUC of 0.736 for discriminating between those with pancreatic cancer and those without.

2.4.2 Respiratory Disease

A review published by Ratiu et al 2021 (88) identified 60 papers looking at exhaled VOCs in the context of asthma and COPD alongside lung cancer. When comparing childhood asthmatics to controls, Gahleitner et al. (89) identified a panel of 8 VOCs which could potentially distinguish between the two groups. However, further analysis of the VOCs identified suggested only 1 as a viable biomarker (2-octenal) with the rest having no logical metabolic connection. Nitric oxide is a VOC produced by airway inflammation and a high level is supportive of a diagnosis of asthma (90). Diagnosis of COPD by VOCs is limited by the impact of smoking on VOC breath profiles (see section 2.5.1) however in a study by Gaida et al., which compared the exhaled VOCs of 31 smokers with moderate COPD, 30 ex-smokers with COPD, 29 healthy smokers and 29 non-smokers, 14 VOCs were associated with COPD independent of smoking (91).

2.4.3 Infectious Diseases

Extensive work has been performed in profiling exhaled VOCs in the context of infection. Specific VOC microbial signatures have been utilised in the diagnosis and monitoring of infections as well as for assessing antimicrobial resistance (92-94). Work by Kamal et al. (95) has also raised the possibility of VOCs being used to discriminate between viral and bacterial infections to help guide appropriate use of antibiotics.

2.5 Confounding Factors for Exhaled VOCs

While differences in VOC profiles between those with and without disease is the theory which under pins the majority of research into exploiting VOCs as a diagnostic tool, there are multiple confounding factors that can influence the profile of VOCs within breath. This can make test standardisation and interpretation of exhaled VOCs challenging. These confounding factors can be broadly divided into patient-related factors and environment-related factors.

2.5.1 Patient Factors

Diet and Nutritional Status

Several studies have interrogated variability in exhaled VOC profiles with differences in diet. Biagini et al. (96) looked at the VOC profile in exhaled breath of those on a vegan diet compared to an omnivore diet and found clear separation between the two cohorts, in part driven by differences in levels of 1-propanol and several alkanes. However, they were unable to establish a definitive causative effect and postulated that lower levels of alkanes may be secondary to lower levels of oxidative stress in those following a vegan diet and a marker of overall improved health in the vegan population rather than a direct consequence of the diet itself. Baranksa et al. (97) performed a longitudinal study investigating the exhaled VOC profiles of 20 individuals when following a gluten-free then subsequently gluten containing diets. They found 12 VOCs which were significantly altered with the change in diet including 2-butanol, octane and nonanal. By using the same individual's breath while consuming both gluten containing and gluten free diets, several confounding factors are controlled for and therefore this study gives more confidence that the VOC profile changes may be caused by the change in diet.

As well as type of diet, the time since an individual last ate, as well as what was consumed can also influence levels of VOCs within exhaled breath. On the most basic level, foods recently consumed can give off extensive volatile signatures, often secondary to flavourings, which can pass directly from the GI tract into breath without needing to undergo metabolism

(98,99). Individuals who have been fasted for a period of time will have elevated levels of ketones in their breath (100) due to the reduction of carbohydrates triggering ketogenesis and the preferential metabolism of fatty acids. A “ketogenic diet” where carbohydrate consumption is limited will also trigger the same effect (101). A study by Statheropoulos et al. (102) in 2006 has further investigated the breath volatile profile after fasting by analysing the breath of a cohort of monks after a 3-day fast. The study was limited by only having seven participants and a study design that meant their breath was only sampled after their fast rather than prior to starting. VOC levels post-fasting were therefore interpreted based on the “alveolar gradient” i.e., the level of VOC in exhaled breath minus the level of VOC within the inhaled atmosphere (discussed further in section 2.5.2). The strength of this study was that all the monks followed the exact same diet, and all had no known co-morbidities. They found the VOCs with the highest alveolar gradient post fasting to be acetone, phenol, 2-pentanone, isoprene and acetaldehyde.

Efforts to mitigate the impact of diet and nutritional status on VOC profiles within studies to date have included food diaries, standardised periods of fasting, standardised meals, and oral water rinses (103). The latter aims to remove contaminant VOCs from the oral cavity.

Exercise

As discussed in section 2.3, levels of VOCs in exhaled breath can be influenced by cardiac output and ventilation rate, both of which will be impacted by any physical exertion prior to sampling exhaled breath. Isoprene, one of the most abundant VOCs in breath has been studied extensively and is a good example to demonstrate how exercise can influence levels of an exhaled VOC. The levels of isoprene detected in breath are likely secondary to release of isoprene from muscle tissue (104). When at rest, isoprene in alveolar air should be at a steady state with the pulmonary blood concentration. Isoprene has a low blood: air partition coefficient and therefore end tidal levels of isoprene should correlate with alveolar levels. As summarised in a review by Mochalski et al. (105), based on a three-compartment model by King et al. (106), isoprene levels in exhaled breath increase with exercise due to the increase in ventilation and perfusion produced by physical exertion. They then quickly fall back to baseline. With a further episode of exercise, the rise is isoprene is less due to the time taken to resynthesise the body’s isoprene stores. This supports collection of breath in a consistent and controlled manner (e.g., in a sitting position after a period of rest) as deviations from this may result in VOCs levels being different due to other physiological parameters than disease state (107).

Medications

Antibiotics, along with probiotics, are known the alter the gut flora which in turn can impact upon VOC levels (108). The bacteria which colonise the human gastrointestinal (GI) tract can produce VOCs via metabolic processes which can then either be absorbed from within the GI tract or pass directly out of the oral cavity. Therefore, by altering the GI flora with antibiotics or probiotics, it will alter the VOCs generated.

Blanchet et al. (109) performed a study looking at confounding factors on breath analysis and included medications. While there were significant differences in the breath VOC profiles of individuals taking certain medications to those not (e.g. proton pump inhibitors, corticosteroids, anti-hypertensives, and statins), it was acknowledged that, without full understanding of the underlying metabolic pathways driving changes in VOC levels, it can be hard to discriminate if a change in VOC level is due to the medication itself or the disease for which the medication is being taken.

Medications may also alter the metabolic pathways which may underly changes in VOC profiles. For example, cytochrome P450 enzymes in the liver are in part responsible for metabolism of endogenous substrates (110). They are also liable to be induced or inhibited by many common place medications (111) and can therefore theoretically increase or decrease levels of VOCs by inducing or inhibits stages in their metabolic pathways.

Smoking

It has been established that the breath VOC profiles of individuals who smoke contain much higher levels of certain VOCs compared to those that are non-smokers secondary to both the cigarette smoke itself and the metabolic impact upon the body from the oxidative stress processes triggered by the harmful constituents of cigarettes (112). Early work by Jordan et al. (113) found higher levels of acetonitrile and benzene in smokers compared to non-smokers and while the concentrations of benzene decreased rapidly once smoking stopped, it took nearly a week for acetonitrile to drop to comparable levels of non-smokers. Pauwels et al. also found individuals who smoke tobacco-based cigarettes had higher concentrations of certain VOCs in breath including toluene, ethylbenzene, furan, o-xylene, p-xylene and 2,5-dimethylfuran (114). Many smoking associated VOCs are highest immediately after of smoking and exhibit washout, whereas others remain persistently higher in individuals who smoke compared to those that do not (112). As part of a study looking at VOCs in the exhaled breath of COPD patients, Pizzini et al. excluded furan, m-xylene, 3-methylfuran, 3-penten-2-one, ethylbenzene, and toluene from their analysis as smoking related compounds

(115). Individuals who use vapes and electronic cigarettes could also be separated from tobacco smokers by esters (e.g., ethyl acetate), terpenes (e.g., α-pinene, β-pinene, d-limonene, p-cymene) and aldehydes (e.g., benzaldehyde, hexanal and decanal) based on a study by Papaefstathiou et al. (116).

Fraction of Breath Sampled

Exhaled breath can be broadly divided into end-tidal, late expiratory, and mixed expiratory portions (83). The latter, also referred to as whole breath, typically contains air from the anatomical dead space i.e., air from between the oral cavity and upper airways where no gas exchange occurs. Late-expiratory breath aims to remove the dead space air by discarding the initial portion of exhaled breath, providing a higher concentration of physiological VOCs. End-tidal breath, sometimes referred to as alveolar breath, aims to collect the breath at the end of exhalation only.

A study by Miekisch et al. suggested that collection of mixed expiratory breath could reduce VOC concentrations by up to 25% compared to techniques which minimise dead space air (117), although work performed by Doran et al. (118) found that the fraction of breath only altered the concentrations of certain VOCs. They found no difference in the concentrations of aldehydes, phenols and fatty acids in alveolar breath compared to whole breath and cyclopentane was the only VOC they identified as being different between the two portions. A review by Lawal et al. identified significant heterogeneity in both the portion of breath collected and the methods for collecting the different portions of breath (119) across studies of exhaled breath VOCs. Breath holding prior to sampling, exhalation flow rate and multiple or single exhalations have also been shown to influence the concentrations of some VOCs (120,121).

Ideally the portion of breath collected would not include anatomical dead space as this may contain VOCs from the mouth and upper airways that may not be reflective of systemic production. However, an advantage of collecting whole breath over alveolar breath is that a larger volume of breath can be collected (theoretically increasing VOC yield) without the need for additional equipment. The collection of the other portions of breath often involves specialised equipment, such as capnography which can make its use in large scale clinical studies more challenging. If capnography is not available, consistency with breathing patterns for collection is vital for sample standardisation.

Gender

A study by Das et al. compared the VOCs in the exhaled breath of biological females to biological males and could separate between sexes with 11 VOCs with AUC of 0.94 (122). However, they were unable to provide any logical explanation for the differences and were limited by a small sample size (47 participants). Lechner et el. performed a study which suggested isoprene levels were significantly higher in the male sex (123) but a more recent review of isoprene by Mochalski et al. has highlighted many other factors which will influence the level of isoprene (105). It may have been that isoprene was increased due to the higher average muscle mass in men compared to women.

Age

Španěl et al. performed a study in 2007 which suggested ammonia was higher in the breath of people aged 60-83 years compared to a group of 4–6-year-olds but acetone levels were comparable (124). However, as there were only 17 volunteers, it is not possible to draw any definitive conclusions. A 2009 study by Smith et al. (125) found concentrations of isoprene significantly increased with age and a more recent study by Mazzatenta et al. in 2015 (65) demonstrated different exhaled VOCs profile in centenarians (n = 5) compared to a group of 20-25-year-olds (n = 35) but did not comment on which VOCs were discriminatory. They also commented on different breathing patterns used by the different groups which makes interpretation of the results challenging.

While there is insufficient evidence at present to present a convincing argument that age or gender is a significant confounder for exhaled breath VOC analysis compared to the other factors discussed, it remains good practice to match control and disease cohorts appropriately where possible.

2.5.2 Environmental Factors

Ambient Room Air of Sampling Locations

Another crucial area of interest for breath analysis standardisation is the potential effect of background VOCs within the ambient room air (126). Previous studies have suggested that background VOCs can influence the levels of VOCs detected within exhaled breath (127). A study by Boshier et al. (73) in 2010 utilising selected-ion-flow-tube mass spectrometry examined the levels of seven VOCs in three clinical environments. Differing ambient VOCs concentrations were identified across the three areas which in turn raised suggestions about the ability of VOCs of high concentrations in room air to be utilised as disease biomarkers. In 2013, Trefz et al. (128) also monitored the ambient room air of an operating theatre over

the course of a working day alongside breath samples from hospital staff. They found levels of exogenous compounds such as sevoflurane had increased in both ambient room air and breath by the end of the working day raising questions as to when and where sampling of patients for breath analysis should be performed to minimise such changes. This was correlated by a study by Castellanos et al. (72) in 2016 who identified sevoflurane in the breath of hospital workers but not in that of workers outside of the hospital.

In 2018, Markar et al. attempted to demonstrate the impact of variation in room air composition on breath analysis as part of their study to assess the diagnostic capability of exhaled breath for oesophagogastric (OG) cancer (129). They utilised steel breath bags and SIFT-MS for their sampling process and identified eight VOCs within room air that differed significantly across sampling locations. These VOCs however were not included within their final diagnostic model of breath VOCs and thus their impact was negated. In 2021, Salman et al. performed a study monitoring the VOC levels across three hospital locations over 27 months. They identified 17 VOCs that acted as seasonal differentiators and proposed a cut off level of exhaled VOC concentrations above 3 $\mu g/m^3$ as being unlikely to be secondary to background VOC contamination (130). A review by Risby et al. (131) in 2006 suggested that that if the concentrations of analyte in the inspiratory air are greater than 25% of the exhaled values, the VOC should be treated with caution.

Aside from setting a cut off level or direct exclusion of established exogenous compounds, alternative methods to negate this background variation include collection of paired samples of room air at the same time as breath sampling so that the level of any VOCs present in high concentrations in the inhaled room air can be subtracted from levels found in the exhaled breath (132) providing what is referred to as an “alveolar gradient”. With this logic, a positive gradient is felt to be suggestive of an endogenous compound (133) and those with a negative gradient (i.e., higher in the room air than in breath) are felt to be exogenous compounds. Work by Unterkofler et al. (134), based on a two-compartment model, has however questioned the alveolar gradient as being too simple and generating false results. VOCs with a negative gradient may simply be stored within tissues depending on their individual physiology and fat: blood and blood: air partition coefficients (135).

Additionally, in 2013, Španěl et al. (136) performed a targeted study utilising selected ion flow tube mass spectrometry to look at the levels of seven VOCs of interest (pentane, isoprene, acetone, ammonia, methanol, and formaldehyde) in exhaled breath of 10 healthy volunteers compared to the room air that they inhaled. Based on the differences in concentration, they calculated a blood: air partition coefficient for each VOC, and with the

exception of ammonia, they found them to be consistent across all volunteers. It was proposed the partition coefficients could be used to correct exhaled breath concentrations of VOCs rather than an alveolar gradient. It is however acknowledged that the coefficients may be influenced by altered gas exchange in those with respiratory disease (e.g., chronic obstructive pulmonary disease).

Another approach is to have participants inhale "scrubbed" or filtered air that is theoretically free from contaminant VOCs (137). However, this is onerous for both participants and researchers. It can involve unpleasant face masks and can cause a dry mouth from inhaling large volumes of non-humidified air. It is also time consuming and the equipment itself can generate additional contaminant VOCs. A study by Maurer et al. in 2014 had participants inhale synthetic air. This reduced the intensity of 39 VOCs, but increased intensity of 29 VOCs compared to inhaling ambient room air (138). The use of synthetic air also significantly limits the portability of equipment for breath sampling. It is also established that the levels of VOCs within ambient air vary throughout the day (139) which could further impact upon standardisation and accuracy of breath sampling.

There is as of yet no strong consensus on the best strategy to mitigate background volatiles and for this reason, some studies choose to ignore it.

2.6 Current Clinical Uses of Exhaled VOCs

Given the issues faced in standardising breath testing and the multiple confounding factors that have been established, it is unsurprising that few VOC-based diagnostic tools have broken through the biomarker pipeline into real life clinical practice. Commercially available VOC based breath tests in current clinical practice include:

^{13}C Urea Breath Test for Helicobacter Pylori Infection

A meal of urea that has been radiolabelled with ^{13}carbon is given to a patient. Breath samples are taken prior to consumption and then 30 minutes afterwards and processed via infrared spectroscopy or less commonly, by mass spectrometry. The radiolabelled urea is metabolised by urease produced by the helicobacter pylori organisms, resulting in ^{13}carbon labelled carbon dioxide (and ammonia) that can be detected in breath. The presence of a pre-determined level of ^{13}carbon dioxide has good sensitivity and specificity for the diagnosis of helicobacter pylori infection (140).

Hydrogen/Methane Breath Test for Small Intestinal Bacterial Overgrowth

Hydrogen and methane are produced by the body via anaerobic metabolism of carbohydrates by the gastrointestinal flora (141). The presence of excess bacteria thus results in increased metabolism and increase in the associated metabolites. In the hydrogen/methane breath test for small intestinal bacterial overgrowth (SIBO), a glucose meal is given, and measurements of hydrogen and methane are taken at baseline and 90 minutes post ingestion. Rises in hydrogen of >20 parts-per-million (ppm) and in methane of >10 ppm are considered diagnostic for SIBO (142).

Fractional Exhaled Nitric Oxide Test for Asthma

Nitric oxide is a VOC produced by airway inflammation and can be useful in differentiating asthma from alternative respiratory conditions. A concentration of fractional exhaled nitric oxide (FeNO) of >50 parts-per-billion (ppb) is considered high and supports a diagnosis of asthma (90).

With the exception of the FeNO test for asthma, both the ^{13}C urea breath test for helicobacter pylori and the hydrogen/methane breath test for SIBO utilise a substrate and measure the delta change in a specific VOC in relation to this, overcoming some of the confounding factors discussed previously.

2.7 Urinary VOCs

VOCs can also be detected from urine and there has been extensive work to date looking at VOC signatures for specific diseases. A recent review by Wen et al. identified 13 published studies looking at urinary VOC analysis for diagnosis of cancer covering prostate cancer, lymphoma, lung cancer and bladder cancer (143). While arguably an easier medium to collect, there has been far fewer published studies interrogating urinary VOCs for analysis of liver disease compared to breath to date. Analysis of urinary VOCs is subject to similar challenges as exhaled VOC analysis with significant variation in urinary VOC profiles based on diet, genetics, hydration status, and the pH of the urine. Urinary VOCs are commonly analysed by headspace analysis by solid phase microextraction (SPME) or solvent extraction (81). While investigation of urinary VOCs for assessment of liver disease was considered as part of this PhD, it was ultimately not pursued due to the paucity of historical literature for urinary volatiles and liver disease, equipment availability and constraints on urinary VOC analysis method development imposed by the COVID-19 pandemic.

3 CURRENT METHODS OF ANALYSING EXHALED VOLATILE ORGANIC COMPOUNDS

VOCs within breath may occur at trace concentrations and specialised analytical techniques are required for accurate characterisation and quantification. Within our group's laboratory, we use several complementary mass spectrometry techniques for identification and quantification of VOCs. The use of multi-platform analytical methods enables both compound identification and high-throughput quantification which are very important in the search and validation of biomarkers.

3.1 General Principals of Mass Spectrometry

Mass spectrometry is an analytical tool that ionises analytes and measures the mass-to-charge (m/z) ratio and intensity of the generated ions. The nominal values and relative intensities of the mass-to-charge spectra allow for the postulation of compound identity. The stability of relative intensities also allows for accurate quantification of the analytes, particularly when used with targeted calibration standards. A typical mass spectrometer consists of an ion source, an analyser, and a detector. The ion source is responsible for ionising the sample. In the case of gas-chromatography mass spectrometry (GC-MS), this is commonly via electron ionisation (EI) whereby the analyte is bombarded by a stream of electrons generated by a thermionic emission from a filament. The electrons will displace an electron from the analyte and generate an ion. The high amount of energy imparted during EI often results in a high degree of fragmentation (144) of the analyte. In the case of proton-transfer-reaction mass spectrometry, chemical ionisation is used with a reagent (e.g., H_3O+) donating a proton. This is considered a "soft" ionisation process as it results in minimal fragmentation (145). The mass analyser sorts ions according to their m/z ratio and the detector then provides a relative abundance of each individual ion. Results are then displayed as a mass spectrum (i.e., a histogram plot of m/z ratio against intensity). Mass spectrometry can be combined with separation techniques, such as gas or liquid chromatography, to enhance specificity.

The most commonly used types of instruments for breath analysis are selected-ion-flow-tube mass spectrometry (SIFT-MS), proton-transfer-reaction mass spectrometry (PTR-MS) and gas chromatography-mass spectrometry (GC-MS).

3.2 Selected-Ion-Flow-Tube Mass Spectrometry

In selected-ion-flow-tube mass spectrometry, H_3O^+, NO^+ and O_2^+ reagent ions are generated by a microwave discharge source and filtered by a quadrupole mass filter according to their m/z and injected into a flowing carrier gas (helium or nitrogen) at an established flow rate. The analyte (i.e., breath) is then introduced to the reagent ions via a heated inlet at a known flow rate which results in a soft ionisation of the sample. A second quadrupole then analyses the product ions. By measuring the count rate of both the precursor ions and the characteristic product ions at the downstream detection system, a real-time quantification is achieved without the use of authentic standards, resulting in the absolute concentration of trace and VOCs at the parts-per-billion level. Additionally, SIFT-MS instruments contain a kinetic library of the various product ions generated by each reagent ion as well as the rate coefficient for several hundred compounds, allowing identification as well as quantification. If the desired compound is not present in the library, chemical standards can be used to confirm the identification (126). When first calibrating a SIFT-MS instrument, authentic standards should be used to confirm compound identity and improve the accuracy of identification. While untargeted analysis can be performed (known as a “full scan analysis”), the sensitivity is reduced compared to a targeted analysis (known as “selected ion monitoring”) of a smaller number of compounds.

By enabling direct analysis, the SIFT-MS technique allows real-time detection and quantification of VOCs within biological samples, such as exhaled breath, without any sample preparation, minimising diagnostic delay. This is particularly advantageous within the clinical environment where samples can be retrieved, and real-time VOC measurements made with negligible concern for sample degradation. To date, SIFT-MS has been utilised in the study of VOCs in breath and urine from patients with conditions including cystic fibrosis (146) and bladder cancer (147). When used for direct analysis, the SIFT-MS is limited by the need for an individual instrument at each recruitment site. SIFT-MS instruments on their own are unable to sample specific portions of breath accurately and generally whole breath is analysed by this technique (126).

A schematic representation of SIFT-MS can be seen in figure 3.1.

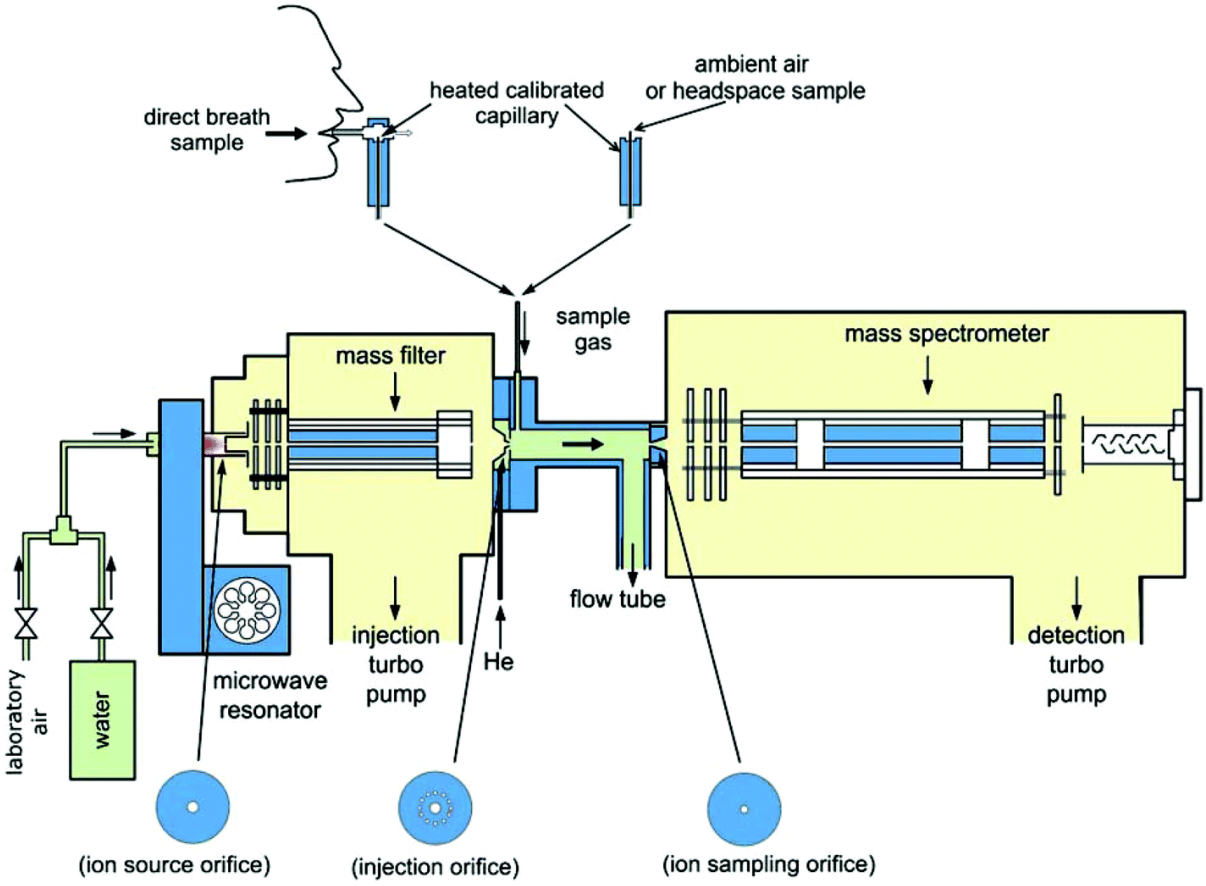


Figure 3.1 Schematic diagram of SIFT-MS technology. Reproduced from Smith et al. (148) with permission from the Royal Society of Chemistry.

3.3 Proton-Transfer-Reaction Mass Spectrometry

Proton-transfer-reaction mass spectrometry is another direct injection-based technology. In contrast to SIFT-MS, PTR-MS connects an ion source to a drift tube without a mass filter. The drift tube consists of a series of electrodes which generate an electric field which in turn draws ions along it (145). The sample enters the drift tube at a constant flow rate via a mass flow controller where it encounters reagent ions (H_3O^+, NO^+ or O_2^+) generated by the ion source. Ionisation then occurs within the drift tube. As with SIFT-MS, this is a soft ionisation process which results in minimal fragmentation of the analyte.

The ionised samples are then separated via a quadrupole analyser or a time-of-flight (ToF) mass analyser. Quadrupole analysers are scanning instruments which allow only certain ions within a specific m/z range to reach the ion detector. A time-of-flight tube separates ions by the time it takes them to pass along the tube (149). Those with lower masses will have a higher velocity and arrive quicker. Key advantages guaranteed by the ToF mass analyser are a higher mass resolution (and therefore increased confidence in identification)

and a higher spectral acquisition (i.e., an increased number of spectra collected per second) without compromising sensitivity (150).

A schematic representation of PTR-MS can be seen in figure 3.2.

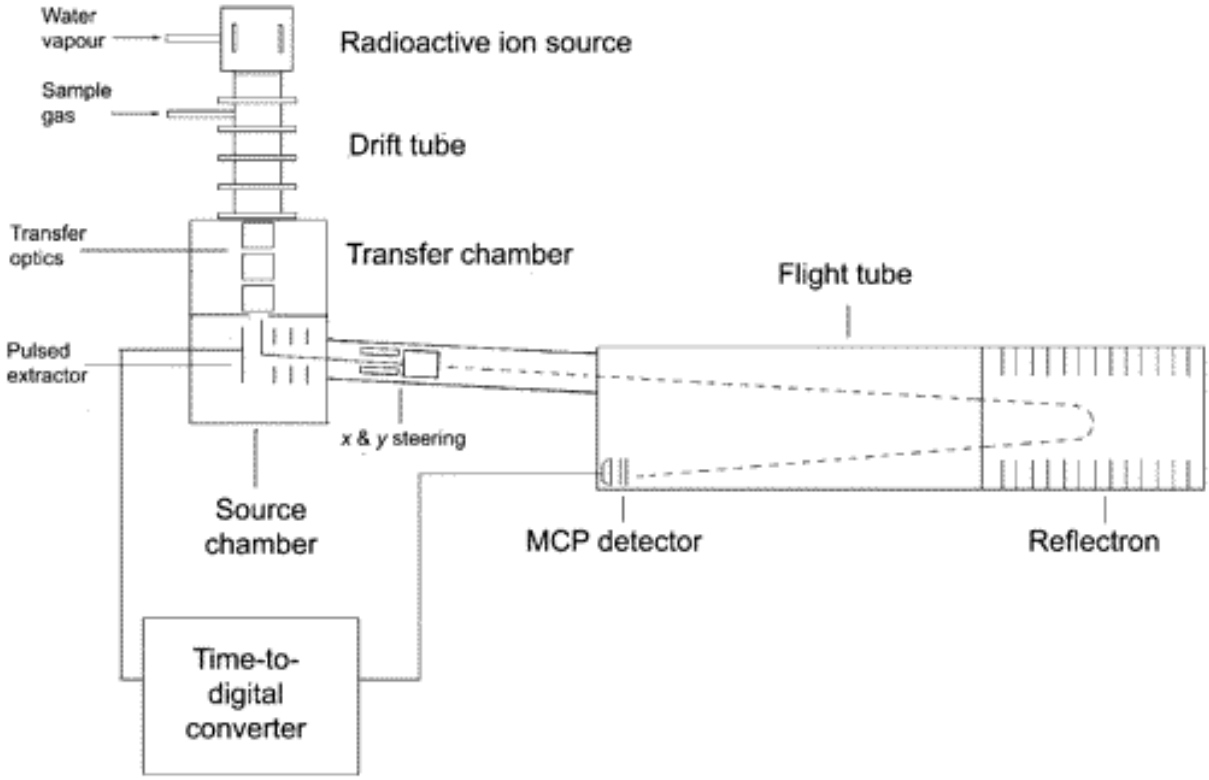


Figure 3.2 Schematic diagram of a PTR-ToF-MS. Reprinted with permission from Blake et al. (151). Copyright 2004 American Chemical Society.

3.4 Gas Chromatography-Mass Spectrometry

Gas chromatography-mass spectrometry (GC-MS) is a well-established analytical technique that allows for the successful separation, identification, and semi-quantification of trace VOCs in unknown complex matrices such as breath by combining mass spectrometry and gas chromatography. This technique has allowed for the identification of hundreds of VOCs in breath.

For breath, the sample enters a gas chromatography capillary column along with an inert carrier gas (usually helium). As the sample passes through the column, its component molecules will separate and elute from the column at different times, based on the volatility and other physicochemical properties of each molecule which influence their affinity to the column's stationary phase. Varying the chemistry of the stationary phase will affect the

relative retention of the analytes. The time taken for each molecule to pass through the column is known as its retention time (RT). The analytes then elute from the GC column into the ion source of the mass spectrometer. They are then ionised and separated by the mass analyser (usually a quadrupole or ToF analyser) (152).

A schematic representation of GC-MS (in conjunction with a thermal desorption unit) can be seen in figure 3.3.

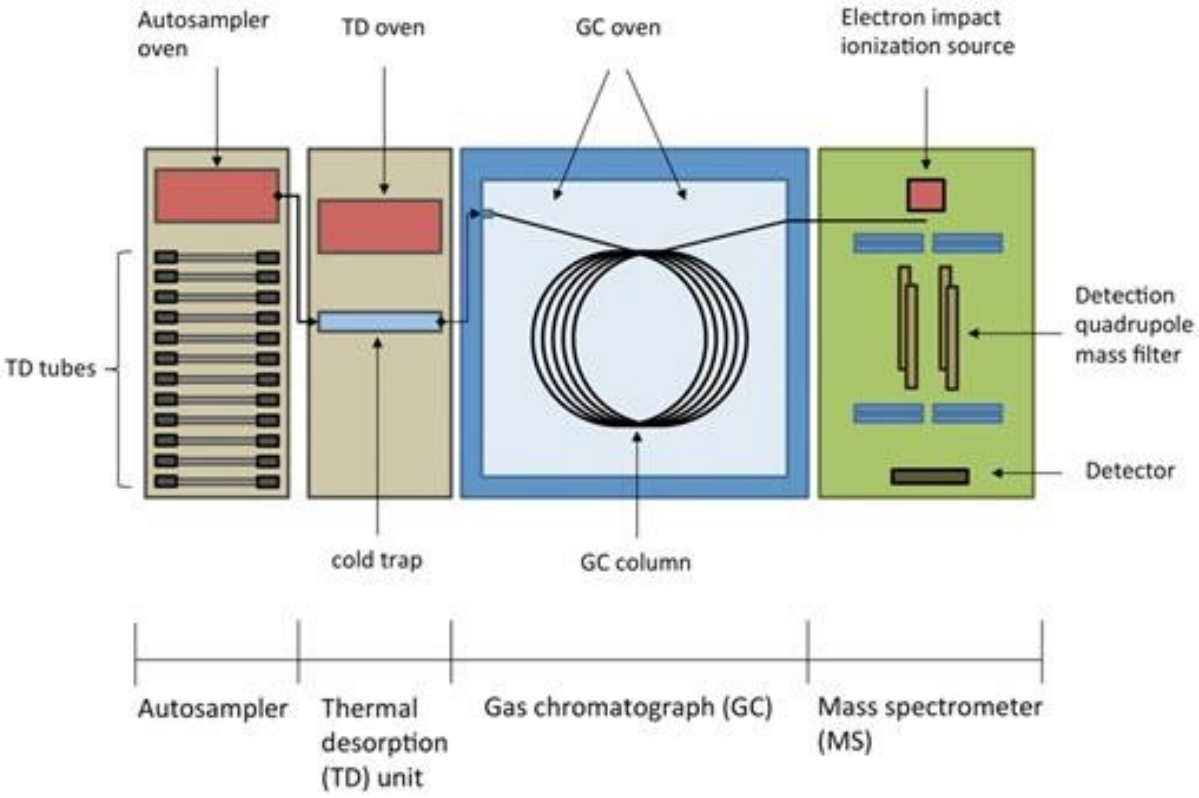


Figure 3.3 Schematic diagram of GC-MS coupled with thermal desorption. While the mass analyser featured here is a quadrupole mass analyser, a ToF analyser can also be used. Figure reproduced with permission from Materic et al. (153) Copyright 2015 Botanical Society of America.

Direct injection is challenging with GC-MS and so an intermediate medium is usually used. The most commonly used method for breath analysis with GC-MS is by coupling the instruments to thermal desorption (TD) techniques, which is discussed in further detail in section 3.6.2.

A technique which is currently growing in the field of mass spectrometry is the use of two-dimensional gas chromatography-mass spectrometry (GC×GC-MS). This connects two

separate gas chromatography columns with different stationary phases, allowing for significantly increased separation of compounds, especially of those in complex mixtures such a breath, compared to conventional GC-MS. It has been shown to detect more VOCs per breath sample than previous techniques (154) and is a valuable technique for breath analysis.

3.5 Limitations of Mass Spectrometry for VOC Analysis

Mass spectrometry analysis of VOCs is not without its limitations. The type of samples used in metabolomic studies often consists of large numbers of VOCs within complex matrices such as breath and it is therefore often not possible to identify and accurately quantify every individual VOC with one instrument. Mass spectrometers are also incredibly complex, requiring specialist knowledge to operate. They can also be liable to external influences. It therefore cannot be assumed that if the same sample is analysed on the same type of mass spectrometer in different laboratories, that the same results will be achieved. Nor if the same instrument is used for large numbers of samples, as the instruments can suffer from instrumental drifts. Instrumental drifts, whereby there is an unintentional change in the reference value with respect to which measurements are made (155), can be caused by instrument ageing and maintenance processes (156). One of the consequences of an instrumental drift is a “batch effect” (157), whereby there are variations in data sets between sample batches due to technical issues. As well as instrumental drift, there can be carry over between samples and build-up of contamination over time. Ideally, all samples for a particular study would be run together with the same experimental procedure without interruption. For large studies, this is rarely possible and therefore batch correction strategies need to be applied, such as the use of calibrations, internal standards, and inter-batch quality control samples. However, it is impractical to calibrate the analysis for every single VOC (158).

The results of mass spectrometry analysis are generated as spectra, with a series of peaks specific to ions with differing m/z ratios and a relative abundance provided by peak area. Identification of the compound correlating with the mass spectra can then be inferred by the use of mass spectral libraries, such as that provided by National Institute of Standards & Technology (NIST) (159). Ions are tentatively identified, and the confidence of the library match is given by a “match factor” (MF), “reverse match factor” (RMF) and probability (%). Generally, a match factor or reverse match factor of >800 correlates with a strong match, 700-800 is a fair match and <600 is a poor match. Therefore, while useful, peak identification often needs to be confirmed with the use of authentic standards. In order to

accurately quantify analytes, rather than relative abundance or peak area, calibration curves are required, whereby a set amount of a chemical standard is run with the instrument.

From a practical perspective, a significant limitation of mass spectrometry is the cost and portability of instruments. Mass spectrometry and the associated equipment can cost in excess of £250,000 per instrument with ongoing costs for maintenance and replacement parts. Additionally, mass spectrometry instruments are rarely portable and often have to remain at one site. For multi-centre studies, this can limit recruitment and also often requires storage of, or rapid couriering of samples, further increasing costs. These limitations should be considered when designing metabolomic studies.

3.6 Current Techniques of Breath Sampling

Current techniques of breath sampling for VOC analysis can be broadly divided into direct and indirect sampling techniques.

3.6.1 Direct Sampling

Direct, or on-line, sampling refers to the direct injection of breath into a mass spectrometer. Advantages to this approach include eliminating any sample preparation or pre-concentration steps which can compromise recovery of VOCs or introduce contaminants to the sample. It also can avoid the need for calibration curves and can provide near-immediate results. However, this approach can be impractical for clinical studies. Only certain mass spectrometry instruments can receive direct injection (e.g., SIFT-MS and PTR-MS) and it often requires the instrument to be located close to a clinical area and to be dedicated to that study only. Consequently, the cost of each mass spectrometry instrument would make multi-centre studies challenging. Although SYFT Technologies Ltd., (Christchurch, New Zealand) have developed the SIFT-MS Voice200ultra (160), a portable SIFT-MS instrument including its own carrier gas cannister which can be moved between areas, most mass spectrometry instruments capable of direct injection are not portable or compatible with use in clinical areas. There is also the risk of inter-operator variability and batch effects, which cannot be easily corrected for (161).

As well as direct exhalation of breath into the instrument via a mouthpiece, direct injection can also be achieved via transfer of samples from gas sampling bags or syringes.

3.6.2 Indirect Sampling

Indirect, or off-line, sampling refers to the process whereby breath samples are collected in to or on to an intermediate device or medium, which can then be stored or transported for analysis at a later time (119). By far the most common method of indirect breath sample collection is via thermal desorption (TD) tubes (162).

Thermal Desorption

TD tubes are generally composed of an inert stainless-steel tube packed with sorbent(s). The sorbent traps the VOCs that pass through them. Glass TD tubes are also available which benefit from excellent inertness but are liable to breakage. Large volumes of breath can be drawn across these tubes. Within my group, we use Tenax™ TA/Carbograph™ inert-coated TD tubes from Markes International Ltd. (Llantrisant, UK). These tubes contain a porous polymer sorbent that has been optimised for trapping both polar and non-polar compounds (163).

VOCs can subsequently be desorbed off the sorbent using high temperatures and an inert gas. A TD unit generally consists of an autosampler, an oven and a focusing trap. TD tubes are placed in the autosampler with diffusion locking caps. The autosampler transfers tubes to an oven which is heated to between 200°C and 400°C. The high temperature desorbs the VOCs from the sorbent within the tube. The desorbed sample is then transferred to a focusing trap (commonly a quartz capillary with a sorbent bed that is matched to the TD tube). This trap provides a preconcentration function. The focussing trap is then rapidly heated, and a secondary desorption occurs, allowing rapid transfer of the VOCs to the GC column.

TD coupled to GC-MS is one of the most widely used techniques for breath biomarker discovery work for a number of reasons. Firstly, VOCs in exhaled breath are present at very low concentrations of parts-per-million to parts-per-trillion (ppt) and pre-concentration is often necessary to allow for detection. The coupling of TD to a GC-MS system allows for the trapped VOCs to be released and rapidly injected on to the GC column for analysis. Although this technique does not allow for real-time analysis, it provides a much more comprehensive analysis of the VOC profile (152). It can therefore complement the results obtained from direct sampling high-throughput platforms, such as SIFT-MS and PTR-MS. Thermal desorption has also been successfully coupled with PTR-MS (164) and SIFT-MS (165).

Another significant benefit of TD tubes is the ability to store and transport them with relative ease, making their use in multi-centre studies convenient. However, caution should be taken with the time and conditions under which tubes are stored. While reassuringly a study by Brown et al. found that TD tubes retained >90% of the sample loaded after 4 weeks (166), a further study by Kang et al. suggested a limit of 6 weeks when stored at -80°C to avoid sample degradation and artefact. They analysed TD tubes loaded with breath at intervals of up to 12.5 months and by 12.5 months, only 41% of the VOC profile demonstrated no significant change. They also noted 33 compounds which were not present on the reference samples, suggesting accumulation of contamination of samples during storage (167).

An additional advantage of using TD is the ability to split and recollect breath samples. The TD100-xr thermal desorption unit by Markes International Ltd. (Llantrisant, UK) has the ability to split a sample and recollect a proportion of the desorbed VOCs back on to another TD tube (either the same tube or a different one). Following collection to the focusing trap, the trap is rapidly heated rapidly and a percentage of the sample is injected into the GC column with the remaining sample being recollected back to a TD tube (168,169). The default settings for our GC-MS setup are for 20% of the sample to be injected into the GC column with the remaining 80% recollected (figure 3.4).

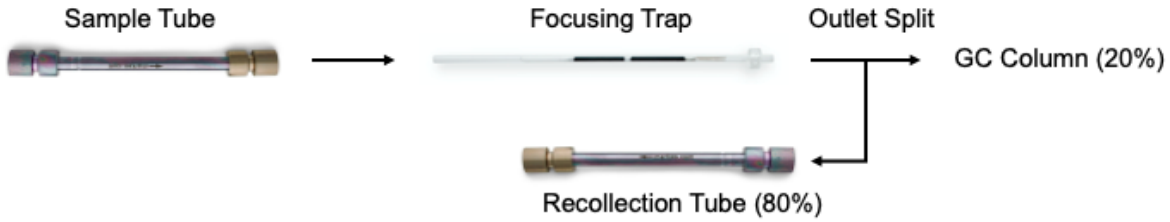


Figure 3.4 A visual representation of the ability to split a desorbed sample and recollect a proportion of the sample back on to another TD tube.

By utilising this recollection option, a single sample could then be analysed on several complementary mass spectrometry techniques, allowing a much more in-depth analysis of samples, and reducing the need to collect multiple, potentially variable, duplicate samples. For example, GC-MS methods can be based on polar or non-polar columns and are optimised for different VOCs. The same sample could now be analysed with a GC-MS using a different polarity column. However, given that most breath collection studies collect two to four samples per patient for the purpose of quality control, it remains unclear what the best recollection strategy would be. Individual samples could be recollected back on to a single

tube, or duplicates could be recollected to a single tube, potentially increasing the yield of VOCs. Another potential advantage of the recollection option is for compound identification. Recollection of multiple breath samples on to a single TD tube may increase the concentration of VOCs of low abundance, which may in turn aid the identification process.

Gas Sampling Bags

The commonest technique of loading breath on to TD tubes involves patients breathing into a bag, with the breath then transferred to the TD tubes by pump technology. Several different types of gas sampling bags have been used for breath analysis studies. Bags used for breath analysis should ideally be inert to reduce background contamination. Commercially available bags for breath analysis include:

Tedlar® Gas Sampling Bags

Tedlar® gas sampling bags are composed of a polyvinyl fluoride (PVF) film and are well validated for VOC collection, having been used in multiple studies (170) to date. They are available in several sizes and have good stability of compounds for up to six hours for the majority of VOCs (171). They do however have background concentrations of phenol and acetamide (172). They cost approximately £10 to 15 per bag.

SKC FlexFoil® PLUS Sample Bags

These are multi-layer bags consisting of nylon, aluminium, and polyethylene with a stainless-steel valve. They are useful for low parts-per-million to high parts-per-billion level VOCs and have been demonstrated to have good 48-hour stability for sulphur compounds (173).

ALTEF® Bags

These are another PVF film-based bag but have the advantage over Tedlar® bags in that they do not have background phenol and acetamide. They have polypropylene mouthpiece for ease of collection of breath. However, the bags are not recommended for collection of ketones, acetate, or hydrogen sulfide (174).

Nalophan™ Bags

Nalophan™ is polyethylene terephthalate (PET) and has been demonstrated to be a good medium for breath storage due to its inert properties and ability to provide compound stability for up to 12 hours (175).

The transfer of breath samples to TD tubes via a handpump will often require additional tubing and equipment which can also be a source of contamination for breath analysis studies. Additional disadvantages to indirect sampling via bags include potential loss of sample during each step and the additional time taken during the pre-processing steps. Gas sampling bags can also be prone to condensation which can compromise VOC recovery (176). There is also currently a push for more environmentally friendly and sustainable approaches to research and therefore single-use plastic gas sampling bags would be seen as undesirable.

Glass Syringes

Several studies have utilised glass syringes for collection of breath samples (177,178). These are often used with capnography which allows diversion of specific portions of breath (e.g., end tidal breath) based on the CO_2 level. Glass syringes provide excellent inertness and accuracy of volume but can limit the volume of breath that is collected and can be difficult to transport due to their comparative fragility.

Breath Collection Instruments

As an alternative to breath bags, there are several commercially available instruments designed to load breath directly on to TD tubes:

BioVOC® (Markes International Ltd.)

The BioVOC® is a plastic, handheld device that consists of a mouthpiece, a 100 mL chamber, and a plunger. Subjects breathe directly into the cardboard mouthpiece. The chamber is open ended, meaning only the last portion of exhaled breath will be collected and will theoretically represent a breath sample that is predominantly alveolar air. Once the subject has filled the chamber, it is capped, and a TD tube is inserted into a specially designed port. The mouthpiece is replaced with a plunger and the sample is directly transferred to the TD tube. While concerns have been raised by the limited volume of breath that is collected with the BioVOC®, a study by Kwak et al. suggested that repeated collections to the same tube to increase the volume could increase the yield of VOCs at lower concentrations (179). While the device is simple and user-friendly, there will be significant variability in the flow rate of the breath across TD tubes which can impact on compound yield and consistency across sampling episodes.

Respiration Collector for In Vitro Analysis (ReCIVA®) Sampler (Owlstone Medical Ltd.)
The ReCIVA® is a device consisting of a face mask and a unit which accommodates four TD tubes connected via polytetrafluoroethylene (PTFE) tubing to a CASPER® portable air supply which provides scrubbed air that has been passed through a carbon filter (180) to remove contaminant VOCs. Both the flow rate of breath and volume of breath collected can be adjusted using a computer connected to the device. The device also contains CO_2 sensors which helps the device collect breath during certain phases of the respiratory cycle. Consequently, the device can collect either whole breath or end-alveolar breath samples. Participants are seated and perform normal tidal breathing into the mask while inhaling scrubbed air at 40 L/min.

The ReCIVA® device, however, is expensive, limiting its use in clinical trials. Work by Doran et al. also identified several contaminant VOCs generated by the ReCIVA® (181). It is also unclear as to whether using scrubbed air is of benefit. Di Gilio et al. (162) identified comparable concentrations of common VOCs in breath collected with and without scrubbed air. They also identified higher concentrations of benzene in whole breath samples collected via the ReCIVA® compared to the background room air, suggesting minimal advantage to using scrubbed air. They also identified lower concentrations of VOCs detected by the ReCIVA® compared to alternative devices for both alveolar and whole breath. Harshman et al. (182) also identified that without manual calibration, there was variability in the flow rates across different TD tubes raising further concerns regarding standardisation of sampling with the ReCIVA®. Flow rate instability has also been correlated by unpublished work by our group.

Mistral Breath Sampler (Predict s.r.l.)
The Mistral is another breath collection device used in conjunction with TD tubes. It consists of a custom mouthpiece through which participants exhale. A volume control system allows the final 150 mL of exhaled breath to be loaded directly to TD tubes at a rate of 200 mL/min (183). There is also a heating element to maintain a constant temperature and reduce breath condensation. There is also the opportunity to sample background room air with the device. A study by Di Gilio et al. however demonstrated several contaminant VOCs from the materials the device is made from (162).

A summary of the advantages and disadvantages of the current commonly used breath sampling methods can be seen in table 3.1.

Table 3.1 Advantages and disadvantages of commonly used breath sampling methods.

Method	Advantages	Disadvantages
Direct Injection	Avoids costly TD tubes. Minimises sample loss and degradation by avoiding the need for an intermediate medium for analysis. No preconcentration steps.	An expensive instrument which needs to be dedicated to a single study. Smaller concentrations of VOCs due to lack of pre-concentration steps. Not portable. Impractical for multi-centre studies.
Gas Sampling Bags	Variable sizes. Cost effective. Simple for patients to use. Widely available. Easily adaptable to study needs.	Cannot easily discriminate between portion of breath collected. Contaminant VOCs from bag materials. Prone to condensation which can lead to VOC loss. Some materials require the bags to be cleaned prior to use. Environmental impact of single use plastics. Can require additional equipment (e.g., hand pump and tubing). Transfer from bag to instrument / TD tubes can result in loss of sample and additional contamination.
Glass Syringe	Can easily connect to capnography to accurate delineate portion of breath collected. Increased accuracy of volume of breath collected and analysed. Excellent inertness with minimal contaminant VOCs.	Impractical to transport long distance. Small volumes. Additional tubing for capnography can generate contaminant VOCs. User variability of flow rates of breath if loaded on to TD tubes.
BioVOC®	Simple to use. Portable. Theoretically designed collect alveolar air.	User variability of flow rates of the breath when loaded on to tubes. Small volume of breath without repeated collections. Potential contaminant VOCs from cardboard mouthpiece.
ReCIVA®	Use of theoretically clean “scrubbed” air. Ability to collect whole and alveolar breath. Clear instructions for patients and recruiters using custom software.	Additional sources of contamination from additional tubing and equipment. Variable flow rates across differing tubes (182). Bulky equipment which can limit

	Variable flow rates.	transportation. Cost of equipment and consumables (i.e., mask or mouthpiece) can be prohibitive for multi-centre studies.
Mistral Breath Sampler	Ability to collect whole and alveolar breath. Temperature controlled. Portable.	Contaminant VOCs from device materials. Cost prohibitive. Fixed flow rate of 200 mL/min.

3.7 Novel Breath Collection Device

Given the limitations of the current breath collection devices, there is a need to develop alternative methods of collecting breath on to TD tubes. In conjunction with Sierra Medical International Inc. (Irvine, California), our research group has developed a breath collection device that uses pump technology to load breath on to TD tubes.

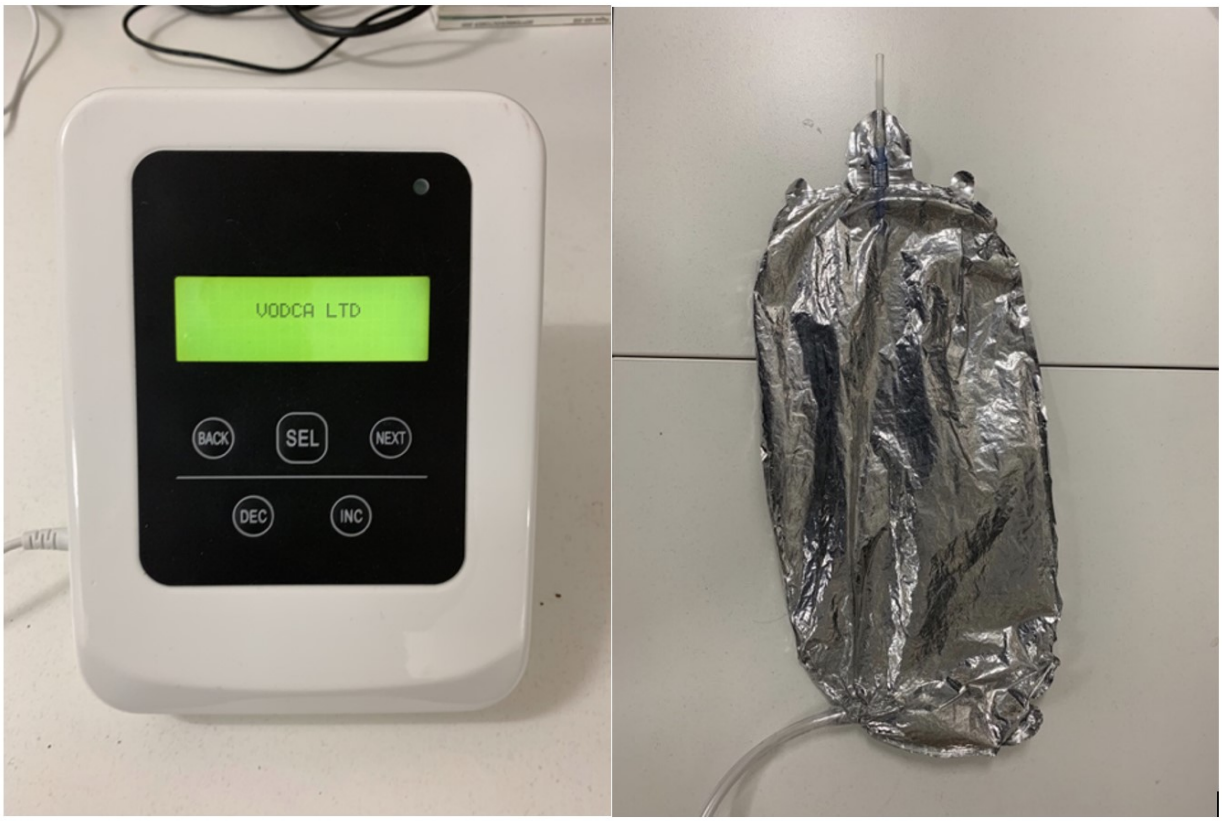


Figure 3.5 Photographs of the initial prototype of novel breath collection device and breath collection bag.

The device consists of a moulded plastic case with two air-tight sockets for individual TD tubes. A custom designed breath bag with a mouthpiece and one-way valve is attached to

the other end of the TD tubes via PTFE tubing (figure 3.5). Within the device, an electric motor draws breath from the bag through the TD tubes, one tube at a time. The device also contains an SGP30 multi-pixel metal oxide gas sensor by Sensirion AG (Stäfa, Switzerland) which provides measurements of the temperature, flow rate, humidity and CO_2 level of the breath with an accuracy of +/- 15% of the value recorded (184). As well as providing real time assurance that breath from the bag is being successfully drawn across the tubes, the data from these sensors can be extracted as part of quality control measures. While the intention at present is for whole breath to be used for analysis with the device, the bags could be adapted to allow capnography and selections of different fractions of breath if required.

Externally, a liquid-crystal display (LCD) screen provides a user-friendly interface with buttons for “BACK”, “SEL” (abbreviation for “select”), “NEXT”, “DEC” (abbreviation for “decrease”) and “INC” (abbreviation for “increase”). A flip switch at the back of the device will turn the device on and off and the LCD screen will provide step by step instructions for adjusting the settings and commencing the breath sampling process. There is the ability to adjust flow rates (from 50 mL/min up to 200 mL/min in 50 mL/min increments) and timings (from 30 s up to 300 s in 30 s increments) which then allows fixed volumes of breath to be determined. The device is powered by two rechargeable Samsung 25R 18650 batteries with 2500 mAh capacities and can also be used with a mains power supply.

The initial prototype suffered significant flow rate instability during sampling episodes. On the initiation of breath collection, there was a spike in the flow rate higher than that which it was programmed to attain. It would then take up to 30 s to stabilise and even then, would fluctuate up to 20 mL/min either side of the programmed flow rate. This was concerning for the consistency of the volume of breath loaded on to each tube. Working with the manufacturer, a new motor alongside a medical grade PET elastic buffer was introduced to the device (figure 3.6) which provided much more stable flow rates.

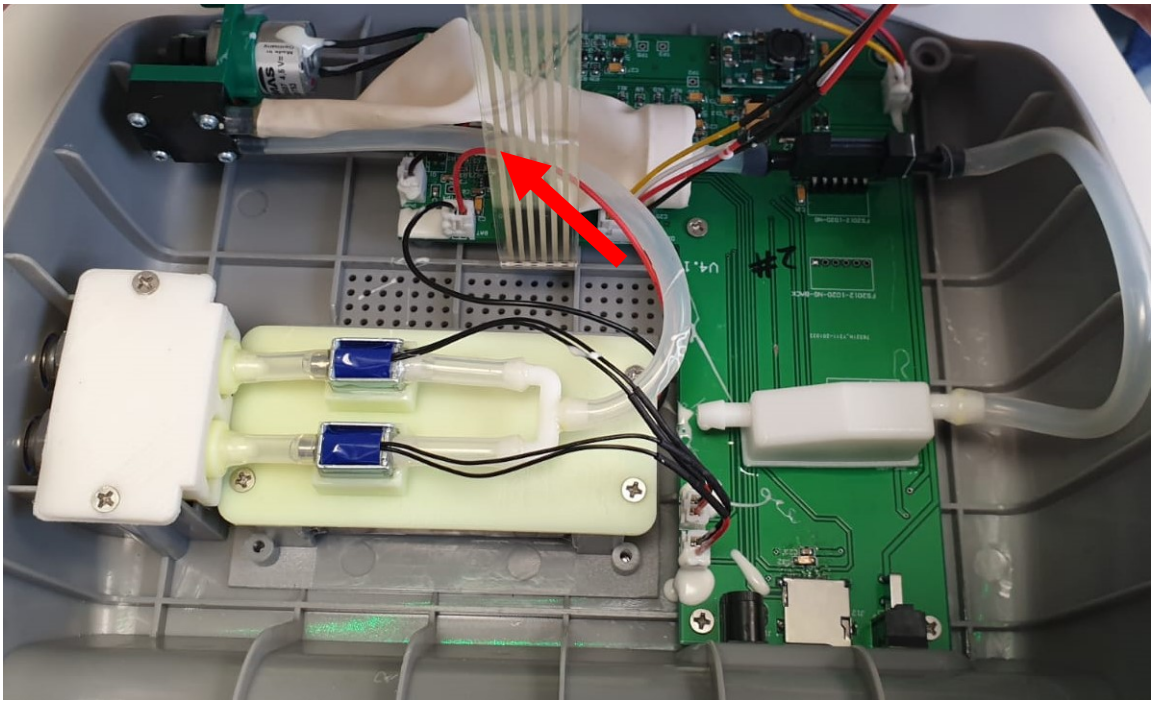


Figure 3.6 Photograph demonstrating addition of elastic buffer to stabilise flow rates.

Additionally, while it was felt that adding the option to provide patient's information into the device for storage was too complex with the limited interface, there was no way of identifying which sampling episode corresponded to which patient, making data extraction from the device difficult. Based on my feedback, the software was adjusted so a unique sampling episode identification number was given at the start of the sampling process which could be recorded with participant's clinician research form and allowing data to be matched to participants.

Alongside the initial prototype bag, several different bag designs were provided by Sierra Medical International, Inc. with different mouth pieces and attachments (figure 3.7). All the prototype bags were made with Nalophan™.

Figure 3.7 Photographs of alternative bag designs for use with novel breath collection device.

The inclusion of a one-way valve with the bag increased condensation so this was removed from the design. Following testing of the prototype bags by a member of my research group with regards to appropriateness of materials and the background VOCs they generated, a final bag design was developed. This final bag design consists of an injection moulded polypropylene mouthpiece and cap, a heat-sealed PTFE bag and polypropylene connectors for the TD tubes (figure 3.8). While Nalophan™ was initially desired as the bag material, it was difficult to heat seal and therefore the decision was made to switch to PTFE which allowed effective heat sealing while still providing excellent inertness (173). The bag has can collect volumes of breath up to 2000 mL. This bag design can also be mass produced which is integral for potential large-scale use of the device.

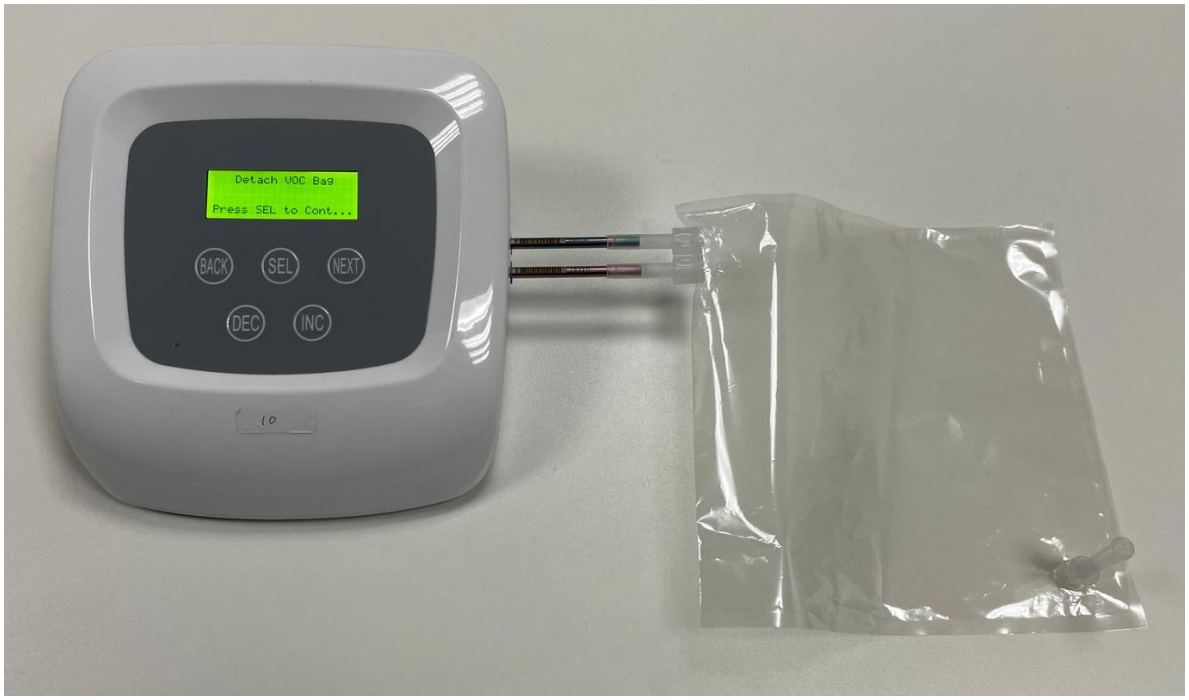


Figure 3.8 Photograph of novel breath collection device connected to two TD tubes and the final design of PTFE breath sampling bag.

3.7.1 Advantages and Limitations of the Novel Breath Collection Device

As well as being portable and easy to use, a significant advantage of the novel breath collection device is the cost. Each device costs approximately £500 to manufacture, which is significantly cheaper than the pre-existing devices. The bags designed for the device (approximately £4 manufacturing cost per bag) also represent a significant cost saving over the consumables associated with the ReCIVA®. The cost of the device however is reflected in the reduced functionality compared to alternatives. At present, with the current bags and programming, only whole breath can be collected. However, there is opportunity to develop alternative bags with connections to capnography to facilitate collection of different fractions of breath, should this be required as part of a study.

It should also be acknowledged that at present, that while temperature data is captured, there is no temperature control of the breath collection process. The current settings are also limited with regards to flow rates and timings. The latter however can easily be corrected with relatively minor software updates. The device is also designed to be used with single-use gas sampling bags which remains an issue with regards to sustainability.

4 RATIONALE FOR PHD AND THESIS AIMS

4.1 Rationale for PhD

Histological assessment of liver biopsy remains the gold standard investigation for diagnosis of chronic liver disease, cirrhosis, and hepatocellular carcinoma. However, liver biopsies are expensive, invasive, unpleasant for patients and have an associated mortality (11) . Abdominal ultrasound remains the first line modality for screening patients for hepatocellular carcinoma, but it has a low sensitivity for detecting early hepatocellular carcinoma and current UK statistics reveal that we are not diagnosing patients early enough for intervention. Thus, there is an important unmet clinical need to diagnosis of CLD and HCC earlier, as highlighted by the Lancet's report into liver disease which recommended the need to "strengthen detection of early liver disease" (8). Earlier diagnosis of liver disease will then allow for earlier referral to secondary care and with the intention of improving treatment outcomes. Ideally such a test should be acceptable to patients (e.g., non-invasive), accessible, accurate and affordable. Breath analysis has the potential to match this brief.

The concept of the character of an individual's breath being used to detect liver disease is not a novel concept. It was first suggested by Hippocrates more than 2500 years ago in ancient Greece when he described the specific sweet, ketotic smell of the breath of a patient with fulminant liver failure (referred to as "foetor hepaticus"). Since then, there has been multiple small-scale studies looking into the different VOC profiles in the breath of patients with various forms of liver disease compared to healthy controls. While the majority of these studies have focussed on the diagnosis of cirrhosis, there is a myriad of potential uses for breath testing in liver disease including:

- Screening for significant fibrosis.
- Monitoring patients with early-stage liver disease at risk of progression to cirrhosis.
- Screening for hepatocellular carcinoma.

Even if breath testing is unable to provide a definitive diagnosis, it has potential to be an invaluable tool in identifying those at risk and requiring further investigations and referral to secondary care.

Breath testing however is not without its own challenges. The multiple confounding factors that can influence the composition of breath, the variation in sampling methods and the limitations of mass spectrometry analysis techniques can distort results. The current

provision of breath sampling devices and equipment can also prohibit large scale, multi-centre studies.

My PhD aims to build on the work already performed by the Imperial College London Volatile Biomarker Group and others to try and address some of these limitations.

4.2 Thesis Aims

- To determine the repeatability of and optimum flow rate and volume settings for the novel breath collection device in order to formulate a standard operating procedure for its use in clinical studies.
- To determine the background concentrations of VOCs within common sampling locations used by my research group and interrogate their potential influence on the breath collection process.
- To determine the most effective recollection strategy for sample splitting with thermal desorption tubes analysed by gas chromatography-mass spectrometry.
- To profile the VOCs in the breath of individuals with cirrhosis and hepatocellular carcinoma compared to controls without liver disease with both direct and indirect breath sampling techniques.
- To independently validate VOCs identified in pre-existing literature as potential biomarkers for liver disease as well as identify potential novel biomarkers of liver disease.

As is a prerequisite for any clinical study, I will begin by performing a critical review of the existing literature on assessment of liver disease via VOCs to help guide the design of my experiments.

5 CRITICAL ANALYSIS OF EXISTING LITERATURE ON ASSESSMENT OF LIVER DISEASE VIA VOLATILE ORGANIC COMPOUNDS

5.1 Introduction

There have been multiple studies to date that have profiled the human volatilome in relation to liver disease. In order to help guide the design of my clinical study, I performed a systematic review and critical analysis of the pre-existing literature. While there has been a recent review of VOCs for assessment of liver disease in 2021 by Stavropolous et al. (185), my review expands on this with inclusion of more recent studies as well as providing a more critical appraisal of the proposed VOCs and of the studies' methodologies. As breath collection is not standardised, the results of breath analysis cannot be pooled and therefore a formal meta-analysis of the results was not possible.

5.2 Literature Search Strategy

I performed the search with Medline via Ovid®SP using the following search terms:

Map Term to Subject Heading (MeSH) terms
Exp volatile organic compounds/ OR
Exp breath tests/ OR
Keywords
Volatil*.ti,ab.
AND
MeSH terms
Exp liver diseases/ OR
Keywords
Hepat*.ti,ab. OR
Bili*.ti,ab.

The terms "AND" and "OR" acted as Boolean operators and the use of truncated search terms with the wild card characters allowed a comprehensive search. Results were limited to English language only. 1672 papers were identified using this search (table 5.1).

Table 5.1 Search strategy used with Medline via Ovid®SP.

	Search Term	**Number of Results**
1	exp volatile organic compounds/	13378
2	exp breath tests/	16276
3	Volatil*.ti,ab.	81562
4	1 or 2 or 3	97967
5	exp liver diseases/	623613
6	Hepat*.ti,ab.	802318
7	Bili*.ti,ab.	150440
8	5 or 6 or 7	1123272
9	4 and 8	1810
10	limit 9 to English language	1672

Papers were selected based on the criteria detailed in table 5.2.

Table 5.2 Inclusion and exclusion criteria for systematic review.

Inclusion Criteria	**Exclusion Criteria**
Assessment of liver disease or HCC via VOCs. Utilises mass spectrometry-based techniques for analysis. Adult human population (i.e., including participants >18 years).	Studies where the focus is not on VOCs in liver disease / HCC. Review articles, abstracts, and letters to the editor. STARD Score <13. Animal studies. Studies utilising sensor-based technologies (e.g., electronic-nose). Studies utilising gas chromatography techniques without mass spectrometry (e.g., flame ionisation detectors). Studies utilising ultraviolet field asymmetric ion mobility spectrometry (UV-FAIMS).

The references of all included articles were also screened for any additional papers not previously identified. Two independent reviewers (MH and CY) reviewed the titles and abstracts of all identified papers followed by the full text of papers for potential inclusion. Any disagreements over paper inclusion were resolved by a third-party arbiter (PB). The search was predominantly focussed on exhaled VOCs as they are the most commonly

researched and the intended focus of my study. However, VOCs from other sources (e.g., urine, blood, faeces, bile, and cerebrospinal fluid) were not excluded.

5.3 Outcome Measures

The following information from included articles were extracted and summarised: year of publication, country of origin, number of participants, disease type, analytical instrument used, breath collection techniques (where applicable), identity of VOCs (where available) and whether increased or decreased in disease, sensitivity and specificity, and area under the receiver operating characteristic curve.

5.4 Quality Assessment

Study quality was assessed using the STARD (STAndards for the Reporting of Diagnostic Accuracy) initiative criteria with the 25-point checklist completed for each study. A paper was deemed of adequate quality for inclusion if it scored a minimum of 13/25 in the STARD checklist (appendix 17.1). Papers which scored >19 were deemed high quality.

5.5 Pioneering Studies

Although they lie outside the inclusion criteria of this review due to an absence of mass spectrometry techniques, it is important to acknowledge the pioneering studies which laid the foundations for future work on VOCs and liver disease. Within my search, the first study to investigate VOCs in the context of liver disease dates back to 1970 when Chen et al. (186) performed a targeted study of specific volatile fatty acids in the breath of those with and without liver disease. They collected the breath of 11 individuals with decompensated cirrhosis (defined as the presence of jaundice or ascites in an individual with established chronic liver disease), 4 individuals with compensated cirrhosis and 8 control subjects. They measured the concentrations of five volatile fatty acids using gas chromatography with a flame ionisation detector (GC-FID). Attempts at standardisation included the participants fasting overnight and refraining from smoking for 2 hours prior to providing a breath sample. The study identified significantly elevated concentrations of acetic and propanoic acid in those with cirrhosis. Not all fatty acids were detected in all participants and given the small recruitment numbers, it is not possible to draw definitive conclusions from their findings.

In 1978, Kaji et al. (187) also used GC-FID to investigate volatile sulfur compounds in the breath of hepatopathic individuals. They performed a targeted analysis of dimethyl sulfide and methyl mercaptan, inspired by previous case report by Challenger et al. in 1955 which raised the possibility of these compounds contributing to foetor hepaticus (188). Kaji et al.

identified higher concentrations of dimethyl sulfide in cirrhotic individuals (20 participants) compared to controls (24 participants), but not in those with acute hepatitis (13 participants) or chronic viral hepatitis without cirrhosis (11 participants). Their samples were collected via syringe after 20 s of breath holding and there was no comment on the fasting or smoking status of those recruited. A further study with gas chromatography only by Tangerman et al. in 1983 (189) also identified higher concentrations of dimethyl sulfide in individuals with cirrhosis.

5.6 Results

5.6.1 Study Characteristics

23 papers were included in the final review. All papers were published between 1981 and 2023. A PRISMA 2020 (190) flow chart of the selection process can be seen in figure 5.1 and a summary of the papers included can be seen in table 5.3.

Study Size

The average number of participants recruited was 91 (range 26-296 participants).

VOC Matrices

The majority of studies included used breath as their VOC matrix (19/23). One study looked at VOCs from urine, one study looked at faecal VOCs, one study looked at blood VOCs and one study looked at VOCs from both blood and cerebrospinal fluid (CSF).

Mass Spectrometry Techniques

The majority of studies (14/23) utilised GC-MS techniques in conjunction TD tubes. Qin et al. (191) and Sukaram et al. (192) used SPME with GC-MS rather than TD. The remaining studies used direct injection techniques with three studies utilising SIFT-MS, three studies utilising PTR-MS and one study utilising ion molecule reaction mass spectrometry (IMR-MS).

Breath Collection Techniques

Of the 19 studies which investigated breath as their VOC matrix, 8 used gas sampling bags (either Tedlar® or Mylar®) to collect breath samples (191-199) , 3 studies used glass syringes (177,200,201) with or without a gas collection bag, 2 studies used the BioVOC® (202,203), 2 studies used the ReCIVA® (185,204), 1 study used glass vials (205), 1 study used a buffered end-tidal (BET) inlet directly into a PTR-MS (206), and 1 study used a custom built breath collection device which loaded breath directly on to TD tubes (207).

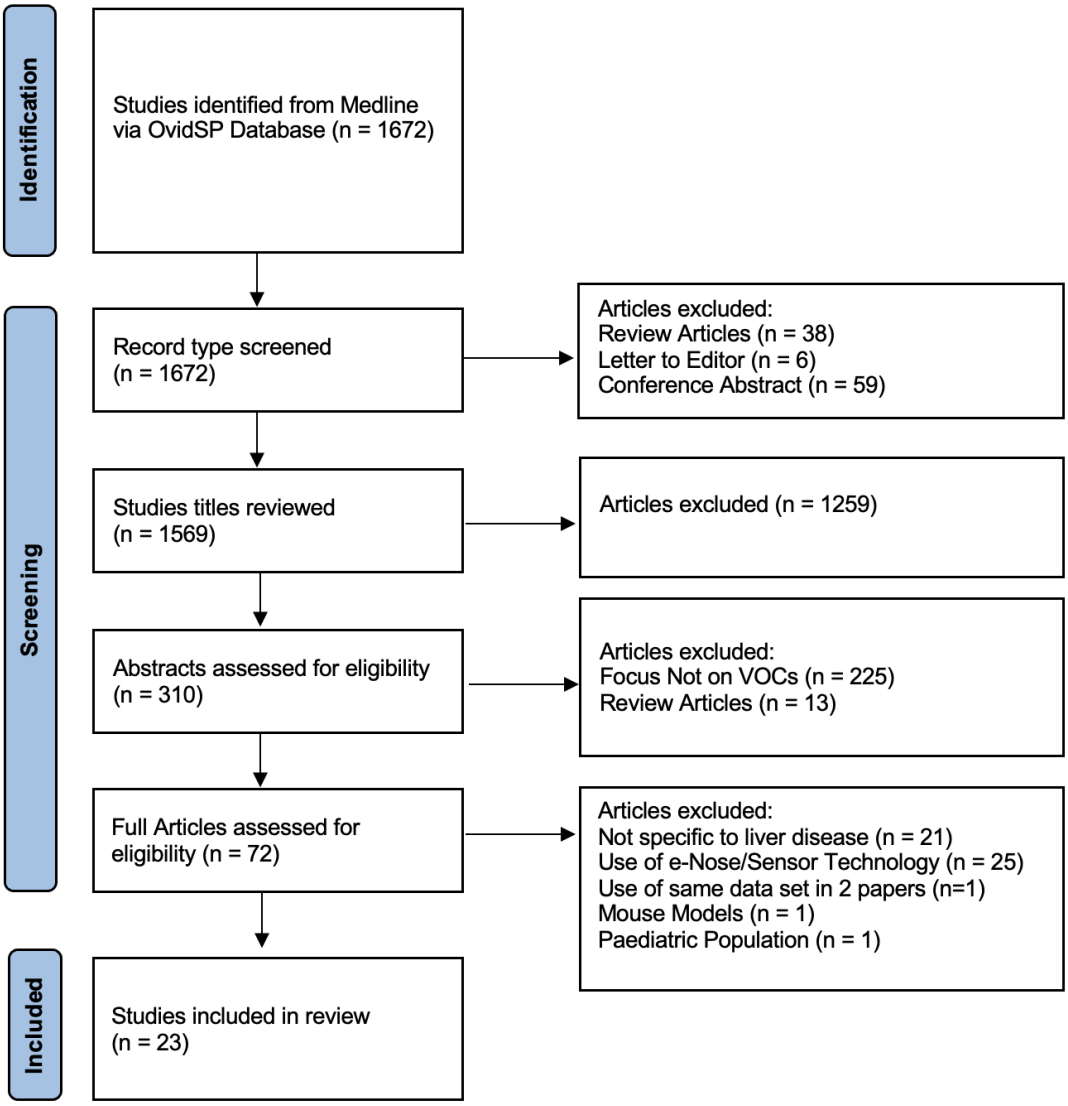


Figure 5.1 PRISMA 2020 Flow diagram demonstrating the selection of papers for inclusion in the literature review.

Fraction of Breath

Of the studies analysing VOCs in exhaled breath, 8 studies collected what was described as alveolar or end-expiratory breath with the remaining 11 studies collecting whole breath (see table 5.3). Several studies commented on how whole breath was selected for practical reasons including a lack of specialist equipment that could facilitate collection of alveolar breath.

Volume of Breath

Six studies did not comment on the volume of breath collected (191-196). For those that did, there was significant variation of the volume of breath analysed per individual. For those using direct injection techniques (i.e., SIFT-MS, PTR-MS, and IMR-MS), the average volume

of breath injected was 65 mL (range 20-100 mL). For those studies using TD-GC-MS, the average volume of breath loaded on to each TD tube was 2310 mL (range 60-5000 mL). While Sinha et al. (199) used a 10 L Tedlar® bag, they only filled this with one forced vital capacity exhalation. On average, a typical forced vital capacity will be approximately 4600 mL (208). Sehnert at el. (201) used the smallest volume of breath with 60 mL loaded from a glass syringe to each TD tube.

Flow Rates

For those studies which transferred breath on to TD tubes, several did not comment upon the flow rate with which the breath was transferred (194,197,198,201-203,207). Both studies by Ferrandino et al. (204,209) utilised the ReCIVA® with TD tubes loaded with breath at 225 mL/min. Sinha et al. (199) were the only other study to comment on flow rate and they transferred breath from gas sampling bags via a pump at 200 mL/min. Flow rates may have an impact upon the quantity of certain volatiles trapped by the sorbent tubes (181) and so it is unfortunate that this information has been omitted from so many studies.

5.6.2 Strategies for Mitigating Confounding Factors

As established in section 2.5, there are multiple confounding factors which can influence the human volatilome. While some studies made no efforts to correct for these confounding factors, several strategies were employed by other studies.

Smoking

Although the majority of studies commented on the smoking status of individuals within their participant demographics and considered this with their interpretation of their results, few studies asked participants to refrain from smoking prior to breath sampling as part of their breath sampling protocol. Exceptions to this include Millonig et al. (205) and Morisco et al. (206) who insisted that participants refrain from smoking in the 24 hours prior to collection, Sukaram et al. (192) who insisted on no smoking from the day prior to breath sampling and Hanouneh et al. (195) who excluded current smokers entirely. While the latter will result in less contaminant volatiles from smoking, it is also highly unrepresentative of populations to be recruited. Miller-Atkins et al. (196) and Verdam et al. (198) made no comment on smoking status at all in their studies. This is unfortunate given the extensive list of volatiles that have been found to be associated with smoking (115).

Exercise

Several studies (177,200,202,205) commented that individuals were in a resting physiological state i.e. seated for a minimum of 10 minutes prior to breath collection. Morisco et al. (206) took this further and asked participants to refrain from any exercise in the 24 hours prior to breath sampling. The remaining studies made no specific comment on the exercise status of those recruited.

Diet and Fasting

Nine studies (191-193,195,196,199,205-207) insisted on a period of fasting prior to breath collection. Three studies commented on an "overnight" fast (191,199,206). Alkhouri et al. (193) and Sukaram et al. (192) insisted on 8 hour fasts and Miller-Atkins et al. (196) reported those recruited were nil by mouth for 2-8 hours prior to breath sampling. They also commented that "no statistically significant differences in VOC profiles between fasting and non-fasting individuals were observed". However, it is unclear what their cut off time to define fasting was. Millonig et al. (205) and Hanouneh et al. (195) reported that subjects were "fasted" but made no comment on the duration of fasting. It is acknowledged that while fasting will reduce contamination from VOCs generated from the gastrointestinal tract, a fasting physiological status in itself will alter concentrations of VOCs as discussed in section 2.5.1.

Several studies (177,191,192,194,200-202,206) completed a food diary with participants so that any volatiles could be correlated with dietary behaviours. Sehnert et al. (201) and Van den Velde et al. (202) took this further by restricting individuals from consuming certain foods including garlic from the night before and Morisco et al. (206) gave all participants a standard meal the night before sampling. While this is commendable, it will not offset the VOC profiles influenced by long-term diet.

Oral Hygiene

It has been established that VOCs generated by bacteria within the oral cavity can influence the VOC profile of exhaled breath. A simple oral water rinse has been shown to reduce VOC contamination (103) and this strategy was employed by several studies (193,195,196,199,202).

Background Room Air Volatiles

Although established as influencing the VOC profile in breath, only six of the studies which used breath as their matrix collected room air as part of their study design. Sehnert et al. (201) collected room air samples and used them to "correct" VOC concentrations in breath

samples according to the background. It is unclear whether this referred to the alveolar gradient approach or if an alternative strategy was employed. Millonig et al. (205), Van den Velde et al. (202), Dadamio et al. (203) and Qin et al. (191) all collected room air samples and used these to generate breath VOC concentrations as an alveolar gradient (as described in section 2.5.2). Fernandez del Rio et al. (177) and O'Hara et al. (200) used a different strategy. They collected paired room air samples at the same time as breath samples but rather than use the alveolar gradient, they only used VOCs in their analysis whose ion signal intensities in breath samples were "at least twice that in the room air samples in at least half of the patients".

Other strategies to mitigate background room air volatile contamination include the use of the CASPER® scrubbed air unit in conjunction with the ReCIVA® by Ferrandino et al. for both of their studies (204,209) and the use of inhalation filters by Miller-Atkins et al. (196) and Sinha et al. (199). The remaining studies (192,194,195,197,198,206,207) made no comment on background room air volatiles.

5.6.3 Selection of Control Cohorts

Due to the various confounding factors influencing VOC profiles, selection of an appropriate "healthy" control cohort is difficult. While it is easy to match for age, race, and gender, controlling for other co-morbidities is challenging. Even if an individual has no evidence of liver disease, they may suffer from alternative pathologies which may alter their VOC profiles. The majority of studies merely commented that "healthy controls" had no evidence of liver disease, often on the basis of normal serological tests with or without ultrasound imaging. However, as discussed in section 1.5.2, it is difficult to diagnose liver disease in its early stages and LFTs alone cannot be relied upon. These individuals may have liver disease that has not been detected with the current diagnostic tools. No studies utilised transient elastography as a reference test for liver health. While histology was used in some cohorts to confirm presence of, or stage of liver disease, the use of liver biopsy to confirm healthy liver parenchyma in the context of a clinical study would be unethical given the risks and invasiveness of the procedure.

The studies by Fernandez del Rio et al. (177) and O'Hara et al. (200) should be noted for their selection of control cohorts. Where possible, they selected the partners of those with hepatopathology on the assumption they would have comparable diets and environmental exposures prior to recruitment. Where there was no partner or the partner did not fulfil the inclusion criteria, members of staff in the hospital where recruitment took place were used

instead. Miller-Atkins et al. (196) and Qin et al. (191) used a similar approach but instead of partners, asked the relatives of the participants with liver disease to act as controls. Khalid et al. (207) used hospital staff as controls to try and mitigate for environmental contaminants. Several studies (195,203,206) gave no assurance as to how their controls were selected, matched, or determined to have a healthy liver.

5.6.4 Comparisons of Hepatopathology

Mixed Type Liver Disease vs Healthy Control

The earliest paper within this review to combine gas chromatography with mass spectrometry to investigate VOCs in exhaled breath for detection of liver disease was by Friedman et al. in 1994 (194). They performed a targeted analysis of limonene concentrations in the breath of 48 participants (24 with liver disease including both cirrhotic and non-cirrhotic aetiologies and 24 controls without liver disease matched for age, gender, and race). Efforts to mitigate for confounding factors included participants not brushing their teeth on the day of collection as well as a food diary to check for any potential dietary sources of limonene. “End-expiratory” breath was collected in a 20 L Tedlar® gas sampling bag and aliquoted. It is unclear from their methods how end-expiratory breath was differentiated from whole breath. The study found elevated limonene concentrations in 50% of those with liver disease. However, following a review of the food diaries, they suggested that the elevated limonene concentrations could be influenced by dietary intake rather than disease process. They also noted higher concentrations of hydrogen sulfide in those with liver disease compared to controls. This study is limited by the heterogeneity in liver disease patients: just over half the participants had what was historically referred to as primary biliary cirrhosis (now referred to as primary biliary cholangitis) with the remaining participants’ aetiologies consisting of primary sclerosing cholangitis, viral hepatitides, and chronic hepatitides of unknown aetiology. In several of these aetiologies, the function of the liver will not be significantly impaired (e.g., chronic viral hepatitis with no evidence of cirrhosis) which may account for why only 50% of those with liver disease had elevated limonene concentrations.

In 2001, Sehnert et al. (201) used GC-MS to analyse the breath of 86 individuals with liver disease compared to 109 individuals with “normal liver function”. Whole breath was collected in gas sampling bags and 60 mL was transferred to a glass syringe with use of respiratory tubing. The breath within the syringe was then transferred to a TD tube. They did not comment on the smoking or fasting status of individuals or refer to any other attempts at mitigating for patient related confounding factors. Room air was sampled to monitor

background volatiles. As with Friedman et al., their cohort of liver disease patients was extremely heterogenous, including individuals with viral hepatitides, biliary obstruction, primary biliary cholangitis and sclerosing cholangitis. While these conditions can lead to cirrhosis, it is unclear from their study how many of these individuals demonstrated clinical or radiological evidence of cirrhosis. To try and provide more homogenous cohorts, they divided the aetiologies into “hepatocellular injury” and “bile duct damage” but the underlying liver function for each participant remains unclear. They further divided liver patients into early-, mid- and end-stage liver disease according to clinical assessment but give no information on what objective measures were used to determine these classifications. When comparing both the hepatocellular damage cohort to controls, they identified elevated concentrations of carbonyl sulfide and carbon disulfide. Carbonyl sulfide also appeared to increase with disease severity. The inverse was found with the bile duct damage cohort, with lower concentrations of both carbonyl sulfide and carbon disulfide compared to controls. They also commented on differences in isoprene concentrations, but as it has now been established, this is unlikely to be a clinically useful biomarker due to the physiological processes which can influence it (105).

In 2010 Millonig et al. (205) used IMR-MS to characterise the breath of 91 patients with liver pathology (34 with non-alcoholic fatty liver disease, 20 with alcohol related fatty liver disease and 37 with cirrhosis) alongside 35 healthy controls. Controls were defined as individuals with no history of liver disease and normal LFTs. Participants fasted and refrained from smoking from the night before and then rested for 15 minutes before providing 20 mL of breath into a glass vial via a plastic straw. Vials were frozen at -20°C prior to analysis. Room air was also sampled, and the “alveolar gradient” technique used to correct for background volatiles. Unfortunately, with their analysis technique, they could only tentatively identity of 11 of the 19 VOCs they found to be significantly different in those with hepatopathology. Comparison of all types of liver disease compared to healthy controls gave an AUC of 0.94 but they did not state which VOCs were driving the separation. When comparing only those with cirrhosis to healthy controls, a model of six VOCs gave an AUC of 0.88. Of those six VOCs, only ethanol and acetaldehyde were identified. Acetaldehyde, along with isoprene also formed part of a five VOC model which discriminated between alcohol-related fatty liver disease and healthy controls with an AUC of 0.97. When comparing non-alcoholic related fatty liver disease to controls, a six VOC model including methane and hydrogen disulphide gave an AUC of 0.96. This study is limited by a lack of identification of the other VOCs making it difficult to comment on whether the other VOCs used in their models make metabolic sense as potential biomarkers.

In 2015, Fernandez Del Rio et al. performed a two-stage study (177). In stage one, they analysed the breath of 31 individuals with all type liver disease (including a mix of cirrhotic and non-cirrhotic aetiologies as well as 10 individuals with HCC) and compared them to 30 healthy controls with alveolar breath collected via a glass syringe and capnography. Samples were analysed with PTR-MS. Smoking status and diet were also recorded although participants were not asked specifically to fast or refrain from smoking prior to sampling. Room air samples were also taken, and product ions were only considered in their analysis if the signal intensities in the breath sample were at least twice that of the room air samples in at least 50% the patients. The investigators found significant differences in eight product ion signals between disease and non-disease participants. Of these eight product ion signals, five VOCs were tentatively identified as methanol, 2-butanone, carbon disulfide, 2-pentanone, and limonene. A model including only methanol, 2-pentanone and limonene only gave an AUC of 0.95 of differentiating between liver disease participants and controls.

In stage two, they followed up 12 participants with liver disease (all but 1 of which had histologically confirmed cirrhosis) who subsequently underwent liver transplantation. They compared the pre-transplantation breath samples to post-transplantation breath samples, affording a unique opportunity to control for several of patient-related confounding factors that can influence VOC concentrations. They found post-transplantation concentrations of the five significant VOCs to be lower than pre-transplantation, with limonene showing the highest decrease for the majority of participants. Limonene also demonstrated a longer wash-out period: while the other five VOCs reached a new baseline at the first measurement post-transplant surgery, limonene concentrations continued to gradually decrease over approximately 2 weeks. This washout period was felt secondary to the lipophilic nature of limonene, thereby taking time for the newly transplanted liver to metabolise the limonene stored within adipose tissue.

Cirrhosis vs Healthy Controls

In 2008, Van den Velde et al. (202) analysed the breath of 52 cirrhotic patients of mixed aetiologies and compared them to 50 age-matched healthy controls. Participants were not fasted but were asked to refrain from consuming garlic, spicy foods, and alcohol in the 24 hours prior to sampling. While smoking status was noted, it is unclear how long it was between a participant's last cigarette and the sampling episode in those who were active smokers. After a mouth rinse, alveolar breath was collected with a BioVoc®. The samples were analysed with GC-MS. Room air was also sampled but it is unclear from the methods

how these results were used. The investigators found elevated concentrations of dimethyl sulfide, acetone, 2-butanone and 2-pentanone and decreased concentrations of indole and dimethyl selenide in those with cirrhosis compared to the control cohort. A model developed with these compounds gave a sensitivity of 1.00 and specificity of 0.70. They also noted that concentrations of 2-pentanone appeared to positively correlate with participants' MELD scores, raising the possibility that 2-pentanone could be used to assess severity of liver disease.

In 2011, Dadamio et al. (203) also used the BioVOC® to collect the alveolar breath of 35 individuals with histologically proven cirrhosis alongside the breath of 49 healthy controls with analysis via GC-MS. It is unclear from the study how healthy controls were defined and there is no comment on any attempts to mitigate patient related confounding factors. Room air samples were taken and used to calculate the alveolar gradient for VOCs of interest. One of the strengths of this study was the inclusion of an additional 12 individuals with cirrhosis to act as a small validation cohort. The investigators found 28 VOCs that were significantly different between disease state and controls. From this, they developed 24 potential models of 8 different VOCs, with the VOCs most frequently incorporated being acetone, styrene, dimethyl sulfide, dimethylselene and an unidentified branched alkane. Sensitivities of the models generated ranged from 0.82-0.88 and specificities from 0.96-1.00. Using up to eight VOCs in a model does risk overfitting (210) and it should be noted that many of the VOCs, including those highlighted, were not found in a large percentage of participants. For example, dimethylselene was only detected in 9% of healthy controls and styrene was only detected in 22% of cirrhotics. While these differences may be accounted for by the presence or absence of liver disease, the complete absence of many supposedly discriminating VOCs in such a large proportion of participants raises concerns about the accuracy of the analytical techniques used and the validity of these VOCs as biomarkers.

In 2013, Morisco et al. (206) published the smallest study within the literature review with only 26 participants (12 cirrhotics of mixed aetiologies and 14 healthy controls). They collected whole breath via a BET inlet attached to a PTR-MS with no preconcentration steps. Participants were fasted and asked to avoid smoking, chewing gum, mouthwash, alcohol, coffee, and pungent foods (e.g., garlic and mint) from the evening before as well as avoiding exercise in the preceding 24 hours. Background volatiles within room air do not appear to have been considered. The investigators found 12 VOCs at significantly different concentrations in the exhaled breath of cirrhotics compared to healthy controls including elevated concentrations of 2-butanone, 2-pentanone, an unidentified monoterpene, an unidentified terpene and several unidentified sulfur and nitrogen compounds. The

unidentified monoterpene, possibly limonene, provided the best sensitivity (0.83) and specificity (0.86) for discriminating between the two groups. As well as comparing all cirrhotics against healthy controls, they also compared compensated cirrhotic patients (broadly classified as Child-Pugh A cirrhotics) with decompensated cirrhotic patients (broadly classified as Child-Pugh B and C cirrhotics) and found that the monoterpene and an unidentified nitrogen compound could differentiate between Child-Pugh A cirrhotics and Child-Pugh B and C cirrhotics. This study is limited by the small number of participants and a lack of confidence with the identification of the proposed VOCs.

Also in 2013, Khalid et al. recruited 69 participants (47 cirrhotics of mixed aetiologies, 7 participants with a history of alcohol excess but no evidence of cirrhosis and 15 healthy controls). Cirrhosis was confirmed histologically in the majority of cases (41/47). Controls consisted of hospital staff with no history of liver disease who were matched for age. Participants were nil by mouth for a minimum of 1 hour prior to sampling. Breath was collected with a custom designed device that loaded alveolar breath directly on to TD tubes. Background volatiles within room air were not considered. TD tubes were analysed within 18 hours by GC-MS. The investigators found that o-cymene, methyl vinyl ketone and an unidentified VOC could discriminate non-alcohol related cirrhosis patients from healthy controls. Those with alcohol-related cirrhosis could be discriminated from healthy controls by heptane, o-cymene, phellandrene, and 2-methylhexane (207).

In 2014, as part of a study looking at alcoholic hepatitis with background cirrhosis versus decompensated cirrhotics (participants with cirrhosis of mixed aetiologies with the decompensated episode triggered by causes other than alcohol), Hanouneh et al. (195) also used the opportunity to compare the exhaled VOC profiles of all individuals with cirrhosis. Individuals were fasted and performed an oral mouth rinse with water prior to breath sampling. 40 individuals with alcoholic hepatitis and 40 participants with decompensated cirrhosis were compared to 43 controls without liver disease with whole breath collection into a Mylar® bag which was then analysed with SIFT-MS. As part of a targeted analysis of 14 hepatopathology related VOCs identified in previous studies, they found elevated concentrations of 2-propanol, acetaldehyde, acetone, ethanol, pentane, and trimethylamine (TMA). Background volatiles in room air were not considered in this study.

In 2020, as part of a larger study including non-hepatic pathology, Miller-Atkins et al. (196) looked at 30 cirrhotics and compared their breath to 54 healthy controls (consisting of the relatives of the cirrhosis cohort). While there was a significant variability in the fasting status of those recruited from 2-8 hours, they did not detect any significant difference between

individuals with differing fasting durations. Following an oral water rinse, individuals inhaled ambient air through an acid gas cartridge which aims to filter out contaminant volatiles (including chlorine, sulfur dioxide, chlorine dioxide, hydrogen chloride, hydrogen fluoride, and formaldehyde) (211) before exhaling whole breath into a Mylar® bag. Breath was then analysed with SIFT-MS. With the exception of the acid gas cartridge filter, room air volatiles were not considered. A targeted analysis of 22 VOCs identified several which were significantly different between the two groups, with methylhexane, decene and acrylonitrile providing the best discriminatory ability. Carbon disulfide, trimethylamine, nonene, pentane and propanol were also elevated in cirrhotics compared to controls. Benzene was included in their list of significant VOCs but there is no comment on the smoking status of the individuals recruited. While the study reported sensitivity and specificity of detecting cirrhosis compared to all other participants (including those with HCC, pulmonary hypertension, and colorectal cancer with hepatic metastases), it did not report a sensitivity or specificity for discriminating between cirrhotics and healthy controls only.

At the time of writing, the most recently published work on exhaled VOCs for detection of cirrhosis is by Ferrandino et al. in 2023 (204). They built upon their group's previous work in 2020 where they investigated the exhaled VOCs of with HCC, cirrhosis, and healthy controls (209). They once again used the ReCIVA® in conjunction with a CASPER® air supply to collect the breath of 46 individuals with cirrhosis, 14 of which also had an HCC, alongside 40 healthy controls without liver disease. Individuals were not fasted. All samples were analysed with GC-MS within 4 weeks of collection. The best performing VOCs for discriminating between cirrhosis and controls were 2-pentanone, 1-hexene, indole, dimethyl selenide, limonene, eucalyptol, and benzene, (1-propylnonyl)-. While individual VOCs gave AUCs of 0.76-0.82, a panel of all 7 VOCs gave an AUC of 0.95 when used with a test set. This many VOCs within a model will be prone to overfitting and may represent noise between the two groups rather than true differences. While several VOCs are reproduced from previous studies and have potential logical underlying metabolic pathways (e.g., limonene, 2-pentanone), the same cannot be said for benzene, (1-propylnonyl)- or indole.

Cirrhosis vs Non-Cirrhotic Chronic Liver Disease

While the majority of studies looking at individuals with cirrhosis compared them to controls without liver disease only, Pijls et al.'s 2015 study (197) compared the VOCs in the exhaled breath of cirrhotics to those with non-cirrhotic chronic liver disease as well as controls with no history of liver disease using GC-MS. They recruited 34 compensated cirrhotics and 87 non-cirrhotic patients with chronic liver disease of all aetiologies alongside 31 individuals

with no serological evidence of liver disease to act as the healthy control cohort. Chronic liver disease was classified as an individual with established liver disease (e.g., NAFLD, autoimmune liver disease or chronic viral hepatitis) without radiological, endoscopic, or serological evidence of cirrhosis. Histological confirmation was also available in 47 cases. Whole breath was collected with a Mylar® gas sampling bag. There were no restrictions on diet, fasting or smoking status prior to collection and there were no attempts to mitigate for background room air volatiles. A model based on 11 VOCs (3-methylbutanal, propanoic acid, octane, an unidentified terpene, α-pinene, 3-carene, branched C16H34, 1-hexadecanol, branched C16H34, dimethyl disulfide) was able to differentiate between compensated cirrhotics and those with only chronic liver disease with a sensitivity of 0.71 and specificity of 0.84. Once again, a model with 11 VOCs is liable to overfitting and for many of the VOCs suggested, there is no logical underlying metabolic pathway to account for the differences.

Alcoholic Hepatitis vs Acutely Decompensated Cirrhosis

As part of their 2014 study, Hanouneh et al. (195) attempted differentiate between alcoholic hepatitis (with background cirrhosis) and acutely decompensated cirrhotics (with decompensation triggered by causes other than alcohol) using SIFT-MS. They recruited 123 patients (40 alcoholic hepatitis with cirrhosis, 40 decompensated cirrhotics with non-alcohol related aetiologies, 43 controls without liver disease). They found TMA, acetone, and pentane could discriminate between alcoholic hepatitis and decompensated cirrhosis from non-alcohol aetiologies with a sensitivity of 0.90 and specificity of 0.80. They suggested TMA may be elevated due to reduced metabolism by the diseased liver and that pentane may be elevated due to increased oxidative stress in the decompensated cohort. As acetone can be influenced by many confounding factors, in particularly fasting status, it is unlikely to be a useful biomarker. While the individuals recruited for this study were “fasted”, the study does not comment further and whether all individuals were fasted for the same duration.

Non-Alcoholic Steatohepatitis vs Healthy Control

In 2013, Verdam et al. were the only group to look specifically at non-alcoholic steatohepatitis (NASH) patients (198). They compared the breath of 39 patients with histologically confirmed NASH to the breath of 26 healthy controls with GC-MS. Whole breath was collected with a 5 L Tedlar® gas sampling bag before being transferred to a TD tube at an unknown flow rate. There is no comment on the fasting status, smoking habits, or diets of individuals, nor consideration of the background room air volatiles. The study found

that discrimination between the two groups was best with three VOCs: n-tridecane, 3-methyl-butanonitrile and 1-propanol. This gave an AUROC of 0.77. This study is limited by a lack of consideration of confounding factors. 3-Methylbutanonitrile is an established component of tobacco smoke (212) and 1-propanol can be influenced by fasting (213) and therefore without knowing the fasting and smoking status of the different cohorts, it is difficult to draw conclusions on their validity as biomarkers.

Non-Alcoholic Fatty Liver Disease vs Healthy Control

In 2020, Sinha et al. looked at the ability of VOCs to differentiate between non-alcoholic fatty liver disease (NAFLD) with and without cirrhosis alongside healthy controls. They recruited 15 participants with Child-Pugh A NAFLD cirrhosis, 14 participants with non-cirrhotic NAFLD and 14 healthy controls. After an overnight fast and following an oral water rinse, participants inhaled via an inspiratory VOC filter before exhaling into a 10 L Tedlar® bag. Breath was transferred to TD tubes via a pump at 200 mL/min and analysed by GC-MS. Background room air volatiles were not analysed. The study identified 3 VOCs which discriminated between groups: terpinene, dimethyl sulfide, and d-limonene. With a model consisting of only dimethyl sulfide and d-limonene, NAFLD cirrhosis could be separated from healthy volunteers with an AUC of 0.98. Patients with NAFLD cirrhosis could be separated from non-cirrhotic NAFLD patients with the same model with an AUC of 0.91 (214).

Raman et al.'s 2013 study (215) was the only study to investigate VOCs from faeces. They analysed the faecal headspace of 30 individuals with clinically suspected NAFLD compared to 30 controls with no evidence of liver disease. They identified 18 VOCs which were significantly higher in those with NAFLD. 13 of these VOCs were ester compounds derived from short chain alcohols and carboxylic acids (e.g., butanoic acid and propyl ester). They also identified 12 VOCs which were significantly lower in individuals with NAFLD compared to controls. This was a more homogenous group of VOCs including ketones (e.g., 2-butanone), aldehydes (e.g., heptanal) and alkane derivatives (e.g., cyclohexane, hexyl-). No models to suggest the diagnostic capabilities of these VOCs were generated and there was no clear metabolic pathway to link the VOCs with NAFLD. Also of note, their control cohort had lower BMIs (<25 kg/m^2) than the NALFD cohort (>30 kg/m^2). There was no comparison of non-obese NAFLD or obese controls which further limits the impact of this study. It is not possible to say if the differences noticed were due to differences in weight and nutritional status over disease process.

HCC vs Cirrhosis vs Healthy Control

A study by Xue et al. (216) in 2008 was the first within this review to have investigated VOCs specific to HCC. They used SPME-GC-MS to assess VOCs within the headspace of blood of 37 participants (19 individuals with HCC, 18 healthy controls). 19 of the 47 VOCs detected were significantly different with hexanal, 1-octen-3-ol and octane providing the best sensitivities and specificities for detecting HCC. Hexanal had a sensitivity of 0.95 and a specificity of 1.00.

In 2010, Qin et al. (191) used GC-MS to compare the breath of 30 participants with untreated and histologically confirmed HCC to 27 Child-Pugh A cirrhotics secondary to hepatitis B alongside 36 healthy controls. Healthy controls were sourced from patient relatives or hospital staff and had no history of chronic disease. After fasting overnight and completing a questionnaire on diet and smoking habits, participants provided whole breath samples into 4 L Tedlar® gas sampling bags. VOCs were analysed via SPME-GC-MS. Room air samples were also collected, and the results of VOCs of interest were reported as an alveolar gradient. The investigators found 3-hydroxy-2-butanone, styrene, and decane were elevated in cirrhosis and further elevated in HCC. Styrene is a well-established component of tobacco smoke (217) and while the smoking status of those recruited was noted, there is no comment on whether smoking status correlated with styrene concentrations. 3-hydroxy-2-butanone as an individual biomarker gave a sensitivity of 0.83 and a specificity of 0.92 in differentiating HCC from healthy controls. It was less successful at differentiating between cirrhotics and HCC with a sensitivity of 0.70 and specificity of 0.70, suggesting that differences in 3-hydroxy-2-butanone are driven by cirrhotic changes rather than HCC.

As part of a study focusing on the presence or absence of hepatic encephalopathy in 2016, O'Hara et al. (200) used the same patient cohort as Fernandez del Rio et al. (177) to perform a subgroup analysis of 11 patients with HCC (deemed suitable for liver transplantation) against patients without HCC (a mix of cirrhotic patients and healthy controls) and found limonene to be lower in patients with HCC. It is acknowledged that this may represent less severe underlying liver disease in those with HCC deemed appropriate for liver transplantation rather than being directly related to the presence of an HCC itself. Ferrandino et al. (209) who also performed a subgroup analysis of only those with HCC compared to cirrhotics and healthy controls as part of their 2020 study. They found limonene to be elevated in participants with cirrhosis whether an HCC was present or not.

Miller-Atkins et al., as part of their 2020 study, compared 112 patients with HCC to 30 patients with cirrhosis. They found lower concentrations of acetaldehyde, acetone and dimethyl sulfide and higher concentrations of ethanol in HCC patients compared to cirrhosis only. When compared to 54 healthy controls, (E)-2-nonene, ethane and benzene were significantly increased, and hydrogen sulfide significantly decreased in patients with HCC (196). Once again, concentrations of benzene were reported without reference to smoking status. While there are logical metabolic pathways which may account for the differing concentrations of acetaldehyde, ketones, and sulfur compounds in varying states of hepatopathology, the same cannot be said for (E)-2-nonene, for which no metabolic explanation is offered.

In 2022, Sukaram et al. (192) published the largest study to date looking at exhaled VOCs for diagnosis of HCC. They sampled the breath of 97 individuals with HCC and compared them to 111 control patients (a mixture of participants with cirrhosis and healthy volunteers without liver disease) using GC-MS. Methods of standardisation of breath collection included fasting for a minimum of 8 hours and avoidance of alcohol and smoking for 24 hours. Antibiotics and probiotics were also restricted for the 3 weeks prior to sampling. Whole breath was collected from individuals via a 1 L Tedlar® bag before analysis with GC-MS. Background volatiles do not appear to have been considered. The model of VOCs which provided the highest accuracy consisted of acetone, 1,4-pentadiene, methylene chloride, benzene, phenol, and allyl methyl sulfide. This model gave an AUC of 0.80 with a sensitivity of 0.77 and specificity of 0.827. When looking only at early-stage HCC, the VOC with the single best diagnostic performance was d-limonene with sensitivity 0.63 and specificity 0.52. As well as comparing HCC to controls, Sukaram et al. also looked at the VOC profiles of HCC participants pre- and post-treatment. Of the 97 individuals with HCC, they re-sampled 34 of them following treatment to determine if there were any consistent changes in VOC profiles associated with response to treatment. There was a reduction in acetone concentrations in the post-treatment cohort but significantly more so in treatment responders compared to non-responders. In those that received TACE and had treatment response there was a significantly higher level of dimethyl sulfide. Individuals who underwent ablation and responded had decreased concentrations of acetone and allyl methyl sulfide. Those who had ablation without response, had elevated concentrations of acetone and allyl methyl sulfide post-treatment. While the longitudinal analysis of participants pre- and post- treatment is admirable, acetone is not a reliable biomarker given the multitude of physiological processes which can influence its concentration (100,101) and there is no logical metabolic pathway identified which would account for the changes in allyl

methyl. The overall study is also limited by the selection of controls. By mixing cirrhotics and individuals with no liver disease, it provides a very heterogenous control group.

Bannaga et al.'s 2021 study (218) was the only identified study to look at urinary volatiles in individuals with HCC. They compared the urinary volatiles of 18 HCC patients to 40 controls both with and without fibrosis using gas chromatography coupled with ion-mobility spectrometry (GC-IMS) followed by GC-MS for compound identification. They identified 7 VOCs which gave a specificity of 0.43 and sensitivity of 0.95 (with overall AUC 0.97) when comparing HCC patients to controls with fibrosis. The included higher concentrations of 2-butanone and lower concentrations of 2-hexanone in individuals with HCC compared to those with fibrosis only but also suggested several compounds which have not been identified in the human volatilome previously (219) making them unlikely to represent HCC biomarkers. Compounds were also only tentatively identified, and the study was further limited by their small sample size.

Early (F0 to F2) vs Advanced (F3 to F4) Fibrosis

Alkhouri et al. (193) were the only group who investigated differing stages of hepatic fibrosis. In their 2015 study, they compared breath samples of histologically confirmed F0-F2 fibrosis with F3-F4 fibrosis. Of the 61 individuals recruited, 40 had F0-F2 fibrosis and 21 had F3-F4 fibrosis. Prior to the breath sampling, individuals fasted for 8 hours and rinsed their mouths with water. Individuals inhaled through an acid gas cartridge to minimise contaminant volatiles from background room air before exhaling into a Mylar® bag. Samples were then analysed with SIFT-MS. Liver biopsies were taken on the same day as the breath sampling. Six VOCs (isoprene, benzene, carbon disulfide, pentane, ethane, and acetone) were found to be significantly lower in patients with advanced fibrosis compared to early fibrosis. Of these six, isoprene had the best discriminative capability with an AUROC of 0.86 but as already discussed section 2.5.1, it is highly unlikely to be a useful biomarker. There was also no consideration of the smoking status of the recruited individuals in this study, which makes reliable interpretation of benzene in this context challenging. Also, as discussed previously, acetone is also unlikely to be a useful biomarker even with a standardised fast due to the impact of other confounding factors e.g., overall nutritional status (220).

Hepatic Encephalopathy vs No Hepatic Encephalopathy

In 1981, Goldberg at al used GC-MS to investigate VOCs in blood plasma and CSF in individuals with hepatic encephalopathy compared to a control group without hepatic encephalopathy (221). They identified elevated 3-methylbutanal in the CSF and plasma in

those with HE, and lower concentrations of furfural in the plasma only of those with HE. They also however noted that 3-methylbutanal were higher in those who had not been fasted, compared to those that had and proposed that the VOC may be of dietary protein origin. While all individuals in the HE cohort had previously had HE, they were of different grades, and it is unclear how many had active HE at the time of sampling. The control group was also extremely heterogenous, with varying degrees of fasting, both cirrhotics and non-cirrhotics as well as some individuals with portocaval shunts.

As part of their 2013 study, Khalid et al. also investigated whether VOCs could be used to detect the presence of hepatic encephalopathy (HE). They found that methyl vinyl ketone was negatively associated with HE and an unconfirmed VOC (possibly isothiocyanato-cyclohexane) was positively associated with HE. Together, these two VOCs gave a PPV of 0.77 and NPV of 0.95 (207). As commented on by the authors themselves, isothiocyanato-cyclohexane is an exogenous environmental contaminant and as the study did not examine the ambient room air during the recruitment process, it requires further investigation before being considered as a biomarker for HE.

In 2016, O'Hara et al. (200) also analysed the breath of 11 participants with HE compared to 18 participants without HE at the time of breath sampling. The latter were further subdivided into individuals who had never been known to have HE (11 participants) and those who had had HE previously (7 participants). Alveolar breath was collected via glass syringe and capnography and analysed with PTR-MS. They found a significant difference in the level of limonene in those with active HE compared to those with a history of HE. Reassuringly they found no association between recent consumption of citrus fruits and juices and limonene concentrations within breath. As those with HE are likely to have more advanced liver disease than those without, it is logical that the hepatic metabolism of limonene would be more impaired in those with HE than those without which may account for the higher concentrations of limonene in those with HE.

Table 5.3. Summary of articles included within systematic review of VOCs and Liver Disease.

Author / Year / Reference	Country	Matrix	Participants	Aetiology	Collection / Analysis Technique	STARD	Comparison	Discriminating VOCs	Sens.	Spec.	AUC	Notes
Goldberg et al. (1981) (221)	Canada	Blood CSF	39 (20 Controls, 18 hepatic encephalopathy patients)	All	GC-MS	14	HE vs no HE	Blood plasma: 3-methylbutanal (+), furfural (-) CSF: 3-methylbutanal (+)	N/A	N/A	N/A	
Friedman et al. (1994) (194)	USA	Breath	48 (24 Liver disease, 24 healthy controls)	All	Tedlar® bags End-expiratory breath GC-MS	19	Cirrhosis vs healthy control	Limonene (+), hydrogen sulfide (+)	N/A	N/A	N/A	
Sehnert et al. (2002) (201)	USA	Breath	195 (86 Liver disease, 109 healthy controls)	All	Glass syringe Whole breath GC-MS	18	All type liver disease vs control	Carbon disulfide (+), carbonyl sulfide (+), isoprene (+)	N/A	N/A	N/A	
							Bile duct damage vs control	Carbon disulfide (+), carbonyl sulfide (+),isoprene (+)	N/A	N/A	N/A	
							Hepatocellular damage vs control	Carbon disulfide (-), carbonyl sulfide (-)	0.48	0.88	N/A	
Xue et al. (2008) (216)	China	Blood	37 (19 HCC, 18 healthy controls)	All	GC-MS	20	HCC vs healthy control	Hexanal (+), 1-octen-3-ol (+), octane (+)	1	1	1	
Van den Velde et al. (2008) (202)	Belgium	Breath	102 (52 Cirrhosis, 50 healthy controls)	All	BioVoc® Alveolar breath GC-MS	21	Cirrhosis vs healthy control	Dimethyl sulfide (+), acetone (+), 2-butanone (+), 2-pentanone (+), indole (-), dimethyl selenide (-)	1	0.7	N/A	
Millonig et al. (2010)	Austria	Breath	126 (NAFLD 34, AFLD 20,	All	Glass vial Whole breath	19	AFLD vs healthy controls	Acetaldehyde (+), M103 (+), isoprene (+), M67 (+), M60 (+), M116 (+), M80	N/A	N/A	0.97	

(205)			cirrhosis 37, healthy controls 35)		IMR-MS			(+), butadiene (+), M114 (+)				
							NAFLD vs healthy controls	Methane (+), M39 (+), M32 (+), hydrogen sulphide (+), M41 (+), M20 (+), M75 (+), M38 (+), acetaldehyde (+), M29 (+)	N/A	N/A	0.96	
							NAFLD vs AFLD	M76 (+), M80 (+), M75 (+), methylamine (+), M79 (+), M102 (+), N20 (+), M19 (+), M33 (+), M67 (+)	N/A	N/A	0.95	
							Cirrhosis vs healthy control	Ethanol, M44, M49, M60, acetaldehyde, M39, propene, M50, M32, propanol	N/A	N/A	0.88	
Qin et al. (2010) (191)	China	Breath	93 (30 HCC, 27 cirrhosis, 36 healthy controls)	Hepatitis B	Tedlar® bag Whole breath GC-MS	22	HCC vs healthy control	3-Hydroxy-2-butanone (+), styrene (+), decane (+)	0.83*	0.92*	0.93*	*3-Hydroxy-2-butanone only
							HCC vs cirrhosis	3-Hydroxy-2-butanone (+), styrene (+), decane (+)	0.70*	0.70*	0.75*	*3-Hydroxy-2-butanone only
Sinha et al. (2010) (214)	UK	Breath	43 (15 Cirrhosis, 14 NAFLD, 14 healthy controls)	NAFLD	Tedlar® Bag Whole breath GC-MS	22	Cirrhosis vs NAFLD	Dimethyl sulfide (+), d-limonene (+)	N/A	N/A	0.91	
							Cirrhosis vs healthy control	Styrene (-), isoprene (-), acetophenone (-), terpinene (-), dimethyl sulfide (+), d-limonene (+)	N/A	N/A	0.98*	*Dimethyl sulfide and d-limonene only
							NAFLD vs healthy control	Isoprene (-), acetophenone (-), terpinene (-)	N/A	N/A	0.84*	*Terpinene only
Dadamio et al. (2011) (203)	Belgium	Breath	84 (35 Cirrhosis, 49 healthy controls)	All	BioVoc® Alveolar breath GC-MS	24	Cirrhosis vs healthy control	β-Pinene (+), acetone (+), isoprene (+), 2-methyl-1-propene (+), caryophyllene (+), dimethyl sulfide (+), propane (+), octane (+), γ-terpinene (+), α-pinene (+), 2-pentanone (+), d-Limonene (+), alkylbenzene (+), 2-butanone (+), nonane (+), tridecane (+), styrene (+), tetradecane (+), phenol (+),	0.83	1	N/A	

								indole (+), dimethylselene (+), hexane (+)				
Raman et al. (2013) (215)	Canada	Faeces	60 (30 NAFLD, 30 healthy controls)	NAFLD	GC-MS	20	NAFLD Vs healthy control	Butanoic acid, propyl ester (+), propanoic acid, propyl ester (+), acetic acid, ethyl ester (+), acetic acid, pentyl ester (+), cyclohexene, 4-ethenyl-4- methyl-3-(1-methylethenyl)- 1-(1-methylethyl)-, (3Rtrans)- (+), butanoic acid, 3-methyl-, butyl ester (+), n-propyl acetate (+), butanoic acid, butyl ester (+), phellandrene (+), propanoic acid, ethyl ester (+), 1,6-octadien-3-ol, 3,7-dimethyl- (+), myrcene (+), pentanoic acid, methyl ester (+), acetic acid, methyl ester (+), 2-propynoic acid methyl ester (+), butanoic acid, 3-methyl-, ethyl ester (+), 1-propanol (+), propanoic acid, 2-methyl-, propyl ester (+), 2-butanone (-), furan, 2-methyl- (-), heptanal (-), 2(3H)-furanone, dihydro-5- methyl- (-), 2-heptanone, 6-methyl- (-), 2,3-pentanedione (-), 1,6-octadien-3-ol, 3,7-dimethyl-, 2-aminobenzoate (-), cyclohexanol, 5-methyl-2-(1-methylethyl)- (-), 2-octene, 3,7-dimethyl-, (Z)- (-), 3-hexanone, 2-methyl- (-), acetic acid,	N/A	N/A	N/A	

								(1,2,3,4,5,6,7,8- octahydro-3,8,8-trimethylnaphth-2-yl) methyl ester (-), cyclohexane, hexyl- (-)				
Verdam et al. (2013) (198)	Netherlands	Breath	65 (39 NASH, 26 controls)	NASH	Tedlar® bag Whole breath GC-MS	21	NASH vs no NASH	n-Tridecane (+), 3-methyl-butanonitrile (+), 1-propanol (+)	0.9	0.69	N/A	
Morisco et al. (2013) (206)	Austria	Breath	26 (12 Cirrhosis, 14 healthy controls)	All	BET mouthpiece Whole breath PTR-MS	19	Cirrhosis vs healthy control	Methanol (+), 2-butanone (+), 2-pentanone (+), 2-octanone (+), NS-compound (+), NS-compound (+), heptadienol (+), s-compound (+), C8-ketone (+), unidentified terpene (+), Sulfoxide-compound (+), unidentified monoterpene (+), C9-ketone (+)	0.83*	0.86*	0.89*	For unidentified monoterpene only
Khalid et al. (2013) (207)	UK	Breath	69 (47 Cirrhosis, 7 ETOH excess without cirrhosis, 15 healthy controls)	All	Custom breath collection device Alveolar breath GC-MS	22	Alcohol Related Cirrhosis vs Healthy Controls	Heptane, 1-methyl-2-(1-methylethyl)-benzene (+), phellandrene (+), 2-methylhexane (+)	0.97	0.93	N/A	
Hanouneh et al. (2014) (195)	USA	Breath	123 (40 Alcoholic hepatitis with cirrhosis, 40 non-ETOH decompensated cirrhosis, 43 healthy controls)	All	Mylar® bag Whole breath SIFT-MS	22	All type liver disease vs healthy control	2-Propanol (+), acetaldehyde (+), acetone (+), ethanol (+), pentane (+), trimethyl amine (+)	N/A	N/A	N/A	
							AH cirrhosis vs non-ETOH cirrhosis	Acetaldehyde (+), acetone (+), pentane (+), trimethylamine (+)	N/A	N/A	0.92*	*Pentane and Trimethyl amine only
Fernandez del Rio et	UK	Breath	61 (31 Liver disease, 30	All	Glass syringe Alveolar breath	20	Cirrhosis vs healthy control	Limonene (+), methanol (+), 2-pentanone (+), 2-butanone (+),	N/A	N/A	0.95*	*Methanol, 2-pentanone and

al. (2015) (177)			healthy controls)		PTR-MS			carbon disulfide (+)				limonene only
							Pre vs post liver transplant	Methanol (-), 2-pentanone (-), Limonene (-)	N/A	N/A	N/A	
Pijls et al. (2015) (197)	Netherlands	Breath	121 (87 CLD, 34 cirrhosis)	All	Mylar® bag Whole breath GC-MS	24	CLD vs cirrhosis	3-Methylbutanal (−), propanoic acid (+), octane (+), terpene (C10H16) (+), α-pinene (+), 3-carene (+), branched C16H34 (+), 1-hexadecanol (−), branched C16H34 (−), dimethyl disulfide (+)	0.83	0.87	0.9	
Alkhouri et al. (2015) (193)	USA	Breath	61 (40 F0-2 fibrosis, 20 F3-4 fibrosis)	All	Mylar® bag Whole breath SIFT-MS	22	F0-2 Fibrosis vs F3-4 fibrosis	Acetone (-), benzene (-), carbon disulfide (-), isoprene (-), pentane (-), ethane (-)	N/A	N/A	0.85*	*Isoprene Only
O'Hara et al. (2016) (200)	UK	Breath	61 (11 HCC, 20 cirrhosis, 30 healthy controls)	All	Glass syringe Alveolar breath PTR-MS	18	HE vs no HE	Limonene (+)	N/A	N/A	N/A	
							HCC vs no HCC	Limonene (-)	N/A	N/A	N/A	
Ferrandino et al. (2020) (209)	UK	Breath	84 (12 HCC, 32 cirrhosis, 40 healthy controls)	All	ReCIVA® Alveolar breath GC-MS	19	HCC vs cirrhosis vs healthy control	Limonene (+)	0.73	0.23	N/A	
Miller-Atkins et al. (2020) (196)	USA	Breath	296 (112 HCC, 30 cirrhosis, 49 pulmonary hypertension, 51 liver metastases, 54 healthy controls)	All	Mylar® bag Whole breath SIFT-MS	22	HCC vs healthy control	(E)-2-nonene (+), ethane (+), benzene (+), hydrogen sulfide (-)	0.62	0.64	0.63	
							Cirrhosis vs healthy control	Trimethylamine (+), propanol (+), methylhexane (+), decene (+), acrylonitrile (+)	0.41	0.96	0.68	
							HCC vs cirrhosis	Acetone (-), Acetaldehyde (-), Dimethyl Sulfide (-), Ethanol (+)	N/A	N/A	N/A	

Bannaga et al. (2021) (218)	UK	Urine	58 (20 HCC, 38 non-HCC)	All	GC-MS	20	HCC vs fibrosis	4-Methyl-2,4-bis(p-hydroxyphenyl)pent-1-ene, 2TMS derivative (-), 2-butanone (+), 2-hexanone (-),Benzene, 1-ethyl-2-methyl- (-),3-Butene-1,2-diol, 1-(2-furanyl)- (-), Bicyclo[4.1.0]heptane, 3,7,7-trimethyl-, [1S-(1a,3ß,6a)]- (-), Sulpiride (-)	0.43	0.95	0.97	Compounds listed for HCC vs no HCC only
							HCC vs no fibrosis		0.60	0.74	0.62	
							Fibrosis vs no fibrosis		0.29	0.90	0.63	
Sukaram et al. (2022) (192)	Thailand	Breath	208 (97 HCC, 78 cirrhosis, 33 healthy controls)	All	Tedlar® bag Whole breath GC-MS	21	HCC vs No HCC (cirrhosis and healthy controls)	Acetone, 1,4-pentadiene, methylene chloride, benzene, phenol, allyl methyl sulfide	0.77	0.83	0.80	
							HCC pre vs post treatment	Acetone (-), Dimethyl Sulfide (+), Butane (+)	N/A	N/A	N/A	
Ferrandino et al. (2023) (204)	UK	Breath	88 (32 Cirrhosis, 14 cirrhosis with HCC, 42 healthy controls)	All	ReCIVA® Alveolar breath GC-MS	20	Cirrhosis vs healthy control	2-Pentanone (+), 1-pentene, Indole, Dimethyl selenide (-), Limonene (+), Eucalyptol, Benzene, (1-propylnonyl)-	N/A	N/A	N/A	

5.6.5 Recurrent Themes

While studies used different collection techniques, different analytical platforms, and varying methods of controlling for confounding factors, several common themes were identified.

Sulfur Compounds

Volatile sulfur compounds were amongst the first VOCs considered as biomarkers for hepatopathology due to an assumption that the specific odour of foetor hepaticus was related to sulfur compounds, in particular dimethyl sulfide. Dimethyl sulfide is the sulfur compound identified most frequently within the literature review as being elevated in hepatopathology. Van den Velde et al. (202), Dadamio et al. (203), Pijls et al. (197) and Sinha et al. (214) all found elevated concentrations of dimethyl sulfide in those with hepatopathology. Alkhouri et al. (193) found lower concentrations of dimethyl sulfide in those with F3-F4 fibrosis compared to those with F0-F2 fibrosis but unfortunately did not have a control set with normal liver parenchyma to compare to.

Carbon disulfide and carbonyl sulfide were found to be elevated in hepatopathology by Sehnert et al. (201). Fernandez del Rio et al. (177) also found elevated concentrations of carbon disulfide in those with cirrhosis and in all but one participant, this demonstrated wash out post liver transplantation. Friedman et al. (194) found higher concentrations of hydrogen sulfide in those with mixed type liver disease and Millonig et al. (205) found that hydrogen sulfide was elevated in patients with NAFLD compared to controls.

The source of elevated sulfur compounds in liver disease can be linked to impaired metabolism of sulfur-containing amino acids (e.g., methionine) by the diseased liver (81,222). The liver plays a vital role in the metabolism of amino acids, and it is logical to assume that in cirrhosis, this function will be impaired, resulting in elevated concentrations of sulfur compounds. Volatile sulfur compounds however can also be generated by bacteria in the oral cavity (i.e., halitosis) (223), highlighting the importance of basic standardisation techniques such as oral water rinses prior to breath sampling (103) to try and mitigate this confounding factor.

Terpenes and Terpenoids

Limonene is an exogenous, cyclic monoterpene that is commonly found in citrus fruits and oils (224) and was first identified as a possible biomarker for hepatopathology by Friedman et al. (194) in 1994 who identified elevated concentrations in the exhaled breath of those with cirrhosis. This was correlated by Pijls et al. (197), Fernandez del Rio et al. (177),

Ferrandino et al. (209) and Sinha et al. (214). Fernandez del Rio et al. added credence to limonene's potential use as a biomarker with the second stage of their study where they identified that the elevated limonene concentrations gradually washed-out following liver transplantation. The established metabolic pathway for limonene additionally adds support to its use as a biomarker. Limonene is metabolised by hepatic cytochrome P450 enzymes CYP2C19 and CYP2C9 into perillyl alcohol, trans-isopiperitenol, and trans-carveol (225). While present elsewhere in the body, cytochrome P450 enzymes are predominantly found in the liver and are vulnerable to reduced function in the context of liver disease (110). Given the consistent findings across multiple studies and a logical underlying metabolic pathway, limonene is a viable biomarker for hepatopathology.

3-Carene, another exogenous monoterpene and constituent of turpentine (226), was found by Pijls et al. (197) to be elevated in cirrhosis. While also metabolised by hepatic cytochrome P450 enzymes (227) and therefore susceptible to reduced metabolism in hepatopathic states, it has also been found in environmental air (1) and as Pijls et al. did not consider background volatiles in their study, it should be interpreted with caution. However, given a logical potential underlying metabolic pathway, it is feasible that it could be used as a biomarker for hepatopathology. α-Pinene, also an exogenous monoterpene present in essential oils, was found to be elevated in cirrhosis by Pijls et al. (197) and Dadamio et al. (203). As also metabolised by hepatic cytochrome P450 enzymes, it is feasible it could be a biomarker for liver disease.

Isoprene, an endogenous and exogenous hemiterpenoid, was found to be different between disease and control states by several studies but with conflicting results. Alkhouri et al. (193) found decreased concentrations of isoprene in F3-F4 fibrosis compared to F0-F2 fibrosis and Sinha et al. (214) found decreased concentrations in those with cirrhosis compared to controls. The inverse was found by Dadamio et al. (203) and Sehnert et al. (201) who identified elevated concentrations of isoprene in liver disease compared to controls. Millonig et al. (205) identified higher concentrations in alcohol related fatty liver disease only. The justification for isoprene as a possible biomarker was in part based on a historical belief that isoprene was a by-product of cholesterol synthesis but as discussed in section 2.5.1, work by Mochalski et al. (105) has since identified that isoprene is predominantly released by muscle and is highly susceptible to changes in exercise status, breathing patterns and cardiac output. The muscle origin of isoprene has further been confirmed by a recent published study by Sukul et al. (228). At present, it is unlikely to represent a biomarker for hepatopathology.

Eucalyptol, an exogenous monoterpenoid found commonly in plant oils and flavourings, was found to discriminate between cirrhosis and controls by Ferrandino et al. (204). Eucalyptol is also likely metabolised by hepatic CYP3A4 enzymes (229) and therefore would be expected to present in higher concentrations in those with impaired liver function.

Aldehydes

Acetaldehyde was found to be elevated in individuals with liver disease by Hanouneh et al. (195), who also found acetaldehyde was further elevated in cirrhotics with concurrent alcoholic hepatitis compared to those without. Millonig et al. (205) also found elevated concentrations of acetaldehyde in those with NAFLD compared to both controls and cirrhotics. Acetaldehyde is a product of ethanol metabolism. Ethanol is metabolised by alcohol dehydrogenase to acetaldehyde. Acetaldehyde, a toxin, is metabolised further by a group of aldehyde dehydrogenase enzymes to carboxylic acids. While present in multiple sites within the human body, acetaldehyde dehydrogenase is most commonly found in the liver. It is therefore presumed that the function of these enzymes will decrease in liver disease states (230), accounting for higher concentrations of acetaldehyde in those with hepatopathology. Consumption of alcohol will also induce cytochrome p450 enzyme CYP2E1 which metabolises ethanol to acetaldehyde (231). This may in part explain Hanouneh et al.'s (195) finding of higher concentrations in those with alcoholic hepatitis. It should also be noted however that acetaldehyde can be present in tobacco smoke (232) and should therefore be interpreted with reference to the smoking status of participants.

Xue et al. (216) identified hexanal, another aldehyde, as having higher concentrations in the blood headspace of individuals with HCC compared to healthy, non-cirrhotic controls. Hexanal is an aldehyde by-product of lipid peroxidation (233). While lipid peroxidation is triggered by oxidative stress which can be caused by chronic liver disease, it can also be triggered by multiple other non-hepatic disease states (234) including respiratory disease (235) and therefore hexanal is unlikely to be viable as a biomarker specific to HCC or chronic liver disease. Co-morbidities, diet and environmental contamination should be considered when interpreting differing concentrations of aldehydes in those with and without liver disease.

Alkanes

Multiple studies identified significant differences in alkane concentrations in those with and without liver disease. Methane was elevated in participants with NAFLD according to Millonig et al. (205) and decane was elevated in patients with cirrhosis and HCC according

to Qin et al. (191). Dadamio et al. (203) found multiple alkanes elevated in the breath of cirrhotics including propane, tridecane, octane and nonane. Verdam et al. (198) found branched tridecane to be elevated in patients with NASH compared to those without and Khalid et al. found elevated concentrations of heptane in patients with alcohol related cirrhosis compared to controls (207). Hanouneh et al. (195) identified pentane as one of the VOCs which could help differentiate between cirrhotics and controls and also between alcohol related cirrhosis and non-alcohol related cirrhosis. Miller-Atkins et al. also found ethane to be elevated in those with HCC compared to healthy controls (196).

Previous studies investigating exhaled alkanes, in particular ethane, have attributed elevated concentrations to the lipid peroxidation process (236,237). Concentrations of various alkanes have also been found to altered in the context of non-hepatic pathology e.g., elevated pentane in respiratory infections (238) and significant differences in ethane, propane, and pentane in the context of inflammatory bowel disease (239). Therefore, while alkanes may be elevated in hepatopathology, their use as a biomarker is likely limited given the multitude of pathological processes which can cause lipid peroxidation and therefore potential changes in concentrations of exhaled alkanes.

Ketones

Acetone, the most abundant ketone in breath, was found to be elevated in liver disease states by Van den Velde et al. (202), Dadamio et al. (203), Hanouneh et al. (195), Alkhouri et al. (193) and Miller-Atkins et al. (196). It is a major product of ketosis, a crucial step in human energy metabolism. In a starved state, the human body will switch from glycolysis (i.e., gaining energy from metabolism glucose), to ketosis which relies on adipose tissue for energy. Fatty acids released from adipose tissue are metabolised by the liver to acetyl-CoA (a process referred to as beta oxidation) and then to ketone bodies including acetone (220). Insulin resistant states, commonly present in those with liver disease independent of aetiology, favour ketosis due to impaired glycolysis (240). This may account for the increased concentrations of ketones in liver disease states. However, as well as nutritional state, ketosis can be triggered by illness and exercise (241).

2-Butanone was found to be elevated in liver disease states by multiple studies (177,202,203,206,218). It is a predominantly exogenous ketone that is found a wide variety of foods, as well as tobacco smoke. It can also be produced by oxidation of butanol, an exogenous alcohol commonly found in dietary sources. While the exact human metabolism of 2-butanone remains unclear, rat models suggest it can induce hepatic cytochrome P450

enzymes (242). It is also likely a product of beta oxidation of fatty acids in ketosis. Therefore, it is theoretically possible that it could represent a biomarker for liver disease, but further work is required to validate this, and it should be interpreted in the context of smoking status and fasting status. Similarly, 2-pentanone was found to be elevated in hepatopathology by multiple studies (177,202-204,206). It is also a predominantly exogenous ketone found in multiple food types including cow's milk and fruits (243). 2-Octanone was found to be elevated in those with cirrhosis by Morisco et al. (206) only and Raman et al. (215) found elevated 2-nonanone in non-alcoholic fatty liver disease from faecal VOC analysis.

Amines

Trimethylalamine (TMA) is a volatile amine which was identified by Miller-Atkins et al. (196) to be elevated in those with cirrhosis. TMA is produced by the metabolism of dietary choline by the hepatic enzyme amine oxidase. In the context of liver disease, function of amine oxidase may be impaired resulting in elevated concentrations of TMA (244) and other amines. However, it is also found in cigarette smoke. Unfortunately, Miller-Atkins et al. did not comment on the smoking status of their recruited participants and therefore while metabolically feasible, further work is required to confirm its viability as a biomarker.

Aromatic Hydrocarbons

Styrene, a benzene derivative that can be found in coffee, fruits, cigarette smoke (245) and environmental pollution, was found to be elevated by Qin et al. (191) and Dadamio et al. (203). It is another VOC that is metabolised by hepatic cytochrome P450 enzymes (246) and therefore reduced activity of these enzymes in the context of liver disease could account for higher concentrations in hepatopathology. However, as with TMA, concentrations of styrene should be interpreted with reference to the smoking status of participants. Qin et al. found higher concentrations of styrene in those with cirrhosis and those with HCC compared to healthy controls, but they also had a higher proportion of active smokers in the HCC cohort compared to their cirrhosis and healthy control cohorts which may account for the differences in concentrations. Dadamio et al. did review their styrene results in the context of smoking and found it to be elevated in cirrhosis participants independent of smoking status. However, it should be noted that they only identified styrene in 22% of their cirrhotic participants. While feasible as a biomarker, further work is required to validate it based on the existing literature.

A summary of the various compounds discussed can be found in table 5.4.

Table 5.4 Table of VOCs repeatedly identified as differing in hepatopathology.

VOC Class	VOCs	Proposed Mechanism(s)	Biomarker Feasibility
Terpenes	Limonene 3-Carene α-Pinene	Reduced metabolism by cytochrome P450 enzymes (225).	Feasible but should be interpreted in the context of dietary intake and background room air volatiles.
Terpenoids	Isoprene	Previously suggested as a by-product of cholesterol biosynthesis (247) but more recent studies have suggested it is predominantly released by muscles (105).	Unlikely to represent a biomarker. More recent work has established isoprene is predominantly released from muscle stores and concentrations fluctuate significantly with exercise and changes in physiological state (105,106).
	Eucalyptol	Reduced metabolism by hepatic CYP3A4 enzymes (229).	Feasible but should be interpreted in the context of dietary intake and background volatiles.
Sulfur Compounds	Dimethyl sulfide Carbon disulfide Carbonyl sulfide Hydrogen sulfide	Impaired metabolism of sulfur-containing amino acids by the liver (81,222).	Feasible but other sources of volatile sulfur compounds (e.g., oral bacteria) should be considered and controlled for where possible.
Aldehydes	Acetaldehyde Hexanal	Reduced metabolism by aldehyde dehydrogenase (248). Increased lipid peroxidation due to oxidative stress from chronic inflammation in the context of liver disease (233).	While feasible, aldehyde concentrations should be interpreted in context of potential confounding factors which can also influence concentrations (e.g., co-morbidities such as respiratory disease, background room air). It is likely that while aldehyde concentrations may alter with liver disease, the underlying metabolic pathway may not be specific enough to liver disease to be a useful biomarker for hepatopathology.
Ketones	Acetone 2-Butanone 2-Pentanone 2-Octanone	Increased beta oxidation of fatty acids due to insulin resistant states in liver disease (220).	While concentrations of ketones within breath may be altered by presence or absence of liver disease, they are also highly influenced (and likely more so than liver disease) by fasting status and overall nutritional status and therefore should be interpreted with caution as biomarkers for liver disease.
Alkanes	Ethane Propane	Increased lipid peroxidation due to oxidative stress from chronic inflammation in the context	While alkanes appear to be elevated in those with liver disease, alkanes have also been shown to be elevated in multiple other disease states and

	Pentane Decane Octane	of liver disease (236,237).	therefore is unlikely to be specific enough in isolation as biomarkers for liver disease.
Amines	Trimethylalamine	Reduced metabolism by hepatic enzyme amine oxidase (244).	Feasible but should be interpreted with reference smoking status.
Aromatic Hydrocarbons	Styrene	Reduced metabolism by cytochrome P450 enzymes (246).	Feasible but should be interpreted with reference smoking status.

5.7 Conclusion

There have been multiple studies to date using mass spectrometry analysis of VOCs to assess liver function and presence or absence of liver disease. While most studies involved relatively small numbers of participants, it is reassuring that several VOCs are repeatedly identified by independent studies using different populations, different breath collection techniques and variable mass spectrometry techniques. Only one study involved a validation cohort (203).

When focussing specifically on those studies which used breath as their VOC matrix, there was huge variation in attempts at mitigating potential confounding factors. In most cases, where standardisation strategies were less robust, it was a case of pragmatism and practicality rather than a lack of consideration. For some confounding factors, it is a balancing act. For example, fasting will reduce potential contaminant VOCs from the GI tract from recently consumed food but will also alter concentrations of other VOCs e.g., higher concentrations of ketones. With regards to smoking status, it is important to acknowledge that many of the population for which a biomarker would be clinically useful, will be smokers. Therefore, while exclusion or restriction of smokers within an analysis is useful for standardisation purposes, it may not be reflective of real-life clinical practice. Any VOCs which are known to relate to tobacco smoke or e-cigarettes are unlikely to be useful biomarkers and it may be more helpful to exclude these VOCs from the analysis, rather than restrict smoking behaviour.

While many potential biomarkers have been suggested, some studies included models of multiple VOCs which will likely cause overfitting and overconfidence with results (210), especially in the absence of external validation. Several suggested biomarkers are also influenced by other disease states or changes in physiological status and therefore their specificity to hepatopathology is limited. Others also have no logical metabolic basis for their changes and may well be incidental findings.

At present, limonene appears to be the strongest candidate biomarker. As an exclusively exogenous monoterpene, limonene is less prone to the influence of changes in physiological status which affect endogenous VOCs. It also is largely metabolised by the liver and has a strong metabolic justification underpinning the changes found between hepatopathology and healthy controls. Other monoterpenes (e.g., α-pinene, 3-carene), while identified less frequently in the existing literature, also fulfil these criteria. Ketones also have a logical metabolic pathway to justify elevated concentrations in hepatopathology, but as they can be

both endogenous and exogenous, they will be more prone to variation for non-hepatopathology reasons (e.g., fasting, and nutritional status).

Overall, this literature review afforded valuable insight into breath collection techniques, strategies for mitigation of confounding factors as well as potential target VOCs for validation which I have incorporated into my own study design where possible.

SECTION 2: MATERIALS AND METHODS

6 DEVELOPMENT OF A NOVEL BREATH COLLECTION DEVICE FOR BREATH VOC ANALYSIS

6.1 Study Design and Participants

I performed a series of experiments on the novel breath collection device (as described in section 3.7) to ascertain the repeatability of results determine the optimum settings for volume and flow rates for breath collection on to TD tubes. All experiments took place between June 2019 and September 2019.

Five healthy individuals with no known co-morbidities were recruited (three females and two males) to provide breath samples for analysis with the novel breath collection device. The average age was 33 years (range 31-36 years).

6.1.1 Ethical Approval

All participants were recruited under the ethical approval for the study “Non-invasive testing for the diagnosis and assessment of gastro-intestinal disease” (IRAS ID 142097, REC 14/LO/1136) which includes the recruitment of healthy individuals for cross platform validation of VOC measurement. All participants provided written informed consent.

6.2 Breath Collection Protocol

6.2.1 TD Tube Preparation

TD tubes were cleaned prior to being loaded with breath using a TC-20 TD conditioning unit by Markes International Ltd. (Llantrisant, UK) for 40 minutes at 330°C with a nitrogen flow of 50 mL/min. In order to confirm that the tubes had minimal background contamination, the tubes were then analysed with a gas chromatography flame ionisation detector instrument. Acceptable limits of background contaminant VOCs were determined by a previous study by members of our research group. 500 TD tubes were conditioned twice and then analysed on the GC-FID. The mean and standard deviation of the total peak area of the background VOC concentrations on the tubes were calculated. It was deemed that anything which was greater than two standard deviations from the mean are classed as a tube background check failure. Should the tube fail the background check, it is re-conditioned and re-run on the GC-FID until the background total peak area is within the determined limits.

6.2.2 Standardisation and Quality Control

Individuals rested for a minimum of 10 minutes prior to sampling. All sampling episodes took place within the same room within Imperial College London's VOC Laboratory, which has VOC filtration with a CodaAir® VOC remover to minimise background volatile contamination. No fasting or smoking restrictions were placed on individuals as all breath samples for all experiments were provided in the same session. Prior to providing breath samples, all individuals undertook an oral water rinse to minimise volatiles from the oral cavity. Individuals were instructed to perform normal tidal breathing and to not exhale with force.

6.2.3 Repeatability

Five individuals filled five 2000 mL PTFE sampling bags with their breath, one after the other. Multiple exhalations were permitted if required. Using the novel breath collection device, two TD tubes were loaded with 500 mL of breath from each bag at 200 mL/min over 150 s.

6.2.4 Flow Rates

Three individuals provided breath samples into two 2000 mL PTFE sampling bags. From each bag, two TD tubes were loaded with 500 mL of breath. From the first bag, two TD tubes were loaded at 100 mL/min for 300s and from the second bag, two tubes were loaded 200 mL/min for 150s. This ensured the same volume of breath (500 mL) was loaded on to the TD tubes regardless of flow rate.

6.2.5 Volumes

Two individuals provided six breath samples into individual 2000 mL PTFE sampling bags. Two TD tubes were taken via the breath collection device for each bag at a fixed flow rate of 200 mL/min. Different timings were used for each bag, providing the following volumes of breath: 60 s = 200 mL, 120 s = 400 mL, 150 s = 500 mL, 180 s = 600 mL, 240 s = 800 mL and 300 s = 1000 mL.

6.3 Sample Analysis

6.3.1 GC-MS Settings

All samples were analysed within 48 hours of collection using GC-MS. An Agilent Technologies (Santa Clara, California, USA) 7890A GC was paired with a TD100-xr thermal desorption unit and a BenchTOF Select mass spectrometer by Markes International Ltd (Llantrisant, UK). TD tubes were initially pre-purged for one minute with the flow at 50 mL/min. Primary desorption was performed at 250 °C for 5 minutes at 50 mL/min He flow to

desorb the VOCs onto a “Material emissions” focusing trap (Markes International Ltd., Llantrisant, UK) at 20°C in split mode (1:10). While some focusing traps require lower temperatures (and are thus can be referred to as a “cold trap”), Markes International Ltd. recommend a temperature of 20-25°C for use with their “Material emissions” focusing trap to minimise ice crystals and solvent retention which can occur at lower temperatures (249,250). These settings have been used in comparable metabolomic studies (251,252) including Ferrandino et al.’s (204) 2023 study on exhaled VOCs for detection of cirrhosis. Focusing trap (secondary) desorption was performed at 250°C (ballistic heating at 60°C/s) for 3 minutes at 5.7 mL/min He flow, with the flow path onto GC heated constantly at 200°C.

The chromatographic column was a high polarity Mega WAX-HT, (20 m × 0.18 mm × 0.18 µm, Chromalytic, Hampshire, USA). The column flow was set at 0.7 mL/min. Oven temperature was initially set at 35 °C for 1.9 min and was increased to 240 °C (20 °C/min with 2 min hold). The MS transfer line was maintained at 260 °C, whilst ion source (70 eV electron impact) was at 260 °C. The MS analyser was set to acquire over the range of 35 to 500 m/z. Focusing trap desorption (with no TD tube) and conditioned, clean TD tube desorption were included at the beginning and at the end of every analytical run to ensure the absence of carryover effects. Blank analyses were performed right before and right after breath sample desorption to ensure that samples can be analysed sequentially without need for TD conditioning.

6.3.2 Data Analysis

The chromatograms were visually inspected and compared to the chromatograms of the blank TD tubes to ensure peaks were present and that samples had been successfully desorbed. The raw data files were then analysed using Chromspace® (Sepsolve Analytical Ltd.). Compounds of interested were identified from representative breath samples. Annotations were performed with the NIST 2017 mass spectral library based on VOC mass spectra and retention indices. From the raw compound list, only those with a reverse match factor (RMF) greater than 800 were kept for analysis. Oxygen, argon, carbon dioxide and siloxanes were also removed. Any compounds with a signal to noise ratio of less than 3 were also excluded. Peak picking was performed using MATLAB R2018b (version 9.5) and Gavin Beta 3.0 software (253). Peak areas were then generated for each VOC identified.

6.3.3 Repeatability

The coefficient of variation (CV%) was calculated for each compound for each individual and then averaged across all participants to give an indicator of repeatability for each compound.

6.3.4 Flow Rates

The total peak area was calculated for all samples from the sum of all individual peak areas for all VOCs detected at each flow rate. The average of all samples collected at 100 mL/min and all samples collected at 200 mL/min was calculated. For individual VOCs, average peak area was compared between the two different flow rates with a two-tailed t-test.

6.3.5 Volumes

The sum of all VOC peak areas for each volume was calculated and plotted against the volume to determine the correlation between flow rate and volume and to check for the presence of a plateau or evidence of VOC saturation / breakthrough.

7 OPTIMUM STRATEGY FOR RECOLLECTION OF BREATH SAMPLES VIA SAMPLE SPLITTING DURING THERMAL DESORPTION

7.1 Study Design and Participants

All experiments took place in March 2022. Four healthy participants were recruited (two male and two female) to provide breath samples, with an average of 32 years (range 31-36 years).

7.1.1 Ethical Approval

All participants were recruited under pre-existing ethics for the study “Volatile Organic Compounds for Assessment of Liver Disease” (IRAS ID 254249, REC 19/NW/0553) which includes the recruitment of healthy participants for cross platform validation of volatile organic compounds measurement using complementary mass spectrometry techniques.

7.2 Breath Collection Protocol

The experiment took place over four non-consecutive days. Each day, each participant was asked to breathe into a single use PTFE breath collection bag connected to the novel breath collection device. Breath from the filled bag was drawn across two blank TD tubes in sequence at 200 mL/min for 300 s resulting in 1000 mL of breath being loaded on to each TD tube. This was repeated six times per participant to provide 12 TD tubes loaded with breath per participant. All participants repeated this a total of four times. Samples were then processed according to three pathways.

7.2.1 Standardisation and Quality Control

The same standardisation and quality control techniques were used as detailed in sections 6.2.1 and 6.2.2.

Additionally, as part of quality control measures for this study, breath collection bags were also filled with grade 5 (99.999% purity) nitrogen and run with the novel breath collection device with the same settings to load 12 TD tubes with nitrogen. Nitrogen samples were also repeated four times. This was to ensure that any separation of samples was not driven by contaminants within the workflow process.

7.2.2 Pathway One

Each tube has 80% of the desorbed sample recollected back on to the same TD tube (figure 7.1). The remaining 20% of the sample is analysed via GC-MS. Advantages to this strategy includes retaining duplicates of samples. All recollected samples were then run by GC×GC-MS with the sample split setting disabled.

7.2.3 Pathway Two

Two tubes each have 80% of the desorbed sample recollected back on to a single TD tube (figure 7.2). The remaining 20% of the sample is analysed via GC-MS. This will provide both duplicates and potentially higher VOC recollection yield per tube. All recollected samples were then run by GCxGC MS with the sample split setting disabled.

7.2.4 Pathway Three

All four samples have 80% of the desorbed sample recollected back on to a single TD tube (figure 7.3). The remaining 20% of the sample is analysed via GC-MS. This will theoretically increase compound yield however the lack of duplicate samples may reduce confidence over results and the higher sample volume may risk VOC breakthrough. All recollected samples were then run by GCxGC MS with the sample split setting disabled.

7.3 Sample Processing

All samples were analysed within 48 hours of collection.

7.3.1 TD Tube Preparation

In addition to the quality control measures detailed in section 6.2.1, an internal standard was loaded to each TD tube. After being loaded with breath but prior to analysis, each tube was spiked with 1 µl of a working internal standard (containing 2-propanol d_8, acetophenone d_8, octane d_{18}, toluene d_8, benzaldehyde d_6 at 5 mg/L each in methanol) as part of a quality control check for sample analysis. Tubes were placed in a calibration solution loading rig (CSLR) which was attached to a supply of grade 5 nitrogen at a flow rate of 100 mL/min and using a clean calibrated syringe, 1 µL of working internal standard was injected across the TD tube.

7.3.2 GC-MS Settings

Primary analysis was performed by GC-MS with the same settings and methods as listed in section 6.3.1. 240 samples (192 breath samples and 48 nitrogen samples) were processed via GC-MS with the sample split option engaged.

7.3.3 GC×GC-MS Settings

An Agilent Technologies (Santa Clara, California, USA) System 7890B GC was paired with a TD100-xr TD unit and a BenchTOF Select (Markes International Ltd., Llantrisant, UK) mass spectrometer, with a SepSolve Analytical Ltd. (Peterborough, UK) INSIGHT flow modulator. TD tubes were once again pre-purged for 1 min with the flow at 50 mL/min. Primary desorption was performed at 280 °C for 8 min at 50 mL/min He flow to desorb the VOCs onto a focusing trap (Material emissions, Markes International, Llantrisant, UK) at 20°C in split mode. Focusing trap (secondary) desorption was performed at 300°C for 3 min at 6 mL/min He flow, with the flow path onto GC heated constantly at 180°C. The primary (^{1}D) chromatographic column was an Agilent J&W DB-HeavyWAX Polyethylene Glycol (PEG) Column. The column flow was set at 0.5 mL/min. The secondary (^{2}D) column was an Agilent VF-200MS column. The ^{2}D column flow was set at 20mL/min. A splitter was used at the end of the column and the flow to the mass spectrometer was set to 5mL/min. The oven temperature was initially set at 50°C for 3.0 min and was increased to 260°C (5°C/min with 10 min hold). The MS transfer line was maintained at 250°C, whilst ion source (70 eV electron impact) was at 250°C. MS analyser was set to acquire over the range of 35 to 500 m/z. 140 recollected samples were analysed with GC×GC-MS.

7.4 Data Analysis

Following visual inspection of the chromatograms, raw data files were analysed using Chromspace® (Sepsolve Analytical Ltd.). Dynamic baseline correction (DBC) was performed, and all files were converted to .lsc format, a proprietary file format by Markes International Ltd (Llantrisant, UK). Using Chromspace®'s alignment software, all samples were aligned together based on a representative sample.

The aligned .lsc files were analysed using a Tile Sum method (254) via Chromspace®'s sequencer before being imported into Chromcompare+ by Chromspace®. This provided an output that consisted of multiple ions labelled as "features" and the peak area count of each feature in each sample. Samples were divided by pathway.

Within Chromcompare+ by Chromspace®, initial filtering was applied to filter out any features with an intensity of less than 10,000 counts. Using the software's feature discovery settings, the top 100 features separating the samples by participant were selected and further filtering was applied via feature selection and validation to provide the top 10 features discriminating between participants. Principal component analysis (PCA) plots for each pathway were viewed with and without normalisation via probabilistic quotient normalization (PQN). Nitrogen samples were removed from the PQN normalised data sets. The software then generated confidence score (%) providing information on how accurately the samples could be separated by participant.

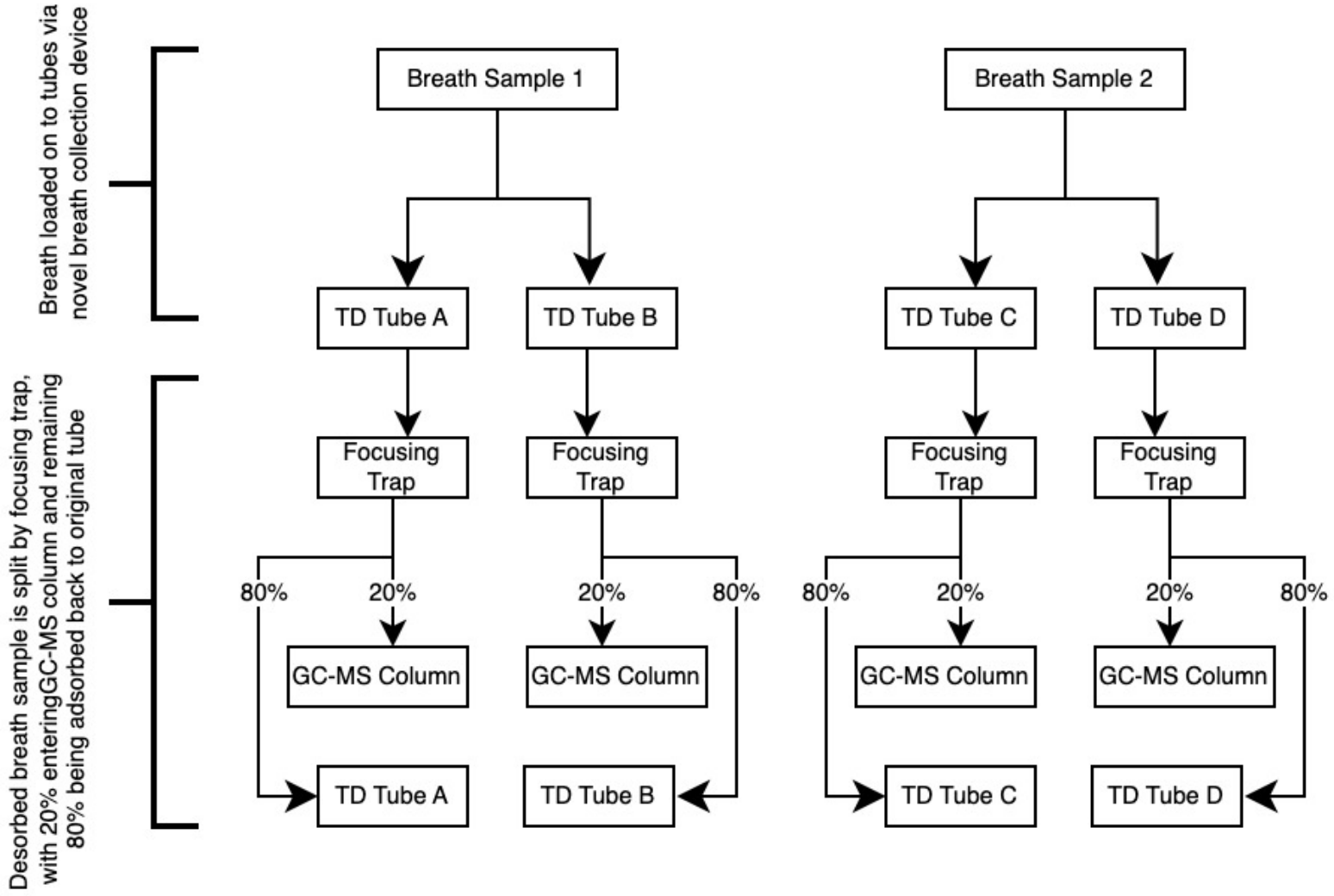


Figure 7.1 Flow chart of sample splitting study pathway 1 where each sample is recollected on to the same tube.

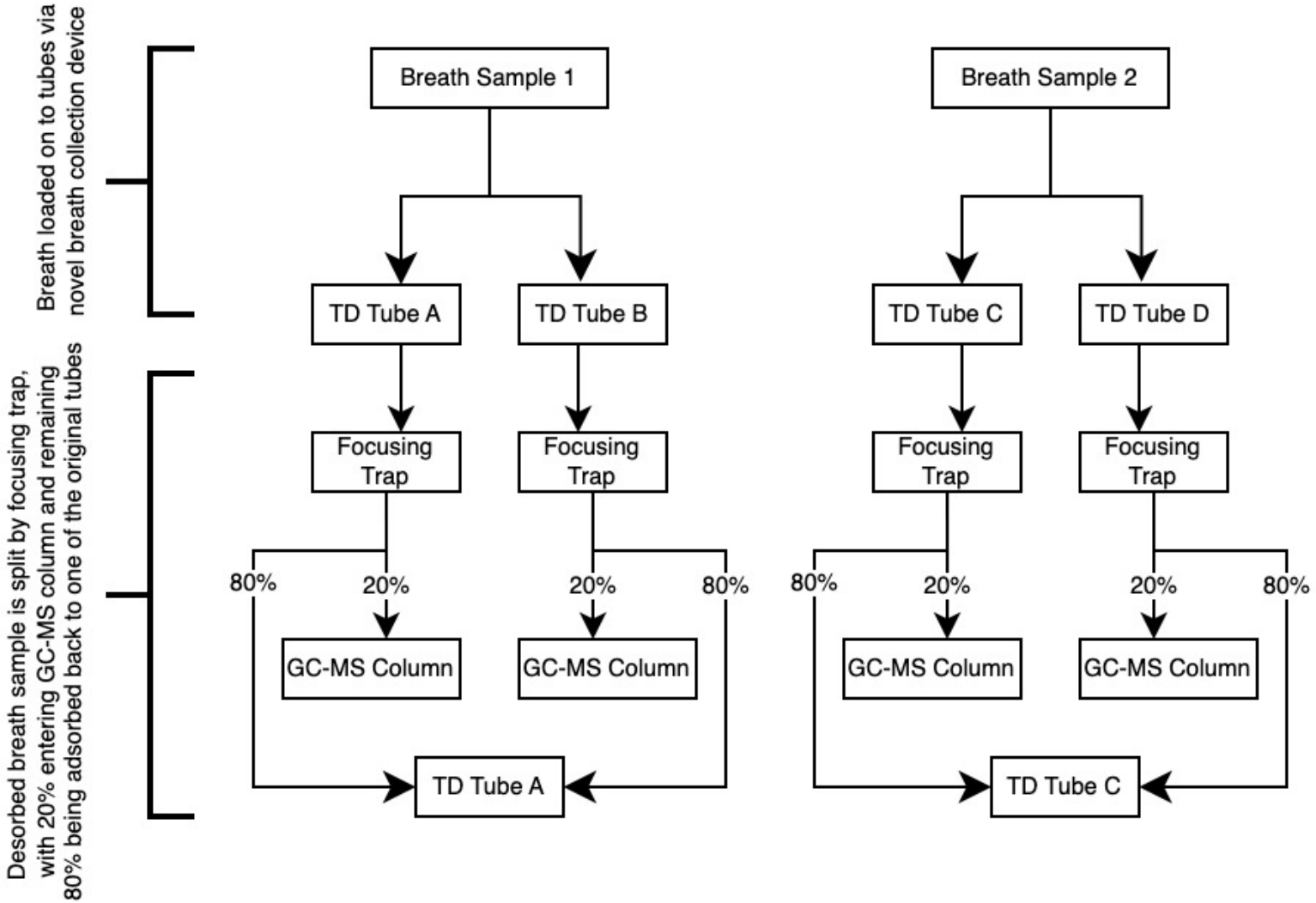


Figure 7.2 Flow chart of sample splitting study pathway 2 where two TD tubes are recollected on to one TD tube.

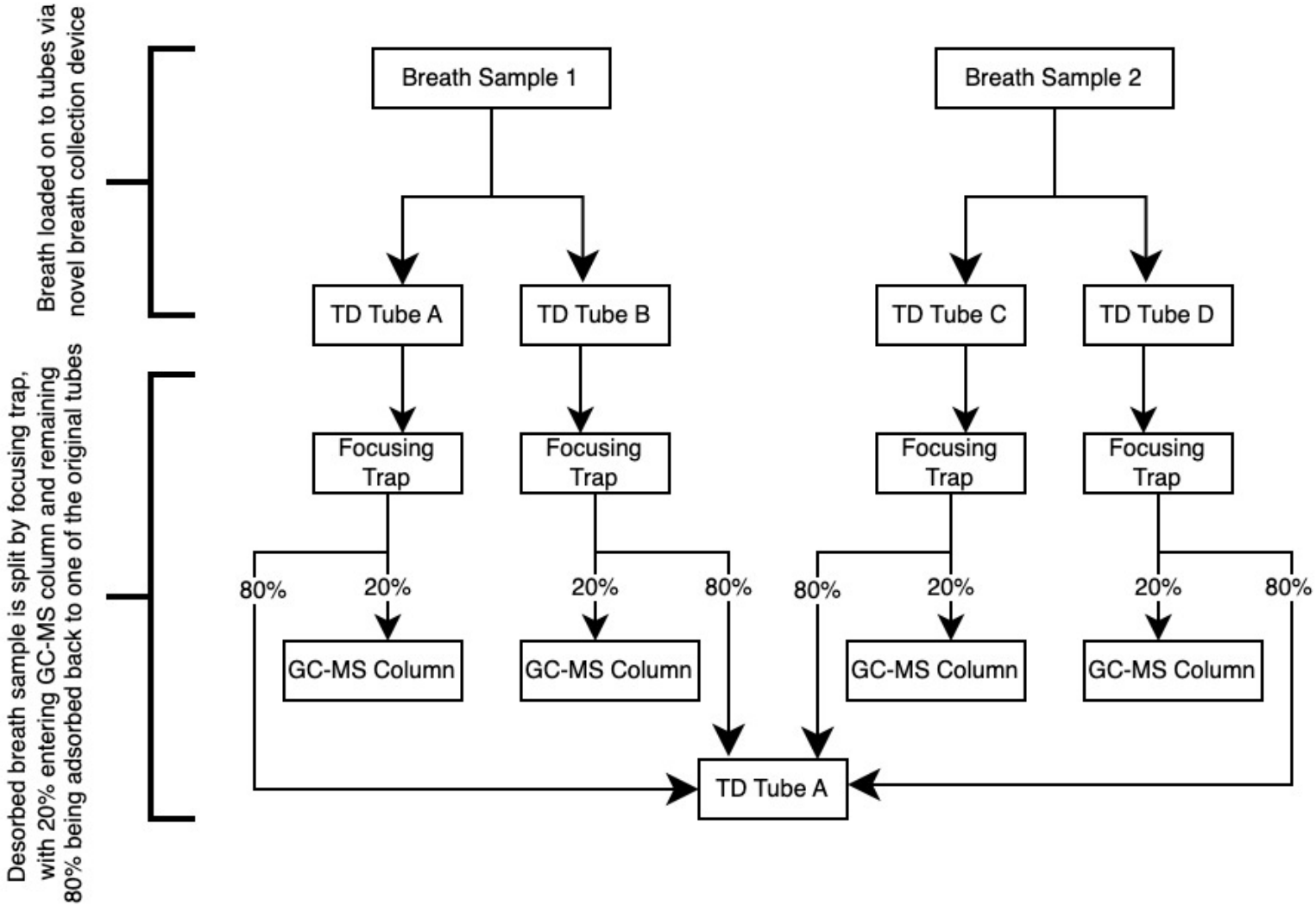


Figure 7.3 Flow chart of sample splitting study pathway 3 where all four TD tubes are recollected to a single TD tube

8 VARIATION OF VOC CONCENTRATIONS WITHIN AMBIENT ROOM AIR AND ITS IMPACT UPON THE STANDARDISATION OF BREATH SAMPLING

8.1 Study Design and Participants

The study took place over 10 non-consecutive weekdays in February 2020 at St. Mary's Hospital, London. Each day, two breath samples and four room air samples were collected in five locations. 300 samples were collected in total.

The five locations selected for collection of room air samples were a mass spectrometry instrument laboratory, a surgical outpatient clinic room, an operating theatres assessment area, an endoscopy assessment area, and a research bay within a clinical area. Each area was selected as they are regularly utilised for participant recruitment for breath analysis by my research group. All five sampling areas were temperature controlled at 25^{o}C.

8.1.1 Ethical Approval

Breath collection was performed under the ethical approval of "Non-invasive diagnostic testing for Gastro-intestinal disease" approved by the NHS Health Research Authority - London - Camden & Kings Cross Research Ethics Committee (reference 14/LO/1136) which includes the collection of breath from healthy volunteers for the purpose of developing the breath collection process. The single participant providing breath samples provided informed, written consent.

8.2 Sample Collection

8.2.1 Room Air

An air sampling pump (SKC Ltd., Dorset, UK) was used to draw ambient room air across Tenax™ TA/Carbograph™ inert-coated TD tubes (Markes International Ltd., Llantrisant, UK) at a rate of 250 mL/min for 2 minutes, loading a total of 500 mL of ambient room air on to each TD tube. The tubes were then sealed with air-tight brass caps (Swagelok Ltd., Kings Langley, UK) for transportation back to the mass spectrometry laboratory. Room air was sampled from each location in sequence between 9am and 11am each day and then again between 3pm and 5pm. Samples were collected in duplicate (i.e., two samples collected at the same time, in an identical manner under identical conditions for each room air measurement).

8.2.2 Breath Sampling

Breath samples were collected from a single subject who also performed the room air sampling. A custom-made, single use Nalophan™ bag with a 2000 mL capacity and a polypropylene syringe acting as a sealable mouthpiece was utilised for the collection of breath as previously described by Belluomo et al. (126). After spending a minimum of 10 minutes in each location at rest, the investigator exhaled into the sample bag via normal tidal breathing. Once filled to maximum volume, the bag was sealed with the syringe plunger. Immediately after collection, the air sampling pump was used to draw breath from the bag across TD tubes. A wide bore needle without a filter was attached to a TD tube via plastic tubing, with the air sampling pump at the other end. The bag was needled, and breath was drawn through each TD tube at a rate of 250 mL/min for 2 min, loading a total of 500 mL of breath on to each TD tube. Samples were once again collected in duplicate to minimise sampling variability. Breath was collected in the morning only.

For additional standardisation purposes, the investigator had nothing to eat or drink from midnight the previous evening and performed an oral water rinse prior to providing breath samples.

8.3 Sample Processing

Samples were processed via GC-MS with the same methods and settings as listed in section 6.3.1.

8.4 Data Analysis

Following visual inspection of the chromatograms, the raw data files were analysed using Chromspace® (Sepsolve Analytical Ltd.). Compounds of interested were identified from representative samples of breath and room air. Annotations were performed using NIST's 2017 mass spectral library based on VOC mass spectra and retention indices. Retention indices were calculated by analysing an alkane mixture (nC_8-nC_{40}, 500 µg/mL in dichloromethane, Merck, USA). 1 µL was spiked onto three conditioned TD tubes via a CSLR and analysed under the same TD-GC-MS conditions. From the raw compound list, only those with a RMF of greater than 800 were kept for analysis. Oxygen, argon, carbon dioxide and siloxanes were also removed. Finally, any compounds with a signal to noise ratio of less than 3 were also excluded. The relative abundance of each compound was then extracted from all data files using the compound list generated. 117 compounds were identified in breath samples. Peak picking was performed using MATLAB R2018b (version 9.5) and Gavin Beta 3.0 software (253). Following further interrogation of the data with visual

inspection of the chromatograms, a further 4 compounds were excluded leaving 113 compounds included in the downstream analysis. The abundance of these compounds was extracted from all 294 samples that were successfully processed. Six samples were removed due to poor data quality (leaked TD tubes). In the remaining dataset, 1-tailed Pearson correlation was calculated between the 113 VOCs in the repeated measurement samples to assess reproducibility. Correlation coefficients were 0.990±0.016 and p-values $2.00X10^{-46}$ ± $2.41X10^{-45}$ (arithmetic mean ± standard deviation).

8.4.1 Statistical Analysis

All statistical analyses were performed on R version 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria). Integrated peaks were first log transformed and then normalised using total area correction (TAC). Samples for which repeated measurements were available were collapsed to the mean. The `ropls` and `mixOmics` packages were used to generate the unsupervised PCA models and supervised PLS-DA models (255,256). PCA allowed for the identification of nine sample outliers. One breath sample clustered with the room air samples and therefore was felt to represent an empty tube secondary to sampling error. The other eight samples were room air samples driven by 1,1'-biphenyl, 3-methyl-. On further inspection, it was identified that all eight samples had significantly lower VOC yields compared to the other samples, suggesting these outliers were due to manual errors in loading the tubes.

Separation due to location was tested in the PCA using PERMANOVA from the `vegan` package. PERMANOVA allows the identification of group separation based on centroids. This technique has been previously used in similar metabolomic studies (257-259). The `ropls` package was used to evaluate PLS-DA models significance using a randomised 7-fold cross validation and 999 permutations. Compounds with a variable importance projection (VIP) score > 1 were considered relevant for the classification and retained as significant. Loadings from the PLS-DA models were also extracted to identify group contribution. Location specific VOCs were identified through consensus of pairwise PLS-DA models. To do so, all locations VOCs profiles were tested against each other and if a VOC with VIP > 1 was constantly significant in the models and attributed to the same location, it was then considered location specific. Comparison between breath and room air samples was investigated only on samples collected during the morning since no breath samples were collected in the afternoon. A Wilcoxon test was used for univariate analysis and false discovery rate was accounted for by applying a Benjamini-Hochberg correction.

9 VOLATILE ORGANIC COMPOUNDS FOR THE ASSESSMENT OF LIVER DISEASE (VOCAL) VIA SIFT-MS

9.1 Study Design

The study took place between February 2019 and April 2019. Individuals who were attending the hepatology unit at St. Mary's Hospital, London for transient elastography were identified as potential participants by their lead clinician. This cohort of participants provided a wide range of pathologies and varying severities of liver disease including those with a normal transient elastography (TE) score who could act as a control. As part of the protocol for TE, individuals were fasted for a minimum of 4 hours prior to their scan. Participants who were attending for routine review from dedicated cirrhosis clinics were also approached for recruitment. These participants had varying durations of fasting. As all dedicated hepatocellular cancer care is undertaken at a different clinical site and movement of the instrument between sites on a regular basis was impractical, no participants with HCC were recruited for this pilot study.

Individuals were invited to participate within the inclusion and exclusion criteria listed in table 9.1.

Table 9.1 Inclusion and exclusion criteria for the SIFT-MS VOCAL study.

Inclusion Criteria	**Exclusion Criteria**
Age 18-90 years. Attending secondary care hepatology services for transient elastography or clinical review. Able to understand and retain the information provided, thereby being able to give informed consent for inclusion in this study.	Individuals unable to provide deep breath exhalations through a narrow mouth-piece due to respiratory discomfort. Alcohol consumption within preceding 6 hours. Pregnancy.

All individuals provided written consent for participation. A case report form was completed for each participant, including any co-morbidities and medications, as well as a food diary of the preceding 24 hours.

Participants were sorted into control, fibrosis or cirrhosis cohorts based on their TE score as well as historical investigations and expert clinician opinion. Demographics and clinical data were recorded for all participants including age, height, weight, alcohol consumption, smoking status including time since last cigarette, regular medications, and co-morbidities.

9.1.1 Ethical Approval

All eligible participants were recruited under the ethical approval for the study "Non-invasive testing for the diagnosis and assessment of gastro-intestinal disease" (IRAS ID 142097, REC 14/LO/1136) which includes the recruitment of participants for breath analysis for GI disease.

9.2 Breath Sampling Protocol

Utilising the SYFT Voice200Ultra's compound library, two methods were developed to measure compounds previously identified in the literature as differing between healthy controls and individuals with liver disease. The two methods were developed to avoid compound overlap. A list of the compounds included in each method are shown in table 9.2 and the full method information (including ionisations and product ions) is available in the appendix 17.6, table 17.5.

Each participant was asked to provide three deep exhalations into the SIFT-MS via a disposable plastic mouthpiece over the course of 3 minutes. This process was repeated twice for two different methods. All participants were sampled in the same location (St. Mary's Hospital, Imperial College Healthcare Trust, London).

Room air samples were also taken on each sampling day by collecting room air in a single use 2000 mL Nalophan™ bag and attaching a needle to the inlet of the SIFT-MS. The bag was then needled to transfer the room air sample to the SIFT-MS.

Table 9.2. The different VOCs included within the two methods used with SYFT Voice200Ultra.

VOCs in Method 1	VOCs in Method 2
2-Nonanone	Ammonia
2-Octanone	Ethanol
2-Pentanone	Methanol
Acetone	Nonane
Butanone	Octane
Carbon Disulfide	Pentane
Dimethyl Sulfide	Tetradecane
Isoprene	Phenol
Limonene	
Styrene	
Acetaldehyde	
Indole	

9.2.1 Standardisation and Quality Control

Daily quality control checks, including instrument calibration, were performed using the SIFT-MS's pre-programmed software and internal calibrant gas. The instrument was kept between 20°C and 28°C. Participants remained in a resting state in the sampling location for a minimum of 10 minutes before sampling and undertook an oral water rinse before providing their breath samples. Additionally, all sampling was performed by the same investigator (MH) using the same processes in the same location.

9.3 Data Analysis

Using its inbuilt "Voice-Series" software, the concentration of target ions was extracted from each exhalation and the three readings were averaged for each participant.

9.4 Statistical Analysis

Potential confounding differences in demographics across groups were assessed using the non-parametric Mann-Whitney U test when comparing two groups only. Non-parametric tests were used as the demographic data was not normally distributed, as per a Shapiro-Wilk test. Statistical analysis was performed with IBM® SPSS® Statistics (Version 28.0.0.0) software. Differences in VOC concentrations between all three cohorts were assessed via a Kruskal-Wallis H test and with a Mann-Whitney U test when comparing only two cohorts.

10 VOLATILE ORGANIC COMPOUNDS FOR THE ASSESSMENT OF LIVER DISEASE (VOCAL) VIA GC-MS

10.1 Study Design

The VOCAL study is a prospective, non-randomised cohort study assessing the breath of participants with suspected or confirmed liver disease including cirrhosis and hepatocellular carcinoma. All recruitment took place between March 2021 and April 2022.

While the literature review demonstrated some recurrent VOCs that appear to correlate with hepatopathology, I decided to perform an untargeted analysis of VOCs to try and identify novel potential biomarker VOCs in addition to validating the VOCs identified in previous studies.

Recruitment was due to commence in April 2020 but was delayed by 11 months due to the COVID-19 pandemic. Recruitment commenced once Imperial College Healthcare Trust approved the resumption of onsite research activities and all recruitment and study activities took place according to local guidelines including use of personal protective equipment and social distancing where necessary. Individuals could only be recruited in line with existing clinical activities (e.g., outpatient appointments and imaging) and could not attend the hospital specifically for the study.

10.2 Participant Recruitment

Participants were recruited according to the criteria listed in table 10.1. Key stakeholders for recruitment were identified as hepatology consultants, specialist registrars and clinical and research nurse specialists working across Imperial College Healthcare Trust. I engaged with them via one-on-one meetings and via departmental presentations to educate them on the study and to aid with identifying opportunities for recruitment. I also engaged the equivalent team members for oncology services for recruiting those with hepatocellular carcinoma. I placed posters (appendix 17.7.6) in clinical areas frequented by individuals with hepatopathology with a brief description of the study and a dedicated contact email address (vocalstudy@imperial.ac.uk) for the study.

Table 10.1 Inclusion and exclusion criteria for the VOCAL study.

Inclusion Criteria	Exclusion Criteria
Age 18 years to 90 years. Seen in secondary care with suspected or confirmed liver disease or hepatocellular carcinoma cancer. Participants able to understand and retain the information provided, thereby being able to give informed consent for inclusion in this study.	Any participant with alcohol consumption within preceding 6 hours. Any participant where active COVID-19 infection is suspected. Any participant with respiratory distress or unable to exhale without respiratory distress. Any participant with an active, non-HCC, malignancy. Pregnancy

10.2.1 Participants with Cirrhosis

Several outpatient clinics based at St. Mary's Hospital for identified for recruitment. Appropriate participants were highlighted by the clinician in charge of the clinics, and I approached them either while they were waiting for their appointment, or after their consultation. I also approached individuals also attending for venesection (where their underlying aetiology was haemochromatosis) or elective large volume paracentesis after they were identified by their relevant clinical team.

10.2.2 Participants with Hepatocellular Carcinoma

A weekly joint hepatology-oncology clinic at Hammersmith Hospital was identified as the main opportunity for recruitment of individuals with hepatocellular carcinoma. A pre-clinic meeting was held with the clinical team to highlight appropriate individuals to approach for recruitment. Potential participants were also identified by interventional radiology teams for those attending for pre-treatment assessment clinics prior to radiofrequency ablation or trans-arterial chemo-embolisation.

10.2.3 Control Participants

While I considered recruiting the partners or relatives of those with pathology as the control cohort as per the studies by Fernandez Del Rio (177), O'Hara et al. (200), Miller-Atkins et al. (196) and Qin et al. (191), this was not possible due the COVID-19 pandemic. Due to social distancing requirements, those attending outpatient services were not permitted to be accompanied unless special circumstances dictated. Khalid et al.'s (207) approach of using

hospital staff was also considered, but during the COVID-19 pandemic there were insufficient members of staff working in the appropriate departments for this to be practical.

Therefore, after discussion with senior hepatology clinicians, it was decided to once again use individuals who had been referred for transient elastography due to concerns about possible liver disease but who had a result that was not consistent with cirrhosis or significant fibrosis, as the most appropriate, available control group. In the absence of histology, TE was felt to provide the best surrogate assessment of the presence or absence of liver disease. Note was made of any potential risk factors for liver disease (e.g., chronic viral hepatitis). Those with a TE score of <8 kPa and normal LFTs were deemed to be an appropriate control participant. In the event that a TE score was >8 kPa but <12.5 kPa, expert opinion was sought with regards to classification of the individual. The strength of this control group is that they all had a reference test.

10.2.4 Ethical Approval

The VOCAL study (IRAS ID 254249) was given research ethics committee (REC) approval on 09/10/2019 (reference 19/NW/0553) by Northwest - Liverpool East REC, and HRA and Health and Care Research Wales (HCRW) approval on 23/10/2019. The ethical approval considered the VOCAL consent form (appendix 17.7.1), participant information leaflet (appendix 17.7.2), case report form (appendix 17.7.3) and protocol (appendix 17.7.4). The study was sponsored by Imperial College London (appendix 17.7.5).

The VOCAL study and its associated protocol have been registered on public database at clinicatrials.gov (reference 19SM5129) (260).

10.3 Breath Sampling

All individuals that were approached were provided with the participant information leaflet and given an appropriate amount of time to read and understand it. An opportunity was given to ask questions before, during and after the consent process. It was made clear to all participants that they were free to withdraw their consent at any point during the process without any negative impact on their clinical care. Those who agreed to participate were asked to sign the consent form (appendix 17.7.1) and were provided with a copy. The case report form was then completed for each participant, taking down details including past medical history, weight, height, medications, and dietary information including oral intake in the preceding 24 hours. At this point several individuals were excluded.

Once consented, participants were sampled with the novel breath collection device and a single use breath collection bag as per the standard operating procedure (appendix 17.3).

The novel breath collection device was turned on and the system purged. Two Tenax™ TA/Carbograph™ inert-coated TD tubes (Markes International Ltd, Llantrisant, UK) were inserted into the sockets of the novel breath collection device and a single use custom breath collection bag was attached to the tubes. The novel breath collection device was set to collect breath at 200 mL/min for 150 s per tube, loading 500 mL of breath on to each tube. The participant was instructed to exhale with normal tidal breathing into the bag through the custom mouthpiece until they felt resistance. The participant was permitted multiple exhalations to fill the bag if needed. They were instructed to not hold their breath or forcefully exhale. In this fashion, whole breath was collected.

Once the bag was filled, the mouthpiece was capped, and the breath drawn across the TD tubes by the novel breath collection device. Once the breath collection device had finished drawing breath across the tubes, they were removed and sealed with air-tight brass caps (Swagelok Ltd., Kings Langley, UK) using spanners. Capped tubes were then sealed in single-use plastic sampling bags which were labelled with the study ID and the date and time of sample collection and transported back to the mass spectrometry laboratory within 4 hours. The unique tube bar code numbers were noted on the case report from to allow matching of tubes to participants. The breath collection device was then purged in preparation for the next sample and the breath bag disposed of within a clinical waste bin. Only two tubes were collected per participant due to limitations to the number of available TD tubes at the time of recruitment.

10.3.1 Quality Control and Standardisation

Prior to providing a breath sample, participants were asked to rinse their mouth with water to reduce contaminant VOCs from the oral cavity (103). While it was desirable to have had individuals fasted for a set amount of time, individuals had to be recruited in line with pre-existing clinical commitments and could not be brought back after a period of fasting specifically for the study. Note was made for all individuals as to how long it had been since they last ate or drank as well as their smoking status and time since their last cigarette.

Recruited participants were rested for a minimum of 5 minutes prior to providing breath samples. Whilst a longer period was desired, it was not possible due to the restrictions in the number of individuals who could be within the department at any one time.

Samples of ambient room air were collected to assess background concentrations of VOCs. The novel breath collection device was run without a breath sampling bag attached so that ambient room air was drawn directly across the TD tubes with the same settings as used for breath collection. While the intention was to collect paired room air samples for every breath sample, due to a persistent shortage of TD tubes and a desire to prioritise the limited number of participants with pathology attending in person appointments, this was not possible.

All sampling episodes at St. Mary's Hospital occurred in the same room and all sampling episodes at Hammersmith Hospital took place in the same department. All breath samples were collected using the same processes and conducted by the same person (MH).

10.4 Sample Processing

All TD tubes used were cleaned prior to being loaded with breath as detailed in section 6.2.1 and spiked with a working internal standard mixture as detailed in section 7.3.1. If immediate analysis was not available, the TD tubes were stored within 4 hours of collection in freezers at -80°C. They were then allowed to return to room temperature prior to analysis. The duration for which they were stored at -80°C was noted.

10.4.1 GC-MS Analysis

Samples were processed via GC-MS with the same methods listed in section 6.3.1 with the exception of the choice of GC column. For this experiment, the chromatographic column was a Rxi-624Sil MS column (Restek UK Ltd., Ripley, UK) (30 m x 0.25 mm x 1.40 µM). This is a column with a mid-polarity stationary phase that has been designed to provide inertness and thermal stability alongside detection of a wide variety of VOCs (261).

The focusing trap's sample splitting setting was engaged. 80% of each duplicate breath sample was recollected on to a single conditioned Tenax™/Carbograph™ TD tube to allow further analysis of samples via another mass spectrometry platform.

A total of 316 samples (312 breath samples from 156 participants, 4 room air samples) were processed via GC-MS. The 312 breath samples were recollected via recollection pathway two (see section 7.2.3, figure 7.2). While duplicate samples were run together, the order of running samples was otherwise randomised.

10.5 Data Analysis

10.5.1 Mass Spectra Data Extraction

Following visual inspection of the chromatograms, the raw data files were analysed using Chromspace® (Sepsolve Analytical Ltd.). Dynamic baseline correction (DBC) was performed with a peak width of 6.0 seconds. All files were then converted to .lsc format. Using Chromspace®'s alignment software, all samples were aligned together based on a representative sample. It was identified that the alignment and deconvolution performed by Chromspace® (Sepsolve Analytical Ltd.) was liable to error, with single peaks being misinterpreted as multiple peaks and consequent incorrect identification. In order to address this, a compilation script was developed by a member of my research group (BD) which overlay all aligned chromatogram files, generating a single file (figure 10.1). This file then underwent a single deconvolution. From this, a peak table was generated. Each peak was manually reviewed to ensure only reliable peaks were included (i.e., not inappropriate selection of background noise). The peak windows were optimised, and the appropriate quantitative/source ions selected. All samples then underwent a targeted integration based on this method, providing peak area values for each ion from individual samples. Using this method, 859 compounds were identified, and their peak areas extracted.

10.5.2 VOC Filtering

Oxygen, argon, and carbon dioxide were removed as they cannot be considered biomarkers for hepatopathology. Siloxanes were also removed as these are generally contaminants from the mass spectrometry instrument. Any compounds with a signal to noise ratio of less than 3 were also excluded. The peak table was imported into Chromcompare+ by Chromspace® (Sepsolve Analytical Ltd.) and initial filtering applied with an absolute minimum ion intensity of 10,000 counts and no maximum intensity. This generated 362 compounds for inclusion in my analysis. Data was also inspected with ion intensity counts of 100 but best separation of data was observed with 10,000 counts.

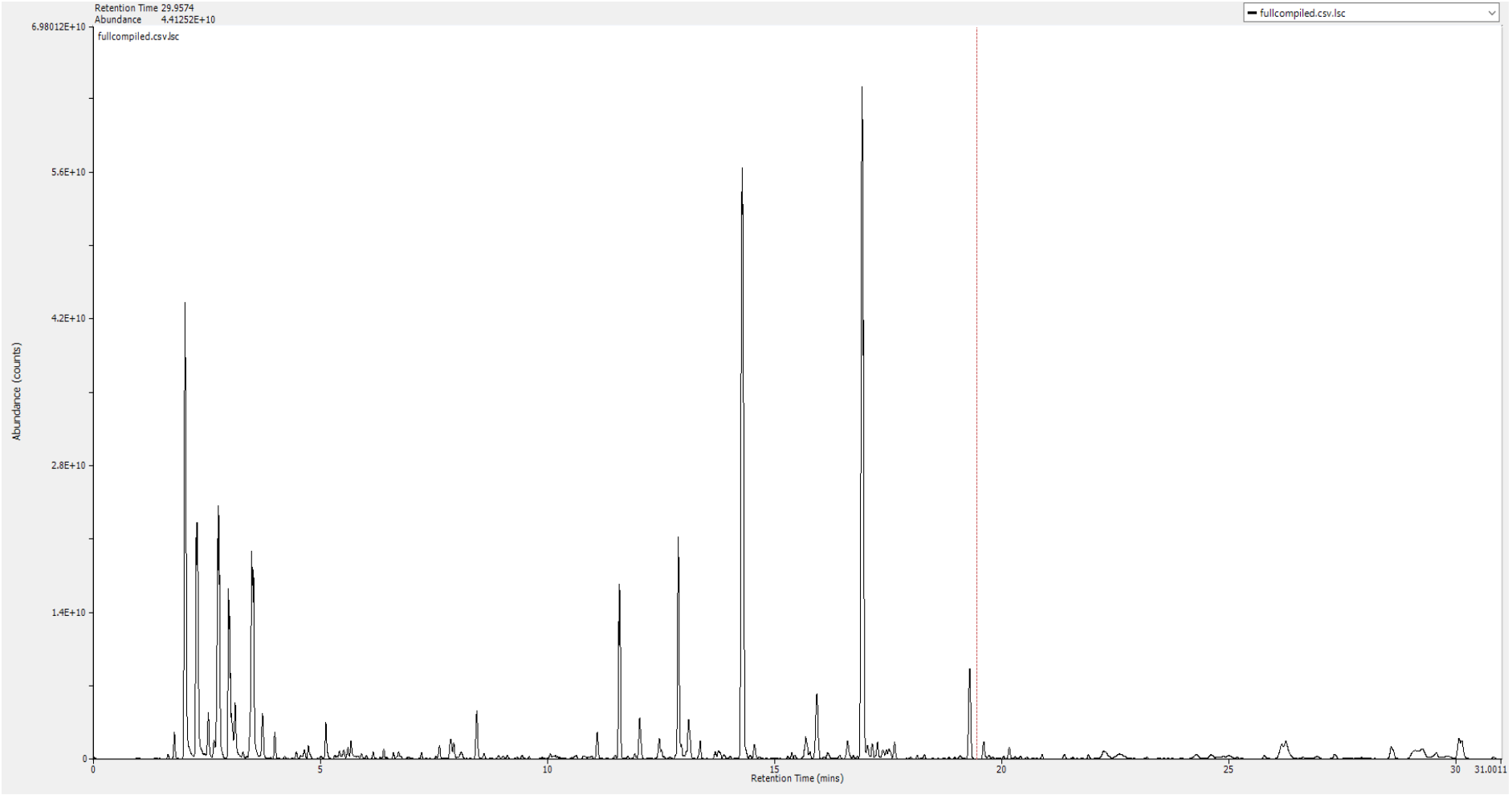


Figure 10.1 Chromatogram consisting of all samples overlayed with each other.

10.5.3 Data Transformation and Scaling

Inspection of the data with SIMCA® (Version 17) revealed a data set that was not normally distributed. Therefore, the data was log transformed to make it less skewed. A log10 (10000*X+1) transformation was used. Pareto and unit variance (UV) scaling were both tested. Visual inspection of the graphs generated demonstrated that UV scaling providing a more normal distribution (see appendix 17.7.7). UV scaling was therefore applied to further provide a more normal distribution of data while giving all variables equal weighting. While Hotelling's T^2 distribution was reviewed, it was ultimately decided to not exclude any samples that were identified as outliers based on this parameter in case these anomalies could be accounted for by disease state.

Multiple different normalisation strategies were explored including probabilistic quotient normalisation (PQN) (169), total area correction (TAC), normalisation by internal standard, (whereby samples are normalised according to the peak area of an internal standard that has been loaded on to all samples at a set quantity) and normalisation by internal standard, followed by PQN. The internal standard selected for our normalisation strategy was toluene d_8, which was loaded on to every tube prior to analysis as part of an internal working standard solution (see section 7.3.1).

The PCA plots from each normalisation strategy can be seen in appendix 17.7.8, figures 17.15 to 17.18. There was minimal difference between the plots and therefore normalisation by internal standard and PQN was selected. This normalisation strategy has been used in previous metabolomic studies (262), shows a homogenous sample distribution in the PCA plot, and also ensures effective elimination of potential sample preparation or instrumental drifts.

The coefficient of variation for each peak for duplicates was calculated. Those with an average CV% >30 were excluded from the analysis due to the variability between duplicates potentially skewing the results. Duplicate samples with an average CV% <30 were averaged to provide a single value per participant for inclusion in the analysis.

10.5.4 Statistical Analysis

Multivariate statistics were performed using SIMCA® (Version 17) software by Sartorious AG. Differences between cohorts were initially assessed via unsupervised principal component analysis (PCA) followed by supervised analysis via orthogonal projections to latent structures discriminant analysis (OPLS-DA). The significance of any separation

between cohorts was assessed by cross-validated analysis of variance (CV ANOVA) testing. Where separation was significant, the VOCs driving the separation were ranked according to their variable importance projection (VIP) and those with a score >1.5 were further interrogated to confirm appropriate identification by Chromspace®.

MetaboAnalyst 5.0 by XiaLab (McGill University, Montreal, Canada) was used to apply the Benjamini-Hochberg correction, a statistical tool which aims to control for the fact that small p-values (i.e., $p < 0.05$) can also happen by chance rather than representing true significance. This is also referred to as the false discovery rate (FDR). Those VOCs with an FDR corrected p value of >0.05 were excluded from analysis. MetaboAnalyst was then used to generate receiver operating characteristic (ROC) curves via random forests classification and support vector machine (SVM) feature ranking for individual VOCs and models containing all VOCs driving separation, as well as smaller models of up to five VOCs. The areas under the ROC curves were used to measure the ability of the VOCs or models of VOCs to discriminate between groups.

Univariate statistics were performed for individual VOCs with IBM® SPSS® Statistics (Version 28.0.0.0) software. As the results were non-parametric, statistical significance was assessed via the Kruskal-Wallis H test when comparing all three groups and via a Mann-Whitney U test when comparing two groups.

SECTION 3: RESULTS

11 DEVELOPMENT OF A NOVEL BREATH COLLECTION DEVICE FOR BREATH VOC ANALYSIS

11.1 Repeatability

After filtering, 97 VOCs remained for inclusion in the analysis. The average coefficient of variation was calculated across all VOCs for all five participants (figure 11.1). A CV% of <30 correlates with acceptable repeatability (263). My results demonstrated:

33 VOCs with CV% >30
64 VOCs with CV% <30

The average of all CV% was 31. VOCs with CV >30 had generally small peak areas (figure 11.2).

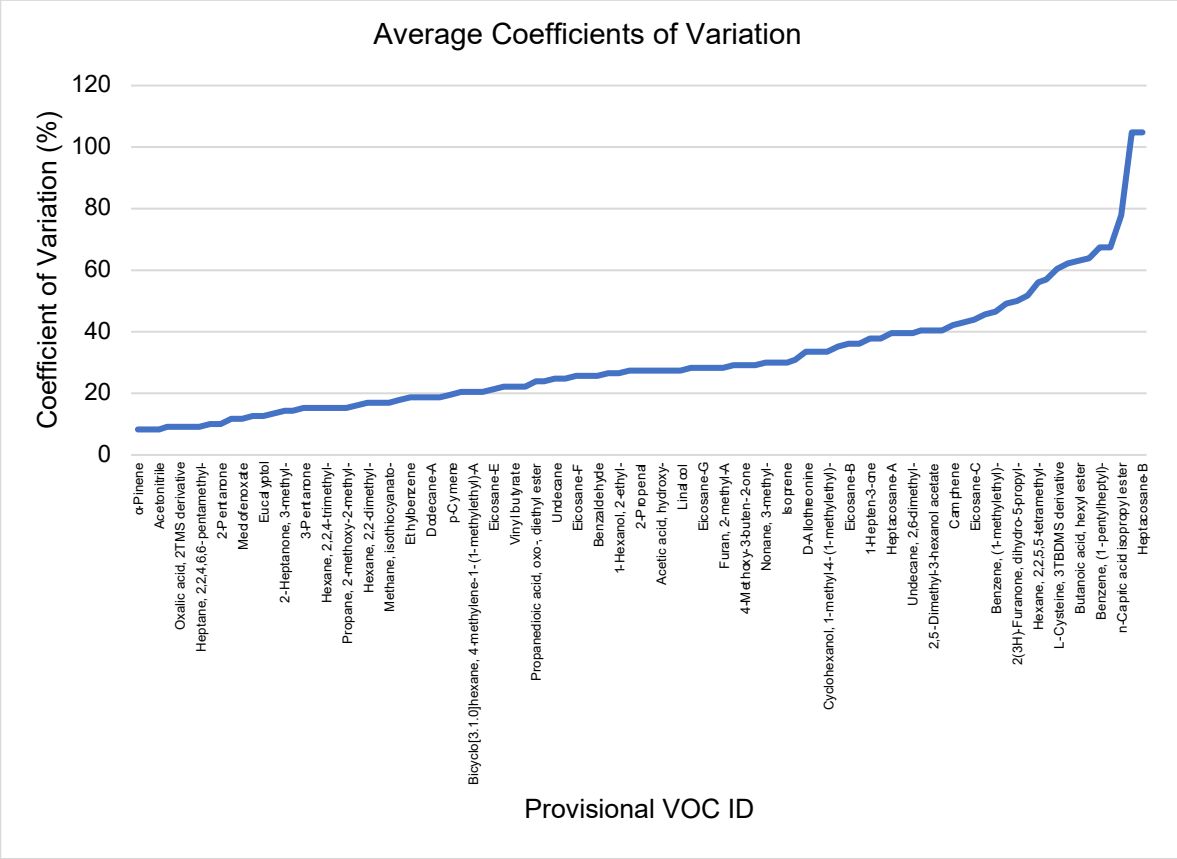


Figure 11.1 Graph of average coefficient of variation for individual VOCs.

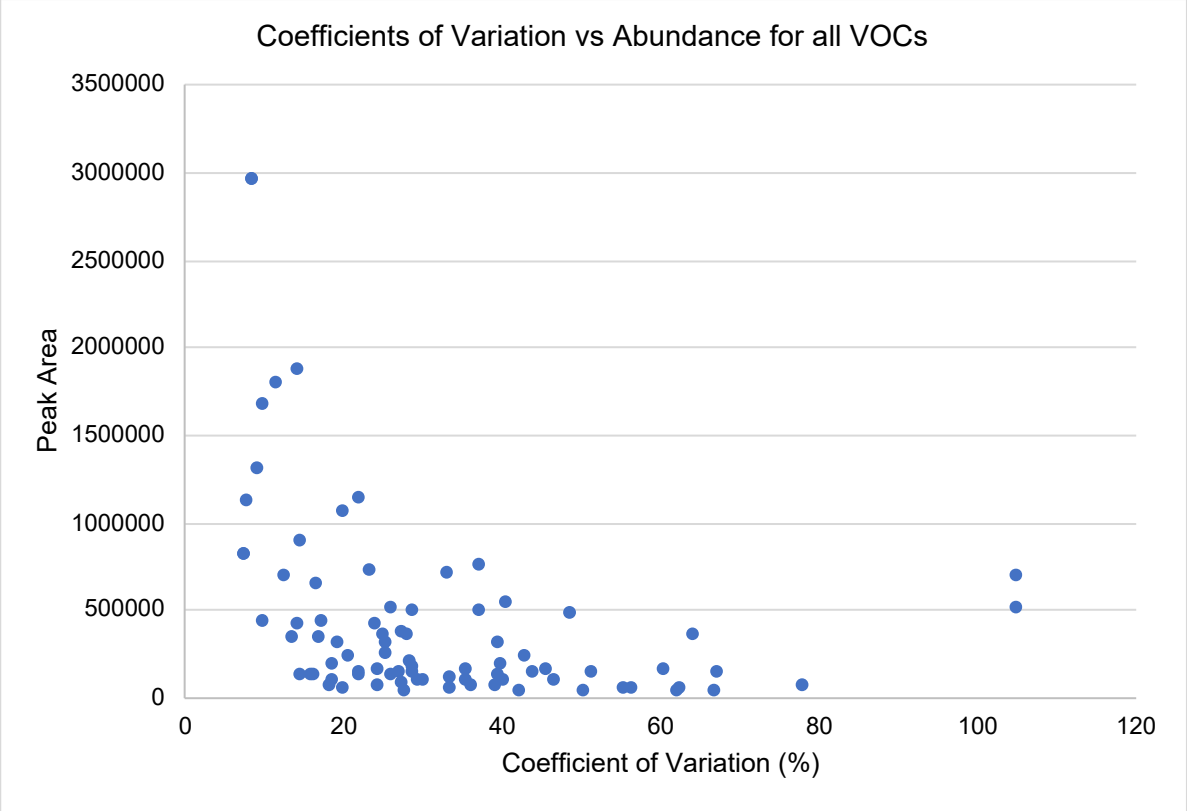


Figure 11.2 Graph of peak area of VOCs against their coefficient of variation demonstrating those with a higher CV% had lower peak areas.

11.2 Flow Rates

The average total peak area of all samples collected at 100 mL/min and of all samples collected at 200 mL/min was calculated (figure 11.3). This revealed a small reduction in the average of the total peak areas with the flow rate of 200 mL/min compared to 100 mL/min. This was not statistically significant (p value 0.12).

According to a two-tailed t-test, only 23 of the 97 VOCs identified had a difference between flow rates that was statistically significant ($p < 0.05$). The p values of all VOCs can be found in appendix 17.2, table 17.1. For the 23 significant compounds, the average total peak area was higher with the slower 100 mL/min flow rate.

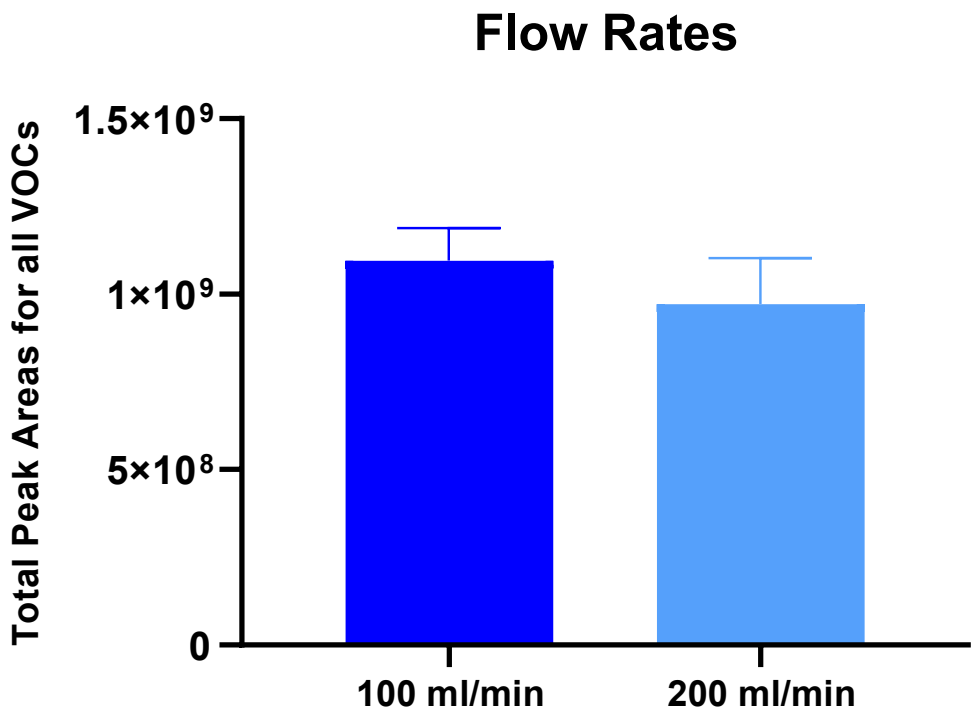


Figure 11.3 Bar chart demonstrating the total peak area for differing flow rates of novel breath collection device with a slightly higher total peak area for 100 mL/min compared to 200 mL/min.

11.3 Volume of Breath

The sum of all VOC peak areas for each volume was calculated and plotted against the volume (figure 11.4). This initially demonstrates linearity between volume and total peak area but at higher volumes it potentially begins to show breakthrough, i.e., the higher volumes lead to VOCs being potentially eluted off the sorbent bed and therefore lost (264).

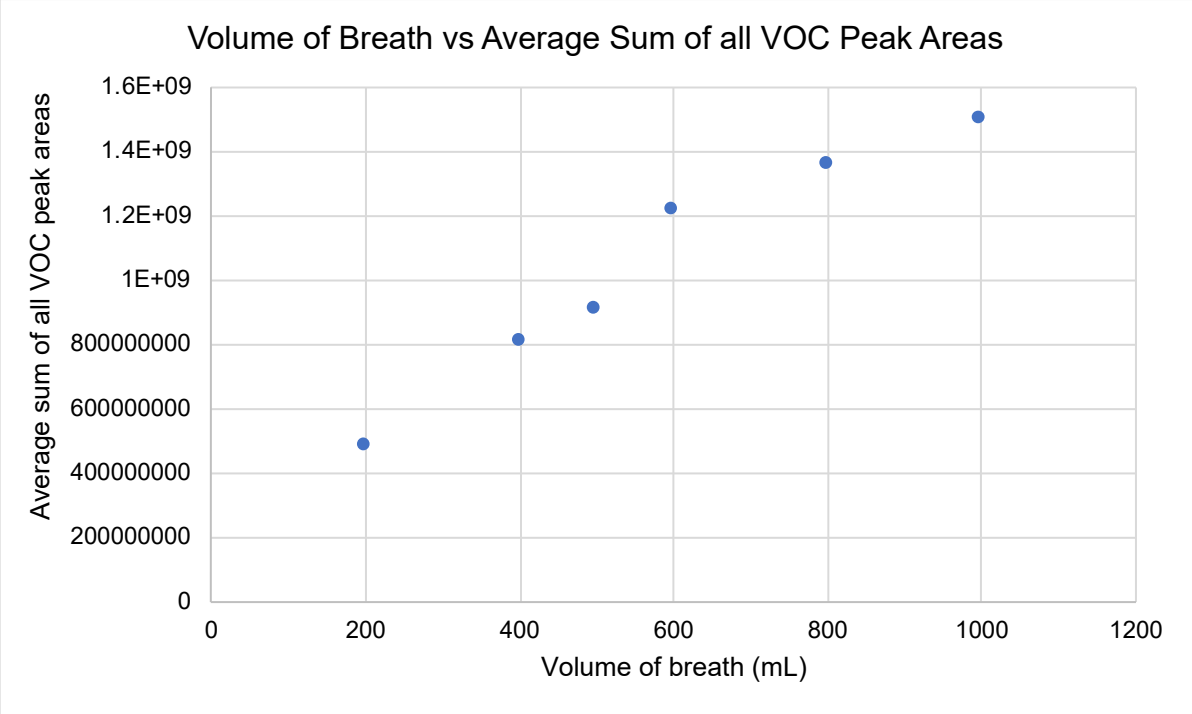


Figure 11.4 Graph of average sum of all total peak areas against volume of breath with evidence of possible breakthrough at higher volumes.

A Spearman rank correlation coefficient was performed for each individual VOC. 62 of the 97 VOCs had a coefficient >0.7 demonstrating a strong positive correlation between peak area and volume (see appendix 17.2, figure 17.1 for individual scatter graphs). For the remaining compounds, visual inspection of graphs demonstrated a degree of linearity for some VOCs but minimal correlation between total peak area and volume for others.

11.4 Discussion

Testing of the novel breath collection device has demonstrated acceptable repeatability for the majority of VOCs identified with the methods used. This includes VOCs of interest for my clinical study including alkanes, ketones, and limonene, all of which showed good repeatability.

While the sum of all total peak areas was higher for a flow rate of 100 mL/min, this was not statistically significant, and the majority of individual VOCs did not have a significantly higher yield with a flow rate of 100 mL/min compared to 200 mL/min. The latter will provide a more efficient sampling process for participant and researcher. 200 mL/min is also comparable to the flow rates which have been used by studies utilising the ReCIVA® (204,209) and also correlates with a previous methodology study by Doran et al. (181) who selected 200 mL/min as the optimum flow rate for loading TD tubes with the ReCIVA®.

VOC yield increased linearly for the majority of VOCs with possible breakthrough at higher volumes. Those VOCs with no correlation between total peak area and volume were comparatively low in abundance and may represent contaminant VOCs.

11.5 Study Limitations

It must be acknowledged that there are limitations to this study which may impact upon the confidence of the results presented.

Firstly, I only tested the novel breath collection device with breath and not with analytical standards. This was primarily due to resource constraints and mass spectrometry instrument availability. Given these restrictions, all experiments were designed with the intention of replicating the real-life function of the device. Additionally, breath cannot easily be replicated artificially due to its complexity and vast numbers of VOCs within it. It is acknowledged that testing the device with analytical standards would provide additional reassurance of the robustness of the flow rates and volumes settings chosen and this will be performed when testing the next generation of devices.

Additionally, due to a paucity of TD tubes (due to instrument workflow issues, storage of tubes and prioritisation of TD tubes for clinical studies), it was only possible to recruit a relatively small numbers of individuals for the study. For flow rates, it was technically possible to test 50 mL/min and 150 mL/min, but due to the lack of TD tubes, it was decided to prioritise testing 100 mL/min and 200 mL/min with a higher number of individuals than a greater number of flow rates with a smaller number of individuals.

It is also acknowledged that the choice of a polar wax GC column for the analysis with GC-MS is controversial. As the intention was for the novel breath collection device to be utilised with all clinical studies going forwards and the Mega WAX-HT column was being used by our research group to assess VOCs in oesophagogastric cancer (due to its ability to detect short chain fatty acids, an important class of VOCs in oesophagogastric cancer), it was deemed acceptable to use this column for assessing the novel breath collection device. Additionally, at the time of the study, only one GC-MS instrument was available and was consequently in significant demand from multiple studies. With the high pressure on the instrument workflow, the existing column was kept for the methodology work to minimise instrument downtime. Additional studies with GC columns with different stationary phases are required to confirm the results are applicable to a wider range of VOCs.

11.6 Conclusion

Within the limitations of the methodology, testing of the novel breath collection device has demonstrated acceptable repeatability for the majority of VOCs detected, allowing the same individual to provide multiple breath samples while still allowing accurate comparison of results. While the choice of column for the GC-MS has compromised the VOCs detected, there was still a wide variety of VOCs detected including VOCs of interest for my clinical study. As the majority of VOCs did not have a significantly higher yield with the slower flow rate and as the higher flow rate provides a more efficient sampling process for the participant and researcher, a flow rate of 200 mL/min was selected for inclusion in our standard operating procedure. Testing of larger volumes of breath is required to confirm breakthrough volumes for individual VOCs. Given the possibility of breakthrough at higher volumes, 500 mL has been selected at present as the volume of breath for inclusion in the standard operating procedure. Reassuringly, despite the limitations in methodology, these parameters correlate with pre-existing literature for alternative breath collection devices (181). Additional studies on the next generation of the device however are required to assess slower flow rates and higher volumes. Given the relative affordability of this device compared to alternative breath collection devices, it can potentially facilitate expansion in research on exhaled VOCs which to date has often been limited by the costs and availability of breath collection equipment that can be restrictive for multi-centre studies.

With the results generated from this study, I have formulated a standard operating procedure for collection of breath samples with the novel breath collection device. This was devised in conjunction with Miss Sara Jamel and can be found in appendix 17.3.

12 OPTIMUM STRATEGY FOR RECOLLECTION OF BREATH SAMPLES VIA SAMPLE SPLITTING DURING THERMAL DESORPTION

12.1 Breath vs Nitrogen

Visual inspection of the initial PCA plots reassuringly demonstrated excellent separation of nitrogen from breath samples (appendix 17.4, figures 17.2-17.4) for all three recollection pathways. Following PQN normalisation, this separation was enhanced (appendix 17.4, figures 17.5-17.7).

12.2 Pathway One vs Pathway Two vs Pathway Three

Nitrogen samples were removed from the normalised data set and the PCA plots of each pathway generated (figures 12.1-12.3). While all three PCA plots demonstrated clustering of individual participants' breath samples and good separation, pathway two appeared to have the best separation. Pathway three featured some overlap of different participants.

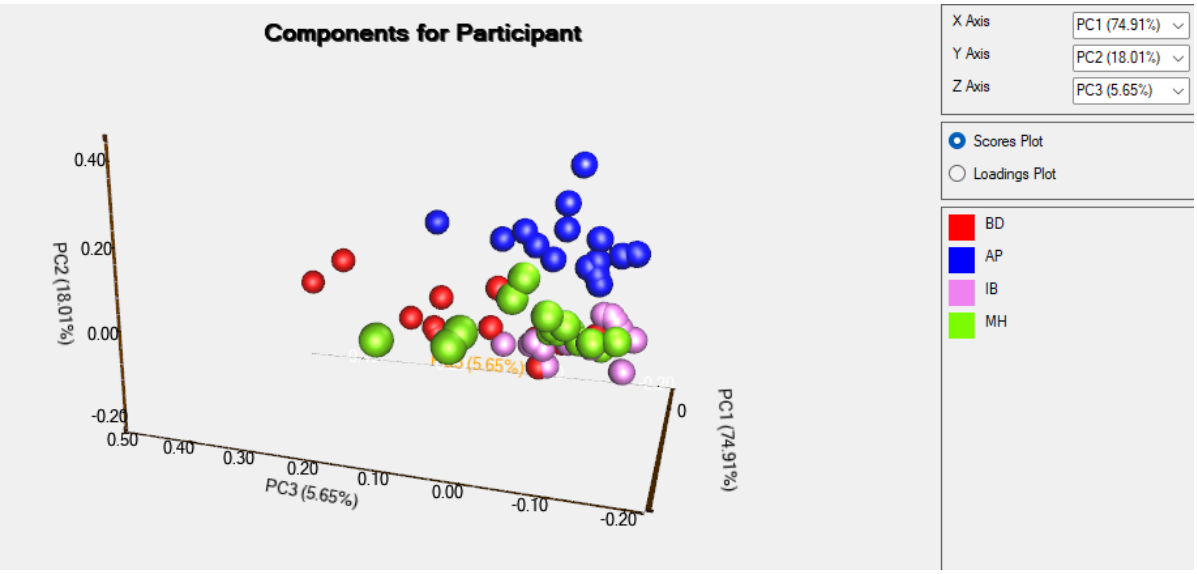


Figure 12.1 PCA plot of recollection pathway one with PQN normalisation and nitrogen blanks removed demonstrating clustering of breath samples for each participant and good separation between different participants.

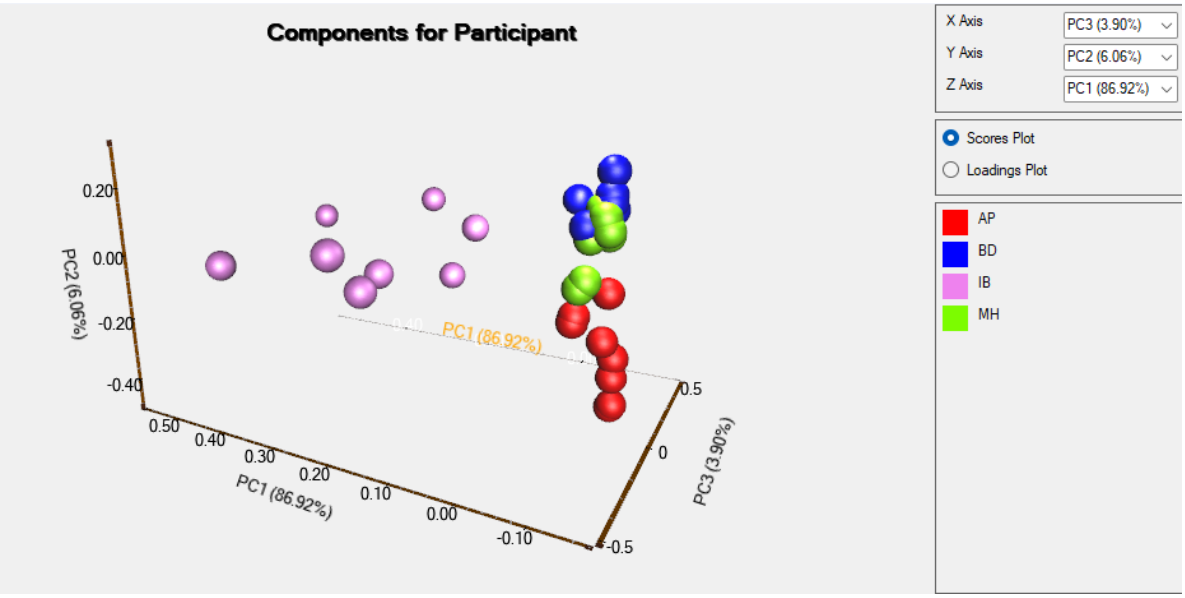


Figure 12.2 PCA plot of recollection pathway two with PQN normalisation and nitrogen blanks removed demonstrating clustering of breath samples for each participant and improved separation between different participants compared to pathway one.

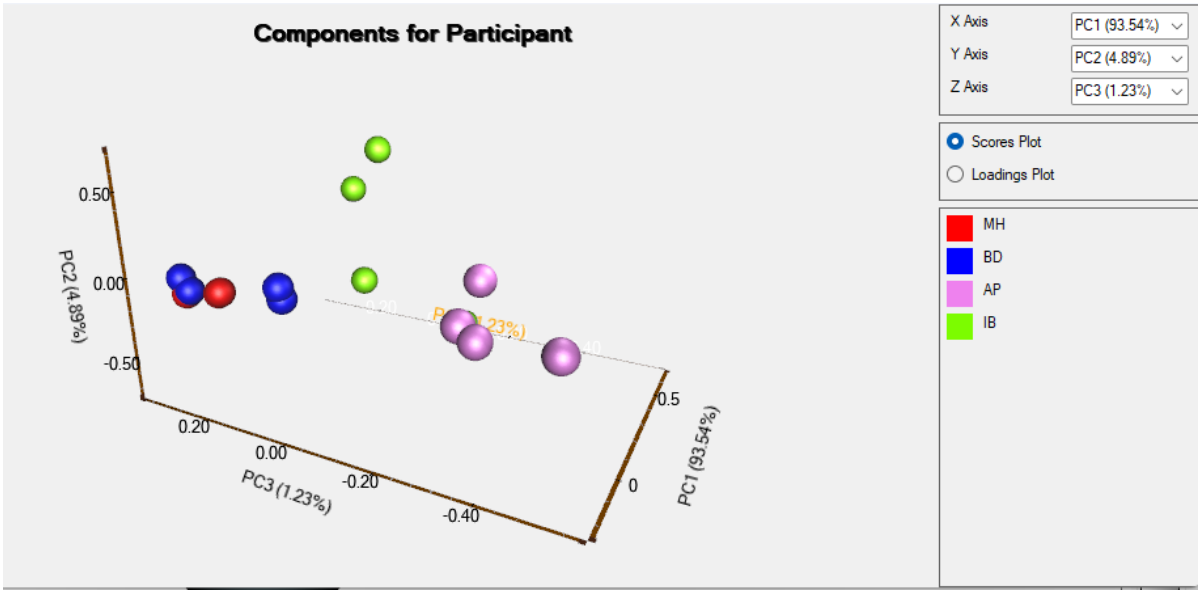


Figure 12.3 PCA plot of recollection pathway three with PQN normalisation and nitrogen blanks removed demonstrating clustering of breath samples for each participant and separation between most participants but some overlap between BD and MH.

Visual inspection of the PCAs was supplemented by the confidence scores generated by Chromcompare+ by Chromspace® which can be seen in table 12.1.

Table 12.1 Summary of confidence scores for each analysis with TD tubes recollected via pathway two demonstrating the best confidence in discriminating between individuals.

Pathway	No Normalisation (%)	PQN Normalisation (%)	PQN Normalisation and Nitrogen Removed (%)
1	60.7	80.9	88.5
2	68.6	93.8	99.4
3	79.0	80.5	92.0

The confidence scores correlated with the visual inspection of the PCAs, with the best discrimination between participants by pathway two i.e., with two samples being recollected on to one tube (99.4%).

12.3 Discussion

In this experiment, I demonstrated three different strategies of splitting breath samples for recollection of breath samples back on to TD tubes. Recollecting two duplicate samples on to one TD tube provided the best discrimination between participants and will be incorporated into standard practice going forwards. Splitting and recollection of samples will allow breath samples to be analysed on multiple complementary mass spectrometry platforms which may enhance compound detection and identification. It may also allow a reduction in the number of breath samples collected per person, providing a more efficient sampling process for our research group.

12.4 Study Limitations

This study was limited by the number of participants recruited for the study. Once again, at the time of the experiment, there was a shortage of TD tubes due to prioritisation of time critical clinical studies over methodology work. Ideally this study would be repeated with additional participants to improve confidence in results. While the predominant focus is on discrimination of participants between samples, it is acknowledged that breakthrough of VOCs due to higher volumes of VOCs being loaded on to recollected samples may pose an issue and further interrogation of specific VOCs is necessary to determine if breakthrough occurs with recollection of multiple samples back on to a single tube.

12.5 Conclusion

Within the limitations established, this study has demonstrated that VOCs from desorbed TD tubes can be successfully split, with proportion recollected back on to a TD tube. This can facilitate multi-platform analysis of the same sample with a reduced number of TD tubes.

Recollecting two TD tubes back on to one TD tube provided the best separation between individuals and therefore it is still recommended that samples are collected in duplicate.

13 VARIATION OF VOC CONCENTRATIONS WITHIN AMBIENT ROOM AIR AND ITS IMPACT UPON THE STANDARDISATION OF BREATH SAMPLING

13.1 Breath vs Room Air

Breath samples were collected in the morning alongside matched room air samples at five different locations and analysed by GC-MS. As expected, visual inspection of the chromatograms demonstrated significant differences between room air and breath samples (appendix 17.5, figures 17.8 and 17.9). Visual comparison of the total peak areas of room air samples to breath samples also reassuringly shows significantly higher peaks for the breath samples (figure 13.1).

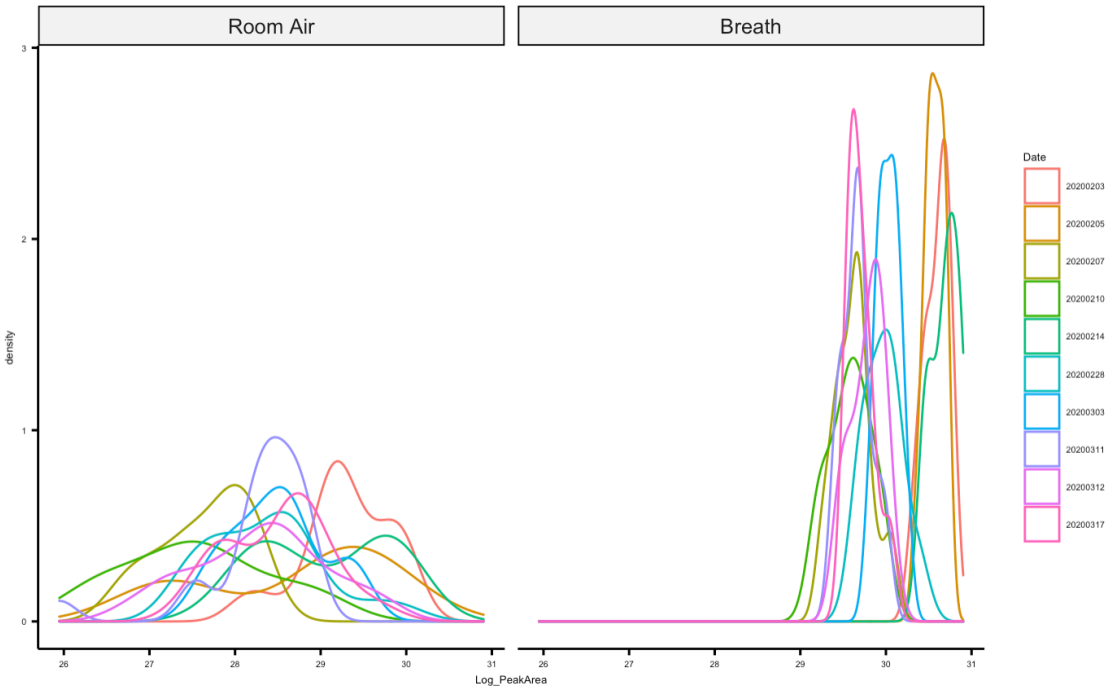


Figure 13.1 Log transformed total peak areas of samples demonstrating clear differences between the size of peaks for room air and breath samples.

A total of 113 VOCs were detected and extracted from the chromatograms. Repeated measures were collapsed to the mean before performing PCA on the extracted and normalised peak areas to identify and remove outliers. Supervised analysis through partial least squares-discriminant analysis (PLS-DA) was then able to show a clear separation between breath and room air samples ($R^2Y = 0.97$, $Q^2Y = 0.96$, $p<0.001$) (figure 13.2). Group separation was driven by 62 different VOCs, with a variable importance projection

(VIP) score >1. A complete list of the VOCs characterizing each sample type and their respective VIP scores can be found in appendix 17.5, table 17.2.

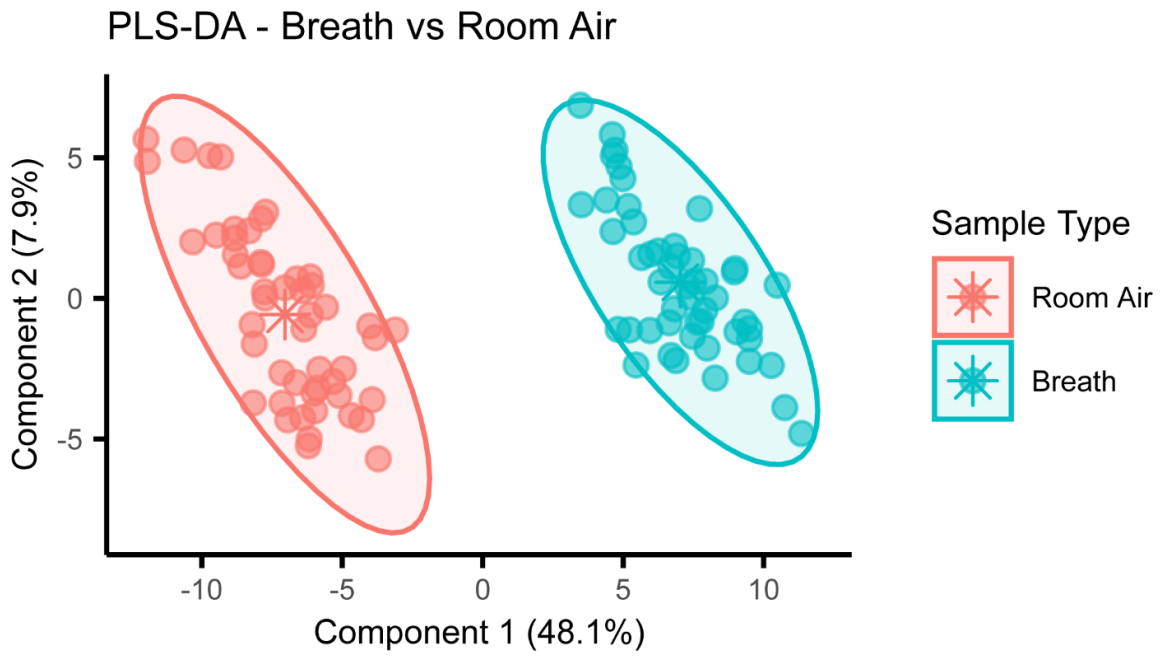


Figure 13.2 PLS-DA of Breath vs Room Air demonstrating breath and room air present distinct VOCs profiles.

13.2 Diurnal Variation in Room Air VOCs Concentrations

Differences in room air VOC profiles between morning and afternoon samples were investigated using PLS-DA. The model identified significant separation between the two timepoints ($R^2Y = 0.46$, $Q^2Y = 0.22$, $p<0.001$) (figure 13.3). This was driven by 47 VOCs with a VIP score >1. VOCs with the highest VIP score characterizing morning samples included multiple branched alkanes, oxalic acid and hexacosane, while afternoon samples presented more 1-propanol, phenol, propanoic acid, 2-methyl-, 2-ethyl-3-hydroxyhexyl ester, isoprene and nonanal. A comprehensive list of VOCs characterising daily variation in room air composition can be found in appendix 17.5, table 17.3.

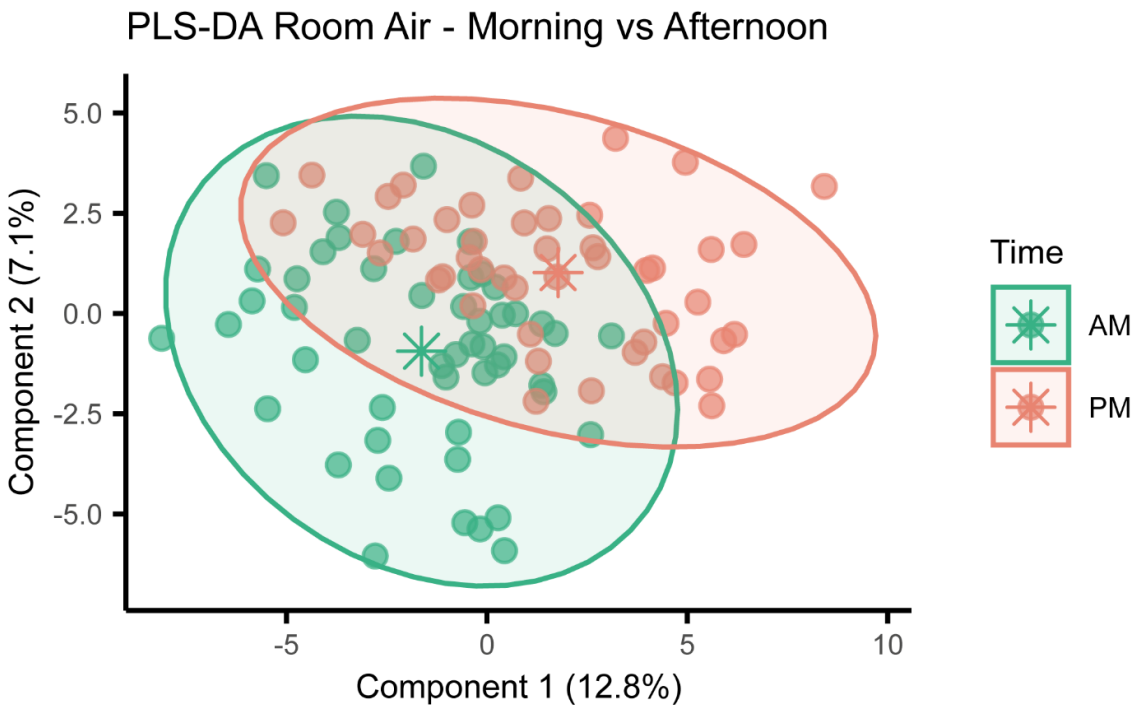


Figure 13.3 PLS-DA for room air in morning vs afternoon demonstrating that room air VOC profiles change during the day.

13.3 Variation by Sampling Location

Samples were collected across five different locations: endoscopy unit, clinical research bay within the hepatology department, operating theatre assessment area, outpatient clinic and a mass spectrometry laboratory within St Mary's Hospital, London. These locations are all commonly used for patient recruitment and breath collection by my research group. Room air was collected both in the morning and afternoon, while breath samples were only collected in the morning.

PCA highlighted a separation of room air samples by location through permutational multivariate analysis of variance (PERMANOVA, R^2 = 0.16, $p<0.001$) (figure 3.25a). Thus, pairwise PLS-DA models were generated, comparing each location against all the others to identify characteristic signatures. All models were significant and VOCs with a VIP score >1 were extracted with respective loading to identify group contribution. The results indicate that the composition of ambient air changed by location, and I identified location characteristic signatures through model consensus. The endoscopy unit was characterized by a higher presence of undecane, dodecane, benzonitrile and benzaldehyde. The clinical research bay (also identified as liver research unit) samples displayed more α-pinene, di-isopropyl phthalate and 3-carene. The operating theatre assessment area air was distinguished by a more abundant presence of branched decane, branched dodecane,

branched tridecane, propanoic acid, 2-methyl-, 2-ethyl-3-hydroxyhexyl ester, toluene and 2-butenal. The outpatient clinic (Paterson building) was marked by higher concentrations of 1-nonanol, vinyl lauryl ether, benzyl alcohol, ethanol, 2-phenoxy-, naphthalene, 2-methoxy-, isobutyl salicylate, tridecane, and branched tridecane. Finally, the room air collected in the mass spectrometry laboratory presented more acetamide, 2,2,2-trifluoro-N-methyl-, pyridine, furan, 2-pentyl-, branched undecane, ethylbenzene, m-xylene, o-xylene, furfural, and ethyl anisate. Varying concentrations of 3-carene were present in all five locations, suggesting this VOC to be a common contaminant, with highest abundance observed in the clinical research bay.

A list of consensus VOCs separating each location can be found in appendix 17.5, table 17.4. In addition, univariate analysis was performed on each VOC of interest, comparing all the locations to each other with a pairwise Wilcoxon test followed by Benjamini-Hochberg correction. Boxplots for each VOC are reported in appendix 17.5, figure 17.10. Overall breath VOC profiles did not appear to be affected by location as observed in PCA followed by PERMANOVA ($p = 0.39$) (figure 13.4b). Additionally, pairwise PLS-DA models were generated between all the different location for the breath samples too, but no significant differences were identified ($p > 0.05$).

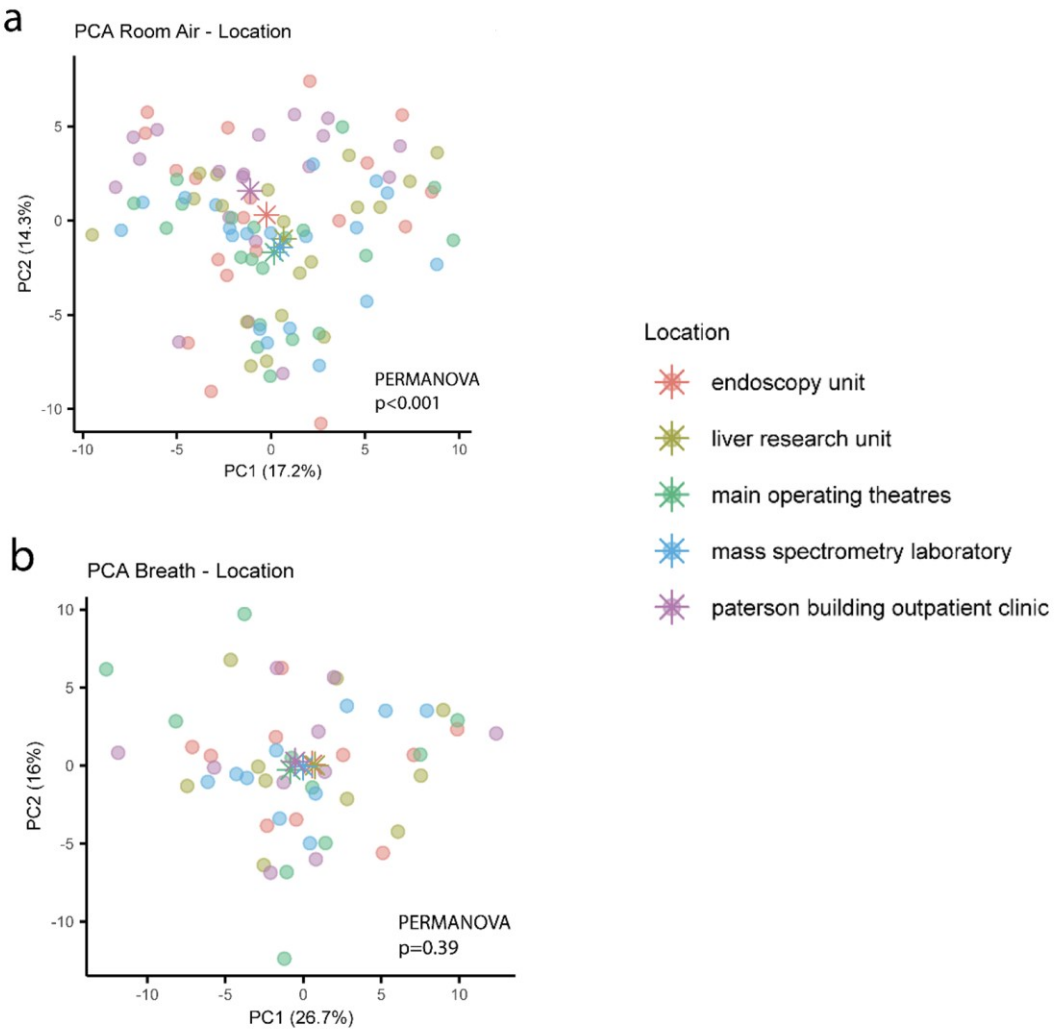


Figure 13.4 PCA plots of room air and breath by location with centroids demonstrating that the VOC profiles of room air differ across location, but overall breath VOC profile does not.

Unsupervised analysis with PCA revealed separation between room air samples collected in different locations but did not show separation for their corresponding breath samples. The asterisks represent group centroids.

13.4 Discussion

13.4.1 Separation of Room Air and Breath Samples

As expected, there was strong separation of room air and breath samples. PLS-DA revealed the separation was driven by 62 of the 113 detected VOCs. Within room air, these VOCs included di-isopropyl phthalate, benzophenone, acetophenone and benzyl alcohol, which are all commonly used within plasticisers and fragrances (265-268), the latter of which can be found extensively in cleaning products (269). The VOCs identified in breath were a

mixture of both endogenous and exogenous VOCs. The endogenous VOCs largely consisted of branched alkanes which are an established by-product of lipid peroxidation (270) and isoprene, which is largely stored within human muscle tissue (105). Exogenous VOCs included monoterpenes such as β-pinene and d-limonene, which can be traced back to essential oils from citrus fruit (also commonly used in cleaning products) and food preservatives (271,272). 1-Propanol can be both endogenous, deriving from amino acid breakdown, and exogenous, as it is present in disinfectants (273). Of the VOCs which were found in higher concentrations in room air compared to breath, several have been suggested as possible disease biomarkers. 1-Propanol, for example, has been found to be elevated in individuals with NASH by Verdam et al. (198) and m-cymene has been found to be higher in patients with active ulcerative colitis (274). Therefore, even if background room air VOCs don't appear to affect the overall breath profile of an individual, they might influence the concentrations of specific VOCs of interest. Background room air monitoring is therefore important, as has been established by previous studies (73,130,136,275).

13.4.2 Separation of Room Air Samples by Location

Separation of room air samples across all five different locations was observed. With the exception of 3-carene which was present in all investigated areas, separation was driven by different VOCs, giving each location a specific volatile signature.

In the endoscopy assessment area, VOCs driving separation were predominantly monoterpenes (e.g., β-pinene), and alkanes (e.g., dodecane, undecane and tridecane) that are commonly found in the essential oils that are commonly used in cleaning products (272). Given the frequency with which the endoscopy unit is cleaned, it is likely these VOCs are a result of frequent cleaning processes within this space. In the clinical research bay, as with endoscopy, separation was predominantly due to monoterpenes (e.g., α-pinene), also most likely originating from cleaning products. In the operating theatres complex, the VOC signature predominantly consisted of branched alkanes. These compounds may originate from surgical instruments which are abundant in oils and lubricants (276). In the surgical outpatient clinic, characteristic VOCs included a selection of alcohols including 1-nonanol, which is commonly found in plant oil and consequently cleaning products, and benzyl alcohol, which can be found in fragrances and local anaesthetics (269,277-279).

VOCs within the mass spectrometry laboratory were broadly different to the other areas which was to be expected given that this was the only non-clinical area that was assessed. While some monoterpenes were present, a more homogenous group of compounds

separated this area from the others (2,2,2-trifluoro-N-methyl-acetamide, pyridine, branched undecane, 2-pentyl-furan, ethylbenzene, furfural, ethyl anisate, o-xylene, m-xylene, isopropyl alcohol, and 3-carene), including aromatic hydrocarbons and alcohols. Some of these VOCs may be secondary to chemicals used within the laboratory, which at the time of assessment, contained seven mass spectrometry systems operating with both TD and liquid injection modes.

13.4.3 Diurnal Variation in Room Air VOC Profiles

Separation between the room air samples collected in the morning and those collected in the afternoon was also observed. The full list of VOCs driving the separation can be found in table 17.3 in appendix 17.5. Morning samples were broadly characterised by branched alkanes, which are commonly found exogenously in cleaning products and waxes (280). The four clinical areas included within this study were all cleaned prior to the sampling of the room air which would account for this. Afternoon samples typically presented mixture of alcohols, hydrocarbons, esters, ketones, and aldehydes in higher concentrations compared to the morning samples. 1-Propanol and phenol can both be found in disinfectants (273,281), which is expected given the regular cleaning that goes on throughout clinical areas during the day. Breath was only collected in the morning due to the multiple confounding factors that can influence VOC level within breath over the course of the day which could not be controlled for. This includes drink and food consumption prior to breath sampling (71,282) and changes in physiological parameters (e.g., exercise status) (283,284).

13.5 Study Limitations

This study was significantly limited by the number of breath samples taken in each location and the use of a single participant and further work is therefore required with a larger number of breath samples and participants to draw assured conclusion on the impact on the composition of human breath on the background environment in which it is samples. While the initial intention was for a minimum of five participants to provide room air samples over a minimum of fourteen days, only one subject provided breath samples over 10 days as the study was curtailed by the COVID-19 pandemic. The decision was made to end the study early and analyse the results which were obtained remotely while on-site research activities were restricted. The paucity of breath samples and having breath samples from one individual only means it is not possible to draw definitive conclusions from this study. The advantage of having a single participant provide breath samples, however, is that it has the potential to reduce variance from other confounding factors influenced by human behaviour

and single subject study design has been successfully used previously in several studies (285).

This study once again utilised the Mega WAX-HT polar column within the GC-MS for sample analysis. As already discussed in section 11.5, this column will limit the VOCs identified in this study and many non-polar compounds may not have been detected. Due to the COVID-19 pandemic and restriction of on-site activities, it was not possible to change the column to a more universal type for this study. While storing the tubes until the column could be changed was considered, the consensus of my research group was to analyse the samples with the instrument available at the time while acknowledging the limitations that this will impose upon the results.

Additionally, relative humidity (RH) data has not been collected and while I acknowledge that differentiations in RH might influence VOC distribution, in large scale studies, the logistical challenge is substantial for both control of RH and for collection of RH data.

13.6 Conclusion

In this study, I analysed VOC profiles within ambient room air across five commonly used locations for breath sample collection with a view to further understanding the impact of background VOCs concentrations within my group's sampling locations on breath analysis.

Within the aforementioned limitations, my study has affirmed the results of previous studies, identifying variation in background VOCs across different locations as well as diurnal variation in VOC concentrations.

While my study raises the possibility that the overall breath VOC profile of an individual is not significantly altered by sampling in different locations, there may still be variations in individual VOCs, especially those of lower abundance which will have less impact upon the overall "breath-print" of an individual. These VOCs may be of significance in metabolomic studies and therefore it is still recommended that background room air is sampled in parallel to breath sampling to detect any specific contaminant VOCs.

Although there are previous studies which have investigated the impact of background room air volatiles on breath sampling (as discussed in section 2.4.2), the results of this study are useful in providing local site-specific information regarding potential contaminant VOCs which has been considered in my clinical study.

14 VOLATILE ORGANIC COMPOUNDS FOR THE ASSESSMENT OF LIVER DISEASE VIA SIFT-MS

14.1 Participant Demographics

62 individuals were invited to participate. Eight individuals declined and one was excluded following initial screening due to recent consumption of alcohol. 53 participants provided breath samples between February 2019 and April 2019. One participant was excluded following recruitment as the SIFT-MS results suggested alcohol consumption prior to sampling which had not been volunteered during the screening process, leaving 52 participants within the analysis. The full demographic data of the participants recruited can be seen in table 14.1. As per Mann-Whitney U comparisons, there were significant differences between the sex distribution of the control and cirrhosis cohorts, the age distribution of the control and fibrosis cohorts and the BMIs of the control cohorts and both the fibrosis and cirrhosis cohorts. There were no significant differences in the average duration of fasting, smoking status, or alcohol status of the three cohorts.

14.2 Cirrhosis vs Fibrosis vs Control

A Kruskal-Wallis H test revealed that there was a statistically significant difference in four VOCs when control, fibrosis and cirrhosis participants were compared against each other. A list of p values for each VOC can be found in appendix 17.6, table 17.6. The four VOCs were limonene, 1-propanol, acetaldehyde, and ethanol.

14.2.1 Limonene

Concentrations of limonene were significantly higher in participants with cirrhosis (mean concentration 11 ppb) compared to those with fibrosis (mean concentration 4.6 ppb) and controls (mean concentration 6 ppb) ($\chi 2(2) = 6.24$, $p = 0.044$) (figure 14.1). The concentration of limonene in the background room air was low (mean concentration 1.5 ppb).

Table 14.1 Demographics of participants recruited for SIFT-MS study.

Characteristic	Classification	Healthy (n = 30)	Fibrosis (n = 7)	Cirrhosis (n = 15)	Mann-Whitney U p-value		
					Control vs Fibrosis	Control vs Cirrhosis	Fibrosis vs Cirrhosis
Sex	Male	23	6	10	0.50	0.01**	0.36
	Female	7	1	5			
Age (Years)	Minimum	24	48	41	0.01**	0.65	0.65
	Maximum	75	67	82			
	Mean	49	59	60			
Ethnicity	Asian	7	1	2	0.44	0.25	0.95
	African-Caribbean	2	0	0			
	Caucasian	21	6	13			
Smoking Status	Current Smoker	7	0	3	0.24	0.43	0.64
	Ex-Smoker	6	2	2			
	Never Smoked	17	5	8			
Alcohol Status	Current	13	7	8	0.19	0.84	0.39
	Previous	1	0	0			
	Never	6	0	7			
Time Fasted Prior to Sampling (Hours)	Minimum	1	2	1	0.42	0.12	0.80
	Maximum	12	12	12			
	Mean	6	6	4			
BMI (kg/m2)	Minimum	16.7	23.0	18.2	0.03**	0.01**	0.46
	Maximum	47.0	32.7	36.9			
	Mean	25.2	28.8	29.1			
Diet	Unrestricted	30	7	14			
	Vegetarian	0	0	1			
	Vegan	0	0	0			
Co-morbidities	None	26	2	7			
	Asthma	0	0	1			
	CKD	1	0	0			
	Crohn's Disease	1	0	0			
	Gout	0	1	0			
	HIV	1	2	0			
	Hypertension	0	0	2			
	Hypothyroidism	0	0	1			
	Renal Transplant	1	0	0			
	Type 2 Diabetes	1	2	4			
Medications	Anticoagulants	0	0	1			
	Laxatives	1	0	5			
	PPI	1	1	4			
Long Term Antibiotics	None	30	7	13			
	Rifaximin	0	0	2			

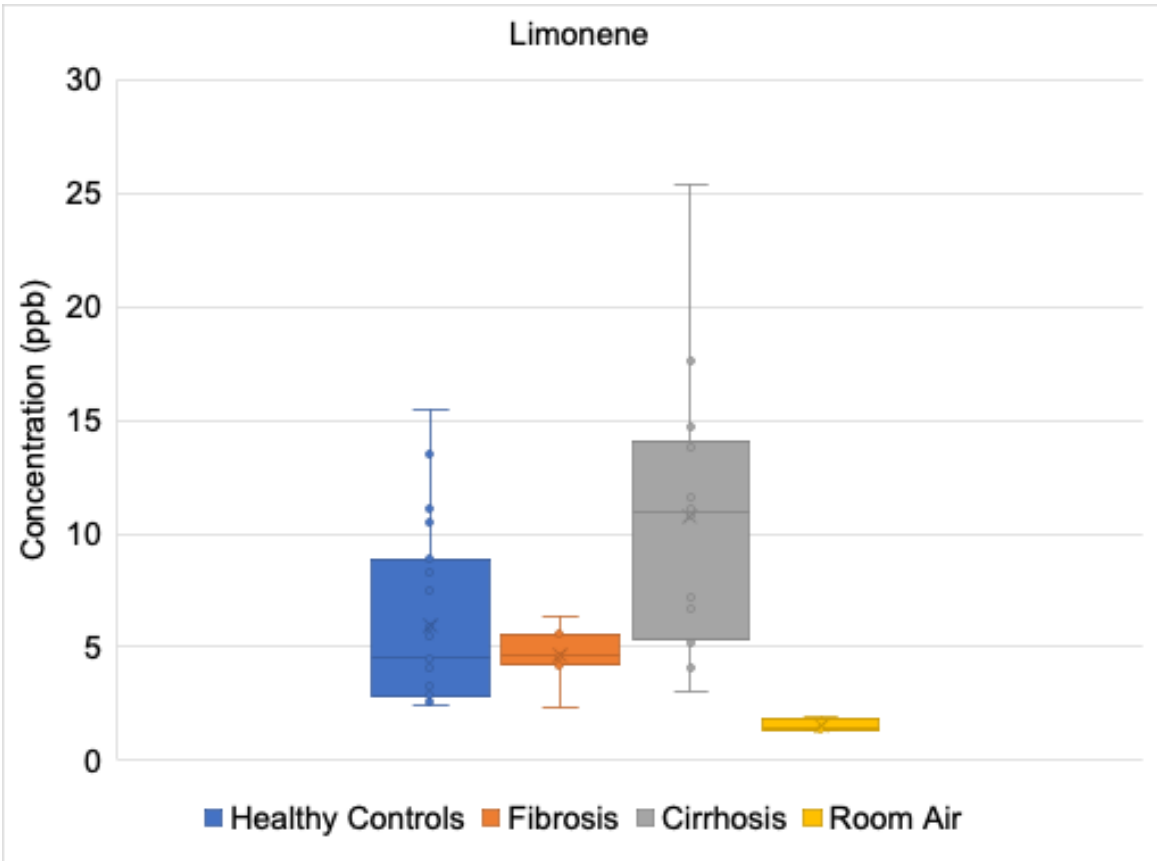


Figure 14.1 Box plot of concentrations of limonene comparing healthy controls, participants with fibrosis and participants with cirrhosis demonstrating higher concentrations of limonene in the cirrhosis cohort. Room air concentrations of limonene were low.

14.2.2 1-Propanol

Concentrations of 1-propanol were lower in those with cirrhosis (mean concentration 37 ppb) and fibrosis (mean concentration 74 ppb) compared to healthy controls (mean concentration 137 ppb) ($\chi 2(2) = 10.12$, $p = 0.006$) (figure 14.2). However, it should also be noted that there were relatively high concentrations of 1-propanol within the ambient room air (mean concentration 111 ppb).

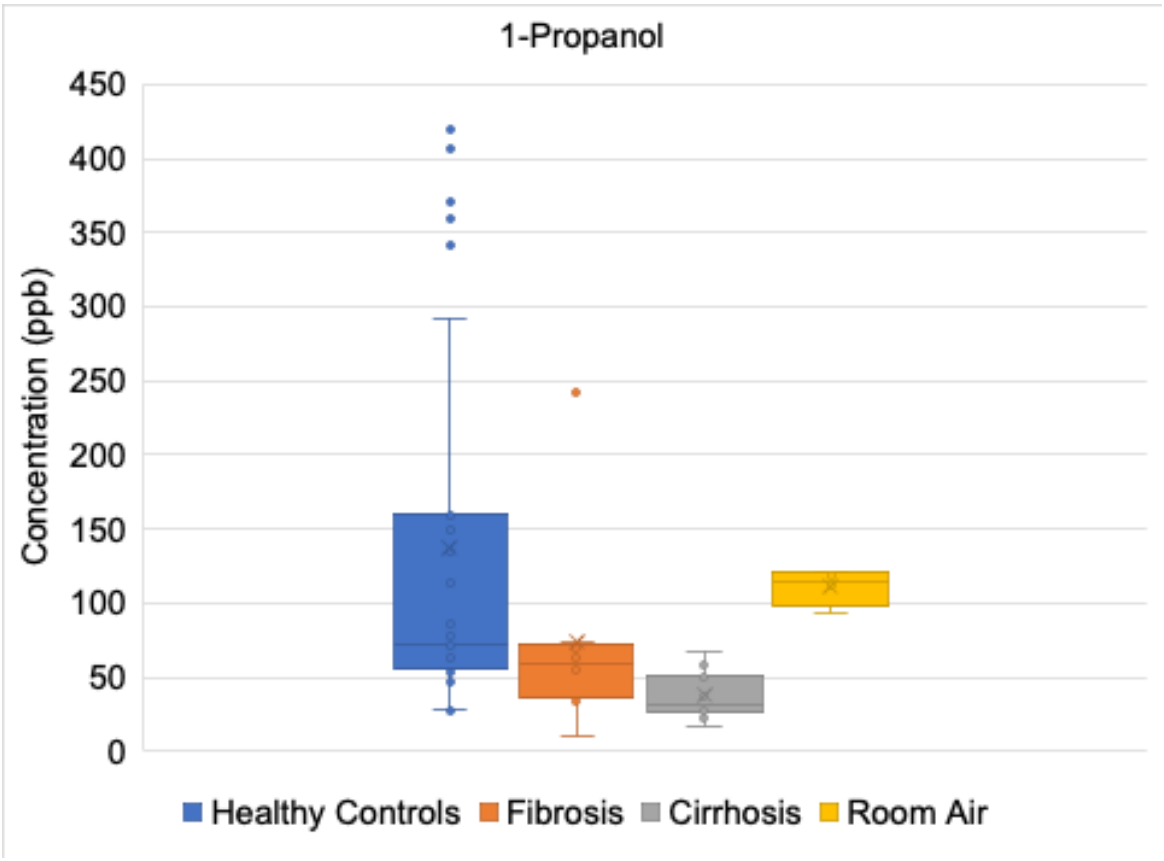


Figure 14.2 Box plot of concentrations of 1-propanol of healthy controls, participants with fibrosis and participants with cirrhosis demonstrating the highest concentrations of 1-propanol in healthy controls but also high concentrations of 1-propanol within the background room air.

14.2.3 Acetaldehyde

Concentrations of acetaldehyde were lower in those with cirrhosis (mean concentration 14 ppb) and fibrosis (mean concentration 14 ppb) compared to healthy controls (mean concentration 27 ppb) ($\chi 2(2) = 7.76$, $p = 0.021$) (figure 14.3). Background room air concentrations of acetaldehyde were low but comparable to that of the cirrhosis and fibrosis cohorts (mean concentration 15 ppb).

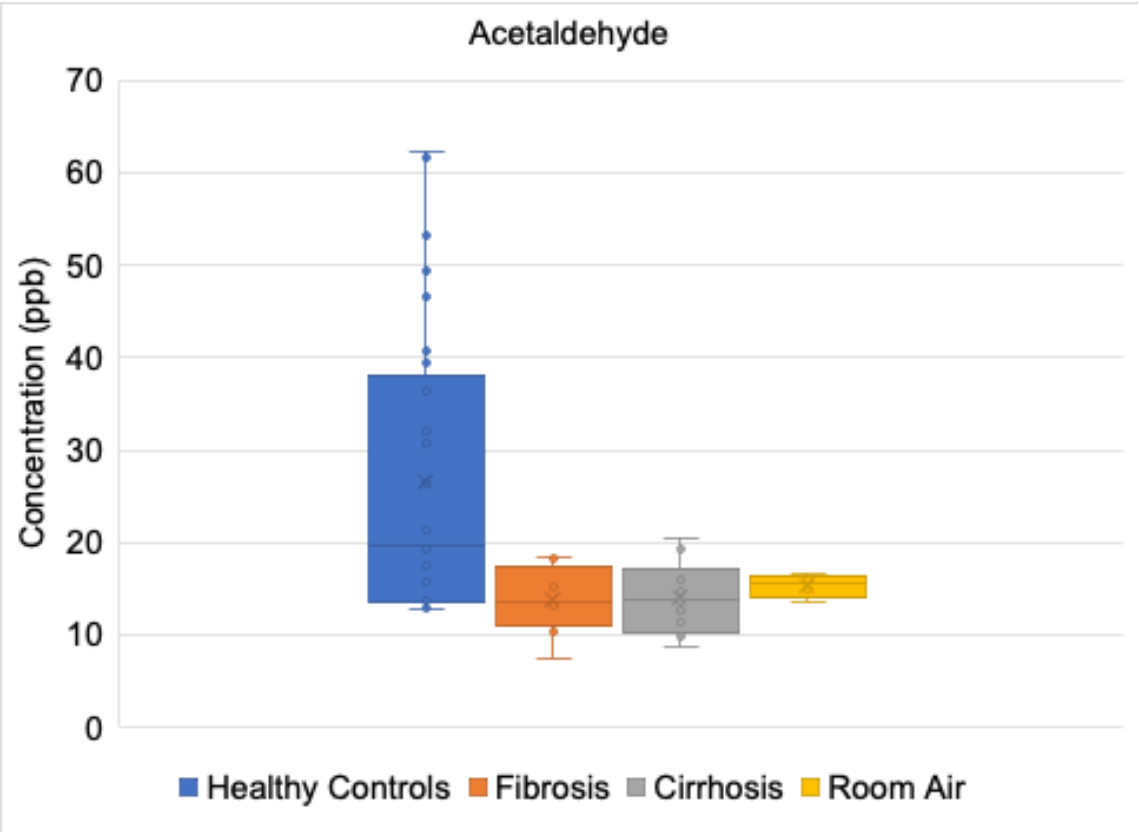


Figure 14.3 Box plot of concentrations of acetaldehyde comparing healthy controls, participants with fibrosis and participants with cirrhosis demonstrating high concentrations of acetaldehyde in healthy controls compared to the other cohorts.

14.2.4 Ethanol

Concentrations of ethanol were higher in healthy controls (mean concentration 540 ppb) compared to those with fibrosis (mean concentration 211 ppb) and cirrhosis (mean concentration 246 ppb) ($\chi 2(2) = 9.09$, $p = 0.011$) (figure 14.4) however there was also very high concentrations of ethanol in the background room air (mean concentration 1084 ppb).

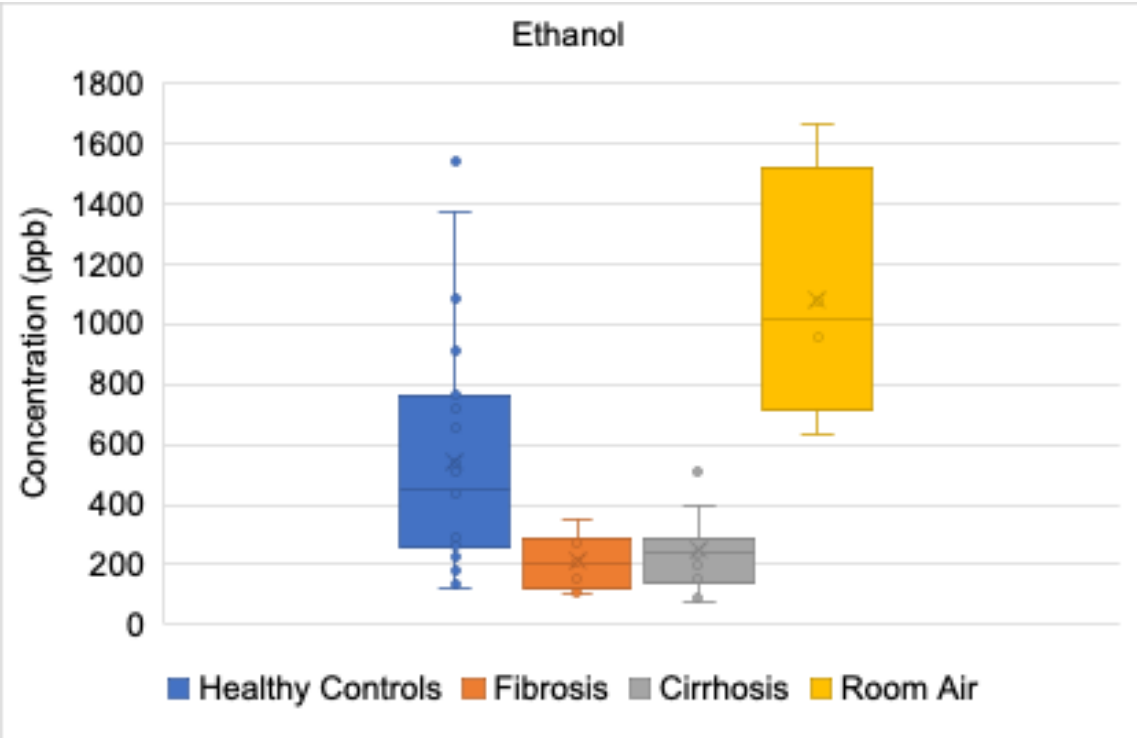


Figure 14.4 Box plot of concentrations of ethanol comparing healthy controls, participants with fibrosis and participants with cirrhosis demonstrating very high concentrations of ethanol within the background room air.

14.3 Fibrosis vs Controls

When comparing only control participants to those with fibrosis, only acetaldehyde and ethanol were significantly different. As with the comparison between all three groups, both acetaldehyde and ethanol were significantly lower in those with fibrosis compared to controls.

14.4 Cirrhosis vs Controls

When comparing cirrhosis participants to controls, alongside limonene, 1-propanol, acetaldehyde and ethanol, phenol was also significantly different.

14.4.1 Phenol

Concentrations of phenol were significantly higher in control participants (mean concentration 8.6 ppb) compared to participants with cirrhosis (mean concentration 3.5 ppb) ($U = 134$, $p = 0.028$) (figure 14.5). Concentrations of phenol in background room air were low (mean concentration 0.5 ppb).

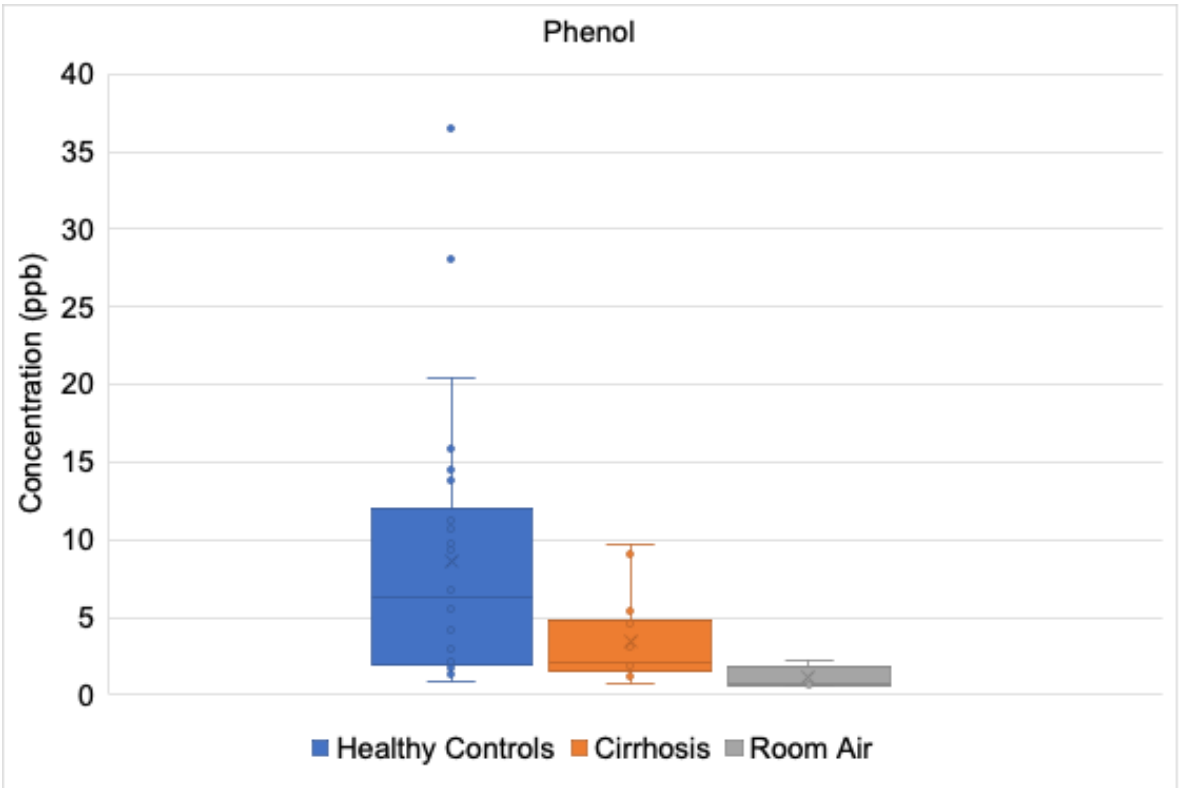


Figure 14.5 Box plot of concentrations of phenol comparing healthy control and participants with cirrhosis. Concentrations of phenol were highest in healthy controls and low in room air.

14.5 Cirrhosis vs Fibrosis

When comparing only participants with fibrosis to those with cirrhosis, limonene, 2-octanone and 2-pentanone were significantly different.

14.5.1 2-Octanone

Concentrations of 2-octanone were generally of low concentration but significantly higher in participants with cirrhosis (mean concentration 1.9 ppb) compared to those with fibrosis (mean concentration 1.2 ppb) ($U = 21$, $p = 0.026$) (figure 14.6). Room air concentrations of 2-octanone were low (mean concentration 1.2 ppb) but comparable to the concentrations of those with fibrosis.

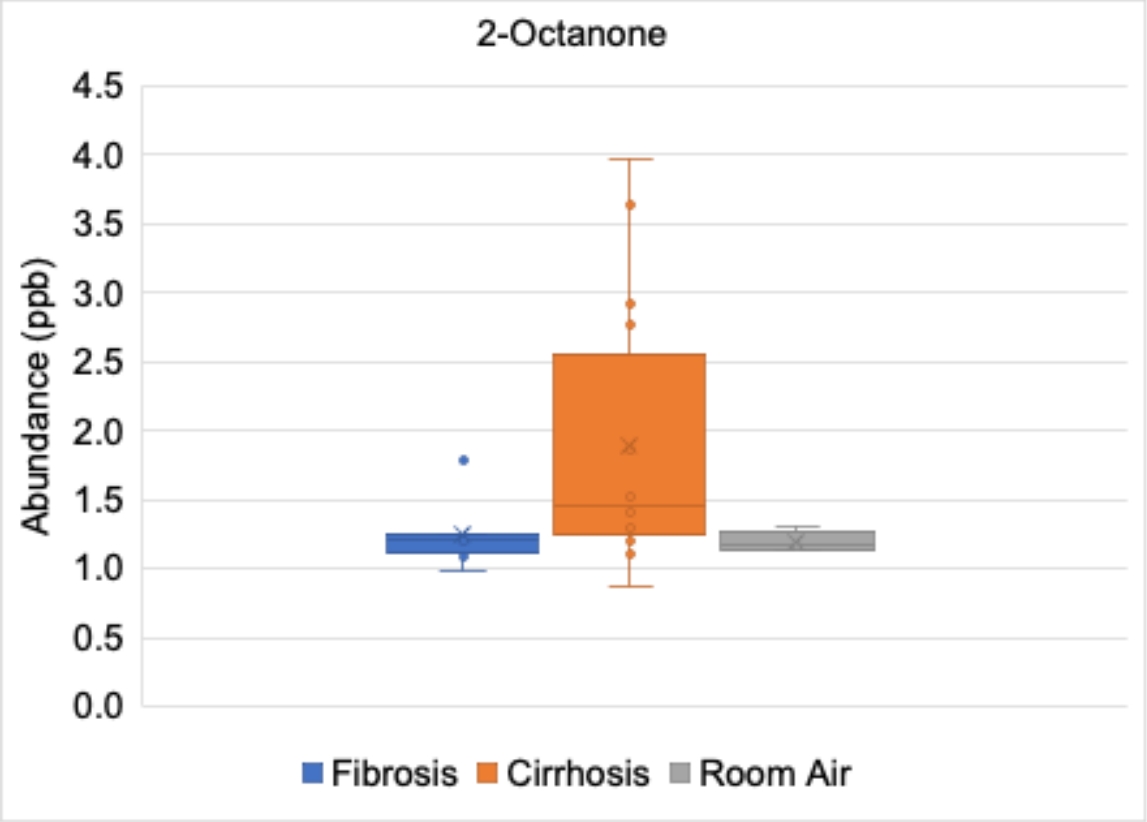


Figure 14.6 Box plot of concentrations of 2-octanone comparing participants with fibrosis and participants with cirrhosis. Concentrations of 2-octenone were highest in those with cirrhosis and low in room air.

14.5.2 2-Pentanone

Concentrations of 2-pentanone were significantly higher in participants with cirrhosis (mean concentration 2.1 ppb) compared to those with fibrosis (mean concentration 1.2 ppb) ($U = 24$, $p = 0.044$) (figure 14.7). The room air concentration of 2-pentanone was low (mean concentration 0.8 ppb).

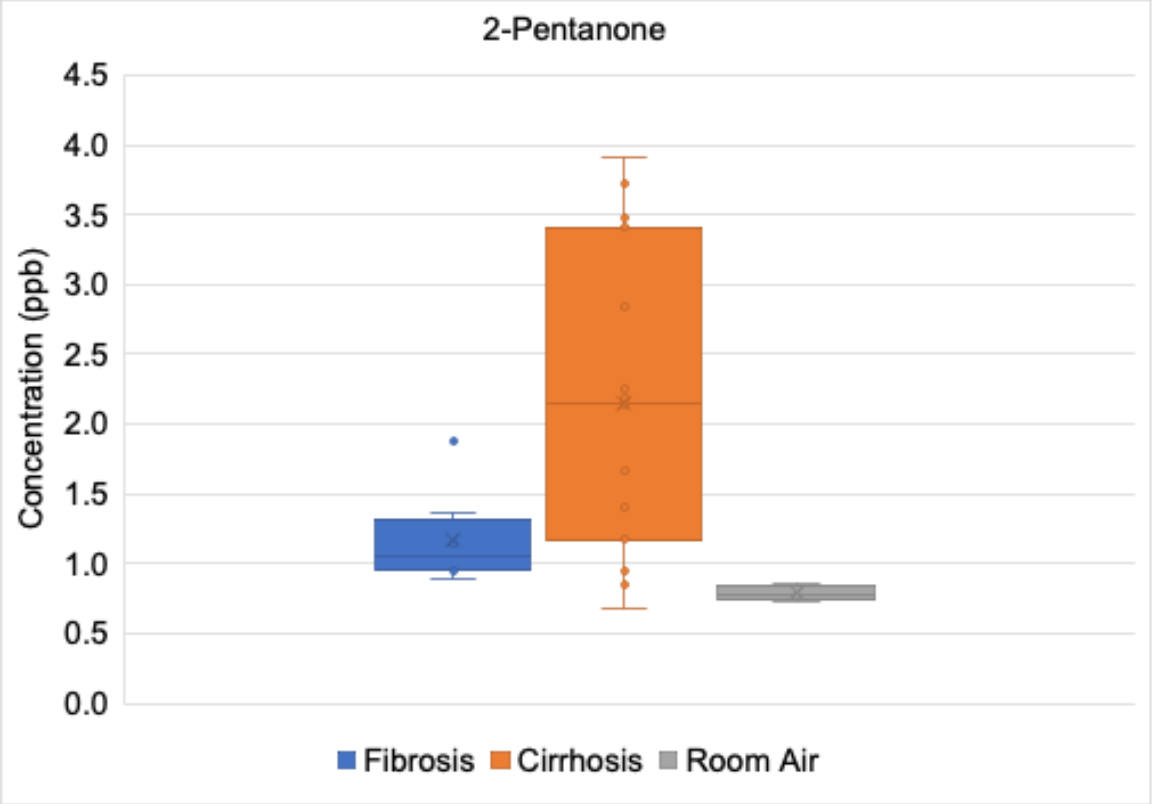


Figure 14.7 Box plot of 2-pentanone comparing participants with fibrosis and participants with cirrhosis demonstrating higher concentrations of 2-pentanone in those with cirrhosis compared to those with fibrosis. The room air concentration of 2-pentanone was low.

14.6 Discussion

14.6.1 Cirrhosis vs Fibrosis vs Controls

Elevated concentrations of limonene in those with liver disease correlates with previous studies by Friedman et al. (194), Pijls et al. (197) Fernandez del Rio et al. (177) , Ferrandino et al. (204,209) and Sinha et al. (214). There is a logical underlying pathway for why limonene may be elevated in those with hepatopathology (i.e., reduced metabolism of limonene due to impaired function of hepatic cytochrome p450 enzymes in the diseased liver). While limonene was identified in room air samples, it was of low concentration. Although the room air samples were limited, this gives some reassurance that limonene concentrations in breath are unlikely to be significantly elevated due to background room air contamination. Additionally, the food diaries of all participants were not supportive of dietary sources of limonene in those with high concentrations detected in their breath.

Elevated concentrations of 1-propanol in healthy controls compared to those with cirrhosis and fibrosis is the inverse of what had been found in previous literature by Miller-Atkins et al. (196) and Verdam et al. (198) who identified higher concentrations of 1-propanol in those

with liver disease. It should be noted that there were elevated concentrations of 1-propanol in the room air of the sampling location for my study, whereas Miller-Atkins et al. used an inhalation filter to try and remove ambient volatiles (but otherwise made no comment on sampling location or concentrations of background volatiles). It may be that removal of background 1-propanol concentrations by Miller-Atkins et al.'s inhalation filters may account for the differing results. Verdam et al. also made no comment on background volatile concentrations. As corroborated by my own work on background VOCs in sampling locations, 1-propanol can be found in the room air of clinical areas (1,286) and it is possible that the varying concentrations in this study are due to use of alcohol-based hand sanitisers. It is acknowledged that not having paired 1-propanol room air measurements for each breath sample makes interpretation of results challenging.

Acetaldehyde was higher in healthy controls compared to those with cirrhosis and fibrosis. Concentrations in those with cirrhosis and fibrosis were comparable to the concentration in background room air. Elevated concentrations in healthy controls is the inverse of what was found by Hanouneh et al. (195) who found higher concentrations of acetaldehyde in those with liver disease. However, Hanouneh et al. excluded smokers from their study to avoid their results being influenced by tobacco smoke. Further review of the data found that acetaldehyde was higher in smokers and ex-smokers (appendix 17.6, figure 17.11) compared to non-smokers which is to be expected given acetaldehyde can be found in tobacco smoke. Although a Mann-Whitney U test did not suggest any significant difference between the smoking demographics across the three cohorts, 7/10 active smokers and 6/10 ex-smokers included in the study were within the healthy control cohort. This may account for higher concentrations of acetaldehydes in the healthy controls compared to cirrhotics in my study. When only non-smokers were compared, the differences in acetaldehyde concentrations were non-significant (p 0.22). It is therefore probable that the discrepancy in results is driven by the differing smoking statuses of the individuals.

Ethanol was elevated in healthy controls compared to those with cirrhosis and fibrosis which is the inverse of what was found by Millonig et al. (205) and Hanouneh at al. (195) who found typically higher concentrations of ethanol in those with cirrhosis. However, the concentrations of ethanol detected in human breath were significantly higher than the concentrations identified in previous studies and review of the room air concentrations identified a high level of ethanol in the background room air which is likely skewing results. Like 1-propanol, the ethanol present in ambient room air is likely secondary to regular use of alcohol hand sanitiser within the clinical area. Therefore, ethanol cannot be considered a

valid biomarker on the basis of these results, and this also has further confirmed the importance of monitoring background room air volatiles.

14.6.2 Cirrhosis vs Controls

Higher concentrations of phenol in control participants contradicts the findings of Dadamio et al. (203) who identified higher concentrations of phenol in those with liver disease. Sukaram et al. (192) also used phenol as part of a model which could discriminate between HCC and non-HCC patients but did not comment if phenol was higher or lower in disease states. Phenol can be found in disinfectants (281) as well as tobacco smoke (287) although the low concentrations of phenol found in the background room air makes the likelihood of disinfectants in the clinical area influencing the results less likely. A more likely explanation for the discrepancy between studies is once again smoking status. The two individuals with the highest level of phenol were both active smokers and in the healthy control cohort (appendix 17.17.6, figure 17.12).

14.6.3 Cirrhosis vs Fibrosis

Elevated concentrations of 2-pentanone in those with cirrhosis validates the work of multiple previous studies (177,202-204,206) and elevated 2-octanone correlates with the work of by Morisco et al. (206). While ketones can be elevated due to fasting status, there was no significant difference between the average fasting duration of each cohort according to Mann-Whitney U comparison of the two groups. As discussed previously, ketones may be elevated due to increased beta oxidation of fatty acids in hepatic disease states. Room air concentrations for both ketones were low, although comparable to the concentrations in healthy controls. Those with cirrhosis, by definition will have increased impairment of liver function compared to those with fibrosis, which may account for why ketone concentrations are higher in those with cirrhosis compared to those with fibrosis only.

14.7 Study Limitations

This study is limited by the number of individuals recruited, especially with regards to the fibrosis cohort. The control cohort is also not well matched for sex, age, or BMI. However, it should be noted that the use of those attending for transient elastography ensures that all have a reference test for presence or absence of liver disease. While not perfect, transient elastography is currently the next best surrogate for assessment of liver disease and a normal result is reassuring that the individual recruited does not have liver disease.

While there was overall no significant difference between the average fasting states of each cohort, the lack of consistency in fasting durations may impact upon results, especially with VOCs such as ketones which were identified as of interest VOCs. Additionally, while smoking status was recorded, the duration of time since the last cigarette was not noted. It is therefore difficult to draw definitive conclusions on VOCs which can also be related to smoking status (e.g., phenol and acetaldehyde). There is also clear evidence of contamination from the ambient room air within the sampling location with significantly elevated concentrations of ethanol, likely secondary to regular use of alcohol hand gel sanitiser within the clinical area.

14.8 Conclusion

This targeted pilot study validated the results of several previous studies, with elevated concentrations of limonene, 2-pentanone and 2-octanone in those with cirrhosis. Those results which conflicted with pre-existing literature could potentially be explained by differences in smoking status and background room air contamination, highlighting the importance of trying to control and correcting for confounding factors where possible.

15 VOLATILE ORGANIC COMPOUNDS FOR THE ASSESSMENT OF LIVER DISEASE VIA GC-MS

15.1 Participant Recruitment and Demographics

212 individuals were approached for inclusion in the study. 27 declined and 33 individuals were excluded following screening at the initial consultation (12 patients were excluded as they unable to fully understand the study in order to provide informed consent due to a language barrier, 9 further patients were excluded due to recent consumption of alcohol and 12 were excluded due to a co-existing, non-HCC malignancy). Breath samples were collected from 156 individuals in total (figure 15.1).

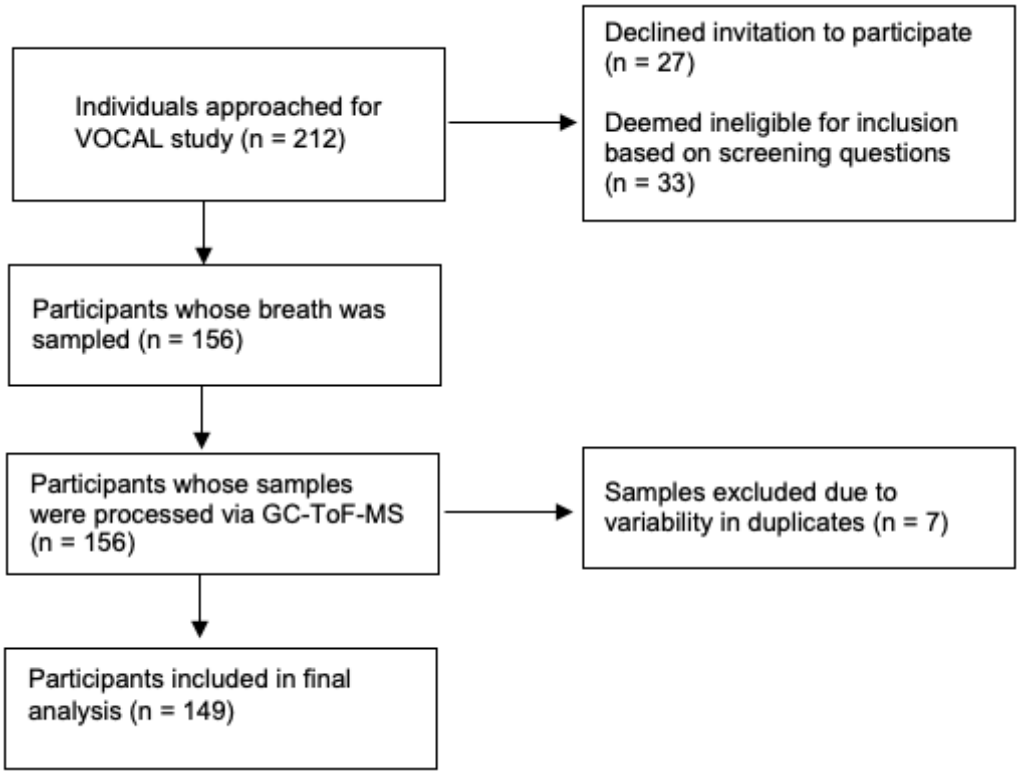


Figure 15.1 Flowchart of participation inclusion.

Following normalisation of the data and calculation of the CV% for duplicate samples, 7 samples (VOCAL2, VOCAL4, VOCAL7, VOCAL12, VOCAL14, VOCAL24 and VOCAL73) were excluded due to a CV% >30, leaving 149 participants for inclusion in the final analysis.

A summary of the demographics of the recruited participants can be seen in table 15.1.

While efforts were made to provide groups matched for demographics, there are some statistically significant differences between the groups.

Sex

There is significantly higher proportion of male participants in the HCC cohort compared to the control (p 0.002) and cirrhosis (p <0.001) cohorts. This fits with the epidemiology of hepatocellular carcinoma which has a male preponderance. A male to female ratio of 3.6:1 was recently reported in the USA (288).

Age

The average age of participants with HCC is significantly higher than that of the control (p <0.001) and cirrhosis participants (p <0.001).

Ethnicity

The ethnic diversity of the control group is significantly different to that of the cirrhosis (p 0.022) and HCC (p 0.01) groups, with a higher proportion of Asian and African-Caribbean participants. There is no significant difference in the ethnicity breakdown between the cirrhosis and HCC cohorts.

Alcohol Status

There is a significantly higher proportion of participants who were regularly consuming alcohol in the control cohort compared to the cirrhotic cohort (p 0.004) and HCC cohort (p 0.022). This is likely due to the control cohort not having liver disease and thus not having abstinence encouraged by their clinicians.

Smoking Status

There is no significant difference in the smoking status in the three cohorts. All individuals who were classed as smokers used tobacco cigarettes. There was no use of vapes reported.

Fasting Status

While the range and average duration for which the participants were fasted is similar for each group, overall, there is a significant difference between controls and cirrhotics (p 0.042), between controls and participants with HCC (p <0.001) and between cirrhotics and participants with HCC (p 0.004). Those in the control cohort were more consistently fasted due to fasting being part of the protocol for transient elastography.

Antibiotic Usage

A significantly higher proportion of cirrhotic participants were taking long-term antibiotics compared to the control cohort (p 0.038). This is to be expected as participants with cirrhosis are commonly placed on rifaximin to treat and prevent hepatic encephalopathy and/or ciprofloxacin as secondary prevention for spontaneous bacterial peritonitis. The same would not be applicable to the control cohort. There was no significant difference between the cirrhosis and HCC cohorts.

Reassuringly there is no statistically significant difference between the height, weight, or body mass index (BMI) of the participants across the three cohorts.

15.1.1 Control Participants

All control participants were recruited from transient elastography lists. As such, all control participants had a reference test with which to provide assurance that they had no evidence of underlying liver disease. The average transient elastography reading was 5.7 kPa (range 3–11.2 kPa). All participants except one (VOCAL153) had a transient elastography score of <8 kPa which is not suggestive cirrhosis for all aetiologies. Participant VOCAL153 had a transient elastography score of 11.2 kPa which cannot exclude cirrhosis or significant fibrosis. However, it was retained within the control cohort based on expert opinion (i.e., based on other assessments and parameters including serology and radiology, it was felt there was no evidence of cirrhosis by the clinician in charge of their care).

15.1.2 Cirrhosis Participants

Aetiology

I recruited participants with cirrhosis without discrimination of the underlying aetiology of their cirrhosis. A breakdown of the differing aetiologies can be seen in figure 15.2. The majority of participants (29/64, 45%) had cirrhosis secondary to alcohol (ETOH) excess.

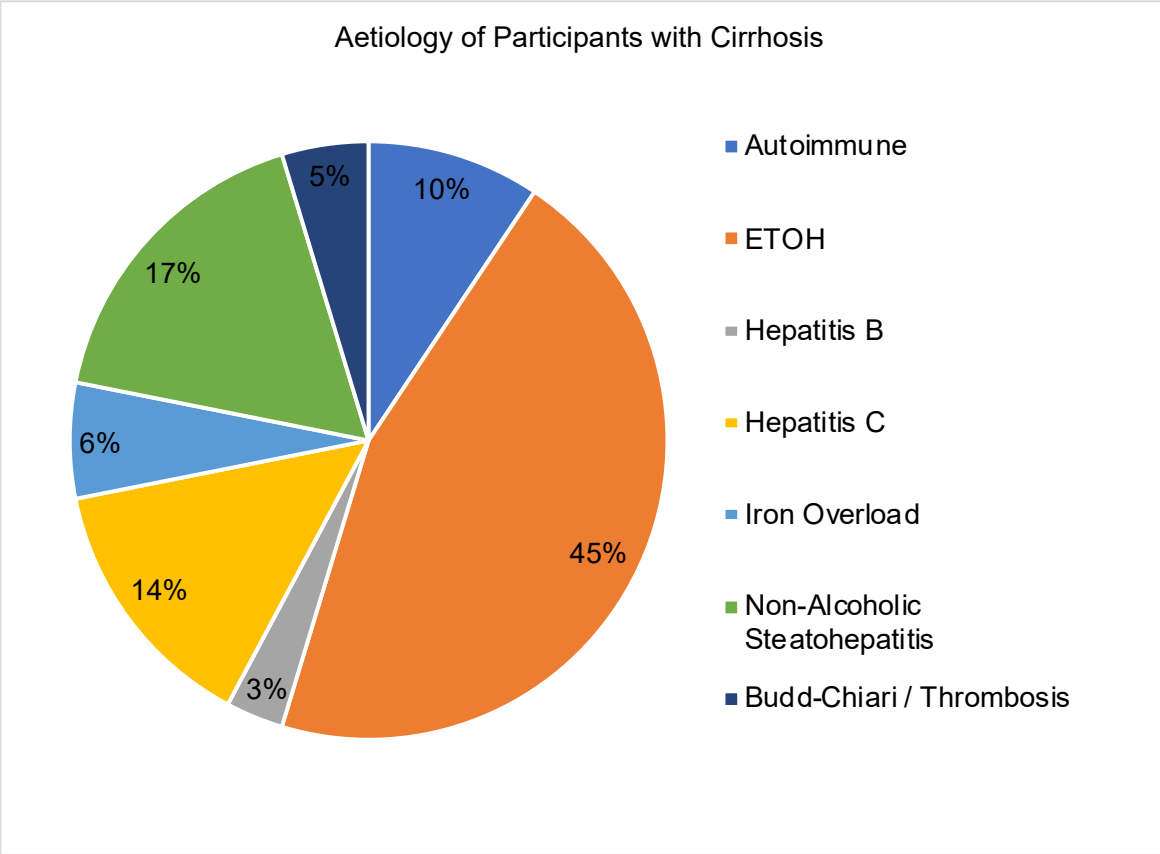


Figure 15.2 Pie chart demonstrating breakdown of aetiology of cirrhosis in participants with cirrhosis, with the majority of participants having alcohol-related cirrhosis.

Child-Pugh Classification

Participants with cirrhosis had varying degrees of severity. Figure 15.3 provides the breakdown of the Child-Pugh classifications of cirrhotic participants. The majority of participants (41/64, 64%) were classified as Child-Pugh A.

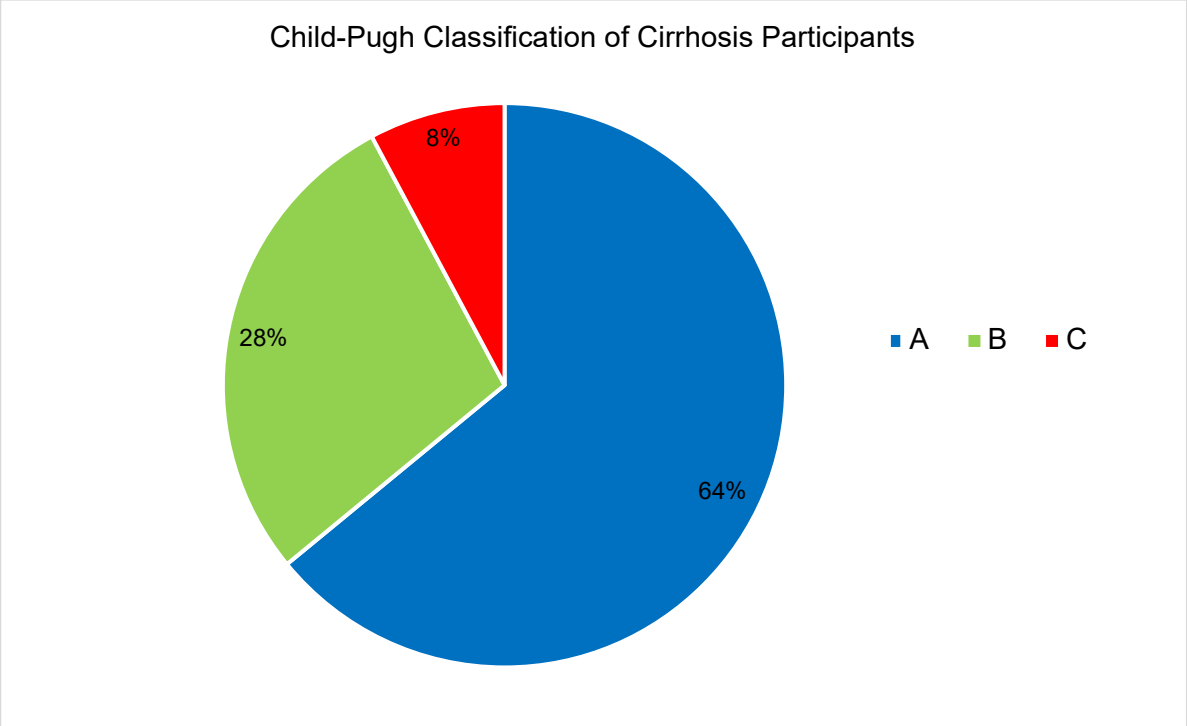


Figure 15.3 Pie chart of Child-Pugh classifications of participants with cirrhosis demonstrating the majority were Child-Pugh A.

Cirrhosis Phenotype

The majority of participants (52/64, 81%) were classified by expert opinion as being in a compensated state at the time of breath collection.

MELD Score

The average MELD score of participants with cirrhosis was 10.3 (range 6-30).

UKELD Score

The average UKELD score of participants with cirrhosis was 48.7 (range 41-61). 34/64 participants with cirrhosis (53%) reached the threshold of >49 that correlates with a 9% 1-year mortality rate, the minimum required for transplant listing.

Presence of Hepatic Encephalopathy

10/64 participants with cirrhosis (16%) had a clinical diagnosis of active grade I or II hepatic encephalopathy at the time of breath collection.

15.1.3 Hepatocellular Carcinoma Participants

Aetiology

Three HCC participants had an HCC that developed in a non-cirrhotic liver. For the remaining HCC participants, a breakdown of the underlying aetiology of their cirrhosis can be seen in figure 15.4. The majority of HCC participants were cirrhotic due to chronic viral hepatitides (14/35, 39%), or alcohol (11/35, 32%).

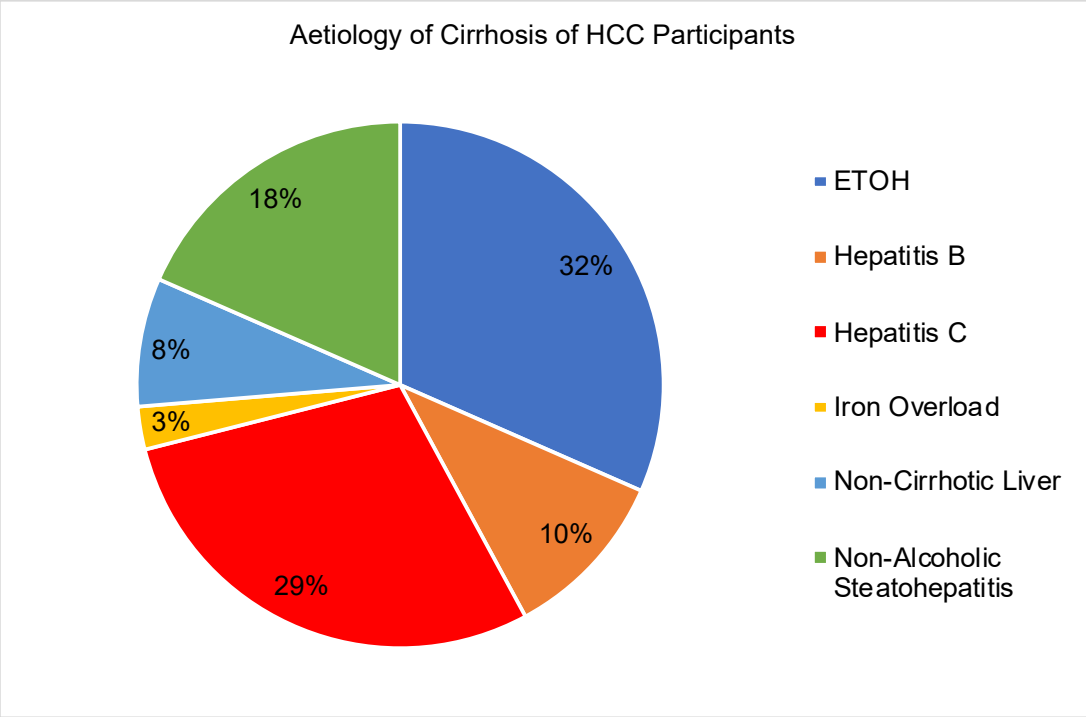


Figure 15.4 Pie chart demonstrating aetiologies of any underlying cirrhosis in participants with HCC with the majority having ETOH or hepatitis C related cirrhosis.

Child-Pugh Score

All participants with HCC that were cirrhotic were Child-Pugh A (29/35, 83%) or Child-Pugh B (6/35, 17%). There were no HCC participants with Child-Pugh C cirrhosis. This is likely because Child-Pugh C cirrhosis makes individuals ineligible for systemic anti-cancer therapy and are thus more often managed by palliative care following initial diagnosis, rather than regular attendance at a specialist HCC clinic.

HCC Stage

7/38 (18%) participants had metastatic disease at the time of breath sampling. According to the BCLC staging system, 10/38 (26%) of participants were BCLC-A, 16/38 (42%) of participants were BCLC-B and 12/38 (32%) of participants were BCLC-C (figure 15.5). There were no participants with BCLC-D stage disease. As with underlying Child-Pugh C cirrhosis, those with BCLC-D stage disease are appropriate for best supportive care only and would be less likely to attend outpatient appointments in the HCC clinic where I recruited.

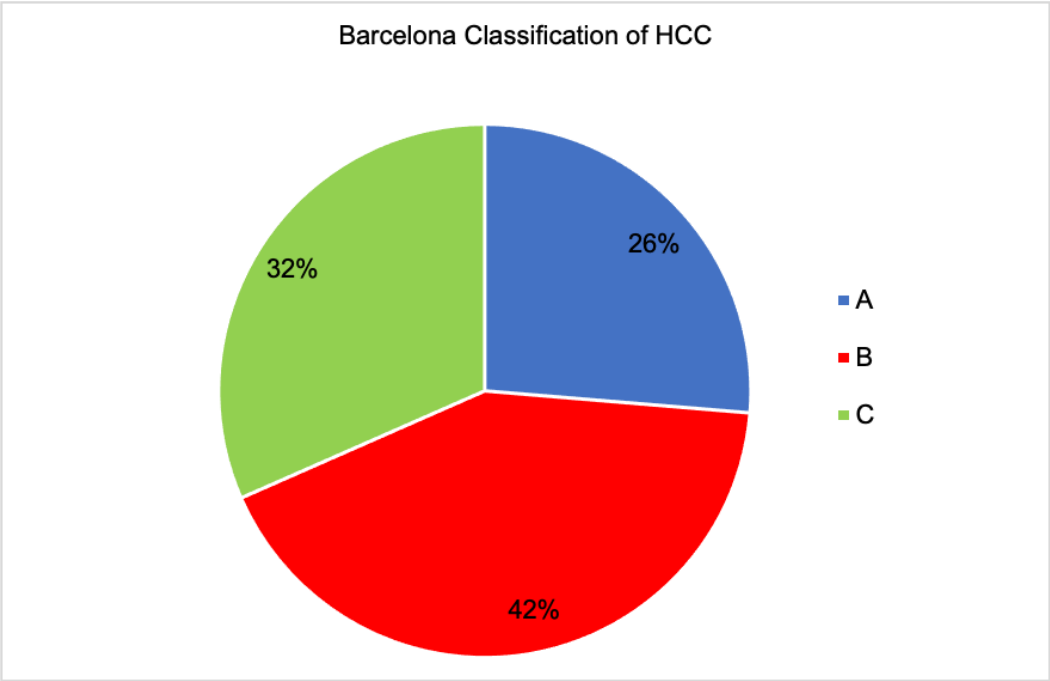


Figure 15.5 Pie chart demonstrating breakdown of BCLC classification of participants with HCC.

HCC Size

17 participants had multifocal HCC. Those with solitary lesions ranged from 0.8-17 cm in size (average 5.8 cm).

HCC Treatment Status

14/38 (37%) participants were treatment naïve at the time of recruitment. 7/38 (18%) of participants were on first line systemic therapy with atezolizumab and bevacizumab, 1/38 (3%) was awaiting transplantation, having had insufficient response to RFA and SBRT. 2/38 (5%) were awaiting further locoregional therapy with TACE or RFA and 12/40 (30%) were on first- or second-line tyrosine kinase inhibitors.

AFP Level

The concentrations of AFP of participants ranged from 2-18585 IU/mL with a mean AFP of 816 IU/L. 19/38 (50%) of participants had an AFP within the normal range (i.e., <10 IU/mL).

Table 15.1 Demographic Information of Recruited Participants

Characteristic	Classification	All Participants (n = 149)		Control (n = 47)		Cirrhosis (n = 64)		HCC (n = 38)		Mann Whitney U p-value of Comparison of Groups		
		Number	% Of Total	Number	% Of Total	Number	% Of Total	Number	% Of Total	Control vs Cirrhosis	Control vs HCC	Cirrhosis vs HCC
Sex	Male	102	68.5	30	63.8	37	57.8	35	92.1	0.524	0.002**	<0.001**
	Female	47	31.5	17	36.2	27	42.2	3	7.9			
Age (Years)	Minimum	27	N/A	27	N/A	39	N/A	36	N/A	0.071	<0.001**	<0.001**
	Maximum	88	N/A	83	N/A	83	N/A	88	N/A			
	Mean	63	N/A	57	N/A	62	N/A	71	N/A			
Ethnicity	Middle Eastern	10	7.4	4	8.5	5	7.8	1	2.6	0.022**	0.01**	0.517
	Asian	29	20.8	12	25.5	11	17.2	6	15.8			
	African-Caribbean	17	11.4	10	21.3	4	6.3	3	7.9			
	Caucasian	93	62.2	21	44.7	44	68.8	28	73.7			
Recruitment Site	St Mary's Hospital	119	79.9	47	100.0	64	100.0	8	21.1			
	Hammersmith Hospital	30	20.1	0	0.0	0	0.0	30	78.9			
Smoking Status	Current Smoker	34	22.8	9	19.1	14	21.9	11	28.9	0.286	0.055	0.272
	Ex-Smoker	54	36.2	14	29.8	22	34.4	16	42.1			
	Never Smoked	61	40.9	24	51.1	28	43.8	11	28.9			
Smokers Time Since Last Cigarette (Hours)	Minimum	0.5	N/A	2	N/A	0.5	N/A	0.5	N/A			
	Maximum	16	N/A	16	N/A	6	N/A	6	N/A			
	Mean	3.3	N/A	6.2	N/A	2.2	N/A	2.4	N/A			
Alcohol Status	Current Consumption	40	26.8	22	46.8	12	18.8	6	15.8	0.004**	0.022**	0.955
	Previous Consumption	36	24.2	2	4.3	26	40.6	8	21.1			
	Never Consumed	73	49.0	23	48.9	26	40.6	24	63.2			
Weekly Units of Current Alcohol Consumers	Minimum Weekly Units	1	N/A	1	N/A	1	N/A	2	N/A			
	Maximum Weekly Units	64	N/A	45	N/A	64	N/A	50	N/A			

	Mean Weekly Units	11.4	N/A	7.9	N/A	14.8	N/A	16.9	N/A			
Time Fasted Prior to Sampling (Hours)	Minimum	1	N/A	1	N/A	1	N/A	1	N/A	0.042**	<0.001**	0.004**
	Maximum	16	N/A	16	N/A	16	N/A	14	N/A			
	Mean	5.4	N/A	7	N/A	5.4	N/A	3.6	N/A			
Height (cm)	Minimum	142	N/A	142	N/A	144	N/A	148	N/A	0.171	0.605	0.334
	Maximum	198	N/A	189	N/A	198	N/A	184	N/A			
	Mean	170	N/A	171	N/A	169	N/A	170	N/A			
Weight (kg)	Minimum	46	N/A	51.1	N/A	46	N/A	56	N/A	0.377	0.301	0.716
	Maximum	150	N/A	128	N/A	150	N/A	126	N/A			
	Mean	82.5	N/A	80.2	N/A	83.6	N/A	83.4	N/A			
BMI (kg/m2)	Minimum	16.4	N/A	16.7	N/A	16.4	N/A	20.6	N/A	0.197	0.356	0.742
	Maximum	58.8	N/A	40.4	N/A	58.8	N/A	45.2	N/A			
	Mean	28.7	N/A	27.7	N/A	29.4	N/A	28.8	N/A			
Diet	Unrestricted	144	96.6	46	97.9	61	95.3	37	97.4			
	Pescatarian	2	1.3	0	0.0	2	3.1	0	0.0			
	Vegetarian	3	2.0	1	2.1	1	1.6	1	2.6			
	Vegan	0	0.0	0	0.0	0	0.0	0	0.0			
Co-Morbidities	None	25	16.8	6	12.8	13	20.3	6	15.8			
	Allergic Rhinitis	1	0.7	1	2.1	0	0.0	0	0.0			
	Asthma	9	6.0	2	4.3	5	7.8	2	5.3			
	Atrial Fibrillation	5	3.4	1	2.1	0	0.0	4	10.5			
	Barrett's Oesophagus	2	1.3	0	0.0	2	3.1	0	0.0			
	Bechet’s Disease	1	0.7	1	2.1	0	0.0	0	0.0			
	Beta Thalassaemia	2	1.3	0	0.0	2	3.1	0	0.0			
	Chronic Kidney Disease	9	6.0	1	2.1	4	6.3	4	10.5			
	Chronic Pancreatitis	3	2.0	2	4.3	1	1.6	0	0.0			
	Coeliac Disease	1	0.7	0	0.0	1	1.6	0	0.0			

	Constipation	2	1.3	1	2.1	1	1.6	0	0.0	
	Chronic Obstructive Pulmonary Disease	12	8.1	1	2.1	6	9.4	5	13.2	
	Crohn's Disease	2	1.3	1	2.1	1	1.6	0	0.0	
	Diverticulosis	3	2.0	1	2.1	1	1.6	1	2.6	
	Epilepsy	1	0.7	0	0.0	1	1.6	0	0.0	
	Gout	2	1.3	0	0.0	1	1.6	1	2.6	
	Heart Failure	2	1.3	0	0.0	0	0.0	2	5.3	
	Human Immunodeficiency Virus	2	1.3	2	4.3	0	0.0	0	0.0	
	Hypertension	39	26.2	13	27.7	14	21.9	12	31.6	
	Hypothyroidism	9	6.0	2	4.3	5	7.8	2	5.3	
	Ischaemic Heart Disease	10	6.7	2	4.3	4	6.3	4	10.5	
	Psoriasis	5	3.4	3	6.4	1	1.6	1	2.6	
	Renal Transplant	1	0.7	0	0.0	1	1.6	0	0.0	
	Rheumatoid Arthritis	4	2.7	2	4.3	1	1.6	1	2.6	
	Sjogren's Syndrome	1	0.7	1	2.1	0	0.0	0	0.0	
	Systemic Lupus Erythematosus	1	0.7	1	2.1	0	0.0	0	0.0	
	Stroke	7	4.7	0	0.0	1	1.6	6	15.8	
	Type 2 Diabetes	39	26.2	9	19.1	19	29.7	11	28.9	
	Ulcerative Colitis	2	1.3	1	2.1	1	1.6	0	0.0	
Previous Cancer	None	132	88.6	43	91.5	60	93.8	29	76.3	
	Acute Myeloid Leukaemia	1	0.7	0	0.0	0	0.0	1	2.6	
	Breast	4	2.7	2	4.3	1	1.6	1	2.6	
	Lung	1	0.7	0	0.0	0	0.0	1	2.6	
	Lymphoma	1	0.7	0	0.0	0	0.0	1	2.6	
	Melanoma	1	0.7	0	0.0	1	1.6	0	0.0	
	Prostate	7	4.7	1	2.1	2	3.1	4	10.5	

	Renal	2	1.3	1	2.1	0	0.0	1	2.6			
Previous Gastrointestinal Surgery	None	135	90.6	42	89.4	63	98.4	37	97.4			
	Appendicectomy	2	1.3	1	2.1	1	1.6	0	0.0			
	Cholecystectomy	3	2.0	1	2.1	2	3.1	0	0.0			
	Hartmann's Resection	1	0.7	1	2.1	0	0.0	0	0.0			
	Ileocaecal Resection	1	0.7	0	0.0	1	1.6	0	0.0			
	Left Hemicolectomy	2	1.3	1	2.1	1	1.6	0	0.0			
	Pancreatectomy	1	0.7	0	0.0	1	1.6	0	0.0			
	Perforated Ulcer	1	0.7	0	0.0	0	0.0	1	2.6			
	Roux-en-Y Bypass	1	0.7	1	2.1	0	0.0	0	0.0			
	Splenectomy	1	0.7	0	0.0	1	1.6	0	0.0			
	Whipple's	1	0.7	0	0.0	1	1.6	0	0.0			
Medications	Anticoagulants	12	8.1	2	4.3	4	6.3	6	15.8			
	H2 Receptor Antagonist	4	2.7	0	0.0	1	1.6	3	7.9			
	Laxatives	22	14.8	1	2.1	21	32.8	0	0.0			
	Proton Pump Inhibitor	56	37.6	12	25.5	33	51.6	11	28.9			
Long Term Antibiotics	None	123	82.6	46	97.9	43	67.2	34	89.5	0.04**	0.290	0.362
	Ciprofloxacin	4	2.7	0	0.0	1	1.6	3	7.9			
	Ciprofloxacin and Rifaximin	4	2.7	0	0.0	2	3.1	2	5.3			
	Doxycycline	1	0.7	0	0.0	0	0.0	1	2.6			
	Phenoxymethylpenicillin	2	1.3	1	2.1	1	1.6	0	0.0			
	Rifaximin	11	7.4	0	0.0	11	17.2	0	0.0			
Time Stored at -80°C (Days)	Minimum	0	N/A	0	N/A	2	N/A	0	N/A	<0.001**	<0.001**	0.766
	Maximum	343	N/A	343	N/A	336	N/A	310	N/A			
	Mean	81.7	N/A	21.5	N/A	114.1	N/A	101.5	N/A			

15.1.4 Room Air Samples

As part of the study protocol, I had intended to collect room air samples for each sampling session. Ultimately this was not possible due to a shortage in TD tubes. The limited TD tubes that were available were therefore prioritised for breath sample collection. Four room air samples were collected from St. Mary's Hospital during the recruitment process. This is discussed further in the limitations section.

15.2 Control vs Cirrhosis vs HCC

An unsupervised PCA of control participants vs cirrhosis participants vs HCC participants demonstrated some clustering of the individual cohorts (figure 15.6). Control participants appeared to separate best, with significant overlap between the cirrhosis and HCC participants. The separation improved with supervised analysis via OPLS-DA (figure 15.7). The separation was significant according to CV ANOVA testing ($p < 0.001$) and permutation plots with 999 permutations (appendix 17.7.9, figure 17.19).

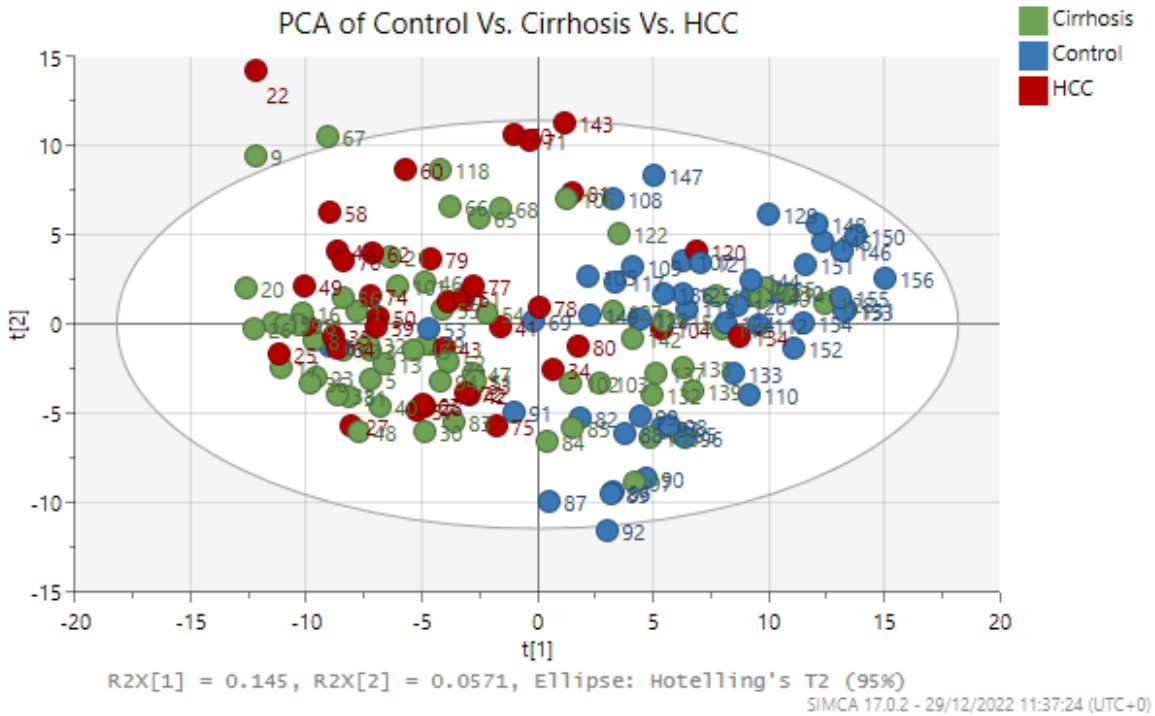


Figure 15.6 PCA of control vs cirrhosis vs HCC demonstrating some clustering of cohorts with the strongest separation being between the control cohort and the other two cohorts.

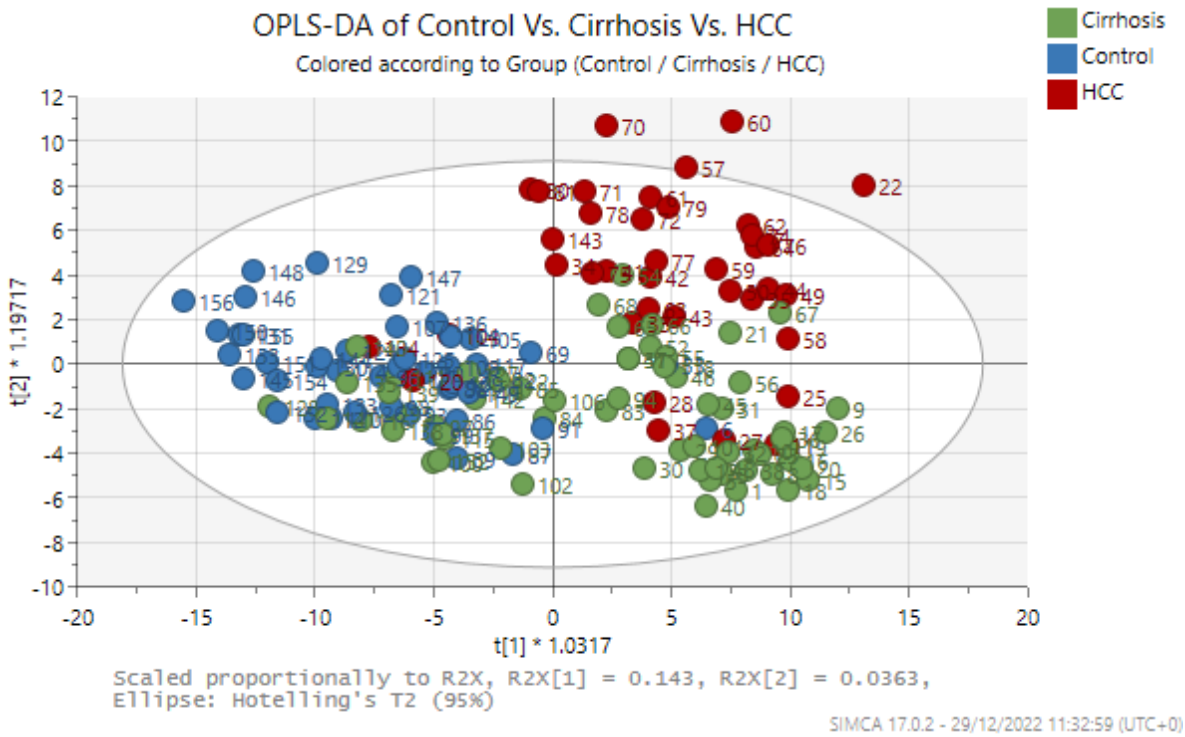


Figure 15.7 OPLS-DA of control vs cirrhosis vs HCC demonstrating improved separation between all three cohorts.

15.3 Cirrhosis vs HCC

An unsupervised PCA of participants with cirrhosis vs participants with HCC demonstrated no separation (figure 15.8). Supervised analysis via OPLS-DA however demonstrated excellent separation between cirrhosis and HCC (figure 15.9). The separation was significant according to CV ANOVA testing ($p < 0.001$) and permutation plots with 999 permutations (appendix 17.7.9, figure 17.20).

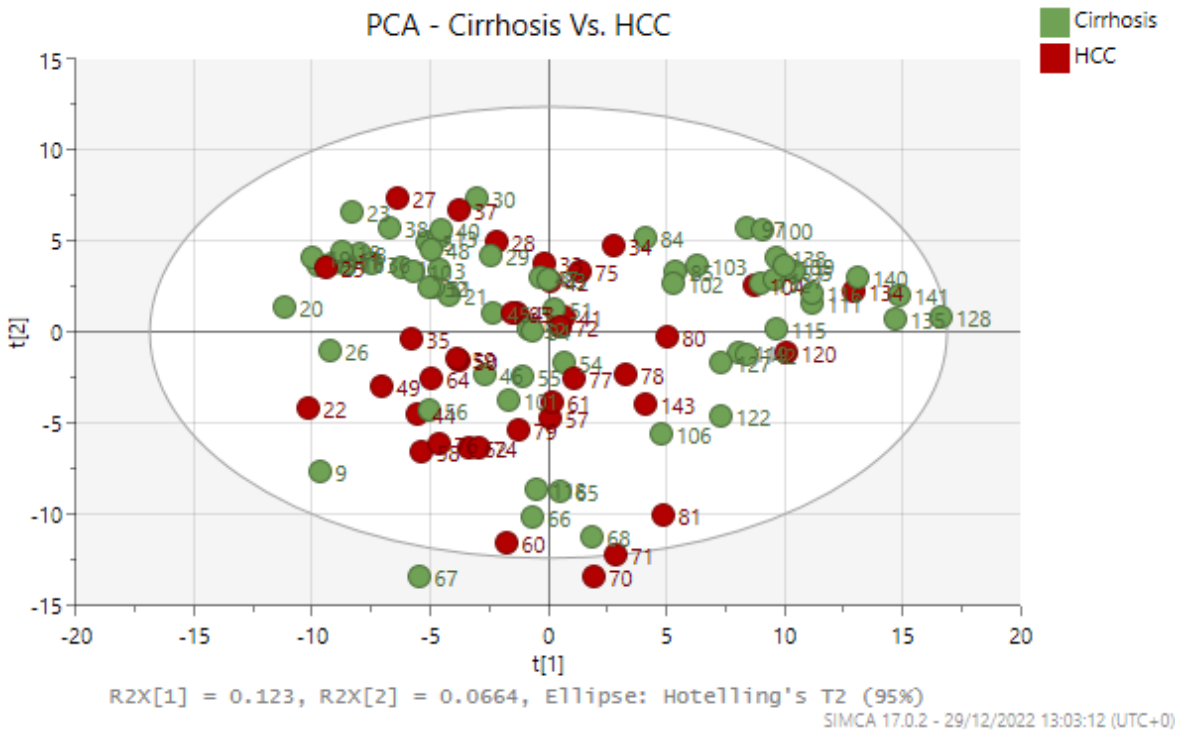


Figure 15.8 PCA of cirrhosis vs HCC demonstrating no separation between the two cohorts.

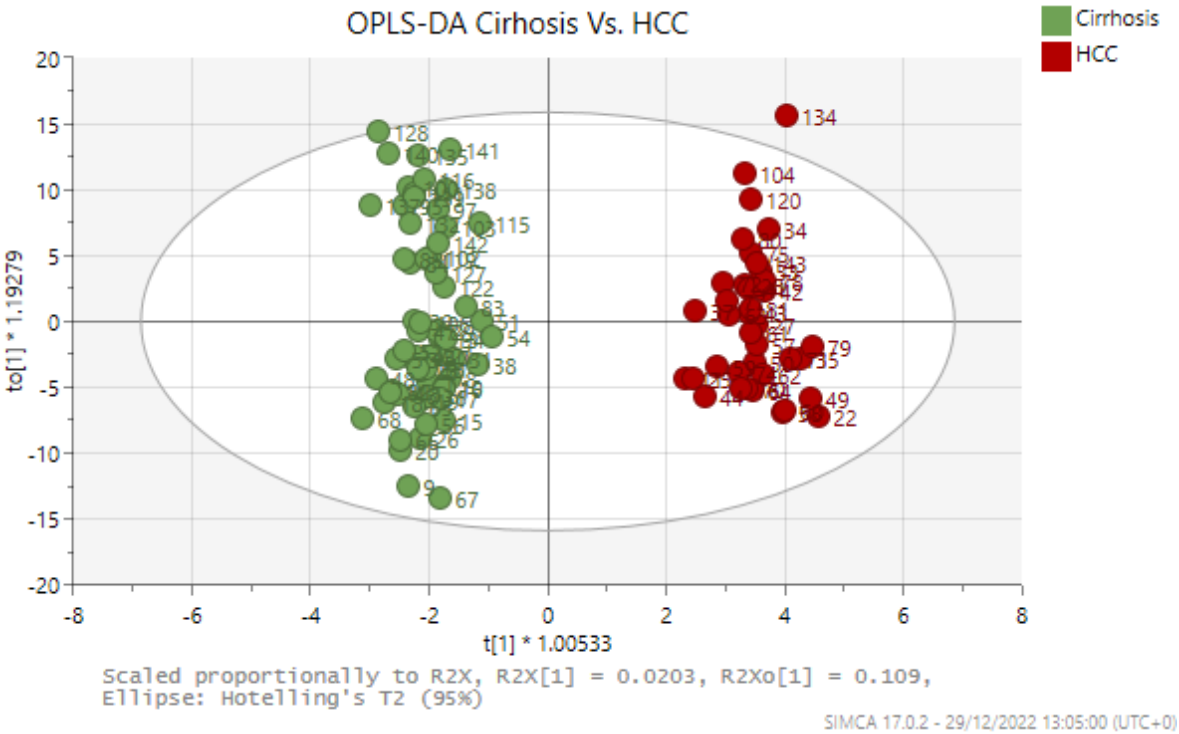


Figure 15.9 OPLS-DA of cirrhosis vs HCC demonstrating excellent separation of the two cohorts.

The chromatogram peaks of the VOCs driving the separation with a VIP score >1.5 were interrogated. Misidentified peaks were corrected, and further contaminants removed. Following FDR correction, five VOCs (cyclopentane, methyl-, hexane, 2,4,4-trimethyl-, hexane, 3-methyl- and neopentylamine) were excluded from analysis (p >0.05). The seven FDR-significant compounds driving the separation with VIP score >1.5 can be seen in table 15.2.

Table 15.2 FDR significant VOCs with VIP score >1.5 of cirrhosis vs HCC.

Compound Identification	CAS Number	RT	FDR p Value	VIP Score	AUC
3-Heptanone	106-35-4	10.66	<0.01	1.86	0.77
1-Hexanol, 2,2-dimethyl-	2370-13-0	12.75	<0.01	1.59	0.70
Octane, 2,2-dimethyl-	15869-87-1	12.04	<0.01	1.64	0.70
1-Propanol, 3,3'-oxybis-	2396-61-4	16.75	<0.01	1.69	0.69
Octane, 2,6-dimethyl-	2051-30-1	13.93	<0.01	1.81	0.67
1-Dodecanol	112-53-8	13.54	0.03	1.74	0.63
1-Hexene, 3,4-dimethyl-	16745-94-1	6.23	0.02	1.68	0.62

A ROC curve generated using all seven compounds (appendix 17.7.10, figure 17.31) generated an AUC of 0.805 (95% CI 0.658-0.913). A model of seven VOCs will be liable to overfitting. When including only the top three discriminating compounds (3-heptanone, 1-hexanol, 2,2-diemethyl- and octane, 2,2-dimethyl), an AUC of 0.765 (95% CI 0.632 to 0.873) was generated (figure 15.10).

Box plots of the top five FDR significant compounds with VIP score >1.5 with the highest AUCs can be found in appendix 17.7.11, figures 17.38 to 17.42. For all five VOCs (3-heptanone, 1-hexanol, 2,2-dimethyl-, octane, 2,2-dimethyl-, 1-propanol, 3,3'-oxybis- and octane, 2,6-dimethyl-), their concentrations were higher in those with cirrhosis compared to those with HCC. The ambient room air samples collected from St. Mary's Hospital reassuringly demonstrated low concentrations of 3-hepatanone and octane, 2,2-dimethyl. 1-Hexanol, 2,2-dimethyl and octane, 2,6-dimethyl- were not detected in the limited room air samples. Unfortunately, no room air samples are available for Hammersmith Hospital, where the majority of HCC participants were recruited.

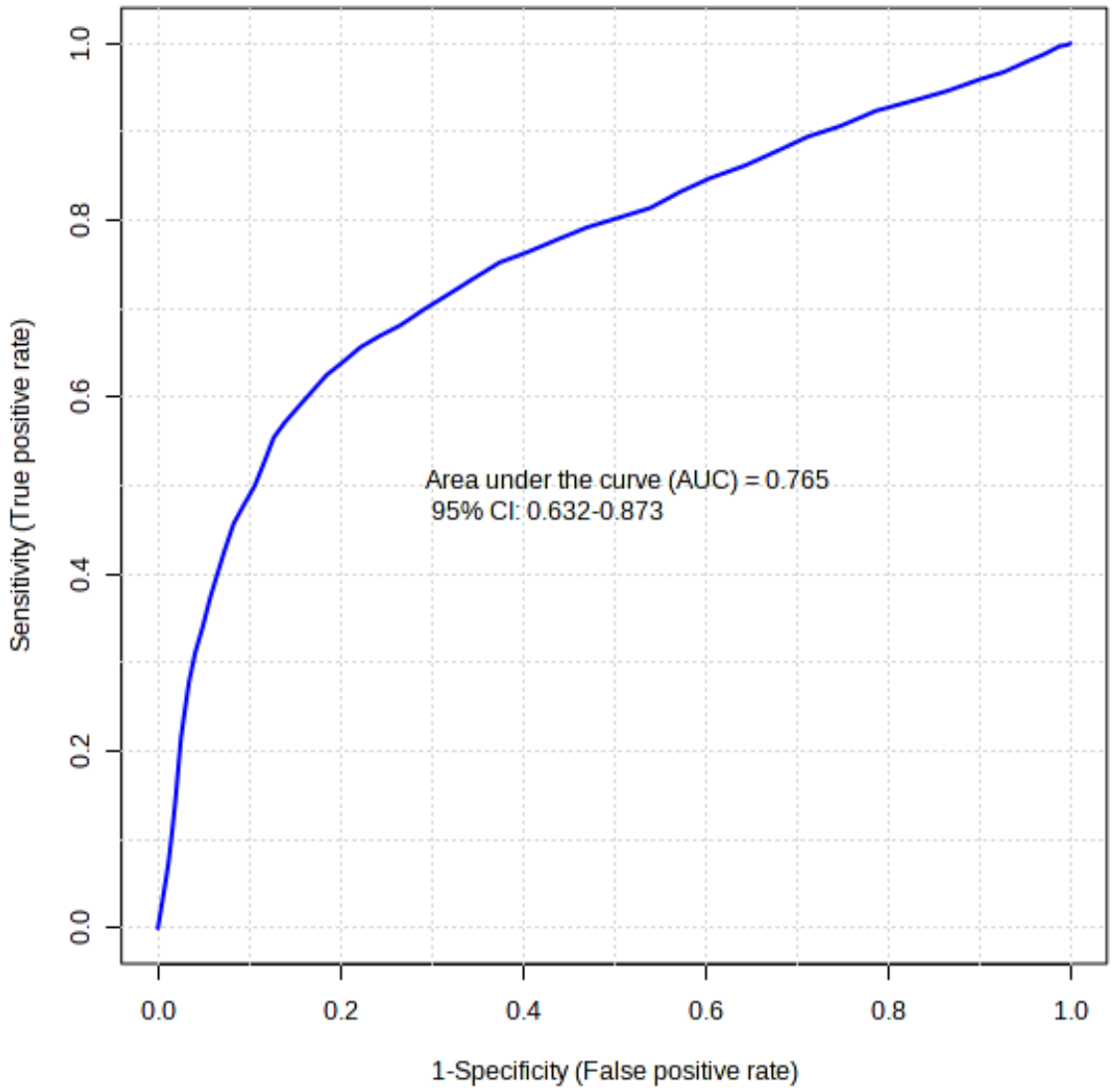


Figure 15.10 ROC curve for Cirrhosis vs HCC with top three FDR significant compounds with VIP score >1.5.

15.4 Cirrhosis vs Control

All participants with cirrhosis only were compared to healthy control participants only. An unsupervised PCA demonstrated some separation (figure 15.11). Supervised analysis via OPLS-DA improved the separation (figure 15.12). The separation was significant according to CV ANOVA ($p < 0.001$) and permutation plots with 999 permutations (appendix 17.7.9, figure 17.21).

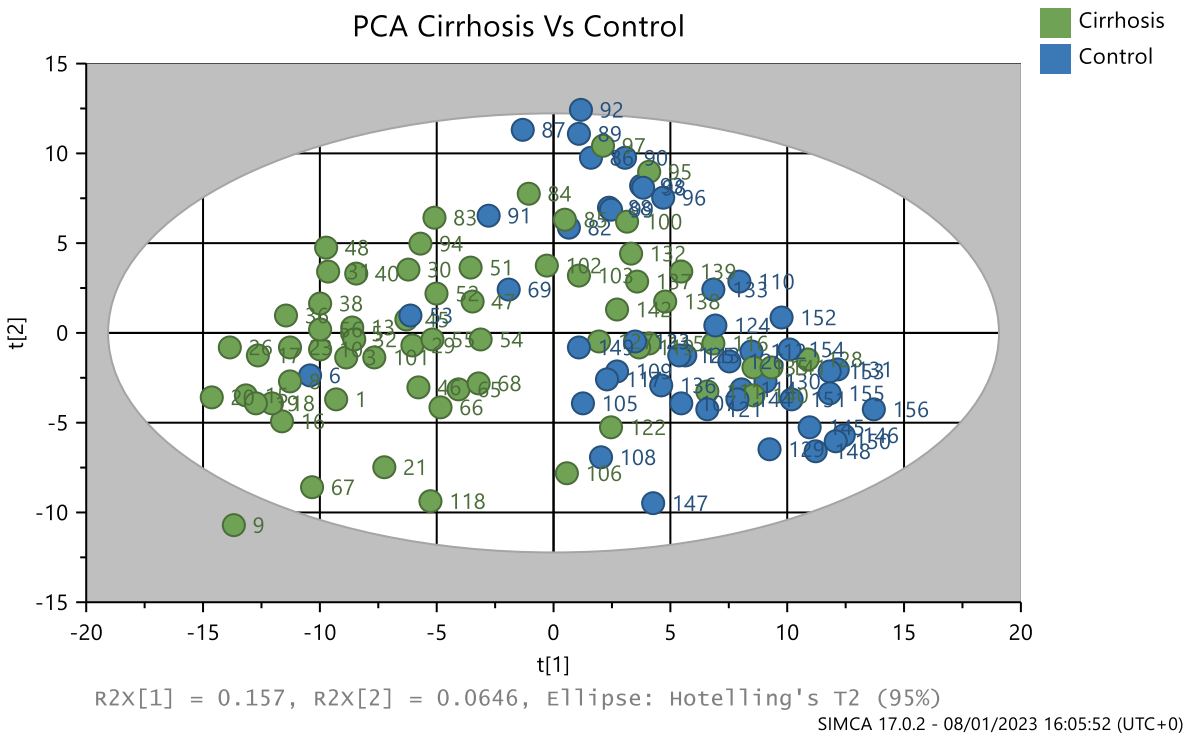


Figure 15.11 PCA of cirrhosis participants vs control participants demonstrating some separation between the two cohorts.

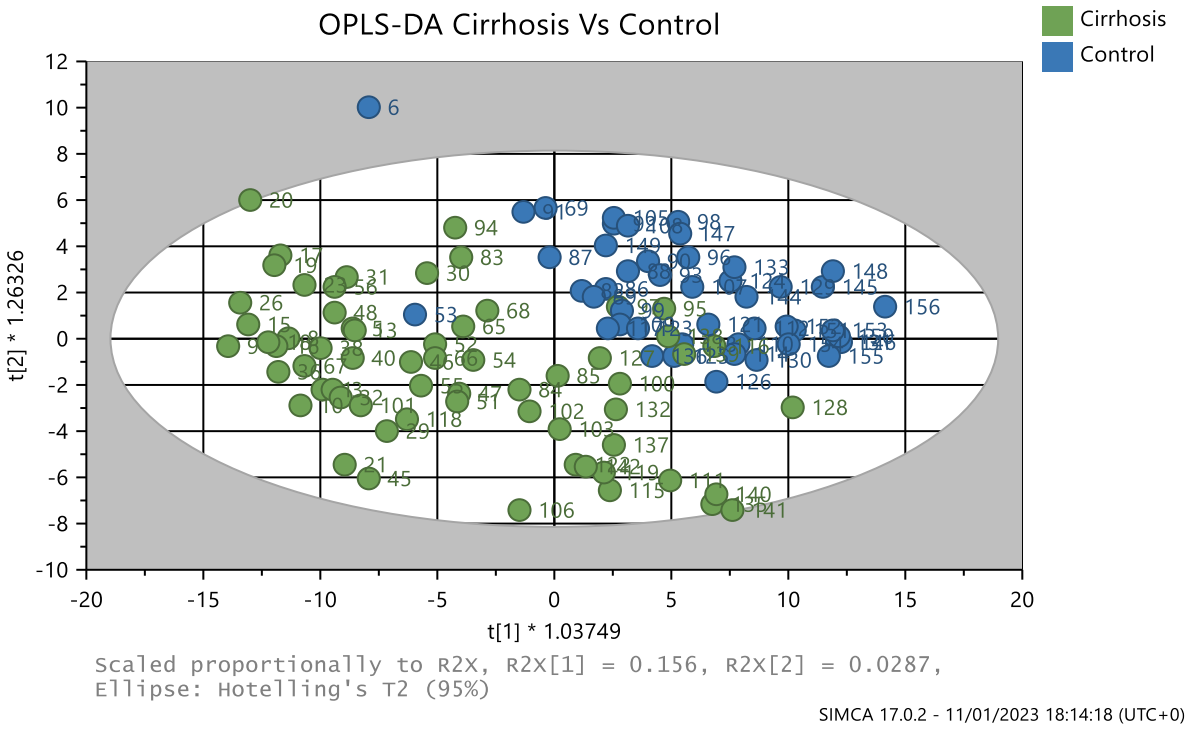


Figure 15.12 OPLS-DA of cirrhosis participants vs control participants demonstrating good separation between the two cohorts.

29 compounds with a VIP score >1.5 were significant after FDR correction (p <0.05) (table 15.3).

Table 15.3 FDR significant VOCs with VIP Score >1.5 of cirrhosis vs control.

Compound Identification	CAS Number	RT	FDR p Value	VIP Score	AUC
m-Cymene	535-77-3	14.19	<0.01	1.88	0.86
α-Terpinene	99-86-5	12.72	<0.01	1.62	0.81
Limonene	138-86-3	12.91	<0.01	1.58	0.81
Terpinolene	586-62-9	13.89	<0.01	1.58	0.81
Cyclopentane, methyl-	96-37-7	4.67	<0.01	1.94	0.80
p-Cymene	99-87-6	12.97	<0.01	1.55	0.79
Cyclopentane	287-92-3	6.15	<0.01	1.94	0.79
2-Hexene, 2-methyl-	2738-19-4	6.16	<0.01	1.74	0.78
Cyclopentanone	120-92-3	4.17	<0.01	1.98	0.78
1-Heptene	592-76-7	5.90	<0.01	1.93	0.76
3-Hexanone, 4,4-dimethyl-	19550-14-2	5.18	<0.01	1.55	0.76
1-Hexene, 4,5-dimethyl-	16106-59-5	3.50	<0.01	1.51	0.76
Propanoic acid, 2,2-dimethyl-, ethyl ester	3938-95-2	4.76	<0.01	1.77	0.76
n-Hexane	110-54-3	4.01	<0.01	1.95	0.76
Hexane, 2,4,4-trimethyl-	16747-30-1	5.43	<0.01	1.86	0.76
2-Propenal	107-02-8	1.78	<0.01	1.60	0.75
2,4,4-Trimethyl-1-pentanol	16325-63-6	17.93	<0.01	1.52	0.74
2-Butene	107-01-7	2.30	<0.01	1.92	0.73
Pentane	109-66-0	2.55	<0.01	1.88	0.73
1-Octanol, 2-butyl-	3913-02-8	19.90	<0.01	1.76	0.73
Tetradecane, 2-methyl-	1560-95-8	17.93	<0.01	1.56	0.72
Nonanal	124-19-6	14.58	<0.01	1.65	0.72
Phenol, 3-ethyl-	620-17-7	16.55	<0.01	1.57	0.72
2-Butene, 2,3-dimethyl-	563-79-1	4.49	<0.01	1.89	0.71
1-Hexene, 3,4-dimethyl-	16745-94-1	6.25	<0.01	1.84	0.70
[1,1'-Bicyclopentyl]-2-one	4884-24-6	4.38	<0.01	1.79	0.69
1-Propyl-1-cyclopentanol	1604-02-0	6.64	<0.01	1.73	0.69
Heptane, 2-methyl-	592-27-8	7.43	<0.01	1.52	0.69
Pentane, 2,2-dimethyl-	590-35-2	4.49	<0.01	1.81	0.68

A ROC curve generated using all 29 compounds with random forests classification and SVM feature ranking gave an AUC of 0.861 (95% CI 0.784-0.93) (appendix 17.7.10, figure 17.30).

A model of the top 5 discriminating compounds (m-cymene, α-terpinene, limonene, terpinolene and cyclopentane, methyl-), gave an AUC of 0.857 (95% confidence interval 0.782-0.921) (figure 15.13).

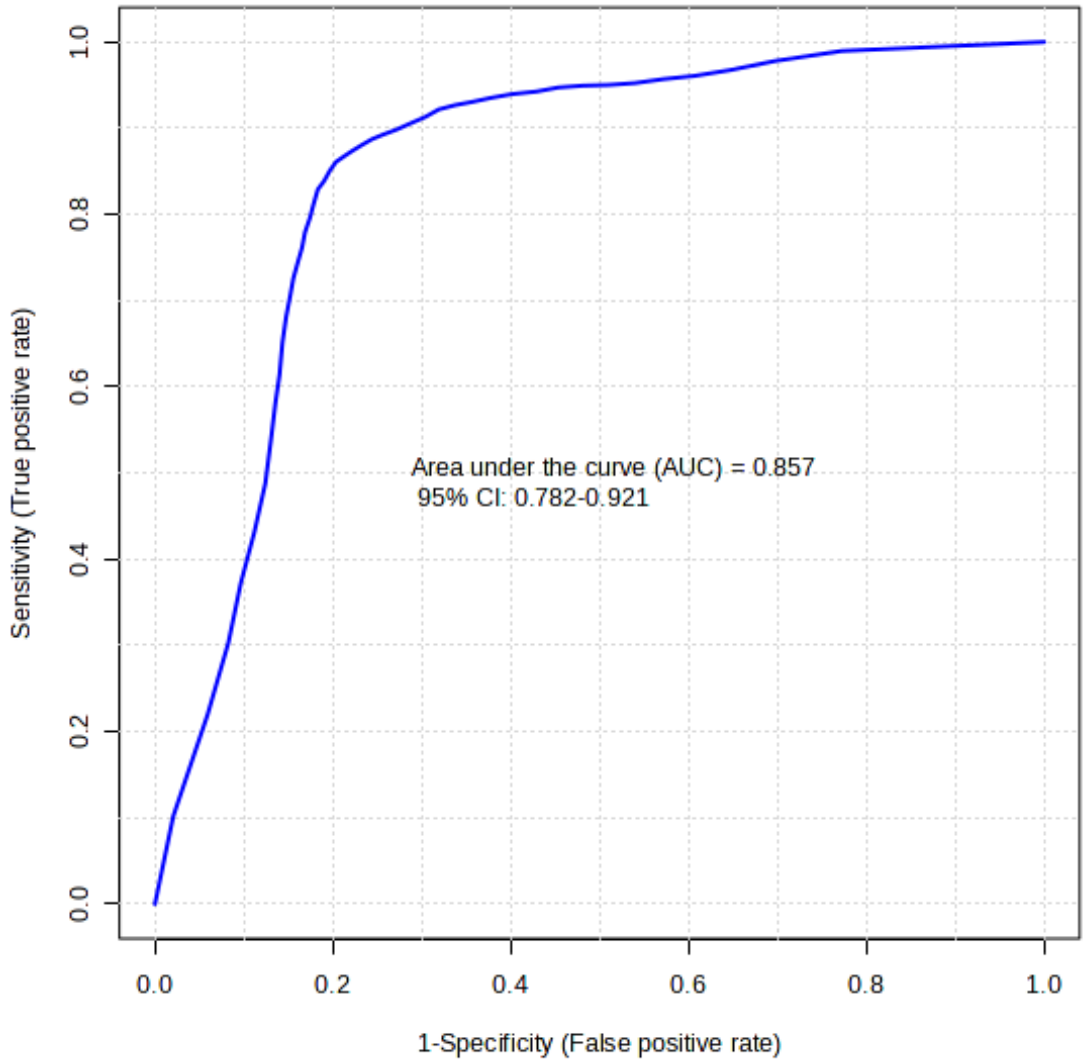


Figure 15.13 ROC curve of cirrhosis vs control with top five FDR significant VOCs with VIP score >1.5.

Box plots of the top five FDR significant compounds with VIP score >1.5 with the highest AUCs can be found in appendix 17.7.11, figures 17.39-17.43. All five VOCs were higher in the cirrhosis cohort compared to the control cohort. m-Cymene, α-terpinene and terpinolene were not detected in room air samples. Limonene and cyclopentane, methyl- were of low abundance in room air.

15.5 HCC vs Controls

All participants with HCC were compared to control participants only. Unsupervised PCA demonstrated good separation (figure 15.14). Supervised analysis via OPLS-DA improved this separation (figure 15.15). The separation was significant according to CV ANOVA (p <0.001) and permutation plots with 999 permutations (appendix 17.7.9, figure 17.22).

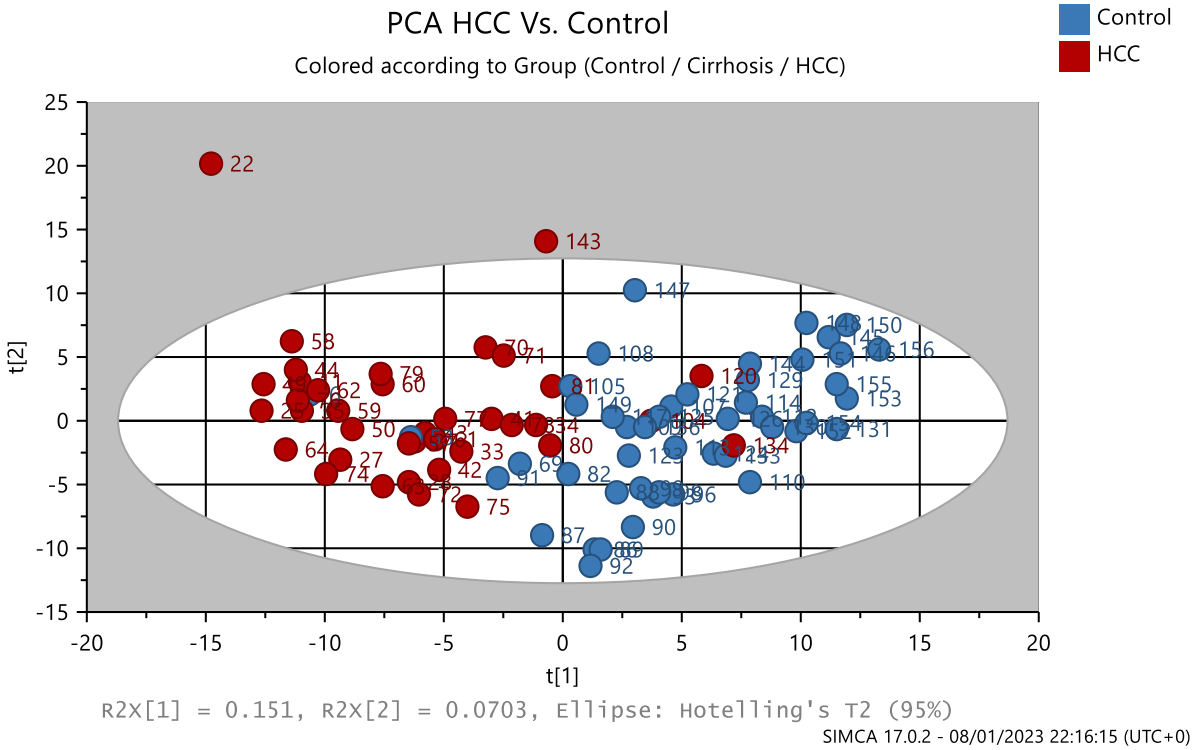


Figure 15.14 PCA of HCC participants vs control participants demonstrating clustering of cohorts and good separation.

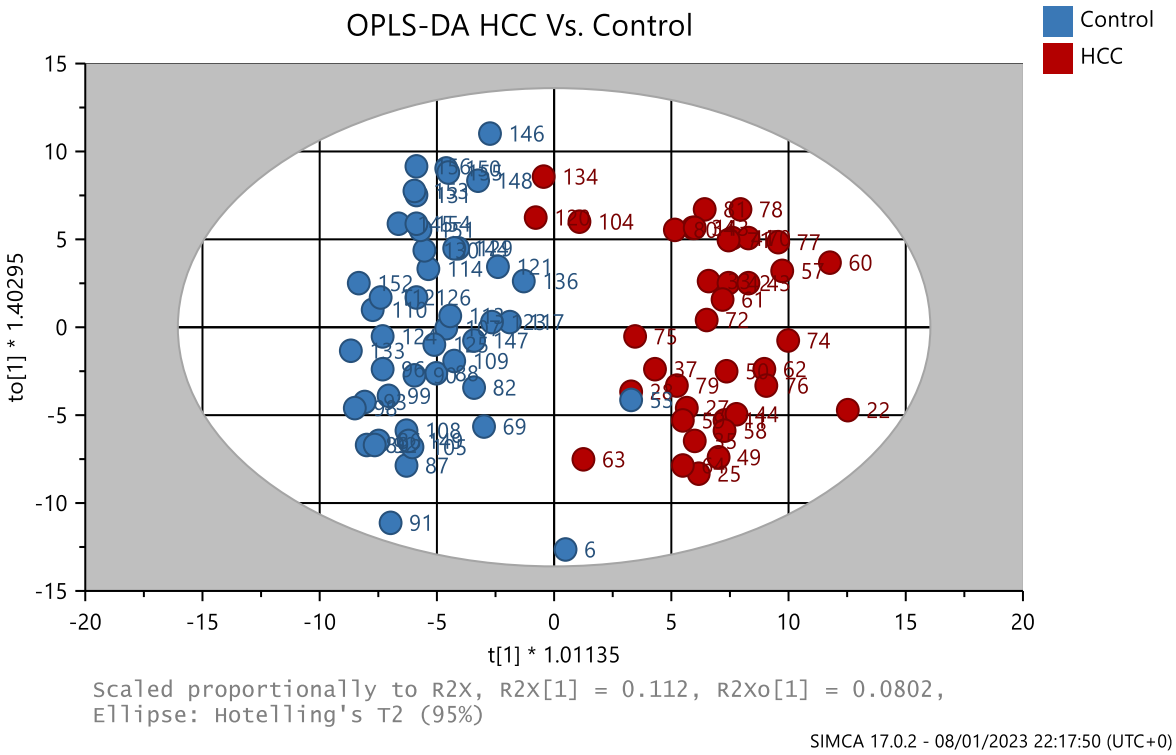


Figure 15.15 OPLS-DA of HCC participants vs control participants demonstrating excellent separation of cohorts.

23 compounds with a VIP score >1.5 were significant according to FDR correction ($p < 0.05$) (table 15.4).

Table 15.4 FDR significant VOCs with VIP score >1.5 of HCC vs controls.

Compound Identification	CAS Number	RT	FDR p Value	VIP Score	AUC
1-Octen-3-one	4312-99-6	3.52	<0.01	1.73	0.93
Cyclopentane, methyl-	96-37-7	4.67	<0.01	2.14	0.92
Cyclopentane	287-92-3	6.15	<0.01	2.11	0.92
Pentane, 2,3-dimethyl-	565-59-3	5.44	<0.01	1.63	0.91
Hexane, 2,4,4-trimethyl-	16747-30-1	5.43	<0.01	2.26	0.90
2-Butene	107-01-7	2.30	<0.01	2.04	0.90
Pentane	109-66-0	2.55	<0.01	2.11	0.89
1-Hexene, 4,5-dimethyl-	16106-59-5	3.50	<0.01	1.80	0.89
n-Hexane	110-54-3	4.01	<0.01	1.70	0.88
3-Heptanone	106-35-4	10.66	<0.01	1.56	0.83
1-Octanol, 2-butyl-	3913-02-8	19.90	<0.01	1.91	0.82
2-Propenal	107-02-8	1.78	<0.01	1.68	0.82
Limonene	138-86-3	12.91	<0.01	1.55	0.82
2-Butanone	78-93-3	4.76	<0.01	1.68	0.82
m-Cymene	535-77-3	14.19	<0.01	1.79	0.81
Propanoic acid, 2,2-dimethyl-, ethyl ester	3938-95-2	4.76	<0.01	1.63	0.81
Heptane, 2-methyl-	592-27-8	7.43	<0.01	1.67	0.81
p-Cymene	99-87-6	12.97	<0.01	1.53	0.80
2-Hexene, 2-methyl-	2738-19-4	6.16	<0.01	1.53	0.79
Octane, 2,6-dimethyl-	2051-30-1	13.93	<0.01	1.60	0.79
Nonanal	124-19-6	14.58	<0.01	1.65	0.77
2-Octene, (E)-	13389-42-9	8.41	<0.01	1.54	0.75
Cyclopentanone	120-92-3	4.17	<0.01	1.52	0.69

A ROC curve generated using all 23 compounds with random forests classification and SVM feature ranking gave an AUC of 0.943 (95% CI 0.71-1.00) (appendix 17.7.10, figure 17.31). The top five discriminating compounds (1-octen-3-one, cyclopentane, methyl-, cyclopentane, pentane, 2, 3-dimethyl- and hexane, 2,4,4-trimethyl) when combined gave an AUC of 0.929 (95% CI 0.868 to 0.998) (figure 15.16).

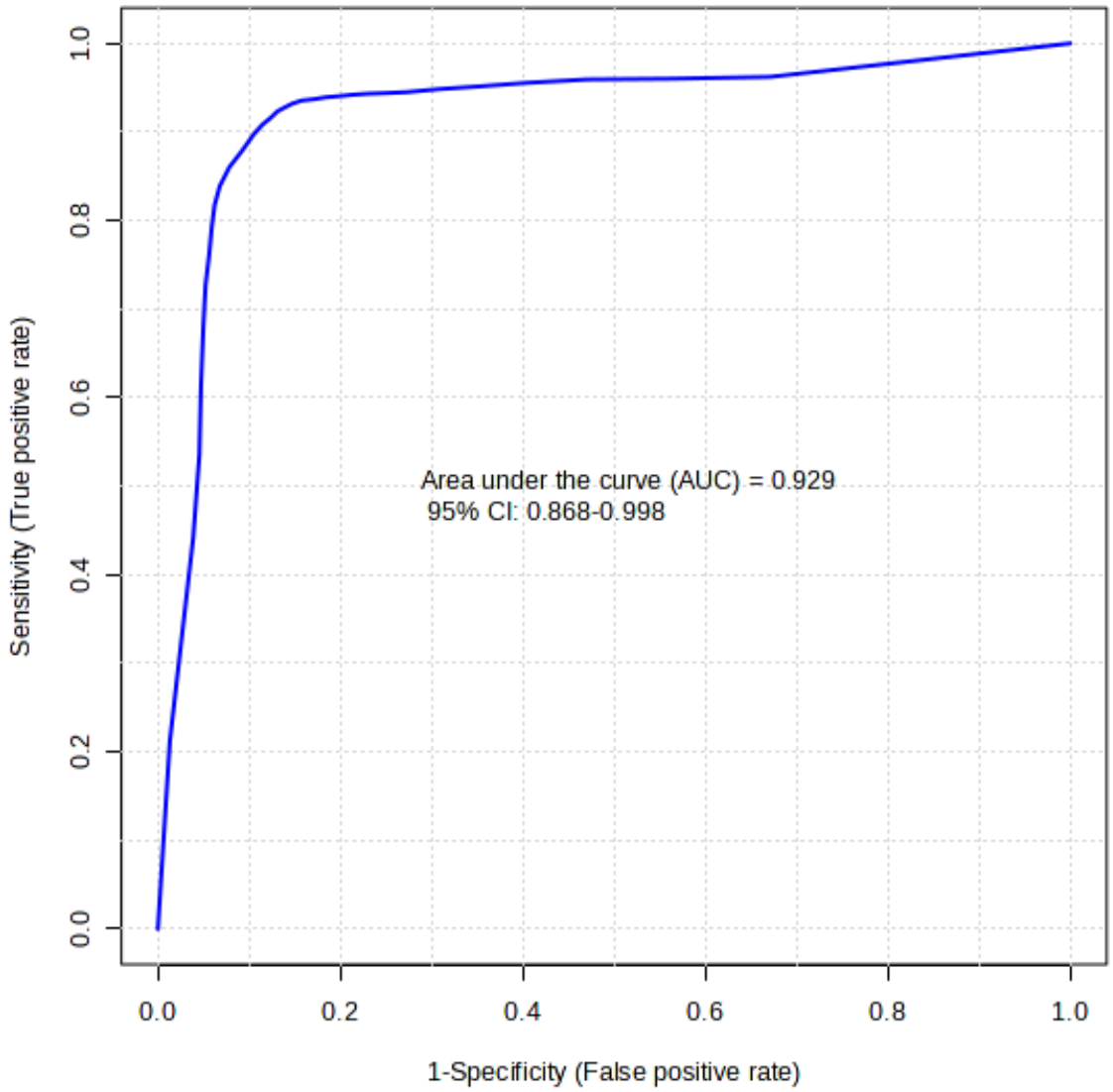


Figure 15.16 ROC curve of HCC vs controls with top five FDR significant VOCs with VIP score >1.5.

Box plots of the top five FDR significant compounds with VIP score >1.5 with the highest AUCs can be found in appendix 17.7.11, figures 17.58 to 17.62. Review of room air samples demonstrated low concentrations of the VOCs of interest with the exception of pentane, 2, 3-dimethyl- whose abundance comparable to that of healthy controls.

15.6 Compensated vs Decompensated Cirrhosis

Subgroup analysis was performed on the cirrhosis cohort by further dividing them into compensated or decompensated cohorts. Unsupervised PCA demonstrated no separation (appendix 17.7.12, figure 17.64). Supervised analysis via OPLS-DA provided some separation (figure appendix 17.7.12, figure 17.65) although the separation was not

significant according to CV ANOVA testing (p = 1.0). Therefore, further analysis was not performed.

15.7 Child-Pugh A vs Child-Pugh B/C Cirrhosis

All participants with Child-Pugh B and C cirrhosis were grouped together and compared to those with the less advanced disease state of Child-Pugh A cirrhosis. Unsupervised PCA demonstrated minimal separation (figure 15.17). Supervised analysis with OPLS-DA improved this separation (figure 15.18). The separation was significant according to CV ANOVA (p <0.001) and permutation plots with 999 permutations (appendix 17.7.9, figure 17.23).

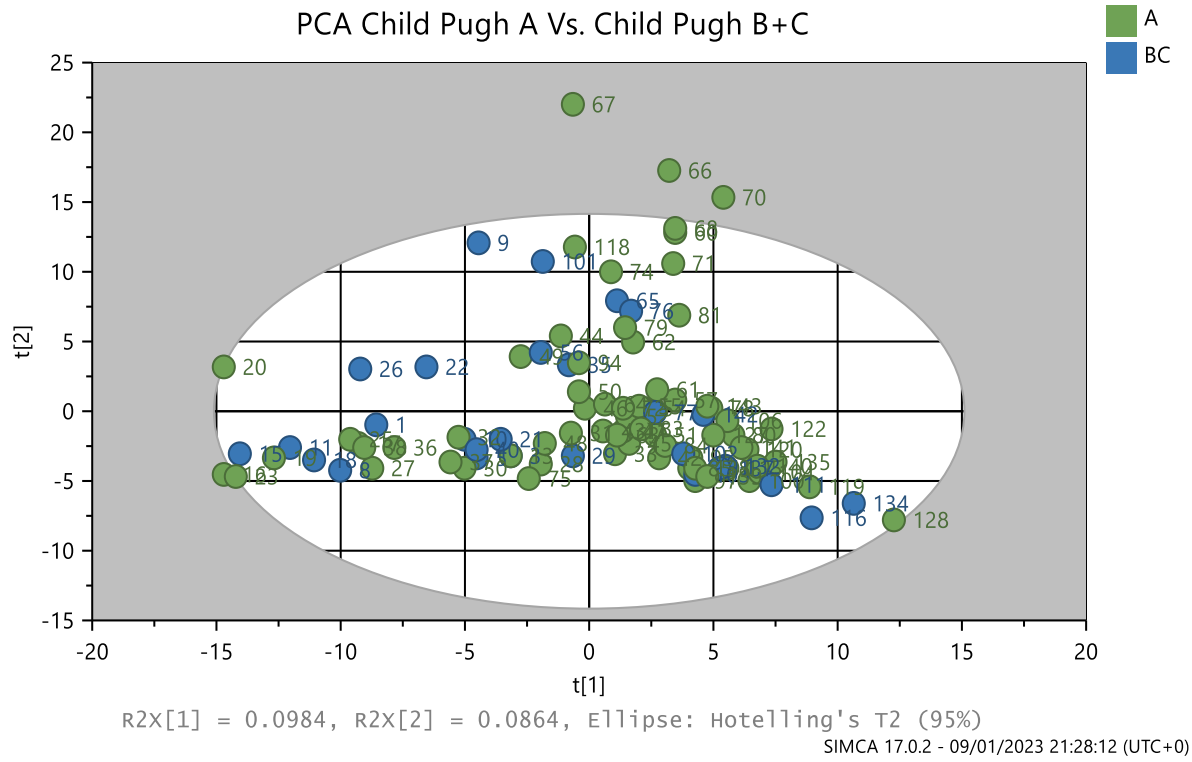


Figure 15.17 PCA of Child-Pugh A cirrhosis vs Child-Pugh B/C cirrhosis demonstrating no separation between the two cohorts.

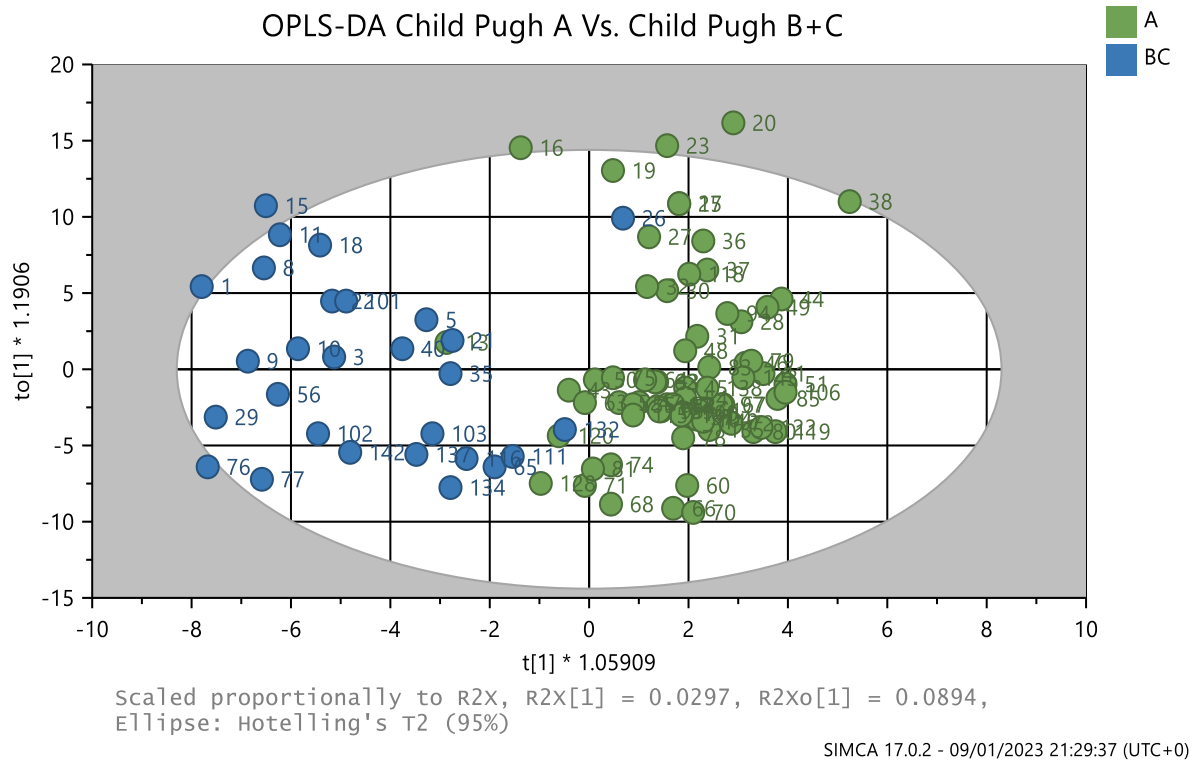


Figure 15.18 OPLS-DA of Child-Pugh A cirrhosis vs Child-Pugh B/C cirrhosis demonstrating good separation between the two cohorts.

VOCs with a VIP score >1.5 were significant according to FDR correction ($p < 0.05$) (table 15.5).

Table 15.5 FDR significant VOCs with VIP Score >1.5 of Child-Pugh A cirrhosis vs Child-Pugh B/C cirrhosis.

Compound Identification	CAS Number	RT	FDR p Value	VIP Score	AUC
3-Carene	13466-78-9	12.51	<0.01	2.71	0.85
Limonene	138-86-3	12.91	<0.01	2.00	0.78
2-Pentanone	107-87-9	6.63	<0.01	2.17	0.75
Dodecanal	112-54-9	19.12	<0.01	1.78	0.71
Pentadecane	629-62-9	19.98	<0.01	1.57	0.69
1-Dodecanol	112-53-8	13.54	0.01	1.70	0.68
3-Hexanone, 4,4-dimethyl-	19550-14-2	5.18	0.01	1.59	0.67
Copaene	3856-25-5	18.52	0.01	1.61	0.66
Heptadecane	629-78-7	20.91	0.02	2.13	0.66
1-Octanol, 2-butyl-	3913-02-8	19.90	0.02	1.50	0.65
Sulfide, allyl methyl	10152-76-8	6.42	0.02	1.70	0.65
1-Hexene, 3,4-dimethyl-	16745-94-1	6.25	0.01	1.75	0.65
2-Butene, 2,3-dimethyl-	563-79-1	4.49	0.02	1.66	0.64
Cyclododecanol	1724-39-6	22.87	0.04	1.64	0.64
Pentane, 2,2-dimethyl-	590-35-2	4.49	0.03	1.64	0.62

A ROC curve generated using all 15 compounds with random forests classification and SVM feature ranking gave an AUC of 0.828 (95% CI 0.708-0.931) (appendix 17.7.10, figure 17.32). The top five discriminating compounds (3-carene, limonene, 2-pentanone, dodecanal and pentadecane), when combined gave an AUC of 0.831 (95% CI 0.718-0.941) (figure 15.19).

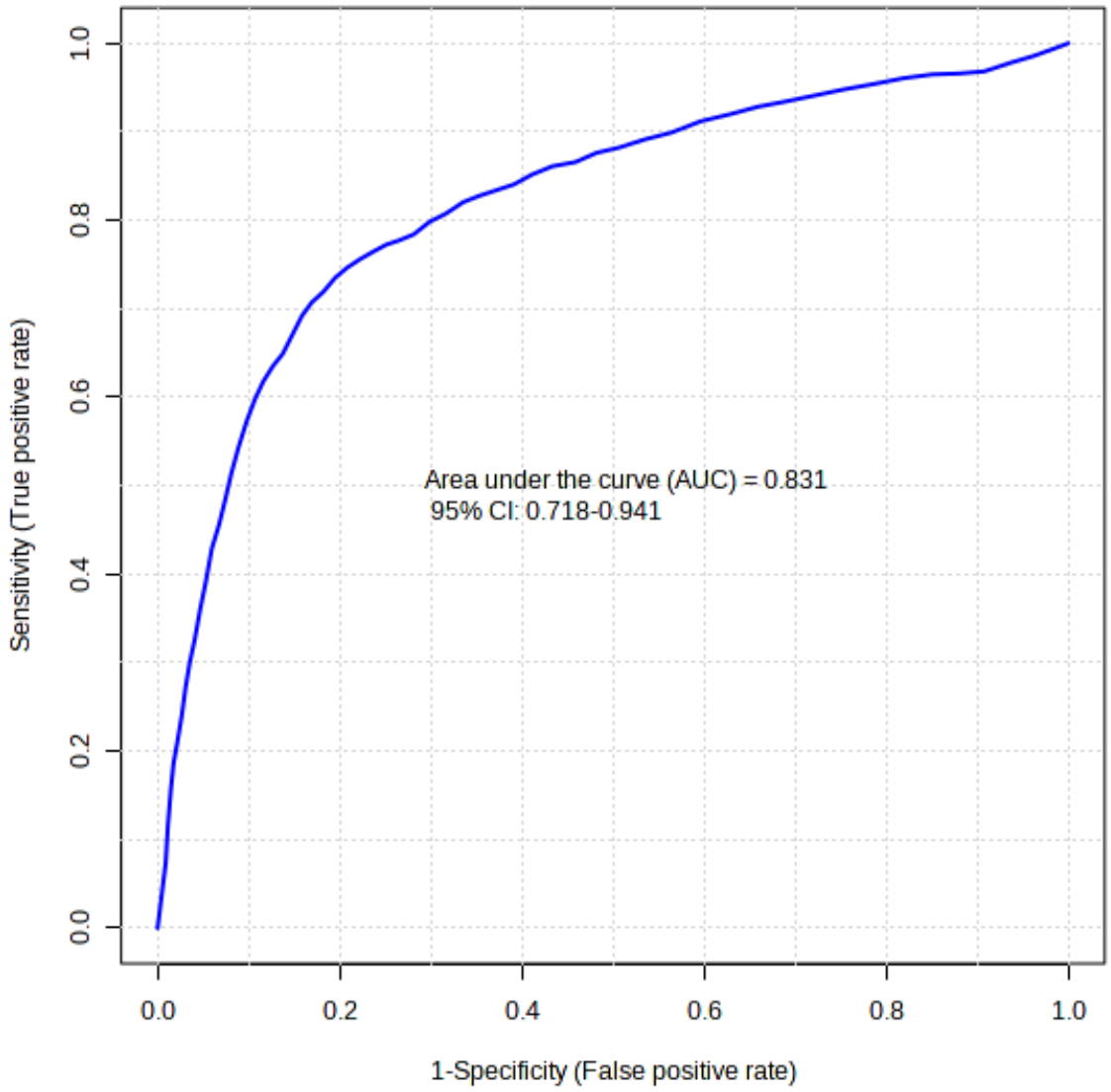


Figure 15.19 ROC curve of Child-Pugh A vs Child-Pugh B/C cirrhosis with top five FDR significant compounds with VIP score >1.5.

Box plots of the top five FDR significant compounds with VIP score >1.5 with the highest AUCs can be found in appendix 17.7.11, figures 17.49-17.53. 3-Carene, limonene, 2-pentanone were elevated in Child-Pugh B/C cirrhotics compared to Child-Pugh A but the inverse was true of pentadecane and dodecanal. Pentadecane was not detected in the

room air samples and all other VOCs were of low abundance compared to breath samples with the exception of dodecanal which had an abundance in room air that was comparable to those in breath samples.

15.8 Cirrhosis Aetiology

All participants with cirrhosis were compared based on their aetiology (autoimmune conditions, alcohol, hepatitis B, hepatitis C, iron overload, non-alcoholic steatohepatitis or thrombosis related). Unsupervised PCA demonstrated no separation (appendix 17.7.13, figure 17.66). Supervised analysis with OPLS-DA did not improve the separation (appendix 17.7.13, figure 17.67) and the separation was not significant according to CV ANOVA (p 1.0) therefore further analysis was not performed.

15.9 Portal Hypertension

While portal hypertension is a common sequela of cirrhosis, it can also develop in non-cirrhotic liver (non-cirrhotic portal hypertension) (289). Those with cirrhosis can also have normal portal pressures. Thus, all participants (control, cirrhosis, and HCC) were classified on the basis of the presence or absence of portal hypertension. Unsupervised PCA demonstrated no separation (figure 15.20). Supervised analysis via OPLS-DA demonstrated some separation (figure 15.21). The separation was significant according to CV ANOVA (p <0.001) and permutation plots with 999 permutations (appendix 17.7.9, figure 17.24).

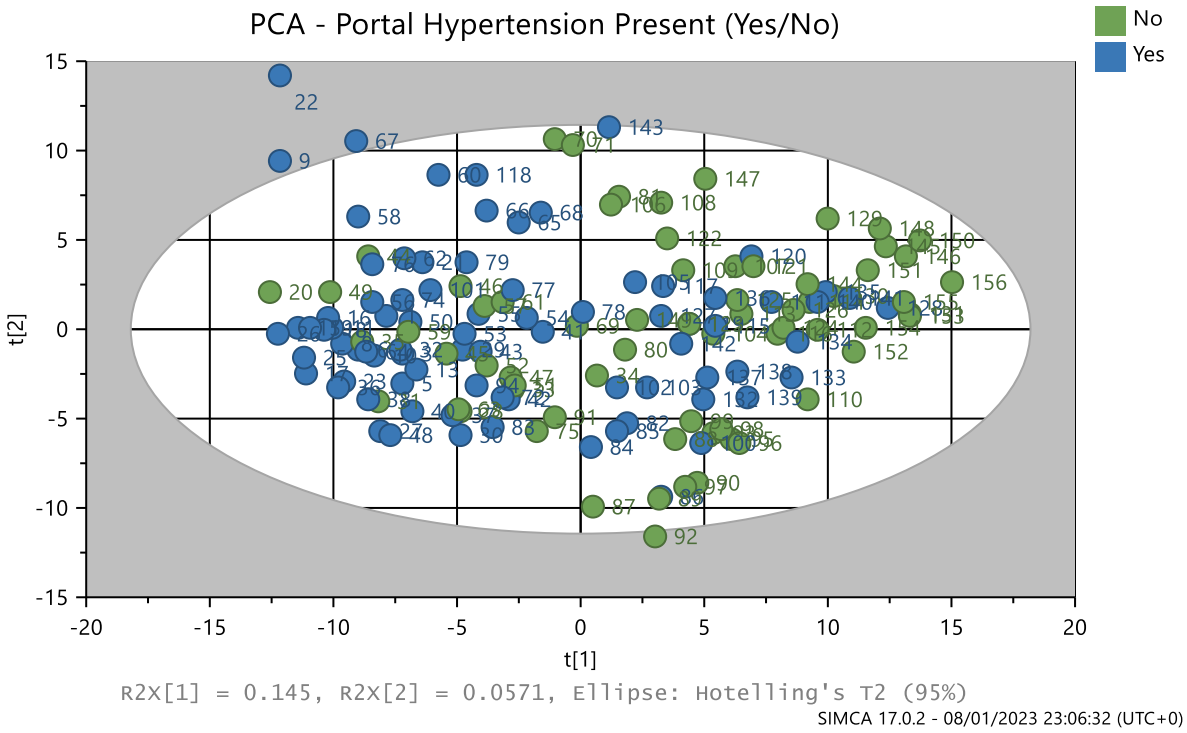


Figure 15.20 PCA of participants with and without portal hypertension demonstrating significant overlap between the two cohorts and no separation.

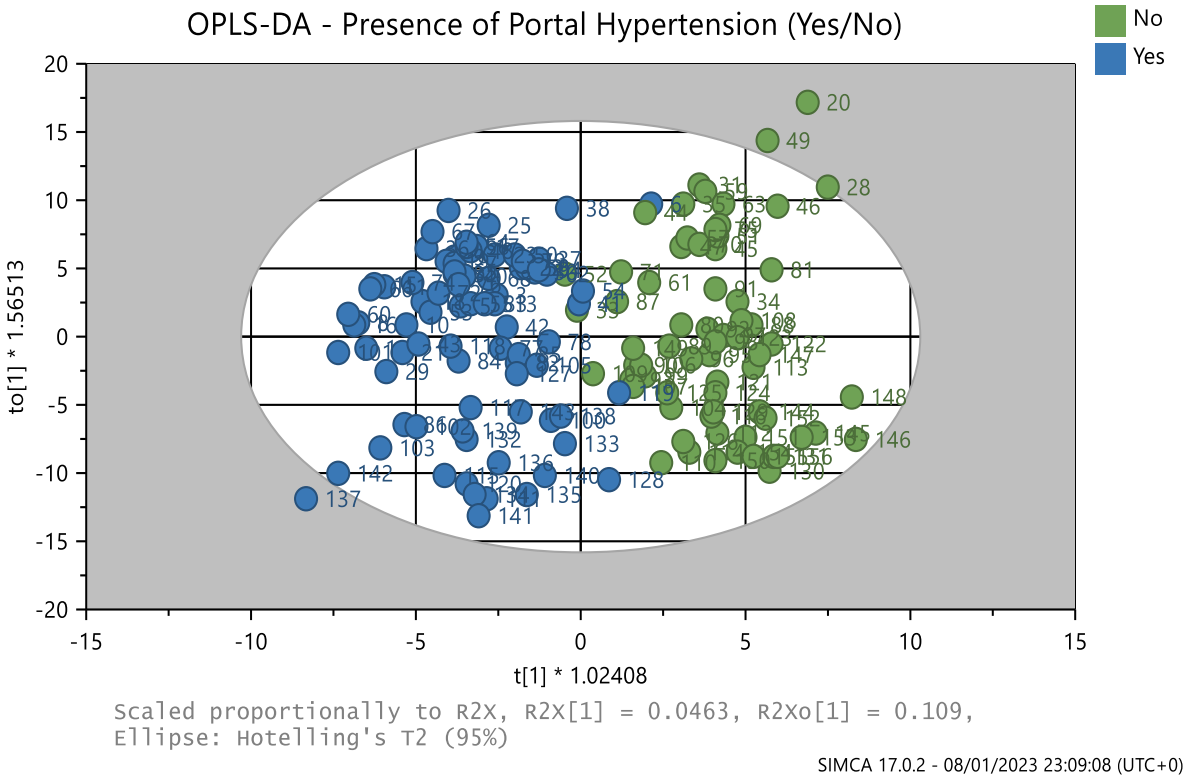


Figure 15.21 OPLS-DA of participants with and without portal hypertension demonstrating good separation between the two cohorts following supervised analysis.

28 compounds with a VIP score >1.5 were significant according to FDR correction ($p < 0.05$) (table 15.6).

Table 15.6 FDR Significant VOCs with VIP Score >1.5 of presence or absence of portal hypertension.

Compound Identification	CAS Number	RT	FDR p Value	VIP Score	AUC
o-Xylene	95-47-6	3.40	<0.01	1.76	0.80
Limonene	138-86-3	12.91	<0.01	2.34	0.79
3-Hexanone, 4,4-dimethyl-	19550-14-2	5.18	<0.01	2.03	0.79
p-Cymene	99-87-6	12.97	<0.01	1.95	0.78
β-Pinene	127-91-3	12.05	<0.01	2.05	0.78
α-Terpinene	99-86-5	12.72	<0.01	2.06	0.77
m-Cymene	535-77-3	14.19	<0.01	2.20	0.76
Terpinolene	586-62-9	13.89	<0.01	2.07	0.76
α-Pinene	80-56-8	11.11	<0.01	1.87	0.75
2-Butanol	78-92-2	4.98	<0.01	1.64	0.74
2-Pentanone	107-87-9	6.63	<0.01	1.89	0.72
2-Hexene, 2-methyl-	2738-19-4	6.16	<0.01	1.75	0.71
1-Decene	872-05-9	11.96	<0.01	1.52	0.71
Cyclopentanone	120-92-3	4.17	<0.01	1.78	0.71
2-Butene, 2,3-dimethyl-	563-79-1	4.49	<0.01	1.68	0.69
Cyclopentane, methyl-	96-37-7	4.67	<0.01	1.79	0.69
1-Heptene	592-76-7	5.90	<0.01	1.63	0.69
Hexane, 2,4,4-trimethyl-	16747-30-1	5.43	<0.01	1.56	0.68
2-Octene, (E)-	13389-42-9	8.41	<0.01	1.55	0.68
1-Propyl-1-cyclopentanol	1604-02-0	6.64	<0.01	1.74	0.68
2-Butanone	78-93-3	4.76	<0.01	1.62	0.67
1-Hexene, 3,4-dimethyl-	16745-94-1	6.25	<0.01	1.65	0.67
n-Hexane	110-54-3	4.01	<0.01	1.71	0.66
1-Octanol, 2-butyl-	3913-02-8	19.90	<0.01	1.60	0.66
[1,1'-Bicyclopentyl]-2-one	4884-24-6	4.38	<0.01	1.50	0.66
Pentane	109-66-0	2.55	<0.01	1.59	0.66
Pentane, 2,2-dimethyl-	590-35-2	4.49	<0.01	1.57	0.66
2-Butene	107-01-7	2.30	<0.01	1.62	0.66

A ROC curve generated using all 28 compounds with random forests classification and SVM feature ranking gave an AUC of 0.833 (95% CI 0.765-0.908) (appendix 17.7.10, figure 17.33). The top five discriminating compounds (o-xylene, limonene, 3-hexanone, 4,4-

dimethyl-, p-cymene and β-pinene), when combined, gave an AUC of 0.772 (95% CI 0.71-0.866) (figure 15.22).

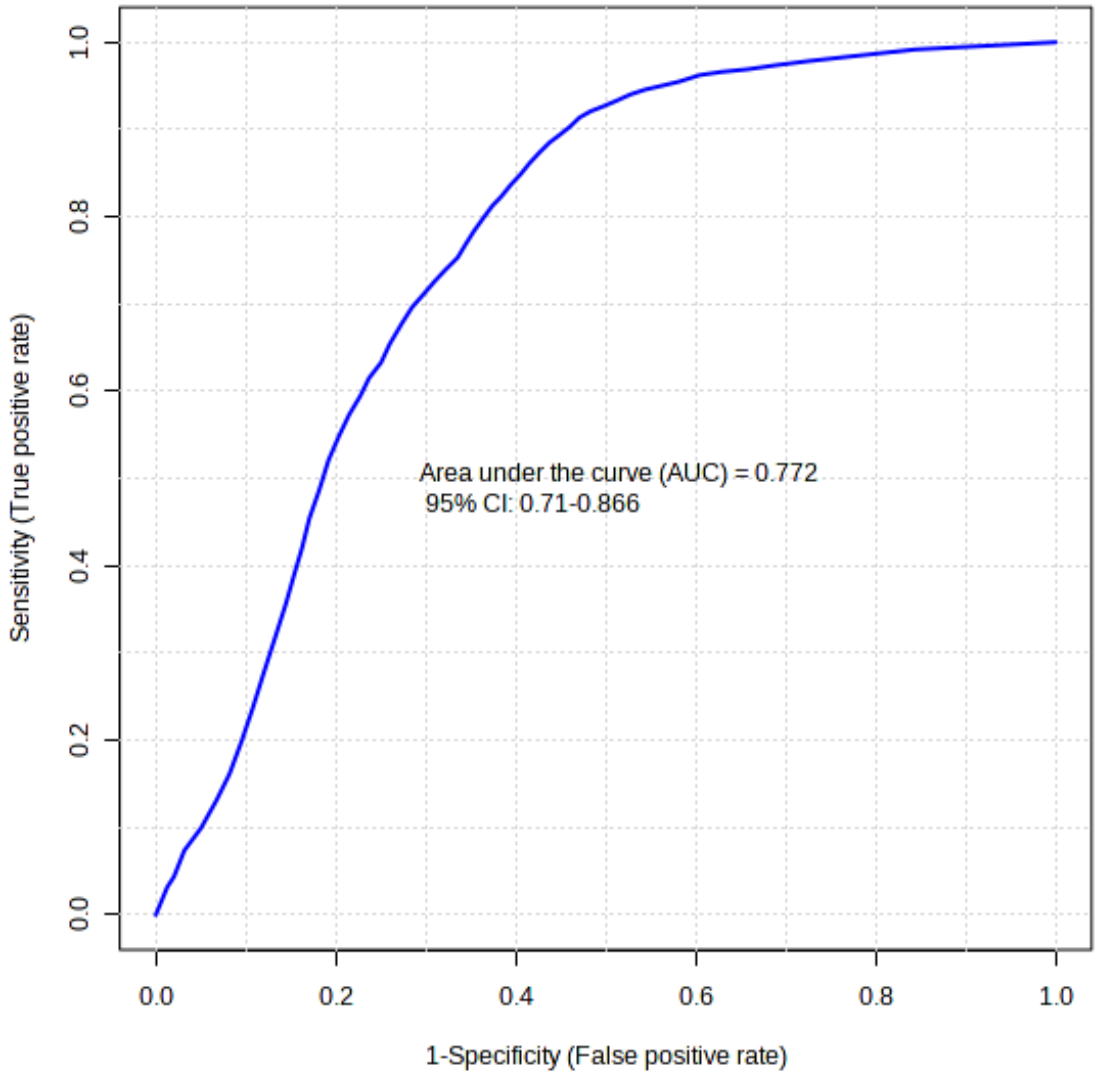


Figure 15.22 ROC of presence or absence of portal hypertension with all FDR significant VOCs with a VIP score >1.5.

Box plots of the FDR significant compounds with VIP score >1.5 with the highest AUCs can be found in appendix 17.7.11, figures 17.54-17.58. All VOCs were higher in those with portal hypertension compared to those without portal hypertension. o-Xylene was not detected in room air and all other VOCs were of low abundance in room air compared to the abundance in breath samples.

15.10 HCC Barcelona Clinic Classification

All participants with HCC were classified on the basis of their Barcelona clinic liver cancer classification. An unsupervised PCA demonstrated no separation (appendix 17.7.14, figure 17.68). Supervised analysis with OPLS-DA did not improve the separation (appendix 17.7.14, figure 17.69). The separation was not significant according to CV ANOVA (p 1.0) and permutation plots with 999 permutations (appendix 17.7.9, figure 17.25). Further analysis was therefore not performed.

15.11 UKELD Score

All participants with cirrhosis (including those with HCC that developed within a cirrhotic liver) were classified on the basis of their UKELD score and whether they met the threshold of >49 required for consideration of transplantation. An unsupervised PCA demonstrated no separation (appendix 17.7.15, figure 17.70). Supervised analysis with OPLS-DA did appear to demonstrate good separation (appendix 17.7.15, figure 17.71). However, the separation was not significant according to CV ANOVA (p1.0) and therefore further analysis was not performed.

15.12 Hepatic Encephalopathy

All cirrhotic and HCC participants with underlying cirrhosis were classified as to whether they had hepatic encephalopathy at the time of recruitment or not. An unsupervised PCA demonstrated no separation (appendix 17.7.16, figure 17.72). Supervised analysis with OPLS-DA (appendix 17.7.16, figure 17.73) improved the separation however this was not significant according to CV ANOVA (p1.0).

15.13 Long Term Antibiotics

All participants (control, cirrhosis, and HCC) were classified on whether they were taking long-term antibiotics at the time of recruitment or not. An unsupervised PCA demonstrated no separation (figure 15.23). Supervised analysis via OPLS-DA demonstrated some separation (figure 15.24). The separation was significant according to CV ANOVA testing with SIMCA (p 0.002) and permutation plots with 999 permutations (appendix 17.7.9, figure 17.27).

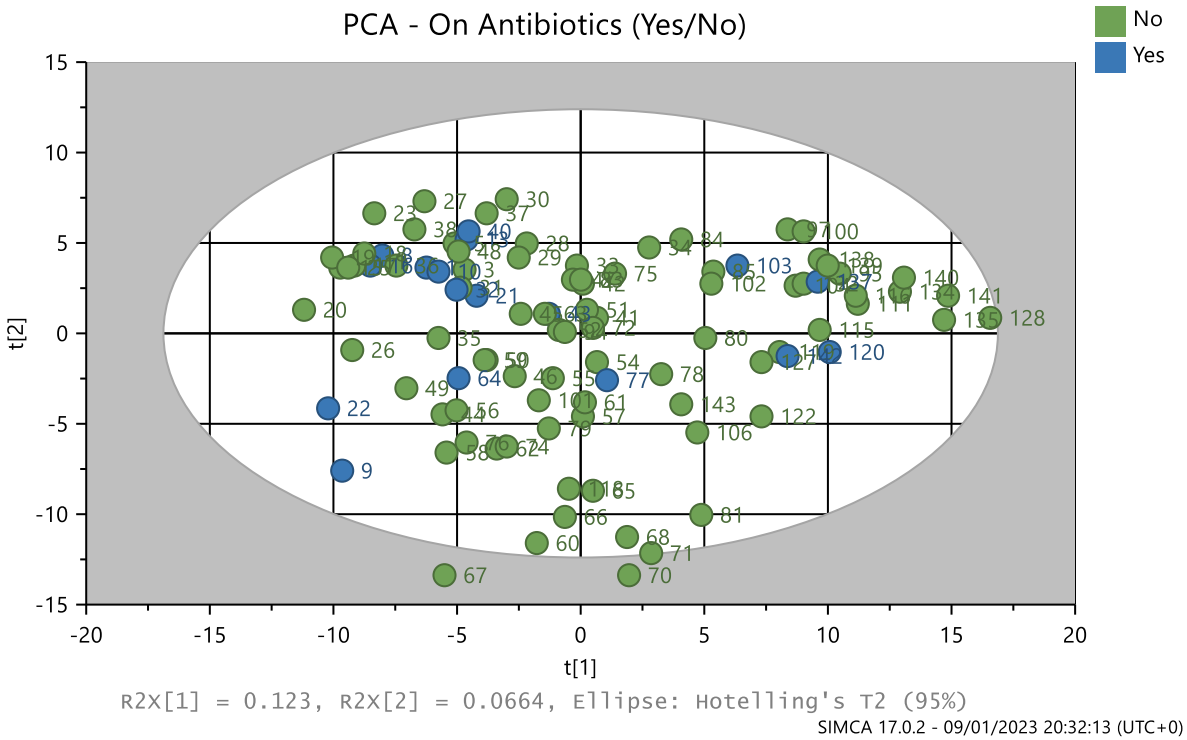


Figure 15.23 PCA of long-term antibiotic use of all participants demonstrating no separation between the two cohorts.

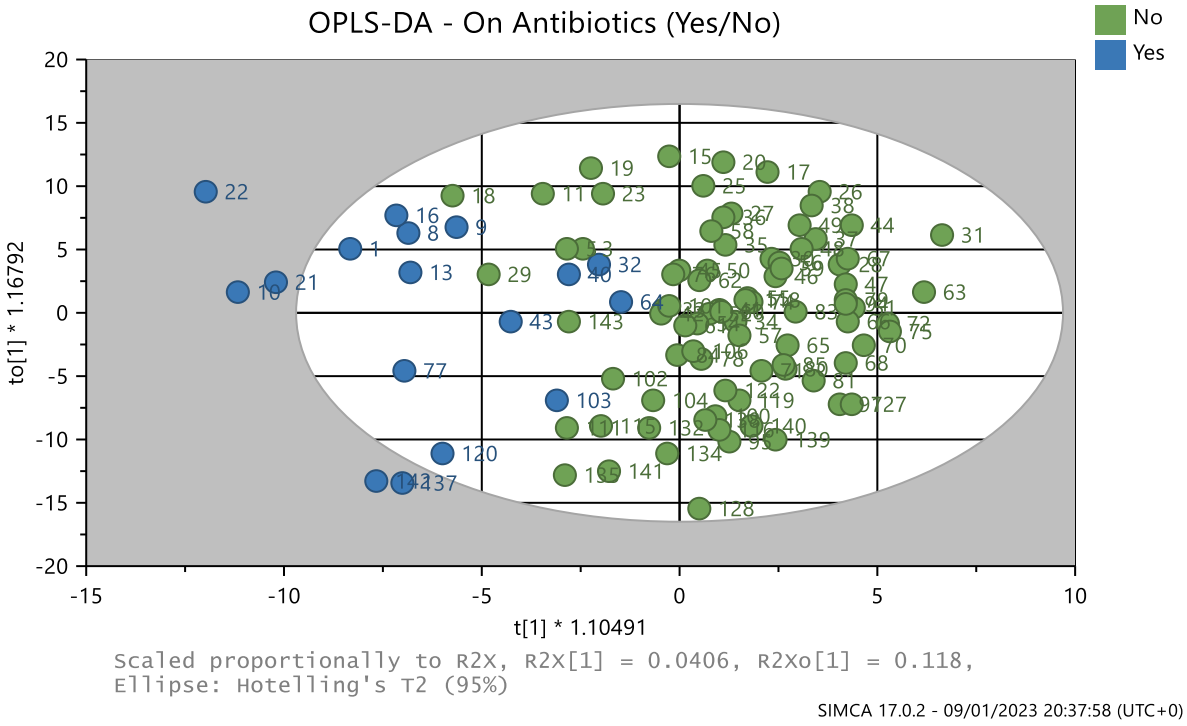


Figure 15.24 OPLS-DA of long-term antibiotic use of all participants demonstrating some separation between the two cohorts.

19 compounds with a VIP score >1.5 were significant after FDR correction (p <0.05) (table 15.7).

Table 15.7 FDR significant VOCs with VIP Score >1.5 driving separation of participants on the basis of whether they were taking long-term antibiotics or not at the time of recruitment.

Compound Identification	CAS Number	RT	FDR p Value	VIP Score	AUC
Limonene	138-86-3	12.91	<0.01	2.56	0.84
β-Pinene	127-91-3	12.05	<0.01	2.33	0.80
o-Xylene	95-47-6	3.40	<0.01	1.52	0.80
α-Terpinene	99-86-5	12.72	<0.01	1.82	0.79
2-Pentanone	107-87-9	6.63	<0.01	2.08	0.79
γ-Terpinene	99-85-4	13.38	0.01	1.51	0.79
Terpinolene	586-62-9	13.89	<0.01	1.73	0.79
m-Cymene	535-77-3	14.19	<0.01	1.58	0.78
α-Pinene	80-56-8	11.11	<0.01	2.00	0.77
p-Cymene	99-87-6	12.97	<0.01	2.05	0.77
1-Octanol, 2-butyl-	3913-02-8	19.90	0.01	1.58	0.69
Undecanal	112-44-7	17.69	<0.01	1.61	0.68
2-Butene	107-01-7	2.30	0.01	1.59	0.67
n-Hexane	110-54-3	4.01	0.01	1.83	0.67
Pentane, 2,2-dimethyl-	590-35-2	4.49	<0.01	1.88	0.66
Nonacosane	630-03-5	17.66	0.03	1.72	0.65
1-Hexene, 3,4-dimethyl-	16745-94-1	6.25	<0.01	1.77	0.65
Hexane, 2-methyl-	591-76-4	5.34	0.03	1.65	0.64
2-Butene, 2,3-dimethyl-	563-79-1	4.49	0.03	1.50	0.63

The top five discriminating compounds based on their individual AUCs were limonene, β-pinene, o-xylene, α-terpinene and 2-pentanone. Box plots of these compounds can be seen in appendix 17.7.11, figures 17.59-17.63. All five VOCs were higher in those taking long-term antibiotics compared to those that weren't. o-Xylene and α-terpinene were not detected in the room air samples and the other VOCs were of low abundance in the room air samples.

15.14 Sampling Location of HCC Participants

Participants with HCC were the only cohort to have samples collected from two different locations. All control participants were recruited from one location (SMH), as were participants with cirrhosis (SMH). HCC participants were collected from both sites, therefore, a comparison of site collected was performed for HCC participants only.

An unsupervised PCA demonstrated no separation (figure 15.31). Supervised analysis via OPLS-DA demonstrated clear separation (figure 15.32). The separation was significant according to CV ANOVA testing with SIMCA ($p<0.001$) but less convincing based on the associated permutation plot with 999 permutations (appendix 17.7.9, figure 17.26).

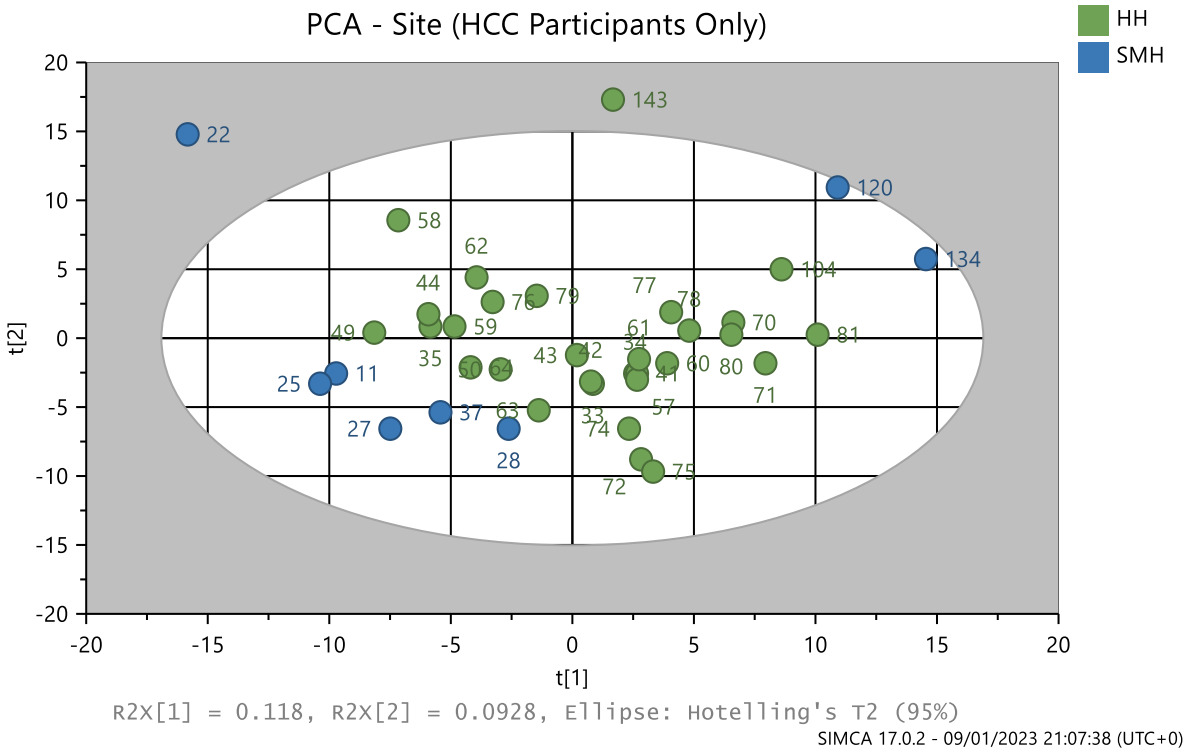


Figure 15.25 PCA of site of HCC participant recruitment (HH or SMH) demonstrating no separation between the two cohorts.

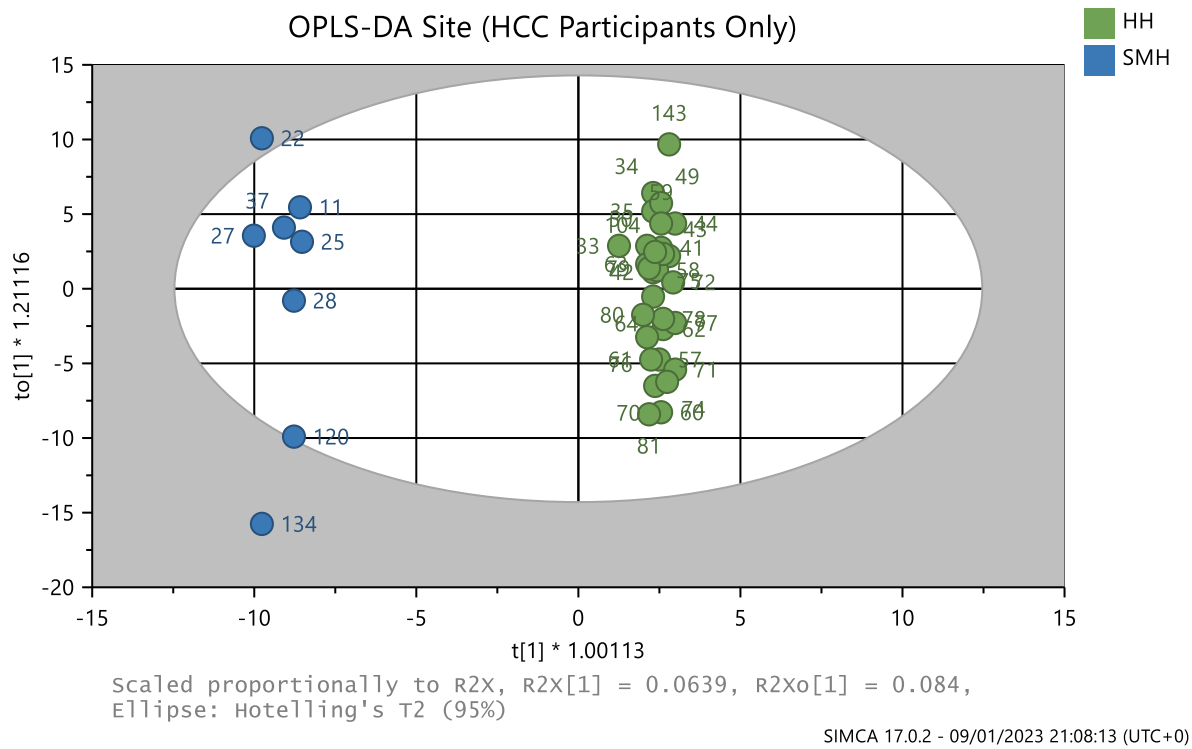


Figure 15.26 OPLS-DA of site of HCC participant recruitment (HH or SMH) demonstrating clear separation between the two cohorts.

21 compounds with a VIP score >1.5 were significant after FDR correction ($p < 0.05$) (table 15.8).

The top 5 based on their individual AUCs were octane, 2,6-dimethyl-, octane, 2,2-dimethyl-, 1-hexene, 3,4-dimethyl-, pentane, 2,2-dimethyl- and cyclohexane, 1,3-dimethyl-, cis-.

Table 15.8 FDR significant VOCs with VIP Score >1.5 of HCC recruitment site (SMH/HH).

Compound Identification	CAS Number	RT	FDR p Value	VIP Score	AUC
Octane, 2,6-dimethyl-	2051-30-1	13.93	<0.01	1.76	0.92
Octane, 2,2-dimethyl-	15869-87-1	12.04	<0.01	2.00	0.88
1-Hexene, 3,4-dimethyl-	16745-94-1	6.23	<0.01	2.46	0.87
Pentane, 2,2-dimethyl-	590-35-2	4.49	<0.01	2.35	0.87
Cyclohexane, 1,3-dimethyl-, cis-	638-04-0	9.01	<0.01	1.67	0.86
2-Butene, 2,3-dimethyl-	563-79-1	4.49	<0.01	1.63	0.83
Cyclohexane, ethyl-	1678-91-7	9.06	<0.01	1.57	0.82
Cyclohexane, methyl-	108-87-2	6.70	0.01	1.70	0.81
Heptane, 2-methyl-	592-27-8	7.43	0.01	1.85	0.80
1H-Indene, octahydro-	496-10-6	10.86	0.01	1.56	0.80
2-Hepten-1-ol, (E)-	33467-76-4	7.12	<0.01	1.75	0.79
3-Heptanone	106-35-4	10.66	0.02	1.61	0.78
1-Hexene, 4,5-dimethyl-	16106-59-5	3.50	0.02	2.12	0.77
2-Butene	107-01-7	2.30	0.02	1.68	0.77
Pentane, 3-methyl-	96-14-0	3.74	0.02	2.24	0.76
1-Octen-3-one	4312-99-6	3.52	0.03	1.63	0.76
n-Hexane	110-54-3	4.01	0.03	2.17	0.76
1-Heptene	592-76-7	5.90	0.03	1.65	0.75
Hexane, 3-methyl-	589-34-4	5.23	0.03	1.75	0.75
Pentane, 2,3-dimethyl-	565-59-3	5.44	0.03	1.81	0.75
Hexane, 2-methyl-	591-76-4	5.34	0.03	1.80	0.75

15.15 Batch Effect

TD tubes were analysed in different batches and therefore, in order to assess if there was any evidence of batch effect, the data sets for each processing batch were compared with and without normalisation.

An unsupervised PCA of the data without normalisation (figure 15.27) demonstrated some separation (R^2 0.512, Q^2 0.234). An unsupervised PCA of the data with normalisation (figure 15.28) demonstrated separation that was less strong (R^2 0.435, Q^2 0.265), suggesting that the data normalisation process has corrected any batch effect present to some degree. The normalised data also reassuringly reduced the number and severity of outlying samples within the data set.

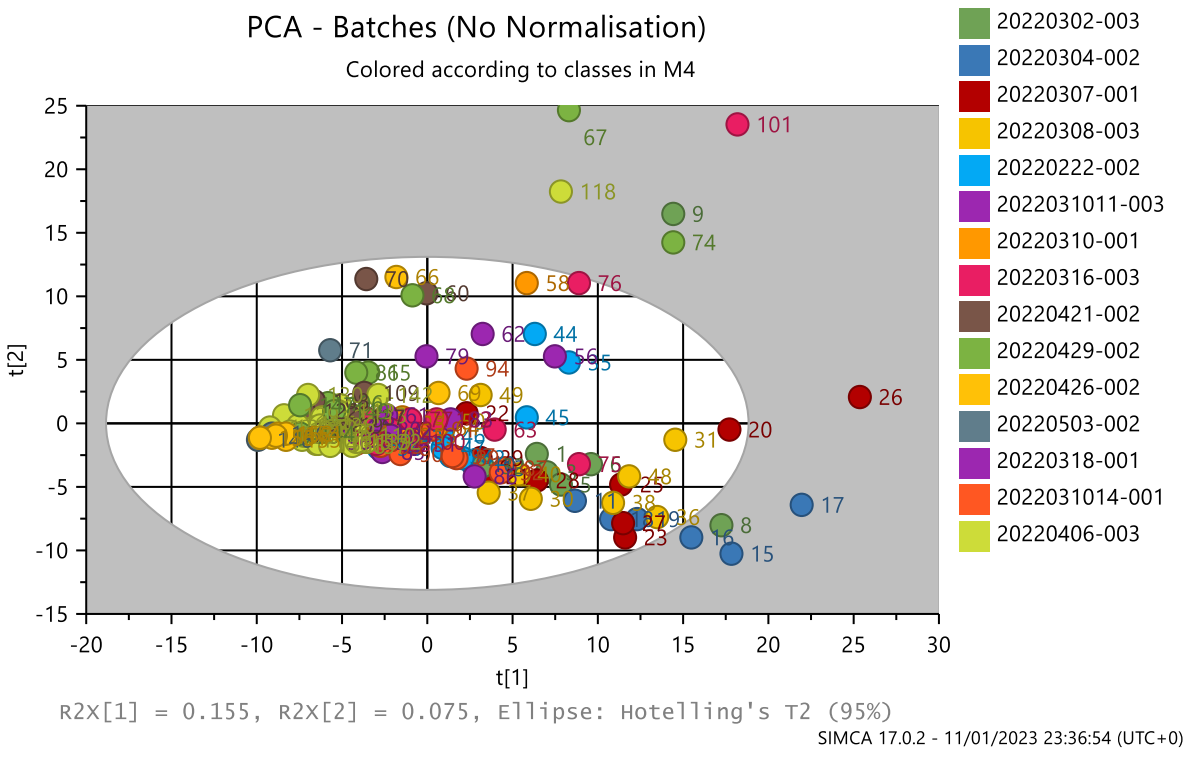


Figure 15.27 PCA of processing batches without normalisation of data.

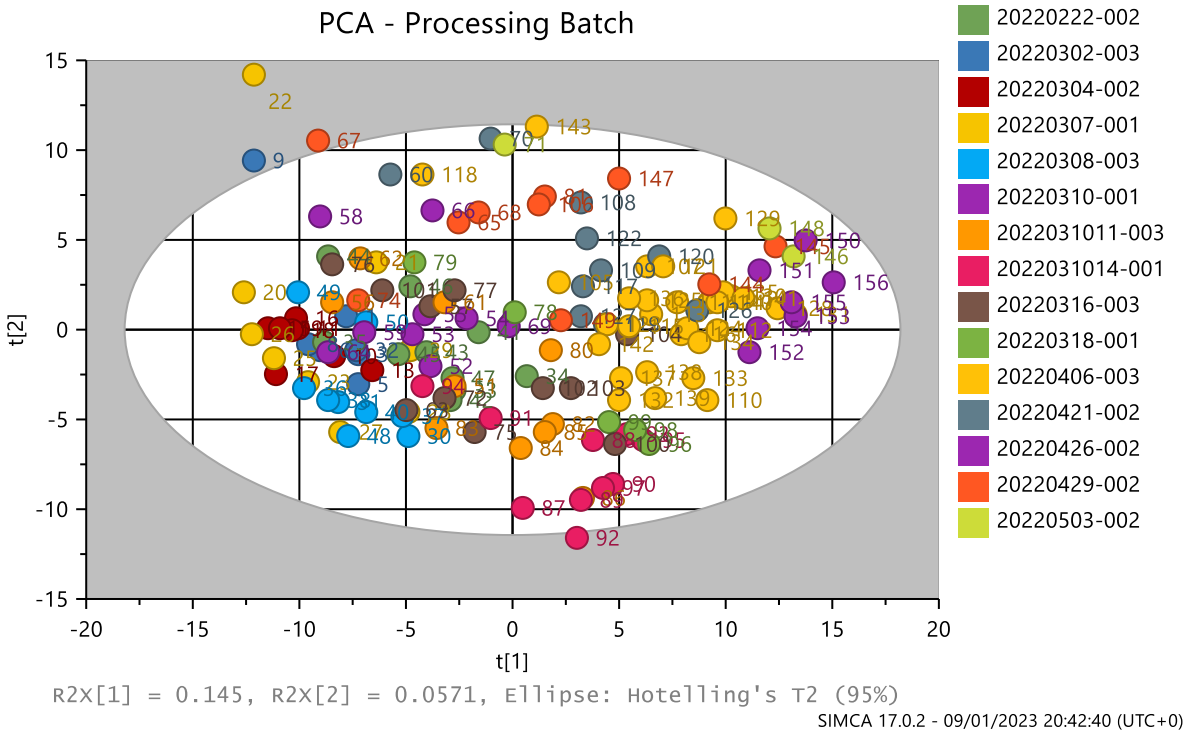


Figure 15.28 PCA of processing batches with normalisation of data.

15.16 Discussion

A summary of the top 5 compounds (based on AUC) for each comparison can be seen in table 15.9.

Table 15.9 Summary of Top 5 FDR Significant Compounds with VIP Scores >1.5 for each comparison.

	Compound ID	RT	Ion	VIP	AUC	CAS ID	Class	Comments	Possible Biomarker
Cirrhosis vs HCC	3-Heptanone	10.66	57	1.86	0.77	106-35-4	Ketone	Higher in cirrhotics compared to HCC. Previously found in human volatilome (219). Previously identified by Mochalski et al. (290) as being released from HCC cells and also identified as a possible breast cancer biomarker by Janfaza et al. (291) and as a potential biomarker in urine for colorectal cancer and leukaemia by Silva et al. (292). As a ketone, there is logical metabolic pathway which may account for the higher abundance in hepatopathology.	Yes
	1-Hexanol, 2,2-dimethyl-	12.75	57	1.59	0.70	2370-13-0	Alcohol	Higher in cirrhotics compared to HCC. Not previously identified in human volatilome (219) although other alcohols have been identified as potential hepatopathology biomarkers (197,293). There is no logical metabolic explanation for this VOC to be elevated in hepatopathology.	No
	Octane, 2,2-dimethyl-	12.04	57	1.64	0.70	15869-87-1	Alkane	Higher in cirrhotics compared to HCC. Not previously identified in human volatilome (219) although other alkanes including octane have been identified as potential hepatopathology biomarkers (203). Alkanes may be elevated due to increased lipid peroxidation in hepatopathology.	Yes
	1-Propanol, 3,3'-oxybis-	16.75	59	1.69	0.69	2396-61-4	Alcohol	Higher with in those with cirrhosis compared to those with HCC. Not previously identified in human volatilome (219) although other alcohols have been identified as potential hepatopathology biomarkers (197,293). There is no logical metabolic explanation for this VOC to be elevated in hepatopathology.	No
	Octane, 2,6-dimethyl-	13.92	57	1.72	0.72	2051-30-1	Alkane	Higher with in those with cirrhosis compared to those with HCC. Previously found in human volatilome (219) in breath but not attributed to a specific disease state. Alkanes may be elevated due to increased lipid peroxidation in hepatopathology.	Yes

Cirrhosis vs Control	m-Cymene	14.19	132	1.88	0.86	535-77-3	Aromatic Hydrocarbon	Higher in cirrhotics compared to controls. Previously found in human volatilome (219) and also as a potential lung cancer biomarker by Peng et al. (294) and disease biomarker for ulcerative colitis by Rondanelli et al. (274). There is no established metabolic pathway at present which would account for a higher abundance in hepatopathology.	No
	α-Terpinene	12.72	121	1.62	0.81	99-86-5	Monoterpene	Higher in cirrhotics compared to controls. Previously found in human volatilome (219) but more commonly in urine than in breath. Identified as a possible cancer biomarker by Silva et al. (292). As a monoterpene that is metabolised by hepatic CP450, there is a logical metabolic explanation for a higher abundance in those with hepatopathology.	Yes
	Limonene	12.90	67	1.58	0.81	138-86-3	Monoterpene	Higher in cirrhotics compared to controls. Previously found in human volatilome (219). Extensively found in literature previously as potential biomarker for hepatopathology (177,194,200,209). Can be explained by reduced CP450 enzyme activity and thus reduced metabolism in those with hepatopathology.	Yes
	Terpinolene	13.89	121	1.58	0.81	586-62-9	Monoterpene	Higher in cirrhotics compared to controls. Previously found in human volatilome (219) but more commonly in urine and faeces than breath. Found by Ahmed et al. to be a possible biomarker for inflammatory bowel disease (295). As a monoterpene that is metabolised by hepatic CP450, there is a logical metabolic explanation for a higher abundance in those with hepatopathology.	Yes
	Cyclopentane, methyl-	4.65	56	1.94	0.80	96-37-7	Alkane	Higher in cirrhotics compared to controls. Previously found in human volatilome (219) and also as a potential biomarker for colorectal cancer by Altomare et al. (296) and lung cancer by D'Amico et al. (297). Alkanes may be elevated due to	Yes

								increased lipid peroxidation in the context of cirrhosis.	
HCC vs Control	1-Octen-3-one	3.53	55	1.73	0.93	4312-99-6	Ketone	Higher in HCC compared to controls. Previously found in human volatilome but more commonly in breast milk VOCs and not in breath (219). Possibly related to food consumption (298).	No
	Cyclopentane, methyl-	4.65	56	2.14	0.92	96-37-7	Alkane	Higher in HCC compared to controls. Previously found in human volatilome (219) and also as a potential biomarker for colorectal cancer by Altomare et al. (296) and lung cancer by D'Amico et al. (297). Alkanes may be elevated due to increased lipid peroxidation in the context of HCC, although it is unclear if this would be specific enough to be useful as a biomarker.	Yes
	Cyclopentane	6.17	42	2.11	0.92	287-92-3	Alkane	Higher in HCC compared to controls. Previously found in human volatilome (219) and also as a potential marker of oxidative stress and lung cancer by Rudnicka et al. (286) and as a breast cancer biomarker by Barash et al. (299). Alkanes may be elevated due to increased lipid peroxidation in the context of HCC, although it is unclear if this would be specific enough to be useful as a biomarker.	Yes
	Pentane, 2,3-dimethyl-	5.45	56	1.63	0.91	565-59-3	Alkane	Higher in HCC compared to controls. Previously found in human volatilome (219) in breath and identified as possible biomarker for early gastric cancer by Chen et al. (300) and for breast cancer subtype by Barash et al. (299). Alkanes may be elevated due to increased lipid peroxidation in the context of HCC, although it is unclear if this would be specific enough to be useful as a biomarker.	Yes
	Hexane, 2,4,4-trimethyl-	5.44	57	2.26	0.90	16747-26-5	Alkane	Higher in HCC compared to controls. Previously found in human volatilome (219) in breath and identified as possible biomarker for Crohn's disease by Bodelier et al. (301). Alkanes may be elevated due to increased lipid peroxidation in the context of	Yes

								HCC.	
Child-Pugh A vs B+C	3-Carene	12.54	93	2.71	0.85	13466-78-9	Monoterpene	Higher in Child-Pugh B+C cirrhotics compared to Child-Pugh A cirrhotics. Previously found in human volatilome (219) and identified as possible biomarker for COPD by Gaida et al. (302). As a monoterpene that is metabolised by hepatic CP450, there is a logical metabolic explanation for a higher abundance in those with more advanced liver disease.	Yes
	Limonene	12.90	67	2.00	0.78	138-86-3	Monoterpene	Higher in Child-Pugh B+C cirrhotics compared to Child-Pugh A cirrhotics. Previously found in human volatilome (219). Extensively found in literature previously as potential biomarker for hepatopathology (177,194,200,209). Can be explained by reduced CP450 enzyme activity and thus reduced metabolism in those with hepatopathology.	Yes
	2-Pentanone	6.63	43	2.17	0.75	107-87-9	Ketone	Higher in Child-Pugh B+C cirrhotics compared to Child-Pugh A cirrhotics. Extensively found in human volatilome (219) in breath including as potential biomarker for hepatopathology by multiple studies (202,203,206) as well as lung cancer (303-306). As a ketone, there is logical metabolic pathway which may account for the higher abundance in more advanced liver disease.	Yes
	Dodecanal	19.11	41	1.78	0.71	112-54-9	Aldehyde	Higher in Child-Pugh A cirrhotics compared to Child-Pugh B+C cirrhotics. Extensively found in human volatilome (219) in breath but not attributed to a specific disease state. Blanchet et al. have suggested concentrations decrease with age (70). As it is also present in high abundance in background room air it is unlikely to be a useful biomarker.	No
	Pentadecane	19.99	71	1.57	0.69	629-62-9	Alkane	Higher in Child-Pugh A cirrhotics compared to Child-Pugh B+C cirrhotics. Previously found in human volatilome (219) in breath and suggested as possible biomarker for	Yes

								oxidative stress by Jalali et al. (307). Alkanes may be elevated due to increased lipid peroxidation in the context of those with more advanced liver disease.	
Portal Hypertension	o-Xylene	13.38	119	1.76	0.80	933-98-2	Aromatic Hydrocarbon	Higher in those with portal hypertension compared to those without. Extensively found in human volatilome (219) and suggested as possible biomarker for lung cancer by Rudnicka et al. (303). Has also been associated with smoking and found in background room air so should be interpreted with caution.	Yes
	Limonene	12.90	67	2.34	0.79	138-86-3	Monoterpene	Higher in those with portal hypertension compared to those without. Previously found in human volatilome (219). Extensively found in literature previously as potential biomarker for hepatopathology (177,194,200,209). Can be explained by reduced CP450 enzyme activity and thus reduced metabolism in those with portal hypertension secondary to hepatopathology.	Yes
	3-Hexanone, 4,4-dimethyl-	5.20	43	2.03	0.79	19550-14-2	Ketone	Higher in those with portal hypertension compared to those without. Not previously identified in human volatilome (219). Origin remains unclear from literature search. While ketones have a logical metabolic explanation for being elevated in hepatopathology, given the paucity of history of literature of this VOC, it seems unlikely to be clinically useful.	No
	p-Cymene	12.97	119	1.95	0.78	99-87-6	Alkylbenzene	Higher in those with portal hypertension compared to those without. Previously found in human volatilome (219) including breath. Urinary p-Cymene has been associated with pancreatic cancer by Wen et al. (169) and with colorectal cancer, lymphoma, and leukaemia by Silva et al. (292). Possibly related to smoking behaviour.	Yes
	β-Pinene	12.06	93	2.05	0.78	127-91-3	Monoterpene	Higher in those with portal hypertension compared to those without. Extensively found in human volatilome (219) and has been linked to smoking and COPD by Gaida et al. (302). Can be explained by reduced CP450 enzyme activity and thus reduced	Yes

								metabolism in those with portal hypertension secondary to hepatopathology.	
Long-term Antibiotics	Limonene	12.90	67	2.56	0.84	138-86-3	Monoterpene	Higher in those on long-term antibiotics than those not on long-term antibiotics. Previously found in human volatilome (219). Extensively found in literature previously as potential biomarker for hepatopathology (177,194,200,209).	N/A
	β-Pinene	12.05	93	2.33	0.80	127-91-3	Monoterpene	Higher in those on long-term antibiotics than those not on long-term antibiotics. Extensively found in human volatilome (219) and has been linked to smoking and COPD by Gaida et al. (302).	N/A
	o-Xylene	13.38	119	1.52	0.80	933-98-2	Aromatic Hydrocarbon	Higher in those with portal hypertension compared to those without. Extensively found in human volatilome (219) and suggested as possible biomarker for lung cancer by Rudnicka et al. (303). Possibly related to smoking (114).	N/A
	α-Terpinene	13.89	121	1.82	0.79	586-62-9	Monoterpene	Higher in those on long-term antibiotics than those not on long-term antibiotics. Previously found in human volatilome (219) but more commonly in urine than in breath. Identified as a possible cancer biomarker by Silva et al. (292)	N/A
	2-Pentanone	6.63	43	2.08	0.79	107-87-9	Ketone	Higher in those on long-term antibiotics than those not on long-term antibiotics. Extensively found in human volatilome (219) in breath including as potential biomarker for hepatopathology by multiple studies (202,203,206) as well as lung cancer (303-306).	N/A

15.16.1 Cirrhosis vs HCC

3-Heptanone was identified in my analysis as the VOC with the highest accuracy in discriminating between HCC and cirrhosis (AUC 0.77). It is a ketone that has been previously identified in the human volatilome (219) in breath and from the headspace of HCC cells by Mochalski et al. (290). In my analysis, 3-heptanone was lower in HCC participants compared to cirrhotics. While 3-heptanone has not been identified as a possible biomarker for hepatopathology previously, comparable ketones (e.g. 2-pentanone) have and there is a logical metabolic pathway which may underpin these changes e.g. insulin resistance and increased ketosis. However, this would not necessarily account for changes between those with cirrhosis and those with HCC. Should HCC cells produce 3-heptanone as suggested by Mochalski et al. (290), one would expect a higher abundance of 3-heptanone in individuals with HCC rather than lower. It is more likely that differences in 3-heptanone between the two cohorts are due to differences in the underlying severity of cirrhosis between the two cohorts rather than being driven by the presence or absence of HCC.

1-Hexanol, 2,2-dimethyl- was the VOC with the second highest AUC (0.70). It is higher in cirrhotic participants compared to those with HCC. It has not previously been identified in human volatilome (219) although other alcohols have been identified as potential hepatopathology biomarkers (197,293). Similarly, 1-propanol, 3,3'-oxybis-, is an alcohol not previously identified in the human volatilome (219). It has an AUC of 0.69 for differentiating between the cirrhosis and HCC cohorts. For both of these VOCs, there is no established metabolic pathway which would explain different levels in those with HCC and those with cirrhosis. Additionally, given that they have not previously been found in the human volatilome, they should be interpreted with caution and their identifications confirmed with chemical standards before they can be considered biomarkers for the presence or absence of HCC.

Octane, 2,2-dimethyl- was the VOC with the third highest AUC (0.70) and octane, 2,6-dimethyl- had the fifth highest AUC (0.67). Both were found to be in higher abundance in cirrhotic participants compared to HCC participants. The former has not previously been identified in human volatilome (219) although the latter has (although not attributed to any specific pathology). Other alkanes (e.g., octane) have been identified as potential hepatopathology biomarkers (203) due to increased lipid peroxidation in liver disease. While there is a logical explanation as to why alkanes in general may be elevated in those with liver disease, there is currently no specific metabolic pathway that would account for why

these two VOCs would be lower in individuals with HCC compared to those with cirrhosis only.

While a pooled AUC of the top three VOCs (3-heptanone, 1-hexanol, 2,2-dimethyl- and octane, 2,2-dimethyl-) gave an AUC of 0.765, which is comparable, to the diagnostic accuracy of pre-existing and in-development tumour markers for HCC (table 1.6), there is a lack of underlying metabolic pathways to account for these changes. Additionally, it may be that the differences in VOC profiles are coming from differences in underlying degrees of cirrhosis rather than the presence or absence of an HCC itself. Therefore, at present, these results do not give confidence that exhaled VOCs can be used as a diagnostic tool for HCC.

15.16.2 Cirrhosis vs Controls

When comparing the cirrhosis cohort to controls, m-cymene was the VOC with the highest individual AUC (0.86). m-Cymene is a monoterpene-related aromatic hydrocarbon that has previously been identified in the human volatilome (219). It was higher in cirrhotics compared to non-cirrhotics. It has previously been suggested as a potential lung cancer biomarker by Peng et al. (294) and biomarker for ulcerative colitis by Rondanelli et al. (274). While not previously identified as a biomarker for hepatopathology, its relation to monoterpenes and its probable metabolism by hepatic CP450 enzymes (308) make it feasible as a biomarker for cirrhosis.

α-Terpinene is a monoterpene found in essential oils and food flavourings (309). It provided the second highest AUC for cirrhosis vs controls (0.81). As with other monoterpenes, α-terpinene was higher in cirrhotics compared to controls. It has previously been found in the human volatilome (219) but more commonly in urine than in breath. It was identified as a possible cancer biomarker by Silva et al. (292) but has not been linked with liver disease in literature to date. As a monoterpene, it is assumed it will be absorbed and metabolised in same manner as limonene (i.e., by hepatic cytochrome P450 enzymes) and therefore it is feasible that it could represent a biomarker for cirrhosis.

Limonene reassuringly appears in the top 3 compounds discriminating between cirrhosis and healthy controls with an AUC of 0.81. This is extremely reassuring as limonene has previously been extensively identified as a potential biomarker for hepatopathology (177,194,200,209) including when specifically comparing cirrhotics and non-cirrhotics. There is logical metabolic explanation for the higher abundance of limonene in cirrhotics. With reduction in hepatic CP450 enzyme activity in those with liver disease, there will be reduced

metabolism of limonene. This validates the work of previous studies on limonene as a biomarker for cirrhosis.

Terpinolene, a terpene that can be found in essential oils and food flavourings (310), was higher in cirrhotics compared to healthy controls with AUC 0.81. It has been found in human volatilome previously (219) but more commonly in urine and faeces than in breath. Faecal terpinolene was found by Ahmed et al. to be a possible biomarker for inflammatory bowel disease (295) but no studies to date have linked it to liver disease. As a terpene that is metabolised by hepatic cytochrome P450 enzymes (311), it is metabolically feasible that terpinolene be used as a biomarker for cirrhosis.

An AUC of 0.857 was generated with a model of m-cymene, limonene, α-terpinene and terpinolene for the comparison of cirrhosis vs healthy controls. This is comparable and improved on several potential biomarkers developed to distinguish between cirrhosis and fibrosis as seen in table 1.1. The results of my study both validate the findings of previous studies for limonene and suggest possible novel biomarkers with logical underlying metabolic pathways.

15.16.3 HCC vs Controls

When comparing those with HCC to healthy controls, 1-octen-3-one, a ketone, provided the strongest separation with an AUC of 0.93. 1-octen-3-one is higher in participants with HCC compared to controls and has previously found in human volatilome but more commonly in breast milk and not in breath (219). While there is a logical metabolic pathway to explain the higher abundance of ketones in those with liver disease, there is not one for HCC and without a robust metabolic explanation for HCC, differences in exhaled VOC profiles are likely to be driven by underlying cirrhosis rather than the HCC itself. Additionally, the fact that 1-octen-3-one has not previously been found in the human volatilome raises concerns for its validity as a biomarker.

Cyclopentane, methyl- and cyclopentane both gave an AUC of 0.92. Both VOCs also drove separation between cirrhotic and control participants. As there were three HCC participants without cirrhosis, it would suggest that the higher abundance of cyclopentane, methyl- in HCC participants is driven by the underlying cirrhosis rather than the HCC itself. Pentane, 2,3-dimethyl- and hexane, 2,4,4-trimethyl- were elevated in the HCC cohort with AUCs of 0.91 and 0.90 respectively. As alkanes, the higher abundance could be explained by increased lipid peroxidation, driven by the underlying cirrhosis. The abundance of pentane,

2,3-dimethyl- in room air, however, was comparable to some breath samples, suggesting this may be a contaminant VOC.

The pooled AUC of the top 5 compounds was 0.94. However, it remains likely that the signal driving the separation is coming from the cirrhosis present in the majority of those with HCC rather than the HCC itself. Further work with a larger cohort of individuals with HCCs arising in non-cirrhotic livers is required to validate candidate VOCs for HCC screening.

15.16.4 Child-Pugh A vs Child-Pugh B/C Cirrhosis

3-Carene, another monoterpene found in essential oils (226) , was higher in those with the more advanced Child-Pugh B+C cirrhosis compared to Child-Pugh A cirrhosis, with an AUC of 0.85. 3-Carene has previously been found in the human volatilome (219) and is metabolised by hepatic cytochrome P450 enzymes (227). It was also previously found to be elevated by Pijls et al. (197) in those with cirrhosis compared to chronic liver disease. It should also be noted however that while low in abundance in the limited room air samples collected during the study, it was a VOC of interest in my room air study. However, at present, it remains feasible as a biomarker for hepatopathology and may correlate with disease severity.

Limonene gave an AUC of 0.78 when distinguishing between Child-Pugh A and Child-Pugh B+C cirrhotics. It was higher in Child-Pugh B+C cirrhotics suggesting that as well as differentiating between healthy controls and cirrhotic participants, it could potentially play a role in assessing the severity of liver disease.

2-Pentanone, a ketone, was higher in Child-Pugh B+C cirrhotics compared to Child-Pugh A cirrhotics and gave an AUC of 0.75. 2-Pentanone has been extensively found in then human volatilome (219) in breath in previous studies including as potential biomarker for hepatopathology by multiple studies (202,203,206). There is a logical metabolic pathway to explain the higher abundance (i.e. insulin resistance states in cirrhosis and a consequent increase in ketosis) in liver disease and I also found it to be elevated in cirrhotics compared to participants with fibrosis in my preliminary study with SIFT-MS. This result validates the finding of previous studies (202,203,206), giving further confidence that 2-pentanone could be utilised as a biomarker for liver disease.

Dodecanal, an aldehyde, was higher in Child-Pugh A cirrhotics compared to Child-Pugh B+C cirrhotics with an AUC of 0.71. Dodecanal has been extensively found in human volatilome (219) in breath previously including as potential biomarker for hepatopathology by multiple

studies (202,203,206). However, there was also a significantly elevated abundance of dodecanal in the limited room air samples collected for my study. Therefore, it should not be considered a biomarker on the basis of this study and further work with additional, paired, room air samples is required to further investigate dodecanal's ability to discriminate between Child-Pugh A and Child-Pugh B/C cirrhotics.

A model of 3-carene, limonene and 2-pentanone only gave AUC of 0.841 (95% CI 0.729-0.924). This is comparable to several potential biomarkers developed to distinguish between cirrhosis and fibrosis as seen in table 1.1.

15.16.5 Portal Hypertension vs No Portal Hypertension

o-Xylene, an aromatic hydrocarbon found in dyes, paints, and polishes (312), was higher in those with portal hypertension compared to those without, with an AUC of 0.80. It has been extensively found in human volatilome previously (219). It has been suggested as possible biomarker for lung cancer by Rudnicka et al. (303) but not yet linked to liver disease. It should be noted that o-xylene was previously associated with smoking by Pauwels et al. (114) however inspection of o-xylene abundance between active smokers and non-smokers demonstrated higher average abundance of o-xylene in non-smokers (appendix 17.7.17, figure 17.88). It was also not detected in room air samples. o-Xylene undergoes metabolism in the liver (312) and therefore it is metabolically feasible that there would be an elevated abundance in those with impaired liver function. While at present there is no obvious connection to portal hypertension in itself, it is more likely to represent the underlying cirrhosis that is driving the portal hypertension in the majority of cases.

Limonene was higher in those with portal hypertension compared to those without with an AUC of 0.79. While it is possible that elevated limonene concentrations could be at least in part caused by portal hypertension, it is once again more likely to be driven by the underlying cirrhosis in those with cirrhosis driving the portal hypertension. β-Pinene, another monoterpene present in essential oils, was higher in those with portal hypertension compared to those without with an AUC of 0.78. It has been extensively found in human volatilome (219) previously and has been linked to smoking and COPD by Gaida et al. (302). A review of the abundance β-pinene however revealed a higher abundance in non-smokers compared to active smokers (appendix 17.7.17, figure 17.74). While β-pinene was previously identified in the room air of sampling locations used by my group (1), the abundance in the room air samples were low. β-Pinene is also metabolised by hepatic cytochrome P450 enzymes and were previously found to be elevated in cirrhosis by Dadamio et al. (203). β-Pinene may therefore represent a biomarker for hepatopathology

but once again is more likely elevated due to underlying cirrhosis rather than specifically the presence of portal hypertension. Similarly, p-cymene, a monoterpene-related alkylbenzene, was higher in those with portal hypertension compared to those without with an AUC of 0.78. It has previously been found in the human volatilome (219) including breath but also suggested as being related to smoking (116). When interrogating the abundance of p-cymene according to smoking status, there was a slight increase in p-cymene abundance in active smokers compared to non-smokers, but this was not statistically significant ($p > 0.05$) (appendix 17.7.17, figure 17.76). The abundance of p-cymene in room air was also low. Therefore, p-cymene has the potential to act as a biomarker for hepatopathology.

3-Hexanone, 4,4-dimethyl-, was higher in those with portal hypertension compared to those without with an AUC of 0.79. However, it has not previously been identified in human volatilome (219) and its origin remains unclear from literature search. Confirmation of its identity with chemical standards is required before it can be considered a biomarker for the present of portal hypertension or hepatopathology.

A panel of limonene, o-xylene, p-cymene and β-pinene gave an AUC of 0.728 (95% CI 0.661-0.83). A larger number of individuals with non-cirrhotic portal hypertension is required to provide more conclusive results on the direct impact of portal hypertension on VOC profile.

15.16.6 Compensated vs Decompensated Cirrhosis

There was no significant difference found between cirrhotics in compensated vs decompensated states. However, there was only 13 participants with decompensated cirrhotics compared to 93 compensated cirrhotics. This reflected the fact that decompensated cirrhotics are more unwell and less likely to be cared for in an outpatient setting, or in the case of HCC, attending outpatient clinics for consideration of systemic treatment. The analysis was therefore likely underpowered and further work is required to establish if there are any biomarker VOCs associated with compensated or decompensated cirrhotic states.

15.16.7 UKELD Score

There were no significant differences identified between cirrhotics with a UKELD score of greater than or less than 49. This is in contrast to the Child-Pugh scoring system where differences were demonstrated between A and B+C cirrhotics. Further work is required as

to why breath may correlate with the Child-Pugh classification of cirrhosis but not the UKELD score.

15.16.8 HCC Barcelona Clinic Classification

There were no significant differences identified between BCLC stage A, B and C hepatocellular carcinomas. However, this analysis is likely to be underpowered due to small numbers of individuals with HCC and would warrant being repeated with a larger number of participants.

15.16.9 Hepatic Encephalopathy vs No Hepatic Encephalopathy

In those participants with established liver disease, the presence or absence of hepatic encephalopathy was assessed, and no significant separation was found between the two cohorts. This is likely due to the small number of participants who were felt to have active hepatic encephalopathy at the time of recruitment. Early-stage grade I/II hepatic encephalopathy can be difficult to detect without an EEG. It requires time consuming PHES testing (9) and while this was planned as part of the study protocol, limitations put in place by COVID-19 restrictions meant this was not feasible for many participants. Therefore there is a possibility that individuals with subclinical hepatic encephalopathy were missed in the analysis. Further dedicated work in profiling the volatilome of patients with and without hepatic encephalopathy is therefore justified.

15.16.10 Cirrhosis Aetiology

Participants with cirrhosis and HCC were recruited without discrimination as to the underlying aetiology of their liver disease. There was no significant separation according to my analysis which suggests that a breath test could potentially be used universally across liver diseases of all aetiologies. However, several aetiologies were only represented by a handful of participants (e.g., 6 participants with cirrhosis secondary to autoimmune conditions and 3 participants with thrombosis related cirrhosis) and therefore further work with additional patients is required to validate this.

15.16.11 Long-Term Antibiotic Use

Participants on long-term antibiotics (typically rifaximin for hepatic encephalopathy (289) or ciprofloxacin as secondary prevention for spontaneous bacterial peritonitis (313)) had higher concentrations of limonene, β-pinene, o-xylene, α-terpinene and 2-pentanone driving separation. With the exception of one control who took phenoxymethylpenicillin for post-

splenectomy prophylaxis, all participants who were on long term antibiotics were in the cirrhosis or HCC cohorts. While antibiotics themselves may cause alteration in an individual's volatilome, those on antibiotics will have more severe liver disease. In order to have hepatic encephalopathy or spontaneous bacterial peritonitis, you generally must have Child-Pugh B/C cirrhosis or portal hypertension. Given the overlap in compounds between this comparison and the portal hypertension analysis, it is likely that the separation in this comparison is driven by liver disease severity rather than the antibiotics themselves.

15.16.12 Sampling Location of HCC Participants

Comparison of HCC participants between sites recruited demonstrated some separation, predominantly driven by octane, 2,6-dimethyl-, octane, 2,2-dimethyl and 1-hexene, 3,4-dimethyl. Octane, 2,6-dimethyl- drove some separation between HCC and all participants without HCC, however its origin remains unclear. Octane, 2,2-dimethyl also drove some separation between cirrhosis and HCC participants. While octane, 2,6-dimethyl- was not detected and octane, 2,2-dimethyl was low in abundance in the limited room air samples collected from St. Mary's Hospital, the lack of room air samples collected from Hammersmith Hospital means these VOCs cannot be excluded as background contaminants. The limitations of this are discussed further in section 15.19.

15.17 Recurrent Themes

Monoterpenes

Monoterpenes were repeatedly identified as differentiating between hepatopathology and controls. This validates the work of previous studies which have previously identified elevated levels of limonene in liver disease (177,194,200,209). Additionally, my findings of elevated 3-carene in more advanced cirrhosis and elevated β-pinene in those with portal hypertension, validate the findings of Pijls et al. (197) who identified higher levels of 3-carene in cirrhotics and Dadamio et al. (203) who found elevated β-pinene in cirrhotics.

The discovery of elevated α-terpinene, and the monoterpene-related p-cymene in those with hepatopathology is novel and raise the possibility of other monoterpenes and related VOCs being able to act as biomarkers for liver disease. Terpenes are generally exogenous compounds and key components of essential oils. They are used commonly as food flavourings and fragrances. As discussed previously, cytochrome P450 enzymes, present within the liver, metabolise terpenes and monoterpenes (225) and thus elevated concentrations of terpenes and monoterpenes within the breath of those with liver disease may be explained in part by reduced cytochrome P450 activity in the diseased liver.

Ketones

While several ketones were identified as potential biomarkers for liver disease including 3-heptanone and 1-octen-3-one, 2-pentanone was the only one which had been identified in previous literature (177,202-204,206) and also by myself in my initial SIFT-MS study. This gives confidence in its ability to act as a biomarker for liver disease. It has been previously hypothesised that elevated ketones can be at least in part attributed to ketosis from insulin resistance states (185), often present in those with liver disease especially when the underlying aetiology is hepatic steatosis.

Alkanes

While multiple alkanes including octane, 2,2-dimethyl and cyclopentane were found to differentiate between hepatopathology and controls, none of them correlate with the pre-existing literature. The underlying metabolic processes (lipid peroxidation) potentially driving the increased abundance are also not specific to hepatopathology. The alkanes in this study are unlikely to represent clinically useful biomarkers for liver disease. A summary of the analysis and associated AUCs for the models generated can be seen in table 15.10.

Table 15.10. Summary of all comparisons and summary AUC.

Comparison	**VOC Model**	**AUC**
Cirrhosis vs HCC	3-Heptanone 1-Hexanol, 2,2-dimethyl- Octane, 2,2-dimethyl	0.777
Cirrhosis vs Healthy Controls	m-Cymene α-Terpinene Limonene Terpinolene	0.857
HCC vs Healthy Controls	1-Octen-3-one Cyclopentane, methyl-, Cyclopentane Pentane, 2, 3-dimethyl- Hexane, 2,4,4-trimethyl	0.940
Portal Hypertension vs No Portal Hypertension	o-Xylene Limonene p-Cymene β-Pinene	0.728
Child-Pugh A Cirrhosis vs Child-Pugh B+C Cirrhosis	3-Carene Limonene 2-Pentanone	0.841

15.18 Study Limitations

The VOCAL study was impacted by two major events. Firstly, the COVID-19 pandemic caused a recruitment delay of 11 months from April 2020 until March 2021. It also resulted in a significantly reduced pool of potential participants due to social distancing guidelines and residual restrictions to on site clinical activities. Individuals had to be recruited in alignment with pre-existing clinical commitments and could not attend hospital purely for the study. This reduced the amount of standardisation processes that could be put in place (e.g., periods of fasting, restrictions to smoking). Social distancing requirements and maximum numbers of individuals within the department where recruitment occurred also meant reduced time with participants. Instead of 10 minutes of rest prior to sampling, this was reduced to 5 minutes and PHES testing for hepatic encephalopathy was not possible within the time restrictions. It also meant that potential participants attended appointments by themselves and therefore partners or relatives could not be used as controls. The control cohort is therefore not well matched, especially for age. Additionally, due to restrictions on on-site activities within Imperial College London, there was also a reduction in mass spectrometer workflow due to delays in servicing of instruments. There were also additional delays in supply chains for replacement parts.

The second major event was the move of the Imperial College VOC laboratory. Between October 2021 and March 2022, the laboratory was moved from St. Mary's Hospital to Hammersmith Hospital. The move had been planned to commence earlier in Summer 2021 but was subject to construction delays. Prior to the move itself, there was a period of time where the instruments were decommissioned in preparation for the move. Due to concerns about batch effects, the decision was made to wait until the move was complete before running any samples, so that all samples could be run in the new laboratory. This meant that minimal TD tubes could be processed between September 2021 and February 2022. The consequences of this is twofold. Firstly, there was a large backlog of TD tubes stored in freezers which resulted in a profound and persistent shortages of TD tubes during the majority of my recruitment window. This resulted in only 2 tubes being collected per participant. Room air sampling was also reluctantly sacrificed in order to prioritise participant recruitment. Secondly, the TD tubes from the VOCAL study were stored for much longer than intended and it is acknowledged this may have resulted in VOC degradation or contamination during the storage process.

Control participants were collected later in the study once issues with instrument workflow were resolved and there was a more robust supply of TD tubes and thus control participants

had their tubes stored at -80°C for a significantly shorter time (average 21.5 days, range 0-343 days) compared to participants with cirrhosis (average 114.1 days, range 2-336 days) and participants with HCC (average 101.5 days, range 0-315 days). There was no statistically significant difference between the length of time that cirrhosis and HCC participants tubes were stored at -80°C. When the sum of all peak areas for each sample were plotted against the duration of storage (figure appendix 17.7.18, figure 17.77), the sum of all peak areas were generally higher in those with longer storage times. While definitive conclusions cannot be drawn, it is supportive that there was not significant degradation over time although contamination may be an issue. As individuals with pathology (i.e., cirrhosis or HCC) had the longest range of storage time, VOCs of interests were individually inspected for those with pathology only. Abundance of limonene did not appear to correlate with the duration of storage (appendix 17.7.18, figure 17.78). Those tubes with the longest storage duration had a comparable distribution of limonene concentrations to those with a shorter storage duration. While contamination may be an issue, limonene has not been identified as a contaminant VOC in previous storage studies. Analysis of the room air in the storage areas, however, is required to interrogate the possibility of contamination driving the increase in VOCs. Comparable results were also found for 2-pentanone, β-pinene and 3-carene (appendix 17.7.18, figures 17.79-17.81).

Additionally, medications were not controlled for. While antibiotic usage appears to be associated with disease severity, I was not able to control for all the other medications that individuals were taking. It is acknowledged that those on medications that will inhibit or induce cytochrome P450 enzymes may impact upon the results presented.

15.19 Conclusions

Within the limitations of the study, I have validated the results of several previous studies by demonstrating an increased abundance of limonene, 3-carene and β-pinene in several analyses of hepatopathology. Limonene was able to discriminate between those with cirrhosis and healthy controls, as well as those with the more advanced Child-Pugh B/C cirrhosis. Limonene is a strong candidate biomarker for cirrhosis given its exogenous nature and logical metabolic pathway and my work adds further weight to the existing body of evidence (177,194,200,209) for its use as a biomarker for hepatopathology.

β-Pinene, along with limonene, was elevated in those with portal hypertension compared to those without. However, this is likely driven by the cirrhosis driving the portal hypertension in majority of participants. Further work is required with a larger cohort of individuals with non-cirrhotic portal hypertension to determine if any changes in VOCs are driven by portal

hypertensive states. Similarly, β-pinene drove separation between those on long-term antibiotics and those not. The use of long-term antibiotics is likely surrogate for more advanced cirrhosis in this setting. Assuming the changes in β-pinene are driven by cirrhosis, this validates the work of Dadamio et al. (203) who have previously found elevated β-pinene in cirrhotics.

3-Carene was able to discriminate between Child-Pugh A and Child-Pugh B/C cirrhotics, suggesting that this could be used to assess severity of cirrhosis. This validates the work of Pijls et al. (197) who found elevated levels in cirrhosis compared to non-cirrhotic chronic liver disease. The same was found for 2-pentanone, which correlated with pre-existing literature that has demonstrated elevated levels in previous literature (177,202-204,206).

While the strength of my studies lies in the validation of previous studies, it has also generated several new VOCs of interest which have a viable metabolic basis including additional monoterpenes (α-terpinene and terpinolene) and monoterpene-related VOCs (m-cymene and xylene). Additional studies are required to validate these VOCs as possible biomarkers.

My study was unsuccessful in convincingly discriminating between participants with HCC and cirrhosis it remains likely that differences in individuals with HCC and controls are driven by the underlying cirrhosis which caused the HCC. Further work with larger number HCCs arising from non-cirrhotics livers would be helpful to determine if there is a specific breath-print from HCCs that could be manipulated for use in screening in at risk questions. Additionally, at present, VOCs do not appear to be able to discriminate between cirrhosis aetiology or HCC stage.

SECTION 4: CONCLUSION

16 SUMMARY, CONCLUSIONS, AND FUTURE WORK

16.1 Thesis Summary

My PhD began with a critical review of the existing literature on the use of volatile organic compounds in the diagnosis of all type liver disease including hepatocellular carcinoma. I also established current methods for detecting VOCs in exhaled breath and highlighted the barriers currently faced both in terms of standardisation of breath testing and equipment availability.

The questions raised from the literature review were addressed in the following experiments and clinical studies. I first performed method development work, initially on the development of a novel breath collection device. Through testing of a prototype device, I detected instrument flaws and worked with the designer to correct these issues and improve the design. I ascertained the optimum settings for the device and devised a standard operating procedure for use of the device with my group. Following this, I designed and performed a study to determine profile the background VOCs within ambient room air of the sampling locations used by my group and the potential impact on the breath sampling process. Through this experiment I was able to profile the background VOC concentrations in multiple sampling locations used by my group and identified that while an individual's overall "breath-print" does not appear to alter with location, there are VOCs of interest present in the background room air and therefore monitoring of background volatiles is an essential part of quality control and standardisation. For the final part of method development work, I designed and performed an experiment to test the feasibility and optimum strategy for recollection of breath samples from TD tubes to allow analysis on multiple complementary mass spectrometry techniques. I found that recollecting two TD tubes on to one TD tube gave the best discriminatory performance and this has since been incorporated into standard practice within my group.

Following on from methodological work, I undertook a preliminary clinical study with SIFT-MS, performing a targeted analysis of VOCs identified from my literature review as being significantly different in hepatopathology. I demonstrated significant differences in several VOCs, validating the work of previous studies including limonene and 2-pentanone. Following COVID-related recruitment delays, I began recruitment for an untargeted study of VOCs with GC-MS in those with cirrhosis and HCC. Alongside potential novel biomarkers for liver disease, I once again validated the findings of previous studies with increased abundances of 2-pentanone, limonene and other monoterpenes in those with cirrhosis.

Limonene remains a strong candidate biomarker VOC for screening of cirrhosis, with an AUC of 0.81 in discriminating between cirrhosis and healthy controls. A model of four VOCs (m-cymene, limonene, α-terpinene and terpinolene) improved this with an AUC of 0.857. Limonene also has a well understood metabolic pathway which would account for an elevated abundance in liver disease and the same may be applicable for other exogenous monoterpenes.

Separation of HCC from cirrhosis participants was less successful. While separation was seen between HCC participants and healthy controls, this separation is likely driven by the underlying cirrhosis in the majority of HCC participants and further work with HCCs arising from non-cirrhotic livers would be helpful in identifying any HCC-specific VOCs.

The VOCAL study has added to the pre-existing body of work on exhaled VOCs for detection of liver disease. Out of all the VOC investigated, limonene remains the strongest biomarker candidate on the basis of my results, pre-existing literature and the established metabolic pathways which can account for the changes.

16.2 Future Work

The VOCAL GC-MS study should be considered a pilot study with a view to developing a larger body of work. The aim of this would be to validate the VOCs not already identified in the existing literature. Power calculations will be performed on the data set generated by the initial VOCAL study to provide a target sample size and the methodologies will be tailored to effectively quantify the markers found in my initial study.

A collaboration has already been setup with The Chinese University of Hong Kong to recruit further participants. 43 participants with cirrhosis have been recruited thus far with efforts ongoing to recruit further HCC participants alongside a cirrhosis control cohort. The tubes collected so far are currently stored at -80°C in Hong Kong and will be transported back to our mass spectrometry lab for analysis. A further collaboration has been set up with Liverpool University Hospitals NHS Foundation Trust and The Clatterbridge Cancer Centre for further recruitment within the United Kingdom. The aim is also to obtain external funding via grant or fellowship using the data from the VOCAL study in order to gain NIHR portfolio adoption and expand recruitment further. By bolstering our recruitment sites, it will allow greater volumes of participants recruited including a fibrosis cohort, additional individuals with non-cirrhotic portal hypertension and individuals with HCCs arising in a non-cirrhotic liver. This will also allow further interrogation of potential confounders and associated

hepatopathologies e.g., presence or absence of hepatic encephalopathy, disease progression and potential risk stratification for decompensation.

Further work is also planned to investigate the metabolic pathways that underpin changes in VOC concentrations between cohorts to give further confidence to any potential biomarkers. Should limonene continue to be identified as a significant biomarker for liver disease, the intention is also to perform a study investigating whether it is possible to perform dynamic testing using a limonene meal, as has also been suggested by Stavropolous et al. (185) and Murgia et al. (314). By taking a baseline measurement of exhaled breath limonene, administering a limonene meal, and then taking serial measurements of exhaled breath limonene over time, we may find a delta change that may provide a more accurate diagnostic test and a more dynamic study of liver function.

16.3 Conclusions

Within my PhD I have tested and developed a standard operating procedure for a novel breath collection device which is significantly more affordable when compared to existing breath collection devices and has the potential to facilitate expansion of breath analysis studies which were previously limited by cost. I profiled the background levels of VOCs within common sampling locations to help further understand their impact upon the breath sampling process and investigated the splitting and recollection of breath samples to allow more in-depth analysis of VOCs. Despite the limitations of the VOCAL study, I have successfully validated the several VOCs as biomarkers of hepatopathology identified by previous studies, namely limonene and 2-pentanone. I have identified further candidate biomarkers for liver disease which require validation.

The results give strength to the concept of using VOCs in exhaled breath as a tool to diagnose liver disease and justify further research into the field as detailed above. Breath testing remains an attractive proposition given that it is non-invasive, quick, and more acceptable to patients that blood or stool testing. Ongoing method development work continues to enhance the standardisation of breath analysis in the hope that a breath test can be developed to diagnose liver disease and hepatocellular carcinoma at earlier and more treatable stages and finally turn the tide on their rising mortality rates.

SECTION 5: APPENDIX

17 APPENDIX

17.1 STARD Checklist

STARD 2015 Checklist (315). Reproduced with permission granted by the Creative Commons Attribution Non-Commercial (CC BY-NC 4.0) license.

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	
	4	Study objectives and hypotheses	
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	
Participants	**6**	Eligibility criteria	
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	
	8	Where and when potentially eligible participants were identified (setting, location and dates)	
	9	Whether participants formed a consecutive, random or convenience series	
Test methods	**10a**	Index test, in sufficient detail to allow replication	
	10b	Reference standard, in sufficient detail to allow replication	
	11	Rationale for choosing the reference standard (if alternatives exist)	
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	
	15	How indeterminate index test or reference standard results were handled	
	16	How missing data on the index test and reference standard were handled	
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	
	18	Intended sample size and how it was determined	
RESULTS			
Participants	**19**	Flow of participants, using a diagram	
	20	Baseline demographic and clinical characteristics of participants	
	21a	Distribution of severity of disease in those with the target condition	
	21b	Distribution of alternative diagnoses in those without the target condition	
	22	Time interval and any clinical interventions between index test and reference standard	
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	
	25	Any adverse events from performing the index test or the reference standard	
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	
	27	Implications for practice, including the intended use and clinical role of the index test	
OTHER INFORMATION			
	28	Registration number and name of registry	
	29	Where the full study protocol can be accessed	
	30	Sources of funding and other support; role of funders	

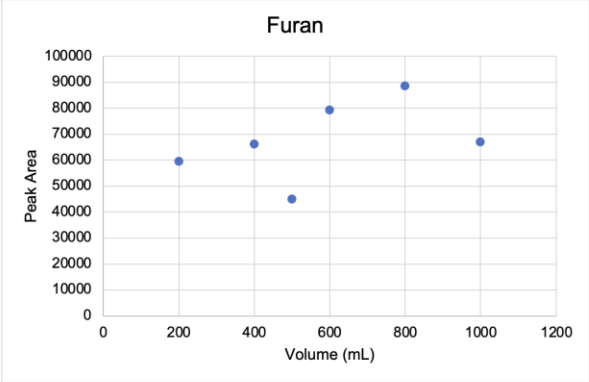
17.2 Development of a Novel Breath Collection Device for VOC Analysis

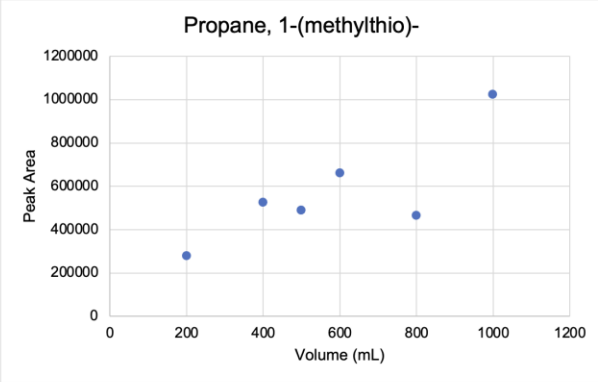
Table 17.1 VOCs and their p-values when comparing yield between flow rates of 100 mL/min and 200 mL/min.

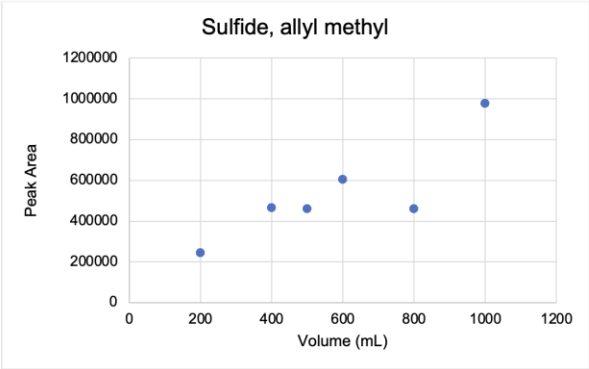
Provisional VOC Identification	2-Tailed T-Test p Value	Spearman Rank Corelation Coefficient
Isoprene	1.000	1.000**
Ethanethioic acid, S-(dihydro-2,5-dioxo-3-furanyl) ester	1.000	1.000**
Vigabatrin	0.386	1.000**
Dimethyl sulfide	1.000	1.000**
Furan	0.248	0.657
Trimethadione	0.386	1.000**
Propanedioic acid, oxo-, diethyl ester	0.386	.943**
2-Propenal	0.021	.886**
Methacrolein	0.149	1.000**
Furan, 2-methyl-	0.021	1.000**
Nonane	0.043	1.000**
Propane, 1-(methylthio)-	0.248	0.600
1-Hepten-3-one	0.021	.943**
Sulfide, allyl methyl	0.248	0.600
Dimethyl ether	0.021	.771*
Tridecane, 2,2,4,10,12,12-hexamethyl-7-(3,5,5-trimethylhexyl)-	0.564	1.000**
2-Pentanone	0.248	1.000**
1,2-Dimethoxy-ethene-A	0.248	.886**
Propanoic acid, 2-methylpropyl ester	0.773	1.000**
Acetone	0.758	0.486
Decane, 2-methyl-	0.773	1.000**
(1S)-2,6,6-Trimethylbicyclo[3.1.1]hept-2-ene	0.248	1.000**
Sulfone, 2-hydroxyoctyl t-butyl	0.386	1.000**
Toluene	0.248	.829*
Propanoic acid, anhydride-A	0.564	1.000**
1-Propanol	0.564	.771*
3,4-Hexanedione, 2,2,5-trimethyl-A	0.149	.943**
Peroxide, bis(1-methylethyl)	0.773	0.657
β-Myrcene	0.149	1.000**
Octane, 1-iodo-	0.021	.943**
Thiophene, 3-methyl-	0.773	0.600

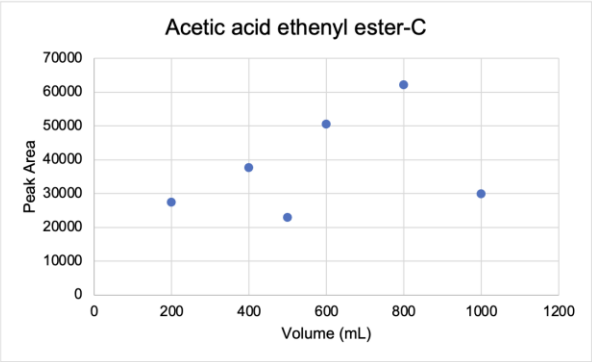
Benzyl methyl disulfide-A	0.386	1.000**
Ethylbenzene-A	0.564	.943**
Bicyclo[3.1.0]hexane, 4-methylene-1-(1-methylethyl)-	0.386	1.000**
Propanoic acid, anhydride-B	0.248	1.000**
Acetone-B	0.564	1.000**
2-Propenoic acid, 1-methylpropyl ester	0.043	1.000**
Pyridine	0.149	-0.029
Limonene	0.043	1.000**
2,3-Dioxabicyclo[2.2.1]heptane, 1-methyl-	0.386	.943**
Octane, 4-ethyl-A	0.564	.943**
Bis(2-furfuryl)disulfide	0.043	0.657
Propanoic acid, pentafluoro-, ethyl ester	0.043	.943**
Benzene, 1-ethyl-2-methyl-	0.386	0.657
Heptacosane-A	0.386	.943**
Eicosane-A	0.564	1.000**
5-Hepten-2-one, 6-methyl-	0.386	.829*
4-Methyl-2,4-bis(p-hydroxyphenyl)pent-1-ene, 2TMS derivative	0.564	1.000**
Octanal, 7-methoxy-3,7-dimethyl-A	0.149	0.029
Undecane, 2,6-dimethyl-	0.083	1.000**
3-Octanol, acetate	0.021	.943**
Nonanal	0.043	.771*
Eicosane-B	0.083	1.000**
Hexanamide, 6-chloro-N-ethyl-N-isobutyl-	0.149	0.314
Eicosane-C	0.149	.943**
1-Hexanol, 2-ethyl-	0.021	0.657
Decanal	0.021	0.429
Hexadecane	0.149	1.000**
Benzaldehyde	0.149	0.657
Heptacosane-C	0.248	1.000**
2-Bromononane	0.149	1.000**
Propanedinitrile, ethyl(1-oxopropoxy)-	0.773	1.000**
1-Propionylethyl acetate	0.773	1.000**
Benzenecarbothioic acid	0.043	0.429
2-Propenoic acid, 1,7,7-trimethylbicyclo[2.2.1]hept-2-yl ester, exo-	1.000	1.000**
2-Undecen-4-ol	0.248	1.000**

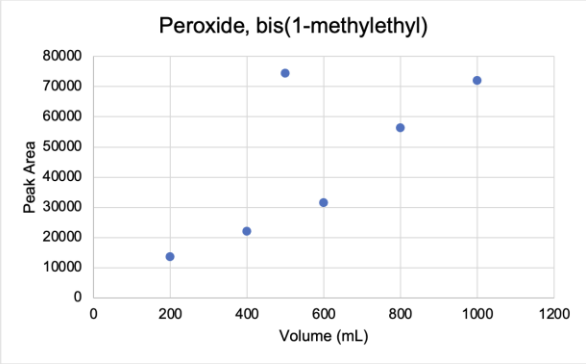
Octanal, 7-methoxy-3,7-dimethyl-C	0.149	.829*
Acetic acid ethenyl ester-C	0.149	0.429
Dodecane, 6-methyl-	0.564	.771*
Benzene, (1-pentylheptyl)-	0.043	.943**
N-Amino-N""-carbobenzyloxy-aspaginylglycine ethyl ester	0.083	-0.200
5,9-Undecadien-2-one, 6,10-dimethyl-	0.083	-0.200
Butanoic acid, 2-methylpropyl ester	0.021	0.143
Butanoic acid, anhydride	0.083	0.086
Octanal, 7-methoxy-3,7-dimethyl-D	1.000	.886**
Dimethyl sulfone	0.021	0.714
Octanal, 7-methoxy-3,7-dimethyl-E	0.043	.771*
1-Heptanol, 6-methyl-	0.083	0.429
3,4-Hexanedione, 2,2,5-trimethyl-B	0.564	0.600
Phenol	0.248	.943**
Benzene, (1-methylpropoxy)-	0.021	-0.314
1-Bromo-3-butene-2-ol	0.248	0.543
Threose triacetate	0.248	0.543
Octanal, 7-methoxy-3,7-dimethyl-H	0.564	.771*
Pentane, 2-methoxy-2,4,4-trimethyl-	0.021	0.029
1,3,7-Nonatriene-1,1-dicarbonitrile, 4,8-dimethyl-, (E)-	0.043	-.771*
Dodecane, 1-iodo-	0.248	0.314
Cyclopentane, 1,1,3-trimethyl-	0.248	-0.257
1,2-Benzenedicarboxylic acid, decyl hexyl ester	0.149	0.714
1,3-Diacetyl-cyclopentane	0.149	-0.257
Heptacosane-E	0.149	0.314
Vinyl benzoate	0.043	-0.314
1-Nitrododecane	0.149	0.314

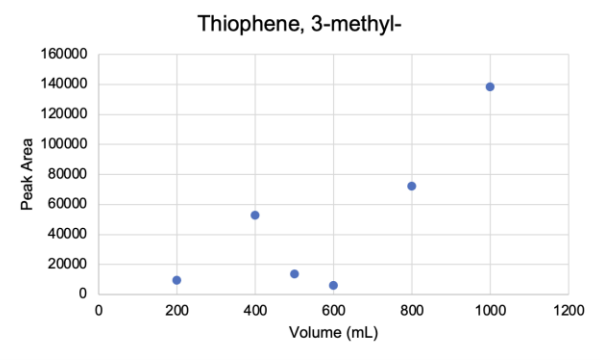

Furan
Peak Area
Volume (mL)

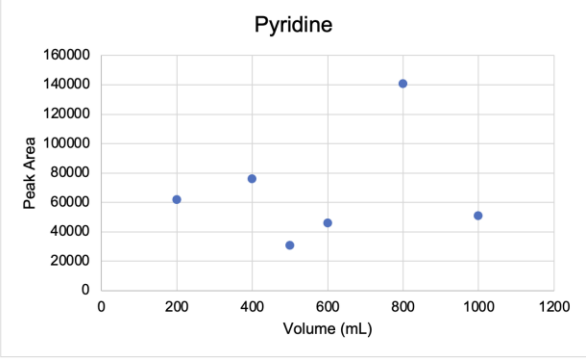

Propane, 1-(methylthio)-
Peak Area
Volume (mL)

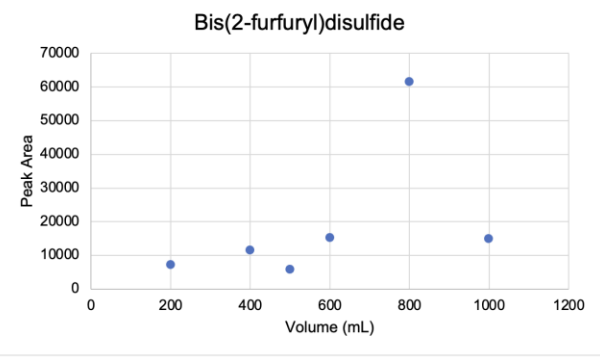

Sulfide, allyl methyl
Peak Area
Volume (mL)


Acetic acid ethenyl ester-C
Peak Area
Volume (mL)


Peroxide, bis(1-methylethyl)
Peak Area
Volume (mL)


Thiophene, 3-methyl-
Peak Area
Volume (mL)


Pyridine
Peak Area
Volume (mL)


Bis(2-furfuryl)disulfide
Peak Area
Volume (mL)

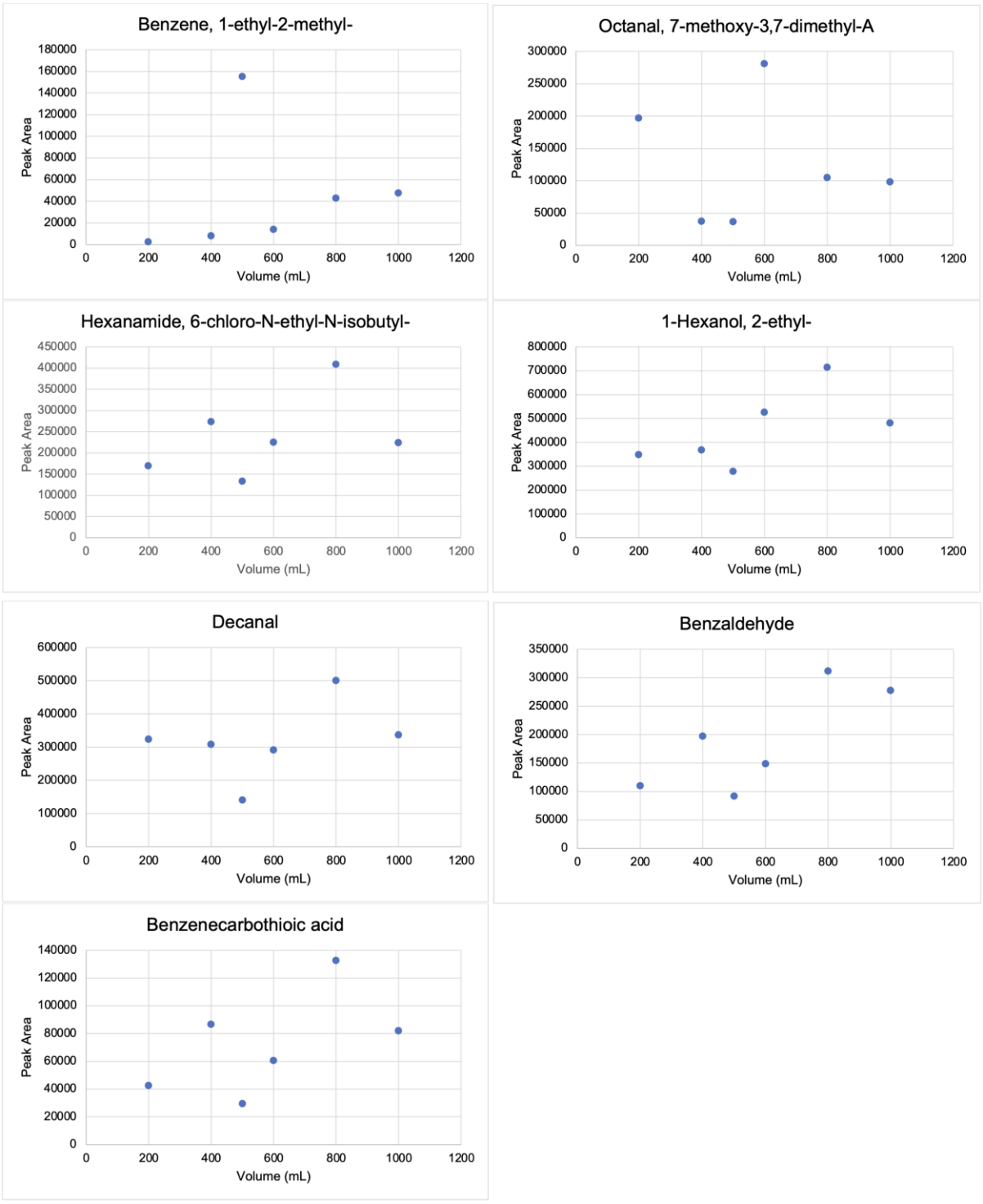


Figure 17.1 Scatter graphs of VOCs with a Spearman Rank coefficient <0.8.

17.3 Standard Operating Procedure for Novel Breath Collection Device

Imperial College London

Breath Collection with Novel Breath Collection Device Utilising Single Use Bags	
SOP Reference: SOP/VOC/VOD1.0	
Version Number: 1.0	
Effective Date: 16/06/2020	Review By: 16/06/2021
Author: Michael Hewitt & Sara Jamel	
Approved By:	Date:
Location:	

Version	Date	Reason For Changes
Version 1.0	16/06/2020	

Imperial College London

Imperial College London

1. PURPOSE:

The purpose of this SOP is to guide the operator in the use of the Novel Breath Collection Device to enable consistent and reliable use of the device to optimise quality of samples.

2. INTRODUCTION:

The novel breath collection device is a custom designed instrument that facilitates loading of breath samples on to thermal desorption (TD) tubes for subsequent analysis via mass spectrometry instruments.

This allows collection of breath samples where direct sampling is not available and also allows standardisation of breath sample collection.

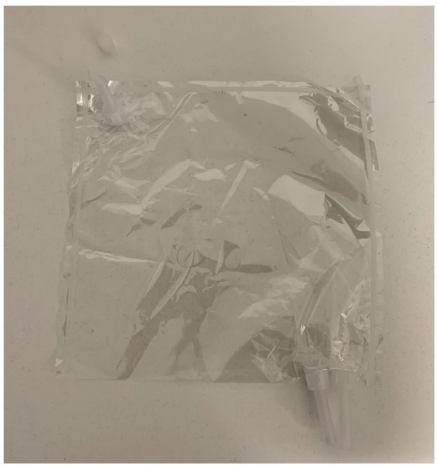
The device has adjustable flow rates and timings which allows differing volumes of breath be loaded on to the TD tubes depending on the operator's requirements.

3. RESPONSIBILITIES:

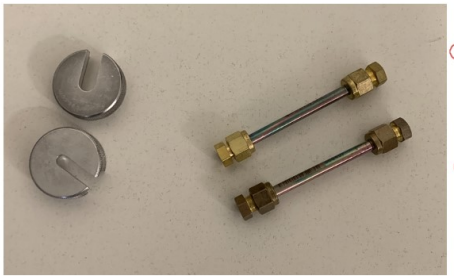
All staff using the novel breath collecting device using bags are responsible for its safe transport, appropriate use of and for checking quality of the samples taken.

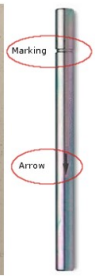
4. EQUIPMENT NEEDED

- Novel Breath Collection Device

Breath Collection Bag with Cap 

- Thermal Desorption Tubes and Spanner (each bag will provide two loaded tubes)





- Disposable Gloves
- Surgical Face Masks

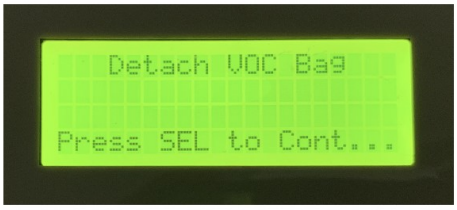
5. PROCEDURES

5.1 Prior to Participant Arrival

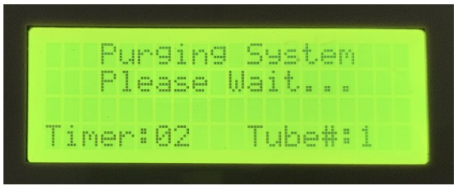
Set up area where sampling will take place to allow 2m social distancing between participant and operator of the device.

Don non-sterile disposable gloves. Turn the Novel Breath Collection Device on using the on/off flip switch at the rear of the unit and ensure that the device is connected to the mains power socket or that there is sufficient battery life for the breath collecting episode.

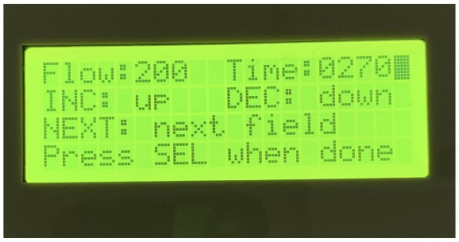
Once the device has been turned on, you will be met by the following screen:



Ensure there is nothing connected to the Novel Breath Collection Device then press the "SEL" button to continue. The system will then purge each of the tube sockets.



Once purging is complete, the settings screen will appear:



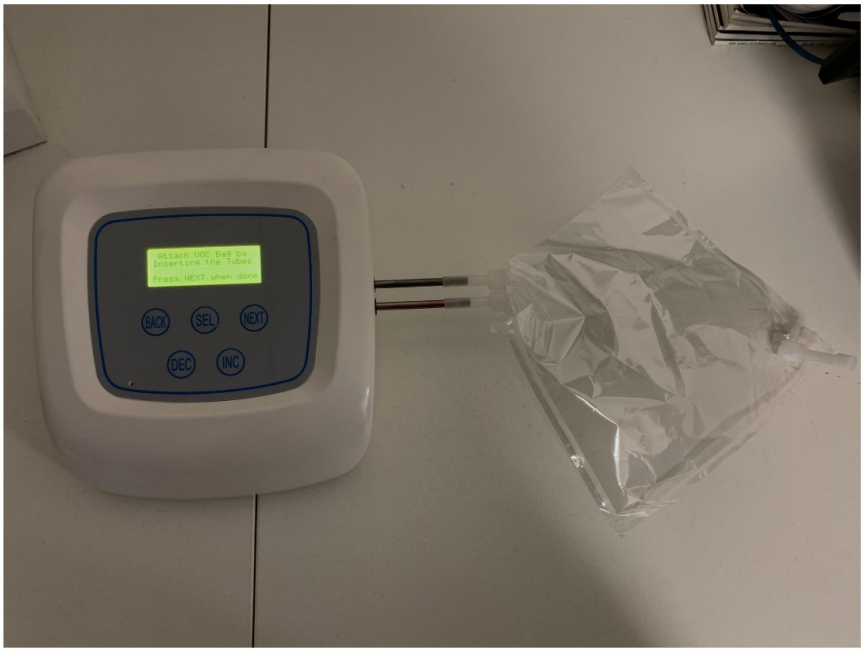
The default settings is for a flow rate of 200mls/min. By pressing "DEC" or "INC" buttons, it allows you to change between 50, 100, 150 and 200mls/min. Once selected, press 'NEXT' to move to the "Time" settings. Time can be adjusted using "DEC" or "INC" in 30 second increments.

The preferred settings for the device is a flow rate of 200mls/min and a time of 150s to give 500mls of breath.

Press 'SEL' when both settings have been adjusted as required. The following screen will then appear:

Next, remove the caps from the TD tubes using the spanner and insert the TD tubes into the sockets, with the arrows on the tube pointing towards the device.

Inspect the disposable bag to ensure it is clean. Attach the disposable bag to the TD tubes using the two ports at the bottom of the bag, ensuring you do not touch the mouthpiece.



5.2 Participant Arrival

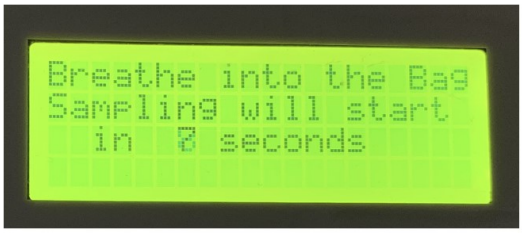
Change your disposable gloves and don your surgical face mask.

Invite the participant whose breath you wish to sample into the room, maintaining 2m distance at all times. Also provide them with a surgical face mask. Explain the study and the test and offer a patient information leaflet. Give them the opportunity to ask questions.

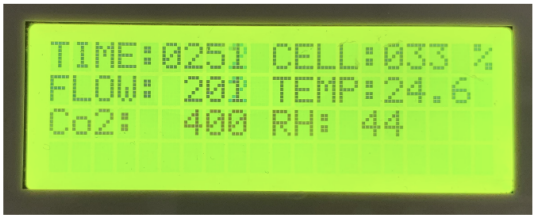
Obtain valid consent from the participant, ensuring they have signed the consent form.

Explain that you will press a button on the device and that they will then be required to exhale into the mouthpiece to fill the bag with their breath. Once they have filled the bag, they should place the plastic cap on the end of the mouthpiece. They should not touch the device itself.

Once explained, press the 'NEXT' button and instruct the patient to breathe into the bag. They will need to remove their face mask to breath into the bag, but this should be replaced as soon as the bag has been filled.



Following this screen, breath will start to be drawn through the TD tubes by the device, one tube at a time. Progress will be demonstrated on the LCD screen:



Minor fluctuations in the flow rate are to be expected but should not exceed +/- 10% of the selected flow rate.

At this time, the participant can fill in the CRF or they can leave depending on the requirements of the device operator.

Once the device has finished loading the tubes, they can be safely removed and re-capped using the spanners.

If the participant has not already left, thank them for their time and ensure they have a copy of their paperwork.

To repeat, press “NEXT” and the purging progress will restart and allow the cycle to be repeated as required.

Ensure changing of disposable gloves between patients.

6. MAINTENANCE

The device itself does not require any specific maintenance but should be wiped down between sessions with universal sanitising wipes.

The unit can be powered by both the mains power adaptor or by battery. The preferred batteries for the device are Samsung 25R 18650 batteries. Each device requires two batteries and they should be fully charged before being inserted into the system.

The battery compartment is located on the underside of the device and is accessed by removing the 4 screws with a crosshead screwdriver.

17.4 Optimum Strategy for Recollection of Breath Samples via Sample Splitting During Thermal Desorption

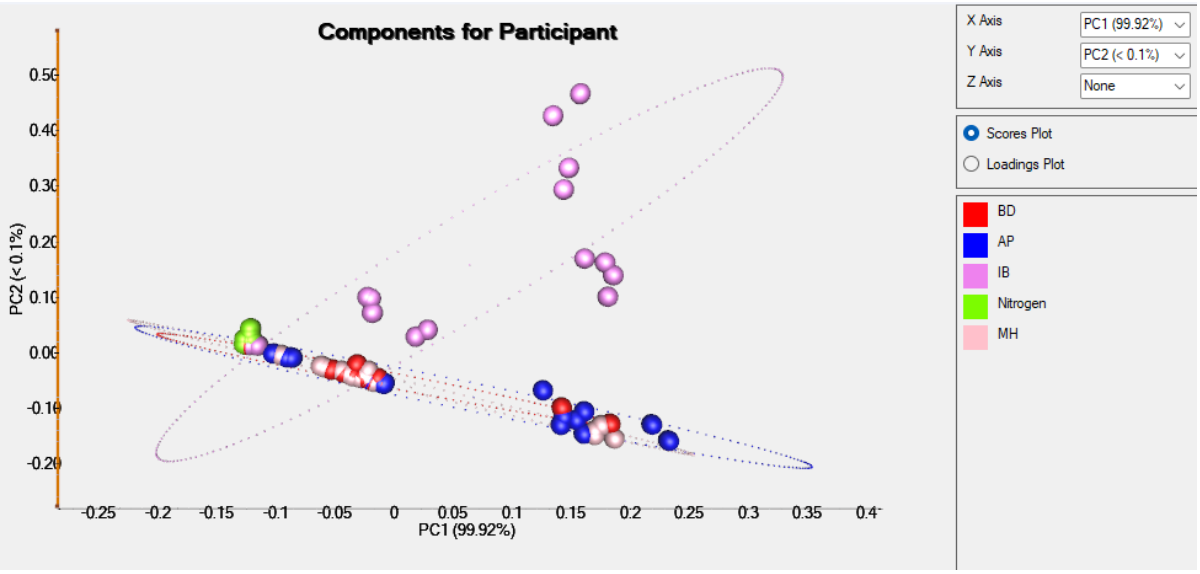


Figure 17.2 PCA plot of recollection pathway one without normalisation.

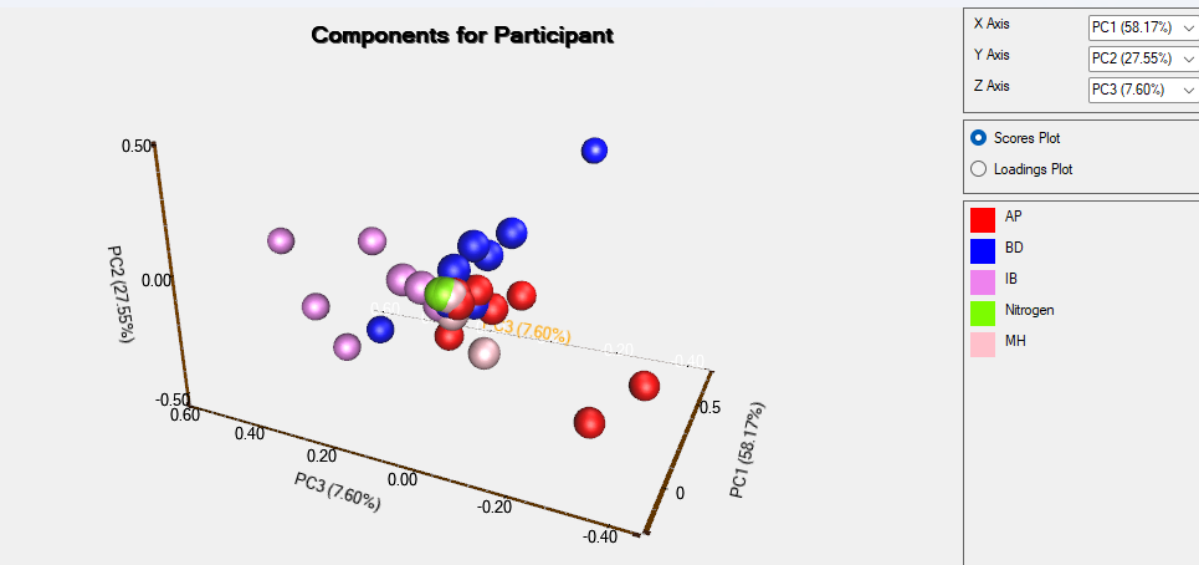


Figure 17.3 PCA plot of recollection pathway two without normalisation.

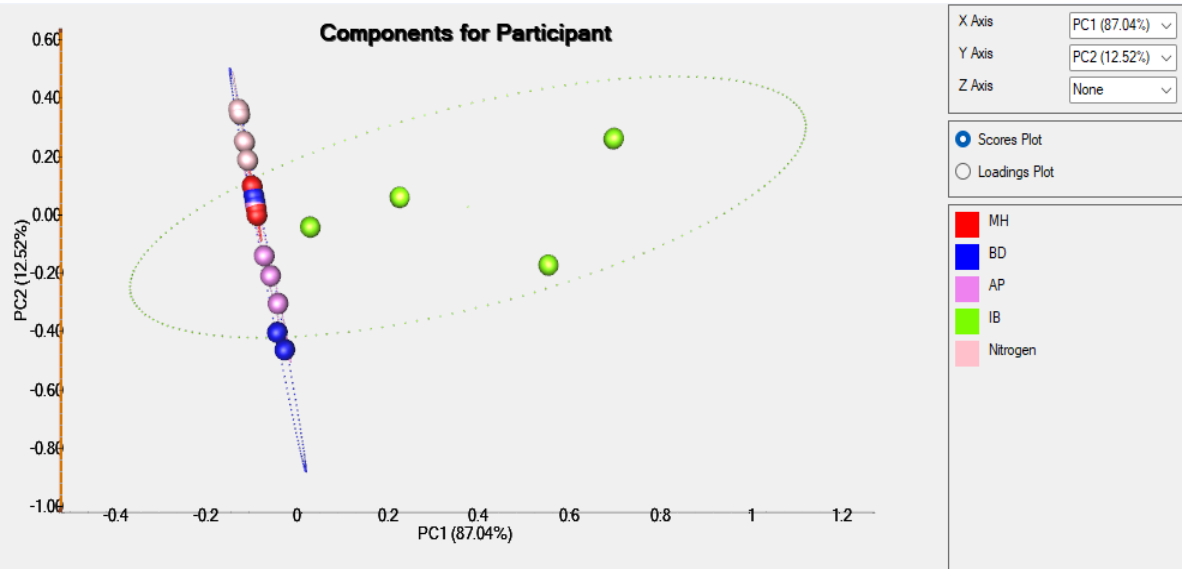


Figure 17.4 PCA plot of recollection pathway three without normalisation.

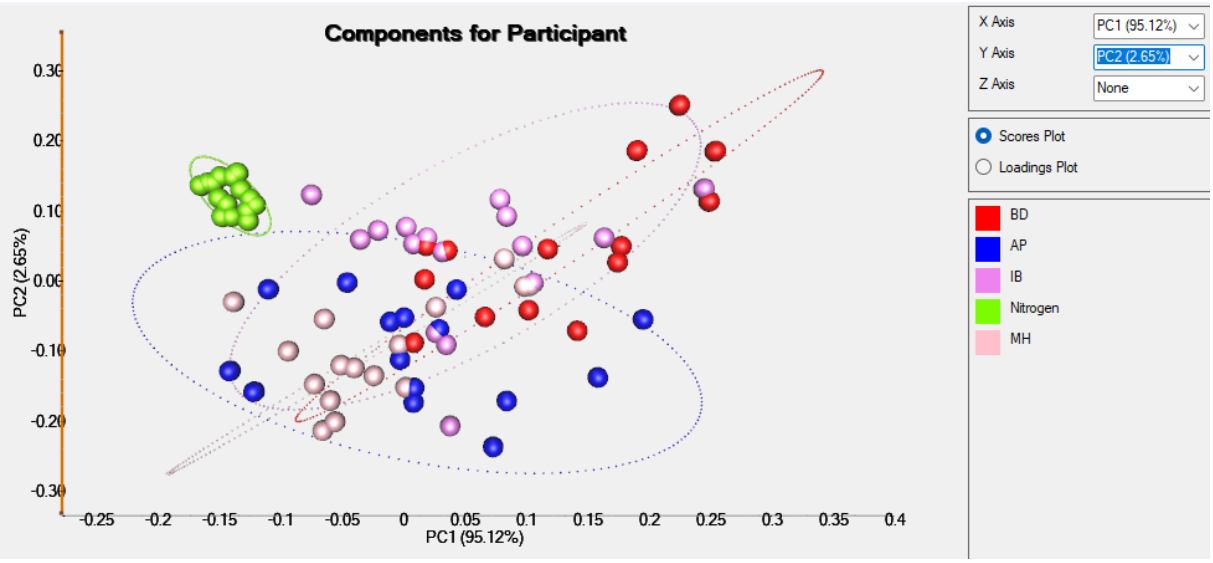


Figure 17.5 PCA plot of recollection pathway one with PQN normalisation demonstrating excellent separation of nitrogen samples from breath samples.

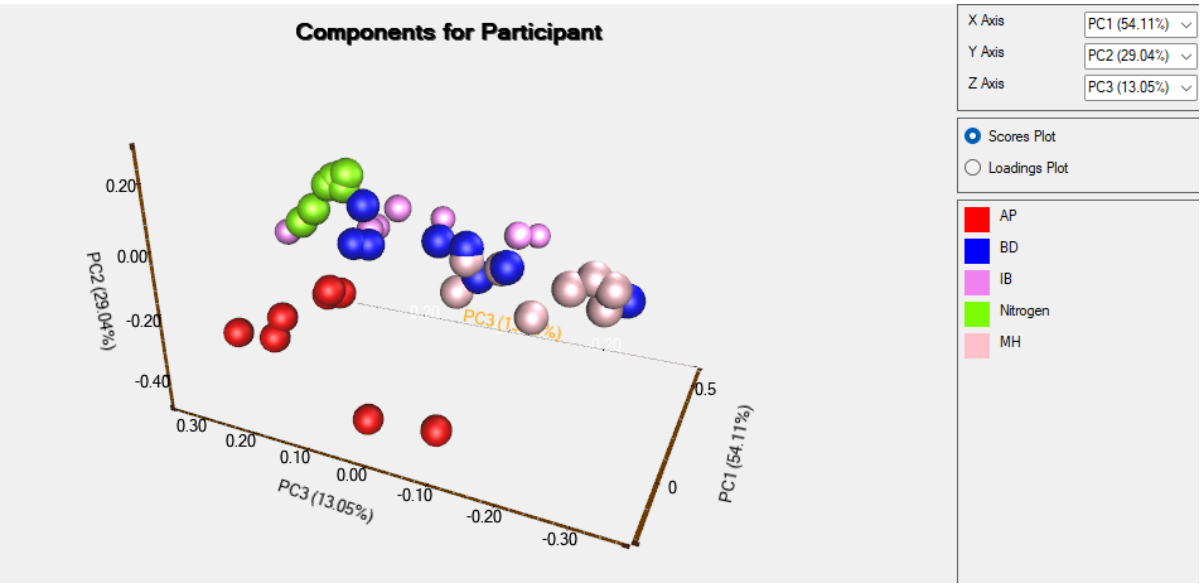


Figure 17.6 PCA plot of recollection pathway two with PQN normalisation demonstrating excellent separation of nitrogen samples from breath samples.

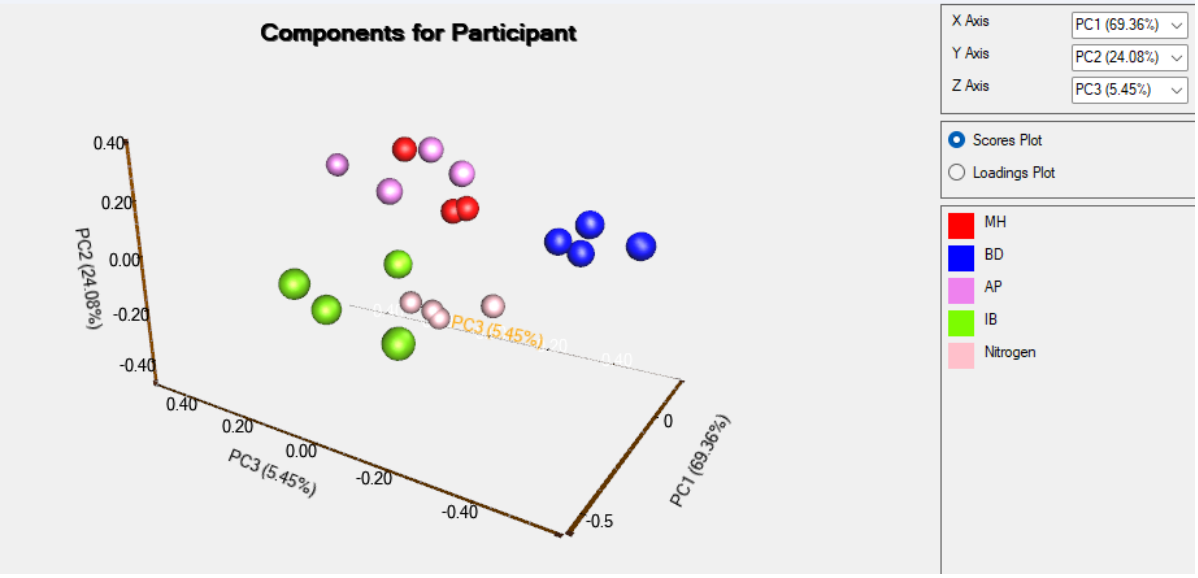


Figure 17.7 PCA plot of recollection pathway two with PQN normalisation demonstrating separation of nitrogen samples from breath samples.

17.5 Variation of VOC Concentrations within Ambient Room Air and its Impact upon the Standardisation of Breath Sampling

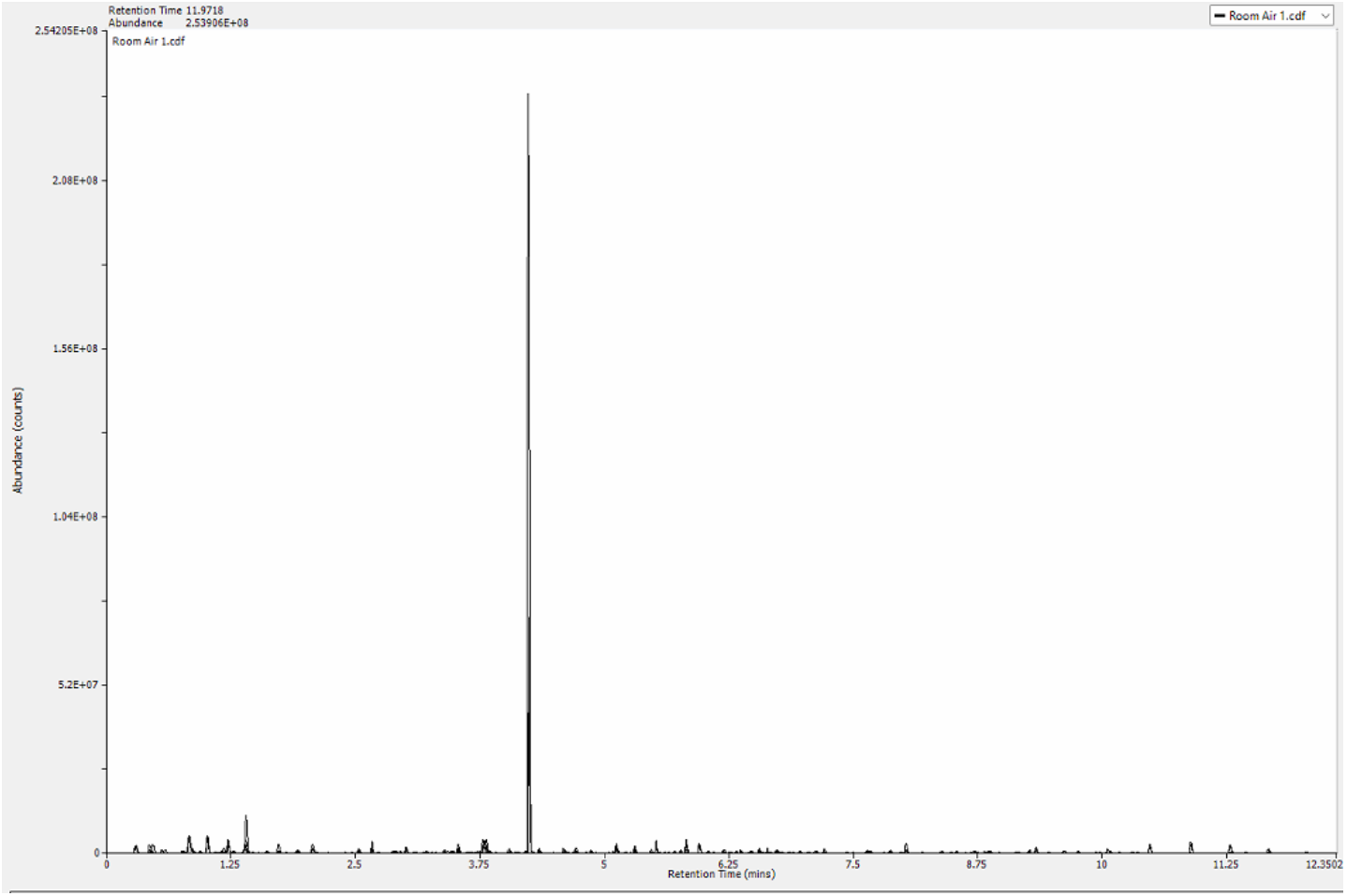


Figure 17.8 Chromatogram of five overlying representative room air samples. The large peak at RT 4.24 represents cyclopentasiloxane, decamethyl-, a contaminant from the mass spectrometry instrument.

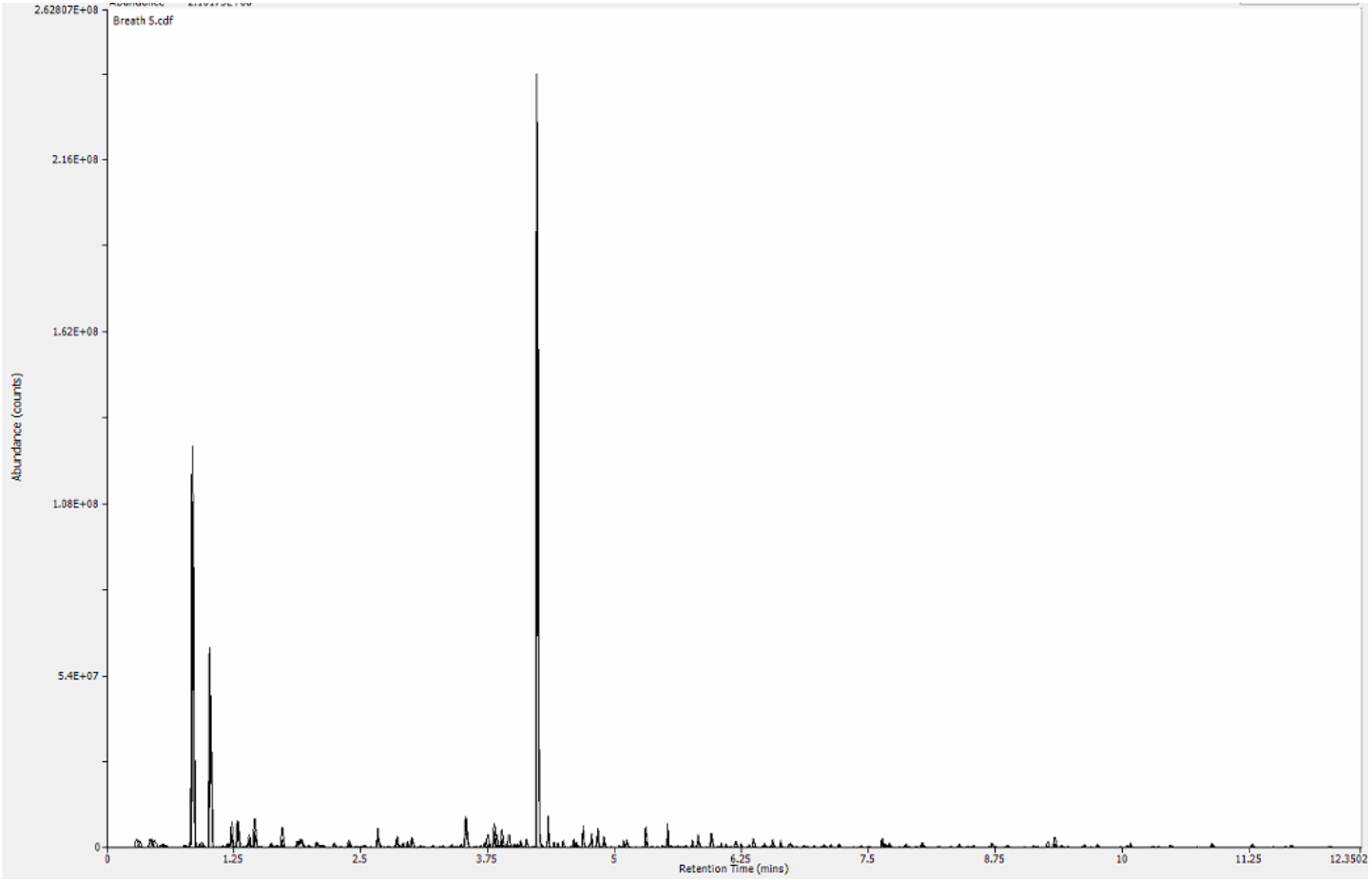


Figure 17.9 Chromatogram of five overlying representative breath samples.

Table 17.2 Identified VOCs driving separation between room air and breath.

Room Air	Compound	CAS ID	Class	VIP Score	Potential Sources
	Di-isopropyl phthalate	605-45-8	Phthalate Ester	1.34	Plasticiser (316), fragrances (265)
	Benzyl alcohol	100-51-6	Alcohol	1.32	Fragrances, solvents (269), soap, local Anaesthetic (278), food (apple, apricot) (279)
	Benzophenone	119-61-9	Ketone	1.31	Plasticisers, fragrances (266)
	Acetophenone	98-86-2	Ketone	1.29	Solvent, plasticiser, fragrance (267), food (beef, plum, raspberry) (268)
	Azulene	275-51-4	Aromatic Hydrocarbon	1.28	Emollients, fragrances, skin conditioner (317)
	Naphthalene, 1-methyl	90-12-0	Aromatic Hydrocarbon	1.27	Fragrances, antimicrobials (318), food (apple, grape, strawberry) (319)
	Benzaldehyde	100-52-7	Aldehyde	1.27	Fragrances, solvent, plastic additives (320), food (apple, apricot, cinnamon) (321)
	Propanoic acid, 2-methyl-, 2-ethyl-3-hydroxyhexyl ester	74367-31-0	Ester	1.26	Food (apricot, plumcot) (322)
	Diphenylacetylene	501-65-5	Alkyne	1.25	Fragrances (323)
	Isobutyl salicylate	87-19-4	Salicylate	1.24	Fragrances, preservatives, antimicrobial (324)
	Naphthalene, 2-methyl-	91-57-6	Aromatic Hydrocarbon	1.24	Cigarette smoke (325)
	Benzothiazole	95-16-9	Heterocyclic Compound	1.22	Fragrances (326) , food (asparagus, cocoa, mango) (327)

	Ethylbenzene	100-41-4	Aromatic Hydrocarbon	1.21	Petroleum, inks, pesticides, paint (328), food (apricot, cherry, peach, kiwi) (329)
	Furfural	98-01-1	Aldehyde	1.20	Tea, coffee, fruits, wholegrain bread (330)
	m-cymene	535-77-3	Aromatic Hydrocarbon	1.19	Fragrances, skin conditioner (331)
	m-Xylene	108-38-3	Aromatic Hydrocarbon	1.190	Solvent (332), fruits (kiwi, apricot) (333)
	1-Dodecanol	112-53-8	Alcohol	1.19	Fragrances (334), toothpaste, detergent (335)
	p-Xylene	106-42-3	Aromatic Hydrocarbon	1.19	Solvents (336)
	1H-Indene, 2,3-dihydro-1,1,3-trimethyl-3-phenyl-	3910-35-8	Hydrocarbon	1.19	Adhesives (337)
	5,9-Undecadien-2-one, 6,10-dimethyl-, (E)-	3796-70-1	Ketone	1.18	Cigarettes, flavourings (338), fragrances, fruit (339)
Breath	Isoprene	78-79-5	Hydrocarbon	1.40	Released from muscles (105)
	Branched Tridecane D		Alkane	1.38	Lipid peroxidation (340)
	Branched Undecane C		Alkane	1.36	Lipid peroxidation (340)
	Branched Tridecane E		Alkane	1.35	Lipid peroxidation (340)
	Branched Tridecane C		Alkane	1.32	Lipid peroxidation (340)
	Branched Dodecane C		Alkane	1.30	Lipid peroxidation (340)
	Oxalic acid	144-62-7	Fatty Acid	1.30	Dehydroascorbic acid and glyoxylate metabolism (341)
	Propane, 1-(methylthio)-	3877-15-4	Sulphur Compound	1.29	Kohlrabi (342)
	Branched Dodecane D		Alkane	1.28	Lipid peroxidation (340)

	Branched Tridecane A		Alkane	1.28	Lipid peroxidation (340)
	Sulfide, allyl methyl	10152-76-8	Sulphur Compound	1.27	Garlic (343)
	1-Propanol	71-23-8	Alcohol	1.25	Acetone reduction, amino acid breakdown, disinfectants (273)
	Branched tridecane I		Alkane	1.24	Lipid peroxidation (340)
	2-Butenal	4170-30-3	Aldehyde	1.23	Lipid peroxidation (344)
	Branched tridecane H		Alkane	1.21	Lipid peroxidation (340)
	Branched undecane A		Alkane	1.20	Lipid peroxidation (340)
	1-heptanol, 2-propyl	10042-59-8	Alcohol	1.18	Soaps, detergents (345)
	β-pinene	127-91-3	Monoterpene	1.15	Plant metabolite, pine trees, essential oils (272)
	D-Limonene	5989-27-5	Monoterpene	1.14	Citrus oils (346)
	1-Hexanol, 5-methyl-2-(1-methylethyl)-	2051-33-4	Monoterpene	1.13	Air freshener (347)

Table 17.3 Identified VOCs driving separation between AM and PM samples.

AM	**Compound**	**CAS ID**	**Class**	**VIP Score**	**Possible Sources**
	Branched tridecane F		Alkane	2.97	Lipid peroxidation (340), cleaning products and waxes (280)
	Branched hexadecane A		Alkane	1.57	Lipid peroxidation (340), cleaning products and waxes (280)
	Branched pentadecane		Alkane	1.57	Lipid peroxidation (340), cleaning products and waxes (280)
	Branched tetradecane A		Alkane	1.49	Lipid peroxidation (340), cleaning products and waxes (280)
	Oxalic acid	144-62-7	Fatty Acid	1.45	Dehydroascorbic acid and glyoxylate (341)
	Branched tridecane H		Alkane	1.40	Lipid peroxidation (340), cleaning products and waxes (280)
	Branched hexadecane B		Alkane	1.38	Lipid peroxidation (340), cleaning products and waxes (280)

	Hexacosane	630-01-3	Alkane	1.37	Flowers, coffee (348)
	Branched tridecane I		Alkane	1.32	Lipid peroxidation (340)
	Branched Tridecane A		Alkane	1.28	Lipid peroxidation (340)
	Pentacosane	629-99-2	Alkane	1.26	Pesticides (349), avocado, coffee (350)
	Heptacosane	593-49-7	Alkane	1.19	Coffee, avocado (351)
	Branched dodecane C		Alkane	1.19	Lipid peroxidation (340), cleaning products and waxes (280)
	Branched tridecane G		Alkane	1.17	Lipid peroxidation (340), cleaning products and waxes (280)
	Branched Tridecane B		Alkane	1.13	Lipid peroxidation (340), cleaning products and waxes (280)
	Branched Undecane C		Alkane	1.13	Lipid peroxidation (340), cleaning products and waxes (280)
	n-pentadecane	629-62-9	Alkane	1.12	Pesticides (352), plants, food (egg, mango, mandarin, papaya) (353)
	Tetracosane	646-31-1	Alkane	1.07	Coffee, mustard, rose, coconut (354)
	1-Pentadecene	13360-61-7	Alkene	1.03	Soap, polishing agents (355)
PM	1-Propanol	71-23-8	Alcohol	1.99	Acetone reduction, amino acid breakdown, disinfectants (273)
	Phenol	108-95-2	Aromatic Hydrocarbon	1.82	Disinfectants (281), cigarette smoke (287)
	Propanoic acid, 2-methyl-, 2-ethyl-3-hydroxyhexyl ester	74367-31-0	Ester	1.66	Plumcot, apricot (356)
	Isoprene	78-79-5	Hydrocarbon	1.52	Released from muscles (105), rubber (357)
	Nonanal	124-19-6	Aldehyde	1.51	Essential oils, flavourings (358)
	5,9-Undecadien-2-one, 6,10-dimethyl-, (E)-	3796-70-1	Ketone	1.50	Cigarettes, flavourings (338), fragrances, fruit (339)
	Acetophenone	98-86-2	Ketone	1.35	Solvents, plasticisers, fragrance (267), food (beef, plum, raspberry) (268)

	Decanal	112-31-2	Aldehyde	1.32	Synthetic citrus oils, fragrance, air freshener (359)
	Benzaldehyde, 3,4-dimethyl-	5973-71-7	Aldehyde	1.32	Wine (360)
	3,4-Difluorobenzaldehyde	34036-07-2	Aldehyde	1.30	Curcumin (361)
	Benzaldehyde	100-52-7	Aldehyde	1.29	Fragrance, solvent, plastic additive (320), food (apple, apricot, cinnamon) (321)
	D-Limonene	5989-27-5	Monoterpene	1.26	Citrus oils (346)
	Isopropyl Alcohol (2-propanol)	67-63-0	Alcohol	1.260	Antimicrobial, solvents (362)
	1-Tetradecanol	112-72-1	Alcohol	1.25	Cosmetics (363)
	Acetone	67-64-1	Ketone	1.21	Fat metabolism, solvents (364)
	2-pentanone	107-87-9	Ketone	1.20	Food additives (243)
	Benzophenone	119-61-9	Ketone	1.16	Plasticisers, fragrances (266)
	Benzyl alcohol	100-51-6	Aromatic Alcohol	1.16	Fragrances, solvents (269), soaps, local anaesthetic (278), food (apple, apricot) (279)
	o-cymene	527-84-4	Hydrocarbon	1.14	Plant oil, seasonings, flavourings (365)
	Furan, 2-pentyl-	3777-69-3	Furan	1.09	Heat-processed food and drink, human urinary metabolite, flavouring agent (366)

Table 17.4 Identified VOCs driving separation between locations.

Endoscopy Unit	**Compound**	**CAS Registry ID**	**Class**	**VIP**	**Possible Sources**
	3-Carene	13466-78-9	Monoterpene	3.21	Plant / citrus oils (226)
	β-pinene	127-91-3	Monoterpene	1.81	Plant metabolite, pine trees, essential oils (272)
	Dodecane	112-40-3	Alkane	1.66	Essential oils (367)

	Benzonitrile	100-47-0	Nitrile	1.62	Specialty solvents (368)
	Undecane	1120-21-4	Alkane	1.54	Essential oils (369)
	5-Hepten-2-one, 6-methyl-	110-93-0	Ketone	1.52	Citronella oil, fragrances, flavourings (370)
	Benzaldehyde	100-52-7	Aldehyde	1.35	Fragrances, solvents, plastic additives (320), food (apple, apricot, cinnamon) (321)
	5,9-Undecadien-2-one, 6,10-dimethyl-, (E)-	3796-70-1	Ketone	1.34	Cigarettes, flavourings (338), fragrances, fruits (339)
	Tridecane	629-50-5	Alkane	1.11	Essential oils (371)
	Branched tridecane J		Alkane	1.11	Lipid peroxidation (340), cleaning products and waxes (280)
Research Bay	3-Carene	13466-78-9	Monoterpene	3.21	Plant, citrus oils (226)
	α-Pinene	80-56-8	Monoterpene	3.1	Essential oils, solvents (372)
	Di-isopropyl phthalate	605-45-8	Phthalate Ester	1.8	Plasticisers (316), fragrances (265)
	Branched Undecane C			1.25	Lipid peroxidation (340), cleaning products and waxes (280)
Main Theatres	Branched decane		Alkane	2.78	Lipid peroxidation (340), cleaning products and waxes (280)
	Propanoic acid, 2-methyl-, 2-ethyl-3-hydroxyhexyl ester	74367-31-0	Ester	2.64	Food (apricot, plumcot) (322)
	Toluene	108-88-3	Monoterpene	2.41	Paint, inks, lacquer, fuel (373)
	3-Carene	13466-78-9	Monoterpene	2.37	Plant and citrus oils (226)
	Branched Dodecane D		Alkane	2.36	Lipid peroxidation (340)

	2-Butenal	4170-30-3	Aldehyde	1.65	Sorbic acid (preservative) (374)
	Branched Tridecane D		Alkane	1.55	Lipid peroxidation (340), cleaning products and waxes (280)
Laboratory	Acetamide, 2,2,2-trifluoro-N-methyl-	815-06-5	Amide	2.87	Antimicrobials (375)
	Pyridine	110-86-1	Heterocyclic Compound	2.71	Medicines, vitamins, food flavourings, paints, dyes (376)
	Branched undecane A		Alkane	2.5	Lipid peroxidation (340), cleaning products and waxes (280)
	Furan, 2-pentyl-	3777-69-3	Furan	1.49	Heat-processed food and drink, human urinary metabolite, flavouring agent (366)
	Ethylbenzene	100-41-4	Aromatic Hydrocarbon	1.42	Petroleum, inks, pesticides, paints (328), food (apricot, cherry, peach, kiwi) (329)
	Furfural	98-01-1	Aldehyde	1.24	Tea, coffee, fruits, wholegrain bread (330)
	Ethyl anisate	94-30-4	Ester	1.21	Food additive, fragrance (377)
	o-Xylene	95-47-6	Aromatic Hydrocarbon	1.21	Solvents (378)
	Isopropyl Alcohol (2-propanol)	67-63-0	Alcohol	1.18	Antimicrobials, solvent (362)
	3-Carene	13466-78-9	Monoterpene	1.17	Plant and citrus oils (226)
	m-Xylene	108-38-3	Aromatic Hydrocarbon	1.13	Solvent (332), fruit (kiwi, apricot) (333)
Surgical Outpatient	1-Nonanol	143-08-8	Alcohol	2.75	Plant oil (277)
	Vinyl lauryl ether	765-14-0	Alkane	2.61	Coating, chemical intermediate (379)
	3-Carene	13466-78-9	Monoterpene	2.28	Plant and citrus oils (226)
	Benzyl alcohol	100-51-6	Aromatic Alcohol	2.16	Fragrances, solvents (269), soap, local anaesthetic (278), food

					(apple, apricot) (279)
	Branched tridecane J		Alkane	1.42	Lipid peroxidation (340)
	Tridecane	629-50-5	Alkane	1.42	Essential oils (371)
	Ethanol, 2-phenoxy-	122-99-6	Glycol Ether	1.28	Insect repellent, antiseptic, solvent, preservative (380)
	Isobutyl salicylate	87-19-4	Ester	1.09	Fragrance, preservative, antimicrobial (324)
	Naphthalene, 2-methoxy-	91-57-6	Aromatic Hydrocarbon	1.07	Cigarette smoke (325)

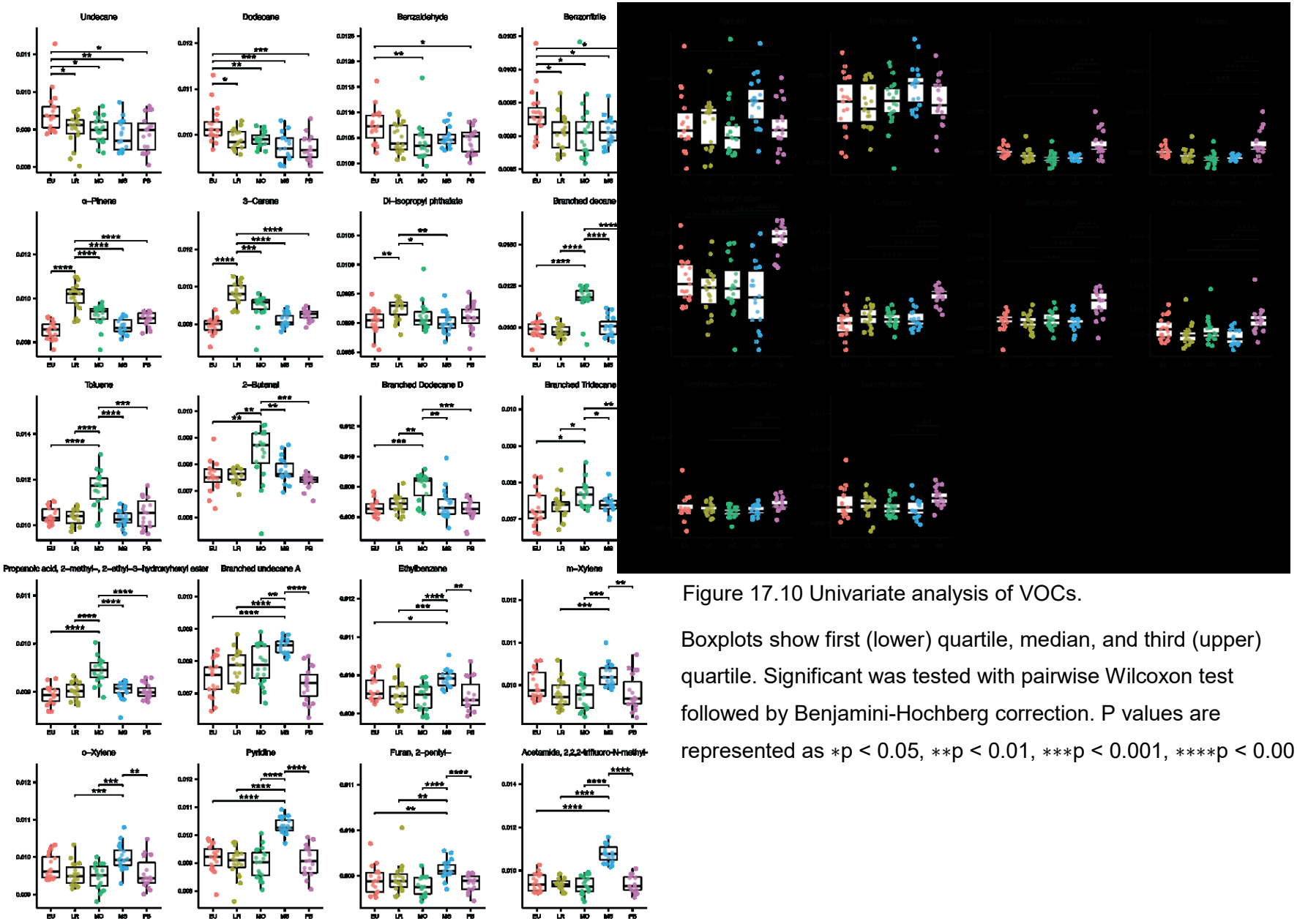


Figure 17.10 Univariate analysis of VOCs.

Boxplots show first (lower) quartile, median, and third (upper) quartile. Significant was tested with pairwise Wilcoxon test followed by Benjamini-Hochberg correction. P values are represented as $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.001$.

17.6 Volatile Organic Compounds for the Assessment of Liver Disease via SIFT-MS

Table 17.5 VOCs used in each SIFT-MS method with ionisations and product ions.

	Compound	Reagent Ion	m/z	Product Ion
Method 1	2-Nonanone	O_2^+	142	$C_9H_{18}O^+$
	2-Nonanone	H_3O^+	143	$C_9H_{18}OH^+$
	2-Nonanone	H_3O^+	161	$C_9H_{18}OH^+.H_20$
	2-Nonanone	NO^+	172	$C_9H_{18}O.NO^+$
	2-Octanone	O_2^+	128	$C_8H_{16}O^+$
	2-Octanone	H_3O^+	129	$C_8H_{16}OH^+$
	2-Octanone	H_3O^+	149	$C_8H_{16}OH^+.H_20$
	2-Octanone	NO^+	158	$C_8H_{16}O.NO^+$
	2-Pentanone	O_2^+	86	$C_5H_{10}O^+$
	2-Pentanone	H_3O^+	87	$C_5H_{11}O^+$
	2-Pentanone	NO^+	116	$NO^+.C_5H_{10}O$
	Acetone	H_3O^+	59	$C_3H_7O^+$
	Acetone	NO^+	88	$NO^+.C_3H_6O$
	Butanone	O_2^+	72	$C_4H_8O^+$
	Butanone	NO^+	102	$NO^+C_4H_8O$
	Carbon Disulfide	O_2^+	76	CS_2+
	Dimethyl Sulfide	NO^+	62	$(CH_3)_2S^+$
	Dimethyl Sulfide	H_3O^+	63	$(CH_3)_2S.H^+$
	Indole	NO^+	117	$C_8H_7N^+$
	Indole	H_3O^+	118	$C_8H_8N^+$
	Isoprene	O_2^+	67	$C_5H_7^+$
	Isoprene	NO^+	68	$C_5H_8^+$
	Isoprene	H_3O^+	69	$C_5H_8.H^+$
	Limonene	NO^+	136	$C_{10}H_{16}^+$
	Limonene	H_3O^+	137	$C_{10}H_{17}^+$
	Styrene	NO^+	104	$C_8H_8^+$
	Styrene	O_2^+	104	$C_8H_{11}^+$
Method 2	1-Propanol	H_3O^+	43	$C_3H_7^+$
	1-Propanol	NO^+	59	$C_3H_7O^+$
	Acetaldehyde	NO^+	43	CH_3CO^+
	Acetaldehyde	H_3O^+	45	$C_2H_5O^+$
	Acetaldehyde	NO^+	61	$CH_3CO^+.H_2O$
	Acetaldehyde	H_3O^+	63	$C_2H_5O.H_2O$
	Acetone	H_3O^+	59	$C_3H_7O^+$
	Acetone	H_3O^+	77	$(CH_3)_2CO.H^+.H_2O$
	Acetone	NO^+	88	$NO^+.C_3H_6O$
	Ammonia	O_2^+	17	NH_3^+
	Ammonia	H_3O^+	18	NH_4^+

	Ammonia	H_3O^+	36	$NH_4^+.H_2O$
	Ethanol	NO^+	45	$C_2H_5O^+$
	Ethanol	H_3O^+	47	C_2H_7O
	Ethanol	NO^+	63	$C_2H_5O^+.H_2O$
	Ethanol	H_3O^+	65	$C_2H_7O^+.H_2O$
	Methanol	H_3O^+	33	CH_5O^+
	Methanol	H_3O^+	51	$CH_3OH_2^+.H_2O$
	Nonane	O_2^+	85	$C_6H_6O^+$
	Nonane	H_3O^+	147	$C_9H_2O.H_3O^+$
	Octane	H_3O^+	113	$C_8H_{17}^+$
	Octane	O_2^+	114	$C_8H_{18}^+$
	Pentane	O_2^+	72	$C_5H_{12}^+$
	Phenol	NO^+	94	$C_6H_6O^+$
	Phenol	O_2^+	94	$C_6H_6O^+$
	Phenol	H_3O^+	95	$C_6H_7O^+$
	Tetradecane	H_3O^+	217	$C_{14}H_3O.H_3O^+$

Table 17.6 p Values of Kruskal-Wallis (*) and Mann-Whitney U (**) tests for each analysis.

VOC	Control vs Fibrosis vs Cirrhosis*	Control vs Fibrosis**	Control vs Cirrhosis**	Fibrosis vs Cirrhosis**
2-Nonanone	0.41	0.88	0.32	0.11
2-Octanone	0.24	0.28	0.69	0.03
2-Pentanone	0.08	0.06	0.31	0.04
Acetone	0.54	0.39	0.85	0.19
Butanone	0.23	0.37	0.19	0.15
Carbon Disulfide	0.43	0.64	0.24	0.38
Dimethyl Sulfide	0.41	0.91	0.21	0.32
Indole	0.57	0.33	0.92	0.31
isoprene	0.33	0.12	0.85	0.31
Limonene	0.04	0.82	0.03	0.04
Styrene	0.49	0.42	0.41	0.34
1-Propanol	0.01	0.06	0.00	0.36
Acetaldehyde	0.02	0.02	0.03	0.53
Acetone	0.99	0.94	0.96	0.79
Ethanol	0.01	0.01	0.02	0.57
Methanol	0.87	0.61	0.94	0.62
Nonane	0.28	0.94	0.13	0.26
Octane	0.68	0.59	0.85	0.23
Pentane	0.45	0.25	0.67	0.29
Phenol	0.07	0.25	0.03	0.53
Tetradecane	0.96	0.76	0.92	0.89

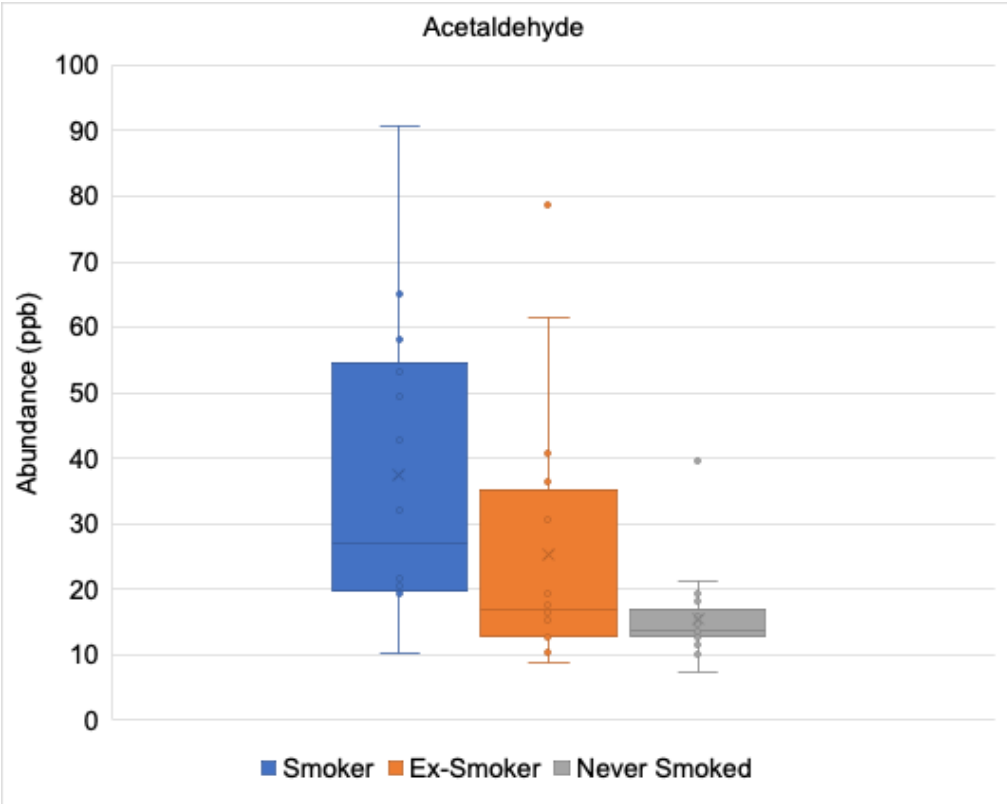


Figure 17.11 Box plot of concentrations of acetaldehyde according to smoking status demonstrating high concentrations of acetaldehyde in smokers and ex-smokers compared to those who have never smoked.

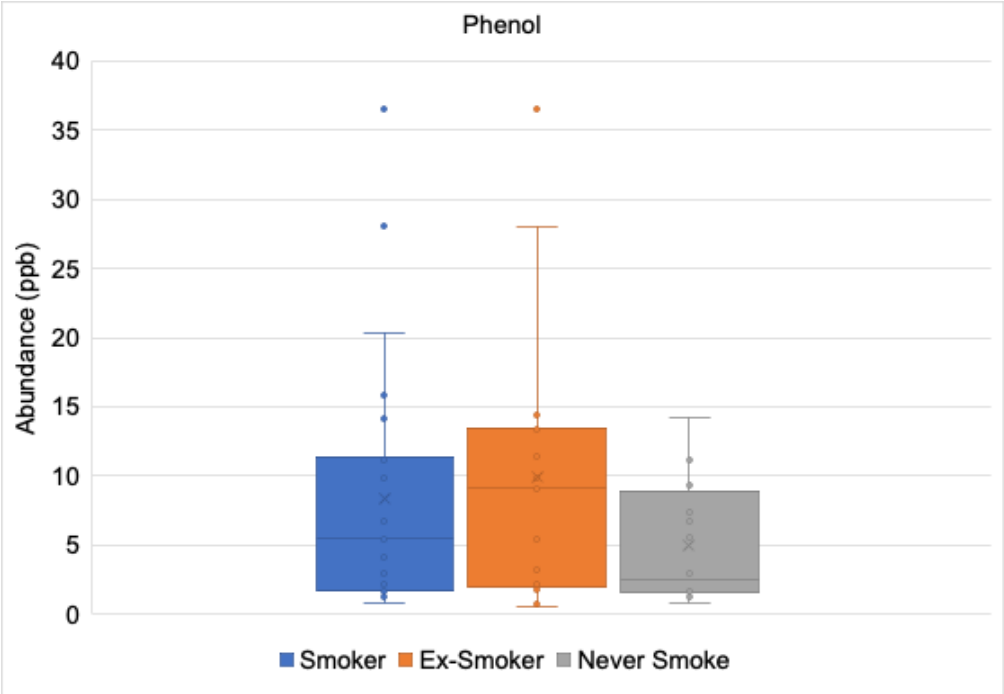


Figure 17.12 Box plot of concentrations of phenol according to smoking status demonstrating the highest concentrations of phenol in smokers and ex-smokers compared to those who have never smoked.

17.7 Volatile Organic Compounds for the Assessment of Liver Disease via GC-MS

17.7.1 Consent Form for VOCAL

Imperial College London

Department of Surgery & Cancer
Imperial College London
10th Floor, QEQM Building
St Mary's Hospital
Praed Street, London W2 1NY
Tel: +44(0)20 7886 2125

Study Protocol Number: 1.2

INFORMED CONSENT FORM
FOR SUBJECTS ABLE TO GIVE CONSENT

Volatile Organic Compounds for the Assessment of Liver Disease (VOCAL)
Principal Investigator: **Professor George Hanna** PhD FRCS
Professor of Surgical Sciences / Consultant Surgeon
g.hanna@imperial.ac.uk

Please initial box

1. I confirm that I have read and understood the subject information leaflet dated ..... version ..... for the above study and have had the opportunity to ask questions and have had these answered fully. ☐
2. I understand that my participation is voluntary and that I may withdraw consent to my samples being included in the study at any time without justifying my decision and without affecting my medical care or legal rights. ☐
3. I understand that sections of any of my medical notes may be looked at by responsible individuals from the Imperial College London research team or from regulatory authorities where it is relevant to my taking part in this research. ☐
4. I give permission for these individuals to have access to my records that are relevant to this research. ☐
5. I give / do not give (*delete as applicable*) consent for information collected about me to be used to support other research in the future, including those outside of the EEA. ☐
6. I give / do not give (*delete as applicable*) consent for samples collected during this study to be used in future ethically approved studies. I give permission for my samples to be sent to other organisations, including those outside of the EEA. ☐
7. I understand that the samples are a gift and that I will not benefit from any intellectual property privilege that result from the use of the samples in research studies. ☐
8. I would like to be contacted if the result of any research carried out on my samples can help me during my treatment course. ☐
9. I agree to take part in the above study and the samples which I hereby consent to donate are breath / urine / ascitic fluid (*delete as applicable*). ☐

Imperial College London

Department of Surgery & Cancer
Imperial College London
10th Floor, QEQM Building
St Mary's Hospital
Praed Street, London W2 1NY
Tel: +44(0)20 7886 2125

10. I agree that I can be contacted in the future for the purpose of providing further breath / urine / ascitic fluid samples up to a maximum of 24 months following the date of enrolment in the study. ☐

11. I consent to take part in the above study. ☐

12. I give / do not give (delete as applicable) consent to being contacted to potentially taking part in other research studies, including follow up to this study. ☐

_____	_____	_____
Name of Participant	Signature	Date
_____	_____	_____
Researcher	Signature	Date

1 copy to be given to the participant, 1 copy in medical notes and original to be filed in trial site file.

17.7.2 VOCAL Participant Information Leaflet

Imperial College
London

Department of Surgery & Cancer
Imperial College London
10th Floor, QEQM Building
St Mary's Hospital
Praed Street, London W2 1NY
Tel: +44(0)20 7886 2125
Fax: +44(0)20 7413 0470

Professor George Hanna PhD FRCS
Professor of Surgical Sciences
g.hanna@imperial.ac.uk
www.imperial.ac.uk

PARTICIPANT INFORMATION LEAFLET

IRAS ID 254249 **Version 1.4** **13/01/2021**

Study Title: Volatile Organic Compounds for the Assessment of Liver Disease (VOCAL)

Chief Investigator: Professor G Hanna

Dear Sir/Madam

The doctors responsible for you care have indicated that you may be eligible to take part in the above named study.

Before you decide if you wish to participate in this study it is important for you to understand why the research is being undertaken and what it will involve. Please take time to read the following information leaflet carefully and discuss it with relatives, friends and your GP if you wish. Please ask us if there is anything that is not clear or if there is anything you would like more information on.

You are free to withdraw at any time from this research study without explanation and without affecting the standard of care you receive.

Thank you for reading this information leaflet.

What is the purpose of the study?
This study is designed to investigate new methods of detecting and monitoring diseases affecting the liver and surrounding organs (including bile ducts and the pancreas). Specifically, we would like to know how volatile organic compounds (VOCs), small molecules produced within the human body, can help create a new non-invasive test for these conditions.

We hope that such tests can one day be used to identify patients with these conditions at an early stage when treatment is likely to be more successful.

Why have I been chosen?
You have been chosen for this research project because you have been referred to hepatology (liver) services for investigation or follow up of a condition affecting your liver, bile ducts or pancreas.

Do I have to take part?
Involvement in research is entirely voluntary and it is up to you to decide whether or not you wish to take part. A decision not to take part or to withdraw from this study in the future will not affect the standard of care you receive.

What will happen to me if I take part?
If you agree to take part in this study, you will first be seen by one of the research doctors who will be able discuss the study with you and answer any questions you may have.

At any time, even after signing a consent form, you will be free to withdrawn from this study without having to give a reason and without affecting the standard of care you receive.

If at any time during the course of this study you are no longer able, for medical reasons, to communicate your wishes in regard to continued participation we will withdraw you from the study and any future sample collection. Any samples or information collected prior to this point will be retained and analysed in accordance with the conditions of this study.

What do I have to do?

As part of the study you will asked to give a breath and urine samples. Breath samples will be collected by asking you to breathe into a sample bag (similar to a balloon) or a special instrument, called a mass spectrometer, that can directly measure different molecules within your breath. In each case we will collected approximately 2 litres of breath. Breath sampling in this way typically takes no more than 5-10 minutes in total. We may ask you to have fasted for at least four hours or to perform basic oral hygiene (e.g. rinsing your mouth with water) before providing the breath sample. We will also ask you to complete a short food diary. You may also be asked to pass a small quantity of urine (20mL) into a sterile container. For some patients who are already undergoing an ascitic drain (otherwise known as abdominal paracentesis) for clinical reasons, we may ask that we take a small sample of ascitic fluid (60mL) of the fluid drained that would otherwise be discarded.

If it has not been undertaken as part of your routine clinical care, we may ask you to complete a short set of written exercises called the “Psychometric Hepatic Encephalopathy Score” or “PHES” for short.

We also request your permission to access your hospital records for the purpose of the research study only, including blood tests, radiology and pathology results. All data will be anonymised and all hospital records shall be handled with strict confidentiality in accordance with the Data Protection Act 2018 and the General Data Protection Regulation (GDPR). Data will be kept for 10 years after the end of the study.

If our study is ongoing at any subsequent follow up appointments you may have, you may be asked to provide repeat breath, urine and ascitic fluid samples. Should this be the case, we will once

again ask your consent and follow the above procedures. You are once again free to decline to take part or withdraw at any time.

What will happen to my samples?
Breath within samples bags will be transferred to special tubes that will then be analysed within 48 hours. Breath samples are destroyed at the time of analysis meaning no sample will be retained for further analysis. After being collected, urine and ascitic fluid samples will be stored within the Upper Gastrointestinal Cancer Tissue Bank of the department of Surgery and Cancer (Imperial College London). At an appropriate time, urine and ascitic fluid samples will be thawed and analysed using mass spectrometry. Any surplus urine or ascitic fluid will be stored until the end of the study in case of the requirement for further analysis. Future use of these samples will be overseen by Imperial College London and will be subject to the standard procedures and regulation of tissue bank access.

What are the possible disadvantages and risks to you of taking part?

Taking part in this study, should have no direct or indirect impact on the care that you receive or the course of your treatment. No samples or test results, from which you can be identified will be shared with individuals outside of your medical team or the local research team.

What are the possible benefits of taking part?
We do not anticipate that taking part in this study will have a direct benefit for yourself, however we hope to be able to use the results of this study to improve the care and quality of life of other patients, like yourself, in the future. In the unlikely event that the analysis of your samples or measurements produces information that is directly relevant to you and your care we will be able to communicate this to the team in charge of your care. You will be required to indicate on the consent for whether you would like us to pass on any such relevant information should it arise.

What if new information becomes available?
Sometimes during the course of a research project, new information becomes available about the condition that is being studied. If this happens, the research team will inform doctor in charge of you care who will discuss with you whether you want to continue in the study. If you decide to continue in the study you will be asked to sign an updated consent form.

What if something goes wrong?
Imperial College London holds insurance policies which apply to this study. If you experience serious and enduring harm or injury as a result of taking part in this study, you may be eligible to claim compensation without having to prove that Imperial College is at fault. This does not affect your legal rights to seek compensation.

If you are harmed due to someone's negligence, then you may have grounds for a legal action. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been treated during the course of this study then you should immediately inform the chief investigator Prof. George Hanna, contact details for whom are provided below.

The normal National Health Service complaints mechanisms are also available to you such as contacting the local Patient Advice Liaison Services (PALS; pals@imperial.nhs.uk, 020 3313 0088). A member of the team will be able to give you their contact information upon request. If you are still not satisfied with the response, you may contact the Imperial College, Joint Research Compliance Office.

Will my taking part in this study be kept confidential?
All patients will be assigned a unique reference code to keep their personal details anonymised. Study associated paperwork and patient unique study identification information will be stored in a key code protected filing cabinet within a locked room within the Department of Surgery and Cancer, St Mary's Campus, Imperial College London. Access will be restricted to the chief investigator and other members of the research team only. Electronic research data will be stored

on an NHS password protected computer at St Mary's Hospital, Imperial College Healthcare NHS Trust with enforced restriction to confidential research information access.

Imperial College London 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. Imperial College London will keep identifiable information about you 10 years after the study has finished in relation to data subject consent forms and primary research data.

Further information on Imperial College London's retention periods may be found at https://www.imperial.ac.uk/media/imperial-college/administration-and-support-services/records-and-archives/public/RetentionSchedule.pdf.

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 Professor George Hanna. Tel: 02033122124; email: g.hanna@imperial.ac.uk

Legal Basis
As a university we use personally-identifiable information to conduct research to improve health, care and services. As a publicly-funded organisation, we have to ensure that it is in the public interest when we use personally-identifiable information from people who have agreed to take part in research. This means that when you agree to take part in a research study, we will use your data in the ways needed to conduct and analyse the research study.

Health and care research should serve the public interest, which means that we have to demonstrate that our research serves the interests of society as a whole. We do this by following the UK Policy Framework for Health and Social Care Research.

International Transfers
There may be a requirement to transfer information to countries outside the European Economic Area (for example, to a research partner). Where this information contains your personal data, Imperial College London will ensure that it is transferred in accordance with data protection legislation. If the data is transferred to a country which is not subject to a European Commission (**EC**) adequacy decision in respect of its data protection standards, Imperial College London will enter into a data sharing agreement with the recipient organisation that incorporates EC approved standard contractual clauses that safeguard how your personal data is processed.

Contact Us
If you wish to raise a complaint on how we have handled your personal data or if you want to find out more about how we use your information, please contact Imperial College London's Data Protection Officer via email at dpo@imperial.ac.uk, via telephone on 020 7594 3502 and via post at Imperial College London, Data Protection Officer, Faculty Building Level 4, London SW7 2AZ.

If you are not satisfied with our response or believe we are processing your personal data in a way that is not lawful you can complain to the Information Commissioner's Office (ICO). The ICO does recommend that you seek to resolve matters with the data controller (us) first before involving the regulator.

Imperial College London and Imperial College Healthcare NHS Trust will collect information from you and your medical records for this research study in accordance with our instructions.

Imperial College London and Imperial College Healthcare NHS Trust will keep your name, and contact details including age confidential. Imperial College Healthcare NHS Trust will use this information as needed, 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. Certain individuals from Imperial College London and regulatory organisations may look at your medical and research records to check the accuracy of the research study. The people who analyse the information will not be able to identify you and will not be able to find out your name, age or contact details. Imperial College London will keep identifiable information about you from this study for 10 years after the study has finished.

Use of Data for Future Research

When you agree to take part in a research study, the information about your health and care may be provided to researchers running other research studies in this organisation and in other organisations. These organisations may be universities, NHS organisations or companies involved in health and care research in this country or abroad. Your information will only be used by organisations and researchers to conduct research in accordance with the UK Policy Framework for Health and Social Care Research.

This information will not identify you and will not be combined with other information in a way that could identify you. The information will only be used for the purpose of health and care research, and cannot be used to contact you or to affect your care. It will not be used to make decisions about future services available to you, such as insurance."

What will happen to the results of the research study?
The results are likely to be published in the six months following completion of the study. Your confidentiality will be ensured at all times and you will not be identified in any publication. At the end of the study, the results of the study can be made available to you should you wish. You will be required to indicate on the consent for whether you would like us to receive a summary of the overall results of this study when this becomes available.

Who is organising and funding the research?
This study is organised by Imperial College London as part of a PhD, Department of Surgery and Cancer, Imperial College London and it is funded by HCA Ltd

Who has reviewed the study?

The Scientific quality of this study was reviewed by independent experts within Imperial College London peer review team This study was reviewed by Health Research Authority (HRA). This study was given a favourable ethical opinion for conduct in the NHS by North West - Liverpool East Research Ethics Committee (REC).

Contact for Further Information

Alternatively, you can seek impartial advice from the Patient Advice and Liaison Service (PALS), at PALS, Ground Floor, QEQM Wing, St Mary's Hospital, London, W2 1NY. Tel: 020 7886 7777, Fax: 020 7886 1753

Lastly, the trust Research and Development service provide an additional point of contact: Ms Ruth Nicholson, Research Governance Manager, Room 221, Level 2, Medical School Building, Norfolk Place, London W2 1PG. Tel: 020759 41862

17.7.3 Case Report Form For VOCAL

Volatile Organic Compounds for the Assessment of Liver Disease (VOCAL Study)

Patient Study ID: …….. **Site:** ……

Tube ID(s)……

Date:…… **Time**:……

Weight:…… **Height:**…… **BMI:**……

Age…… **Gender:** Female ☐ Male ☐ Other ☐

Ethnic Origin:
White ☐ Black / African / Caribbean / Black British ☐
Asian / Asian British ☐ Arab ☐
Other Ethnicity (Please State) ……

Specific Diet?: N/A ☐ Pescatarian ☐ Vegetarian ☐ Vegan ☐ Gluten Free ☐ Other……

How long ago did you last eat?……

What have you eaten in the last 12 hours?

Breakfast	
Lunch	
Dinner	
Other	

Past Medical History:

Confirmed Liver Disease? Y/N Underlying Aetiology……

Previous Cancer ☐ Barrett's Oesophagus ☐ Heart Disease ☐
Diabetes ☐ Kidney Disease ☐
Hypertension ☐ Asthma ☐ COPD/Emphysema ☐

Other……

If previous cancer, which type?……

Current Medications:
Omeprazole/Lansoprazole ☐ Ranitidine ☐ Aspirin/Ibuprofen/Nurofen ☐ Clopidogrel ☐
Blood Thinner e.g Warfarin ☐ Laxatives ☐ Antibiotics inc. Rifaximin ☐
Other Medications:

Previous Abdominal Surgery

Smoking:
Never ☐ Ex ☐ Current ☐ Cigarettes/Day……

Alcohol
Never ☐ Ex ☐ Current ☐ Units/Week……

VOCAL Study CRF 1.2 13.01.21

Recent Investigations:

Ultrasound Abdomen:

Fibroscan:

CT/MRI:

Bloods:

Test	Result	Date
Bilirubin		
ALT		
ALP		
AST		
GGT		
INR		
Prothrombin Time (PT)		
Platelets		
AFP		
Albumin		
Sodium		
Creatinine		
Other e.g ELF, Hepatitis Serology / Viral Load if established viral Hepatitis		

Liver Biopsy? Date:

Recent Endoscopy? Date:

Known Varices – Y/N Previous Variceal Bleed – Y/N

Previous Hospital Admission due to Liver Disease / Decompensating Episode?

PHES Score (if available): Date:

Child's Pugh Score:

17.7.4 Approved Protocol for VOCAL

Imperial College Healthcare NHS Trust | Imperial College London

Joint Research Compliance Office

Volatile Organic Compounds for the Assessment of Liver Disease (VOCAL)

Assessment of Hepatobiliary Disease through Non-Invasive Detection of Exhaled Volatile Organic Compounds

Version 1.3 09/05/2022

MAIN SPONSOR: Imperial College London
STUDY COORDINATION CENTRE: St Mary's Hospital

IRAS Project ID: 254249
REC Reference: 19/NW/0553

Study Management Group

Chief Investigator: Professor George Hanna

Co-investigators: Professor Mark Thursz

Statistician: Dr Alberto Vidal-Diez

Study Management: Dr Michael Hewitt, Mr Piers Boshier, Dr Ilaria Belluomo

Study Coordination Centre

For general queries, supply of study documentation, and collection of data, please contact:

Study Coordinator: Dr Michael Hewitt
Address: 10th Floor QEQM Building, St. Mary's Hospital, Praed Street, London, W2 1NY
E-mail: vocalstudy@imperial.ac.uk

Clinical Queries

Clinical queries should be directed to Dr Michael Hewitt who will direct the query to the appropriate person.

Sponsor

Imperial College London is the main research Sponsor for this study. For further information regarding the sponsorship conditions, please contact the Head of Regulatory Compliance at:

Joint Research Compliance Office
Imperial College London and Imperial College Healthcare NHS Trust
Room 215, Level 2, Medical School Building
Norfolk Place
London, W2 1PG
Tel: **0207 594 9459/ 0207 594 1862**
http://www3.imperial.ac.uk/clinicalresearchgovernanceoffice

Funder

Equipment utilised for sample collection will be funded as part of Professor Hanna's research funding. Laboratory analysis of the biological specimens (breath, urine and ascitic fluid samples) obtained will also be funded through Professor Hanna's research group with external funding from HCA Ltd.

This protocol describes the VOCAL study and provides information about procedures for entering participants. Every care was taken in its drafting, but corrections or amendments may be necessary. These will be circulated to investigators in the study. Problems relating to this study should be referred, in the first instance, to the Chief Investigator.

This study will adhere to the principles outlined in the UK Policy Frame Work for Health and Social Care Research. It will be conducted in compliance with the protocol, the Data Protection Act and other regulatory requirements as appropriate.

Table of Contents

GLOSSARY OF ABBREVIATIONS

ELF	Enhanced Liver Fibrosis Test
CLD	Chronic Liver Disease
GC	Gas Chromatography
HCC	Hepatocellular Carcinoma
HRA	Health Research Authority
LFT	Liver Function Test
MELD	Model for End Stage Liver Disease
MRE	Magnetic Resonance Elastography
MS	Mass Spectrometry
NAFLD	Non-Alcoholic Fatty Liver Disease
PHES	Psychometric Hepatic Encephalopathy Score
PTR	Proton Transfer Reaction
REC	Research Ethics Committee
SIFT	Selected Ion Flow Tube
TD	Thermal Desorption
VOC	Volatile Organic Compounds

KEYWORDS
Liver, Cirrhosis, Fibrosis, Hepatocellular Carcinoma, Mass Spectrometry, Volatile Organic Compounds

STUDY SUMMARY

TITLE	Assessment of hepatobiliary disease through non-invasive detection of exhaled volatile organic compounds
DESIGN	Cross-sectional and cohort observation study
AIMS	To determine the diagnostic accuracy of volatile organic compounds in exhaled breath for the detection of liver disease
RESEARCH SITE	Imperial College Healthcare NHS Trust
INCLUSION CRITERA	Patients aged 18 - 90 years; Patients seen in secondary care with suspected liver disease, confirmed liver disease or hepatocellular carcinoma; Patients able to understand and retain the information provided, thereby being able to give informed consent for inclusion in this study.
DURATION	1 June 2019 – 1 June 2025

1. INTRODUCTION

1.1 BACKGROUND

While most diseases are demonstrating improved mortality and morbidity, chronic liver disease is demonstrating the opposite: with increasing obesity rates and increasing alcohol consumption, chronic liver disease and cirrhosis has demonstrated a 400% increase in mortality rate since the 1970s in the UK population(1) representing a significant financial burden to the NHS. Multiple underlying aetiologies, a lack of confidence among clinicians and unclear diagnostic pathways along with late presentation of clinical signs are all contributing towards this trend.

Chronic inflammation of the liver due to steatosis, alcohol consumption and viral hepatitides amongst other causes result in extracellular matrix deposition and development of fibrosis(2). Early stages of fibrosis have been shown to be reversible once the insult is removed due to the regenerative capacity of the liver, but ongoing insult and inflammation will result in development of fibrosis which can then progress to irreversible cirrhosis.

Cirrhosis can be divided into four clinically significant stages according to the Baveno IV consensus workshop which divides cirrhosis on the basis of the presence or absence of complications from cirrhosis:

Stage 1: No varices or ascites
Stage 2: Varices, no ascites
Stage 3: Ascites +/- Varices
Stage 4: Variceal bleeding +/- ascites(3)

Patients in stage 1 and 2 are classed as stable “compensated” states and 3 and 4 as unstable “decompensated” states with increasing mortality as patients progress through these stages. There is currently no accurate way of predicting which patients will progress through the stages and at what rate. A diagnostic stool that identifies those at higher risk of decompensation would allow targeted intervention to prevent progression and reduce mortality.

Current diagnostic methods of liver disease are inadequate. So-called “liver function tests” will usually remain within normal parameters until decompensation (4) and thus there is difficulty in diagnosing liver disease at an early stage within primary care where intervention to prevent disease progression may be possible. Up to 75% of patients with cirrhosis are diagnosed during a decompensation episode requiring hospital admission, at which point the disease is advanced and carries a poor prognosis (1) . While known alcohol consumption will often trigger screening for liver disease in primary care, the same cannot be said for hepatic steatosis: those with metabolic syndrome are often screened for cardiovascular disease but not liver disease. Earlier detection of liver disease at the fibrosis stage would allow for potential intervention (e.g. tighter diabetic control in hepatic steatosis, reduction of alcohol consumption) to prevent disease progression and allow possible regression.

Liver disease is currently assessed for by:

Histology: Histological assessment via liver biopsy is the gold standard for diagnosis and staging of fibrosis and cirrhosis but is also invasive and flawed. The sample of tissue taken represents 0.002% of the liver parenchyma(5) and given that certain aetiologies can cause patchy distribution of fibrosis and inflammation within the liver, it may underestimate the true disease stage. It is unpleasant experience for patients - 80% of patients undergoing liver biopsy report pain, 20% of which was “severe”(6). There is also risk of infection, bleeding and a 1/10,000 mortality rate(7).

Radiology: While basic ultrasound imaging can detect changes in the liver parenchyma and macroscopic appearance, it is not sensitive for detecting fibrosis. Transient elastography (e.g. “fibroscan”) has been demonstrated as a sensitive and specific imaging modality for diagnosis of fibrosis, it also has its limitations: there can be significant inter-user variability (8) and there is reduced

accuracy with increased body habitus and with fibrosis compared to cirrhosis (9). Magnetic Resonance Elastography (MRE) is another imaging modality that has been demonstrated sensitivity of up to 98% and specificity of 100%(10) in diagnosis of fibrosis, but is time consuming, expensive and relatively inaccessible at present.

Serology: Surrogate scores such as the "MELD" score (Model for End stage Liver Disease) allow a non-invasive assessment of disease severity based on serological tests predicting their 90 day mortality. This is currently utilised in the prioritisation of liver transplants, rather than as a useful screening tool. The Enhanced Liver Fibrosis "ELF" test has been shown to have reasonable sensitivity and specificity but less accurate in discriminating between early and late stage fibrosis.

Psychometric Assessment: Hepatic encephalopathy is often a sign of a decompensated liver and while it's later stages ("overt" hepatic encephalopathy) can be detected clinically, the earlier stages ("covert" hepatic encephalopathy) are often subtler and require formal testing to elicit it. Psychometric hepatic encephalopathy scoring (PHES) is one of several paper and pen tests that can be used to diagnose covert hepatic encephalopathy, but can take up to 20 minutes per patient to complete accurately(11).

Patients with cirrhosis, depending on the underlying aetiology, have an up to 8% annual risk of developing hepatocellular carcinoma (HCC). Those who smoke, consume alcohol or have diabetes mellitus have an enhanced risk (12) . All type liver cancer is a major public health issue, ranking 6th worldwide in cancer incidence and 4th for cancer mortality(13). Early diagnosis is vital for improving survival rates.

While there is no national screening programme for HCC, current NICE guidance recommends costly and time consuming 6-monthly ultrasound imaging with or without alpha-fetoprotein testing as screening for HCC (14) in patient with known cirrhosis (or significant fibrosis in the context of chronic hepatitis B infection). However, a recent meta-analysis demonstrated ultrasound has poor sensitivity (47%) in detecting early stage HCC(15). A recent nationwide survey also identified significant variation in the provision of this recommendation, with many centres having no local screening programme and difficulty in accessing radiology services to provide what was often perceived as a service that was "not a priority"(16). The same survey identified that 60% of HCC cases discussed at MDT were inoperable at the time of diagnosis.

At Imperial College London, we have developed a breath test for the detection of oesophagogastric adenocarcinoma from biomarker discovery to clinical trials. The assessment of volatile organic compound (VOCs) within exhaled breath by selected ion flow tube mass spectrometry (SIFT-MS) (17) was followed by the construction and validation of a diagnostic model using a panel of 9 VOCs in 225 patients with an area under the ROC curve (±SE) of 0.92±0.01 and 0.87±0.03 for the model and validation subsets respectively(18). We also performed cross-platform chemical validation using GC-MS. A NIHR-funded multicentre blind validation study on 396 patients (19) showed a sensitivity of 86% and specificity of 80%. In parallel with this work we have also investigated the exhalation kinetics and molecular drivers of VOCs in oesophagogastric adenocarcinoma (20-22). Work is currently ongoing to determine if a similar diagnostic model could be constructed for use in diagnosis of colorectal cancer. Initial results are suggestive that there are discriminatory VOCs between polyps and colorectal cancer(23).

Volatile Organic Compounds (VOCs)

VOCs are carbon based compounds that are gaseous at room temperature. When emitted from the human body, they can be detected in breath, stool and urine. They have been of interest to researchers for several decades. In 1971, Pauling et al (24), reported that breath and urine contained approximately 250 and 280 VOCs respectively in normal human subjects. Analysis of these VOCs in their various forms have been integrated into several common place clinical investigations including ^{13}C urea breath test for H. pylori and the hydrogen/methane test for small intestinal bacterial overgrowth.

Preliminary work into VOCs in the context of liver disease have raised the possibility of variation in profiles of VOCs in individuals with liver disease. VOCs related to the liver travel within the

bloodstream to reach pulmonary alveoli where they are exhaled and can be quantified. A study by Fernandez Del Rio et al (25) demonstrated significant difference in the levels of limonene, methanol and 2-pentatone in cirrhotic patients compared to healthy controls. Alkhouri et al (26) identified isoprene as a potential predictor of fibrosis in non-alcoholic fatty liver disease and sulphur based compounds have also been implicated (27). All of these studies involved small patient cohorts and require validation. While there has been minimal published work to date in the role of VOCs in hepatocellular carcinoma, there has been reports in using VOCs in the detection of other cancers, such a lung, breast, bladder and prostate(28,29).

These VOCs have been identified using a multitude of mass spectrometry based techniques including Proton Transfer Reaction-Mass Spectrometry (PTR-MS), Selected Ion Flow Tube-Mass Spectrometry (SIFT-MS) and Gas Chromatography in conjunction with Mass Spectrometry (GC-MS). These techniques are complimentary and have their own strengths and limitations. Using multiple techniques for analysis allows both identification and quantification of ions producing a more robust data set.

GC-MS can be couple with thermal-desorption (TD) for analysis of breath VOCs. Pre-concentration of breath samples on to specially designed TD tubes allows for identification and partial quantification of VOCs in breath samples. The technique is capable of separating and detecting hundreds of compounds. SIFT-MS allows real time sampling of breath. Patients can breathe directly into the SIFT-MS equipment which then allows real-time detection and quantification. No pre-concentration or sample preparation is required.

Proton-Transfer-Reaction Time-of-Flight Mass Spectrometry (PTR-ToF-MS) is a relatively new technique that is gaining interest in breath research applications. Similar to SIFT-MS, Proton-Transfer-Reaction Mass Spectrometry (PTR-MS) is a direct injection-based technology, removing the problems related to pre-concentration and separation steps typical of GC-MS. The most important advantage guaranteed by the ToF mass analyzer over quadrupole-based instruments (GC-MS and SIFT-MS) is represented by the high time resolution. In a ToF analyser a complete mass spectrum is generated at each ionisation event and spectra can be recorded at a frequency of 1 Hz or lower. This improves the accuracy of measurement during *in vivo* studies; moreover, full spectral scans can be performed without compromising sensitivity, thus enabling simultaneous targeted and untargeted analysis.

The use of multi-platform analytical methods will enable both compound identification and quantification which are very important in the search and validation of biomarkers.

1.2 RATIONALE FOR CURRENT STUDY

Liver biopsy remains the gold standard investigation for diagnosis of chronic liver disease and cirrhosis. Liver biopsies are expensive, invasive, unpleasant for patients and have an associated mortality.

Abdominal ultrasound remains the first line modality for patients for hepatocellular carcinoma but as discussed previously, has a low sensitivity for detecting early hepatocellular carcinoma and current UK statistics demonstrate we are not diagnosing these patients early enough for the intervention. There is therefore much need for a non-invasive investigation that can screen for liver disease and hepatocellular carcinoma earlier in their disease processes to allow prompt diagnosis and improved treatment outcomes.

A breath-test is a non-invasive investigation and could be used to screen patients for risk of liver fibrosis, cirrhosis and hepatocellular carcinoma. It could also potentially be used for follow up of patients with early fibrosis for surveillance of disease progression to severe fibrosis and cirrhosis. If utilized in primary care, it could increase the number of appropriate referrals into secondary care. Given traditional liver function tests can remain normal until liver disease is advanced, it would allow an earlier screen for liver disease in those with risk factors (e.g. excessive alcohol consumption, metabolic syndrome). If a difference in VOC profiles between cirrhotic patients with and without hepatocellular carcinoma, it would allow potential screening in secondary care for development of HCCs without the costly and time consuming 6 monthly ultrasound imaging. Other potential uses include assessment of liver function pre and post clearance of a viral hepatitis.

2. STUDY OBJECTIVES

Principle Aim

- To determine the diagnostic accuracy of exhaled breath, urine and ascitic fluid volatile organic compounds (VOC) for the detection of hepatobiliary disease.

Objectives

- Perform a cross platform validation of volatile organic compounds measurement;
- Compare the VOC profile in patients with liver fibrosis, compensated liver cirrhosis, decompensated liver cirrhosis and hepatocellular carcinoma;
- Compare the VOC profile in patients who have received treatment for liver disease;
- Identify the metabolic pathways underlying any change in VOC profiles within liver disease patients.
- To assess the acceptability of using breath, urine and ascitic fluid testing

3. STUDY DESIGN

This is a prospective cross-sectional study aimed at determining the diagnostic accuracy of exhaled breath volatile organic compounds (VOC) for the detection of liver disease.

Patients will be recruited from outpatient clinics and hospital wards based across Imperial College Healthcare Trust. They may also be recruited from the radiology department when attending for liver biopsies or hepatocellular carcinoma surveillance ultrasound scans. When attending for their appointment, a member of the research team will discuss the research project with the patient, explain that is requires the patient to provide a breath +/- urine sample. Patients will be provided with a patient information leaflet. If the patient is to agree to this, they will be asked to sign a written consent form. If a recent PHES test has not been performed, one will be conducted otherwise no additional investigations, invasive or otherwise will be required of the patients. If patients are also undergoing planned abdominal paracentesis for drainage or sampling of ascitic fluid (for reasons outside of the study), patients will also be asked if we could take 60mls of fluid that would otherwise be routinely discarded.

For patients under regular follow up within the trust, they may be asked to consent to further breath samples being taken at future appointments to facilitate longitudinal follow up. Patients will be entitled to withdraw at any point.

Patients attending for endoscopy, fibroscan or liver biopsy will have been nil by mouth for 4-6 hours prior to the procedure as per local guidelines. For those attending outpatient appointments, if identified prior to their appointment by their clinical team, we will aim to contact the patients and request that they fast. Should they not be fasted, we may be asked to perform basic oral hygiene techniques (e.g. rinsing mouth with water), prior to the breath sample being taken. A food diary of the preceding 24 hours will be taken as part of the pre-sampling questionnaire.

As part of the consent process, patients will also be asked permission to access their hospital records for the purpose of the research study only, including their blood tests, radiology and pathology results. All data will be anonymised and all hospital records shall be handled with strict confidentiality in accordance with the Data Protection Act 1998 and the General Data Protection Regulation (GDPR).

Patients will be asked to give a number of exhalations into a breath sampling device until a defined volume of breath has been sampled. This may be done via one of several techniques. Our preferred option is that they breathe into a custom designed nalophan bag which is attached to thermal desorption tubes and a breath sampling device. Alternatively, patient may be asked to perform normal tidal breathing whilst wearing a face mask, which takes around 2 minutes on average. Patients located at St Mary's Hospital may also be asked to breathe directly into a mass spectrometer machine (SYFT

Voice200Ultra). This mass spectrometer instrument is a self-contained mobile device which can be moved between clinical areas within St Mary's Hospital. Sampling by this method will involve a single deep nasal inhalation followed by complete exhalation via their mouth into a plastic cylinder connected to a tube leading to the mass spectrometer itself. This is repeated three times within a 60 second window. We expect the breath sampling process to take no longer than 5-10 minutes in total. Patients will be requested to pass urine into a standardised 60ml urine specimen vial, which will be immediately sealed. 20ml of urine will be aliquoted into a standard 60ml specimen vial. Any ascitic fluid collected will also be collected into a standardised 60ml vial and sealed.

The VOCs from the breath will be trapped onto adsorbent tubes inserted into the breath taking device. After sampling the adsorbent tube is removed from the device, capped and transported back to the laboratory at St Mary's Hospital London for analysis. Samples will be transported by courier or by a member of the research team. The sample(s) shall undergo analysis via our mass spectrometry instruments, at St Mary's Hospital London.

Breath VOCs sampled on adsorbent tubes will be analysed within 48 hours of collection. The sample constituents are destroyed within the mass spectrometer for detection meaning that no clinical material will remain after analysis by the mass spectrometry instruments. Urine samples and ascitic fluid samples will be stored within a registered tissue bank (REC Ref no. 04/Q0403/119, custodian Prof G Hanna) in accordance with the Human Tissue Act and disposed of following analysis.

As part of our study, we will aim to follow patients up longitudinally to ascertain if a participant's VOC profile changes with progression or regression of liver disease. At each appointment, before sample is taken, the participant's consent will be reviewed and if the participant is still willing to continue in the study, further breath, urine and ascitic fluid samples will then be collected. This would be for a maximum of 24 months.

This study has been specifically designed to ensure that any disruption to patient's routine care and management is kept to an absolute minimum.

3.1 SAMPLE ANALYSIS

Collected breath samples will be analysed by one or more complimentary mass spectrometry techniques:

Selective Ion Flow Tube Mass Spectrometry (SIFT–MS)
SIFT-MS permits online, real-time Volatile Organic Compound (VOC) quantification. SIFT-MS has been utilised in the study of VOCs in breath and urine from patients with conditions including cystic fibrosis and bladder cancer. The principle of SIFT-MS is selected precursor ions are formed in a microwave discharge source and are selected according to their mass-to-charge ratio, m/z, by a mass filter and injected into a helium carrier gas where they are convected as a thermalised swarm along a flow tube. H_3O^+, NO^+, O_2^+ precursor ions are used to ionize the trace gases in an air sample that is introduced into the helium at a known flow rate, these ions selectively ionise VOCs present within the sample resulting in characteristic product ions. By measuring the count rate of both precursor ions and the characteristic product ions at the downstream detection system, a real-time quantification is achieved, realising the absolute concentration of trace and volatile compounds at the parts-per-billion by volume or parts-per-million by volume. The SIFT-MS technique allows real-time detection and quantification of Volatile Organic Compounds within biological samples such as exhaled breath without any sample preparation minimising diagnostic delay. This is particularly advantageous within the clinical environment where samples can be retrieved and real-time VOC measurements made with negligible concern for sample degradation.

Thermal Desorption (TD) and Gas Chromatography Mass Spectrometry (GC/MS)
Thermal desorption coupled to GC/MS is one of the most widely used techniques for breath biomarker discovery work for a number of reasons. Firstly, VOCs in exhaled breath are present at very low concentrations of parts-per-million (ppm) to parts-per-trillion (ppt) and pre-concentration is often necessary to allow for detection. This can be achieved by adsorption onto sorbent beds (tubes packed with tenax, carbon molecular sieve, or graphitized carbon) followed by release through thermal

desorption. The coupling of a thermal desorber to a GC/MS system allows for the trapped VOCs to be released and rapidly injected on to the GC column for analysis. GC/MS is a very well established analytical technique that allows for the successful separation, identification, and semi-quantification of trace VOCs in unknown complex mixtures such as breath. This technique has allowed for the identification of over 800 compounds in breath. Although this technique does not allow for real-time analysis it provides a much more comprehensive analysis of the VOC profile. It will complement results obtained with SIFT-MS and can allow for cross-platform validation.

Proton-Transfer-Reaction Time-of-Flight Mass Spectrometry (PTR-ToF-MS)
Proton-Transfer-Reaction Mass Spectrometry (PTR-MS) is a direct injection-based technology, obviating problems related to pre-concentration and separation steps typical of GC-MS. In PTR-MS the ion beam generated at the source does not undergo selection by means of a mass filter as in SIFT-MS so the detection limits are on average two orders of magnitude lower than those obtained with SIFT-MS. However, equipped with a Time-of-Flight (ToF) mass analyzer, a higher mass resolution is achieved, signals represented by a single mass peak in a quadrupole-generated spectrum, can be resolved in two or more distinct mass peaks, generally resulting in an improved discrimination performance. In addition to this, accurate masses can be determined with at least two decimal digits; this means that the data generated by PTR-ToF-MS, especially when integrated with GC-MS, can be used for compound identification and thus permit the discovery of new biomarkers. The most important advantage guaranteed by the ToF mass analyzer over quadrupole-based instruments (as those in GC-MS and SIFT-MS) is represented by the high time resolution. In a ToF analyzer a complete mass spectrum is generated at each ionisation event and spectra can be recorded at a frequency of 1 Hz or lower. This improves the accuracy of measurement during *in vivo* studies; moreover, full spectral scans can be performed without compromising sensitivity, thus enabling simultaneous targeted and untargeted analysis.

Two-Dimensional Gas Chromatography (GCxGC MS)
Two-dimensional gas chromatography connects two separate gas chromatography columns (usually one-polar and one non-polar) allowing significantly increased separation of compounds, especially of those lower in abundance, compared to single GC-MS. It has been shown to detect more volatile organic compounds per breath sample than previous techniques and is viable as a high throughput technique for breath analysis(30,31) . It can also be successfully utilised with thermal desorption.

4. PARTICIPANT ENTRY

Patients will be identified and recruited from Imperial College Healthcare NHS Trust inpatient admissions, routine outpatient hepatobiliary clinics, the radiology departments (e.g. for patients undergoing liver biopsy or imaging of their liver) and endoscopy units (e.g. liver disease patients undergoing endoscopic surveillance for varices). Patients with hepatocellular carcinoma may also be recruited from oncology outpatient clinics.

Patients that are recruited will be divided into the following cohorts:

1. Healthy Controls with or without risk factors for Chronic Liver Disease (i.e. patients referred due to concerns re: liver disease but found to have no evidence of chronic liver disease following assessment)
2. Patients with Liver Fibrosis
3. Patients with Compensated (i.e. stable) Cirrhosis
4. Patients with Decompensated (i.e. unstable) Cirrhosis
5. Patients with Hepatocellular Carcinoma

We will aim initially to recruit 200 patients from each cohort (i.e. 1200 patients in total).

A member of the direct care team will first approach patients in the various settings and discuss the study with them. Patients will be informed that participation is voluntary and that if they do not want to participate that it will not affect their usual care. Patients will also be informed that even if they take part, they can withdraw at any time without giving any reasons and without affecting their usual care

and treatment. Patients who are willing to participate will be asked by a member of direct care team for permission for their details to be forwarded to members of the study research team.

A member of the research team will then approach patients after receiving information on their willingness to participate from the direct care team. This will be a study team member who is GCP trained with experience in taking consent. This member of the study team will discuss the study with patients and re-emphasise that participation is voluntary and even if they do not want to participate it will not affect their usual care and treatment and that even when they agreed to take part they can still withdraw from the study at any time without giving any reason. Patients will be allowed at least 30 minutes to decide on participation so that same day participation can be facilitated. Should the patient require longer to decide, arrangements can be made for the patient to return at a later date to participate should they wish to.

They will be given opportunity to ask questions and asked to sign consent form if their queries are fully addressed and they are willing to participate. Patients will then be assigned unique identification code to keep their personal details anonymised.

4.1 INCLUSION CRITERIA

- Patients aged 18 - 90 years
- Patients seen in secondary care with suspected or confirmed liver / pancreatic disease or hepatocellular carcinoma / pancreatic cancer
- Patients able to understand and retain the information provided, thereby being able to give informed consent for inclusion in this study

4.2 EXCLUSION CRITERIA

- Patients who lack capacity or unable to provide informed consent.
- Any patient outside the established age range (18-90years).
- Patients unable to fast.

4.3 WITHDRAWAL CRITERIA

- Withdrawal of consent from patient.

5. ADVERSE EVENTS

5.1 DEFINITIONS

Adverse Event (AE): any untoward medical occurrence in a patient or clinical study subject.

Serious Adverse Event (SAE): any untoward and unexpected medical occurrence or effect that:

- **Results in death**
- **Is life-threatening** – *refers to an event in which the subject was at risk of death at the time of the event; it does not refer to an event which hypothetically might have caused death if it were more severe*
- **Requires hospitalisation, or prolongation of existing inpatients' hospitalisation**
- **Results in persistent or significant disability or incapacity**
- **Results in a congenital anomaly or birth defect**

We do not believe any risks or adverse events are associated with patients providing samples of urine, ascitic fluid and breath or in the consent process for these procedures. All samples will be disposed of following analysis. All hospital records will be handled with strict confidentiality in accordance with the Data Protection Act 1998. All other risks will be in accordance with usual clinical practice.

5.2 REPORTING PROCEDURES

Given the nature of this investigation, it is unlikely that any will occur. Nonetheless, should an unexpected adverse event occur, the investigators will make an assessment of severity and report the event to the Research Ethics Committee and the Sponsor as appropriate. Usually, this will mean within 24 hours in the case of severe or moderate events, and within 14 days for mild events.

5.2.1 Non Serious AEs

All such events, whether expected or not, should be recorded.

5.2.2 Serious AEs

An SAE form will be completed and faxed to the Chief Investigator within 24 hours. However, relapse and death due to liver disease, and hospitalisations for elective treatment of a pre-existing condition do not need reporting as SAEs.

All SAEs will be reported to the Research Ethics Committee where in the opinion of the Chief Investigator, the event was:

- 'related', i.e. resulted from the administration of any of the research procedures; and
- 'unexpected', i.e. an event that is not listed in the protocol as an expected occurrence

Reports of related and unexpected SAEs will be submitted within 15 days of the Chief Investigator becoming aware of the event, using the NRES SAE form for non-IMP studies. The Chief Investigator will also notify the Sponsor of all SAEs.

Local investigators will report any SAEs as required by their Local Research Ethics Committee, Sponsor and/or Research & Development Office.

Contact details for reporting SAEs
jrco@imperial.ac.uk
CI email
g.hanna@imperial.ac.uk

Please send SAE forms to:
10th Floor QEQM Building, St Mary's Hospital, Praed St, London, W2 1YU
Tel: 02078862125 (Mon to Fri 09.00 – 17.00)

6. ASSESSMENT AND FOLLOW-UP

A small cohort of patients will be followed up longitudinally to see if there is any change in their VOC breath profile with disease progression or regression. Otherwise, no routine follow up is required.

7. STATISTICS AND DATA ANALYSIS

Both univariate and multivariate data analysis techniques will be applied to the results to 1) identify VOC components with the best discriminating ability between the groups and 2) to develop a multivariate discriminant analysis model.

Mann-Whitney U test will be used to compare the measured metabolite concentrations between the study groups. A p value <0.05 will be taken as the level to indicate statistical significance.

Non-parametric (Kruskal-Wallis) ANOVA test will be used to compare the measured metabolite levels between the study groups. Statistically significant metabolites will be selected using False Discovery Rate (FDR) adjusted p value threshold such that less than 5% of discovered candidate markers are false positives (i.e. q-value <0.05).

The FDR strategy will be based on Benjamini–Hochberg procedure. Receiver operating characteristic (ROC) curves will be used to determine the accuracy of a diagnostic test in classifying those with and without liver disease.

Our initial study will be a discovery phase with a pragmatic estimate of patient numbers. Following this, we will perform a formal power calculation.

Data and all appropriate documentation will be stored for a minimum of 10 years after the completion of the study, including the follow-up period.

8. REGULATORY ISSUES

8.1 ETHICS APPROVAL

Ethical approval has been received by the Research Ethics Committee (REC) and Health Regulator Authority (HRA) (reference 19/NW/0553). The study will be conducted in accordance with the recommendations for physicians involved in research on human subjects adopted by the 18th World Medical Assembly, Helsinki 1964 and later revisions. The study must also receive confirmation of capacity and capability from each participating NHS trust before accepting participants into the research.

8.2 CONSENT

A member of the direct care team will first approach patients in the various settings and discuss the study with them. Patients will be informed that participation is voluntary and that if they do not want to participate that it will not affect their usual care. Patients will also be informed that even if they take part, they can withdraw at any time without giving any reasons and without affecting their usual care and treatment. Patients who are willing to participate will be asked by a member of direct care team for permission for their details to be forwarded to members of the study research team.

A member of the research team will then approach patients after receiving information on their willingness to participate from the direct care team. This will be a study team member who is GCP trained with experience in taking consent. This member of the study team will discuss the study with patients and re-emphasise that participation is voluntary and even if they do not want to participate it will not affect their usual care and treatment and that even when they agreed to take part they can still withdraw from the study at any time without giving any reason. Patients will be allowed at least 30 minutes to decide on participation so that same day participation can be facilitated. Should the patient require longer to decide, arrangements can be made for the patient to return at a later date to participate should they wish to.

They will be given opportunity to ask questions, provided with a written information leaflet (Participant Information Leaflet 1.1) and asked to sign consent form if their queries are fully addressed and they are willing to participate. Patients will then be assigned unique identification code to keep their personal details anonymised.

8.3 CONFIDENTIALITY

All patients will be assigned a unique reference code to keep their personal details anonymised. Patient consent, other study associated paperwork and patient unique study identification information will be stored in a key code protected filing cabinet within a locked room within the Department of Surgery and Cancer, St Mary's Campus, Imperial College London. Access will be restricted to the chief investigator and other members of the research team only. Electronic research data will be stored on an NHS password protected computer at St Mary's Hospital, Imperial College Healthcare NHS Trust with enforced restriction to confidential research information access. Personal addresses and contact details will also remain on the patient's personal electronic record on Imperial College Healthcare NHS Trust computer systems.

Research data will be stored securely for 10 years.

8.4 INDEMNITY
Imperial College London holds negligent harm and non-negligent harm insurance policies which apply to this study. Medical co-investigators will also be covered by their own medical defence insurance for non-negligent harm.

8.5 SPONSOR
Imperial College London will act as the main Sponsor for this study. Delegated responsibilities will be assigned to NHS trusts taking part in this study.

8.6 FUNDING
Equipment utilised for sample collection will be funded as part of Professor Hanna's research funding. Laboratory analysis of the biological specimens (breath, urine and ascitic fluid samples) obtained will also be funded through Professor Hanna's research group with external funding from HCA Hospitals.

8.7 AUDITS
The study will be subject to inspection and audit by Imperial College London under their remit as sponsor and other regulatory bodies to ensure adherence to GCP and the UK Policy Frame Work for Health and Social Care Research.

9. STUDY MANAGEMENT

The day-to-day management of the study will be co-ordinated through Dr Michael Hewitt.

10. REFERENCES

(1) Williams R, Aspinall R, Bellis M, Camps-Walsh G, Cramp M, Dhawan A, et al. Addressing liver disease in the UK: a blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity, and viral hepatitis. *Lancet (London, England).* 2014; 384 (9958): 1953-1997. Available from: doi: 10.1016/S0140-6736(14)61838-9 [Accessed Oct 1, 2018].

(2) Bonfrate L, Grattagliano I, Palasciano G and Portincasa P. Dynamic carbon 13 breath tests for the study of liver function and gastric emptying. *Gastroenterology Report.* 2015; 3 (1): 12-21. Available from: doi: 10.1093/gastro/gou068 [Accessed Sep 28, 2018].

(3) de Franchis R. Evolving Consensus in Portal Hypertension Report of the Baveno IV Consensus Workshop on methodology of diagnosis and therapy in portal hypertension. *Journal of Hepatology.* 2005; 43 (1): 167-176. Available from: doi: 10.1016/j.jhep.2005.05.009 Available from: https://www.sciencedirect.com/science/article/pii/S0168827805003041 .

(4) Cuperus FJ, Drenth JP and Tjwa ET. *Mistakes in liver function test abnormalities and how to avoid them.* Available from: https://www.ueg.eu/education/latest-news/article/article/mistakes-in-liver-function-test-abnormalities-and-how-to-avoid-them/ [Accessed Oct 1, 2018].

(5) Coral GP, Antunes ADP, Serafini APA, Araujo FB and Mattos AAd. LIVER BIOPSY: IMPORTANCE OF SPECIMEN SIZE IN THE DIAGNOSIS AND STAGING OF CHRONIC VIRAL HEPATITIS. *Revista do Instituto de Medicina Tropical de Sao Paulo.* 2016; 58 (1): 10. Available from: doi: 10.1590/S1678-9946201658010 Available from: https://www.ncbi.nlm.nih.gov/pubmed/26910447 .

(6) Sanai FM, Keeffe EB. Liver Biopsy for Histological Assessment – The Case Against. *Saudi Journal of Gastroenterology : Official Journal of the Saudi Gastroenterology Association.* 2010; 16 (2): 124-132. Available from: doi: 10.4103/1319-3767.61244 Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3016504/ [Accessed Oct 3, 2018].

(7) West J, Timothy R. Reduced Mortality Rates Following Elective Percutaneous Liver Biopsies. *Gastroenterology.* 2010; 139 (4): 1230-1237. Available from: doi: 10.1053/j.gastro.2010.06.015 Available from: https://www.clinicalkey.es/playcontent/1-s2.0-S0016508510008747 .

(8) Lupsor Platon M, Stefanescu H, Feier D, Maniu A and Badea R. Performance of unidimensional transient elastography in staging chronic hepatitis C. Results from a cohort of 1,202 biopsied patients from one single center. *Journal of gastrointestinal and liver diseases: JGLD.* 2013; 22 (2): 157-166. [Accessed Oct 1, 2018].

(9) Tsochatzis EA, Gurusamy KS, Ntaoula S, Cholongitas E, Davidson BR and Burroughs AK. Elastography for the diagnosis of severity of fibrosis in chronic liver disease: a meta-analysis of diagnostic accuracy. *Journal of Hepatology.* 2011; 54 (4): 650-659. Available from: doi: 10.1016/j.jhep.2010.07.033 [Accessed Oct 1, 2018].

(10) Venkatesh SK, Yin M and Ehman RL. Magnetic resonance elastography of liver: clinical applications. *Journal of computer assisted tomography.* 2013; 37 (6): 887-896. Available from: doi: 10.1097/RCT.0000000000000032 Available from: https://www.ncbi.nlm.nih.gov/pubmed/24270110 .

(11) Princivalle A, Monasta L, Butturini G, Bassi C and Perbellini L. Pancreatic ductal adenocarcinoma can be detected by analysis of volatile organic compounds (VOCs) in alveolar air. *BMC cancer.* 2018; 18 (1): 529. Available from: doi: 10.1186/s12885-018-4452-0 Available from: https://www.ncbi.nlm.nih.gov/pubmed/29728093 .

(12) Schwartz JM, Carrithers RL. *Epidemiology and etiologic associations of hepatocellular carcinoma.* Available from: https://www.uptodate.com/contents/epidemiology-and-etiologic-associations-of-hepatocellular-carcinoma [Accessed 1st October 2018].

(13) Akinyemiju T, Abera S, Ahmed M, Alam N, Alemayohu MA, Allen C, et al. The Burden of Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level: Results From the Global Burden of Disease Study 2015. *JAMA oncology.* 2017; 3 (12): 1683-1691. Available from: doi: 10.1001/jamaoncol.2017.3055 [Accessed Oct 3, 2018].

(14) *Cirrhosis in over 16s: assessment and management | Guidance and guidelines | NICE.* NICE. Available from: https://www.nice.org.uk/guidance/ng50. [Accessed Oct 1, 2018].

(15) Tzartzeva K, Obi J, Rich NE, Parikh ND, Marrero JA, Yopp A, et al. Surveillance Imaging and Alpha Fetoprotein for Early Detection of Hepatocellular Carcinoma in Patients With Cirrhosis: A Meta-analysis. *Gastroenterology.* 2018; 154 (6): 1718.e1. Available from: doi: 10.1053/j.gastro.2018.01.064 [Accessed Oct 3, 2018].

(16) Cross TJS, Villaneuva A, Shetty S, Wilkes E, Collins P, Adair A, et al. A national survey of the provision of ultrasound surveillance for the detection of hepatocellular carcinoma. *Frontline gastroenterology.* 2016; 7 (2): 82-89. Available from: doi: 10.1136/flgastro-2015-100617 Available from: https://www.ncbi.nlm.nih.gov/pubmed/28840911 .

(17) Kumar S, Huang J, Abbassi-Ghadi N, Španěl P, Smith D and Hanna GB. Selected ion flow tube mass spectrometry analysis of exhaled breath for volatile organic compound profiling of esophago-gastric cancer. *Analytical Chemistry.* 2013; 85 (12): 6121-6128. Available from: doi: 10.1021/ac4010309 [Accessed Oct 4, 2018].

(18) Kumar S, Huang J, Abbassi-Ghadi N, Mackenzie HA, Veselkov KA, Hoare JM, et al. Mass Spectrometric Analysis of Exhaled Breath for the Identification of Volatile Organic Compound Biomarkers in Esophageal and Gastric Adenocarcinoma. *Annals of Surgery.* 2015; 262 (6): 981-990. Available from: doi: 10.1097/SLA.0000000000001101 [Accessed Oct 4, 2018].

(19) Markar SR, Lagergren J and Hanna GB. Research protocol for a diagnostic study of non-invasive exhaled breath analysis for the prediction of oesophago-gastric cancer. *BMJ open.* 2016; 6 (1): e009139. Available from: doi: 10.1136/bmjopen-2015-009139 [Accessed Oct 4, 2018].

(20) Atonowicz S. *Aldehyde dysfunction in oesophageal adenocarcinoma.* Imperial College London; 2016.

(21) Adam M. *Role of microbiome in volatile production in oesophageal adenocarcinoma.* Imperial College London; 2018.

(22) Wiggins T. *Phenol and aromatic acid dysfunction in oesophageal adenocarcinoma.* Imperial College London; 2016.

(23) Woodfield G, Belluomo I, Lin G, Boshier P, Romano A, Martin J, et al. PTU-072 Breath testing for colorectal polyps and cancer- a load of hot air? *Abstracts.* 2018; 67 (Suppl 1): A208. Available from: https://gut.bmj.com/content/67/Suppl_1/A206.3 .

(24) Pauling L, Robinson AB, Teranishi R and Cary P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proceedings of the National Academy of Sciences of the United States of America.* 1971; 68 (10): 2374-2376. [Accessed Oct 4, 2018].

(25) Fernández Del Río R, O'Hara ME, Holt A, Pemberton P, Shah T, Whitehouse T, et al. Volatile Biomarkers in Breath Associated With Liver Cirrhosis - Comparisons of Pre- and Post-liver Transplant Breath Samples. *EBioMedicine.* 2015; 2 (9): 1243-1250. Available from: doi: 10.1016/j.ebiom.2015.07.027 [Accessed Oct 10, 2018].

(26) Alkhouri N, Singh T, Alsabbagh E, Guirguis J, Chami T, Hanouneh I, et al. Isoprene in the Exhaled Breath is a Novel Biomarker for Advanced Fibrosis in Patients with Chronic Liver Disease: A Pilot Study. *Clinical and Translational Gastroenterology.* 2015; 6 e112. Available from: doi: 10.1038/ctg.2015.40 [Accessed Oct 10, 2018].

(27) Tangerman A, Meuwese-Arends MT and Jansen JB. Cause and composition of foetor hepaticus. *Lancet (London, England).* 1994; 343 (8895): 483. [Accessed Oct 10, 2018].

(28) Phillips M, Gleeson K, Hughes JM, Greenberg J, Cataneo RN, Baker L, et al. Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. *Lancet (London, England).* 1999; 353 (9168): 1930-1933. Available from: doi: 10.1016/S0140-6736(98)07552-7 [Accessed Oct 10, 2018].

(29) Phillips M, Cataneo RN, Ditkoff BA, Fisher P, Greenberg J, Gunawardena R, et al. Prediction of breast cancer using volatile biomarkers in the breath. *Breast Cancer Research and Treatment.* 2006; 99 (1): 19-21. Available from: doi: 10.1007/s10549-006-9176-1 [Accessed Oct 10, 2018].

(30) Wilde MJ, Cordell RL, Salman D, Zhao B, Ibrahim W, Bryant L, et al. Breath analysis by two-dimensional gas chromatography with dual flame ionisation and mass spectrometric detection – Method optimisation and integration within a large-scale clinical study. *Journal of Chromatography A.* 2019; Available from: doi: 10.1016/j.chroma.2019.02.001 Available from: https://www.sciencedirect.com/science/article/pii/S0021967319301311 .

(31) Michael Phillips, Renee N Cataneo, Anirudh Chaturvedi, Peter D Kaplan, Mark Libardoni, Mayur Mundada, et al. Detection of an Extended Human Volatome with Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry. *PLoS One.* 2013; 8 (9): e75274. Available from: doi: 10.1371/journal.pone.0075274 Available from: https://www.ncbi.nlm.nih.gov/pubmed/24086492 .

17.7.5 VOCAL Study Sponsorship

Imperial College London

Imperial College Healthcare NHS Trust

Joint Research Compliance Office
Academic Health Science Centre
Imperial College London and
Imperial College Healthcare NHS Trust
Room 215, Medical School Building
St Mary's Hospital
Pread Street
London W2 1NY
Tel: +44 (0)20 759 49459
Becky.ward@imperial.ac.uk
www.ic.ac.uk/clinicalresearchgovernanceoffice

Becky Ward
Research Governance Manager

08/12/2022

Professor George Hanna
Head of Division of Surgery
Imperial College London
St Mary's Hospital
Praed Street
London W1 1NY

Dear George

RE: Volatile Organic Compounds for Assessment of Liver disease [VOCAL]

Joint Research Compliance Office Reference number: 19SM5129

This is to confirm that the above named research project utilises human participants, their organs, tissue and/or data as defined under the sponsorship requirements of the UK policy framework for health and social care, incorporating the Medicines for Human Use (Clinical Trials) Regulations 2004.

On behalf of Imperial College of Science, Technology and Medicine, we undertake to act as the identified Research Sponsor for this project.

This letter confirms:

- The research proposal has been discussed, assessed and registered with the Joint Research Compliance Office, Imperial College Academic Health Science Centre, Imperial College London and provisional sponsor approval granted.
- The Chief Investigator has undergone a process of scientific critique commensurate with the scale of the project.
- Indemnity and insurance arrangements have been put in place to cover the project.
- Resources and support are available to the research team to aid delivery of the research as proposed.
- Management, monitoring and reporting responsibilities for the research have been approved.
- Imperial College will undertake and enforce those sponsor duties set out in the UK policy Framework for Health and Social Care.

Imperial College Sponsorship is conditional on the project receiving applicable ethical and regulatory approval for all research related aspects of its conduct. It is also conditional on successful contract and agreement negotiations and sign off via the Joint Research Office, where relevant, and before the study commences.

A copy of the ethics approval letter **must** be sent to the Research Governance Manager prior to the study commencing. Sponsorship is dependant on obtaining R&D Office approval for all NHS sites where the research is being conducted.

Yours sincerely

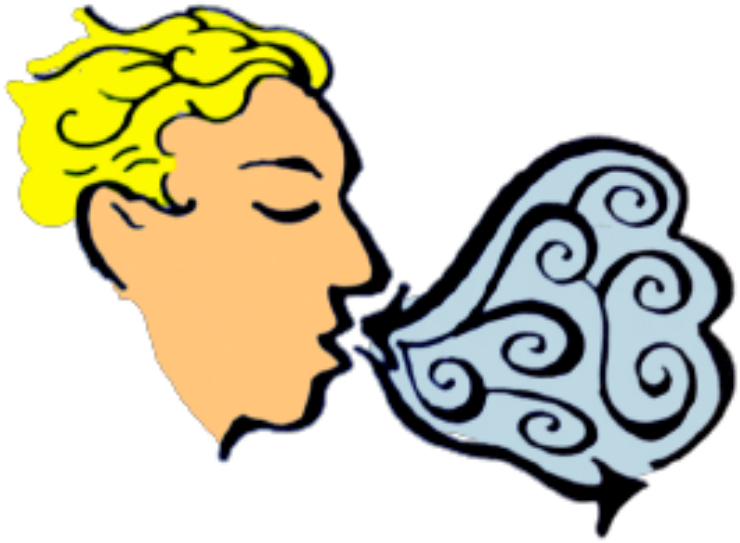
17.7.6 VOCAL Recruitment Poster

Can you spare two breaths?

We are currently recruiting participants for a new study looking into breath analysis to assess liver health.

It should require no more than 5 minutes of your time.

Please ask for more information.



Dr Michael Hewitt
Clinical Research Fellow
mjh206@imperial.ac.uk

17.7.7 VOCAL Data Scaling Strategies

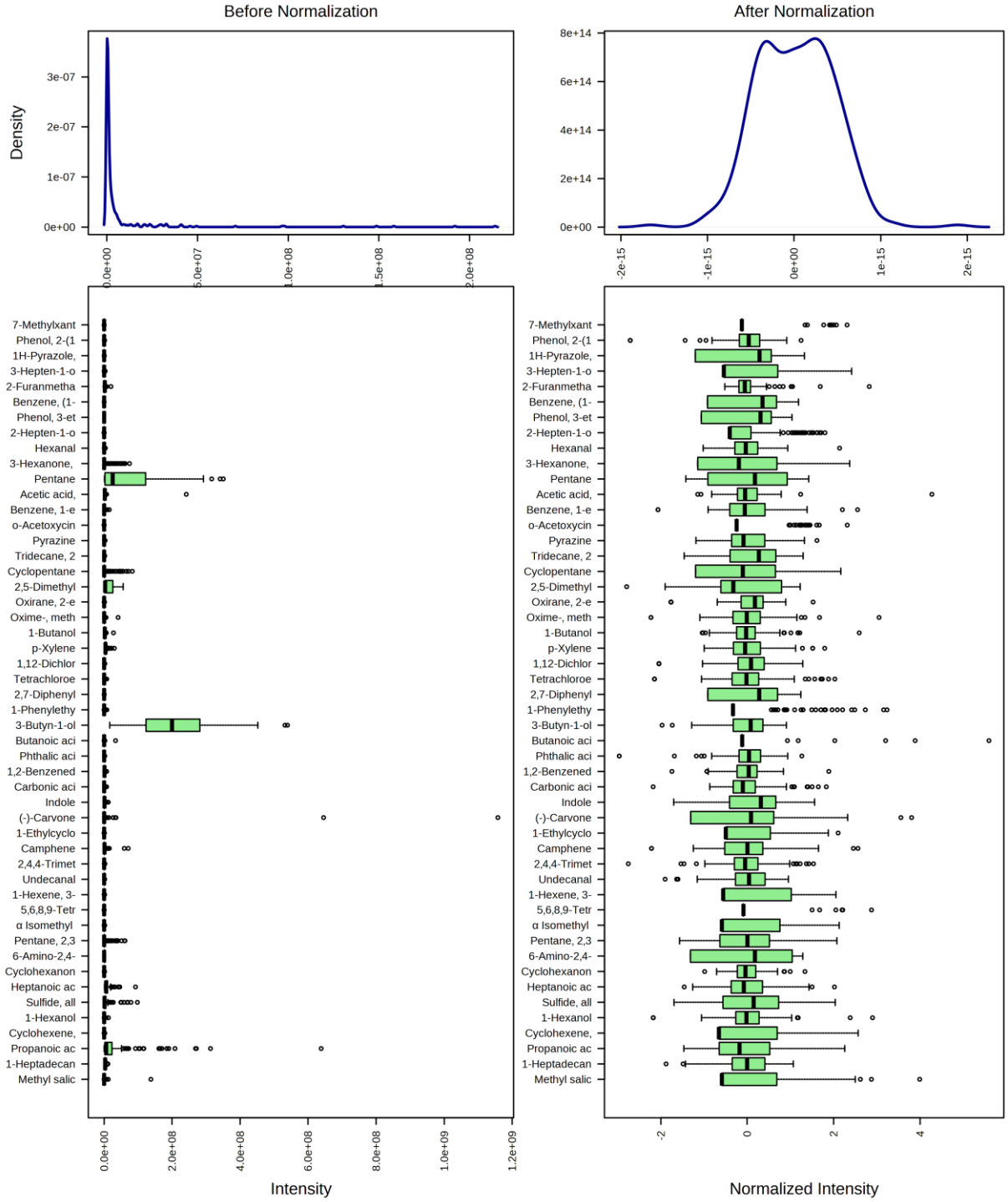


Figure 17.13 VOCAL GC-MS data set pre- and post-Pareto scaling.

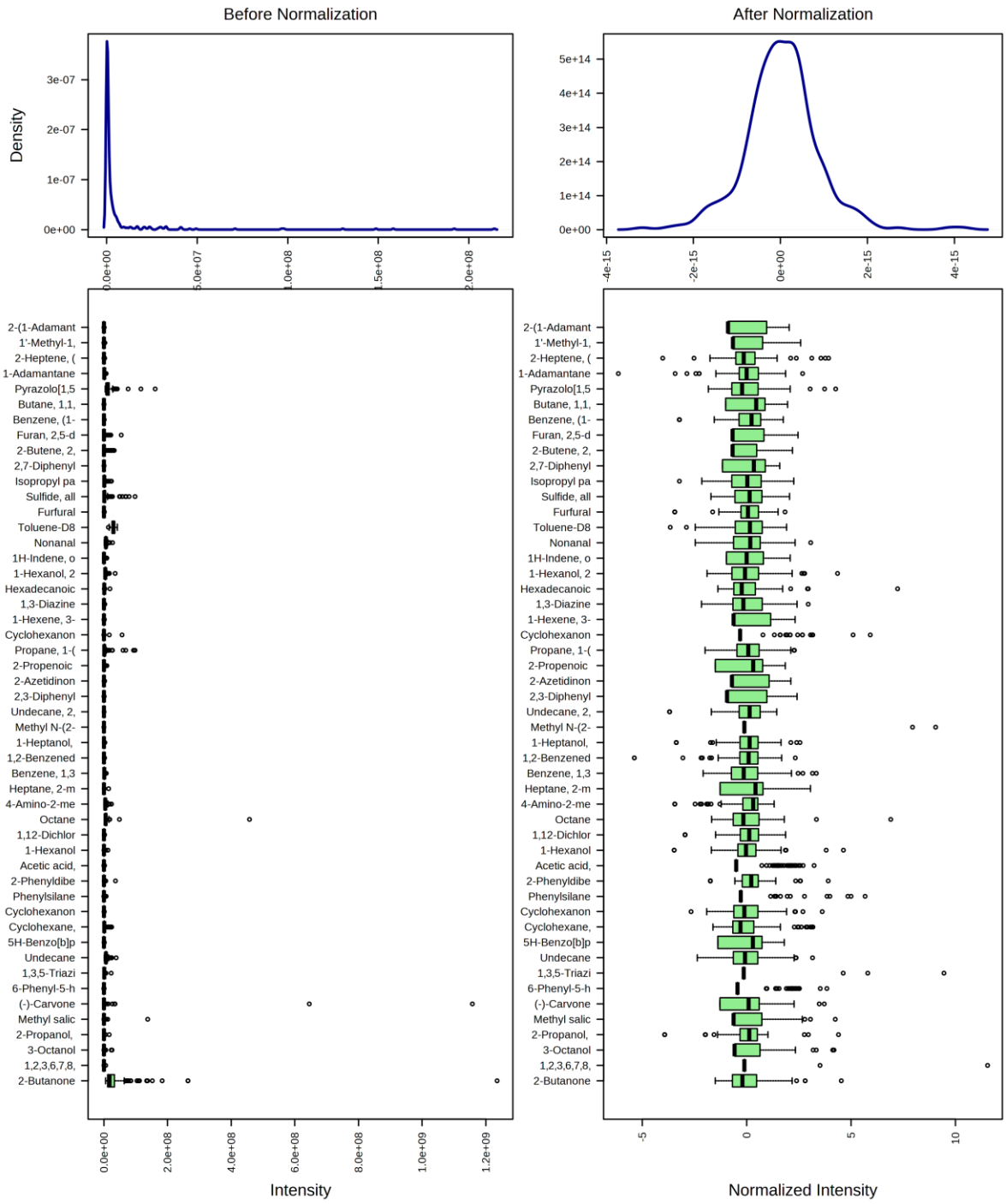


Figure 17.14 VOCAL GC-MS data set pre- and post-unit variance (UV) scaling.

17.7.8 VOCAL Data Normalisation Strategies

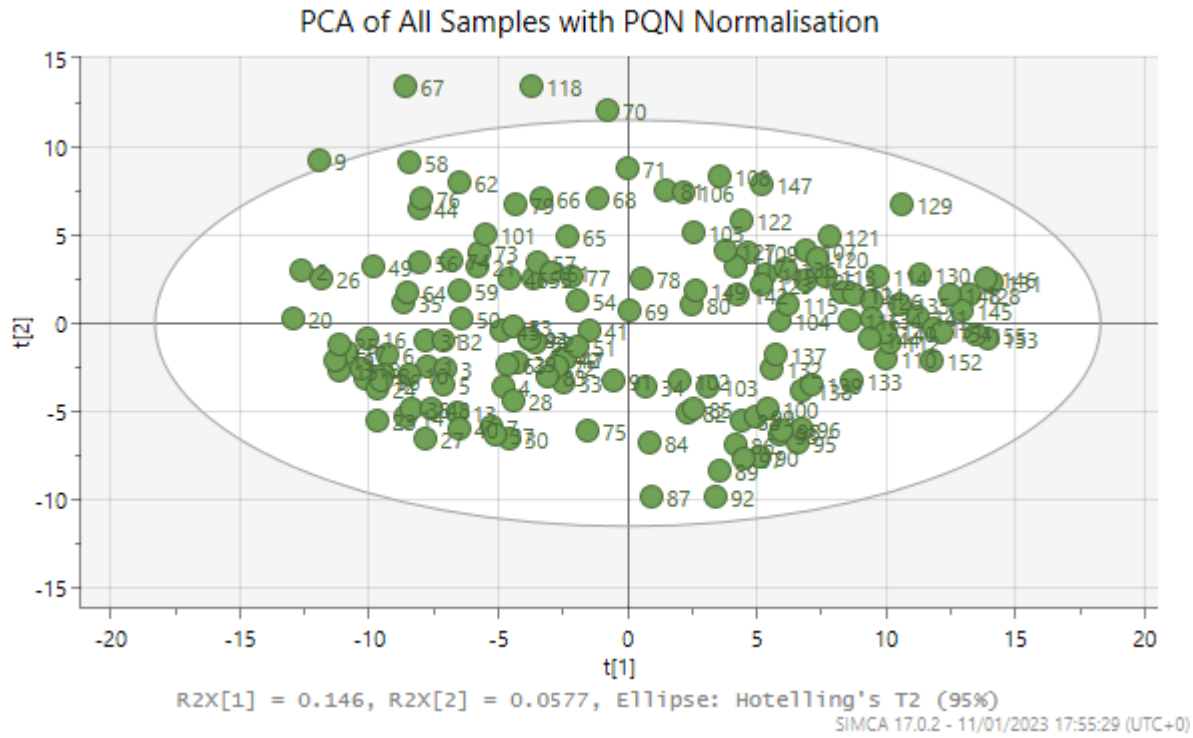


Figure 17.15 PCA of all VOCAL samples with PQN normalisation.

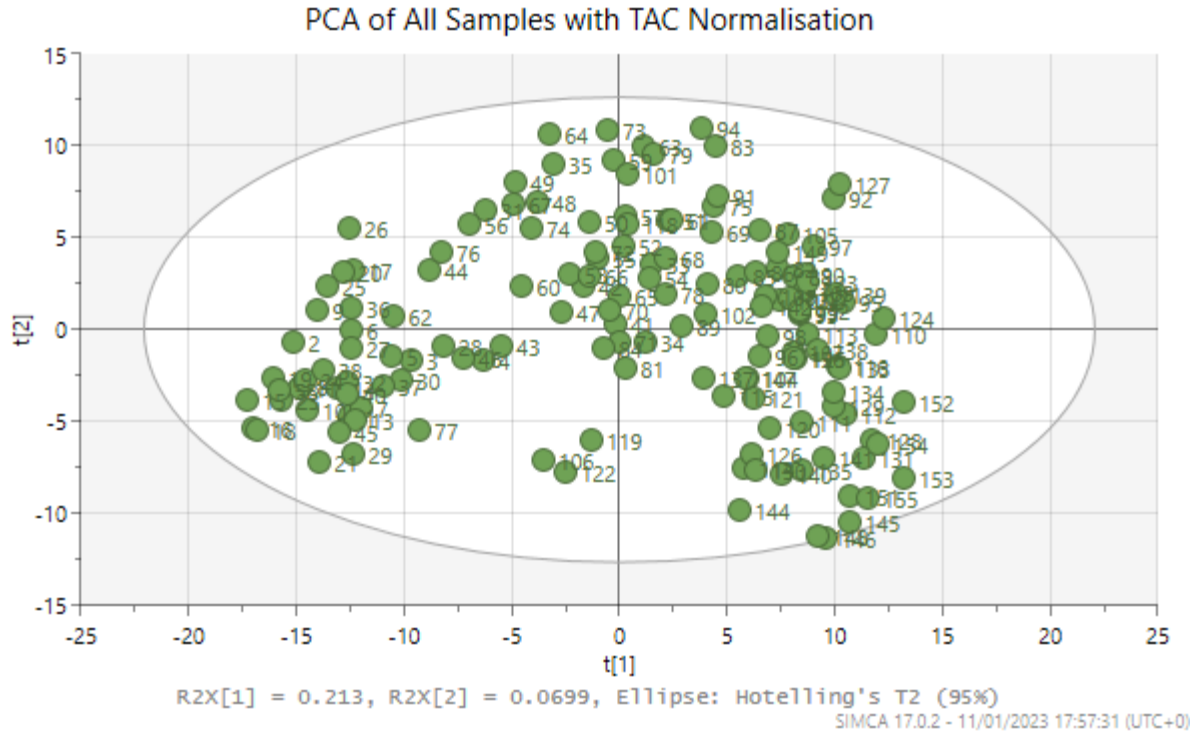


Figure 17.16 PCA of all VOCAL samples with normalisation by total area correction (TAC).

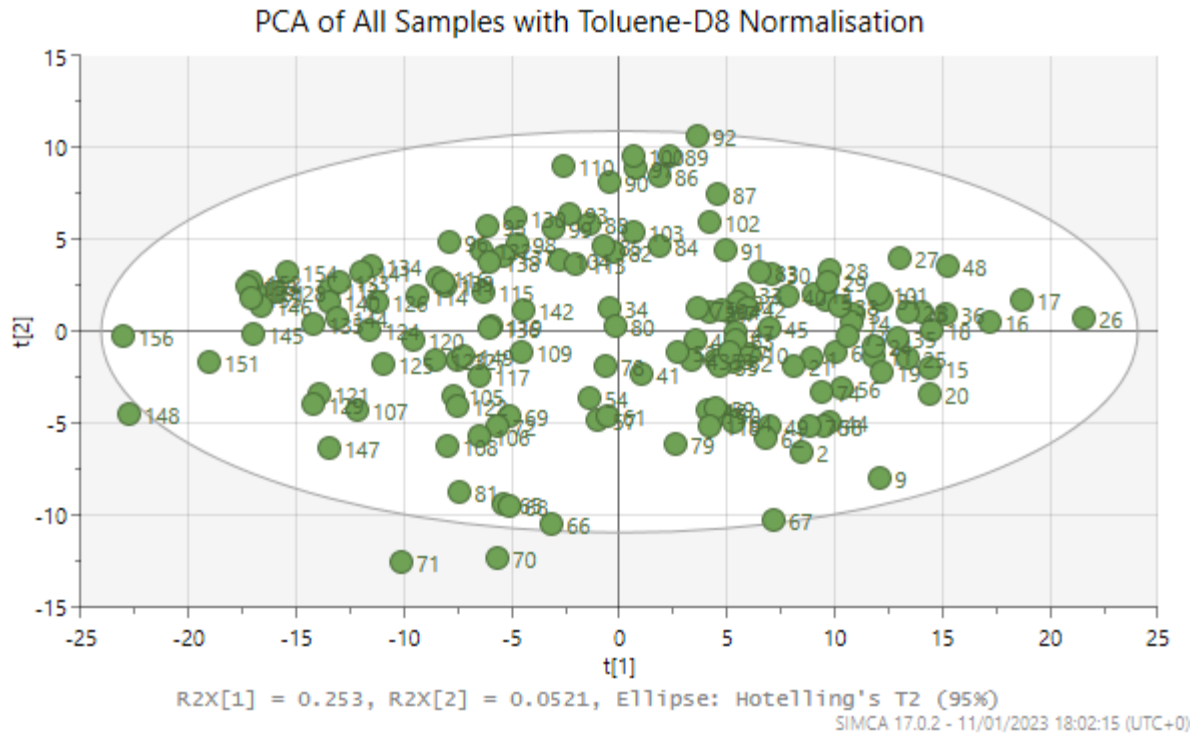


Figure 17.17 PCA of all VOCAL Samples with normalisation by internal standard toluene-d_8.

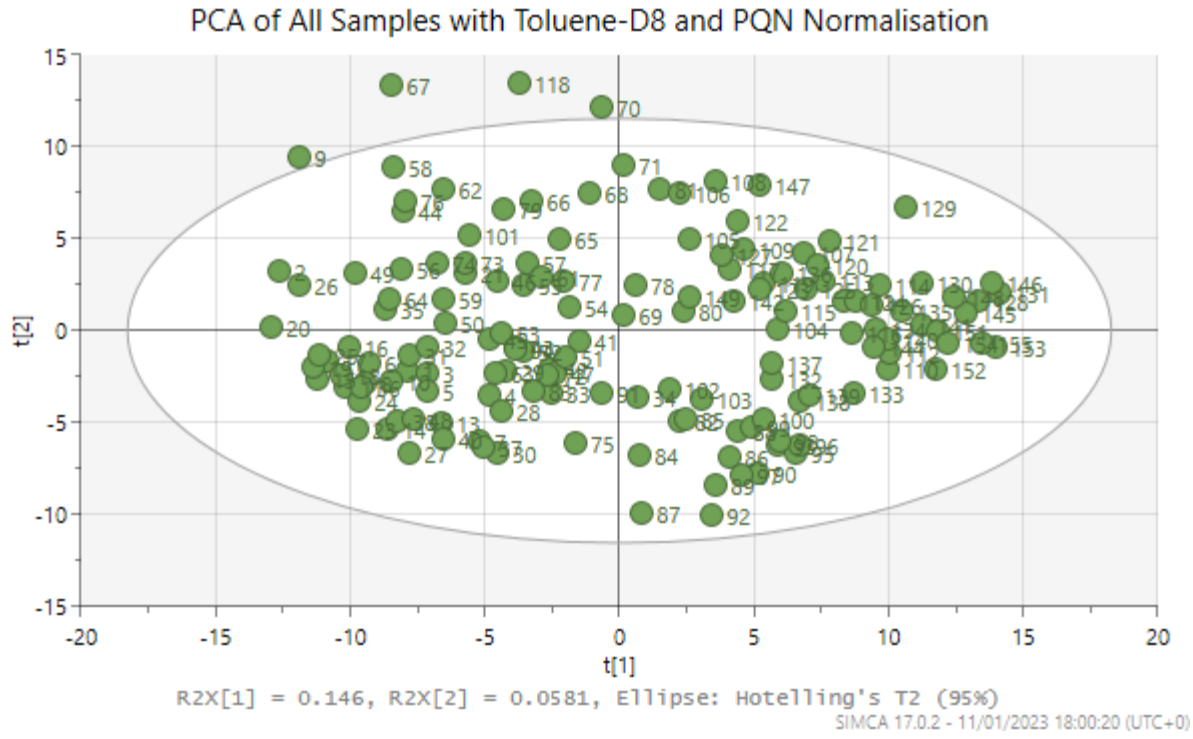


Figure 17.18 PCA of VOCAL Samples with normalisation by internal standard toluene-d_8 and PQN.

17.7.9 Permutation Plots

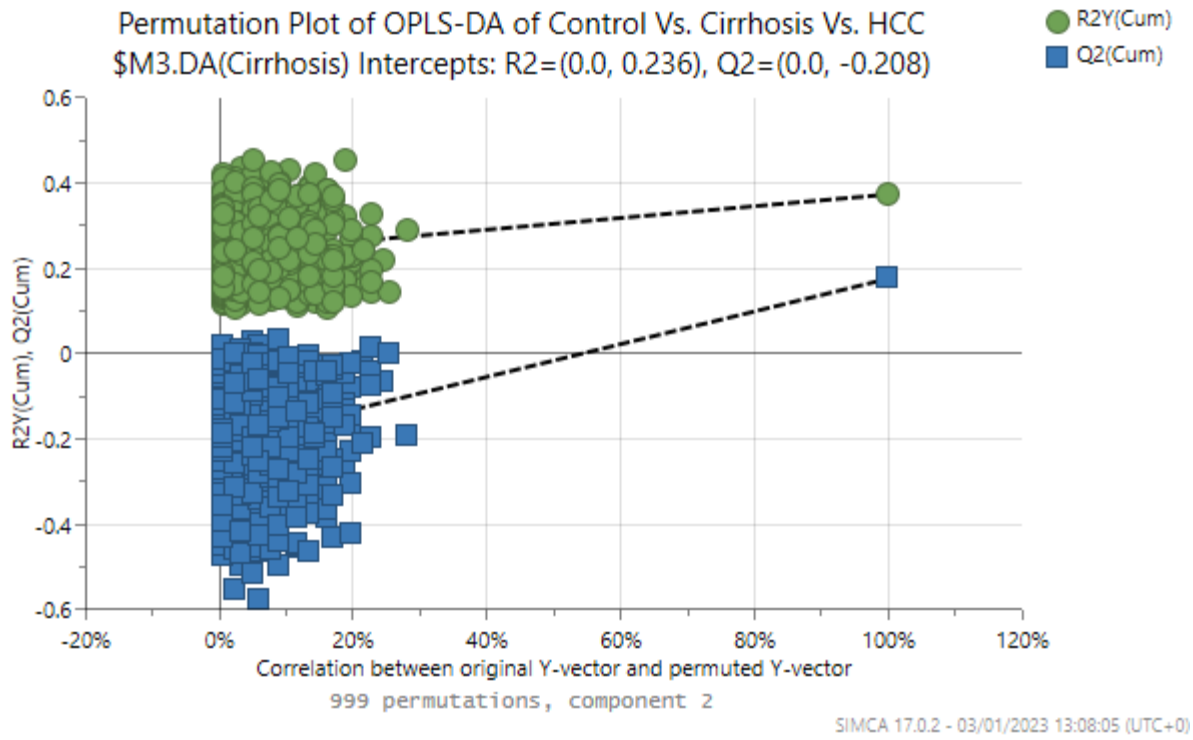


Figure 17.19 Permutation plot from OPLS-DA of control vs cirrhosis vs HCC.

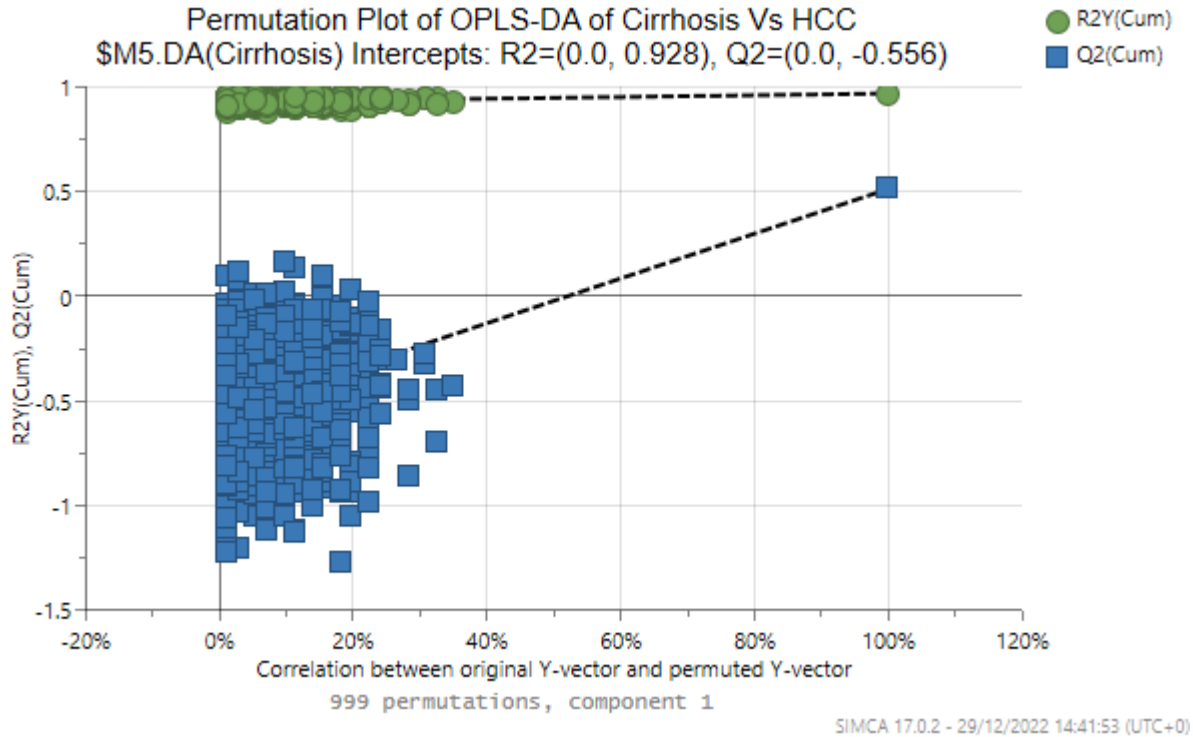


Figure 17.20 Permutation plot from OPLS-DA of cirrhosis vs HCC.

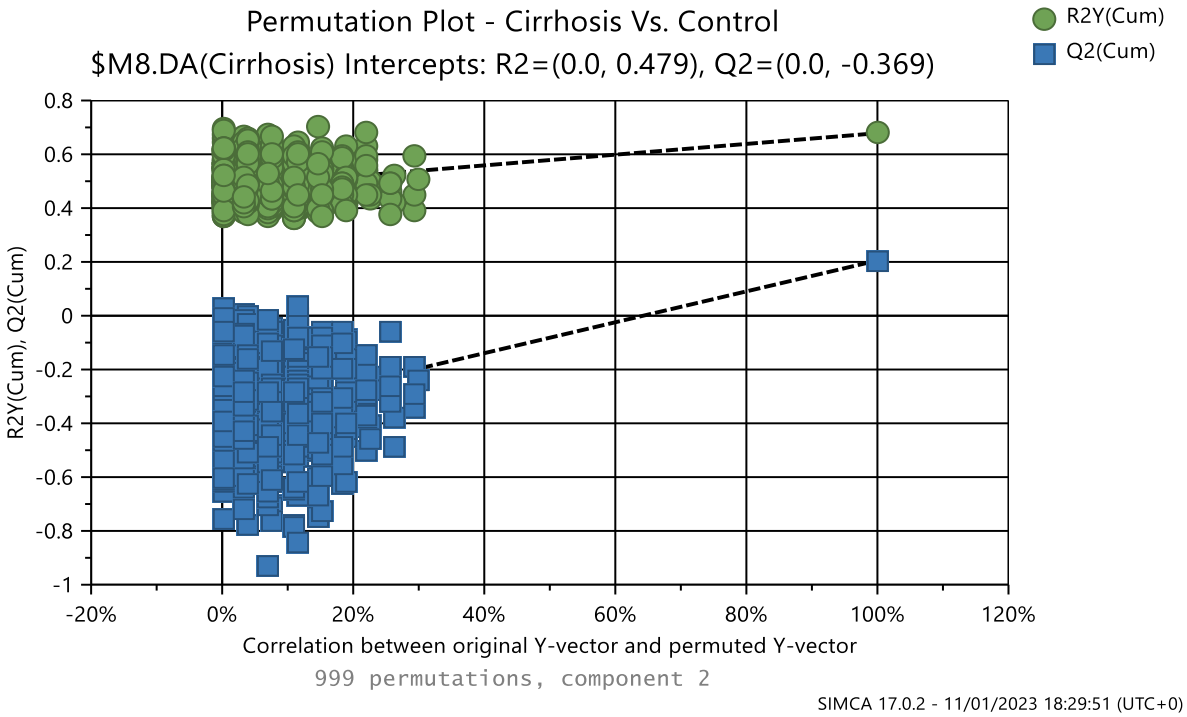


Figure 17.21 Permutation plot from OPLS-DA of cirrhosis vs control.

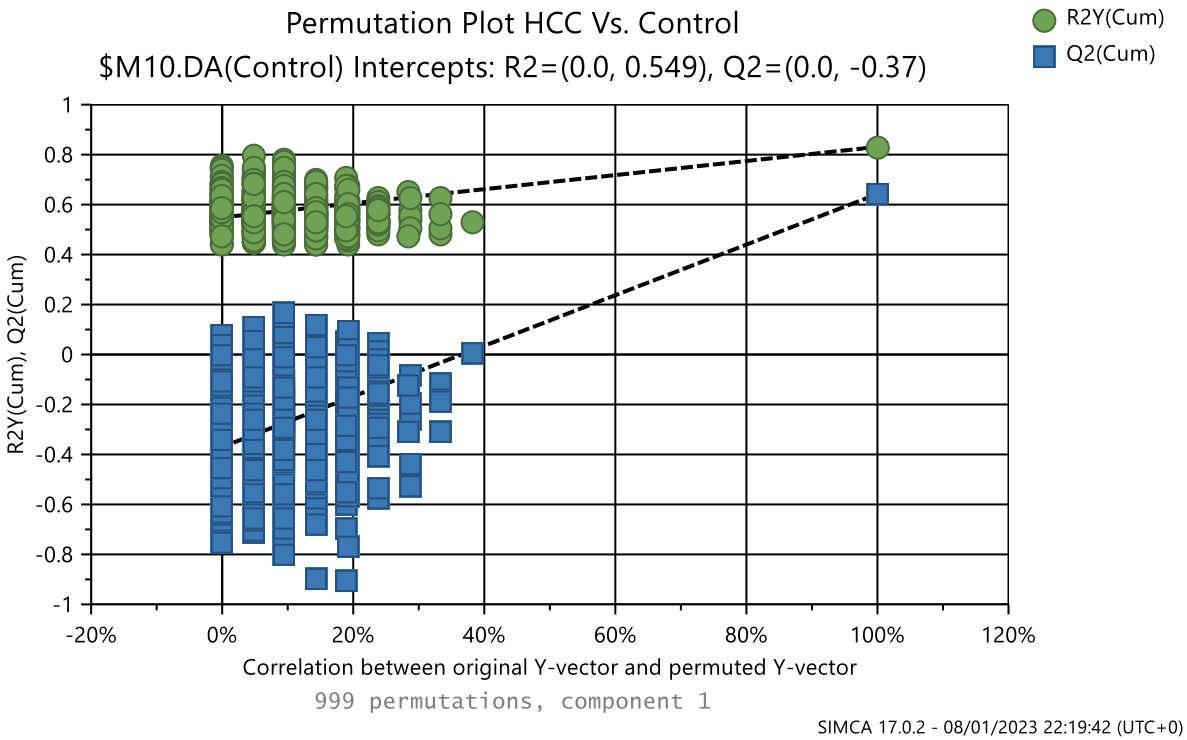


Figure 17.22 Permutation plot from OPLS-DA of HCC vs control.

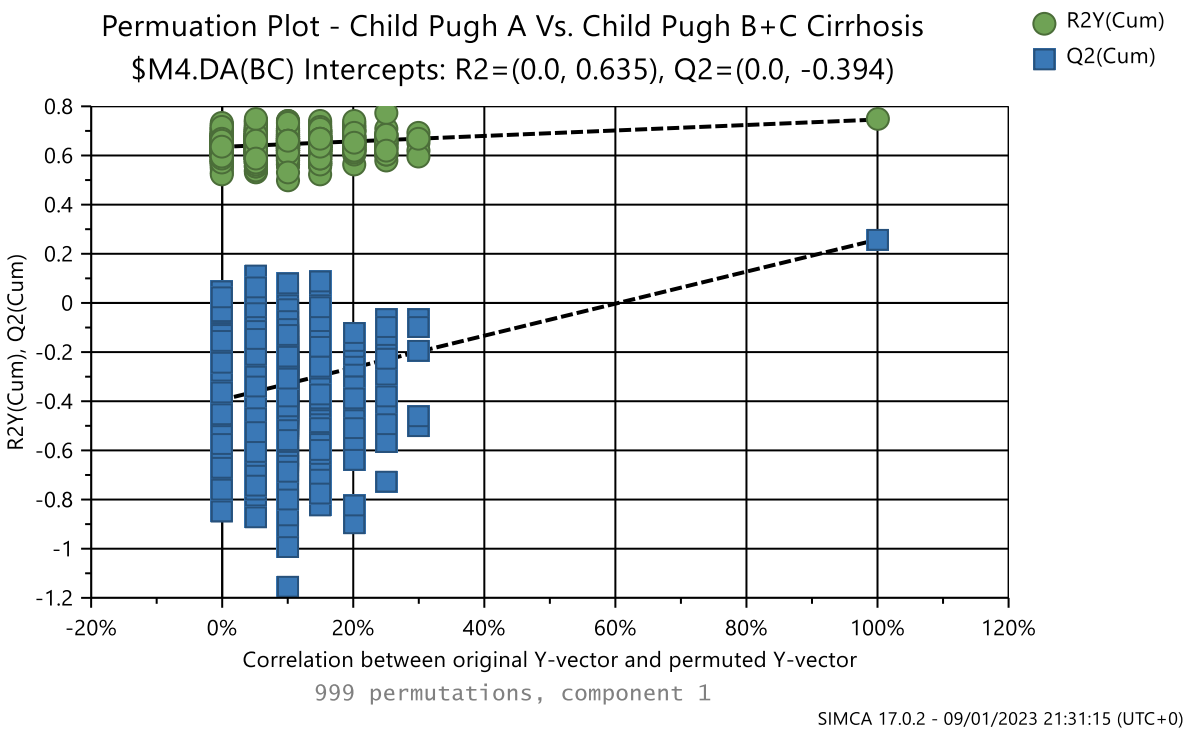


Figure 17.23 Permutation plot from OPLS-DA of Child-Pugh A vs Child-Pugh B/C cirrhosis.

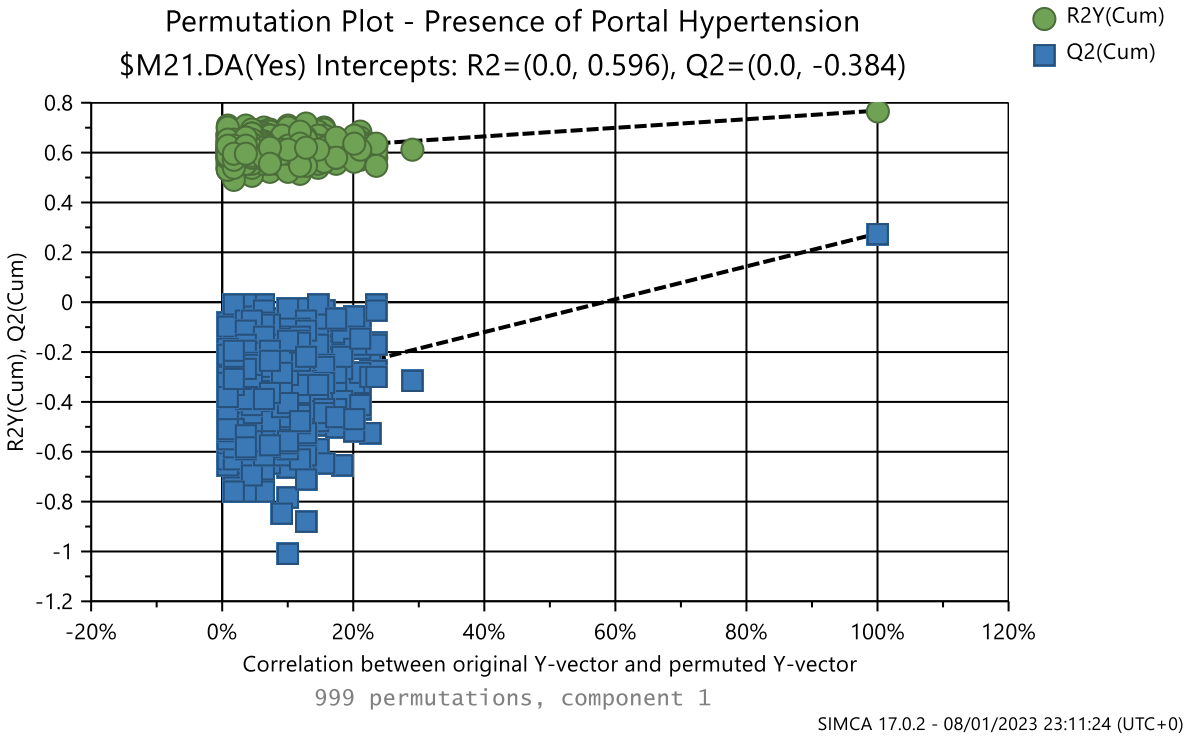


Figure 17.24 Permutation plot from OPLS-DA of presence of portal hypertension.

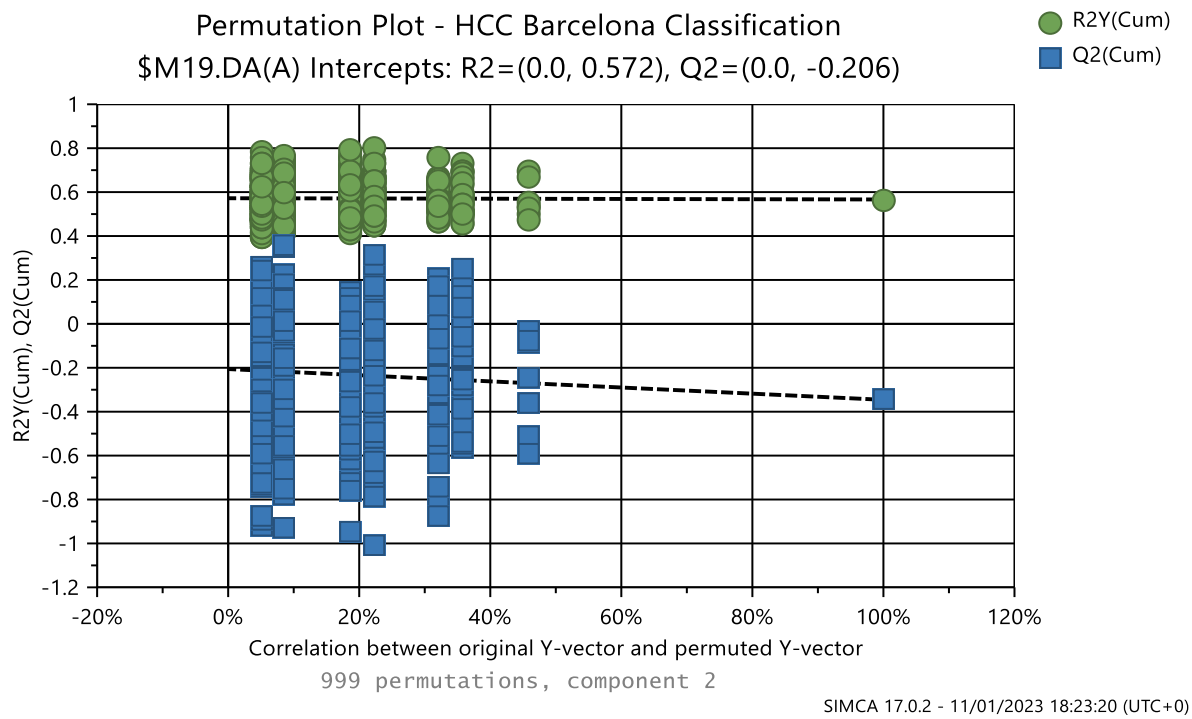


Figure 17.25 Permutation plot from OPLS-DA of HCC Barcelona clinic liver cancer classification.

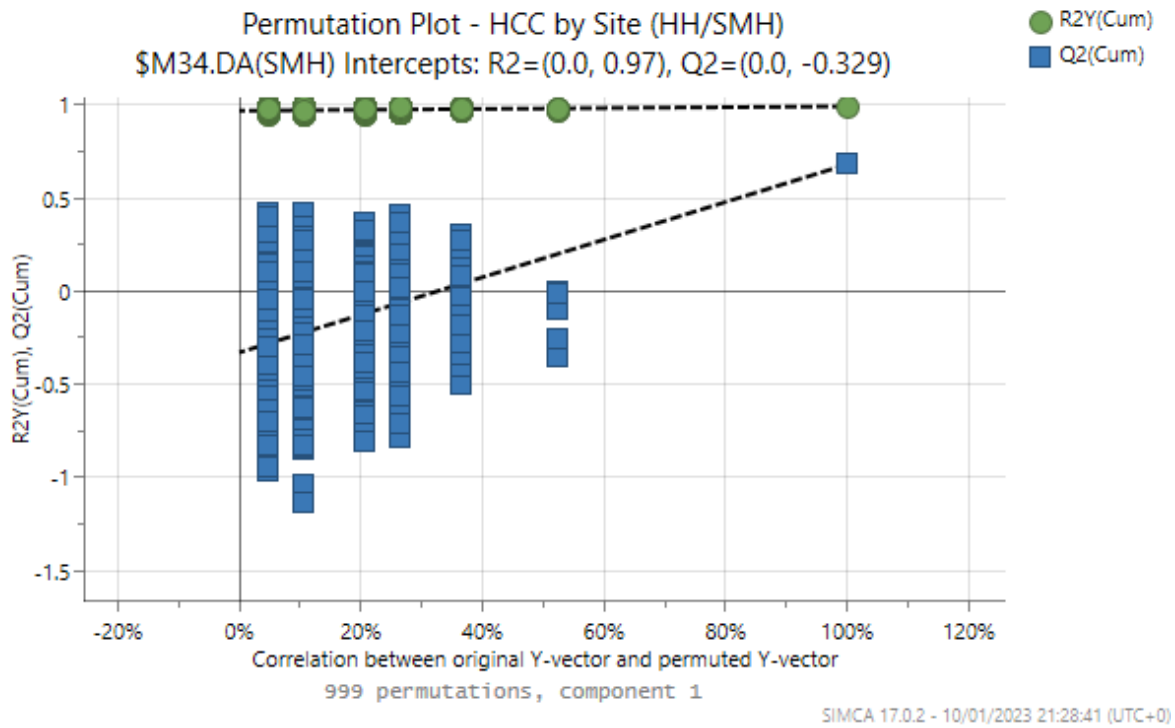


Figure 17.26 Permutation plot of HCC by site.

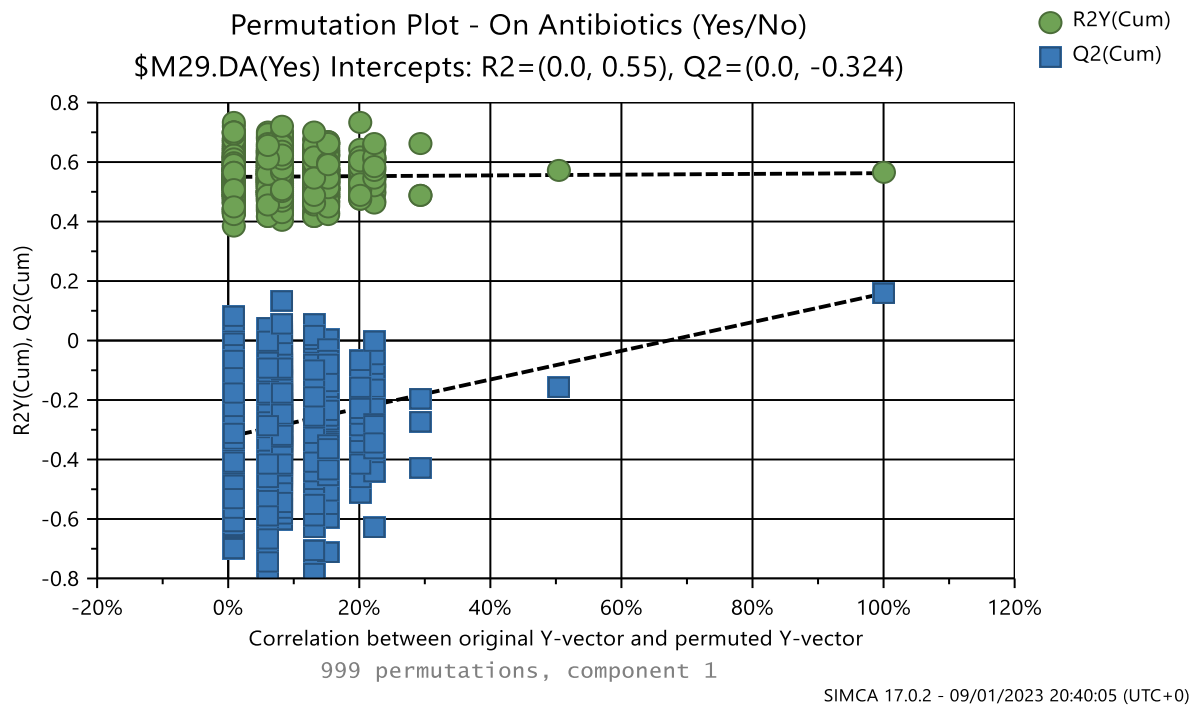


Figure 17.27 Permutation plot of long-term antibiotic use.

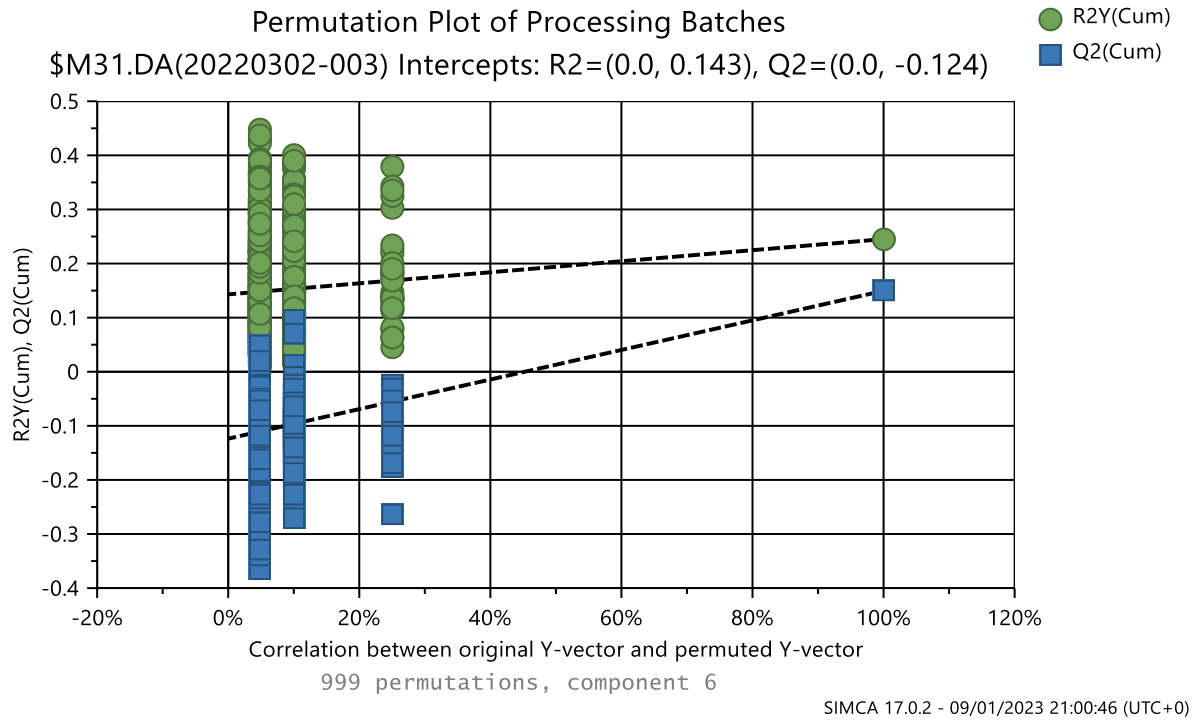


Figure 17.28 Permutation plot of processing batches.

17.7.10 VOCAL ROC Curves

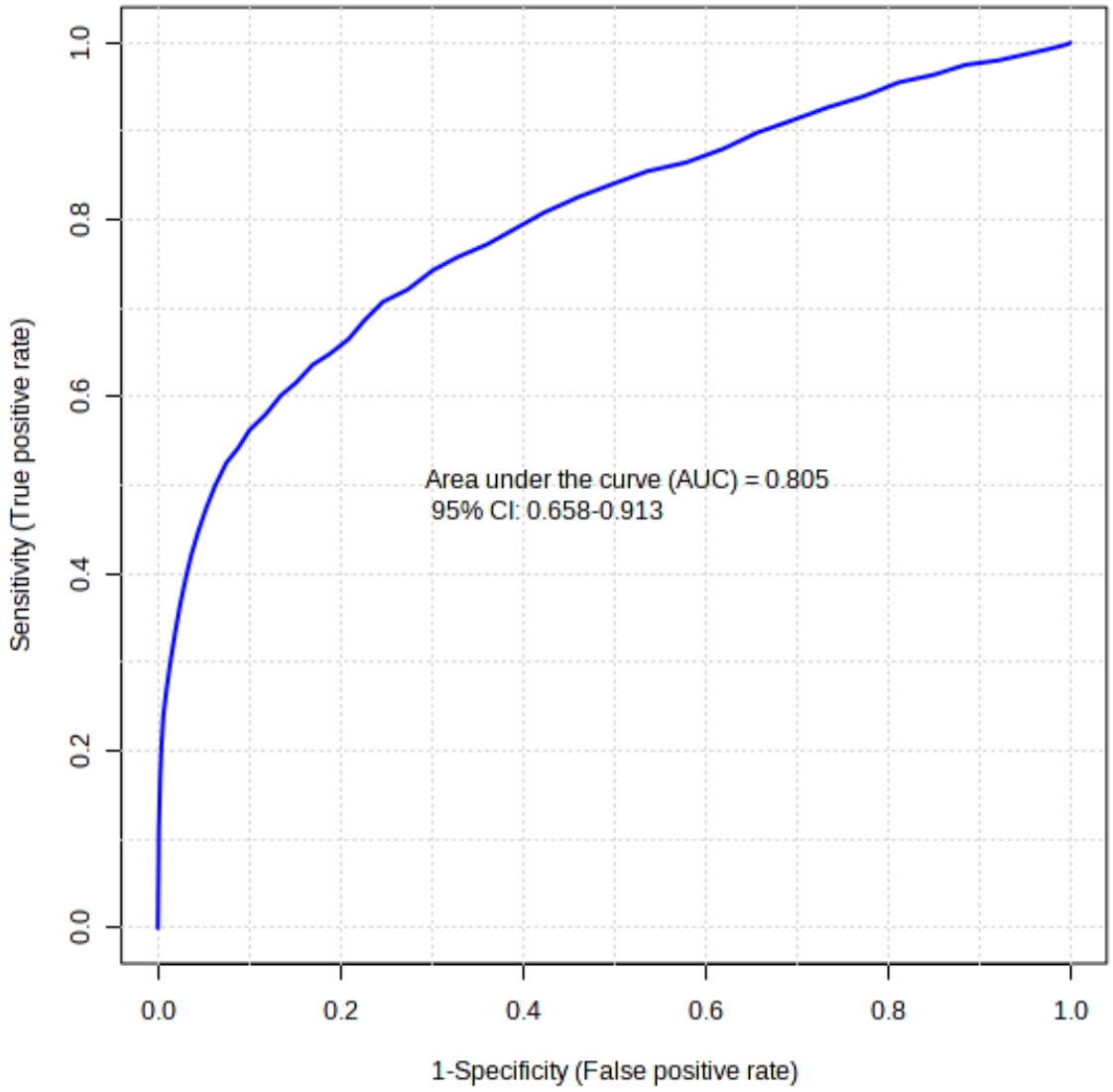


Figure 17.29 ROC curve of cirrhosis vs HCC with all FDR significant VOCs with a VIP score >1.5.

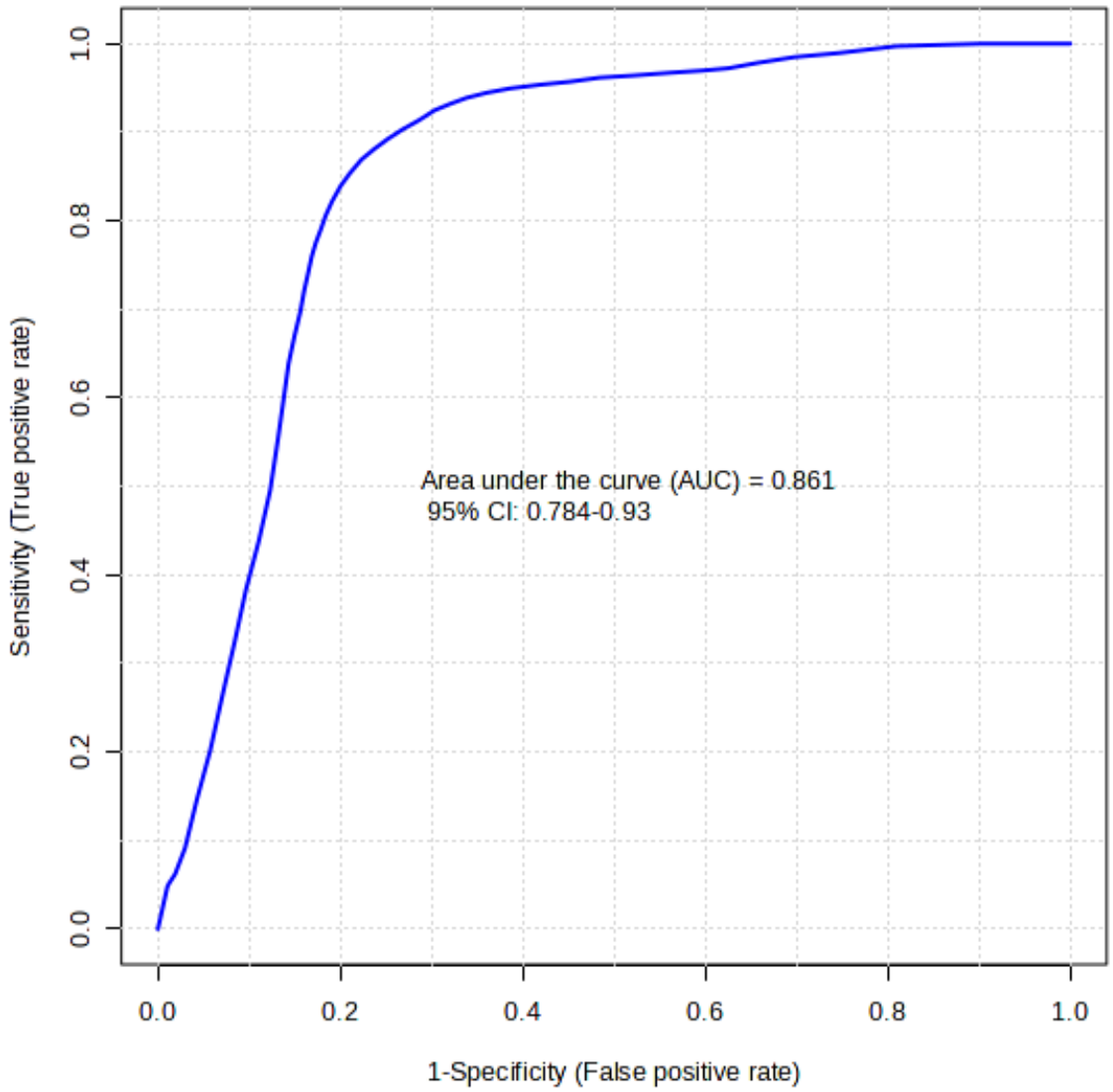


Figure 17.30 ROC curve of cirrhosis vs control with all FDR significant VOCs with a VIP score >1.5.

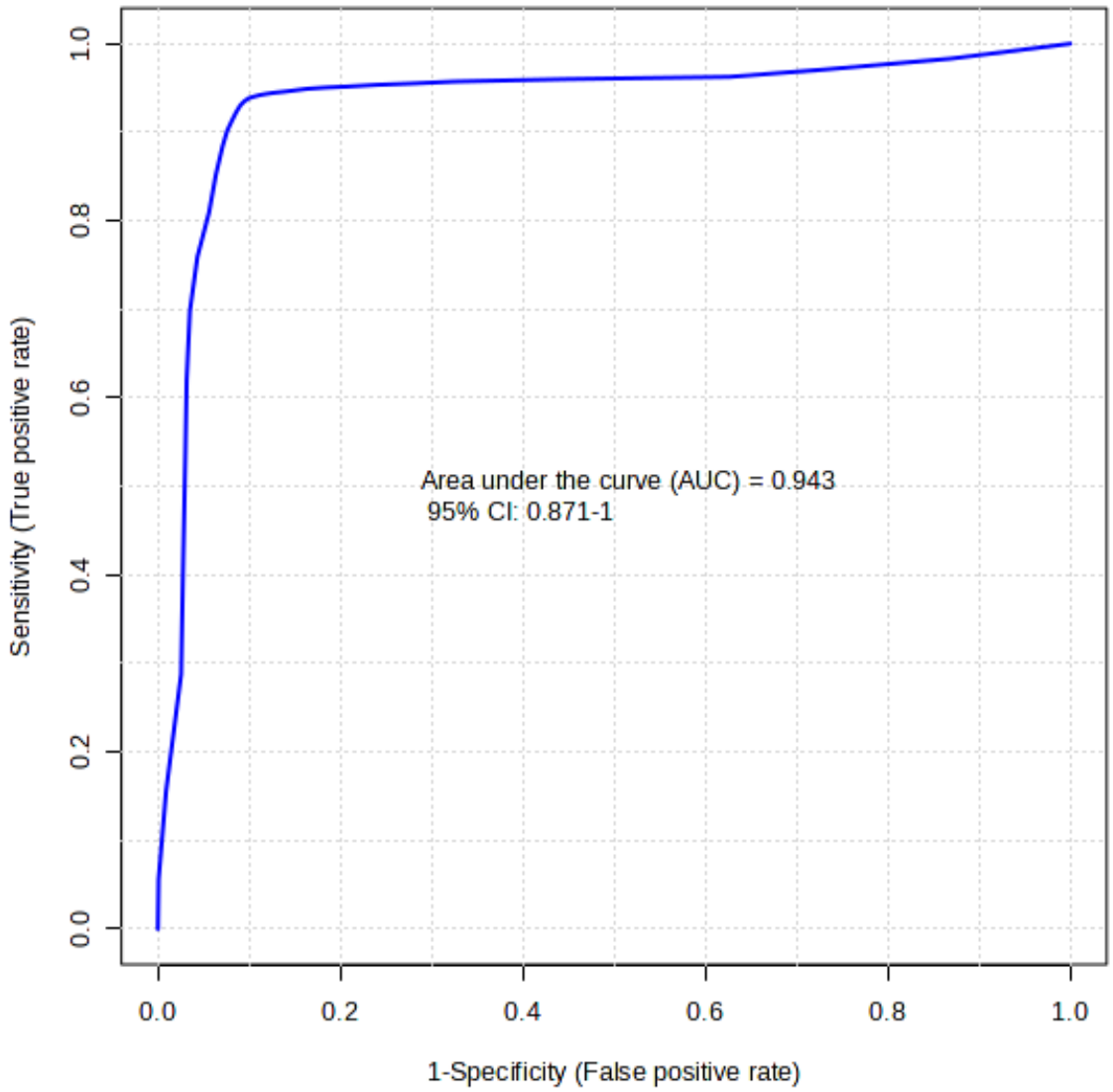


Figure 17.31 ROC curve of HCC vs control with all FDR significant VOCs with a VIP score >1.5.

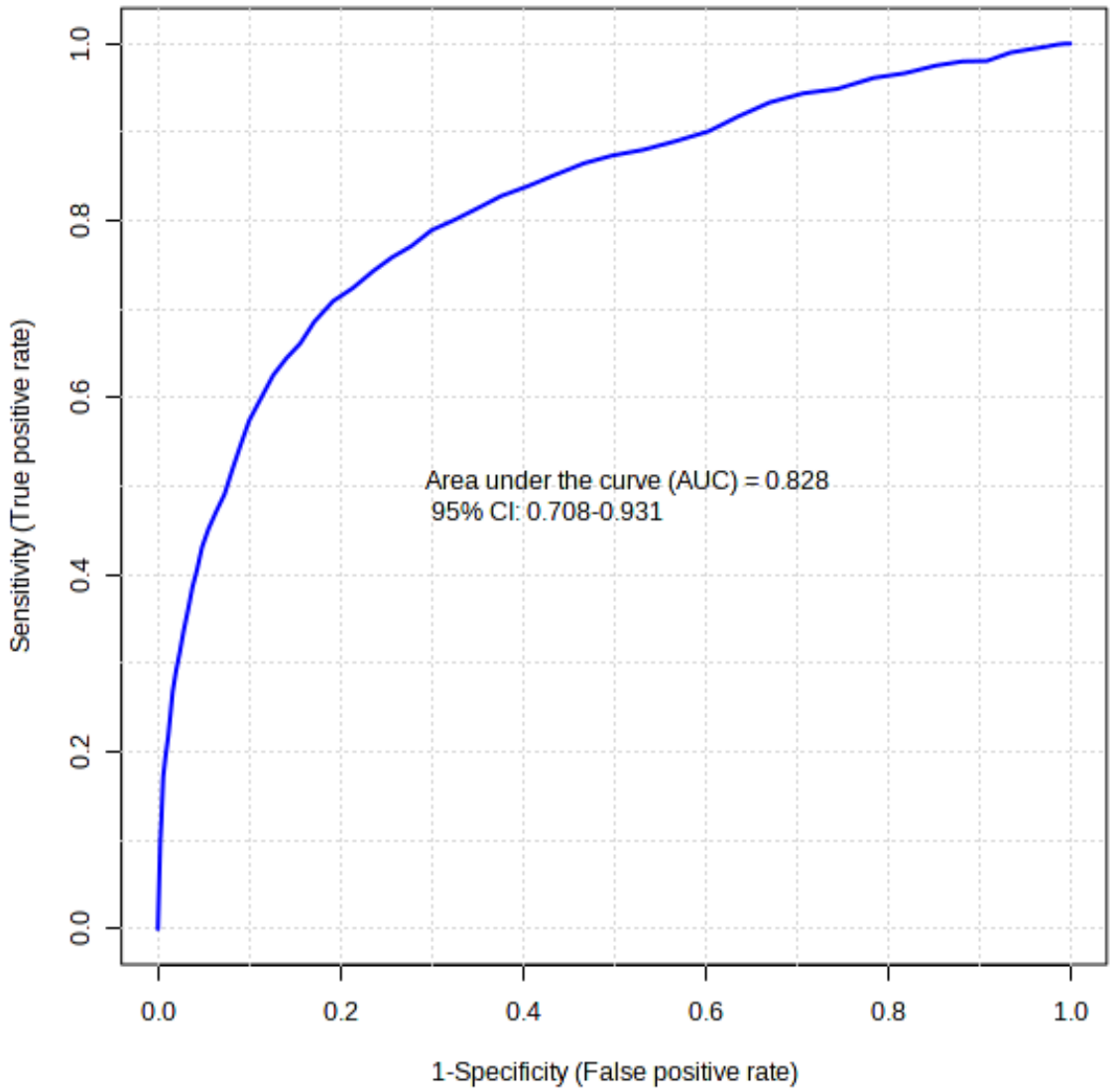


Figure 17.32 ROC curve of Child-Pugh A vs Child Pugh B cirrhotics with all FDR significant VOCs with a VIP score >1.5.

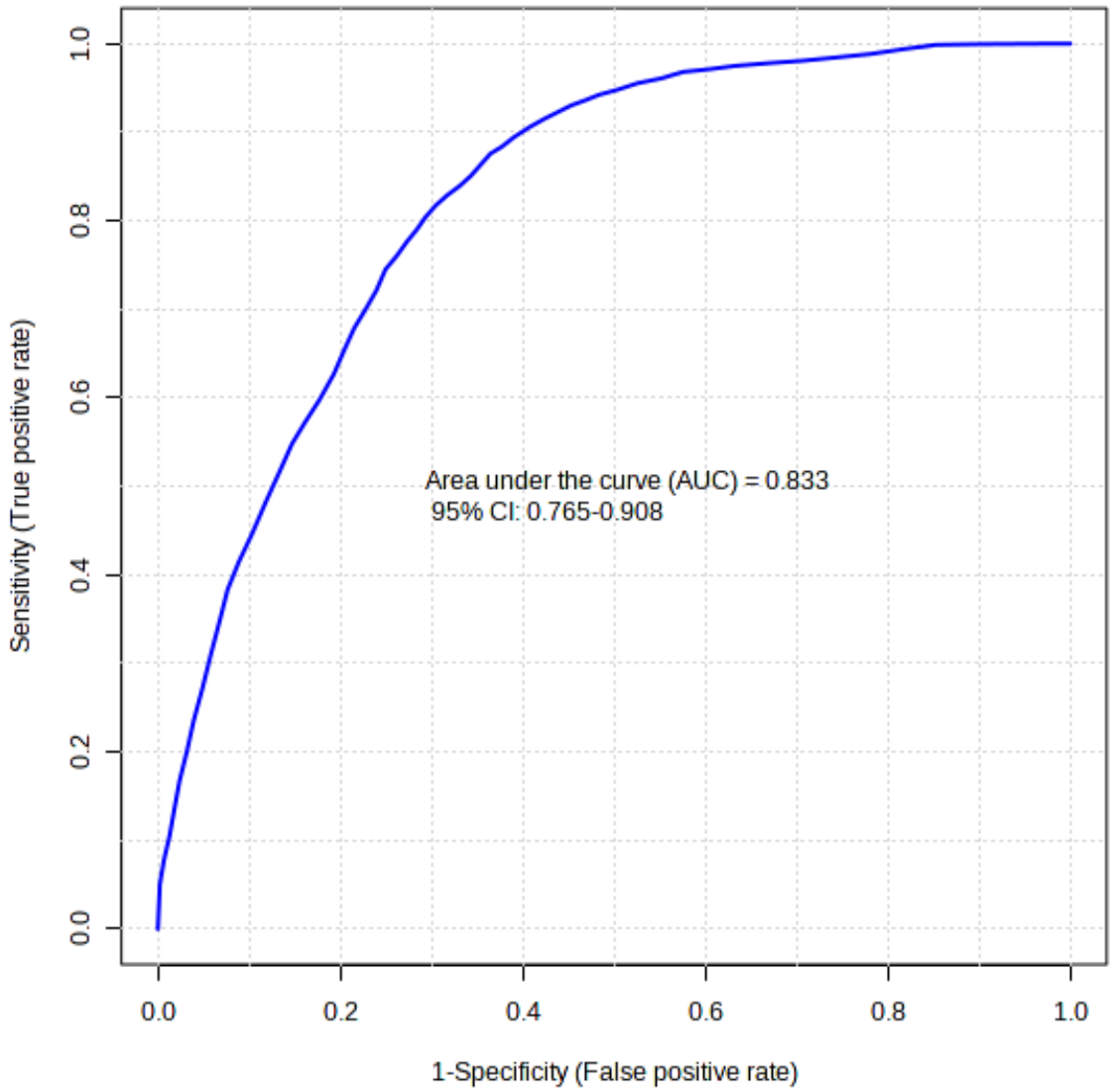


Figure 17.33 ROC curve of those with portal hypertension vs those without portal hypertension with all FDR significant VOCs with a VIP score >1.5.

17.7.11 Box Plots of VOCs with VIP Score >1.5 and High AUCs

Cirrhosis vs HCC

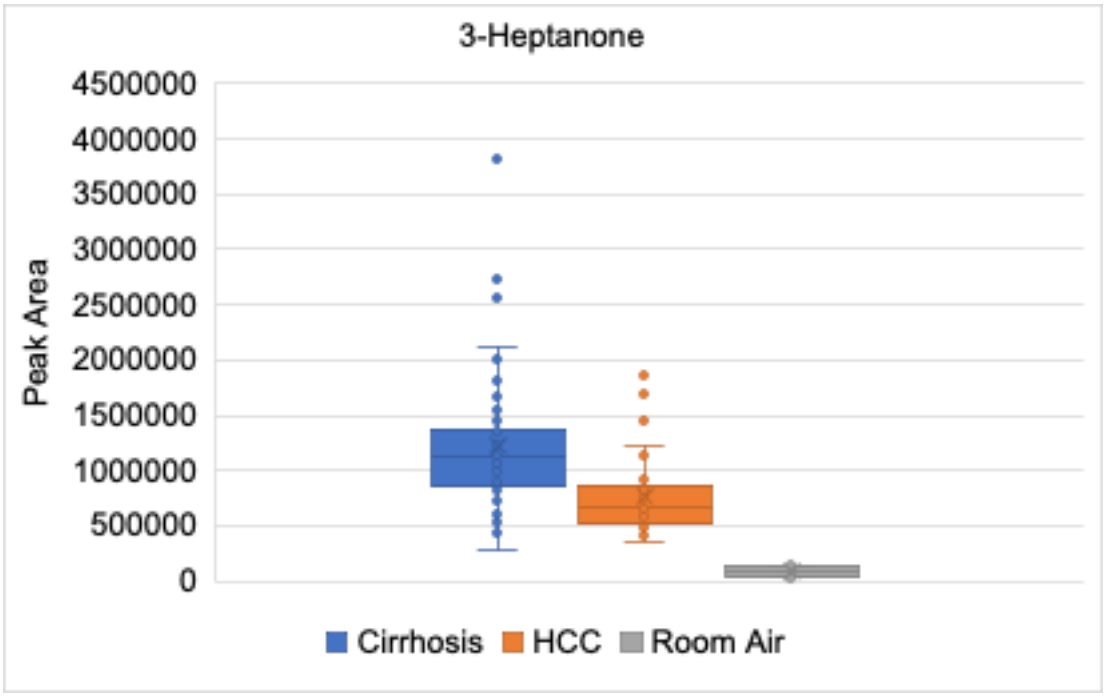


Figure 17.34 Box plot of 3-heptanone for cirrhosis vs HCC demonstrating a higher abundance of 3-heptanone in those with cirrhosis compared to those with HCC. Room air abundance was low.

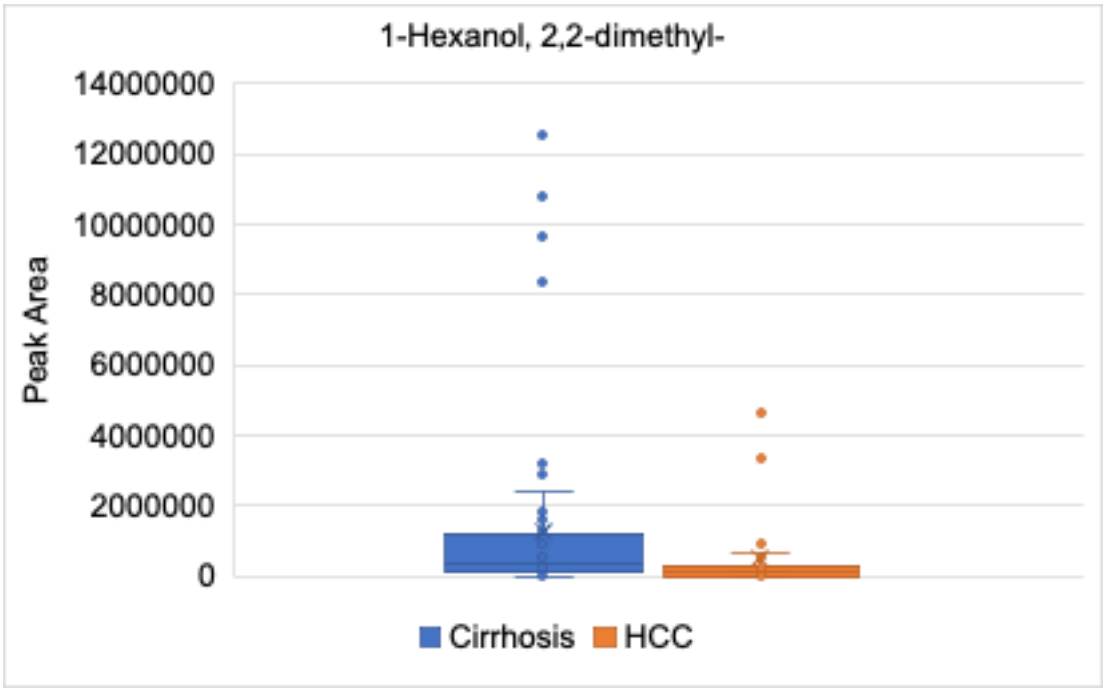


Figure 17.35 Box plot of 1-hexanol, 2,2-dimethyl- for cirrhosis vs HCC demonstrating a higher abundance in those with cirrhosis compared to HCC. 1-Hexanol, 2,2-dimethyl- was not detected in room air samples.

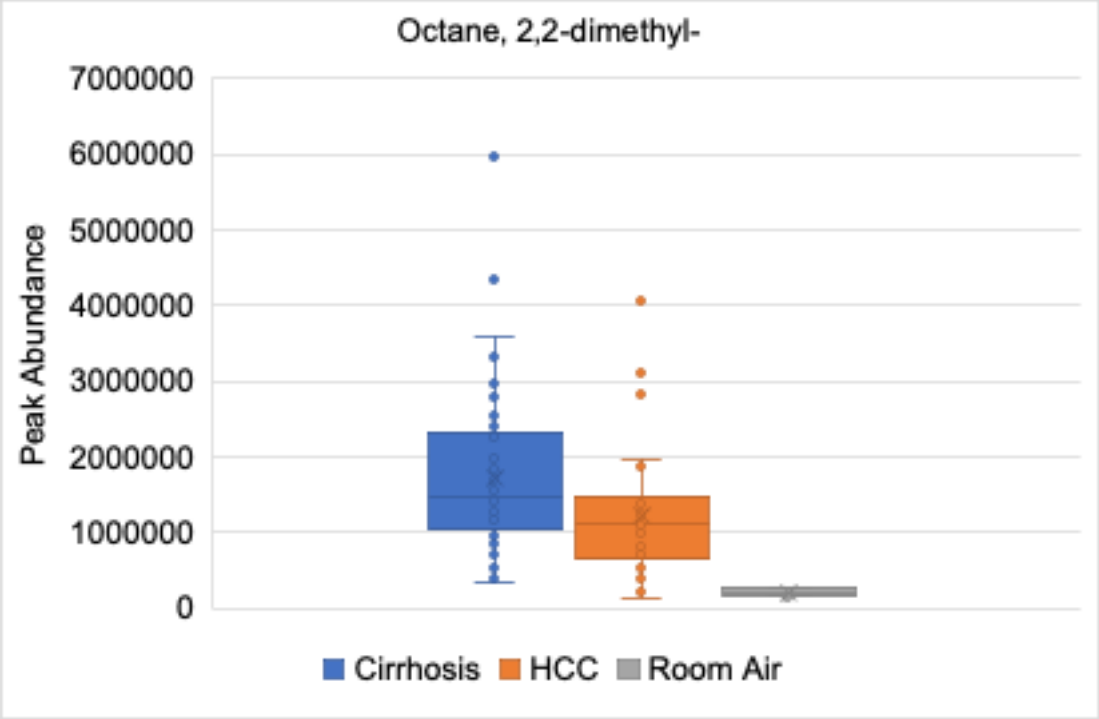


Figure 17.36 Box plot of octane, 2,2-dimethyl- for cirrhosis vs HCC demonstrating higher abundance in those with cirrhosis compared to HCC. Octane, 2,2-dimethyl was low in abundance in room air.

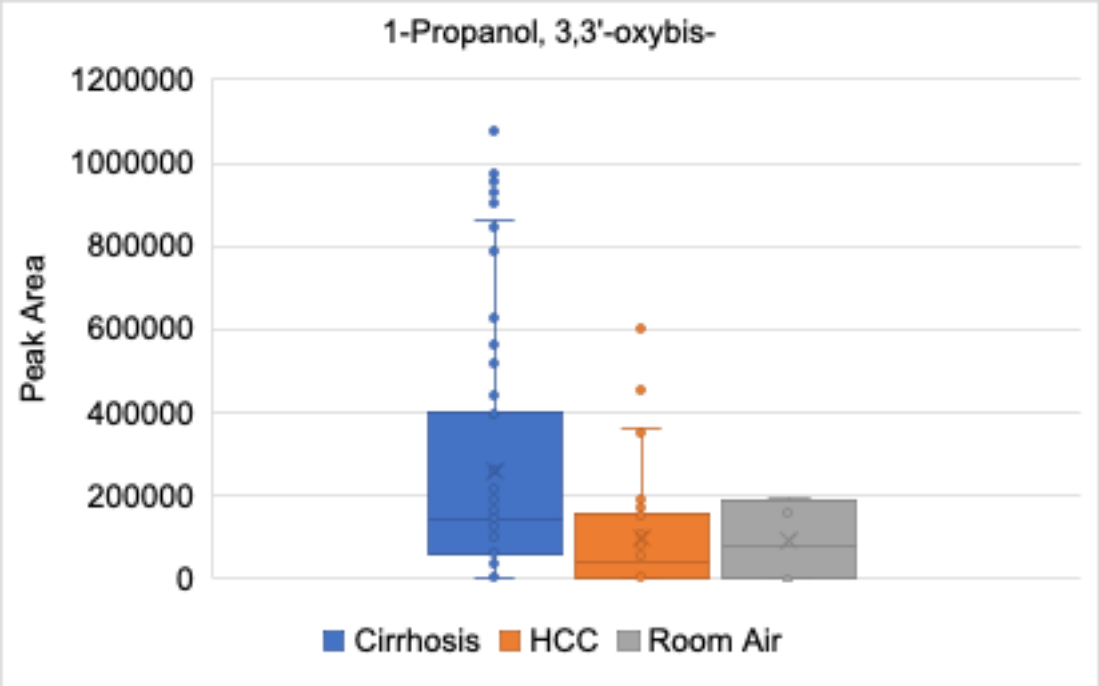


Figure 17.37 Box plot of 1-propanol, 3,3'-oxybis- for cirrhosis vs HCC demonstrating a higher abundance in those with cirrhosis compared to HCC. Room air abundance of 1-propanol, 3,3'-oxybis- was comparable to some breath samples.

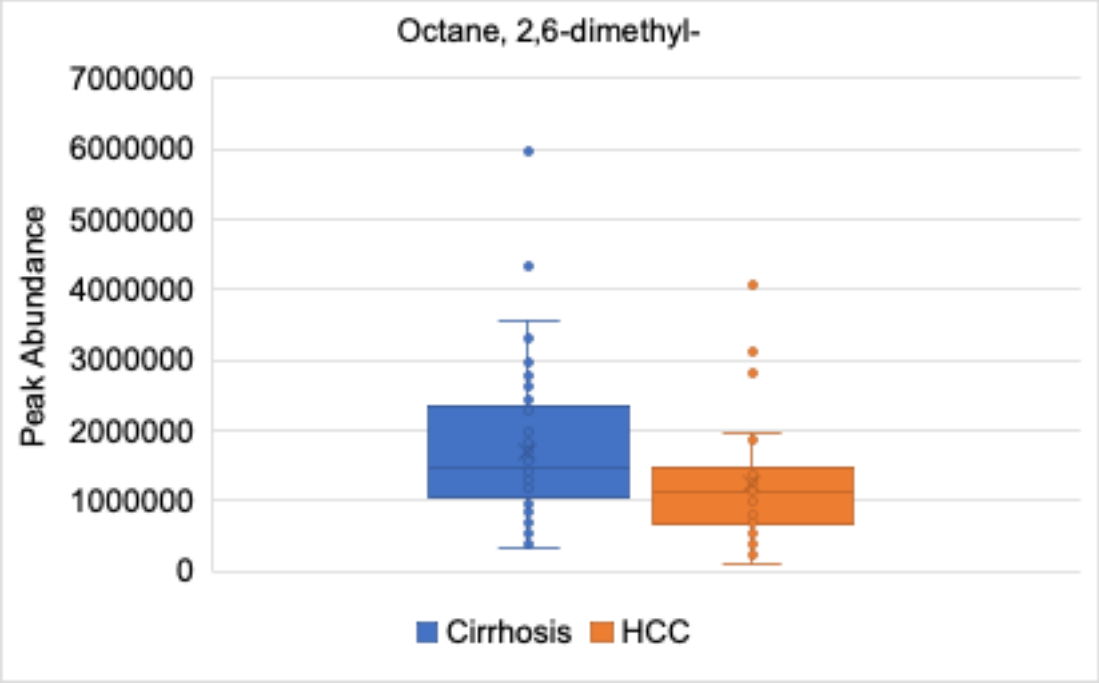


Figure 17.38 Box plot of Octane, 2,6-dimethyl- for cirrhosis vs HCC demonstrating a higher abundance in those with cirrhosis compared to HCC. Octane, 2,6-dimethyl- was not detected in room air samples.

Cirrhosis vs Control

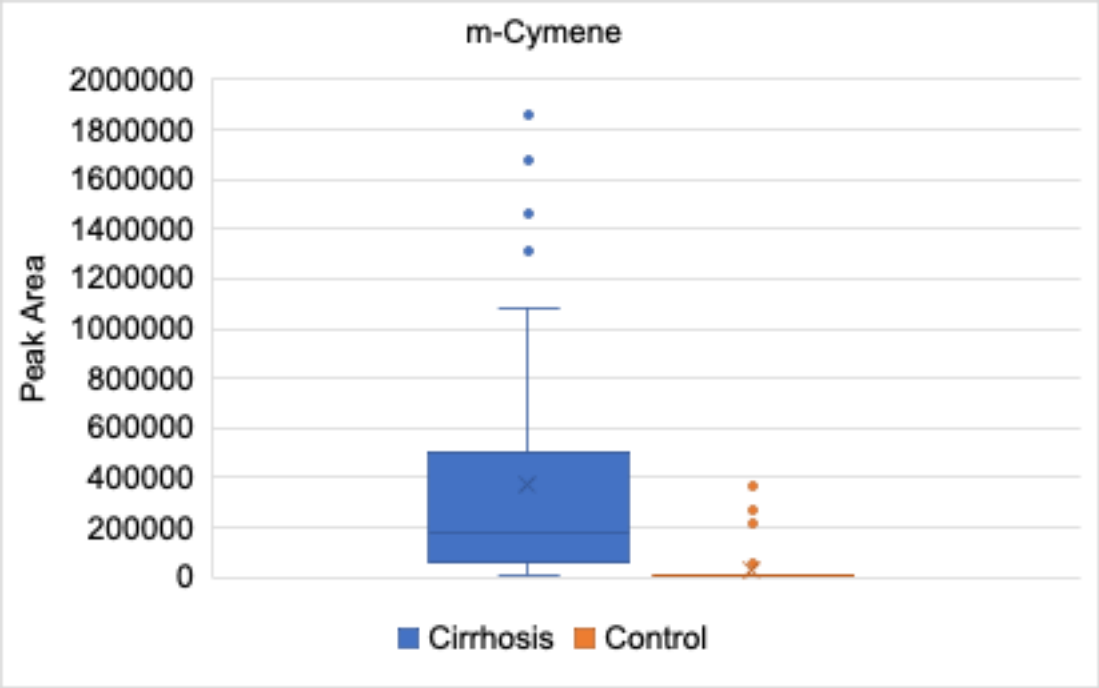


Figure 17.39 Box plot of m-cymene for cirrhosis vs control demonstrating a higher abundance in those with cirrhosis compared to controls. m-Cymene was not detected in room air samples.

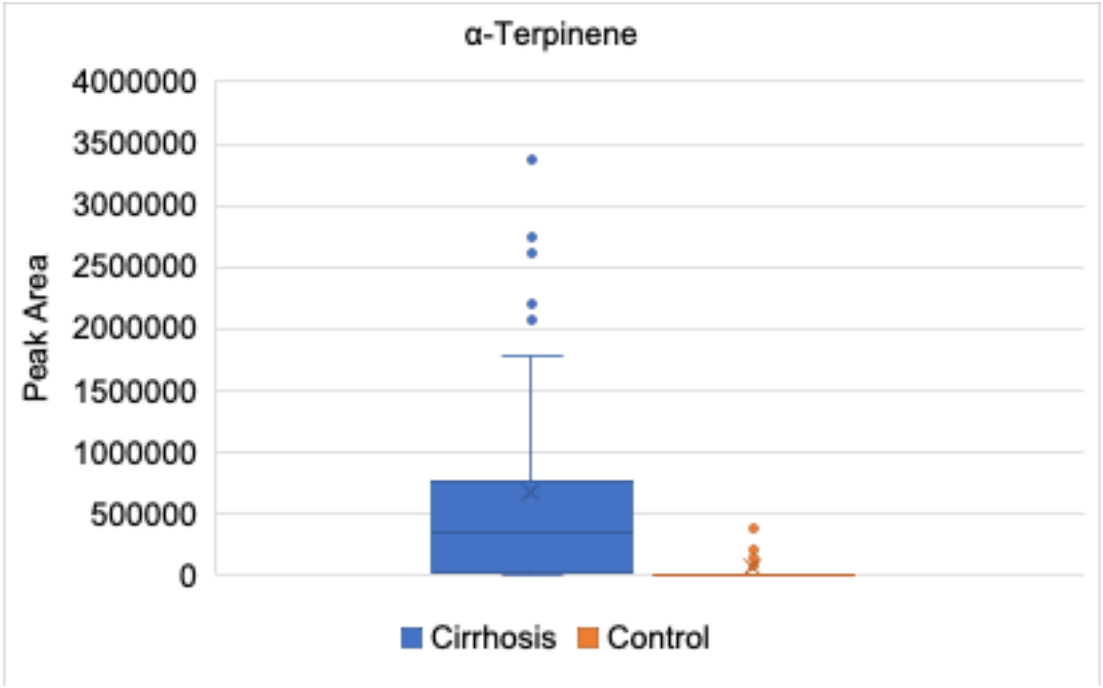


Figure 17.40 Box plot of α-terpinene for cirrhosis vs control demonstrating a higher abundance in those with cirrhosis compared to controls. α-Terpinene was not detected in room air.

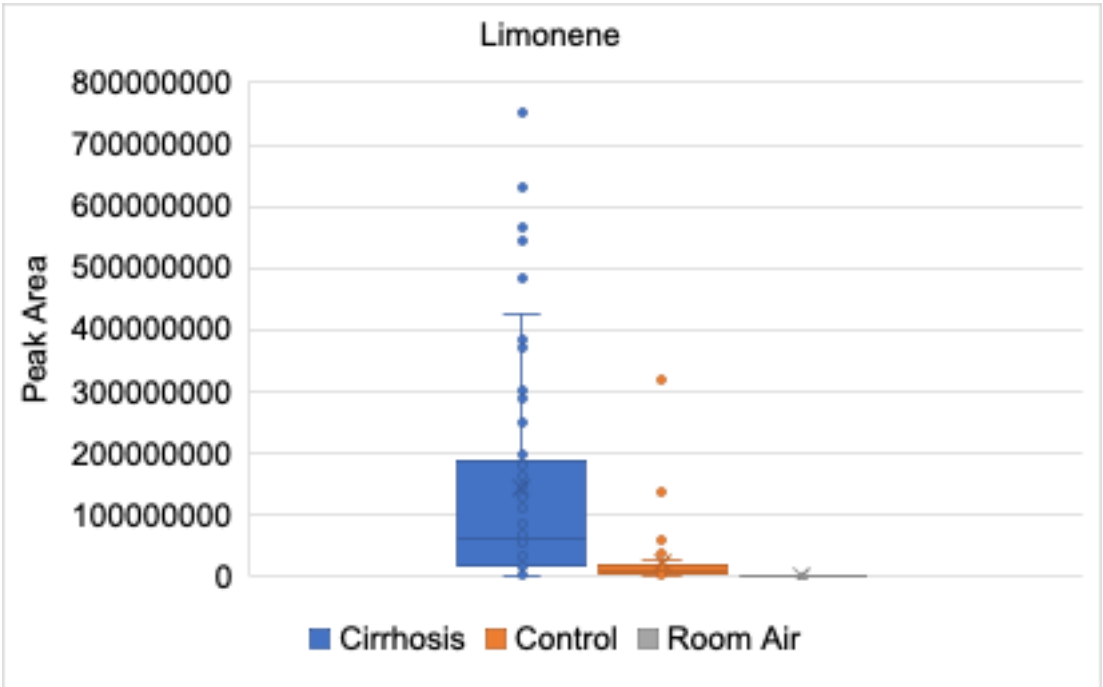


Figure 17.41 Box plot of limonene for cirrhosis vs control demonstrating a higher abundance in those with cirrhosis compared to controls. Room air abundance of limonene was low.

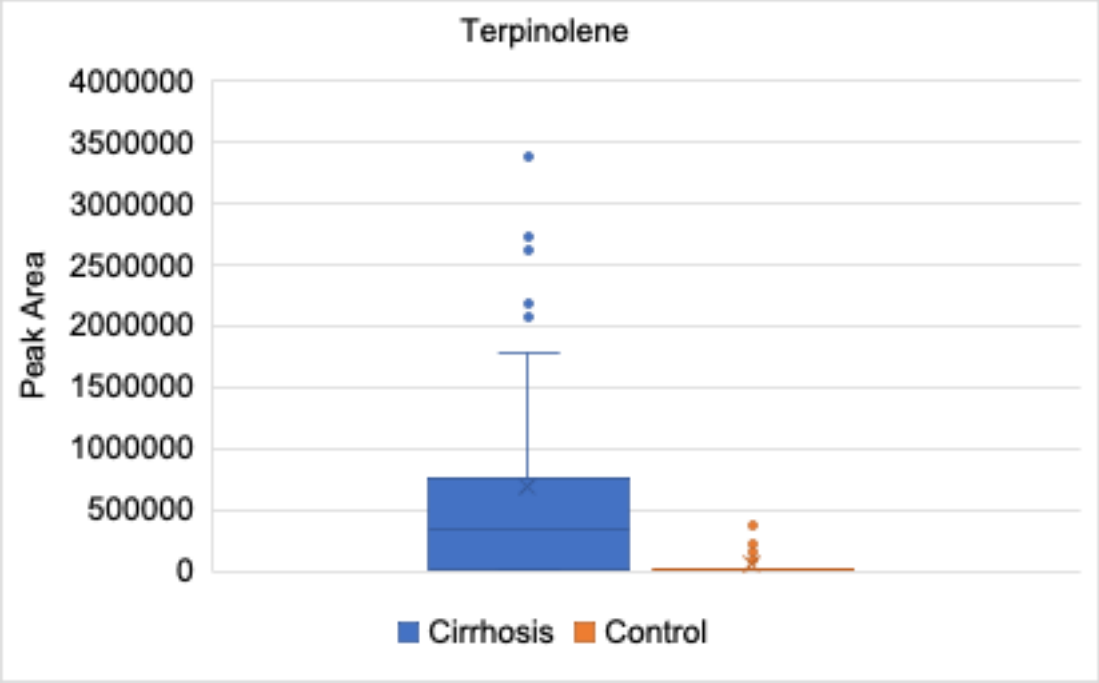


Figure 17.42 Box plot of terpinolene for cirrhosis vs control demonstrating a higher abundance in those with cirrhosis compared to controls. Terpinolene was not detected in room air samples.

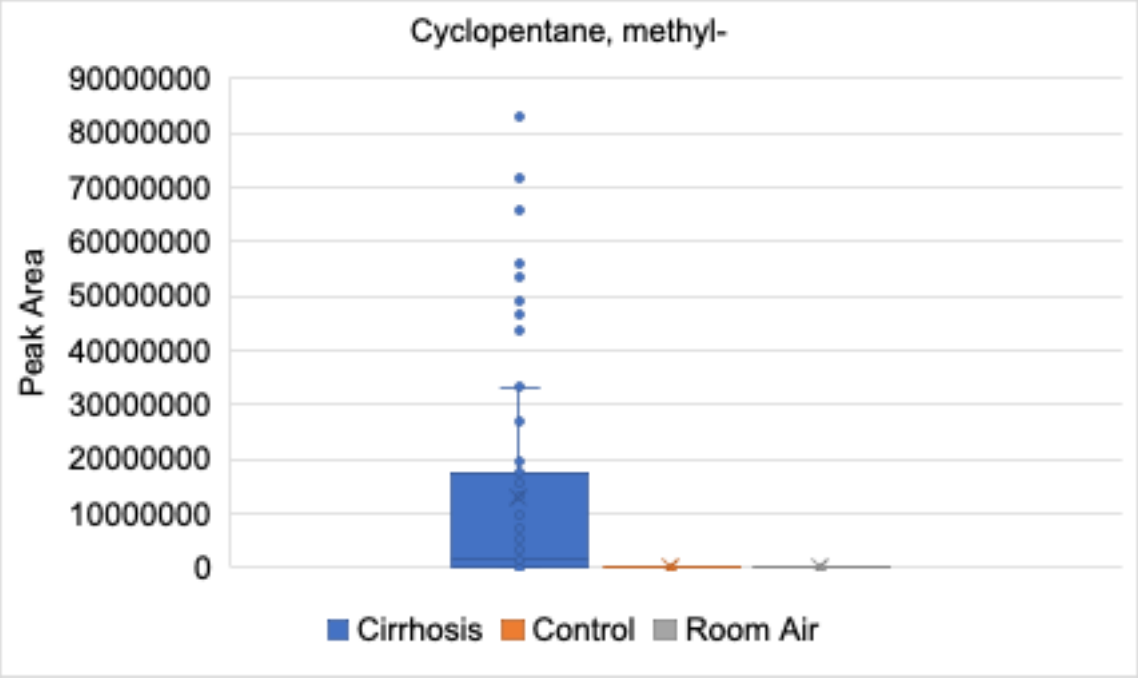


Figure 17.43 Box plot of cyclopentane, methyl- for cirrhosis vs control demonstrating a higher abundance in those with cirrhosis compared to controls. Room air abundance was low.

HCC vs Control

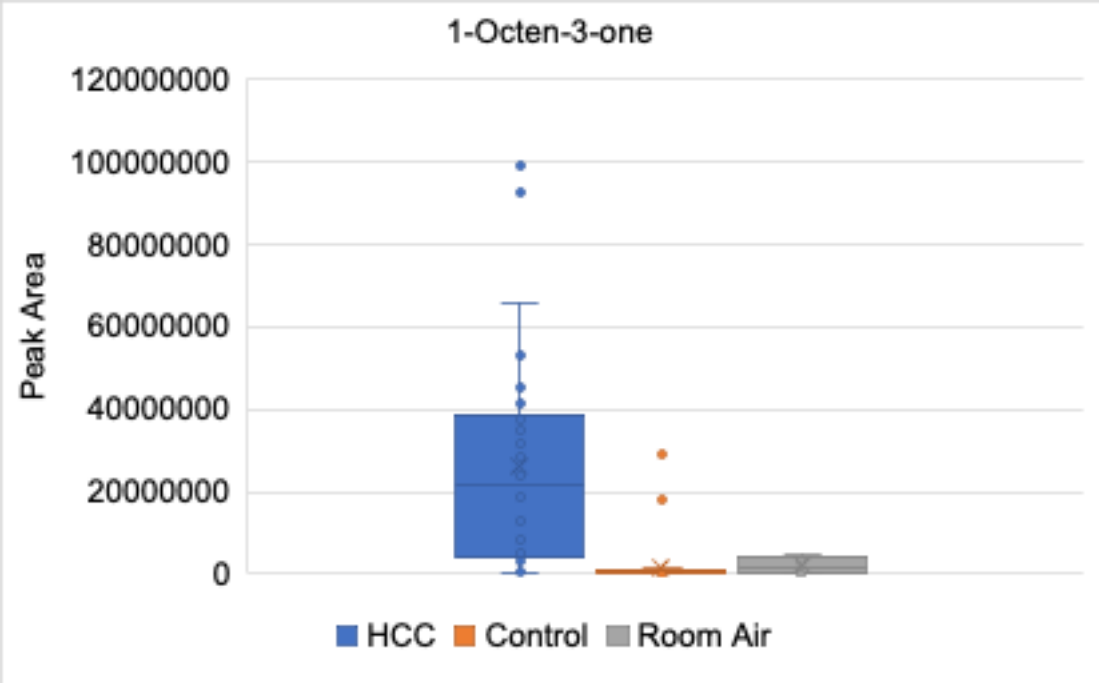


Figure 17.44 Box plot of 1-octen-3-one for HCC vs controls demonstrating a higher abundance in those with HCC compared to controls. Room air abundance was generally low but comparable to some breath samples.

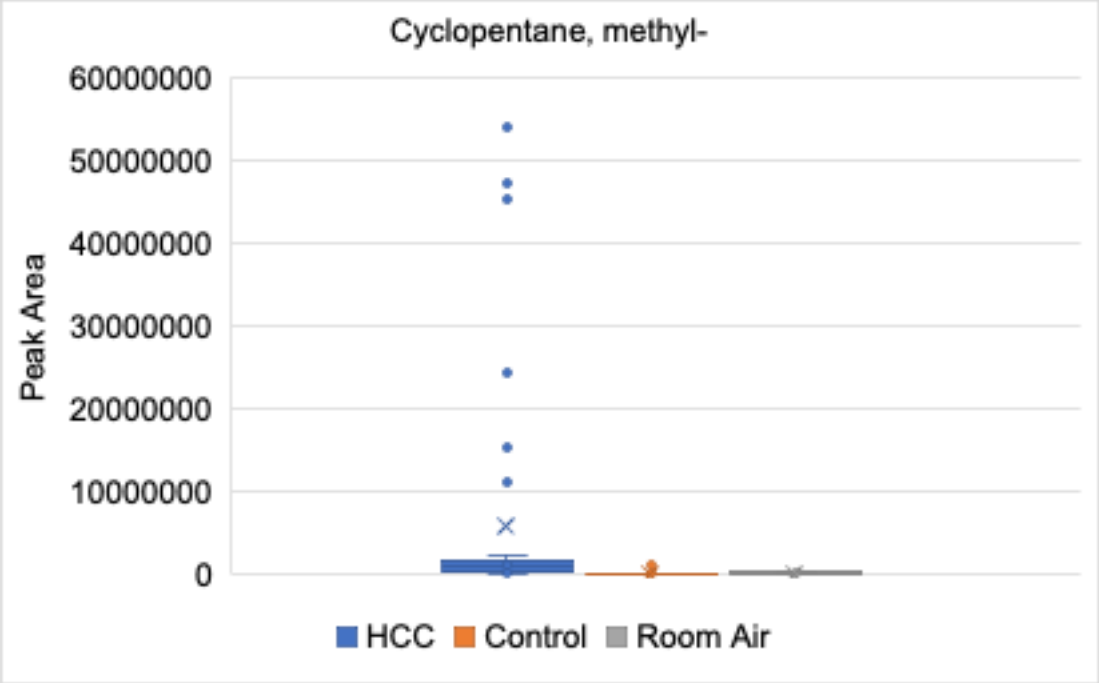


Figure 17.45 Box plot of cyclopentane, methyl- for HCC vs controls demonstrating a higher abundance in those with HCC compared to controls. Abundance was very variable, however. Room air abundance was low.

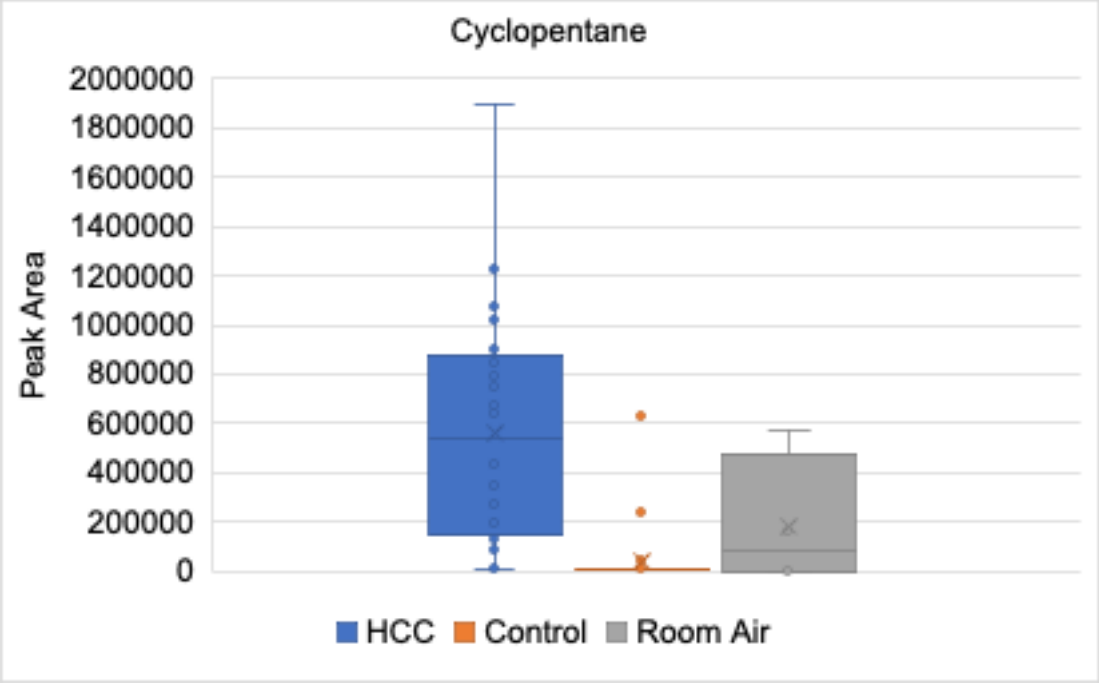


Figure 17.46 Box plot of cyclopentane for HCC vs controls demonstrating a higher abundance in those with HCC compared to controls. The abundance of cyclopentane within room air was also high compared to controls.

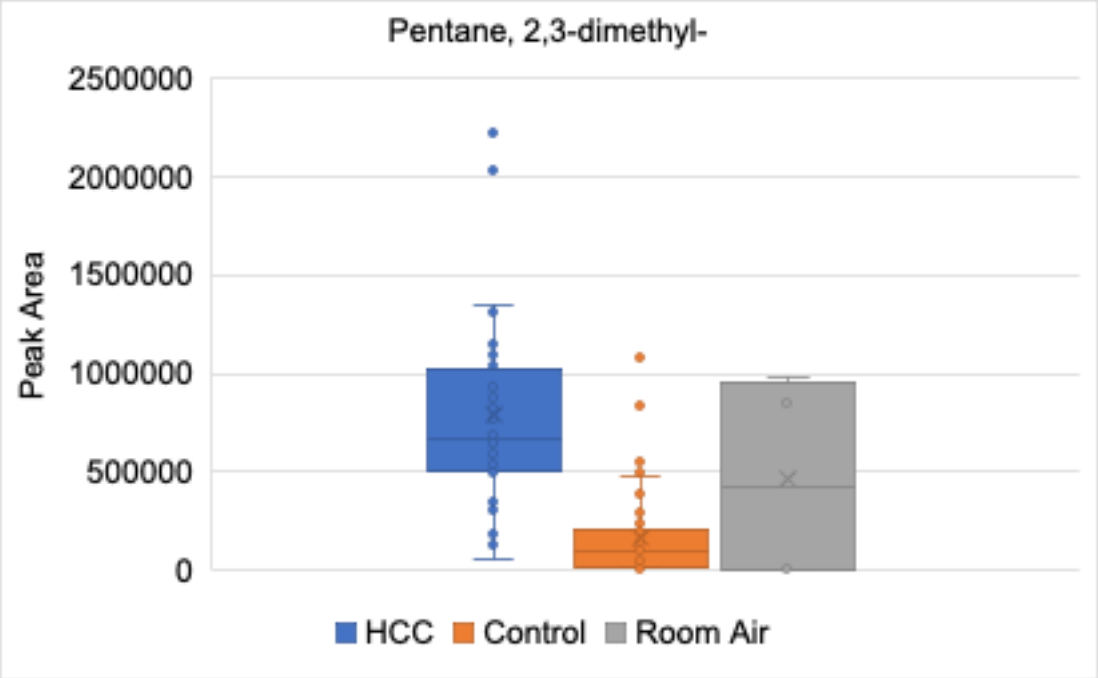


Figure 17.47 Box plot of pentane, 2,3-dimethyl- for HCC vs controls demonstrating a higher abundance in those with HCC compared to controls. The abundance within room air levels was also comparable to controls.

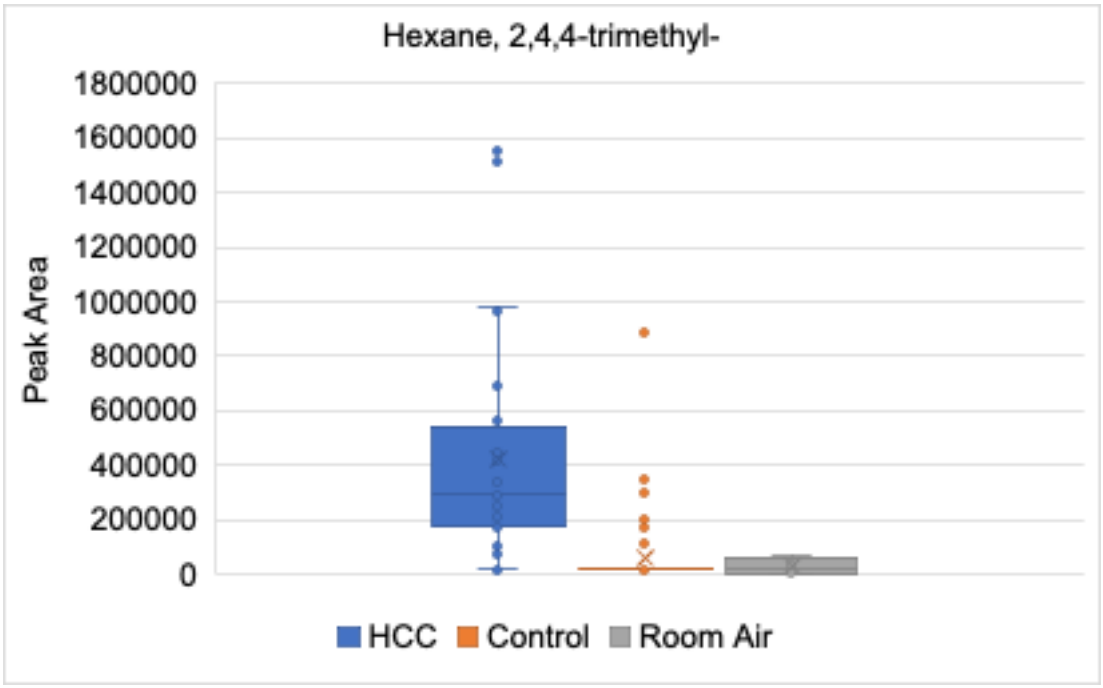


Figure 17.48 Box plot of hexane, 2,4,4-trimethyl- for HCC vs controls demonstrating a higher abundance in those with HCC compared to controls. The abundance in room air was low.

Child-Pugh A vs Child-Pugh B/C Cirrhosis

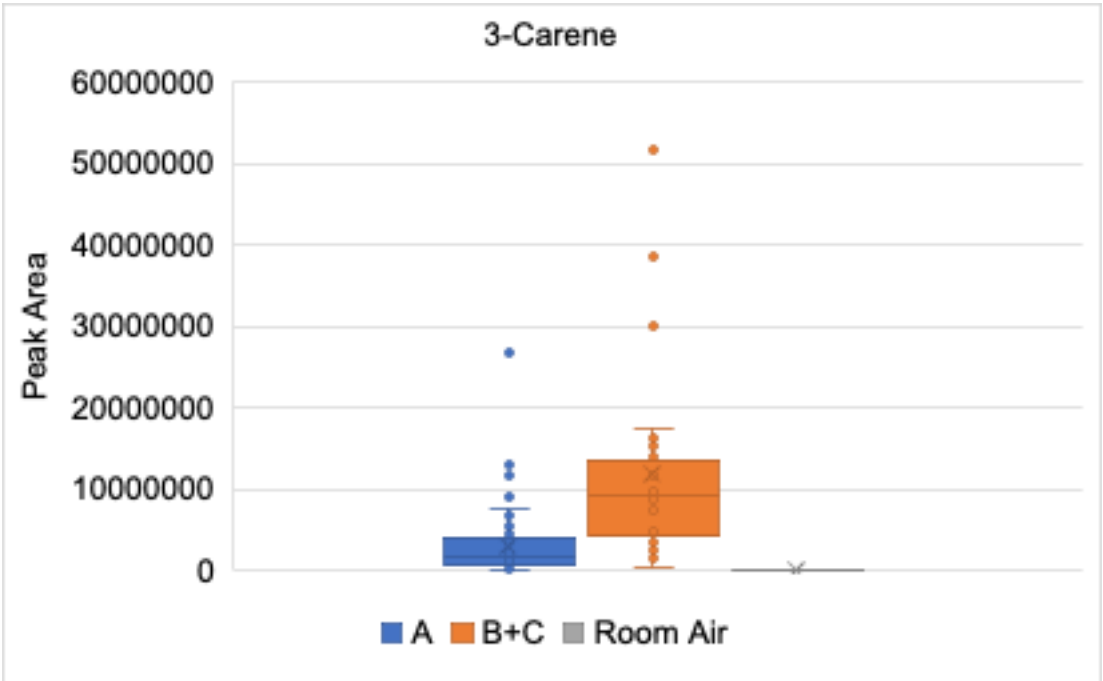


Figure 17.49 Box plot of 3-carene for Child-Pugh A cirrhosis vs Child-Pugh B cirrhosis demonstrating a higher abundance in those with Child-Pugh B/C cirrhosis. Abundance in room air was low.

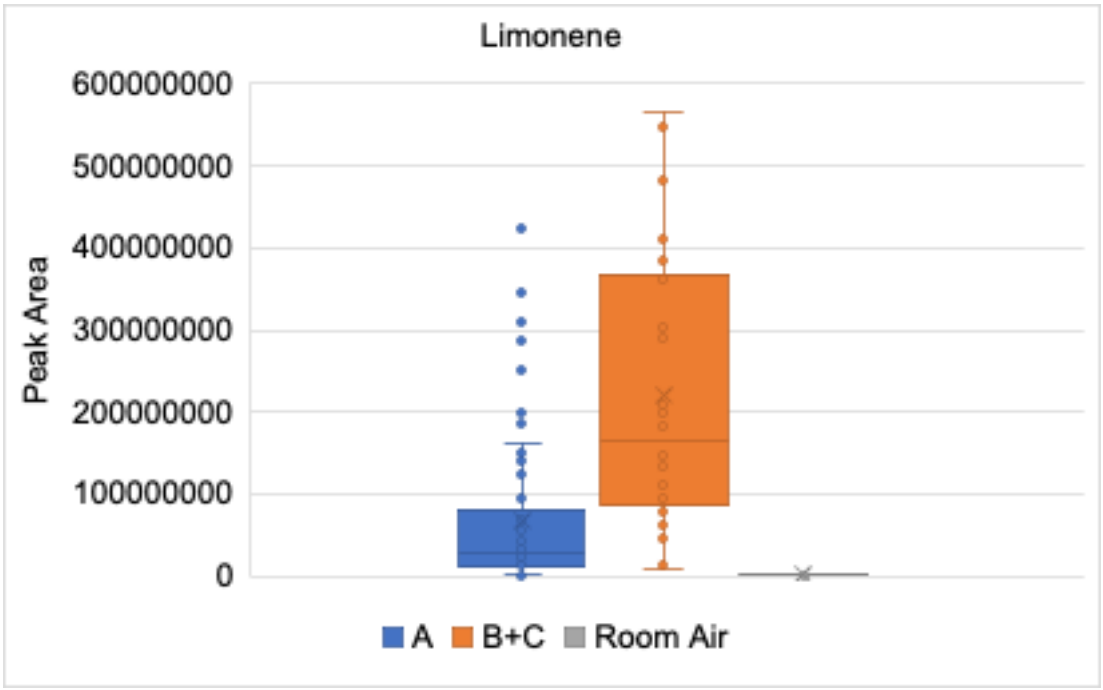


Figure 17.50 Box plot of limonene for Child-Pugh A cirrhosis vs Child-Pugh B cirrhosis demonstrating a higher abundance in those with Child-Pugh B/C cirrhosis. The abundance of limonene in room air was low.

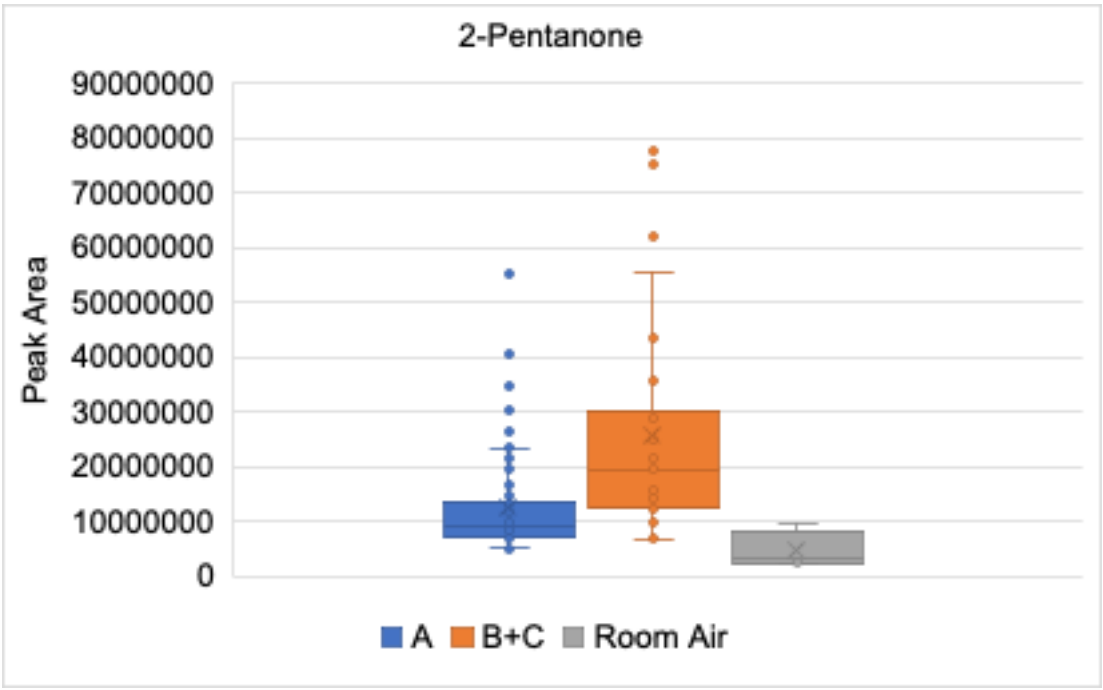


Figure 17.51 Box Plot of 2-Pentanone for Child-Pugh A cirrhosis vs Child-Pugh B cirrhosis demonstrating a higher abundance in those with Child-Pugh B/C cirrhosis. The room air abundance was low.

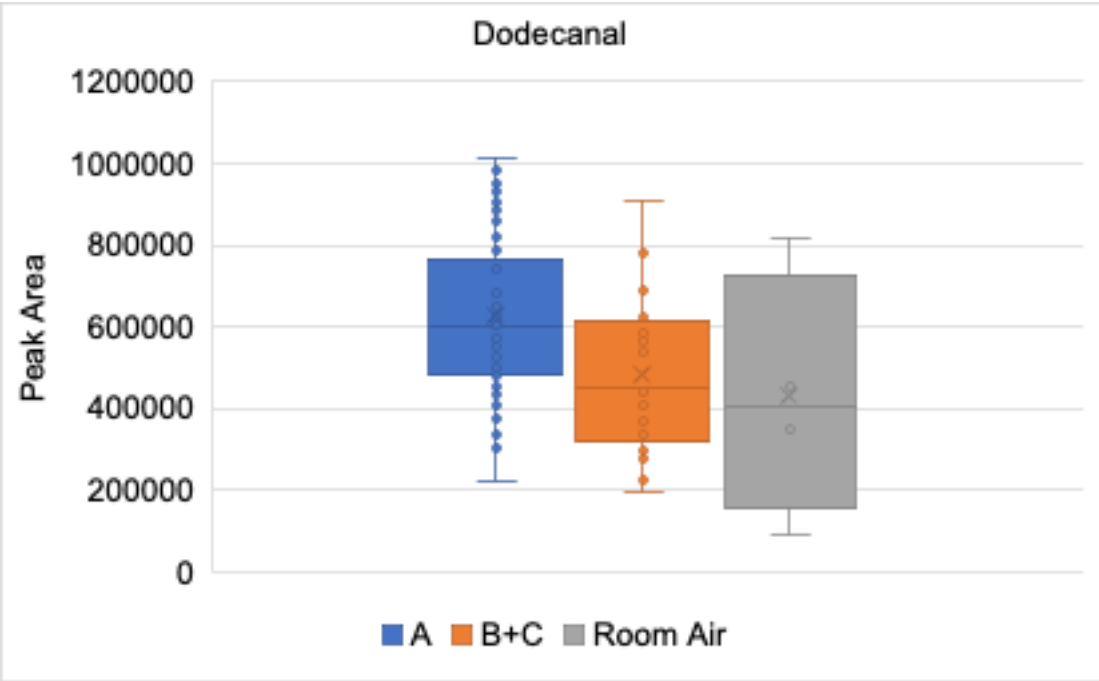


Figure 17.52 Box Plot of Dodecanal for Child-Pugh A cirrhosis vs Child-Pugh B cirrhosis demonstrating a higher abundance in those with Child-Pugh B/C cirrhosis. Dodecanal was also of high abundance in room air.

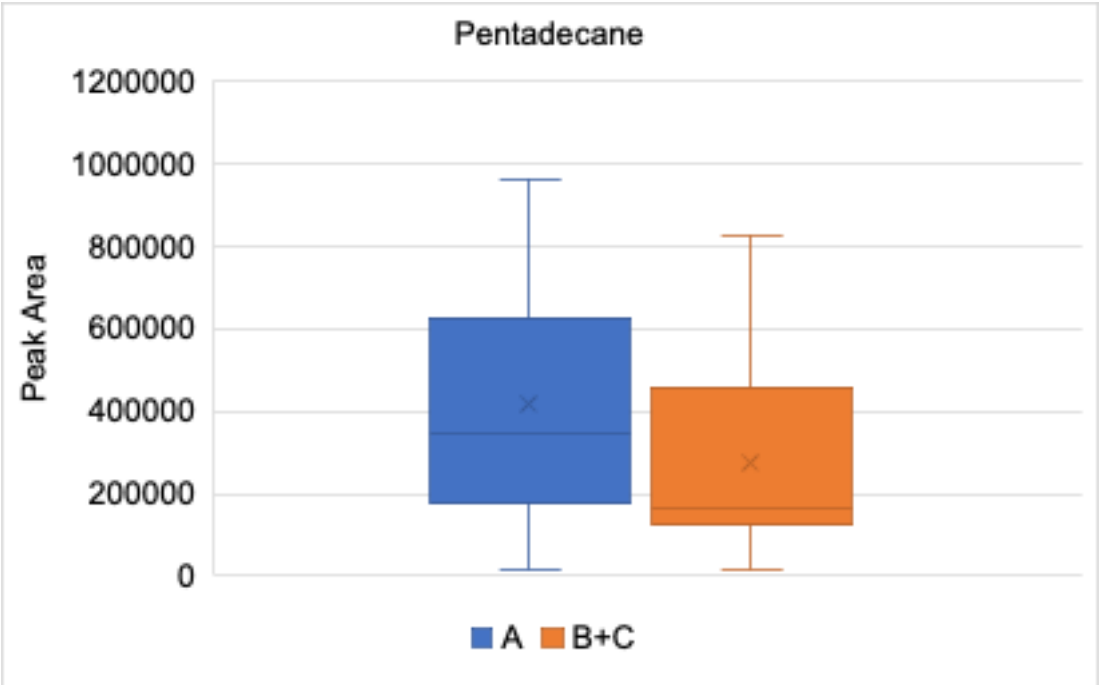


Figure 17.53 Box plot of pentadecane for Child-Pugh A cirrhosis vs Child-Pugh B cirrhosis demonstrating a higher abundance in those with Child-Pugh A cirrhosis. Pentadecane was not detected in the room air samples.

Presence or Absence of Portal Hypertension

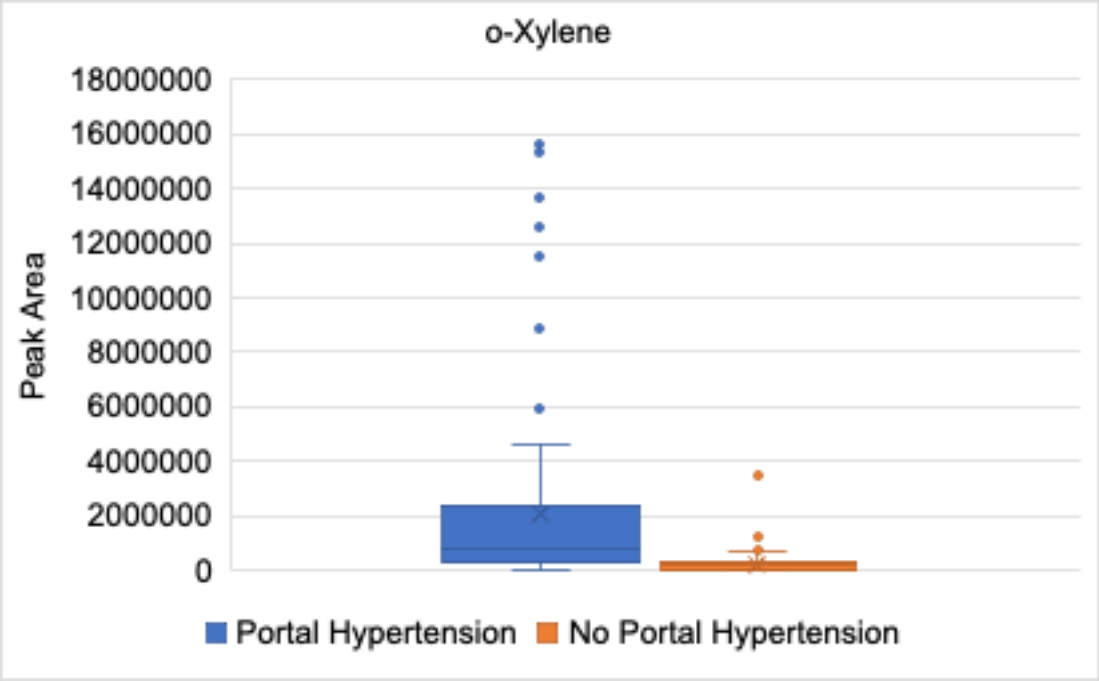


Figure 17.54 Box plot of o-xylene for presence or absence of portal hypertension demonstrating a higher abundance in those with portal hypertension. o-Xylene was not detected in room air samples.

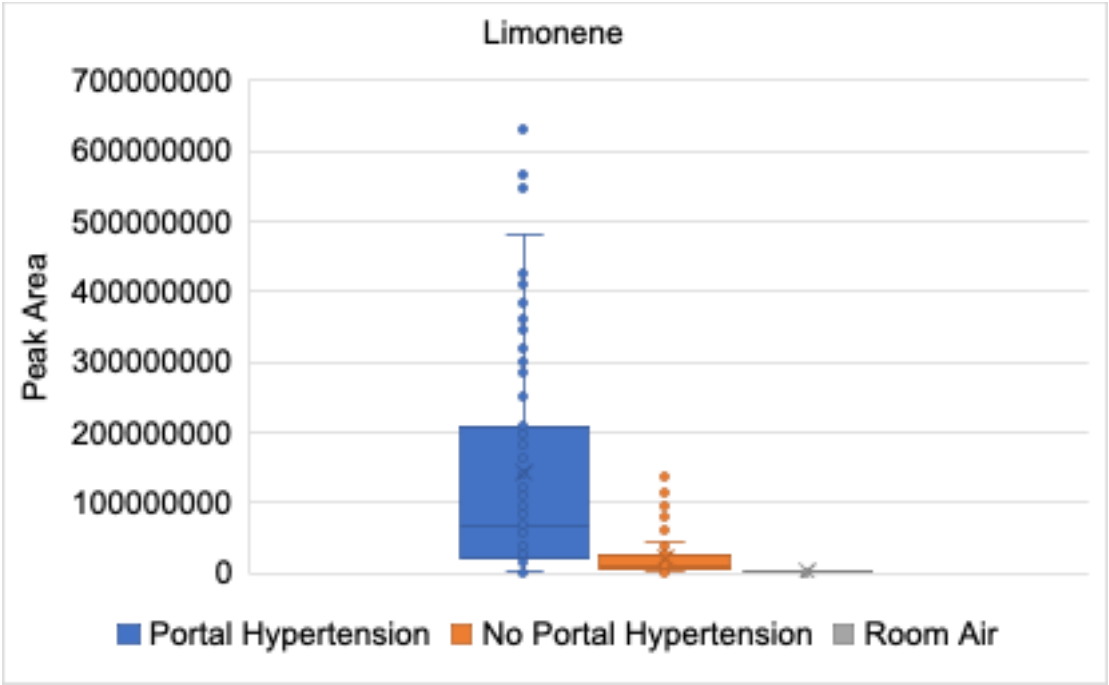


Figure 17.55 Box plot of limonene for presence or absence of portal hypertension demonstrating a higher abundance in those with portal hypertension. The abundance in room air was low.

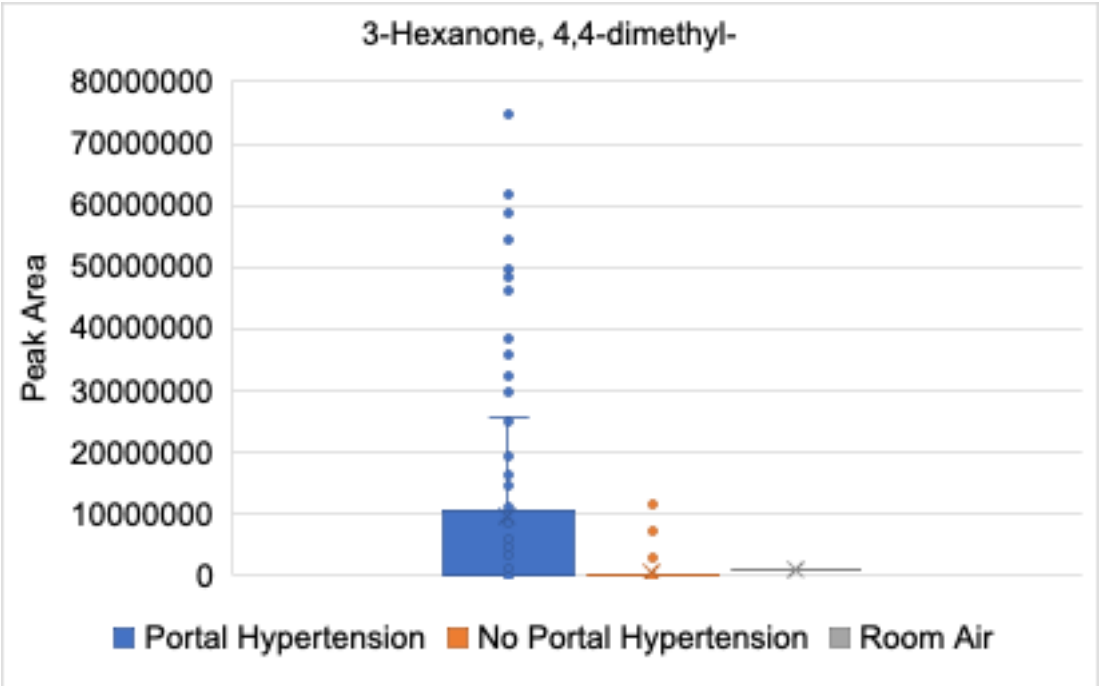


Figure 17.56 Box plot of 3-hexanone, 4,4-dimethyl- for presence or absence of portal hypertension demonstrating a higher abundance in those with portal hypertension. The abundance in room air was low.

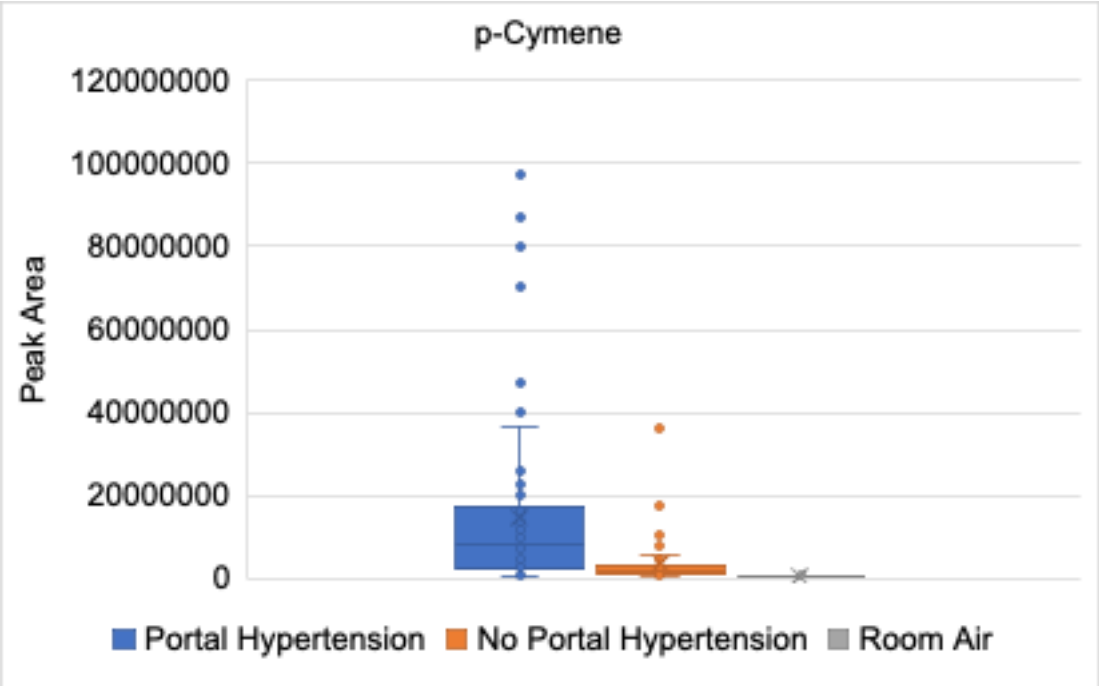


Figure 17.57 Box plot of p-cymene for presence or absence of portal hypertension demonstrating a higher abundance in those with portal hypertension. The abundance in room air was low.

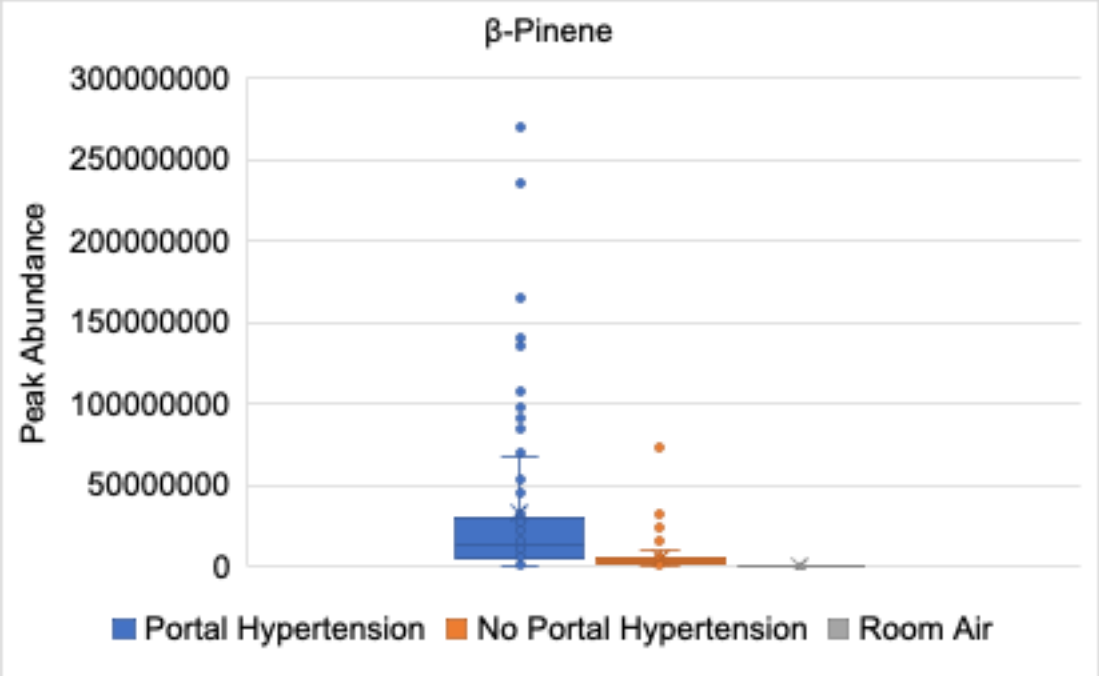


Figure 17.58 Box plot of β-pinene for presence or absence of portal hypertension demonstrating a higher abundance in those with portal hypertension. The abundance in room air was low.

Use of Long-Term Antibiotics

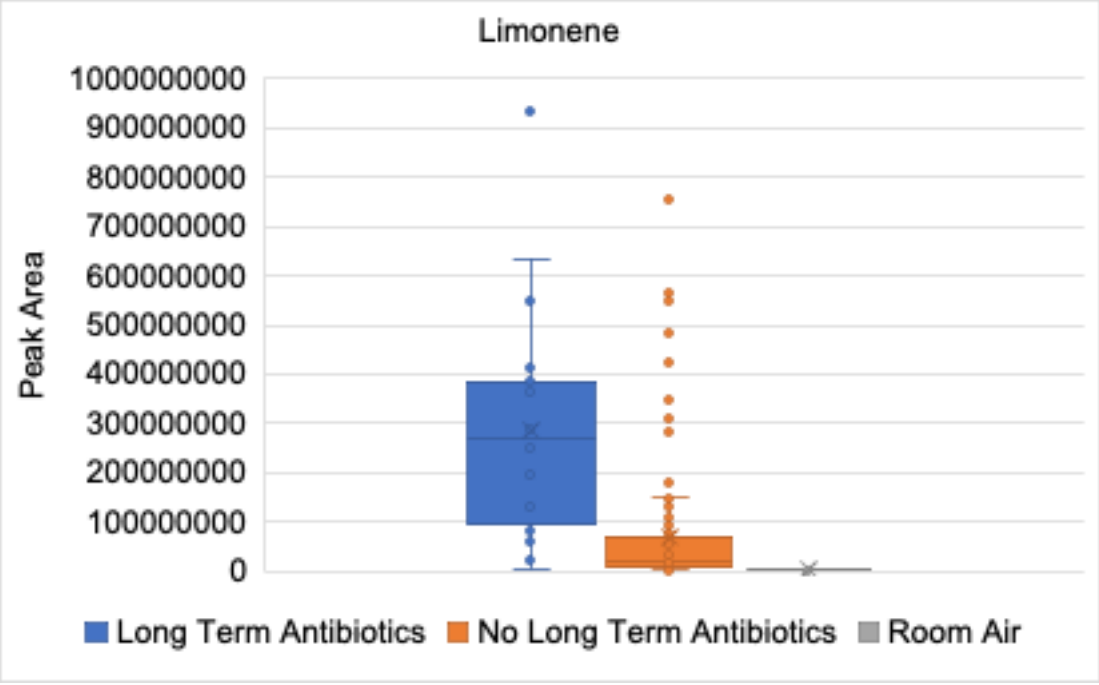


Figure 17.59 Box plot of limonene for those taking long-term antibiotics at the time of recruitment compared to those that were not. Those on long-term antibiotics had a higher abundance of limonene compared to those that were not. Abundance in room air was low.

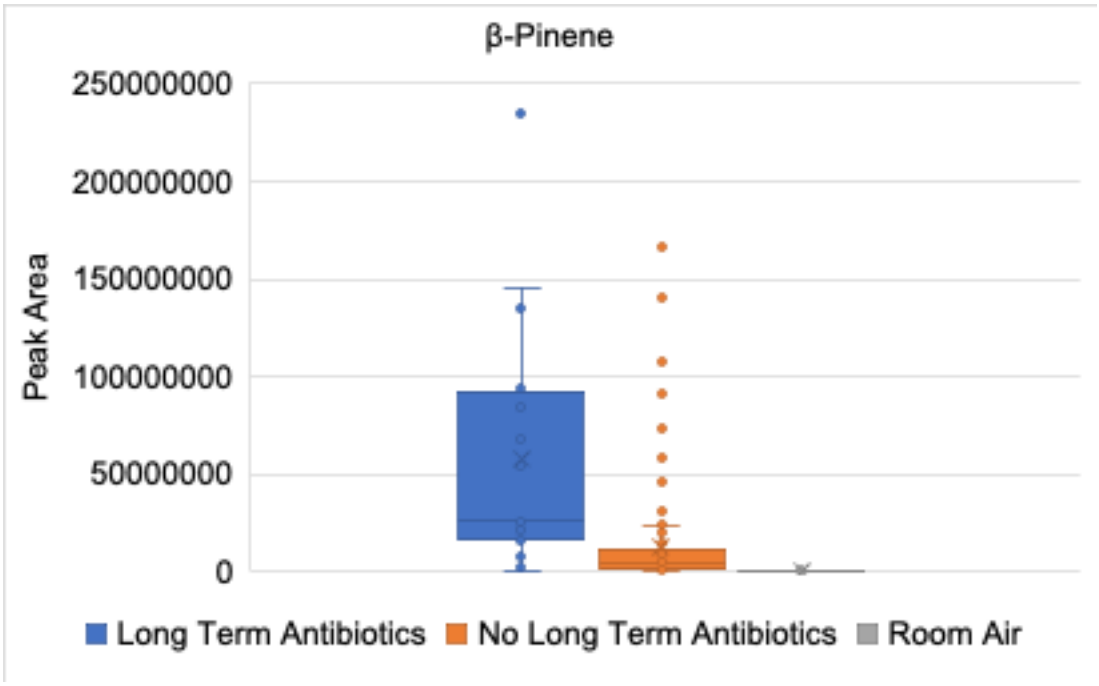


Figure 17.60 Box plot of β-pinene for those taking long-term antibiotics at the time of recruitment compared to those that were not. Those on long-term antibiotics had a higher abundance of β-pinene compared to those that were not. The room air abundance was low.

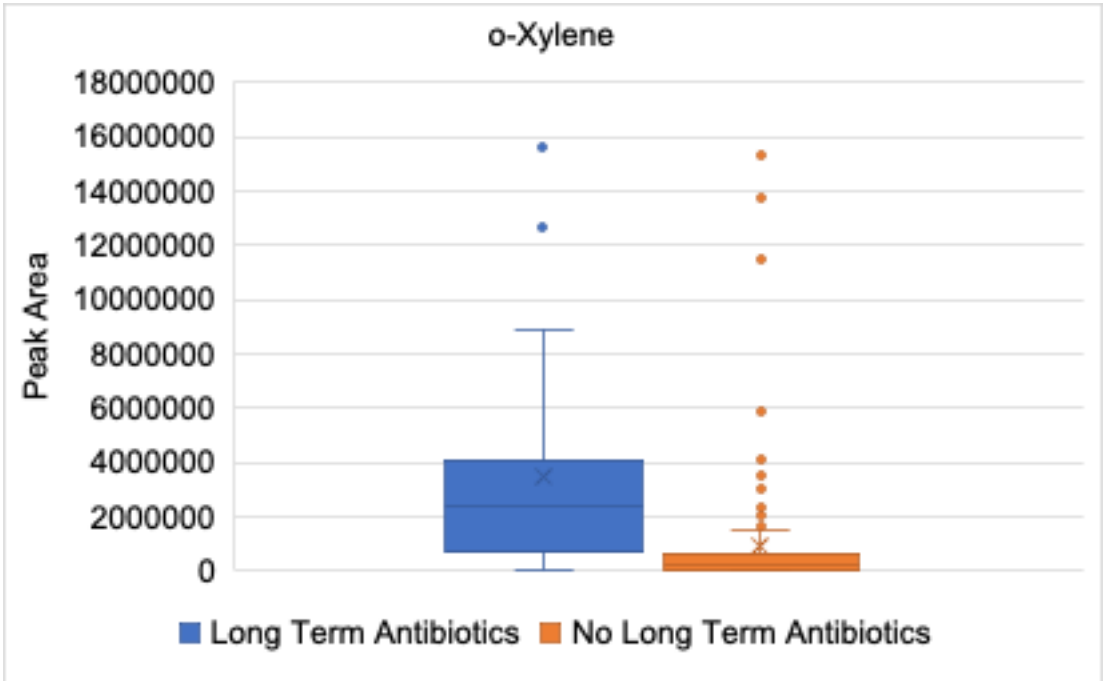


Figure 17.61 Box plot of o-Xylene for those taking long-term antibiotics at the time of recruitment compared to those that were not. Those on long-term antibiotics had a higher abundance of limonene compared to those that were not. o-Xylene was not detected in room air.

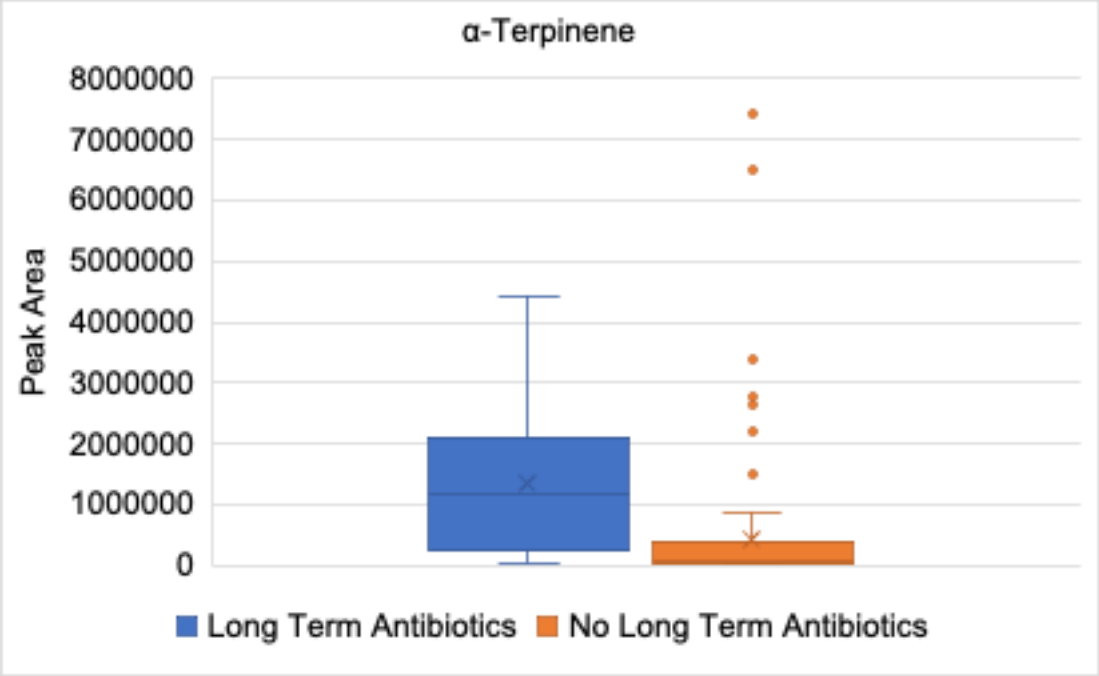


Figure 17.62 Box plot of α-terpinene for those taking long-term antibiotics at the time of recruitment compared to those that were not. Those on long-term antibiotics had a higher abundance of α-terpinene compared to those that were not. α-Terpinene was not detected in room air.

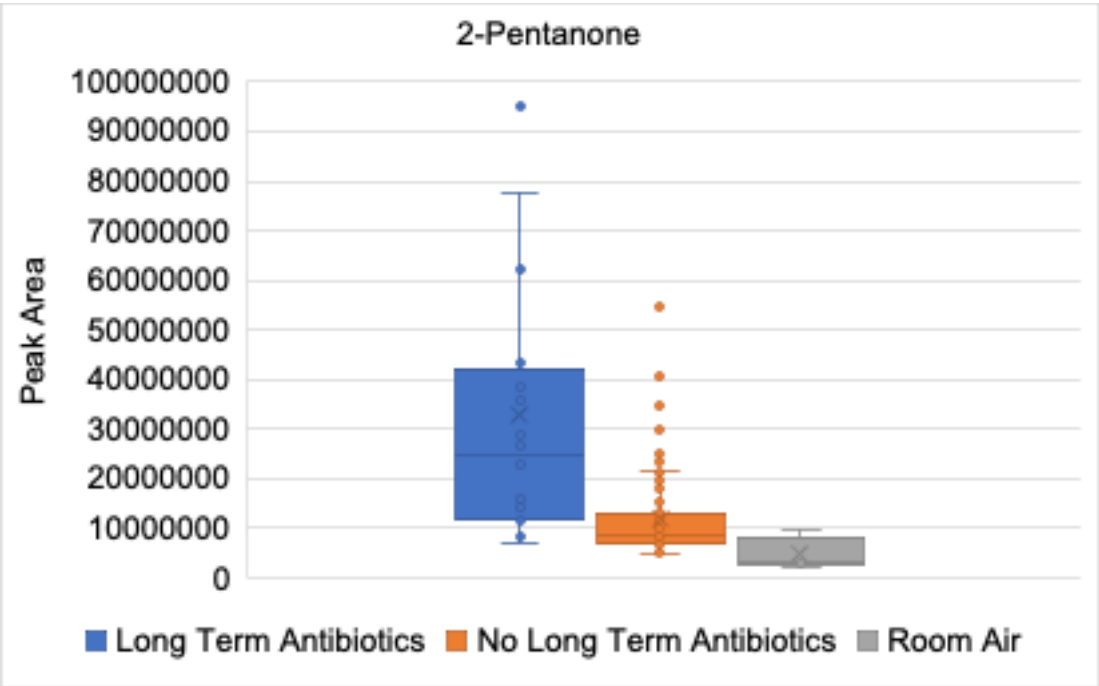


Figure 17.63 Box plot of 2-pentanone for those taking long-term antibiotics at the time of recruitment compared to those that were not. Those on long-term antibiotics had a higher abundance of 2-pentanone compared to those that were not. The room air abundance was low.

17.7.12 Compensated vs Decompensated Cirrhosis

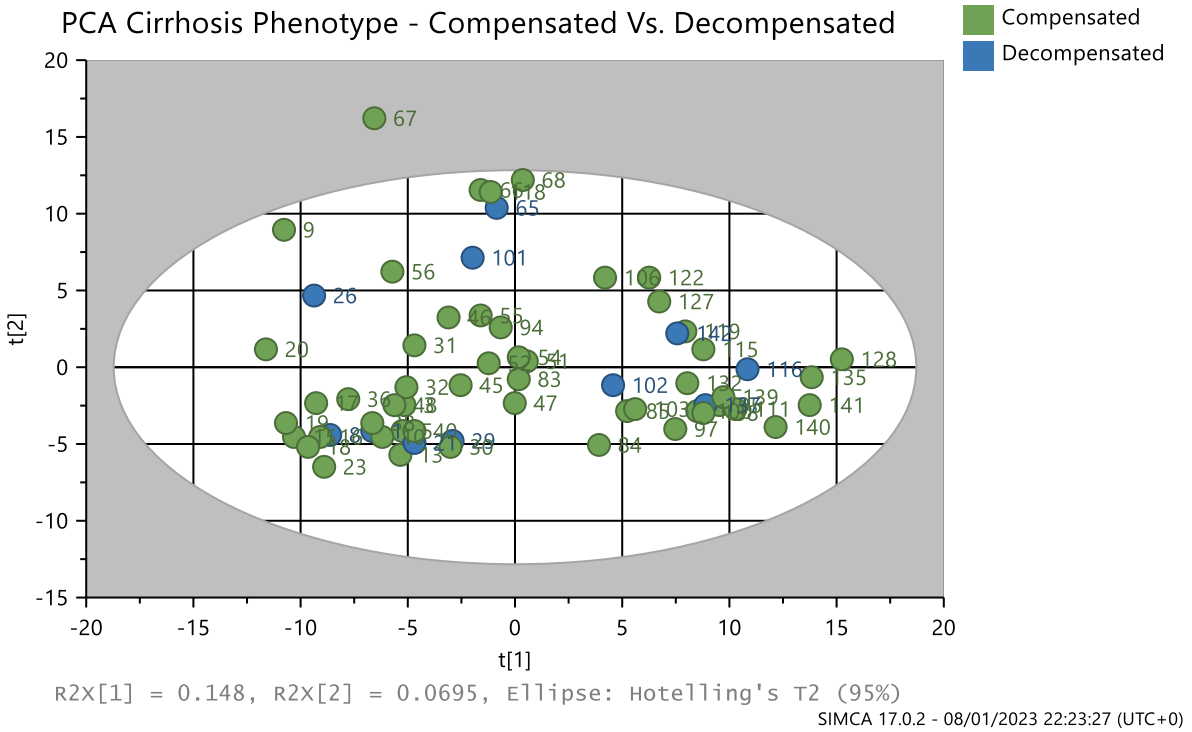


Figure 17.64 PCA of compensated vs decompensated cirrhosis demonstrating no separation.

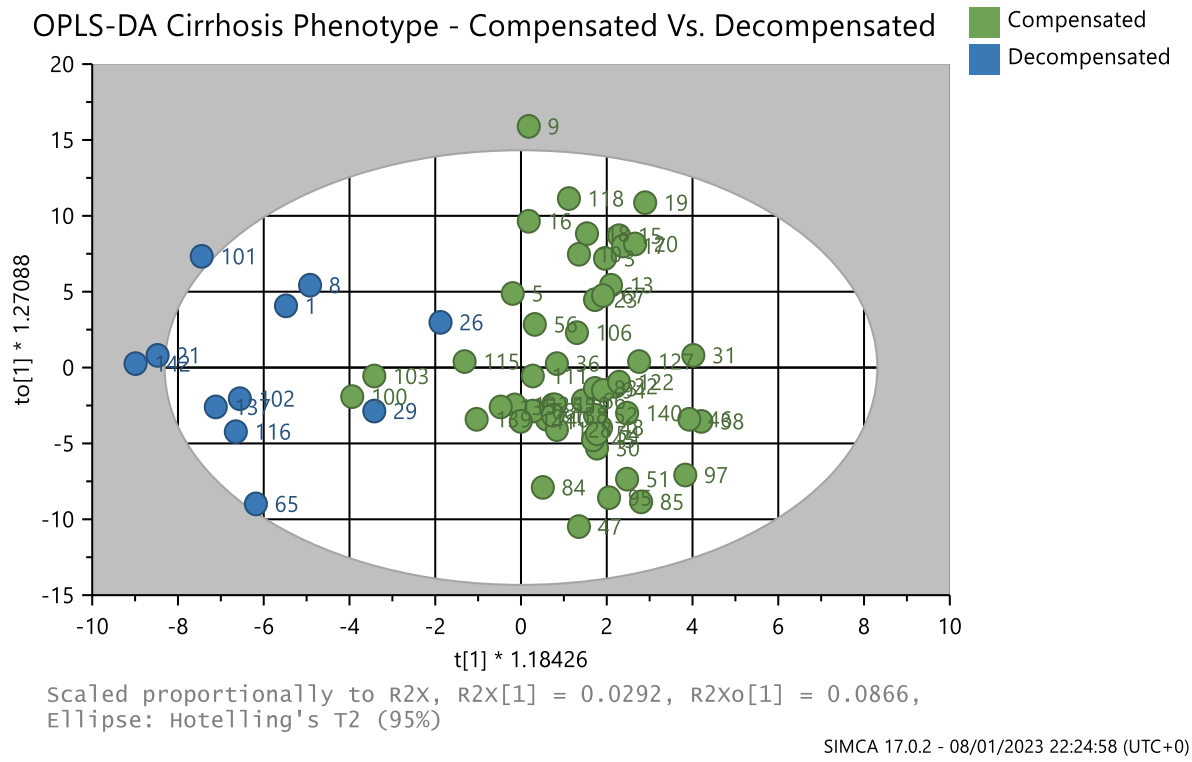


Figure 17.65 OLPS-DA compensated vs decompensated cirrhosis demonstrating good separation. However, this separation is not significant according to CV ANOVA (p = 1.0).

17.7.13 Cirrhosis Aetiology

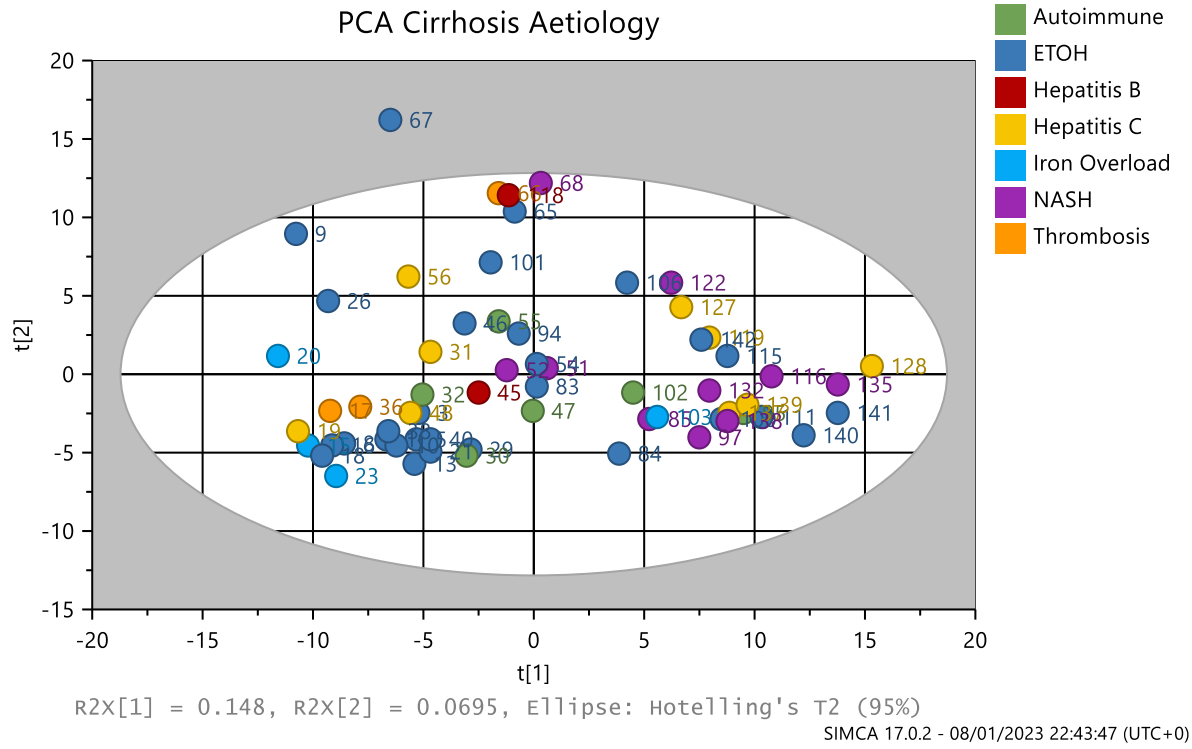


Figure 17.66 PCA of cirrhosis divided by aetiology demonstrating no separation.

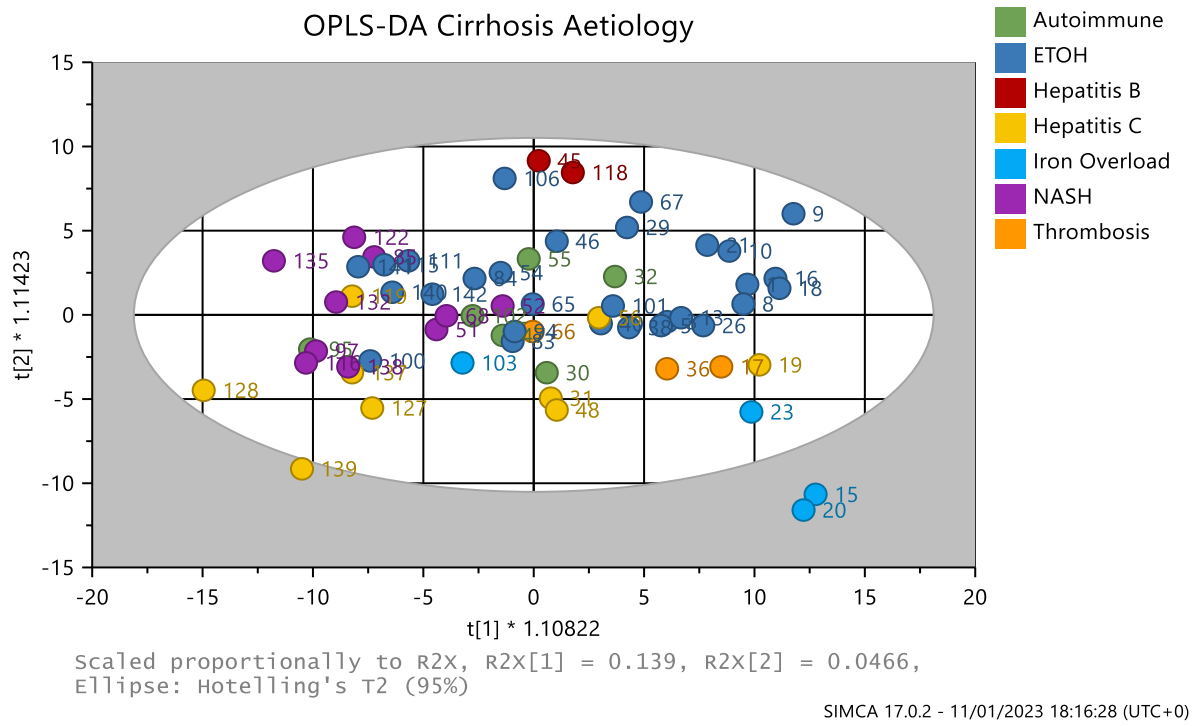


Figure 17.67 OPLS-DA of cirrhosis divided by aetiology demonstrating minimal separation which was not significant.

17.7.14 HCC Barcelona Clinic Classification

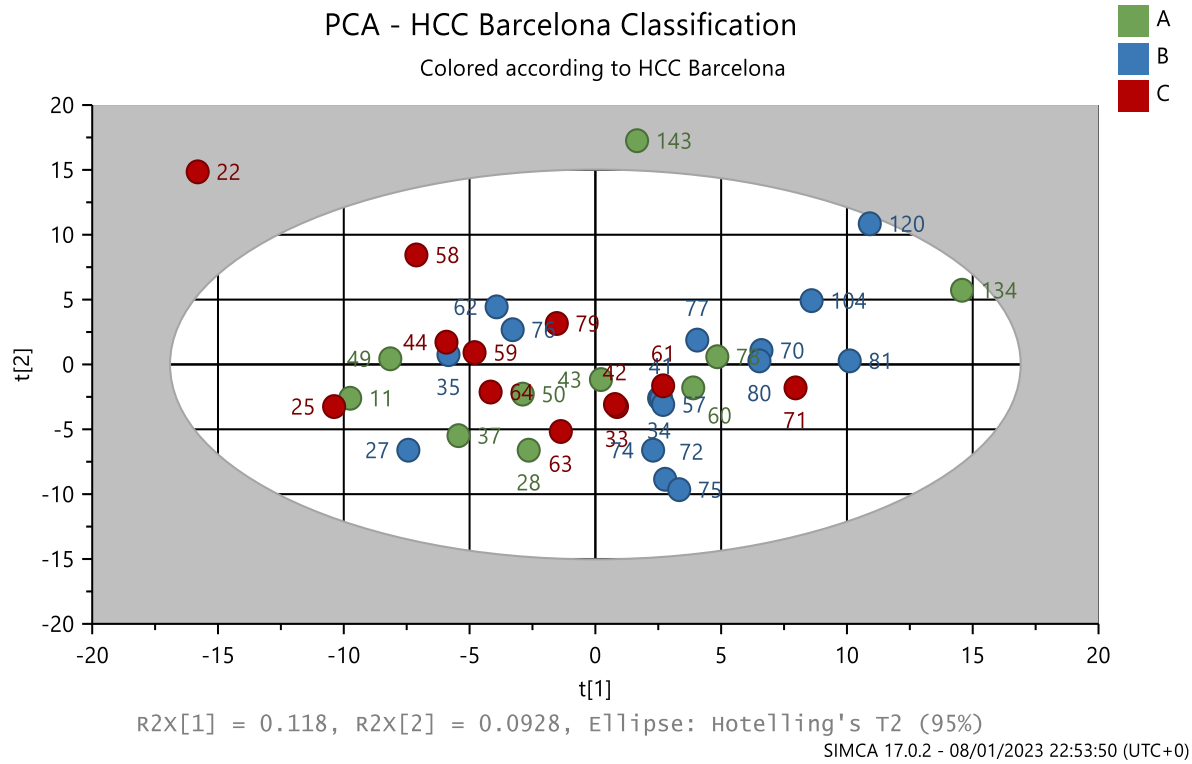


Figure 17.68 PCA of Barcelona classification of HCC participants demonstrating no separation.

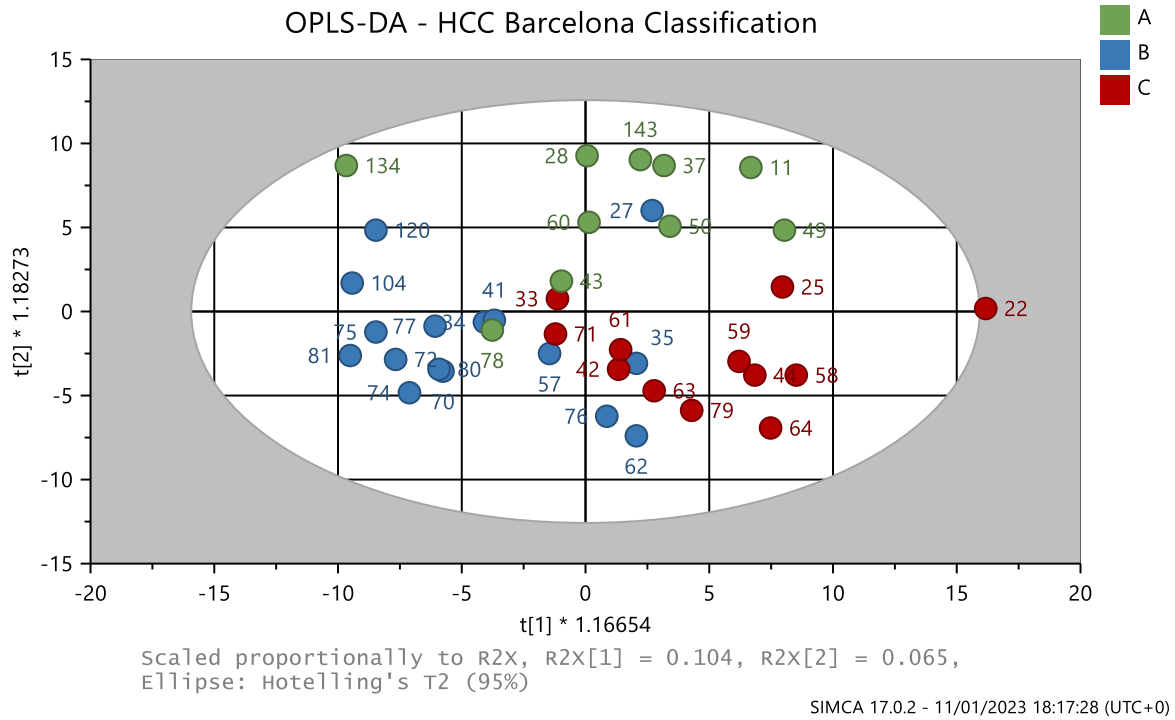


Figure 17.69 OPLS-DA of Barcelona classification of HCC participants demonstrating some separation which is not statistically significant.

17.7.15 UKELD Score

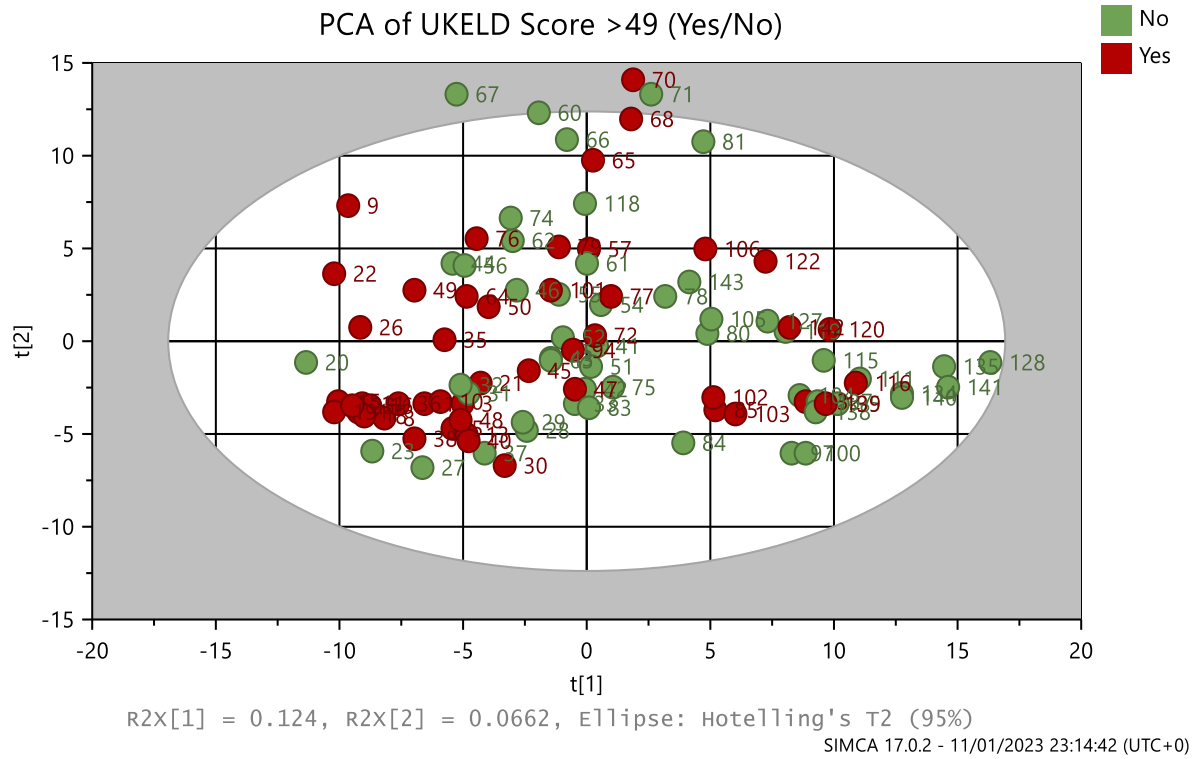


Figure 17.70 PCA of UKELD Score >49 for those individuals with cirrhosis demonstrating no separation between the two cohorts.

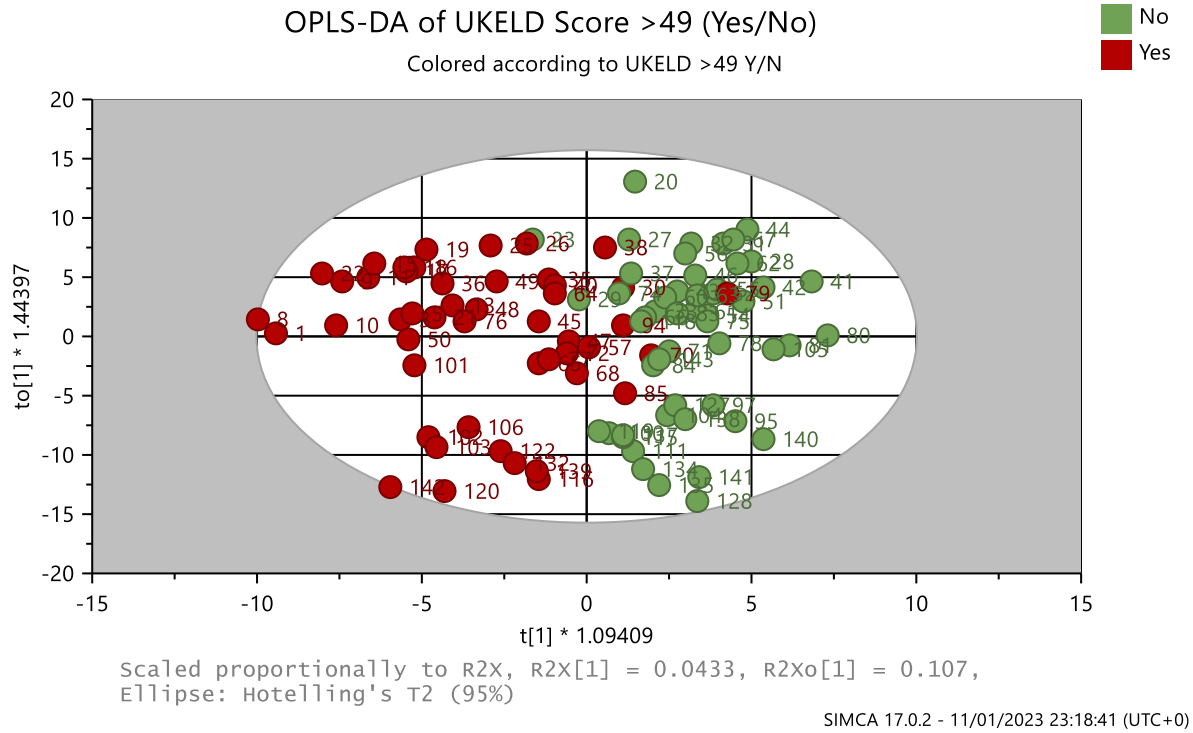


Figure 17.71 OPLS-DA of UKELD Score >49 for those individuals with cirrhosis demonstrating separation between the two cohorts, which was not statistically significant.

17.7.16 Hepatic Encephalopathy

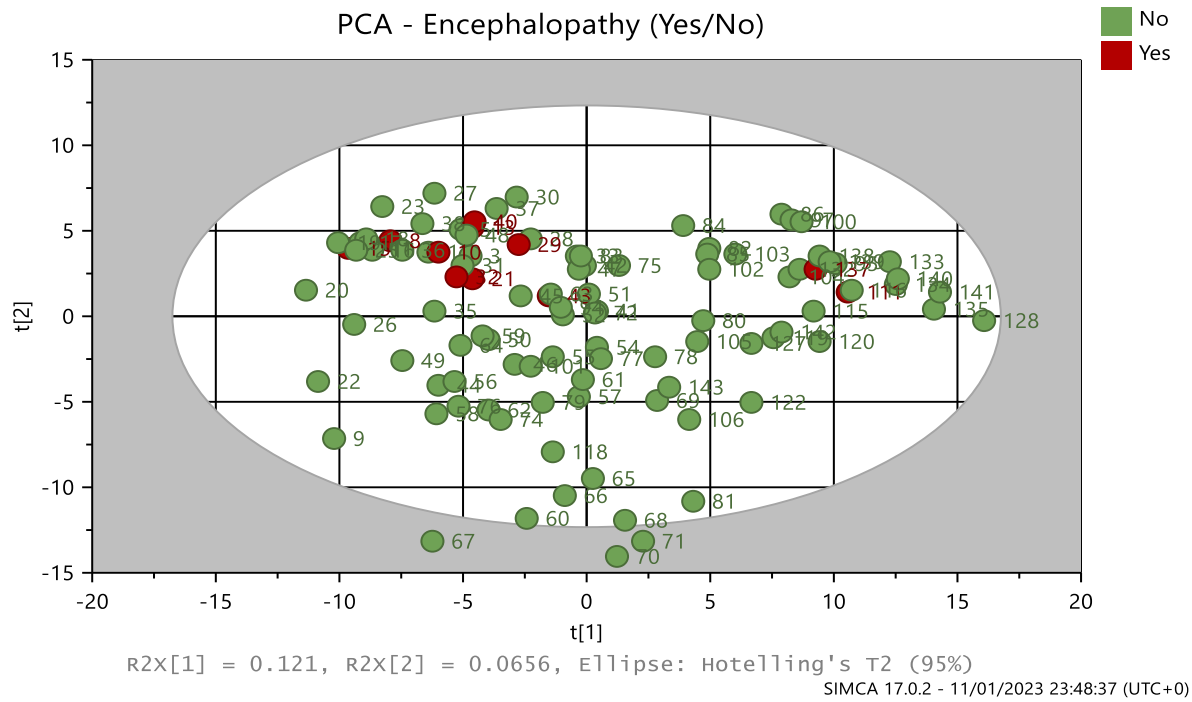


Figure 17.72 PCA of presence or absence of hepatic encephalopathy demonstrating no separation.

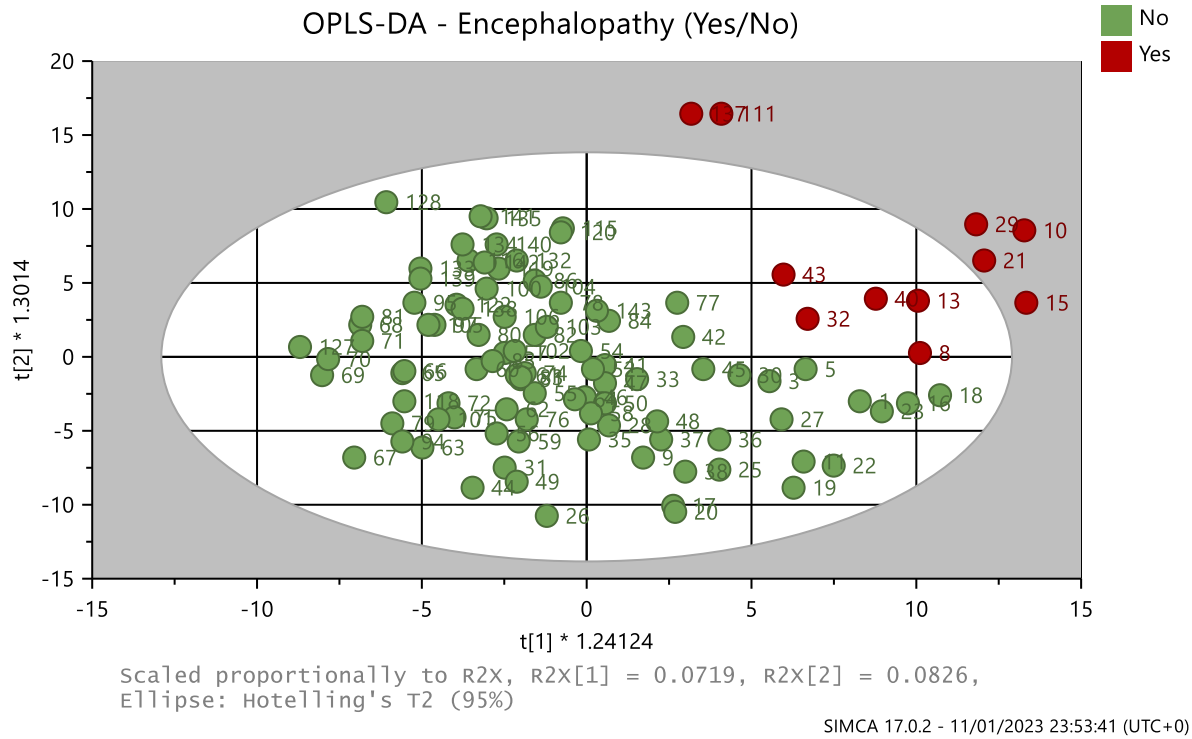


Figure 17.73 OPLS-DA of presence or absence of hepatic encephalopathy demonstrating some separation. This separation however was not significant.

17.7.17 Box Plots of Active Smokers vs Non-Smokers

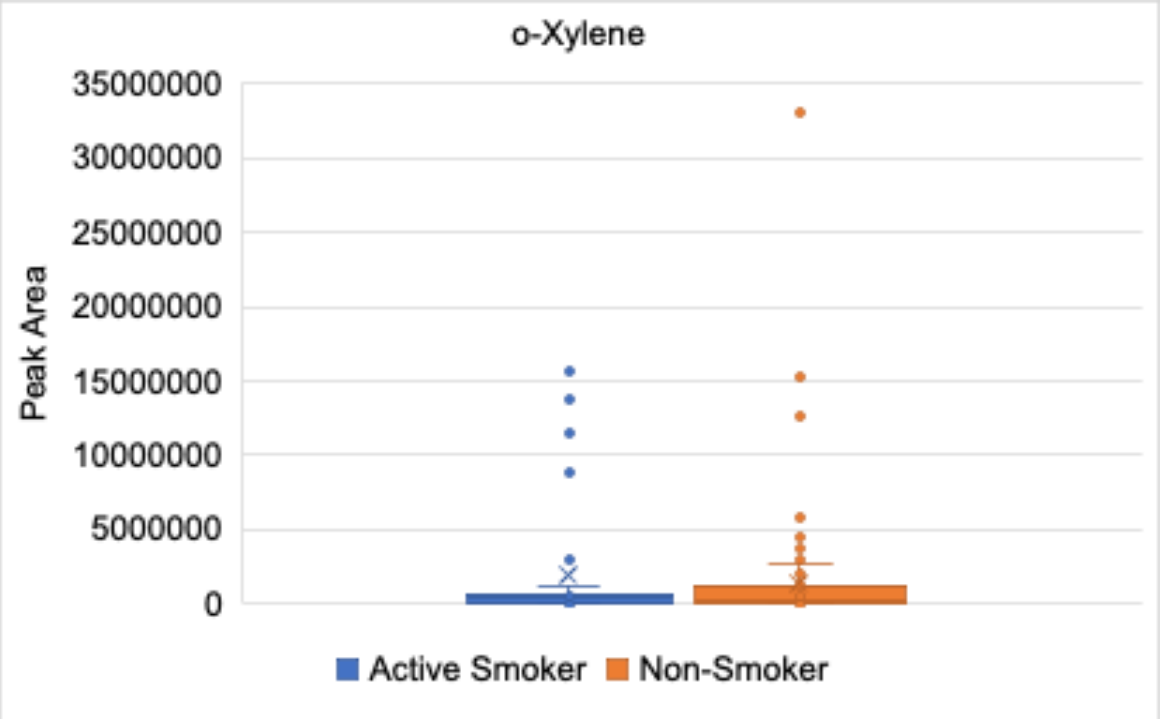


Figure 17.74 Box plot of o-xylene in active smokers and non-smokers demonstrating a higher abundance in non-smokers. o-Xylene was not detected in room air.

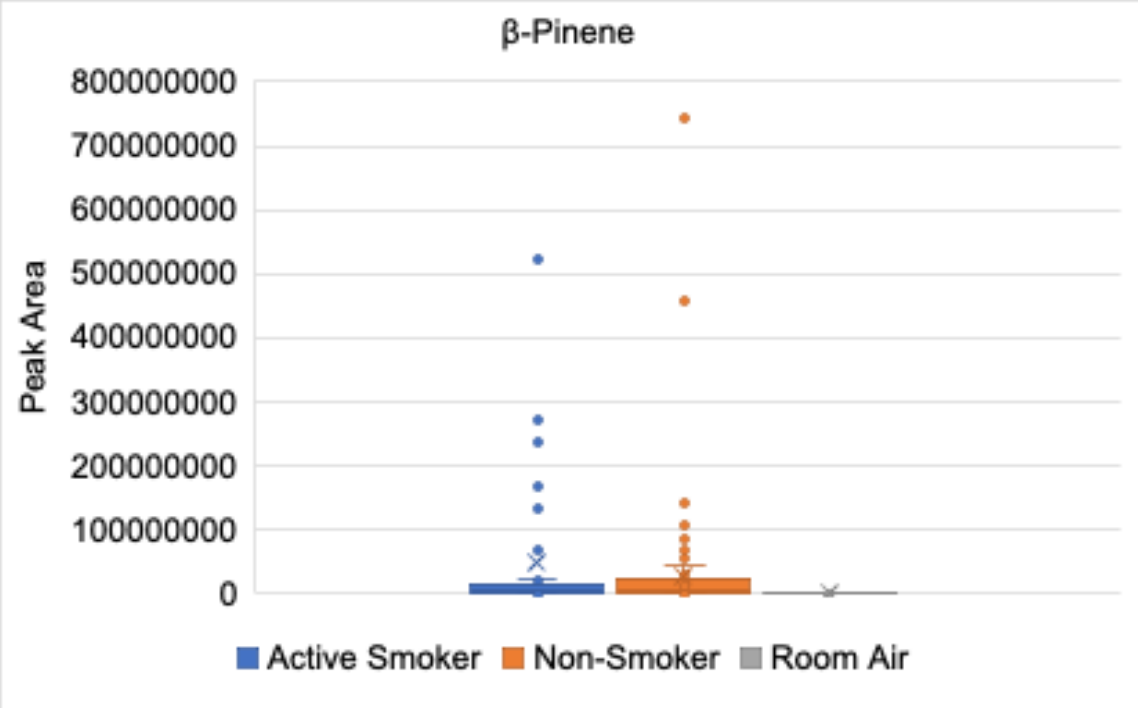


Figure 17.75 Box plot of β-Pinene in active smokers and non-smokers demonstrating a higher abundance in non-smokers. The abundance in room air was low.

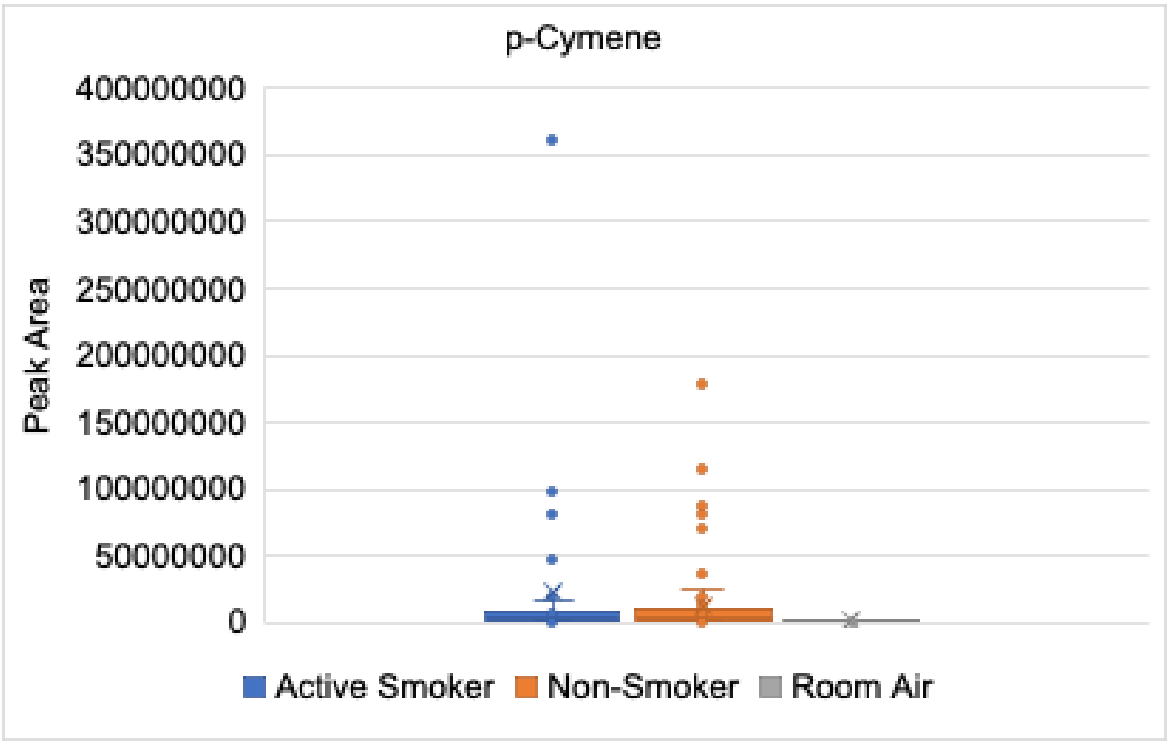


Figure 17.76 Box plot of p-cymene in active smokers and non-smokers demonstrating a higher abundance in non-smokers. The abundance in room air was low.

17.7.18 TD Tube Storage

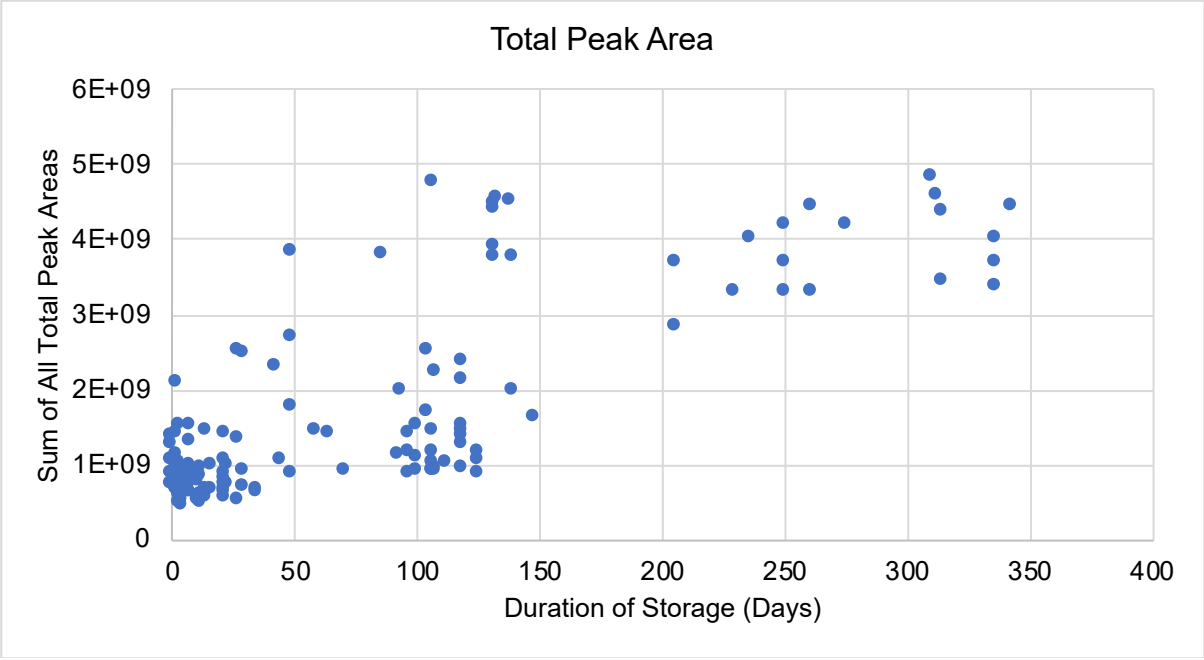


Figure 17.77 Scatter graph of total of all VOC peak areas for each sample against the time stored at -80°C demonstrating higher concentrations of VOCs on those TD tubes stored for longer periods of time.

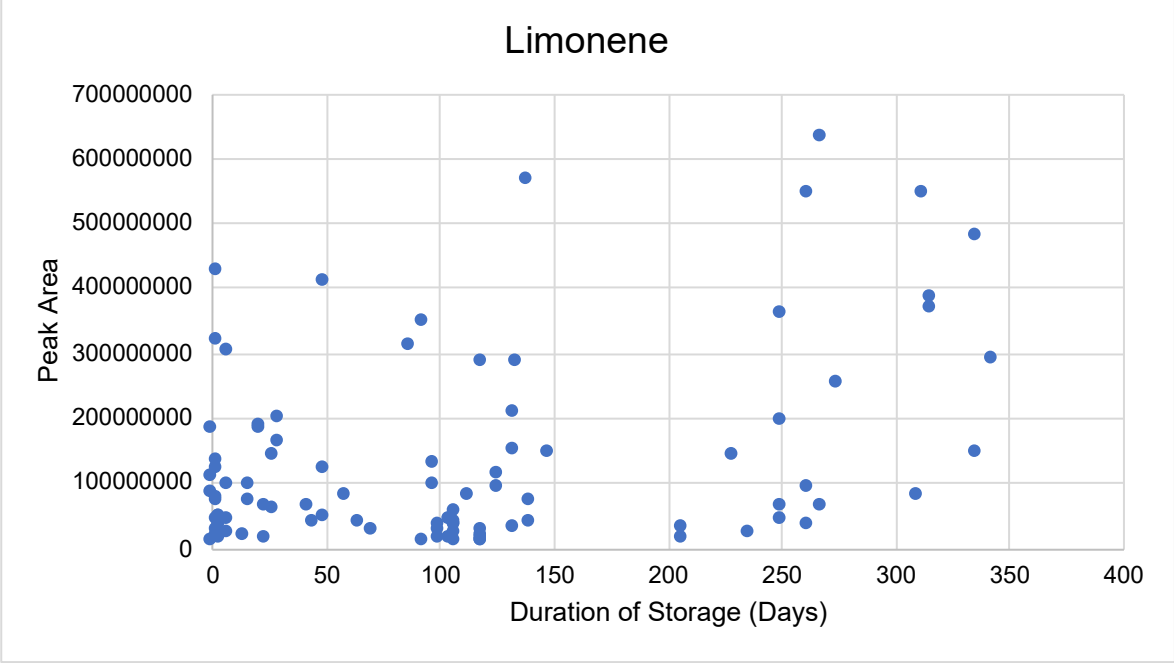


Figure 17.78 Scatter graph of peak areas for limonene of cirrhosis and HCC participant samples, plotted against time stored.

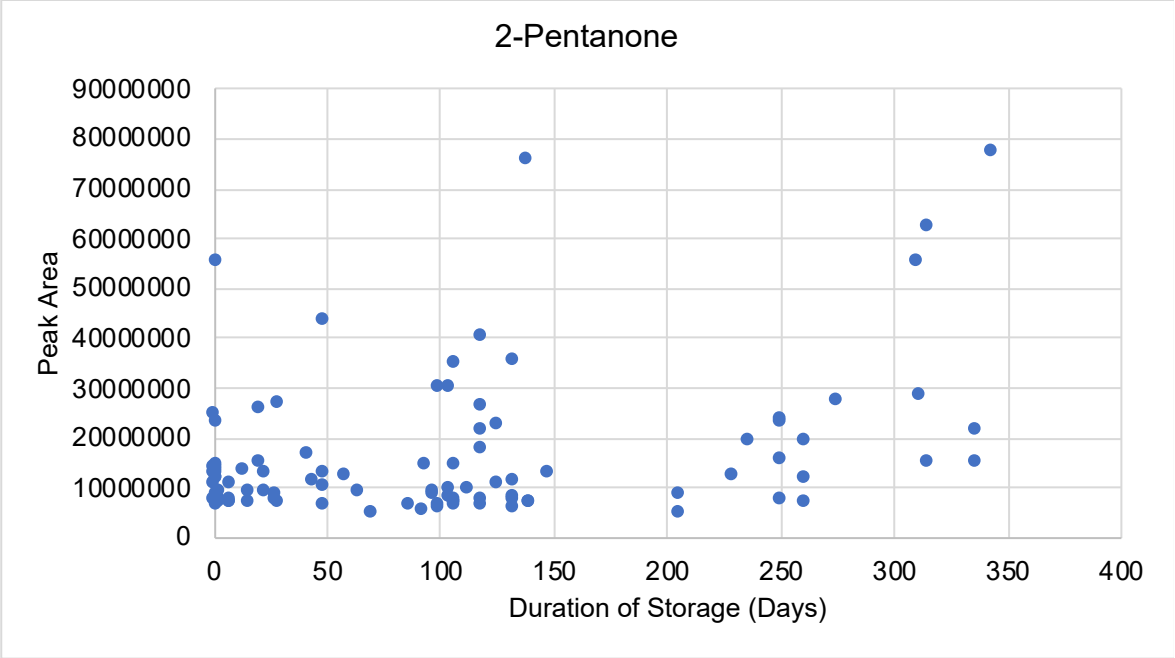


Figure 17.79 Scatter graph of peak areas for 2-pentanone of cirrhosis and HCC participant samples, plotted against time stored.

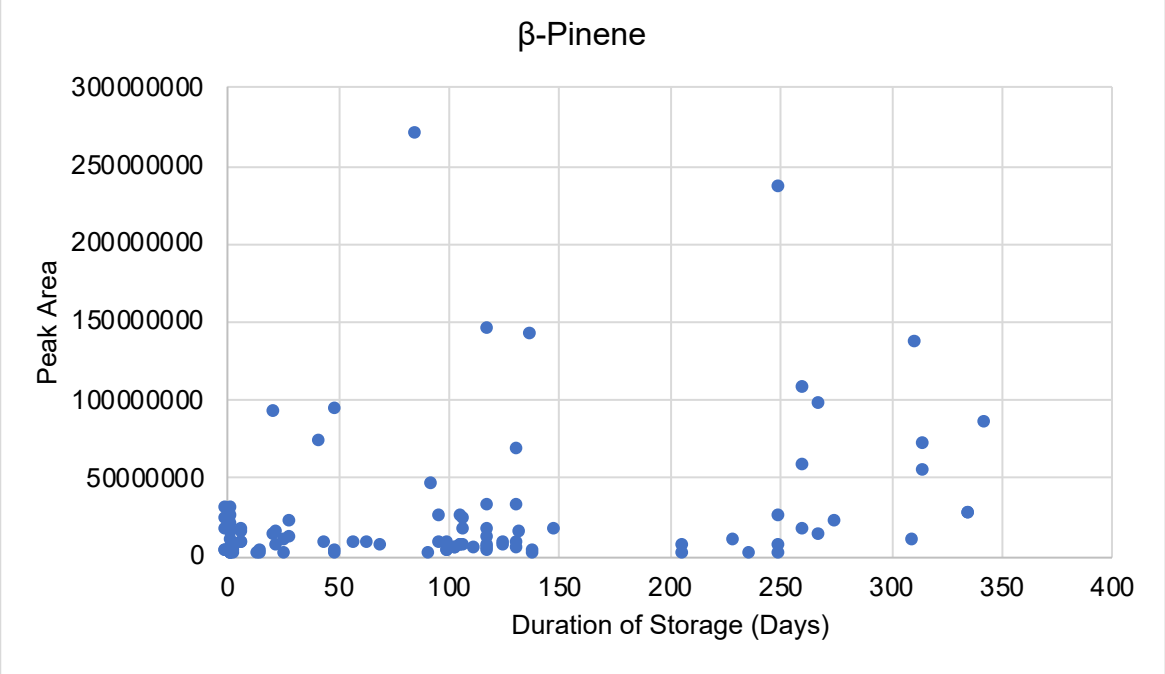


Figure 17.80 Scatter graph of peak areas for β-Pinene of cirrhosis and HCC participant samples, plotted against time stored.

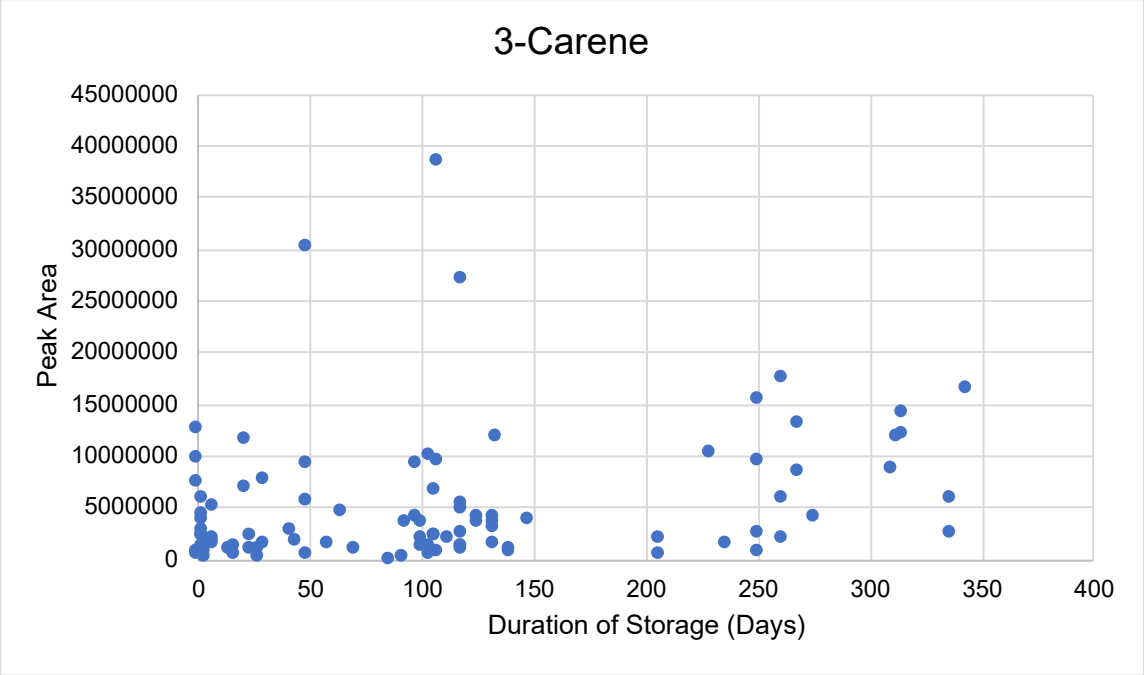


Figure 17.81 Scatter graph of peak areas for 3-carene of cirrhosis and HCC participant samples, plotted against time stored.

17.8 Table of Permissions

Page No.	Type of Work	Name of Work	Source of Work	Copyright Holder and contact	Date Permission Requested	Permission Granted (Yes/No)	Permission Note
32	Figure	Figure 1.2 Graph of UK standardised mortality rate (SMR) data for CLD, demonstrating a significant increase in SMR since 1970 for liver disease.	Addressing liver disease in the UK: a blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity, and viral hepatitis. Lancet (London, England), 384(9958), 1953-1997. 10.1016/S0140-6736(14)61838-9	© 2014 The Lancet	20.07.22	Yes	Permission via email – reference 5353020210179.
37	Table	Table 1.3 3-month mortality based on MELD score	Model for end-stage liver disease (MELD) and allocation of donor livers. Gastroenterology (New York, N.Y. 1943), 124(1), 91-96. 10.1053/gast.2003.50016	© 2003 American Gastroenterological Association	20.07.22	Yes	Permission via email.
39	Figure	Figure 1.3 EASL's diagnostic algorithm and recall policy in cirrhotic livers.	EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. (2018). Journal of Hepatology, 69(1), 182-236. 10.1016/j.jhep.2018.03.019	© 2018 Journal of Hepatology	22.10.22	Yes	Permission via email – reference 5414190467582.

40	Table	Table 1.4 Classification of HCC via TNM staging algorithm and AJCC classification.	The 8th Edition American Joint Committee on Cancer Staging for Hepato-pancreato-biliary Cancer: A Review and Update. Archives of Pathology & Laboratory Medicine (1976), 145(5), 543-553. 10.5858/arpa.2020-0032-RA	© 2021 College of American Pathologists	22.10.22	Yes	Permission via email.
41	Table	Table 1.5 Barcelona Clinic Liver Cancer (BCLC) Classification of HCC.	Hepatocellular Carcinoma (HCC) Staging Systems. Chinese Clinical Oncology, 2(4), 33. 10.3978/j.issn.2304-3865.2013.07.05	Copyright © 2009 - 2023 AME Publishing Company	22.10.22	Yes	Permission via email.
42	Figure	Figure 1.4 EASL HCC treatment algorithm.	EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. (2018). Journal of Hepatology, 69(1), 182-236. 10.1016/j.jhep.2018.03.019	© 2018 Journal of Hepatology	22.10.22	Yes	Permission via email – reference 5414190467582.
47	Figure	Figure 2.1 A diagram of the compartment-based model for distribution of endogenous VOCs	Amann, A., & Smith, D. (2013). Volatile Biomarkers. Elsevier.	© 2013 Elsevier	12.05.23	Yes	Permission via email.

59	Figure	Figure 3.1 Schematic diagram of SIFT-MS technology.	SIFT-MS and FA-MS methods for ambient gas phase analysis: developments and applications in the UK. The Analyst, 140(8), 2573-2591. 10.1039/c4an02049a	© 2015 Royal Society of Chemistry	N/A	Yes	Creative Commons Attribution 3.0 Unported Licence.
60	Figure	Figure 3.2 Schematic diagram of a PTR-ToF-MS.	Demonstration of Proton-Transfer Reaction Time-of-Flight Mass Spectrometry for Real-Time Analysis of Trace Volatile Organic Compounds. Analytical Chemistry, 76(13), 3841-3845. 10.1021/ac0498260	© 2004 American Chemical Society	20.07.22	Yes	Permission via email.
61	Figure	Figure 3.3 Schematic diagram of GC-MS coupled with thermal desorption.	Methods in plant foliar volatile organic compounds research1. Applications in Plant Sciences, 3(12), apps.1500044. 10.3732/apps.1500044	© 2015 Botanical Society of America	12.05.23	Yes	Permission via email.
232	Figure	Appendix 17.1 STARD 2015 Checklist	STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open, 6(11), e012799. 10.1136/bmjopen-2016-012799	© 2023 BMJ Publishing Group Ltd.	20.07.22	Yes	Creative Commons Attribution Non-Commercial (CC BY-NC 4.0) license.

18 REFERENCES

(1) Hewitt MJ, Belluomo I, Zuffa S, Boshier PR, Myridakis A. Variation of volatile organic compound levels within ambient room air and its impact upon the standardisation of breath sampling. *Scientific Reports.* 2022; 12 (1): 15887. 10.1038/s41598-022-20365-7.

(2) Sharma A, Nagalli S. Chronic Liver Disease. *StatPearls.* Treasure Island (FL): StatPearls Publishing; 2022.

(3) Zhou W, Zhang Q, Qiao L. Pathogenesis of liver cirrhosis. *World Journal of Gastroenterology : WJG.* 2014; 20 (23): 7312-7324. 10.3748/wjg.v20.i23.7312.

(4) Schwartz JM, Carrithers RL. *Epidemiology and etiologic associations of hepatocellular carcinoma.* https://www.uptodate.com/contents/epidemiology-and-etiologic-associations-of-hepatocellular-carcinoma [Accessed 1st October 2018].

(5) Banales JM, Marin JJG, Lamarca A, Rodrigues PM, Khan SA, Roberts LR, et al. Cholangiocarcinoma 2020: the next horizon in mechanisms and management. *Nature Reviews Gastroenterology & Hepatology.* 2020; 17 (9): 557-588. 10.1038/s41575-020-0310-z.

(6) Poynard T, Mathurin P, Lai C, Guyader D, Poupon R, Tainturier M, et al. A comparison of fibrosis progression in chronic liver diseases. *Journal of Hepatology.* 2003; 38 (3): 257-265. 10.1016/S0168-8278(02)00413-0.

(7) Asrani SK, Kamath PS. Natural history of cirrhosis. *Current Gastroenterology Reports.* 2013; 15 (2): 308. 10.1007/s11894-012-0308-y.

(8) Williams R, Aspinall R, Bellis M, Camps-Walsh G, Cramp M, Dhawan A, et al. Addressing liver disease in the UK: a blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity, and viral hepatitis. *Lancet (London, England).* 2014; 384 (9958): 1953-1997. 10.1016/S0140-6736(14)61838-9.

(9) Nabi E, Bajaj JS. Useful Tests for Hepatic Encephalopathy in Clinical Practice. *Current gastroenterology reports.* 2014; 16 (1): 362. 10.1007/s11894-013-0362-0.

(10) Weissenborn K. Hepatic Encephalopathy: Definition, Clinical Grading and Diagnostic Principles. *Drugs.* 2019; 79 (Suppl 1): 5-9. 10.1007/s40265-018-1018-z.

(11) Schindler P, Heinzow H, Trebicka J, Wildgruber M. Shunt-Induced Hepatic Encephalopathy in TIPS: Current Approaches and Clinical Challenges. *Journal of Clinical Medicine.* 2020; 9 (11): 3784. 10.3390/jcm9113784.

(12) Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. *Hepatology (Baltimore, Md.).* 1996; 24 (2): 289-293. 10.1002/hep.510240201.

(13) CORAL GP, ANTUNES ADP, SERAFINI APA, ARAUJO FB, de MATTOS AA. LIVER BIOPSY: IMPORTANCE OF SPECIMEN SIZE IN THE DIAGNOSIS AND STAGING OF CHRONIC VIRAL HEPATITIS. *Revista do Instituto de Medicina Tropical de São Paulo.* 2016; 58 10.1590/S1678-9946201658010.

(14) Sanai FM, Keeffe EB. Liver Biopsy for Histological Assessment – The Case Against. *Saudi Journal of Gastroenterology : Official Journal of the Saudi Gastroenterology Association.* 2010; 16 (2): 124-132. 10.4103/1319-3767.61244.

(15) West J, Timothy R. Reduced Mortality Rates Following Elective Percutaneous Liver Biopsies. *Gastroenterology.* 2010; 139 (4): 1230-1237. 10.1053/j.gastro.2010.06.015.

(16) Srivastava A, Jong S, Gola A, Gailer R, Morgan S, Sennett K, et al. Cost-comparison analysis of FIB-4, ELF and fibroscan in community pathways for non-alcoholic fatty liver disease. *BMC Gastroenterology.* 2019; 19 (1): 122. 10.1186/s12876-019-1039-4.

(17) Lupsor Platon M, Stefanescu H, Feier D, Maniu A, Badea R. Performance of unidimensional transient elastography in staging chronic hepatitis C. Results from a cohort of 1,202 biopsied patients from one single center. *Journal of gastrointestinal and liver diseases: JGLD.* 2013; 22 (2): 157-166.

(18) Tsochatzis EA, Gurusamy KS, Ntaoula S, Cholongitas E, Davidson BR, Burroughs AK. Elastography for the diagnosis of severity of fibrosis in chronic liver disease: a meta-analysis of diagnostic accuracy. *Journal of Hepatology.* 2011; 54 (4): 650-659. 10.1016/j.jhep.2010.07.033.

(19) Venkatesh SK, Yin M, Ehman RL. Magnetic resonance elastography of liver: clinical applications. *Journal of computer assisted tomography.* 2013; 37 (6): 887-896. 10.1097/RCT.0000000000000032.

(20) Cuperus FJ, Drenth JP, Tjwa ET. *Mistakes in liver function test abnormalities and how to avoid them.* https://www.ueg.eu/education/latest-news/article/article/mistakes-in-liver-function-test-abnormalities-and-how-to-avoid-them/ [Accessed Oct 1, 2018].

(21) Peck-Radosavljevic M. Thrombocytopenia in chronic liver disease. *Liver International.* 2017; 37 (6): 778-793. 10.1111/liv.13317.

(22) Amitrano L, Guardascione MA, Brancaccio V, Balzano A. Coagulation disorders in liver disease. *Seminars in Liver Disease.* 2002; 22 (1): 83-96. 10.1055/s-2002-23205.

(23) Gatta A, Verardo A, Bolognesi M. Hypoalbuminemia. *Internal and Emergency Medicine.* 2012; 7 Suppl 3 193. 10.1007/s11739-012-0802-0.

(24) Day JW, Rosenberg WM. The enhanced liver fibrosis (ELF) test in diagnosis and management of liver fibrosis. *British Journal of Hospital Medicine (London, England: 2005).* 2018; 79 (12): 694-699. 10.12968/hmed.2018.79.12.694.

(25) Amorim TGF, Staub GJ, Lazzarotto C, Silva AP, Manes J, Ferronato MdG, et al. Validation and comparison of simple noninvasive models for the prediction of liver fibrosis in chronic hepatitis C. *Annals of hepatology.* 2012; 11 (6): 855-861. 10.1016/S1665-2681(19)31410-3.

(26) Teshale E, Lu M, Rupp LB, Holmberg SD, Moorman AC, Spradling P, et al. APRI and FIB-4 are good predictors of the stage of liver fibrosis in chronic hepatitis B: the Chronic Hepatitis Cohort Study (CHeCS). *Journal of viral hepatitis.* 2014; 21 (12): 917-920. 10.1111/jvh.12279.

(27) Shaheen AAM, Myers RP. Diagnostic accuracy of the aspartate aminotransferase-to-platelet ratio index for the prediction of hepatitis C–related fibrosis: A systematic review. *Hepatology (Baltimore, Md.).* 2007; 46 (3): 912-921. 10.1002/hep.21835.

(28) Jin W, Lin Z, Xin Y, Jiang X, Dong Q, Xuan S. Diagnostic accuracy of the aspartate aminotransferase-to-platelet ratio index for the prediction of hepatitis B-related fibrosis: a leading meta-analysis. *BMC Gastroenterology.* 2012; 12 (1): 14. 10.1186/1471-230X-12-14.

(29) Xiao G, Zhu S, Xiao X, Yan L, Yang J, Wu G. Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: A meta-analysis. *Hepatology (Baltimore, Md.).* 2017; 66 (5): 1486-1501. 10.1002/hep.29302.

(30) Martinez SM, Fernández-Varo G, González P, Sampson E, Bruguera M, Navasa M, et al. Assessment of liver fibrosis before and after antiviral therapy by different serum marker panels in patients with chronic hepatitis C. *Alimentary pharmacology & therapeutics.* 2011; 33 (1): 138-148. 10.1111/j.1365-2036.2010.04500.x.

(31) Kim BK, Kim DY, Park JY, Ahn SH, Chon CY, Kim JK, et al. Validation of FIB-4 and comparison with other simple noninvasive indices for predicting liver fibrosis and cirrhosis in hepatitis B virus-infected patients. *Liver international.* 2010; 30 (4): 546-553. 10.1111/j.1478-3231.2009.02192.x.

(32) Trembling PM, Lampertico P, Parkes J, Tanwar S, Viganò M, Facchetti F, et al. Performance of Enhanced Liver Fibrosis test and comparison with transient elastography in the identification of liver fibrosis in patients with chronic hepatitis B infection. *Journal of Viral Hepatitis.* 2014; 21 (6): 430-438. 10.1111/jvh.12161.

(33) Anstee QM, Lawitz EJ, Alkhouri N, Wong VW, Romero-Gomez M, Okanoue T, et al. Noninvasive Tests Accurately Identify Advanced Fibrosis due to NASH: Baseline Data From the STELLAR Trials. *Hepatology (Baltimore, Md.).* 2019; 70 (5): 1521-1530. 10.1002/hep.30842.

(34) Shaheen AAM, Wan AF, Myers RP. FibroTest and FibroScan for the Prediction of Hepatitis C-Related Fibrosis: A Systematic Review of Diagnostic Test Accuracy. *The American journal of gastroenterology.* 2007; 102 (11): 2589-2600. 10.1111/j.1572-0241.2007.01466.x.

(35) Salkic NN, Jovanovic P, Hauser G, Brcic M. FibroTest/Fibrosure for Significant Liver Fibrosis and Cirrhosis in Chronic Hepatitis B: A Meta-Analysis. *The American journal of gastroenterology.* 2014; 109 (6): 796-809. 10.1038/ajg.2014.21.

(36) Boursier J, Vergniol J, Guillet A, Hiriart J, Lannes A, Le Bail B, et al. Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. *Journal of hepatology.* 2016; 65 (3): 570-578. 10.1016/j.jhep.2016.04.023.

(37) Feng S, Wang Z, Zhao Y, Tao C. Wisteria floribunda agglutinin-positive Mac-2-binding protein as a diagnostic biomarker in liver cirrhosis: an updated meta-analysis. *Scientific reports.* 2020; 10 (1): 10582. 10.1038/s41598-020-67471-y.

(38) Child CG, Turcotte JG. Surgery and portal hypertension. *Major problems in clinical surgery.* 1964; 1 1-85.

(39) Goldberg E, Chopra S. *Cirrhosis in adults: Overview of complications, general management, and prognosis.* https://www.uptodate.com/contents/cirrhosis-in-adults-overview-of-complications-general-management-and-prognosis?search=Child%20Pugh%20classification&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1 .

(40) Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, ter Borg PCJ. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology (Baltimore, Md.).* 2000; 31 (4): 864-871. 10.1053/he.2000.5852.

(41) Wiesner R, Edwards E, Freeman R, Harper A, Kim R, Kamath P, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology (New York, N.Y. 1943).* 2003; 124 (1): 91-96. 10.1053/gast.2003.50016.

(42) Neuberger J, Gimson A, Davies M, Akyol M, O'Grady J, Burroughs A, et al. Selection of patients for liver transplantation and allocation of donated livers in the UK. *Gut.* 2008; 57 (2): 252-257. 10.1136/gut.2007.131730.

(43) Yilma M, Saxena V, Mehta N. Models to Predict Development or Recurence of Hepatocellular Carcinoma (HCC) in Patients with Advanced Hepatic Fibrosis. *Current Gastroenterology Reports.* 2022; 24 (1): 1-9. 10.1007/s11894-022-00835-8.

(44) Akinyemiju T, Abera S, Ahmed M, Alam N, Alemayohu MA, Allen C, et al. The Burden of Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level: Results From the Global Burden of Disease Study 2015. *JAMA oncology.* 2017; 3 (12): 1683-1691. 10.1001/jamaoncol.2017.3055.

(45) Schafer DF, Sorrell MF. Hepatocellular carcinoma. *Lancet (London, England).* 1999; 353 (9160): 1253-1257. S0140-6736(98)09148-X [pii].

(46) Desai A, Sandhu S, Lai J, Sandhu DS. Hepatocellular carcinoma in non-cirrhotic liver: A comprehensive review. *World Journal of Hepatology.* 2019; 11 (1): 1-18. 10.4254/wjh.v11.i1.1.

(47) Ryder SD, British Society of Gastroenterology. Guidelines for the diagnosis and treatment of hepatocellular carcinoma (HCC) in adults. *Gut.* 2003; 52 Suppl 3 (Suppl 3): iii1-8. 10.1136/gut.52.suppl_3.iii1 [doi].

(48) Chartampilas E, Rafailidis V, Georgopoulou V, Kalarakis G, Hatzidakis A, Prassopoulos P. Current Imaging Diagnosis of Hepatocellular Carcinoma. *Cancers.* 2022; 14 (16): 10.3390/cancers14163997.

(49) EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *Journal of Hepatology.* 2018; 69 (1): 182-236. 10.1016/j.jhep.2018.03.019.

(50) Liao X, Zhang D. The 8th Edition American Joint Committee on Cancer Staging for Hepato-pancreato-biliary Cancer: A Review and Update. *Archives of pathology & laboratory medicine (1976).* 2021; 145 (5): 543-553. 10.5858/arpa.2020-0032-RA.

(51) Subramaniam S, Kelley RK, Venook AP. A Review of Hepatocellular Carcinoma (HCC) Staging Systems. *Chinese clinical oncology.* 2013; 2 (4): 33. 10.3978/j.issn.2304-3865.2013.07.05.

(52) Fan Y, Xue H, Zheng H. Systemic Therapy for Hepatocellular Carcinoma: Current Updates and Outlook. *Journal of Hepatocellular Carcinoma.* 2022; 9 233-263. 10.2147/JHC.S358082.

(53) *Cirrhosis in over 16s: assessment and management | Guidance and guidelines | NICE.* NICE. https://www.nice.org.uk/guidance/ng50.

(54) Tzartzeva K, Obi J, Rich NE, Parikh ND, Marrero JA, Yopp A, et al. Surveillance Imaging and Alpha Fetoprotein for Early Detection of Hepatocellular Carcinoma in Patients With Cirrhosis: A Meta-analysis. *Gastroenterology.* 2018; 154 (6): 1706-1718.e1. 10.1053/j.gastro.2018.01.064.

(55) Cross TJS, Villaneuva A, Shetty S, Wilkes E, Collins P, Adair A, et al. A national survey of the provision of ultrasound surveillance for the detection of hepatocellular carcinoma. *Frontline gastroenterology.* 2016; 7 (2): 82-89. 10.1136/flgastro-2015-100617.

(56) Galle PR, Foerster F, Kudo M, Chan SL, Llovet JM, Qin S, et al. Biology and significance of alpha-fetoprotein in hepatocellular carcinoma. *Liver International.* 2019; 39 (12): 2214-2229. 10.1111/liv.14223.

(57) Zhang J, Chen G, Zhang P, Zhang J, Li X, Gan D, et al. The threshold of alpha-fetoprotein (AFP) for the diagnosis of hepatocellular carcinoma: A systematic review and meta-analysis. *PLOS ONE.* 2020; 15 (2): e0228857. 10.1371/journal.pone.0228857.

(58) Marrero JA, Feng Z, Wang Y, Nguyen MH, Befeler AS, Roberts LR, et al. α-Fetoprotein, Des-γ Carboxyprothrombin, and Lectin-Bound α-Fetoprotein in Early Hepatocellular Carcinoma. *Gastroenterology.* 2009; 137 (1): 110-118. 10.1053/j.gastro.2009.04.005.

(59) Li D, Mallory T, Satomura S. AFP-L3: a new generation of tumor marker for hepatocellular carcinoma. *Clinica Chimica Acta; International Journal of Clinical Chemistry.* 2001; 313 (1-2): 15-19. 10.1016/s0009-8981(01)00644-1.

(60) Xi L, Yang C. Evaluation of alpha-l-fucosidase for the diagnosis of hepatocellular carcinoma based on meta-analysis. *Journal of Laboratory Medicine.* 2020; 44 (4): 183-189. 10.1515/labmed-2019-0152.

(61) Xu D, Su C, Sun L, Gao Y, Li Y. Performance of Serum Glypican 3 in Diagnosis of Hepatocellular Carcinoma: A meta-analysis. *Annals of Hepatology.* 2019; 18 (1): 58-67. 10.5604/01.3001.0012.7863.

(62) Zhou F, Shang W, Yu X, Tian J. Glypican-3: A promising biomarker for hepatocellular carcinoma diagnosis and treatment. *Medicinal Research Reviews.* 2018; 38 (2): 741-767. 10.1002/med.21455.

(63) Amann A, Costello BdL, Miekisch W, Schubert J, Buszewski BI, Pleil J, et al. The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *Journal of Breath Research.* 2014; 8 (3): 034001. 10.1088/1752-7155/8/3/034001.

(64) Pauling L, Robinson AB, Teranishi R, Cary P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proceedings of the National Academy of Sciences of the United States of America.* 1971; 68 (10): 2374-2376.

(65) Mazzatenta A, Di Giulio C, Pokorski M. Pathologies currently identified by exhaled biomarkers. *Respiratory Physiology & Neurobiology.* 2013; 187 (1): 128-134. 10.1016/j.resp.2013.02.016.

(66) Fenske JD, Paulson SE. Human Breath Emissions of VOCs. *Journal of the Air & Waste Management Association.* 1999; 49 (5): 594-598. 10.1080/10473289.1999.10463831.

(67) Mazzatenta A, Pokorski M, Di Giulio C. Volatile organic compounds (VOCs) in exhaled breath as a marker of hypoxia in multiple chemical sensitivity. *Physiological Reports.* 2021; 9 (18): e15034. 10.14814/phy2.15034.

(68) Oguma T, Nagaoka T, Kurahashi M, Kobayashi N, Yamamori S, Tsuji C, et al. Clinical contributions of exhaled volatile organic compounds in the diagnosis of lung cancer. *PloS One.* 2017; 12 (4): e0174802. 10.1371/journal.pone.0174802.

(69) Jia Z, Zhang H, Ong CN, Patra A, Lu Y, Lim CT, et al. Detection of Lung Cancer: Concomitant Volatile Organic Compounds and Metabolomic Profiling of Six Cancer Cell Lines of Different Histological Origins. *ACS omega.* 2018; 3 (5): 5131-5140. 10.1021/acsomega.7b02035.

(70) Blanchet L, Smolinska A, Baranska A, Tigchelaar E, Swertz M, Zhernakova A, et al. Factors that influence the volatile organic compound content in human breath. *Journal of Breath Research.* 2017; 11 (1): 016013. 10.1088/1752-7163/aa5cc5.

(71) Krilaviciute A, Leja M, Kopp-Schneider A, Barash O, Khatib S, Amal H, et al. Associations of diet and lifestyle factors with common volatile organic compounds in exhaled breath of average-risk individuals. *Journal of Breath Research.* 2019; 13 (2): 026006. 10.1088/1752-7163/aaf3dc.

(72) Castellanos M, Xifra G, Fernández-Real JM, Sánchez JM. Breath gas concentrations mirror exposure to sevoflurane and isopropyl alcohol in hospital environments in non-occupational conditions. *Journal of Breath Research.* 2016; 10 (1): 016001. 10.1088/1752-7155/10/1/016001.

(73) Boshier PR, Cushnir JR, Priest OH, Marczin N, Hanna GB. Variation in the levels of volatile trace gases within three hospital environments: implications for clinical breath testing. *Journal of Breath Research.* 2010; 4 (3): 031001. 10.1088/1752-7155/4/3/031001.

(74) LoPachin RM, Gavin T. Molecular Mechanisms of Aldehyde Toxicity: A Chemical Perspective. *Chemical Research in Toxicology.* 2014; 27 (7): 1081-1091. 10.1021/tx5001046.

(75) Sharma R, Zang W, Zhou M, Schafer N, Begley LA, Huang YJ, et al. Real Time Breath Analysis Using Portable Gas Chromatography for Adult Asthma Phenotypes. *Metabolites.* 2021; 11 (5): 265. 10.3390/metabo11050265.

(76) Haick HD. *Volatile Biomarkers for Human Health : From Nature to Artificial Senses.* ; 2022.

(77) Farhi LE. Elimination of inert gas by the lung. *Respiration Physiology.* 1967; 3 (1): 1-11. 10.1016/0034-5687(67)90018-7.

(78) *Henry's Law.* https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Physical_Properties_of_Matter/Solutions_and_Mixtures/Ideal_Solutions/Dissolving_Gases_In_Liquids%2C_Henry's_Law [Accessed Jun 13, 2023].

(79) Kramer C, Mochalski P, Unterkofler K, Agapiou A, Ruzsanyi V, Liedl KR. Prediction of blood:air and fat:air partition coefficients of volatile organic compounds for the interpretation of data in breath gas analysis. *Journal of breath research.* 2016; 10 (1): 017103. 10.1088/1752-7155/10/1/017103.

(80) Mochalski P, King J, Mayhew CA, Unterkofler K. Modelling of Breath and Various Blood Volatilomic Profiles-Implications for Breath Volatile Analysis. *Molecules (Basel, Switzerland).* 2022; 27 (8): 2381. 10.3390/molecules27082381.

(81) Amann A, Smith D. *Volatile Biomarkers.* : Elsevier; 2013.

(82) Sutaria SR, Gori SS, Morris JD, Xie Z, Fu X, Nantz MH. Lipid Peroxidation Produces a Diverse Mixture of Saturated and Unsaturated Aldehydes in Exhaled Breath That Can Serve as Biomarkers of Lung Cancer-A Review. *Metabolites.* 2022; 12 (6): 561. 10.3390/metabo12060561.

(83) Leemans M, Bauër P, Cuzuel V, Audureau E, Fromantin I. Volatile Organic Compounds Analysis as a Potential Novel Screening Tool for Breast Cancer: A Systematic Review. *Biomarker Insights.* 2022; 17 11772719221100709. 10.1177/11772719221100709.

(84) Dima AC, Balaban DV, Dima A. Diagnostic Application of Volatile Organic Compounds as Potential Biomarkers for Detecting Digestive Neoplasia: A Systematic Review. *Diagnostics (Basel, Switzerland).* 2021; 11 (12): 2317. 10.3390/diagnostics11122317.

(85) Woodfield G, Belluomo I, Laponogov I, Veselkov K, Cross AJ, Hanna GB. Diagnostic Performance of a Noninvasive Breath Test for Colorectal Cancer: COBRA1 Study. *Gastroenterology.* 2022; 163 (5): 1447-1449.e8. 10.1053/j.gastro.2022.06.084.

(86) Adam ME, Fehervari M, Boshier PR, Chin S, Lin G, Romano A, et al. Mass-Spectrometry Analysis of Mixed-Breath, Isolated-Bronchial-Breath, and Gastric-Endoluminal-Air Volatile Fatty Acids in Esophagogastric Cancer. *Analytical Chemistry.* 2019; 91 (5): 3740-3746. 10.1021/acs.analchem.9b00148.

(87) Markar SR, Brodie B, Chin S-, Romano A, Spalding D, Hanna GB. Profile of exhaled-breath volatile organic compounds to diagnose pancreatic cancer. *The British Journal of Surgery.* 2018; 105 (11): 1493-1500. 10.1002/bjs.10909.

(88) Ratiu IA, Ligor T, Bocos-Bintintan V, Mayhew CA, Buszewski B. Volatile Organic Compounds in Exhaled Breath as Fingerprints of Lung Cancer, Asthma and COPD. *Journal of Clinical Medicine.* 2020; 10 (1): 32. 10.3390/jcm10010032.

(89) Gahleitner F, Guallar-Hoyas C, Beardsmore CS, Pandya HC, Thomas CP. Metabolomics pilot study to identify volatile organic compound markers of childhood asthma in exhaled breath. *Bioanalysis.* 2013; 5 (18): 2239-2247. 10.4155/bio.13.184.

(90) Miskoff JA, Dewan A, Chaudhri M. Fractional Exhaled Nitric Oxide Testing: Diagnostic Utility in Asthma, Chronic Obstructive Pulmonary Disease, or Asthma-chronic Obstructive Pulmonary Disease Overlap Syndrome. *Cureus.* 11 (6): e4864. 10.7759/cureus.4864.

(91) Gaida A, Holz O, Nell C, Schuchardt S, Lavae-Mokhtari B, Kruse L, et al. A dual center study to compare breath volatile organic compounds from smokers and non-smokers with and without COPD. *Journal of Breath Research.* 2016; 10 (2): 026006. 10.1088/1752-7155/10/2/026006.

(92) Sethi S, Nanda R, Chakraborty T. Clinical application of volatile organic compound analysis for detecting infectious diseases. *Clinical Microbiology Reviews.* 2013; 26 (3): 462-475. 10.1128/CMR.00020-13.

(93) Ratiu I, Ligor T, Bocos-Bintintan V, Buszewski B. Mass spectrometric techniques for the analysis of volatile organic compounds emitted from bacteria. *Bioanalysis.* 2017; 9 (14): 1069-1092. 10.4155/bio-2017-0051.

(94) Palma SICJ, Traguedo AP, Porteira AR, Frias MJ, Gamboa H, Roque ACA. Machine learning for the meta-analyses of microbial pathogens' volatile signatures. *Scientific Reports.* 2018; 8 (1): 1-15. 10.1038/s41598-018-21544-1.

(95) Kamal F, Kumar S, Edwards MR, Veselkov K, Belluomo I, Kebadze T, et al. Virus-induced Volatile Organic Compounds Are Detectable in Exhaled Breath during Pulmonary Infection. *American Journal of Respiratory and Critical Care Medicine.* 2021; 204 (9): 1075-1085. 10.1164/rccm.202103-0660OC.

(96) Biagini D, Fusi J, Vezzosi A, Oliveri P, Ghimenti S, Lenzi A, et al. Effects of long-term vegan diet on breath composition. *Journal of Breath Research.* 2022; 16 (2): 026004. 10.1088/1752-7163/ac4d41.

(97) Baranska A, Tigchelaar E, Smolinska A, Dallinga JW, Moonen EJC, Dekens JAM, et al. Profile of volatile organic compounds in exhaled breath changes as a result of gluten-free diet. *Journal of Breath Research.* 2013; 7 (3): 037104. 10.1088/1752-7155/7/3/037104.

(98) Cheng H. Volatile flavor compounds in yogurt: a review. *Critical Reviews in Food Science and Nutrition.* 2010; 50 (10): 938-950. 10.1080/10408390903044081.

(99) Belizário JE, Faintuch J, Malpartida MG. Breath Biopsy and Discovery of Exclusive Volatile Organic Compounds for Diagnosis of Infectious Diseases. *Frontiers in Cellular and Infection Microbiology.* 2020; 10 564194. 10.3389/fcimb.2020.564194.

(100) Ruzsányi V, Kalapos MP. Breath acetone as a potential marker in clinical practice*. *Journal of Breath Research.* 2017; 11 (2): 024002. 10.1088/1752-7163/aa66d3.

(101) Spaněl P, Dryahina K, Rejšková A, Chippendale TWE, Smith D. Breath acetone concentration; biological variability and the influence of diet. *Physiological Measurement.* 2011; 32 (8): 23. 10.1088/0967-3334/32/8/N01.

(102) Statheropoulos M, Agapiou A, Georgiadou A. Analysis of expired air of fasting male monks at Mount Athos. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences.* 2006; 832 (2): 274-279. 10.1016/j.jchromb.2006.01.017.

(103) Vadhwana B, Belluomo I, Boshier PR, Pavlou C, Španěl P, Hanna GB. Impact of oral cleansing strategies on exhaled volatile organic compound levels. *Rapid communications in mass spectrometry: RCM.* 2020; 34 (9): e8706. 10.1002/rcm.8706.

(104) King J, Kupferthaler A, Unterkofler K, Koc H, Teschl S, Teschl G, et al. Isoprene and acetone concentration profiles during exercise on an ergometer. *Journal of Breath Research.* 2009; 3 (2): 027006. 10.1088/1752-7155/3/2/027006.

(105) Mochalski P, King J, Mayhew CA, Unterkofler K. A review on isoprene in human breath. *Journal of Breath Research.* 2023; 17 (3): 10.1088/1752-7163/acc964.

(106) King J, Koc H, Unterkofler K, Mochalski P, Kupferthaler A, Teschl G, et al. Physiological modeling of isoprene dynamics in exhaled breath. *Journal of Theoretical Biology.* 2010; 267 (4): 626-637. 10.1016/j.jtbi.2010.09.028.

(107) Amann A, Mochalski P, Ruzsanyi V, Broza YY, Haick H. Assessment of the exhalation kinetics of volatile cancer biomarkers based on their physicochemical properties. *Journal of Breath Research.* 2014; 8 (1): 016003. 10.1088/1752-7155/8/1/016003.

(108) Patangia DV, Anthony Ryan C, Dempsey E, Paul Ross R, Stanton C. Impact of antibiotics on the human microbiome and consequences for host health. *MicrobiologyOpen.* 2022; 11 (1): e1260. 10.1002/mbo3.1260.

(109) Blanchet L, Smolinska A, Baranska A, Tigchelaar E, Swertz M, Zhernakova A, et al. Factors that influence the volatile organic compound content in human breath. *Journal of Breath Research.* 2017; 11 (1): 016013. 10.1088/1752-7163/aa5cc5.

(110) Sadler NC, Nandhikonda P, Webb-Robertson B, Ansong C, Anderson LN, Smith JN, et al. Hepatic Cytochrome P450 Activity, Abundance, and Expression Throughout Human Development. *Drug Metabolism and Disposition.* 2016; 44 (7): 984-991. 10.1124/dmd.115.068593.

(111) *Cytochrome P450 3A inhibitors and inducers - UpToDate.* https://www.uptodate.com/contents/image?imageKey=CARD%2F76992 [Accessed May 22, 2023].

(112) Capone S, Tufariello M, Forleo A, Longo V, Giampetruzzi L, Radogna AV, et al. Chromatographic analysis of VOC patterns in exhaled breath from smokers and nonsmokers. *Biomedical chromatography: BMC.* 2018; 32 (4): 10.1002/bmc.4132.

(113) Jordan A, Hansel A, Holzinger R, Lindinger W. Acetonitrile and benzene in the breath of smokers and non-smokers investigated by proton transfer reaction mass spectrometry (PTR-MS). *International Journal of Mass Spectrometry and Ion Processes.* 1995; 148 (1): L1-L3. 10.1016/0168-1176(95)04236-E.

(114) Pauwels CGGM, Hintzen KFH, Talhout R, Cremers HWJM, Pennings JLA, Smolinska A, et al. Smoking regular and low-nicotine cigarettes results in comparable levels of volatile organic compounds in blood and exhaled breath. *Journal of Breath Research.* 2020; 15 (1): 016010. 10.1088/1752-7163/abbf38.

(115) Pizzini A, Filipiak W, Wille J, Ager C, Wiesenhofer H, Kubinec R, et al. Analysis of volatile organic compounds in the breath of patients with stable or acute exacerbation of

chronic obstructive pulmonary disease. *Journal of Breath Research.* 2018; 12 (3): 036002. 10.1088/1752-7163/aaa4c5.

(116) Papaefstathiou E, Stylianou M, Andreou C, Agapiou A. Breath analysis of smokers, non-smokers, and e-cigarette users. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences.* 2020; 1160 122349. 10.1016/j.jchromb.2020.122349.

(117) Miekisch W, Kischkel S, Sawacki A, Liebau T, Mieth M, Schubert JK. Impact of sampling procedures on the results of breath analysis. *Journal of Breath Research.* 2008; 2 (2): 026007. 10.1088/1752-7155/2/2/026007.

(118) Doran SLF. Scaling up breath testing for the diagnosis and assessment of oesophago-gastric cancer in the clinical environment. 2021; 10.25560/98288.

(119) Lawal O, Ahmed WM, Nijsen TME, Goodacre R, Fowler SJ. Exhaled breath analysis: a review of 'breath-taking' methods for off-line analysis. *Metabolomics: Official Journal of the Metabolomic Society.* 2017; 13 (10): 110. 10.1007/s11306-017-1241-8.

(120) Bikov A, Hernadi M, Korosi BZ, Kunos L, Zsamboki G, Sutto Z, et al. Expiratory flow rate, breath hold and anatomic dead space influence electronic nose ability to detect lung cancer. *BMC pulmonary medicine.* 2014; 14 202. 10.1186/1471-2466-14-202.

(121) Thekedar B, Oeh U, Szymczak W, Hoeschen C, Paretzke HG. Influences of mixed expiratory sampling parameters on exhaled volatile organic compound concentrations. *Journal of Breath Research.* 2011; 5 (1): 016001. 10.1088/1752-7155/5/1/016001.

(122) Das MK, Bishwal SC, Das A, Dabral D, Varshney A, Badireddy VK, et al. Investigation of gender-specific exhaled breath volatome in humans by GCxGC-TOF-MS. *Analytical Chemistry.* 2014; 86 (2): 1229-1237. 10.1021/ac403541a.

(123) Lechner M, Moser B, Niederseer D, Karlseder A, Holzknecht B, Fuchs M, et al. Gender and age specific differences in exhaled isoprene levels. *Respiratory Physiology & Neurobiology.* 2006; 154 (3): 478-483. 10.1016/j.resp.2006.01.007.

(124) Spaněl P, Dryahina K, Smith D. Acetone, ammonia and hydrogen cyanide in exhaled breath of several volunteers aged 4-83 years. *Journal of Breath Research.* 2007; 1 (1): 011001. 10.1088/1752-7155/1/1/011001.

(125) Smith D, Španěl P, Enderby B, Lenney W, Turner C, Davies SJ. Isoprene levels in the exhaled breath of 200 healthy pupils within the age range 7–18 years studied using SIFT-MS. *Journal of Breath Research.* 2009; 4 (1): 017101. 10.1088/1752-7155/4/1/017101.

(126) Belluomo I, Boshier PR, Myridakis A, Vadhwana B, Markar SR, Spanel P, et al. Selected ion flow tube mass spectrometry for targeted analysis of volatile organic compounds in human breath. *Nature protocols.* 2021; 16 (7): 3419-3438. 10.1038/s41596-021-00542-0.

(127) Hanna GB, Boshier PR, Markar SR, Romano A. Accuracy and Methodologic Challenges of Volatile Organic Compound–Based Exhaled Breath Tests for Cancer Diagnosis. *JAMA Oncology.* 2019; 5 (1): 10.1001/jamaoncol.2018.2815.

(128) Trefz P, Schmidt M, Oertel P, Obermeier J, Brock B, Kamysek S, et al. Continuous real time breath gas monitoring in the clinical environment by proton-transfer-reaction-time-

of-flight-mass spectrometry. *Analytical Chemistry.* 2013; 85 (21): 10321-10329. 10.1021/ac402298v.

(129) Markar SR, Wiggins T, Antonowicz S, Chin S, Romano A, Nikolic K, et al. Assessment of a Noninvasive Exhaled Breath Test for the Diagnosis of Oesophagogastric Cancer. *JAMA Oncology.* 2018; 4 (7): 970-976. 10.1001/jamaoncol.2018.0991.

(130) Salman D, Ibrahim W, Kanabar A, Joyce A, Zhao B, Singapuri A, et al. The variability of volatile organic compounds in the indoor air of clinical environments. *Journal of Breath Research.* 2021; 16 (1): 016005. 10.1088/1752-7163/ac3565.

(131) Risby TH, Solga SF. Current status of clinical breath analysis. *Applied Physics B.* 2006; 85 (2): 421-426. 10.1007/s00340-006-2280-4.

(132) Phillips M, Cataneo RN, Ditkoff BA, Fisher P, Greenberg J, Gunawardena R, et al. Volatile Markers of Breast Cancer in the Breath. *The Breast Journal.* 2003; 9 (3): 184-191. 10.1046/j.1524-4741.2003.09309.x.

(133) Phillips M, Greenberg J, Sabas M. Alveolar Gradient of Pentane in Normal Human Breath. *Free Radical Research.* 1994; 20 (5): 333-337. 10.3109/10715769409145633.

(134) Unterkofler K, King J, Mochalski P, Jandacka M, Koc H, Teschl S, et al. Modeling-based determination of physiological parameters of systemic VOCs by breath gas analysis: a pilot study. *Journal of Breath Research.* 2015; 9 (3): 036002. 10.1088/1752-7155/9/3/036002.

(135) Haick H, Broza YY, Mochalski P, Ruzsanyi V, Amann A. Assessment, origin, and implementation of breath volatile cancer markers. *Chemical Society Reviews.* 2014; 43 (5): 1423-1449. 10.1039/c3cs60329f.

(136) Spaněl P, Dryahina K, Smith D. A quantitative study of the influence of inhaled compounds on their concentrations in exhaled breath. *Journal of Breath Research.* 2013; 7 (1): 017106. 10.1088/1752-7155/7/1/017106.

(137) Harshman SW, Pitsch RL, Davidson CN, Scott AM, Hill EM, Smith ZK, et al. Characterization of standardized breath sampling for off-line field use. *Journal of Breath Research.* 2019; 14 (1): 016009. 10.1088/1752-7163/ab55c5.

(138) Maurer F, Wolf A, Fink T, Rittershofer B, Heim N, Volk T, et al. Wash-out of ambient air contaminations for breath measurements. *Journal of Breath Research.* 2014; 8 (2): 027107. 10.1088/1752-7155/8/2/027107.

(139) Lin C, Lin C, Hsieh L, Chen C, Wang J. Vertical and diurnal characterization of volatile organic compounds in ambient air in urban areas. *Journal of the Air & Waste Management Association (1995).* 2011; 61 (7): 714-720. 10.3155/1047-3289.61.7.714.

(140) Savarino V, Vigneri S, Celle G. The 13C urea breath test in the diagnosis of Helicobacter pylori infection. *Gut.* 1999; 45 (Suppl 1): I18-I22. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1766662/.

(141) Petrone P, Sarkisyan G, Fernández M, Coloma E, Akopian G, Ortega A, et al. Small intestinal bacterial overgrowth in patients with lower gastrointestinal symptoms and a history

of previous abdominal surgery. *Archives of Surgery (Chicago, Ill.: 1960).* 2011; 146 (4): 444-447. 10.1001/archsurg.2011.55.

(142) Rezaie A, Buresi M, Lembo A, Lin H, McCallum R, Rao S, et al. Hydrogen and Methane-Based Breath Testing in Gastrointestinal Disorders: The North American Consensus. *The American Journal of Gastroenterology.* 2017; 112 (5): 775-784. 10.1038/ajg.2017.46.

(143) Wen Q, Boshier P, Myridakis A, Belluomo I, Hanna GB. Urinary Volatile Organic Compound Analysis for the Diagnosis of Cancer: A Systematic Literature Review and Quality Assessment. *Metabolites.* 2020; 11 (1): 17. 10.3390/metabo11010017.

(144) *Mass Spectrometry Ionization Methods.* http://chemistry.emory.edu/msc/tutorial/mass-spectrometry-ionization.html [Accessed Apr 14, 2023].

(145) Ellis AM, Mayhew CA. *Proton transfer reaction mass spectrometry.* ; 2014.

(146) Pabary R, Kumar S, Huang J, Alton E, Bush A, Hanna GB, et al. Spoken sessions. *Thorax.* 2002; 57 (suppl 3): iii3-47. 10.1136/thorax.57.suppl_3.iii3.

(147) Spanel P, Smith D, Holland TA, Al Singary W, Elder JB. Analysis of formaldehyde in the headspace of urine from bladder and prostate cancer patients using selected ion flow tube mass spectrometry. *Rapid communications in mass spectrometry: RCM.* 1999; 13 (14): 1354-1359. 10.1002/(SICI)1097-0231(19990730)13:143.0.CO;2-J.

(148) Smith D, Španěl P. SIFT-MS and FA-MS methods for ambient gas phase analysis: developments and applications in the UK. *The Analyst.* 2015; 140 (8): 2573-2591. 10.1039/c4an02049a.

(149) *GC-MS Principle, Instrument and Analyses and GC-MS/MS.* http://www.technologynetworks.com/analysis/articles/gc-ms-principle-instrument-and-analyses-and-gc-msms-362513 [Accessed Apr 25, 2023].

(150) Romano A, Capozzi V, Spano G, Biasioli F. Proton transfer reaction-mass spectrometry: online and rapid determination of volatile organic compounds of microbial origin. *Applied Microbiology and Biotechnology.* 2015; 99 (9): 3787-3795. 10.1007/s00253-015-6528-y.

(151) Blake RS, Whyte C, Hughes CO, Ellis AM, Monks PS. Demonstration of Proton-Transfer Reaction Time-of-Flight Mass Spectrometry for Real-Time Analysis of Trace Volatile Organic Compounds. *Analytical Chemistry.* 2004; 76 (13): 3841-3845. 10.1021/ac0498260.

(152) Chin S, Romano A, Doran SLF, Hanna GB. Cross-platform mass spectrometry annotation in breathomics of oesophageal-gastric cancer. *Scientific Reports.* 2018; 8 (1): 5139. 10.1038/s41598-018-22890-w.

(153) Materić D, Bruhn D, Turner C, Morgan G, Mason N, Gauci V. Methods in plant foliar volatile organic compounds research1. *Applications in Plant Sciences.* 2015; 3 (12): apps.1500044. 10.3732/apps.1500044.

(154) Phillips M, Cataneo RN, Chaturvedi A, Kaplan PD, Libardoni M, Mundada M, et al. Detection of an extended human volatome with comprehensive two-dimensional gas

chromatography time-of-flight mass spectrometry. *PloS One.* 2013; 8 (9): e75274. 10.1371/journal.pone.0075274.

(155) *Gravity Notes: Instrument Drift.* https://pburnley.faculty.unlv.edu/GEOL452_652/gravity/notes/GravityNotes13InstrumentDrift.htm [Accessed Jun 12, 2023].

(156) Märtens A, Holle J, Mollenhauer B, Wegner A, Kirwan J, Hiller K. Instrumental Drift in Untargeted Metabolomics: Optimizing Data Quality with Intrastudy QC Samples. *Metabolites.* 2023; 13 (5): 665. 10.3390/metabo13050665.

(157) Wehrens R, Hageman JA, van Eeuwijk F, Kooke R, Flood PJ, Wijnker E, et al. Improved batch correction in untargeted MS-based metabolomics. *Metabolomics.* 2016; 12 88. 10.1007/s11306-016-1015-8.

(158) Hendriks MMWB, Eeuwijk FAv, Jellema RH, Westerhuis JA, Reijmers TH, Hoefsloot HCJ, et al. Data-processing strategies for metabolomics studies. *TrAC, Trends in analytical chemistry (Regular ed.).* 2011; 30 (10): 1685-1698. 10.1016/j.trac.2011.04.019.

(159) *Mass Spectrometry Data Center, NIST.* https://chemdata.nist.gov/ [Accessed May 16, 2023].

(160) *Syft Technologies SIFT-MS - Voice200ultra Real Time.* https://anatune.co.uk/product/syft-technologies-sift-ms-voice200ultra-real-time/ [Accessed Nov 12, 2022].

(161) Bruderer T, Gaisl T, Gaugg MT, Nowak N, Streckenbach B, Müller S, et al. On-Line Analysis of Exhaled Breath Focus Review. *Chemical Reviews.* 2019; 119 (19): 10803-10828. 10.1021/acs.chemrev.9b00005.

(162) Di Gilio A, Palmisani J, Ventrella G, Facchini L, Catino A, Varesano N, et al. Breath Analysis: Comparison among Methodological Approaches for Breath Sampling. *Molecules.* 2020; 25 (24): 5823. 10.3390/molecules25245823.

(163) *Tenax, Carboxen and other sorbents used for thermal desorption sampling | Markes International.* https://markes.com/content-hub/blog/tenax-carboxen-and-other-sorbents-used-for-thermal-desorption-sampling [Accessed Nov 12, 2022].

(164) Romano A, Doran S, Belluomo I, Hanna GB. High-Throughput Breath Volatile Organic Compound Analysis Using Thermal Desorption Proton Transfer Reaction Time-of-Flight Mass Spectrometry. *Analytical Chemistry.* 2018; 90 (17): 10204-10210. 10.1021/acs.analchem.8b01045.

(165) Perkins M. *THERMAL DESORPTION (TD)-SIFT-MS: A NEW APPROACH TO BREATH ANALYSIS.* 2019. https://gcms.labrulez.com/labrulez-bucket-strapi-h3hsga3/application::paper.paper/SnApp31-1.pdf.

(166) Brown VM, Crump DR, Plant NT, Pengelly I. Evaluation of the stability of a mixture of volatile organic compounds on sorbents for the determination of emissions from indoor materials and products using thermal desorption/gas chromatography/mass spectrometry. *Journal of Chromatography. A.* 2014; 1350 1-9. 10.1016/j.chroma.2014.05.011.

(167) Kang S, Paul Thomas CL. How long may a breath sample be stored for at -80 °C? A study of the stability of volatile organic compounds trapped onto a mixed Tenax:Carbograph trap adsorbent bed from exhaled breath. *Journal of Breath Research.* 2016; 10 (2): 026011. 10.1088/1752-7155/10/2/026011.

(168) *Improving thermal desorption (TD) applications with sample re-collection | Markes International.* https://markes.com/content-hub/instant-insight/improving-thermal-desorption-td-applications-with-sample-re-collection [Accessed Nov 1, 2022].

(169) Wen Q, Myridakis A, Boshier PR, Zuffa S, Belluomo I, Parker AG, et al. A Complete Pipeline for Untargeted Urinary Volatolomic Profiling with Sorptive Extraction and Dual Polar and Nonpolar Column Methodologies Coupled with Gas Chromatography Time-of-Flight Mass Spectrometry. *Analytical Chemistry.* 2023; 10.1021/acs.analchem.2c02873.

(170) Beauchamp J, Herbig J, Gutmann R, Hansel A. On the use of Tedlar® bags for breath-gas sampling and analysis. *Journal of Breath Research.* 2008; 2 (4): 046001. 10.1088/1752-7155/2/4/046001.

(171) Mochalski P, King J, Unterkofler K, Amann A. Stability of selected volatile breath constituents in Tedlar, Kynar and Flexfilm sampling bags. *The Analyst.* 2013; 138 (5): 1405-1418. 10.1039/c2an36193k.

(172) Ghosh S, Kim K, Sohn JR. Some Insights into Analytical Bias Involved in the Application of Grab Sampling for Volatile Organic Compounds: A Case Study against Used Tedlar Bags. *The Scientific World Journal.* 2011; 11 2160-2177. 10.1100/2011/529532.

(173) Mochalski P, Wzorek B, Śliwka I, Amann A. Suitability of different polymer bags for storage of volatile sulphur compounds relevant to breath analysis. *Journal of Chromatography B.* 2009; 877 (3): 189-196. 10.1016/j.jchromb.2008.12.003.

(174) *ALTEF (PVDF) breath collection bags.* MediSense | Smelltest.eu. https://www.smelltest.eu/en/product/altef-pvdf-breath-collection-bags/ [Accessed Nov 20, 2022].

(175) Lourenço C, Turner C. Breath analysis in disease diagnosis: methodological considerations and applications. *Metabolites.* 2014; 4 (2): 465-498. 10.3390/metabo4020465.

(176) Beauchamp J, Herbig J, Gutmann R, Hansel A. On the use of Tedlar® bags for breath-gas sampling and analysis. *Journal of Breath Research.* 2008; 2 (4): 046001. 10.1088/1752-7155/2/4/046001.

(177) Fernandez Del Rio R, O'Hara ME, Holt A, Pemberton P, Shah T, Whitehouse T, et al. Volatile Biomarkers in Breath Associated With Liver Cirrhosis - Comparisons of Pre- and Post-liver Transplant Breath Samples. *EBioMedicine.* 2015; 2 (9): 1243-50. https://dx.doi.org/10.1016/j.ebiom.2015.07.027.

(178) Sukul P, Trefz P, Schubert JK, Miekisch W. Advanced setup for safe breath sampling and patient monitoring under highly infectious conditions in the clinical environment. *Scientific Reports.* 2022; 12 (1): 1-14. 10.1038/s41598-022-22581-7.

(179) Kwak J, Fan M, Harshman SW, Garrison CE, Dershem VL, Phillips JB, et al. Evaluation of Bio-VOC Sampler for Analysis of Volatile Organic Compounds in Exhaled Breath. *Metabolites.* 2014; 4 (4): 879-888. 10.3390/metabo4040879.

(180) *ReCIVA® Breath Sampler.* https://www.owlstonemedical.com/products/reciva/ [Accessed Nov 12, 2022].

(181) Doran SLF, Romano A, Hanna GB. Optimisation of sampling parameters for standardised exhaled breath sampling. *Journal of breath research.* 2017; 12 (1): 16007-016007. 10.1088/1752-7163/aa8a46.

(182) Harshman SW, Pitsch RL, Davidson CN, Lee EM, Scott AM, Hill EM, et al. Evaluation of a standardized collection device for exhaled breath sampling onto thermal desorption tubes. *Journal of Breath Research.* 2020; 14 (3): 036004. 10.1088/1752-7163/ab7e3b.

(183) *Mistral.* Mistral. https://www.mistral-breath.it/en/mistral/ [Accessed Apr 25, 2023].

(184) *Multi-Pixel Gas Sensor SGP - Sensirion | Mouser.* https://eu.mouser.com/new/sensirion/sensirion-sgp-gas-sensors/?_gl=1*13aum5g*_ga*MTc4NjYyNzEzNy4xNjg2MDU5NDUy*_ga_15W4STQT4T*MTY4NjA1OTQ1Mi4xLjAuMTY4NjA1OTQ1Mi4wLjAuMA.. [Accessed Jun 6, 2023].

(185) Stavropoulos G, van Munster K, Ferrandino G, Sauca M, Ponsioen C, van Schooten F, et al. Liver Impairment—The Potential Application of Volatile Organic Compounds in Hepatology. *Metabolites.* 2021; 11 (9): 10.3390/metabo11090618.

(186) Chen S, Mahadevan V, Zieve L. Volatile fatty acids in the breath of patients with cirrhosis of the liver. *The Journal of Laboratory and Clinical Medicine.* 1970; 75 (4): 622-627. 0022214370901605.

(187) Kaji H, Hisamura M, Saito N, Murao M. Gas chromatographic determination of volatile sulfur compounds in the expired alveolar air in hepatopathic subjects. *Journal of chromatography.* 1978; 145 (3): 464-8. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med1&NEWS=N&AN=659533.

(188) Challenger F, Walshe JM. Methyl mercaptan in relation to foetor hepaticus. *The Biochemical Journal.* 1955; 59 (3): 372-375.

(189) Tangerman A, Meuwese-Arends MT, van Tongeren JH. A new sensitive assay for measuring volatile sulphur compounds in human breath by Tenax trapping and gas chromatography and its application in liver cirrhosis. *Clinica Chimica Acta; International Journal of Clinical Chemistry.* 1983; 130 (1): 103-110. 10.1016/0009-8981(83)90263-2.

(190) Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.).* 2021; 372 n71. 10.1136/bmj.n71.

(191) Qin T, Liu H, Song Q, Song G, Wang H, Pan Y, et al. The screening of volatile markers for hepatocellular carcinoma. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology.* 2010; 19 (9): 2247-2253. 10.1158/1055-9965.EPI-10-0302.

(192) Sukaram T, Apiparakoon T, Tiyarattanachai T, Ariyaskul D, Kulkraisri K, Marukatat S, et al. VOCs from Exhaled Breath for the Diagnosis of Hepatocellular Carcinoma. *Diagnostics (Basel, Switzerland).* 2023; 13 (2): 257. 10.3390/diagnostics13020257.

(193) Alkhouri N, Singh T, Alsabbagh E, Guirguis J, Chami T, Hanouneh I, et al. Isoprene in the Exhaled Breath is a Novel Biomarker for Advanced Fibrosis in Patients with Chronic Liver Disease: A Pilot Study. *Clinical and Translational Gastroenterology.* 2015; 6 e112. 10.1038/ctg.2015.40.

(194) Friedman MI, Preti G, Deems RO, Friedman LS, Munoz SJ, Maddrey WC. Limonene in expired lung air of patients with liver disease. *Digestive Diseases and Sciences.* 1994; 39 (8): 1672-1676. 10.1007/bf02087774.

(195) Hanouneh IA, Zein NN, Cikach F, Dababneh L, Grove D, Alkhouri N, et al. The breathprints in patients with liver disease identify novel breath biomarkers in alcoholic hepatitis. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association.* 2014; 12 (3): 516-23. https://dx.doi.org/10.1016/j.cgh.2013.08.048.

(196) Miller-Atkins G, Acevedo-Moreno L, Grove D, Dweik RA, Tonelli AR, Brown JM, et al. Breath Metabolomics Provides an Accurate and Noninvasive Approach for Screening Cirrhosis, Primary, and Secondary Liver Tumors. *Hepatology communications.* 2020; 4 (7): 1041-1055. https://dx.doi.org/10.1002/hep4.1499.

(197) Pijls KE, Smolinska A, Jonkers DMAE, Dallinga JW, Masclee AAM, Koek GH, et al. A profile of volatile organic compounds in exhaled air as a potential non-invasive biomarker for liver cirrhosis. *Scientific Reports.* 2016; 6 19903. 10.1038/srep19903.

(198) Verdam FJ, Dallinga JW, Driessen A, de Jonge C, Moonen EJC, van Berkel J,B.N., et al. Non-alcoholic steatohepatitis: a non-invasive diagnosis by analysis of exhaled breath. *Journal of hepatology.* 2013; 58 (3): 543-8. https://dx.doi.org/10.1016/j.jhep.2012.10.030.

(199) Sinha AE, Johnson KJ, Prazen BJ, Lucas SV, Fraga CG, Synovec RE. Comprehensive two-dimensional gas chromatography of volatile and semi-volatile components using a diaphragm valve-based instrument. *Journal of chromatography.A.* 2003; 983 (1-2): 195-204. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=12568382.

(200) O'Hara ME, Fernandez Del Rio R, Holt A, Pemberton P, Shah T, Whitehouse T, et al. Limonene in exhaled breath is elevated in hepatic encephalopathy. *Journal of breath research.* 2016; 10 (4): 046010. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med13&NEWS=N&AN=27869108.

(201) Sehnert SS, Jiang L, Burdick JF, Risby TH. Breath biomarkers for detection of human liver diseases: preliminary study. *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals.* 2002; 7 (2): 174-87. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=12101636.

(202) Van den Velde S, Nevens F, Van Hee P, van Steenberghe D, Quirynen M. GC-MS analysis of breath odor compounds in liver patients. *Journal of Chromatography. B,*

Analytical Technologies in the Biomedical and Life Sciences. 2008; 875 (2): 344-348. 10.1016/j.jchromb.2008.08.031.

(203) Dadamio J, Van den Velde S, Laleman W, Van Hee P, Coucke W, Nevens F, et al. Breath biomarkers of liver cirrhosis. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences.* 2012; 905 17-22. 10.1016/j.jchromb.2012.07.025.

(204) Ferrandino G, De Palo G, Murgia A, Birch O, Tawfike A, Smith R, et al. Breath Biopsy® to Identify Exhaled Volatile Organic Compounds Biomarkers for Liver Cirrhosis Detection. *Journal of Clinical and Translational Hepatology.* 2023; 11 (3): 638-648. 10.14218/JCTH.2022.00309.

(205) Millonig G, Praun S, Netzer M, Baumgartner C, Dornauer A, Mueller S, et al. Non-invasive diagnosis of liver diseases by breath analysis using an optimized ion-molecule reaction-mass spectrometry approach: a pilot study. *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals.* 2010; 15 (4): 297-306. https://dx.doi.org/10.3109/13547501003624512.

(206) Morisco F, Aprea E, Lembo V, Fogliano V, Vitaglione P, Mazzone G, et al. Rapid "breath-print" of liver cirrhosis by proton transfer reaction time-of-flight mass spectrometry. A pilot study. *PloS one.* 2013; 8 (4): e59658. https://dx.doi.org/10.1371/journal.pone.0059658.

(207) Khalid TY, Costello BDL, Ewen R, White P, Stevens S, Gordon F, et al. Breath volatile analysis from patients diagnosed with harmful drinking, cirrhosis and hepatic encephalopathy: a pilot study. *Metabolomics.* 2013; 9 (5): 938-948. 10.1007/s11306-013-0510-4.

(208) *Lung Volumes and Ventilation--Objectives.* https://www.meddean.luc.edu/lumen/meded/medicine/pulmonar/physio/pf1.htm [Accessed May 21, 2023].

(209) Ferrandino G, Orf I, Smith R, Calcagno M, Thind AK, Debiram-Beecham I, et al. Breath Biopsy Assessment of Liver Disease Using an Exogenous Volatile Organic Compound-Toward Improved Detection of Liver Impairment. *Clinical and translational gastroenterology.* 2020; 11 (9): e00239. https://dx.doi.org/10.14309/ctg.0000000000000239.

(210) Taylor JMG, Ankerst DP, Andridge RR. Validation of Biomarker-based risk prediction models. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2008; 14 (19): 5977-5983. 10.1158/1078-0432.CCR-07-4534.

(211) *Acid Gases Cartridge for Respirators Honeywell North N7500-2A-18.* https://www.hantover.com/Acid-Gases-Cartridge-for-Respirators-Honeywell-North-N7500-2A-18/551120/p [Accessed May 18, 2023].

(212) Klupinski TP, Strozier ED, Friedenberg DA, Brinkman MC, Gordon SM, Clark PI. Identification of New and Distinctive Exposures from Little Cigars. *Chemical research in toxicology.* 2016; 29 (2): 162-168. 10.1021/acs.chemrestox.5b00371.

(213) Raninen KJ, Lappi JE, Mukkala ML, Tuomainen T, Mykkänen HM, Poutanen KS, et al. Fiber content of diet affects exhaled breath volatiles in fasting and postprandial state in a pilot crossover study. *Nutrition Research (New York, N.Y.).* 2016; 36 (6): 612-619. 10.1016/j.nutres.2016.02.008.

(214) Sinha R, Lockman KA, Homer NZM, Bower E, Brinkman P, Knobel HH, et al. Volatomic analysis identifies compounds that can stratify non-alcoholic fatty liver disease. *JHEP reports: innovation in hepatology.* 2020; 2 (5): 100137. 10.1016/j.jhepr.2020.100137.

(215) Raman M, Ahmed I, Gillevet PM, Probert CS, Ratcliffe NM, Smith S, et al. Fecal microbiome and volatile organic compound metabolome in obese humans with nonalcoholic fatty liver disease. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association.* 2013; 11 (7): 868-3. https://dx.doi.org/10.1016/j.cgh.2013.02.015.

(216) Xue R, Dong L, Zhang S, Deng C, Liu T, Wang J, et al. Investigation of volatile biomarkers in liver cancer blood using solid-phase microextraction and gas chromatography/mass spectrometry. *Rapid communications in mass spectrometry : RCM.* 2008; 22 (8): 1181-6. https://dx.doi.org/10.1002/rcm.3466.

(217) *Styrene (ethenylbenzene) - DCCEEW.* https://www.dcceew.gov.au/environment/protection/npi/substances/fact-sheets/styrene-ethenylbenzene [Accessed Jun 12, 2023].

(218) Bannaga AS, Tyagi H, Daulton E, Covington JA, Arasaradnam RP. Exploratory Study Using Urinary Volatile Organic Compounds for the Detection of Hepatocellular Carcinoma. *Molecules (Basel, Switzerland).* 2021; 26 (9): 2447. 10.3390/molecules26092447.

(219) Drabińska N, Flynn C, Ratcliffe N, Belluomo I, Myridakis A, Gould O, et al. A literature survey of all volatiles from healthy human breath and bodily fluids: the human volatilome. *Journal of Breath Research.* 2021; 15 (3): 034001. 10.1088/1752-7163/abf1d0.

(220) Boshier PR, Fehervari M, Markar SR, Purkayastha S, Spanel P, Smith D, et al. Variation in Exhaled Acetone and Other Ketones in Patients Undergoing Bariatric Surgery: a Prospective Cross-sectional Study. *Obesity Surgery.* 2018; 28 (8): 2439-2446. 10.1007/s11695-018-3180-5.

(221) Goldberg EM, Blendis LM, Sandler S. A gas chromatographic--mass spectrometric study of profiles of volatile metabolites in hepatic encephalopathy. *Journal of chromatography.* 1981; 226 (2): 291-9. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med2&NEWS=N&AN=7320160.

(222) Wu G. Amino Acid Metabolism in the Liver: Nutritional and Physiological Significance. *Advances in experimental medicine and biology.* 2020; 1265 21-37. 10.1007/978-3-030-45328-2_2.

(223) Roslund K, Lehto M, Pussinen P, Metsälä M. Volatile composition of the morning breath. *Journal of Breath Research.* 2022; 16 (4): 10.1088/1752-7163/ac8ec8.

(224) *Limonene.* https://pubchem.ncbi.nlm.nih.gov/compound/22311 [Accessed May 23, 2023].

(225) Miyazawa M, Shindo M, Shimada T. Metabolism of (+)- and (-)-limonenes to respective carveols and perillyl alcohols by CYP2C9 and CYP2C19 in human liver microsomes. *Drug Metabolism and Disposition: The Biological Fate of Chemicals.* 2002; 30 (5): 602-607. 10.1124/dmd.30.5.602.

(226) *The Good Scents Company - delta-3-carene.* http://www.thegoodscentscompany.com/data/rw1014471.html [Accessed Sept 25, 2021].

(227) Duisken M, Benz D, Peiffer TH, Blömeke B, Hollender J. Metabolism of Delta(3)-carene by human cytochrome p450 enzymes: identification and characterization of two new metabolites. *Current Drug Metabolism.* 2005; 6 (6): 593-601. 10.2174/138920005774832614.

(228) Sukul P, Richter A, Junghanss C, Schubert J, Miekisch W. *Origin of breath isoprene in humans is revealed via multi-omic investigation.* 2023.

(229) Miyazawa M, Shindo M, Shimada T. Oxidation of 1,8-cineole, the monoterpene cyclic ether originated from eucalyptus polybractea, by cytochrome P450 3A enzymes in rat and human liver microsomes. *Drug Metabolism and Disposition: The Biological Fate of Chemicals.* 2001; 29 (2): 200-205.

(230) Takase S, Takada A, Yasuhara M, Tsutsumi M. Hepatic aldehyde dehydrogenase activity in liver diseases, with particular emphasis on alcoholic liver disease. *Hepatology (Baltimore, Md.).* 1989; 9 (5): 704-709. 10.1002/hep.1840090508.

(231) *NIAAA Publications - Alcohol Metabolism: An Update.* https://pubs.niaaa.nih.gov/publications/aa72/aa72.htm [Accessed May 24, 2023].

(232) Seeman JI, Dixon M, Haussmann H. Acetaldehyde in Mainstream Tobacco Smoke: Formation and Occurrence in Smoke and Bioavailability in the Smoker. *Chemical Research in Toxicology.* 2002; 15 (11): 1331-1350. 10.1021/tx020069f.

(233) *Hexanal.* https://pubchem.ncbi.nlm.nih.gov/compound/6184 [Accessed May 24, 2023].

(234) Ayala A, Muñoz MF, Argüelles S. Lipid Peroxidation: Production, Metabolism, and Signaling Mechanisms of Malondialdehyde and 4-Hydroxy-2-Nonenal. *Oxidative Medicine and Cellular Longevity.* 2014; 2014 360438. 10.1155/2014/360438.

(235) Floss MA, Fink T, Maurer F, Volk T, Kreuer S, Müller-Wirtz LM. Exhaled Aldehydes as Biomarkers for Lung Diseases: A Narrative Review. *Molecules.* 2022; 27 (16): 5258. 10.3390/molecules27165258.

(236) Kazui M, Andreoni KA, Norris EJ, Klein AS, Burdick JF, Beattie C, et al. Breath ethane: a specific indicator of free-radical-mediated lipid peroxidation following reperfusion of the ischemic liver. *Free Radical Biology & Medicine.* 1992; 13 (5): 509-515. 10.1016/0891-5849(92)90145-7.

(237) Risby TH, Sehnert SS. Clinical application of breath biomarkers of oxidative stress status. *Free Radical Biology & Medicine.* 1999; 27 (11-12): 1182-1192. 10.1016/s0891-5849(99)00212-9.

(238) Schubert JK, Müller WP, Benzing A, Geiger K. Application of a new method for analysis of exhaled gas in critically ill patients. *Intensive Care Medicine.* 1998; 24 (5): 415-421. 10.1007/s001340050589.

(239) Pelli MA, Trovarelli G, Capodicasa E, De Medio GE, Bassotti G. Breath alkanes determination in ulcerative colitis and Crohn's disease. *Diseases of the Colon and Rectum.* 1999; 42 (1): 71-76. 10.1007/BF02235186.

(240) Müller MJ, Willmann O, Rieger A, Fenk A, Selberg O, Lautz HU, et al. Mechanism of insulin resistance associated with liver cirrhosis. *Gastroenterology.* 1992; 102 (6): 2033-2041. 10.1016/0016-5085(92)90329-w.

(241) Zhang X, Ren X, Zhong Y, Chingin K, Chen H. Rapid and sensitive detection of acetone in exhaled breath through the ambient reaction with water radical cations. *Analyst.* 2021; 146 (16): 5037-5044. 10.1039/D1AN00402F.

(242) *Toxicological Profile for 2-Butanone.* Atlanta (GA): Agency for Toxic Substances and Disease Registry (US); 2020. http://www.ncbi.nlm.nih.gov/books/NBK590518/.

(243) *2-Pentanone.* https://pubchem.ncbi.nlm.nih.gov/compound/7895 [Accessed Sep 28, 2021].

(244) Marks R, Dudley F, Wan A. Trimethylamine metabolism in liver disease. *Lancet (London, England).* 1978; 1 (8073): 1106-1107. 10.1016/s0140-6736(78)90958-3.

(245) *Styrene.* https://www.niehs.nih.gov/health/topics/agents/styrene/index.cfm [Accessed May 27, 2023].

(246) Nakajima T, Elovaara E, Gonzalez FJ, Gelboin HV, Vainio H, Aoyama T. Characterization of the human cytochrome P450 isozymes responsible for styrene metabolism. *IARC scientific publications.* 1993; (127): 101-108.

(247) Stone BG, Besse TJ, Duane WC, Dean Evans C, DeMaster EG. Effect of regulating cholesterol biosynthesis on breath isoprene excretion in men. *Lipids.* 1993; 28 (8): 705-708. 10.1007/BF02535990.

(248) Takase S, Takada A, Yasuhara M, Tsutsumi M. Hepatic aldehyde dehydrogenase activity in liver diseases, with particular emphasis on alcoholic liver disease. *Hepatology (Baltimore, Md.).* 1989; 9 (5): 704-709. 10.1002/hep.1840090508.

(249) *Optimising the analytical performance of sorbent tube sampling and thermal desorption–GC–MS for disease diagnosis via breath and bio-monitoring | Markes International.* https://markes.com/content-hub/application-notes/application-note-154 [Accessed May 13, 2023].

(250) *Tube-Based Thermal Desorption Methods | Markes International.* https://markes.com/content-hub/application-notes/application-note-021 [Accessed May 13, 2023].

(251) Yishai Aviram L, Marder D, Prihed H, Tartakovsky K, Shem-Tov D, Sinelnikov R, et al. pyAIR—A New Software Tool for Breathomics Applications—Searching for Markers in TD-GC-HRMS Analysis. *Molecules.* 2022; 27 (7): 2063. 10.3390/molecules27072063.

(252) Di Gilio A, Palmisani J, Ventrella G, Facchini L, Catino A, Varesano N, et al. Breath Analysis: Comparison among Methodological Approaches for Breath Sampling. *Molecules.* 2020; 25 (24): 5823. 10.3390/molecules25245823.

(253) Behrends V, Tredwell G, Bundy J. *Gavin, a add-on to AMDIS, new GUI-driven version.*

(254) Cain CN, Trinklein TJ, Ochoa GS, Synovec RE. Tile-Based Pairwise Analysis of GC × GC-TOFMS Data to Facilitate Analyte Discovery and Mass Spectrum Purification. *Analytical Chemistry.* 2022; 94 (14): 5658-5666. 10.1021/acs.analchem.2c00223.

(255) Rohart F, Gautier B, Singh A, Lê Cao K. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS computational biology.* 2017; 13 (11): e1005752. 10.1371/journal.pcbi.1005752.

(256) Ruiz-Perez D, Guan H, Madhivanan P, Mathee K, Narasimhan G. So you think you can PLS-DA? *BMC bioinformatics.* 2020; 21 (Suppl 1): 2. 10.1186/s12859-019-3310-7.

(257) Radzieta M. *Multi-omic analysis of host-microbe interactions.* Mendeley; 2020.

(258) Federici S, Kredo-Russo S, Valdés-Mas R, Kviatcovsky D, Weinstock E, Matiuhin Y, et al. Targeted suppression of human IBD-associated gut microbiota commensals by phage consortia for treatment of intestinal inflammation. *Cell.* 2022; 185 (16): 2879-2898.e24. 10.1016/j.cell.2022.07.003.

(259) Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature.* 2019; 569 (7758): 655-662. 10.1038/s41586-019-1237-9.

(260) *Volatile Organic Compounds for the Assessment of Liver Disease: Assessment of Hepatobiliary Disease Through Non-Invasive Detection of Exhaled Volatile Organic Compounds.* 2022. https://clinicaltrials.gov/ct2/show/NCT04620538.

(261) *Rxi-624Sil MS Columns—Exceptionally Inert, Low Bleed Columns for Volatiles Analysis.* https://www.restek.com/row/technical-literature-library/articles/rxi-624sil-ms-columnsexceptionally-inert-low-bleed-columns-for-volatiles-analysis/ [Accessed May 30, 2023].

(262) Blaise BJ, Correia GDS, Haggart GA, Surowiec I, Sands C, Lewis MR, et al. Statistical analysis in metabolic phenotyping. *Nature Protocols.* 2021; 16 (9): 4299-4326. 10.1038/s41596-021-00579-1.

(263) Brown CE. Coefficient of Variation. In: Brown CE. (ed.) *Applied Multivariate Statistics in Geohydrology and Related Sciences.* Berlin, Heidelberg: Springer Berlin Heidelberg; 1998. pp. 155-157.

(264) Kim K, Lee M, Szulejko JE. Simulation of the breakthrough behavior of volatile organic compounds against sorbent tube sampler as a function of concentration level and sampling volume. *Analytica Chimica Acta.* 2014; 835 46-55. 10.1016/j.aca.2014.05.042.

(265) *CompTox Chemicals Dashboard - Diisopropyl phthalate.* https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID2040731 [Accessed Sep 22, 2021].

(266) Humans, IARC Working Group on the Evaluation of Carcinogenic Risks to. *BENZOPHENONE.* : International Agency for Research on Cancer; 2013. https://www.ncbi.nlm.nih.gov/books/NBK373188/.

(267) *U.S. Environmental Protection Agency - Acetophenone.* 1992.

(268) *The Good Scents Company - acetophenone.* http://www.thegoodscentscompany.com/data/rw1000131.html#tooccur [Accessed Sept 22, 2021].

(269) *CompTox Chemicals Dashboard - Benzyl alcohol.* https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID5020152#chemical-functional-use [Accessed 22, September 2021].

(270) Van Gossum A, Decuyper J. Breath alkanes as an index of lipid peroxidation. *The European respiratory journal.* 1989; 2 (8): 787-91. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med3&NEWS=N&AN=2680586.

(271) *D-Limonene.* https://pubchem.ncbi.nlm.nih.gov/compound/440917 [Accessed Sep 28, 2021].

(272) Salehi B, Upadhyay S, Erdogan Orhan I, Kumar Jugran A, L.D. Jayaweera S, A. Dias D, et al. Therapeutic Potential of α- and β-Pinene: A Miracle Gift of Nature. *Biomolecules.* 2019; 9 (11): 10.3390/biom9110738.

(273) Koureas M, Kirgou P, Amoutzias G, Hadjichristodoulou C, Gourgoulianis K, Tsakalof A. Target Analysis of Volatile Organic Compounds in Exhaled Breath for Lung Cancer Discrimination from Other Pulmonary Diseases and Healthy Persons. *Metabolites.* 2020; 10 (8): 10.3390/metabo10080317.

(274) Rondanelli M, Perdoni F, Infantino V, Faliva MA, Peroni G, Iannello G, et al. Volatile Organic Compounds as Biomarkers of Gastrointestinal Diseases and Nutritional Status. *Journal of Analytical Methods in Chemistry.* 2019; 2019 7247802. 10.1155/2019/7247802.

(275) Ager C, Unterkofler K, Mochalski P, Teschl S, Teschl G, Mayhew CA, et al. Modeling-based determination of physiological parameters of systemic VOCs by breath gas analysis, part 2. *Journal of Breath Research.* 2018; 12 (3): 036011. 10.1088/1752-7163/aab2b6.

(276) Meyers RA(A. *Handbook of petroleum refining processes.* New York: McGraw-Hill Education; 2016.

(277) *1-Nonanol.* https://pubchem.ncbi.nlm.nih.gov/compound/8914 [Accessed Sep 29, 2021].

(278) *Alfa Aesar - L03292 Benzyl alcohol, 99%* https://www.alfa.com/en/catalog/L03292/ [Accessed 22, September 2021].

(279) *The Good Scents Company - benzyl alcohol.* http://www.thegoodscentscompany.com/data/rw1001652.html [Accessed Sep 22, 2021].

(280) Maroni M, Seifert B, Lindvall T. *Indoor Air Quality: A Comprehensive Reference Book.* : Elsevier Science; 1995. https://books.google.co.uk/books?id=qsyLaKnn-nYC.

(281) *Phenol.* https://pubchem.ncbi.nlm.nih.gov/compound/996 [Accessed Sep 24, 2021].

(282) Ajibola OA, Smith D, Španěl P, Ferns GAA. Effects of dietary nutrients on volatile breath metabolites. *Journal of Nutritional Science.* 2013; 2 10.1017/jns.2013.26.

(283) Königstein K, Abegg S, Schorn AN, Weber IC, Derron N, Krebs A, et al. Breath acetone change during aerobic exercise is moderated by cardiorespiratory fitness. *Journal of Breath Research.* 2020; 15 (1): 016006. 10.1088/1752-7163/abba6c.

(284) Raninen K, Nenonen R, Järvelä-Reijonen E, Poutanen K, Mykkänen H, Raatikainen O. Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry Analysis of Exhaled Breath Compounds after Whole Grain Diets. *Molecules (Basel, Switzerland).* 2021; 26 (9): 10.3390/molecules26092667.

(285) Schork NJ, Goetz LH. Single-Subject Studies in Translational Nutrition Research. *Annual Review of Nutrition.* 2017; 37 395-422. 10.1146/annurev-nutr-071816-064717.

(286) Rudnicka J, Kowalkowski T, Ligor T, Buszewski B. Determination of volatile organic compounds as biomarkers of lung cancer by SPME-GC-TOF/MS and chemometrics. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences.* 2011; 879 (30): 3360-3366. 10.1016/j.jchromb.2011.09.001.

(287) *Phenol | Public Health Statement | ATSDR.* https://wwwn.cdc.gov/TSP/PHS/PHS.aspx?phsid=146&toxid=27 [Accessed Jun 7, 2023].

(288) Liu P, Xie S, Hu S, Cheng X, Gao T, Zhang C, et al. Age-specific sex difference in the incidence of hepatocellular carcinoma in the United States. *Oncotarget.* 2017; 8 (40): 68131-68137. 10.18632/oncotarget.19245.

(289) Kmeid M, Liu X, Ballentine S, Lee H. Idiopathic Non-Cirrhotic Portal Hypertension and Porto-Sinusoidal Vascular Disease: Review of Current Data. *Gastroenterology Research.* 2021; 14 (2): 49-65. 10.14740/gr1376.

(290) Mochalski P, Sponring A, King J, Unterkofler K, Troppmair J, Amann A. Release and uptake of volatile organic compounds by human hepatocellular carcinoma cells (HepG2) in vitro. *Cancer Cell International.* 2013; 13 (1): 72. 10.1186/1475-2867-13-72.

(291) Janfaza S, Khorsand B, Nikkhah M, Zahiri J. Digging deeper into volatile organic compounds associated with cancer. *Biology Methods & Protocols.* 2019; 4 (1): bpz014. 10.1093/biomethods/bpz014.

(292) Silva CL, Passos M, Câmara JS. Investigation of urinary volatile organic metabolites as potential cancer biomarkers by solid-phase microextraction in combination with gas chromatography-mass spectrometry. *British Journal of Cancer.* 2011; 105 (12): 1894-1904. 10.1038/bjc.2011.437.

(293) Xue R, Dong L, Zhang S, Deng C, Liu T, Wang J, et al. Investigation of volatile biomarkers in liver cancer blood using solid-phase microextraction and gas chromatography/mass spectrometry. *Rapid communications in mass spectrometry: RCM.* 2008; 22 (8): 1181-1186. 10.1002/rcm.3466.

(294) Peng G, Hakim M, Broza YY, Billan S, Abdah-Bortnyak R, Kuten A, et al. Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. *British Journal of Cancer.* 2010; 103 (4): 542-551. 10.1038/sj.bjc.6605810.

(295) Ahmed I, Greenwood R, Costello B, Ratcliffe N, Probert CS. Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease. *Alimentary Pharmacology & Therapeutics.* 2016; 43 (5): 596-611. 10.1111/apt.13522.

(296) Altomare DF, Di Lena M, Porcelli F, Trizio L, Travaglio E, Tutino M, et al. Exhaled volatile organic compounds identify patients with colorectal cancer. *The British Journal of Surgery.* 2013; 100 (1): 144-150. 10.1002/bjs.8942.

(297) D'Amico A, Pennazza G, Santonico M, Martinelli E, Roscioni C, Galluccio G, et al. An investigation on electronic nose diagnosis of lung cancer. *Lung Cancer (Amsterdam, Netherlands).* 2010; 68 (2): 170-176. 10.1016/j.lungcan.2009.11.003.

(298) Liu W, Hill HH. High-Performance Ion Mobility Spectrometry. *Comprehensive Analytical Chemistry.* 2015; 68 275-305. 10.1016/B978-0-444-63340-8.00005-4.

(299) Barash O, Zhang W, Halpern JM, Hua Q, Pan Y, Kayal H, et al. Differentiation between genetic mutations of breast cancer by breath volatolomics. *Oncotarget.* 2015; 6 (42): 44864-44876. 10.18632/oncotarget.6269.

(300) Chen Y, Zhang Y, Pan F, Liu J, Wang K, Zhang C, et al. Breath Analysis Based on Surface-Enhanced Raman Scattering Sensors Distinguishes Early and Advanced Gastric Cancer Patients from Healthy Persons. *ACS Nano.* 2016; 10 (9): 8169-8179. 10.1021/acsnano.6b01441.

(301) Bodelier AGL, Smolinska A, Baranska A, Dallinga JW, Mujagic Z, Vanhees K, et al. Volatile Organic Compounds in Exhaled Air as Novel Marker for Disease Activity in Crohn's Disease: A Metabolomic Approach. *Inflammatory Bowel Diseases.* 2015; 21 (8): 1776-1785. 10.1097/MIB.0000000000000436.

(302) Gaida A, Holz O, Nell C, Schuchardt S, Lavae-Mokhtari B, Kruse L, et al. A dual center study to compare breath volatile organic compounds from smokers and non-smokers with and without COPD. *Journal of Breath Research.* 2016; 10 (2): 026006. 10.1088/1752-7155/10/2/026006.

(303) Rudnicka J, Kowalkowski T, Ligor T, Buszewski B. Determination of volatile organic compounds as biomarkers of lung cancer by SPME–GC–TOF/MS and chemometrics. *Journal of Chromatography B.* 2011; 879 (30): 3360-3366. 10.1016/j.jchromb.2011.09.001.

(304) Filipiak W, Filipiak A, Sponring A, Schmid T, Zelger B, Ager C, et al. Comparative analyses of volatile organic compounds (VOCs) from patients, tumors and transformed cell lines for the validation of lung cancer-derived breath markers. *Journal of Breath Research.* 2014; 8 (2): 027111. 10.1088/1752-7155/8/2/027111.

(305) Buszewski B, Kesy M, Ligor T, Amann A. Human exhaled air analytics: biomarkers of diseases. *Biomedical chromatography : BMC.* 2007; 21 (6): 553-66. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med6&NEWS=N&AN=17431933.

(306) Li M, Yang D, Brock G, Knipp RJ, Bousamra M, Nantz MH, et al. Breath carbonyl compounds as biomarkers of lung cancer. *Lung Cancer (Amsterdam, Netherlands).* 2015; 90 (1): 92-97. 10.1016/j.lungcan.2015.07.005.

(307) Jalali M, Zare Sakhvid MJ, Bahrami A, Berijani N, Mahjub H. Oxidative Stress Biomarkers in Exhaled Breath of Workers Exposed to Crystalline Silica Dust by SPME-GC-MS. *Journal of Research in Health Sciences.* 2016; 16 (3): 153-161. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7191029/.

(308) Krause ST, Liao P, Crocoll C, Boachon B, Förster C, Leidecker F, et al. The biosynthesis of thymol, carvacrol, and thymohydroquinone in Lamiaceae proceeds via cytochrome P450s and a short-chain dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America.* 2021; 118 (52): e2110092118. 10.1073/pnas.2110092118.

(309) Tadtong S, Kamkaen N, Watthanachaiyingcharoen R, Ruangrungsi N. Chemical Components of Four Essential Oils in Aromatherapy Recipe. *Natural Product Communications.* 2015; 10 (6): 1091-1092.

(310) *Terpinolene.* https://pubchem.ncbi.nlm.nih.gov/compound/11463 [Accessed Jun 15, 2023].

(311) Zehetner P, Höferl M, Buchbauer G. Essential oil components and cytochrome P450 enzymes: a review. *Flavour and fragrance journal.* 2019; 34 (4): 223-240. 10.1002/ffj.3496.

(312) Niaz K, Bahadar H, Maqbool F, Abdollahi M. A review of environmental and occupational exposure to xylene and its health concerns. *EXCLI Journal.* 2015; 14 1167-1186. 10.17179/excli2015-623.

(313) Komolafe O, Roberts D, Freeman SC, Wilson P, Sutton AJ, Cooper NJ, et al. Antibiotic prophylaxis to prevent spontaneous bacterial peritonitis in people with liver cirrhosis: a network meta-analysis. *The Cochrane Database of Systematic Reviews.* 2020; 2020 (1): CD013125. 10.1002/14651858.CD013125.pub2.

(314) Murgia A, Ahmed Y, Sweeney K, Nicholson-Scott L, Arthur K, Allsworth M, et al. Breath-Taking Perspectives and Preliminary Data toward Early Detection of Chronic Liver Diseases. *Biomedicines.* 2021; 9 (11): 1563. 10.3390/biomedicines9111563.

(315) Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open.* 2016; 6 (11): e012799. 10.1136/bmjopen-2016-012799.

(316) *Diisopropyl phthalate analytical standard | Sigma-Aldrich.* http://www.sigmaaldrich.com/ [Accessed Sep 22, 2021].

(317) *CompTox Chemicals Dashboard - Azulene.* https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID2059770#chemical-functional-use [Accessed Sept 22, 2021].

(318) *CompTox Chemicals Dashboard - 1-Methylnaphthalene.* https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID9020877 [Accessed Sept 22, 2021].

(319) *The Good Scents Company - 1-methyl naphthalene.* http://www.thegoodscentscompany.com/data/rw1008021.html#tooccur [Accessed Sept 22, 2021].

(320) *CompTox Chemicals Dashboard - Benzaldehyde .* https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID8039241#chemical-functional-use [Accessed Sept 22, 2021].

(321) *The Good Scents Company - Benzaldehyde .* http://www.thegoodscentscompany.com/data/rw1001492.html [Accessed Sept 22, 2021].

(322) *The Good Scents Company - Propanoic acid, 2-methyl-, 2-ethyl-3-hydroxyhexyl ester.* http://www.thegoodscentscompany.com/data/rw1656781.html [Accessed Sept 22, 2021].

(323) *CompTox Chemicals Dashboard - Benzene, 1,1'-(1,2-ethynediyl)bis-.* https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID4060109#exposure [Accessed Sept 22, 2021].

(324) *CompTox Chemicals Dashboard - Isobutyl Salicylate.* https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID3047186#chemical-functional-use [Accessed Sept 25, 2021].

(325) U.S. EPA. *IRIS Toxicological Review of 2-Methylnaphthalene.* 2003.

(326) *CompTox Chemicals Dashboard - Benzothiazole .* https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID7024586#chemical-functional-use [Accessed Sept 25, 2021].

(327) *The Good Scents Company - Benzothiazole .* http://www.thegoodscentscompany.com/data/rw1007521.html#tooccur [Accessed Sept 25, 2021].

(328) *Ethylbenzene- ToxFAQs™.* https://www.atsdr.cdc.gov/toxfaqs/tfacts110.pdf [Accessed Sept 25, 2021].

(329) *The Good Scents Company - Ethylbenzene .* http://www.thegoodscentscompany.com/data/rw1022281.html#touses [Accessed Sept 25, 2021].

(330) Hoydonckx HE, Rhijn WMV, Rhijn WV, Vos DED, Jacobs PA. Furfural and Derivatives. *Ullmann's Encyclopedia of Industrial Chemistry.* American Cancer Society; 2007.

(331) *CompTox Chemicals Dashboard - m-cymene.* https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID2060206#chemical-functional-use [Accessed Sept 25, 2021].

(332) *CompTox Chemicals Dashboard - m-Xylene .* https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID6026298#chemical-functional-use [Accessed Sept 25, 2021].

(333) *The Good Scents Company - meta-Xylene.* http://www.thegoodscentscompany.com/data/rw1257341.html [Accessed Sept 25, 2021].

(334) *CompTox Chemicals Dashboard - 1-Dodecanol.* https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID5026918 [Accessed Sept 25, 2021].

(335) *112-53-8 - 1-Dodecanol, 98% - Lauryl alcohol - Dodecyl alcohol - A12228 - Alfa Aesar.* https://www.alfa.com/en/catalog/A12228/ [Accessed Sept 25, 2021].

(336) *p-Xylene.* https://pubchem.ncbi.nlm.nih.gov/compound/7809 [Accessed Sep 24, 2021].

(337) *1-Phenyl-1,3,3-trimethylindan.* https://pubchem.ncbi.nlm.nih.gov/compound/19793 [Accessed Sep 24, 2021].

(338) *Geranylacetone.* https://pubchem.ncbi.nlm.nih.gov/compound/1549778 [Accessed Sep 28, 2021].

(339) *Comptox Chemicals Dashboard - (E)-6,10-Dimethylundeca-5,9-dien-2-one.* https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID4052053#exposure [Accessed Sept 25, 2021].

(340) Phillips M, Cataneo RN, Greenberg J, Grodman R, Gunawardena R, Naidu A. Effect of oxygen on breath markers of oxidative stress. *The European Respiratory Journal.* 2003; 21 (1): 48-51. 10.1183/09031936.02.00053402.

(341) Kohlmeier M. Oxalate. In: Kohlmeier M. (ed.) *Nutrient Metabolism.* London: Academic Press; 2003. pp. 232-235.

(342) *ContaminantDB - 1-(Methylthio)propane.* https://contaminantdb.ca/contaminants/CHEM033742 [Accessed Sept 25, 2021].

(343) *The Good Scents Company - allyl methyl sulfide.* http://www.thegoodscentscompany.com/data/rw1038781.html#tooccur [Accessed Sept 25, 2021].

(344) Łuczaj W, Skrzydlewska E. DNA damage caused by lipid peroxidation products. *Cellular & Molecular Biology Letters.* 2003; 8 (2): 391-413.

(345) *2-Propyl-1-heptanol.* https://pubchem.ncbi.nlm.nih.gov/compound/24847 [Accessed Sep 24, 2021].

(346) *D-Limonene.* https://pubchem.ncbi.nlm.nih.gov/compound/440917 [Accessed Sep 24, 2021].

(347) Steinemann A, Nematollahi N, Weinberg J, Flattery J, Goodman N, Kolev S. Volatile chemical emissions from car air fresheners. *Air Quality, Atmosphere & Health.* 2020; 13 10.1007/s11869-020-00886-8.

(348) *The Good Scents Company - hexacosane.* http://www.thegoodscentscompany.com/data/rw1249621.html#toeuus [Accessed Sept 25, 2021].

(349) *Pentacosane.* https://pubchem.ncbi.nlm.nih.gov/compound/12406 [Accessed Sep 24, 2021].

(350) *The Good Scents Company - pentacosane .* http://www.thegoodscentscompany.com/data/rw1272351.html#tooccur [Accessed Sept 25, 2021].

(351) *The Good Scents Company - Heptacosane.* http://www.thegoodscentscompany.com/data/rw1249251.html#tooccur [Accessed Sept 25, 2021].

(352) *Pentadecane.* https://pubchem.ncbi.nlm.nih.gov/compound/12391 [Accessed Sep 24, 2021].

(353) *The Good Scents Company - pentadecane.* http://www.thegoodscentscompany.com/data/rw1272391.html [Accessed Sept 25, 2021].

(354) *The Good Scents Company - Tetracosane .* http://www.thegoodscentscompany.com/data/rw1286971.html#tooccur [Accessed Sept 25, 2021].

(355) *1-Pentadecene.* https://pubchem.ncbi.nlm.nih.gov/compound/25913 [Accessed Sep 24, 2021].

(356) *The Good Scents Company - 2-ethyl-3-hydroxyhexyl 2-methyl propanoate.* http://www.thegoodscentscompany.com/data/rw1656781.html [Accessed Sept 25, 2021].

(357) *Isoprene.* https://pubchem.ncbi.nlm.nih.gov/compound/6557 [Accessed Sep 28, 2021].

(358) *Nonanal.* https://pubchem.ncbi.nlm.nih.gov/compound/31289 [Accessed Sep 28, 2021].

(359) *Decanal.* https://pubchem.ncbi.nlm.nih.gov/compound/8175 [Accessed Sep 28, 2021].

(360) *The Good Scents Company - Benzaldehyde, 3,4-dimethyl-.* http://www.thegoodscentscompany.com/data/rw1415161.html#tooccur [Accessed Sept 25, 2021].

(361) *Chemical Book - 3,4-Difluorobenzaldehyde.* https://www.chemicalbook.com/ChemicalProductProperty_EN_CB9275911.htm [Accessed Sept 25, 2021].

(362) *Isopropyl alcohol.* https://pubchem.ncbi.nlm.nih.gov/compound/3776 [Accessed Sep 28, 2021].

(363) *112-72-1 - 1-Tetradecanol, 97+% - Myristyl alcohol - n-Tetradecyl alcohol - A19638 - Alfa Aesar.* https://www.alfa.com/en/catalog/A19638/ [Accessed Sept 25, 2021].

(364) *Acetone.* https://pubchem.ncbi.nlm.nih.gov/compound/180 [Accessed Sep 28, 2021].

(365) *O-Cymene.* https://pubchem.ncbi.nlm.nih.gov/compound/10703 [Accessed Sep 28, 2021].

(366) *2-Pentylfuran.* https://pubchem.ncbi.nlm.nih.gov/compound/19602 [Accessed Sep 28, 2021].

(367) *Dodecane.* https://pubchem.ncbi.nlm.nih.gov/compound/8182 [Accessed Sep 28, 2021].

(368) *Benzonitrile.* https://pubchem.ncbi.nlm.nih.gov/compound/7505 [Accessed Sep 28, 2021].

(369) *Undecane ≥99% | Sigma-Aldrich.* http://www.sigmaaldrich.com/ [Accessed Sep 28, 2021].

(370) *6-Methyl-5-hepten-2-one.* https://pubchem.ncbi.nlm.nih.gov/compound/9862 [Accessed Sep 28, 2021].

(371) *Tridecane.* https://pubchem.ncbi.nlm.nih.gov/compound/12388 [Accessed Sep 28, 2021].

(372) *alpha-Pinene.* https://pubchem.ncbi.nlm.nih.gov/compound/6654 [Accessed Sep 28, 2021].

(373) *Toluene.* https://pubchem.ncbi.nlm.nih.gov/compound/1140 [Accessed Sep 28, 2021].

(374) *Crotonaldehyde.* https://pubchem.ncbi.nlm.nih.gov/compound/447466 [Accessed Sep 28, 2021].

(375) *CompTox Chemicals Dashboard - Acetamide, 2,2,2-trifluoro-N-methyl-.* https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID6061153#exposure [Accessed Sept 25, 2021].

(376) *Pyridine.* https://pubchem.ncbi.nlm.nih.gov/compound/1049 [Accessed Sep 28, 2021].

(377) *Ethyl 4-methoxybenzoate.* https://pubchem.ncbi.nlm.nih.gov/compound/60979 [Accessed Sep 29, 2021].

(378) *o-Xylene.* https://pubchem.ncbi.nlm.nih.gov/compound/7237 [Accessed Sep 29, 2021].

(379) *Dodecyl vinyl ether,USD,7.50/Kilogram,cas No:765-14-0,Formula:C14H28O,purity:98%min- Shanghai Jinhong Chemical Co.,Ltd.* https://www.guidechem.com/trade/dodecyl-vinyl-ether-id3448477.html [Accessed Sep 29, 2021].

(380) *2-Phenoxyethanol.* https://pubchem.ncbi.nlm.nih.gov/compound/31236 [Accessed Sep 29, 2021].