Imperial College London

Biochemical histology analysis of tissue samples by Desorption Electrospray Ionization (DESI) mass spectrometry imaging

Thesis submitted by

Anna Karolina Mróz

For the degree of Doctor of Philosophy (PhD)

Imperial College London

Department of Metabolism, Digestion and Reproduction

Faculty of Medicine

2021

Declaration of Originality

I certify that this thesis, and the research to which it refers, are the product of my work and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Anna Mróz

Copyright Declaration

For my parents,
who never question, only believe.

Acknowledgements

I would like to thank my supervisor, Professor Zoltan Takats, for this opportunity and all the guidance he has offered me throughout my PhD. It was a fantastic experience to work in his lab and interact with all scientists possessing invaluable expertise.

I would like to say a huge Thank you to my postdoc mentors – Dr Renata Soares and Dr Maria Luísa Dória. All your support is invaluable and so much appreciated! I am so grateful for your guidance regarding my research and for always being there for me as a friend.

I would also like to thank Dr Paolo Inglese, Dr James McKenzie and Dr Kirill Veselkov for the development of the data analysis tools for this project and their assistance with the data analysis.

I cannot forget to thank Dr Evagelia Liaskou and Dr James Alexander for all the clinical support, and most importantly, for providing the samples for this study.

I would also like to acknowledge all Professor Takats lab members, past and present, for all their support during various stages of my PhD. I would never get to this point without the help I have received from all of you. Thank you!

I would like to thank Waters Corporation for offering their instrumentation and support.

Finally, I would like to thank my family for their never-ending support and faith in me.

It has been a long journey and one of a kind. There were moments of doubt and frustration, but most of the time, it was a marvellous and unforgettable experience. Once again, thank you all who played a part in it!

Abstract

For over 100 years, the histopathological analysis of cytology, biopsy or resection specimens has been the final step in the process of diagnosing multiple diseases, including cancer. In recent years, standard clinical care is continuously becoming more complex, and as a result, diagnostic pathology workup is also more complex and extensive. Moreover, despite being considered a gold standard in making a diagnosis, histopathological investigations can be time-consuming. Additionally, an examination of the stained slides is subject to intra-observer error. Therefore, it is evident that some additional techniques are required to complement making a diagnosis.

Desorption electrospray ionisation mass spectrometric imaging (DESI-MSI) is an emerging mass spectrometry technique with great potential in tissue analysis, especially in histological settings. DESI-MSI enables visualising the spatial distribution of lipid species across tissue sections allowing a direct correlation of the metabolomic information with the morphological features. However, this technique has always relied on frozen sections, which are not required in routine histopathology settings very often. Moreover, some embedding media, e.g. OCT, a common choice in diagnostic laboratories, have been proven not to be very well suited for MSI.

The main aim of this study was to make DESI-MSI more compatible with the standard pathology procedures.

Therefore, the first step was to assess OCT's impact on the quality of DESI-MSI data. The acquired data suggested that this embedding medium could be used for histopathological and mass spectrometric analyses. There were no clear polymeric signals causing differences in the negative mode data, but some reduction in intensities might be attributable to polymer-induced ion suppression. In positive mode data, the interferences due to OCT were more overt but could be negated by removing the regular peaks of the various polymeric distributions.

As formalin-fixed, paraffin-embedded (FFPE) samples are the gold standard in histopathology laboratories worldwide, the next step was to optimise the pre-DESI-MSI protocol to allow the analysis of specimens that have been processed that way. A new protocol has been adapted and successfully tested on FFPE mouse and human tissue samples for tissue classification.

Additionally, DESI-MSI has been used to analyse fresh-frozen and FFPE colorectal samples. 88.5% accuracy for normal samples and 91.7% for tumours was achieved when a batch of 38 fresh-frozen samples was analysed. Tissue microarray (TMA) consisting of 54 cores was used further to test the application of DESI-MSI to FFPE samples. A 10µm thick sections were subjected to analysis in negative and positive modes, and accuracy of over 80% and 92% for tissue prediction was achieved, respectively. Equally good results were obtained for TMA sections which were 5µm thick. This last observation was crucial in the light of making DESI-MSI as histology-friendly as possible, as 10µm tissue sections are not routinely prepared in histopathology laboratories.

Lastly, a new statistical approach based on ion colocalisation features has been applied to DESI-MSI data acquired for cirrhotic liver diseases. It allowed to identify top correlations of ions, and their distribution within analysed tissue sections was visualised. It is possible that using this approach, some biochemical interactions that are distinguishing the three classes of cirrhotic liver diseases (metabolic, hepatitis and cholangiopathy) could be captured. The colocalisation patterns can potentially be used for data-driven hypothesis generation, suggesting possible local molecular mechanisms characterising the samples of interest.

Table of contents

List of Figures

List of Tables

Abbreviations

AIH	Autoimmune hepatitis
ALD	Alcoholic liver disease
ANOVA	Analysis of variance
AP-MALDI	Atmospheric pressure MALDI
ATP	Adenosine triphosphate
CE	Capillary electrophoresis
Cer	Ceramide
CI	Chemical ionization
CL	Cardiolipin
CRC	Colorectal cancer
CT	Computerized tomography
DA	Discriminant analysis
DESI	Desorption electrospray ionisation
DESI-MSI	Desorption electrospray ionisation mass spectrometry imaging
D/I	Desorption/ionization
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
EI	Electron ionisation
EMVI	Extramural vascular invasion
ER	Endoplasmic reticulum
ESI	Electrospray ionisation
FAB	Fast atom bombardment
FF	Formalin-fixed
FFPE	Formalin-fixed, paraffin-embedded
FGF	Fibroblast growth factors
FT-ICR	Fourier transform ion cyclotron resonance
FT-Orbitrap	Fourier transform Orbitrap
GC/MS	Gas chromatography/mass spectrometry
GPCR	G protein-coupled receptor
HBV	Hepatitis B virus
HCV	Hepatitis C virus

HCA	Hierarchical clustering analysis
HILIC	Hydrophilic interaction liquid chromatography
HPLC	High-performance liquid chromatography
H&E	Hematoxylin and eosin
IBD	Inflammatory bowel disease
ID	Inner diameter
IHC	Immunohistochemistry
IRMS	Isotope ratio mass spectrometry
LC	Liquid chromathography
LC/MS	Liquid chromatography/mass spectrometry
LDA	Linear discriminant analysis
LOD	Limit of detection
LOQ	Limit of quantification
LSIMS	Liquid secondary ion mass spectrometry
MALDI	Matrix-assisted laser desorption ionisation
MMC-DA	Maximum margin criterion discriminant analysis
MRI	Magnetic resonance imaging
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MSI	Mass spectrometry imaging
m/z	mass-to-charge ratio
NADPH	Nicotinamide adenine dinucleotide phosphate
NAFLD	Non-alcoholic fatty liver disease
NASH	Non-alcoholic steatohepatitis
OCT	Optimal cutting temperature compound
OD	Outer diameter
PA	Phosphatidic acid
PBC	Primary biliary cholangitis
PC	Phosphatidylcholine
PC1	Principal component 1
PCA	Principal component analysis
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PI	Phosphatidylinositol

PLS	Partial least squares
ppm	Parts per million
PS	Phosphatidylserine
PSC	Primary sclerosing cholangitis
QC	Quality control
REC	Research Ethics Committees
RMMC	Recursive maximum margin criterion
RNA	Ribonucleic acid
ROI	Region of interest
RP	Reversed-phase
RPM	Revolutions per minute
SIMS	Secondary ion mass spectrometry
SM	Sphingomyelin
TIC	Total ion count
TMA	Tissue microarray
TNM	Tumour Nodes Metastases
TOF	Time-of-flight
TQ	Triple quadrupole instrument
uPAR	Urokinase-type plasminogen activator-receptor
VEGF	Vascular endothelial growth factor

Chapter 1
Introduction

1.1 Cancer

Cancer has been estimated the second leading cause of death in the United States and is a significant health problem in more and less economically developed countries worldwide. It had been projected that in the USA alone, 1 806 590 new cases and 606 520 cancer-related deaths were expected in 2020 [1, 2]. The occurrence of cancer continues to increase further due to the population's growth and ageing. Moreover, a rising prevalence of established risk factors such as overweight, smoking, and physical inactivity also plays a significant role in the cancer burden [3].

This condition's complexity is linked to cancer being a highly heterogeneous collection of different diseases with different risk factors, clinical presentations, pathological features, responses to therapy, and outcomes. Moreover, over 200 types of cancer have been described, each with different causes, symptoms and treatments [4].

The fundamental abnormality resulting in cancer development is the continual unregulated proliferation of cancer cells. Instead of responding appropriately to the signals that control normal cell behaviour, cancer cells continue to grow and divide in an uncontrolled manner. As a result, they invade healthy tissues and organs and eventually start to spread throughout the body. The loss of growth control exhibited by cancer cells directly results from accumulated abnormalities in multiple cell regulatory systems. Furthermore, several aspects of cell behaviour distinguish cancer cells from their healthy counterparts [5].

These aspects can be described as six essential alterations in cell physiology: insensitivity to growth-inhibitory signals, self-sufficiency in growth signals, evasion of apoptosis, sustained angiogenesis, limitless replicative potential, and tissue invasion and metastasis (Figure 1.10) [6].

Two distinct mechanisms can be used to block proliferation. Cells may be induced to surrender their proliferative potential by being induced to enter a postmitotic state. Alternatively, they might be forced out of the active proliferative cycle into the G_0 state. At the molecular level, many of the antiproliferative signals are guided by the retinoblastoma protein pRb and its relatives, p107 and p130. Therefore, disruption of the pRb pathway allows cell proliferation [6].

Healthy cells rely on mitogenic growth signals before they can move into an active proliferative state. Tumour cells, however, show a significantly reduced dependence on exogenous growth stimulation. As a result, these cells must generate many of their own growth signals, allowing them to reduce their dependence on stimulation from the normal tissue microenvironment. That

leads to disturbances in a critically important homeostatic mechanism that ensures the proper behaviour of various cell types within a tissue [6].

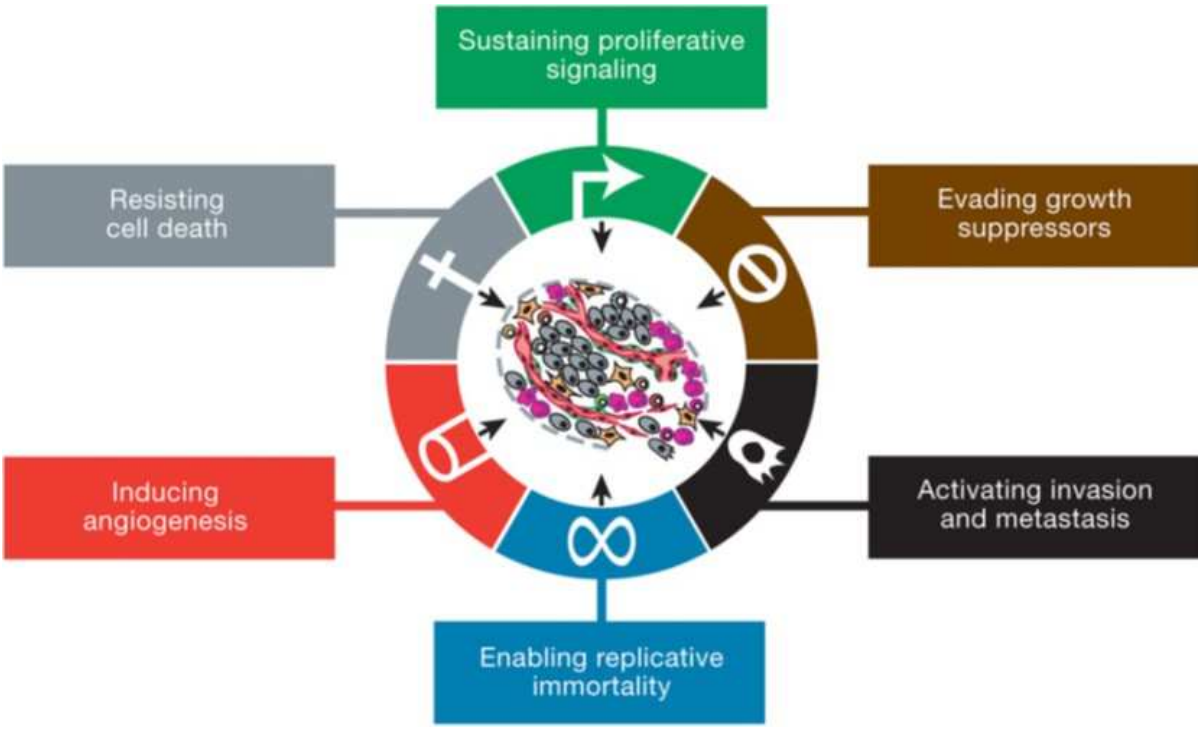


Figure 1.1. Acquired capabilities of cancerous cells. Reproduced with permission [7].

Many strategies allow cancer cells to become resistant to apoptosis, and the p53 tumour suppressor gene is involved in the most commonly occurring loss of a proapoptotic regulator. The p53 protein is seen in more than 50% of human cancers and results in removing a vital component of the DNA sensor responsible for inducing the apoptotic effector cascade [8]. Signal linked to other abnormalities like oncogene hypertension and hypoxia is also funnelled in part via p53 to the apoptotic machinery. When p53 function is lost, these are also impaired at eliciting apoptosis [9].

Nutrients and the oxygen supplied by the vascular system are essential for cell function and survival. Coordinated growth of parenchyma and vessels occurs during organogenesis and is transitory and carefully regulated once the tissue is formed. Interestingly, cells within aberrant proliferative lesions cannot initially grow new blood vessels (angiogenesis). Therefore, to progress to a larger size, angiogenic ability must be developed by incipient neoplasias [10, 11]. Angiogenesis is encouraged or blocked by counterbalancing positive and negative signals. Soluble factors and their receptors displayed on the surface of endothelial cells convey one

class of these signals. Adhesion molecules and integrins responsible for mediating cell-cell and cell-matrix association also play crucial roles [6]. The signals initiating angiogenesis are exemplified by acidic and basic fibroblast growth factors (FGF1/2) and vascular endothelial growth factor (VEGF). Each of them binds to transmembrane tyrosine kinase receptors displayed by endothelial cells [12, 13]. Thrombospondin-1 is a prototypical angiogenesis inhibitor. It binds to CD36, a transmembrane receptor localised on endothelial cells coupled to intracellular Src-like tyrosine kinases [14].

Moreover, in order to generate macroscopic tumours, cancer cells require unlimited replicative potential. There is considerable evidence that telomeres protecting the ends of chromosomes are involved in the capability for unlimited proliferation [15, 16]. The telomeres are composed of multiple tandem hexanucleotide repeats, and they shorten progressively in non-immortalised cells when propagated in culture. Eventually, this leads to the loss of ability to protect the ends of chromosomal DNAs from end-to-end fusion. Those fusions generate unstable dicentric chromosomes, and their resolution leads to a scrambling of karyotype that threatens cell viability [7]. Consequently, the telomeric DNA length in a cell dictates how many successive cell generations its progeny can pass through before telomeres are significantly eroded and lost their protective functions [7].

Lastly, it is clear that as carcinomas arising from epithelial tissues progress to a higher pathological grade of malignancy, the associated cancer cells usually develop alterations in their shape and attachment to other cells as well as the extracellular matrix (ECM). The best-described alteration involves the loss by carcinoma cells of E-cadherin, a critical cell-to-cell adhesion molecule. Increased expression of this molecule is well established as an antagonist of metastasis and invasion, while the reduction of E-cadherin expression is known to potentiate these phenotypes. Frequently observed downregulation of E-cadherin and its occasional mutational inactivation in human carcinomas provides strong support for its role as a critical suppressor of activating invasion and metastasis [17, 18].

With further research into cancer development, additional hallmarks have been added to the original list, including the adjustment and dysregulation of cancer cells metabolism to effectively support proliferation and cancer cells' capability to evade immunological destruction [7].

1.2 Lipids

The lipid classification system currently most frequently referenced was proposed in 2005 by Fahy *et al.* [19]. It was updated in 2009 to encompass lipid structures from non-mammalian sources such as plants, bacteria, and fungi [20]. As a result, there are eight categories, which cover both eukaryotic and prokaryotic lipids: glycerolipids, glycerophospholipids or phospholipids, sphingolipids, saccharolipids, fatty acyls, prenol lipids, sterol lipids, and polyketides.

Lipids, which are fundamental components of biological membranes, are functionally and structurally diverse classes of metabolites. Due to their diversity and combinatorial structures, there are already over 37,500 lipids described by Fahy *et al.* in the LIPID MAPS Structure Database [21]. They play essential roles in biological systems, including composing membrane bilayers, signal transduction, storing energy, providing functional implementations of membrane proteins, and their interactions.

Sphingolipids, glycerophospholipids, triglycerides and sterols are the most common lipid classes which play important roles in membrane structure and energy storage. There is also another classification system used to categorise the most abundant lipids. It references their different charge properties: anionic (phosphatidic acid (PA), phosphatidylinositol (PI), phosphatidylglycerol (PG), phosphatidylserine (PS), and cardiolipin (CL)), weakly anionic (phosphatidylethanolamine (PE) and ceramide (Cer)) and 'neutral' lipids (mono, di, triacylglycerol, cholesterol esters, phosphatidylcholine (PC), sphingomyelin (SM)) [22].

It has been estimated that 5% of eukaryotic cell genes are directly involved in cellular lipid metabolism, underlining the importance of lipids in cellular structure and function. The lipidome, often considered a sub-fraction of the metabolome, contributes to around 70% of entries in the Human Metabolome Database [23].

Therefore, it is crucial to understand lipids' role in healthy and diseased cells as they may have prognostic value [24].

1.3 Lipid metabolism in cancer

There is increasing evidence that cancerous cells exhibit specific alterations in various aspects of lipid metabolism. Those alterations can impact the availability of structural lipids to synthesise and degrade lipids, contributing to energy homeostasis, the synthesis of membranes, and the abundance of lipids responsible for signalling functions [25].

In order to meet their increased demands for proliferation and survival, cancer cells can alter their metabolic pathways. Their accelerated proliferation rate requires a higher metabolism rate [26, 27], and the 'Warburg effect', the most known metabolic change, was first described by Otto Warburg in the 1920s [28]. It was then reported that compared to normal cells, cancer cells take up and utilise much more glucose for glycolysis [28]. It has been proposed that aerobic glycolysis should be considered the core cellular metabolism to supply cancer cells with energy, and also the building blocks required the synthesis of macromolecule proteins, nucleic acids, lipids and carbohydrates [29]. In recent years, more and more evidence confirms that altered lipid metabolism is another common property of cancerous cells [27]. Similar to glucose metabolism, common oncogenic signalling pathways regulate lipid metabolism in malignant cells. It is believed that in that case, altered lipid metabolism plays a crucial role in the process of initiation and progression of tumours [30]. In order to synthesize fatty acids and their derivatives, many lipogenic enzymes utilize reduced acetyl-CoA and nicotinamide adenine dinucleotide phosphate (NADPH) generated from glucose and glutamine metabolism. As a result, the exacerbated lipogenesis in cancerous cells is linked to the upregulated lipid metabolizing enzymes, and additionally, is also directly coupled not only to other common metabolic pathways but also their associated cell signalling pathways [26, 27].

Therefore, the signalling pathways responsible for controlling the altered metabolism in malignant cells are promising targets for cancer therapy [31].

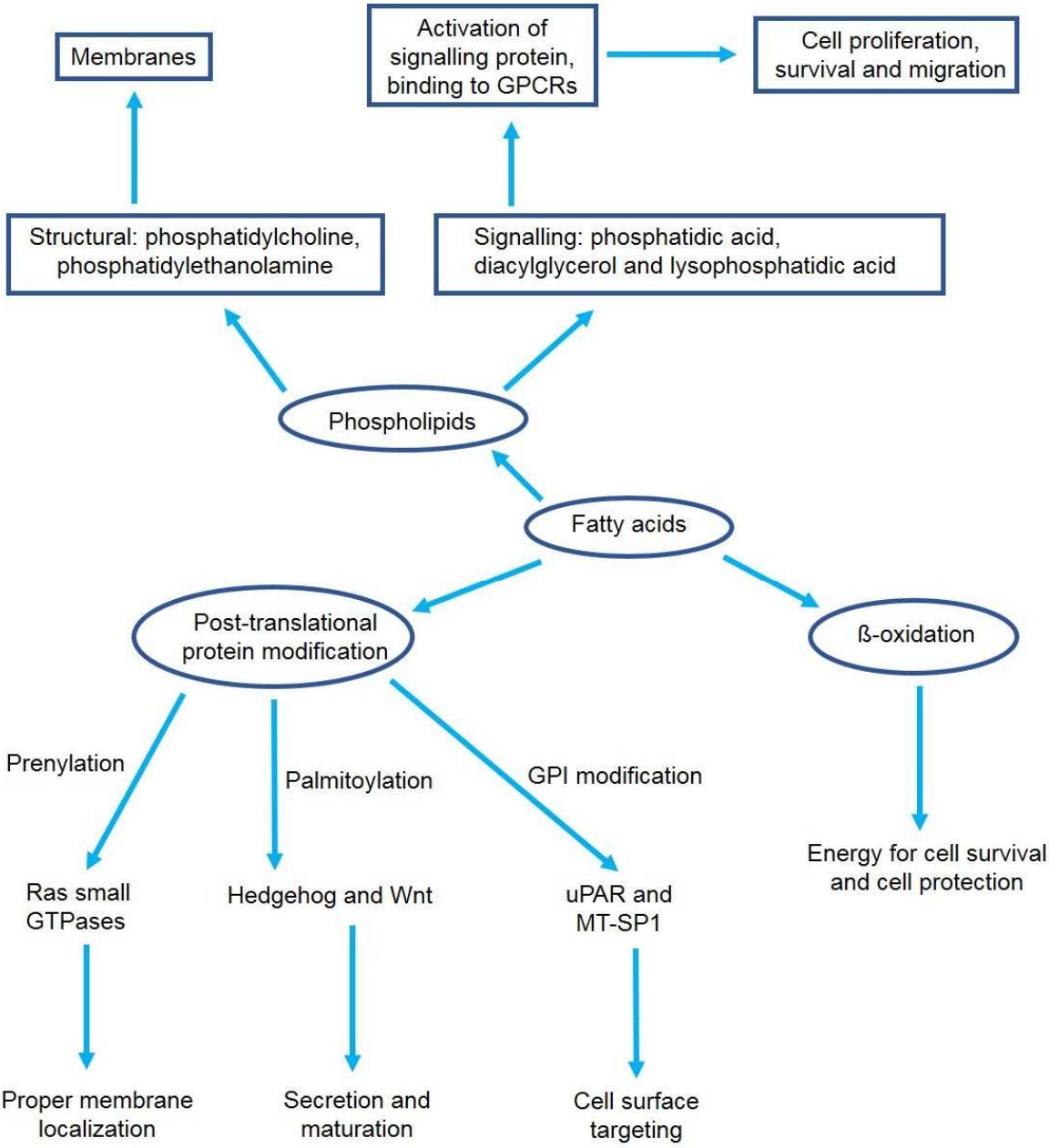


Figure 1.2. Functions of lipids in cancerous cells. Lipids play many crucial roles in cancer cells. They are responsible for providing those cells with signalling molecules, membrane building blocks, energy supply to support rapid cell proliferation, and posttranslational modifications of proteins. uPAR – Urokinase-type plasminogen activator-receptor; GPCRs – G protein-coupled receptors. Adapted from [31].

1.4 Cancer diagnosis

Cancer is described as a physical alteration of the relation cells and their tissues, resulting in aberrant social organization. These alterations are detected as tumours, and a cancer diagnosis is made through manual histological evaluation of biopsies taken from the mass and its

surrounding regions. This evaluation is one of the key prognostic factors for most cancers, providing information on the tumour type, grade and any morphological abnormalities [32].

Grading and staging of tumours are used to decide on the treatment for each particular diagnosis and predict the outcome of individual cancers. Grading relies on the histological criteria of neoplasm. It also includes information about the degree of differentiation and proliferation of individual cells and the degree of deviation from healthy tissue architecture. Extensive studies have contributed to the correlation of these microscopic characteristics with clinical outcomes. Grading of invasive and preinvasive neoplasms is helpful not only in prognosis but also in therapy decisions. Staging is based on the local extent of tissue involvement combined with microscopic confirmation of the neoplastic cells' presence in distant sites [32].

Traditional histological evaluation of the post-operative tissue sections is the standard gold method that has been used over the decades to enable clinicians to understand and characterise tumours. Whilst this approach is relatively simple and powerful, it has some significant drawbacks. There is a broad spectrum in cancer morphology; therefore, many tumours can be atypical or lack morphological features crucial for future diagnosis [33]. For example, bone tumours are classified as a challenging field in pathology, and often they require a global assessment of clinical and imaging data. However, even under perfect conditions where a representative biopsy and imaging documents are available, diagnostic difficulties persist [34]. Another aspect that should also be mentioned here is an inter-observer error [35-37]. These two studies looked at the rate of agreement between specialist histopathologists evaluating liver biopsy specimens. Depending on the scoring system used, that level varied between 87%-65%. Lastly, manual examinations are very time consuming and hence limit the speed and throughput of diagnosis. All the steps required before the final histopathological report is issued are discussed in the next paragraph.

1.5 Why can a histopathological assessment be time-consuming?

Tissues are usually taken either from the biopsies of patients in clinics or during surgical operations. The tissue samples are immediately immersed in a fixative or rapidly frozen and then fixed. The tissue sections are prepared for a histopathological assessment by a pathologist, who examines their cellular appearance [38]. However, there are multiple steps required before the slides are ready for the assessment, and the whole process is summarized in Figure 1.3.

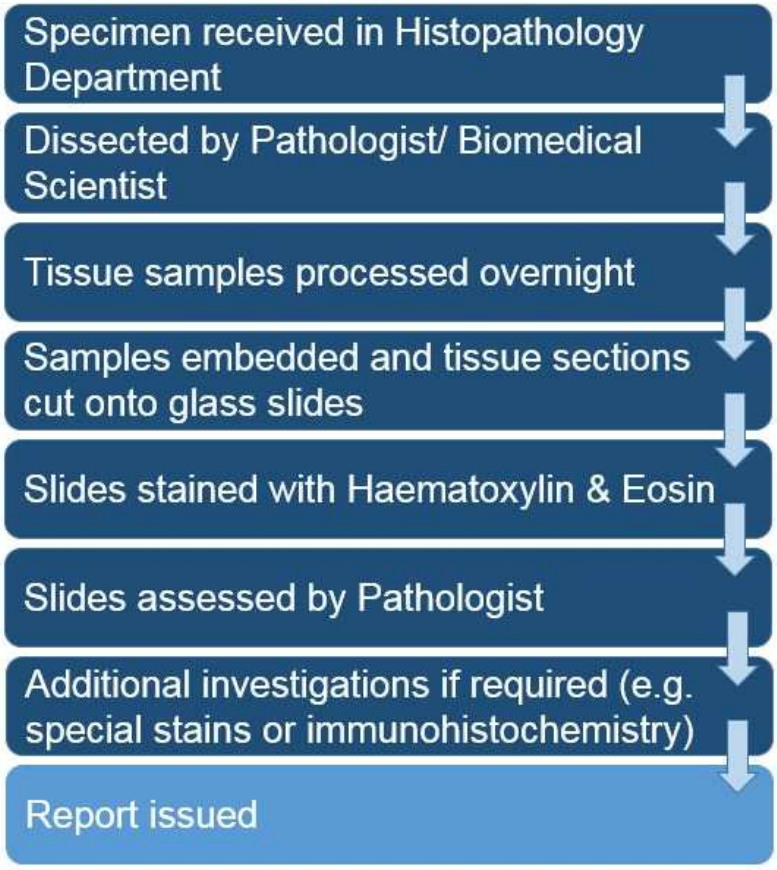


Figure 1.3. All the steps required from a specimen's arrival in a Histopathology Department until the final report can be issued. The total number of days required to complete the whole procedure depends on whether additional tests are requested (e.g. special stains or immunohistochemistry (IHC) staining).

1.5.1 Fixation

Living cells must receive food and oxygen as well as discard waste products. They decline and die when removed from the body because neither food nor oxygen is available. The DNA, sugars, proteins and fats start to deteriorate through waste products build up and the uncontrolled release of proteases. Thus, the first part of the histopathology process is not to allow cells' shape and structure to deteriorate; otherwise, diagnosis becomes impossible as those changes are irreversible. This can be done by either fixing or freezing the tissue sample to keep cells in as natural a state as possible [39].

1.5.1.1 Fixatives – what do they do?

Most fixatives are enzyme poisons because fixation should prevent abnormal enzyme activity and metabolism and inactivate the lysosomal enzymes to stop autolysis. Destroying enzymes

is essential to prevent fungal and bacterial growth, so putrefactive changes do not occur [40]. In terms of maintaining morphology, probably the most important reactions are those which stabilize the proteins. Fixatives form cross-links between proteins, forming a gel, so everything is kept in their *in vivo* relations to each other [41]. Many chemicals act as enzyme poisons, but only some of them are useful as histological fixatives. Aqueous solutions are usually used, but occasionally vapour may be employed. The most commonly used fixative in the UK is formalin, a 4% formaldehyde solution [40], but there are about 2000 known fixatives and fixatives mixtures [38].

Various types of fixatives and their effects on tissue samples are listed in Table 1.1.

Type of fixative	Examples	Other effects on tissues
Aldehydes	Formaldehyde Glutaraldehyde	Shrinkage
Early fixatives	Ethanol Acetic acid	Dehydration
Oxidising agents	Potassium dichromate Potassium permanganate Osmium tetroxide	Enzyme inhibition Proteins breakdown
Cold	Freezing to -200°C	Dehydration
Heat	Microwave Boiling	Enzyme inhibition Release of gases
Others	Picric acid Mercuric chloride Carbodiimides	Various

Table 1.1. Different types of fixatives and their effects on the chemistry of tissues. In general, all the fixatives inhibit enzymes, denature proteins, and change the volume of tissues. Adapted from [38].

1.5.1.2 Formaldehyde

Formaldehyde is a pungent toxic gas, and it is soluble in water. In the solution, it is mainly in the hydrated form of methylene glycol, which acts as a fixative. Fixation with formaldehyde is a complex process in which a very rapid penetration stops autolysis. It is then followed by covalent bonding and cross-linking. Methylene bridges are formed between protein molecules,

mainly with the basic amino acid lysine. Lysine residues that react are only those which are on the exterior of the protein molecule [42, 43]. The reaction between formaldehyde and proteins is pH-dependent and more rapid at high pH. Formaldehyde has many advantages as a routine fixative. It causes tissues to swell slightly, so the organs remain soft, which is suitable for dissection and trimming. Tissues cannot be over fixed and can be stored in formalin for a long time. Formaldehyde is also the only fixative that does not alter tissues' colour, but cytoplasmic staining is duller compared with other fixatives. Formalin is cheap and easy to prepare; however, it has several health and safety considerations [40, 44].

Moreover, several studies reported various issues regarding formalin fixation. Abe *et al.* described changes in collagen's cross-links in nucleic acid fragmentation and degradation due to the development of formic acid by and elastin in formalin-fixed tissue [45]. Schulz *et al.* established that three-dimensional strain fields in the human brain result from this type of fixation and are related to the shrinkage of soft tissues [46]. Chen *et al.* reported that formalin fixation might cause shrinkage of head and neck tumours, leading to underestimating tumour staging [47].

1.5.1.3 Alcohols

These are non-additive fixatives that denature nucleic acids and proteins by removing their bound water so it can be replaced with alcohol. Only the shape of molecules is altered, not the reactive groups. Therefore, the chemical reactivity of many cellular materials is preserved what makes alcohols fixation popular in histochemical investigations. However, the tissue shrinks considerably and may become brittle and overhard if this type of fixation is used. Due to this fact, alcohol is commonly used for fixing fresh sections or smears rather than fixing blocks of tissue [40]. Numerous studies revealed that alcohol-based fixatives could offer superior morphology and RNA quality [48, 49]. Gillespie *et al.* reported that fixation in 70% ethanol provided superior DNA quality to that recovered from formalin-fixed tissue [50]. This approach allowed to perform techniques that require relatively large DNA fragments. There was a significant increase in the number of successful PCR amplification that could be performed per cell number [50].

1.5.1.4 Limitations of fixation

Although fixation is fundamental in histopathology investigations, it is crucial to realize that it may cause artefacts. Good examples are changes in the volume of tissue or formalin pigment artefact produced under acid conditions. Another group of artefacts is related to unfixed

material diffusion to give false localization by resting in someplace different from its original location, e.g. wrong location occurring with glycogen [41]. Materials may also diffuse out of the tissue, which is likely to happen with small molecules during fixation. The reverse of this situation is also possible and leads to the false fixation of extraneous material to the tissue [41]. Moreover, formaldehyde contact with the tissue components leads to the initial formation of highly reactive hydroxy-methyl groups, resulting in methylene bridges between amino groups of proteins. The outcome of this process is the intramolecular loss of antigen availability for the paratope to bind, referred to as antigen masking [51-53].

Fixation is an essential step in the histopathology process. It is the foundation for all the subsequent stages in preparing the sections (processing, embedding, cutting sections, performing different stains) to make a diagnosis. However, it is essential to remember that no fixative can preserve any tissue precisely the same way it was in life. There is no single fixative that can be suitable for all preparation.

1.5.2 Dehydration

Acetone, ethanol, and dioxane are used for dehydration. The first two solvents extract lipids; therefore, most likely, not many of them would be left after the exposure of relatively small pieces of tissue to large volumes of reagents. Since Singer and Nicholson's publication in 1972 [54], it has been accepted that cell membranes are 'fluid' and that the fluids are mainly composed of water and lipids. Therefore, the dehydration step is likely to alter membrane dynamics significantly. Additionally, each phase of the cells and the extracellular fluid contain various amounts of water. Dehydration alters all the constituents' relative concentrations in each of the phases, too [38].

1.5.3 Clearing

Reagents like xylene, benzene, toluene, chloroform, trichloroethylene, clove and cedar oil can be used for clearing. The first three are also used for infiltration. These reagents make tissues more transparent by increasing their refractive index. All of these reagents, except for clove and cedar oil as well as ethanol and methanol, are routinely used in biochemical procedures for lipids extractions [38].

1.5.4 Infiltration

As the dehydrating agents are not miscible with the embedding ones (wax, epoxy or acrylic resins), intermediate substances such as xylene, polyethene glycol, 2-hydroxymethyl acrylate are used to infiltrate tissue samples before proceeding to the embedding step.

Infiltration and dehydration at least twice each extract lipids, water, lipid-soluble, and water-soluble constituents [38].

1.5.5 Embedding

Fixed tissue samples need to be embedded to make them hard enough to allow thin tissue sections to be cut. This step involves enclosing adequately processed and correctly oriented specimens in a support medium that supports microtomy. The embedding medium must fill the matrix within the tissue, providing support for cellular components.

Most histopathology laboratories use modular embedding centres, consisting of a cold plate, a paraffin dispenser, and a heated storage area for moulds and tissue cassettes. Melted paraffin wax is dispensed automatically into a suitably sized mould, but a qualified staff member does the embedding itself [55].

1.5.6 Cutting tissue sections

Microtomy is how an embedded tissue specimen is sectioned and attached to a glass slide surface for further microscopic examination. The primary instrument used to cut FFPE blocks is a microtome. An advancing mechanism moves the block for a predetermined distance until it is in contact with a blade or a knife. The specimen moves vertically past the cutting tool, and a tissue section is produced. Routine surgical material is cut at 3-4μm, while sections for the IHC staining are usually cut at 1.5μm [55].

1.5.7 Floating and mounting tissue sections

The thin, cut tissue sections are floated in the water bath and mounted on glass slides. The sections need to stick to the slides firmly so they do not come off when subjected to other reagents later on. After the tissue sections have been mounted, they usually do not shrink [38].

1.5.8 Rehydration

As most staining solutions are aqueous, tissue sections need to be rehydrated before proceeding to that step. In order to stain the sections, the wax has to be dissolved and replaced with water. It is achieved by passing the tissue sections through xylene, decreasing the strengths of alcohol

(100% to 0%) and finally water. Essentially, this is the reverse of the dehydration process. As a result, any substances soluble in water or ethanol present in the sections will be removed [38, 55].

1.5.9 Staining

There are about 2500 stains and staining procedures available. They can be classified in many different ways, e.g. the colour, whether they are basophil or acidophil, the types of cells, organelles or granules they stain, or the tissue types they are thought to be specific [38].

The most widely used histological stain is the hematoxylin and eosin (H&E) stain. Its popularity is based on its simplicity and ability to demonstrate an impressive number of different tissue structures. The hematoxylin component stains the cell nuclei blue-black and shows good intracellular detail. On the other hand, the eosin stains cell cytoplasm and most connective tissue fibres in varying intensities and shades of red, pink, and orange [55].

Special stains remain essential tools for pathologists, providing a powerful complement to flow cytometry, IHC, in situ hybridization and other diagnostic technologies that define patients' medical profiles. Many special stains play an essential role in diagnosing and monitoring cancer, while others are used to detect and identify pathogens [55].

IHC is a technique used to identify cellular or tissue antigens utilizing antigen-antibody interactions. The antibody binding site is identified using a secondary labelling method or by direct labelling of the antibody. Many antibodies are available these days to identify epitopes that survive formalin fixation and processing paraffin wax. IHC is invaluable in all those cases where morphology combined with clinical data do not allow a firm diagnosis. The increasing use of predictive and prognostic markers allows pathologists to make decisions that could significantly affect patients' management [55].

All the above procedures need to be fulfilled before the final histopathology report can be issued. Some steps of the process are automated, but most of them are still performed by hand by staff members.

1.5.10 Limitations of histology and the need for new alternative techniques

As the result of the increasing complexity in pathology workup, it has been estimated that the number of tissue blocks per patient and the number of required spare slides per block has increased by more than 60% over the last decade [56]. However, this enormous increase in workup is not the only issue that pathology services are facing. Despite being considered a gold

standard in making a diagnosis, histopathological investigations can be time-consuming. Depending on the required stains, it can take days before the final pathology report is issued. Additionally, an examination of the stained slides is subjected to intra-observer error [36, 37]. Therefore, it is clear that alternative methods are needed for accurate and timely diagnosis and tumour grade and stage characterisation. Further research and new modalities involving automated diagnosis are required to overcome challenges to lessen the traditional pathology-based tissue diagnosis burden. Currently, mass spectrometry imaging (MSI) is extensively applied to *in situ* molecular analysis of tissues to classify primary tumour tissues, identify tumour margins as well as prognostic and diagnostic markers, and analyse the rates and resistance of drug responses. Elucidation of various molecules' abundance and spatial distribution with malignant cells and tissues can significantly contribute to the staging, diagnosis, and treatment of various diseases [57].

1.6 Mass spectrometry – general principles

Mass spectrometry (MS) is described as a microanalytical technique that can be selectively applied to detect and determine a given analyte amount. MS can also be used to determine the elemental composition and reveal some aspects of an analyte's molecular structure. This is accomplished by the experimental analysis of the mass of gas-phase ions produced from an analyte's molecules [58]. Advantages of this approach include its capacity to produce and detect fragments of the molecule corresponding to discrete groups of different elements' atoms that reveal structural features and the possibility of a direct determination of the nominal mass of an analyte. Importantly, this microanalytical technique can generate more structural information per unit quantity of an analyte that can be achieved by any other analytical approach [58].

Before ions can be separated according to their *m/z* and detected, they must be in the gas phase. Before the 1970s, only analytes characterized by a significant vapour pressure were amendable to MS. This was because techniques like chemical ionization (CI) or electron ionization (EI) could be used to produce gas-phase ions only from gas-phase molecules. Thermally labile and non-volatile molecules could not be subjected to those gas-phase ionization techniques. Nowadays, both CI and EI play a significant role in the combined methods of liquid chromatography/mass spectrometry (LC/MS) and gas chromatography/mass spectrometry (GC/MS) [58].

After the 1970s, there was a further development in the MS capabilities. Desorption/ionization (D/I) techniques were invented, allowing for the analysis of gas-phase ions directly from a sample in the condensed phase. Fast atom bombardment (FAB), which required analyte's nanomoles to produce an interpretable mass spectrum, was the first widely accepted technique for D/I. In the 1980s, techniques like matrix-assisted laser desorption/ionization (MALDI) and Electrospray ionization (ESI) became more widely applied than FAB. One reason contributing to their increasing popularity was that they only required analyte's picomoles for analysis [58]. Nowadays, MS plays a significant role in the biological sciences. MALDI and ESI have hugely contributed to that as they can be applied to analyse femtomole quantities of non-volatile and thermally labile analytes [58].

However, MS applications are not limited to only measurements of organic molecules. This approach is suitable for the analysis of any element as long as it can be ionized. MS, for example, can be applied to study silicon wafers to establish the presence of iron and lead (any of them can cause microprocessors' semiconductor failure). Moreover, MS is widely applied to material sciences and geology studies. Each of these two disciplines has developed unique analytical capabilities: secondary ion mass spectrometry (SIMS) in material sciences and isotope ratio mass spectrometry (IRMS) in geology [58].

1.6.1 The concept of MS

Since ions are charged particles, the use of magnetic and electric fields can manipulate their position in space. When only individual ions are present, their unique properties, such as mass and the number of charges, can be used to group and move them from one point to another. Therefore, ions need to be analysed in a vacuum, as that way, the individual ions are free from any other forms of matter. As a result, ions must be in the gas phase.

MS analysis relies on ions in the gas phase and at a low pressure to separate and detect them based on their mass-to-charge ratio (*m/z*). That means that each ion's mass on the atomic scale is divided by the number of charges that the ion possesses [58]. Dimensionless numbers represent *m/z* values. A mass spectrum is a recording of the number (abundance) of ions of a given *m/z* value as a function of the m/z value. MS only detects ions; all other particles are removed from the instrument by a continuous pumping that maintains the vacuum [58].

MS consists of three major components: the ionisation source, mass analyser and detector.

The ionization sources used in MSI are described in *Chapter 1.7*; only the last two components will be discussed here.

1.6.2 Mass analyser

Once the ions have been produced in the gas phase, they are separated according to their *m/z* ratio by the mass analysers. Mass analysers are designed to utilise static or dynamic, magnetic or electric fields, either individually or as a combination of both. Quadrupoles, Time-of-Flight (TOF) instruments, ion traps, Fourier transform ion cyclotron resonance (FT-ICR), and Fourier transform Orbitrap (FT-Orbitrap) instruments are the most commonly used mass analysers for the analysis of lipids and metabolites. Each type of mass analyser has its unique characteristics, summarised in Table 1.2 [58].

1.6.2.1 Quadrupole ion traps

This type of mass spectrometer was commercially introduced in the mid-1980s; however, the first use of quadrupole electric fields to manipulate ions was reported in the early 1950s [59]. In this type of mass spectrometer, an electric field is created between four opposing electrical poles. The electric field's shape is described as a function of the geometric arrangement of these four surfaces. The primary component of the electric field surrounding the ions is based on radio frequency potentials which are applied to two pairs of opposing electrodes [58]. The transmission quadrupole instrument relies on a two-dimensional electric field that pulls and pushes ions in the x- and y-directions while travelling along the z-axis. This allows to filter out ions of all *m/z* values except for those of interest. The quadrupole ion trap is different from all other types of mass spectrometers because it operates at a relatively high pressure of ~10^{-1} Pa instead of 10^{-7} Pa for *re*TOF and 10^{-4} for the transmission quadrupole [58]. The quadrupole analysers are one of the most common mass analysers.

1.6.2.2 The Orbitrap

Alexandor Makarov developed this device, and it is the latest development in trapping devices used as an *m/z* analyzer [60]. However, it is helpful to think about the orbitrap as a modified Knight-style Kingdon trap with specially shaped outer coaxial (a barrel) and inner axial (a spindle) electrodes. While the quadrupole ion trap uses a dynamic electric field oscillating at ~1 MHz, the orbitrap relies on a static electrostatic field to sustain ion trapping following the specialized dynamic injection pulse [60]. Two electrodes in the form of coaxial axisymmetric electrodes, an inner spindle-shaped electrode oriented, and an outer barrel-shaped surface form the orbitrap. A constant electric potential is imposed between these two axisymmetric electrodes. No oscillating electric potentials or magnetic field are involved. The opposing surfaces of the axisymmetric coaxial electrodes are nonparallel. Therefore, the electric field

between the two surfaces varies and reaches a minimum at the point of greatest separation of the electrode surfaces (the centre of the orbitrap). Advantages of the orbitrap mass spectrometer include resolving power of 70 000 +, comparable performance to FT-ICR without the need for cryogen, fell get an advantage as in FT-ICR. On the other hand, the main disadvantages are inefficient trapping of the product ions, low pressure is necessary as a mean-free path of ~100 km is required, and high cost compared to 3D and linear QIT [60].

1.6.2.3 TOF *m/z* analysers

The operating principle of this type of mass spectrometer relies on measuring the time required for an ion to travel from an ion source to a detector. The detector is usually 1 to 2 meters away from the source. Therefore, these mass spectrometers require a high vacuum. The same kinetic energy is applied to all the ions during instantaneous acceleration, but since the ions have various *m/z* values, they also have correspondingly different velocities. When the ions traverse the 'field-free' area between the ion source and the detector, they are separated into groups or packets based on velocity [61-63].

1.6.2.4 FT-ICR and FT-Orbitrap

The highest mass resolution characterizes these analysers. In FT-ICR, ions are trapped in a chamber by the Lorentzian force originating from a high strength magnetic field. The ions are orbiting at an angular frequency known as ion cyclotron frequency which is inversely proportional to the *m/z* ratio [58]. FT-Orbitrap relies on a similar principle to FT-ICR, except that an electric field replaces the magnetic field. This mass spectrometer consists of two electrodes, an inner spindle-shaped and an outer barrel-shaped surface electrode. Constant electric potential is applied between these two electrodes for ion trapping. Ion packets of different *m/z* ratios execute their axial oscillations at their respective frequencies [58].

Type of analyser	Principle of separation	Mass range limit	Mass resolution	Accuracy	Pressure
Quadrupole (Q)	*m/z* (trajectory stability)	2000 Th	2000	100 ppm	10^{-5} Torr
Time-of-Flight (ToF)	Velocity (flight time)	10 000 Th	20 000	10 ppm	10^{-6} Torr
Fourier transform ion cyclotron resonance (FT-ICR)	*m/z* (resonance frequency)	30 000 Th	>1 000 000	1 ppm	10^{-10} Torr
Fourier transform Orbitrap (FT-Orbitrap)	*m/z* (resonance frequency)	50 000 Th	1 000 000	<5 ppm	10^{-10} Torr

Table 1.2. Characteristics of most often used types of mass analysers [58].

1.6.3 Ion Detector

Once the ions have been filtered through the mass analyser, the detector is responsible for recording each *m/z* value's abundance. There are many types of detectors, but most work by producing an electronic signal based on the charge or current induced when an ion strikes the surface. This is then converted into a graph demonstrating the relative current produced by ions of different *m/z* ratios [58]. For example, the Faraday cup is a conventional electrical detector, where the measured ion current is directly proportional to the number of ions and the number of charges per ion. It is an inexpensive, simple, rugged and robust detector. Electrical multiplier detectors, another category of detectors, use the principle of secondary-electron emission to amplify the signal's intensity. In the case of FT-ICR and FT-Orbitrap analysers, the detection is already incorporated into the analyser. The ions are detected by the image current they produce between the metal parts within the mass analyser region [58].

Other types of ion detectors include Negative-ion detection, Channel electron multiplier array, Electro-optical ion detection, the Daly detector, Cryogenic detectors, Ion detection in FTMS, Post-acceleration detection and Detection of high-mass ions [58].

1.7 Mass spectrometry imaging

Over the past decade, MSI was demonstrated to be a tremendous scientific tool that can unambiguously detect multiple analytes from complex biological samples during a single analysis and provide information about the spatial distribution of the detected analytes. MSI is a novel method that has proven to be an invaluable tool for localising peptides, metabolites, drugs, and proteins in biological tissues without any prior labelling [64]. Probing the tissue surface with an ionization beam is involved in a typical MSI experiment. It aims to acquire thousands of mass spectra at defined x, y coordinates from specific locations throughout the sample surface. As a result, each sampling location corresponds to a unique mass spectrum. Reconstruction of an image from the collective dataset in which each pixel consists of a mass spectrum is achieved by employing dedicated software. Therefore, mapping the distribution of individual compounds is possible. An alternative to MSI is IHC, which is widely used to study the localisation and distribution of biomarkers in cancer cells in cancer tumour cells' diagnosis. Whilst this provides a sensitive and selective technique, it relies on the availability of the antibodies specific to the protein of interest, resulting in low throughput characterisation [65]. MSI has been shown to provide an accurate technique for analysing metabolic changes within different tissue regions and can provide information on anticancer drug treatments' efficacy. Furthermore, MSI can impact the prognostic and therapeutic stages of cancer phenotyping and provide a platform for identifying and developing novel biomarkers [66, 67].

Multiple ionisation techniques can be used in MSI, and each has its strength and weakness based on its uses, spatial resolution and sample preparation requirements. The three most commonly used are SIMS, MALDI, and DESI [68]. Their characteristics are summarized in Figure 1.4.

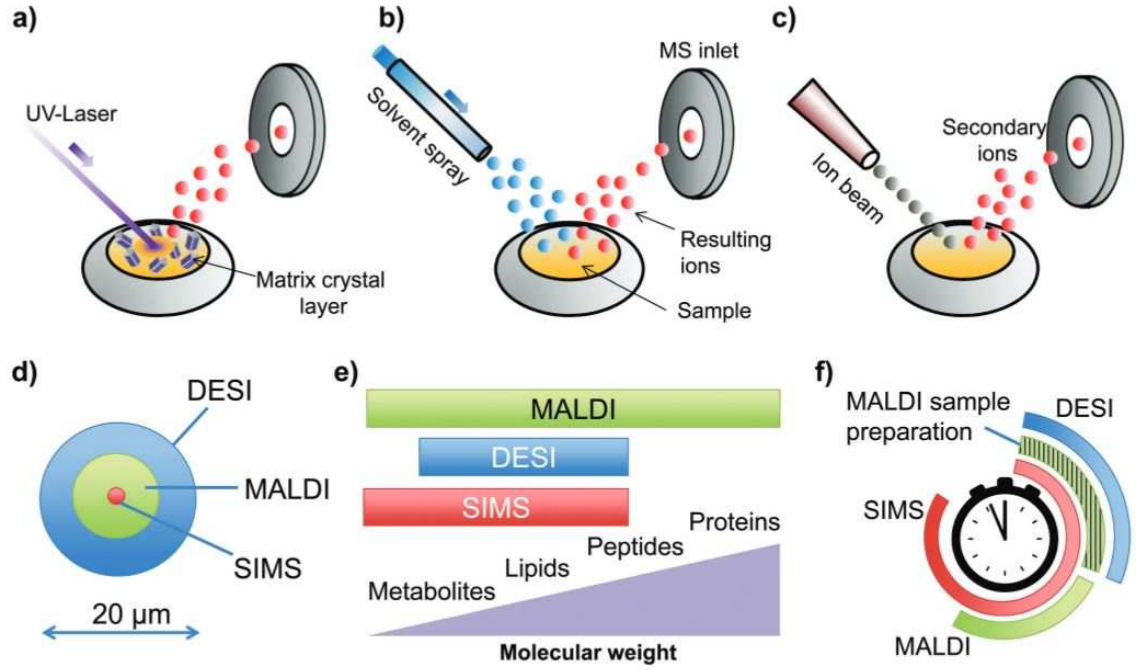


Figure 1.4. Comparison of the three commonly used MSI technologies. MALDI requires an application of an organic matrix across the sample (A). In DESI, a solvent spray is directed towards the sample for desorption (B). SIMS relies on pulsed ion beams to locally desorb compounds (C). SIMS offers the highest spatial resolution, followed by MALDI and DESI (D). MALDI offers the highest versatility in the mass range and hence the type of investigable molecular classes (E). Contrary to MALDI, one of the benefits of DESI is reasonably straightforward and quick sample preparation. As a result, the total time needed to perform DESI-MSI analysis is relatively short. (F). Reproduced with permission [69].

1.7.1 SIMS

SIMS utilizes beams of primary monoatomic or polyatomic ions in a vacuum for ionization [70]. During analysis, the surface of the sample is bombarded with a beam of primary ions, resulting in the release of the molecules from the analysed sample's surface. This process is described as sputtering, and during it, energy from the primary ions is transferred to the analyte molecules [57]. It has been estimated that around 1% of the sputtered analyte molecules has an electric charge, and these are the ions that a mass analyser can later detect. Compared to the covalent bond energies of the analyte molecules, the energy of the primary ion beam is usually high. This causes the fragmentation of the molecules from the analyte. Because of this property, SIMS is referred to as a 'hard' ionization technique [71]. No matrix is required for SIMS experiments. This technique is characterised by higher depth and spatial resolution (<10 µm) compared with MALDI. However, MALDI, on the other hand, is more sensitive. The highest detectable mass range in the case of SIMS is around 1kDa [70].

1.7.2 MALDI

MALDI was developed in the late 1980s. The routine techniques for this ionisation method were developed and described by Karas and Hillenkamp *et al.* [72-75].

In this technique, the sample is mixed with an organic matrix compound, e.g. dihydroxybenzoic acid. The matrix plays multiple roles, but the key one is absorbing the radiation and protecting the analyte from radiation damage. Another purpose of the matrix is the separation by diluting the analyte molecules to prevent analyte-analyte molecular interactions during the analysis. The matrix's essential features that need to be considered when selecting one include its capacity to absorb the laser energy and solubility characteristics similar to the analyte [72, 73]. The matrix's two desirable attributes are as follows: it must be capable of forming a fine crystalline solid during co-deposition with the analyte, and its molecules must be characterized by a high absorptivity for the laser radiation. Most likely, the sample preparation is the most time-consuming part of the whole MALDI analysis [58].

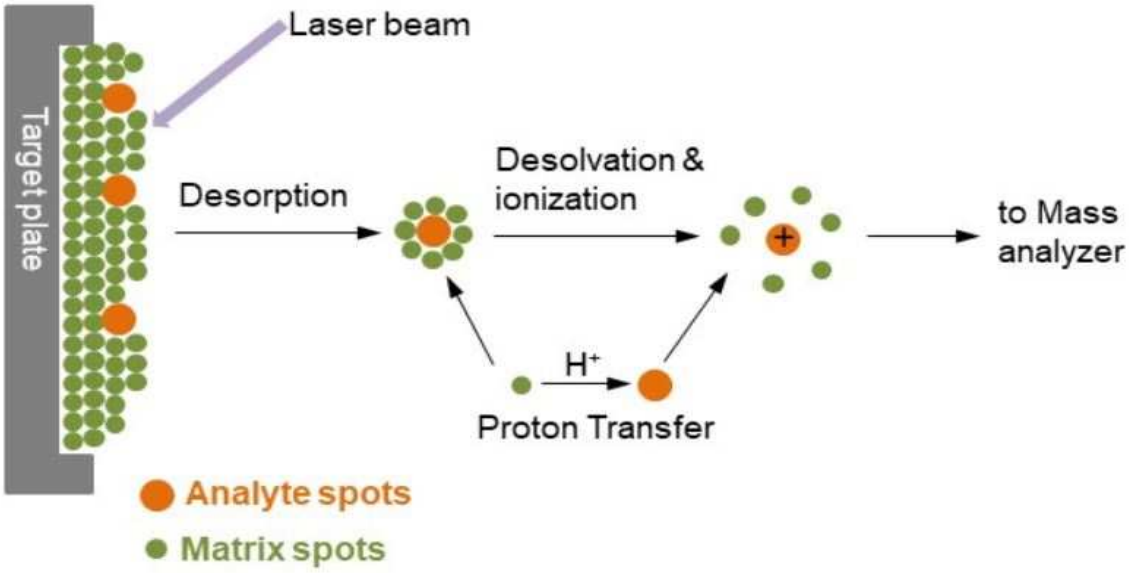


Figure 1.5. Schematic representation of the ionisation process in MALDI. The analyte is mixed with a matrix which is then deposited onto a target plate. The mixture is placed in a high vacuum source region and is irradiated with a short-pulsed laser beam. The excitation of the matrix by the high-intensity laser pulse causes simultaneous desorption and ionisation of the sample resulting in the formation of predominantly singly charged ions.

For the analysis, a sample is placed on a conductive surface. Then, depending on the type of analysis, a digestion step might be needed. The analysis of metabolites, lipids, intact proteins and endogenous peptides performed directly from tissues does not require enzyme treatment

[76]. Before analysis, the matrix is suspended in an acidified solvent buffer and then deposited over the sample. Incorporating the matrix into the analyte molecules results in the formation of crystals [77]. Analyte ionization is induced by the application of a focused laser beam to the sample surface. The ions are accelerated through a strong electric field in the ion source. The charged analyte molecules hit the detector, and detailed information about the number of events is recorded within a time period. This is the source of information about the intensity. The detected analyte ion's mass can be calculated by comparing the measured information to calibration standards [57]. A schematic representation of the ionisation process used in this technique is presented in Figure 1.5.

In general, MALDI is described as a soft ionization technique, meaning there is little, if any, fragmentation of the protonated molecule created during the analysis process. The spatial resolution is around 20 μm, and this technique can accommodate a molecular mass range of over 100 kDa [57].

1.7.3 DESI

DESI was the first described ambient MS method, which was implemented by directing a pneumatically assisted solvent electrospray onto the surface of interest [78].

The impact of multiply charged droplets of solvent on the analysed surface results in the formation of secondary charged molecules. The ions released from the surface are transported through the air at atmospheric pressure for some distance before reaching the atmospheric interface of the mass spectrometer [78]. However, the sample needs to be either isolated from the ground or be non-conductive [79]. These secondary droplets contain molecular species that are dissolved on the surface and then due to solvent evaporation from multiply charged droplets. The molecular species are converted into gaseous ions [80].

Figure 1.6 shows a schematic representation of the ionisation process in DESI.

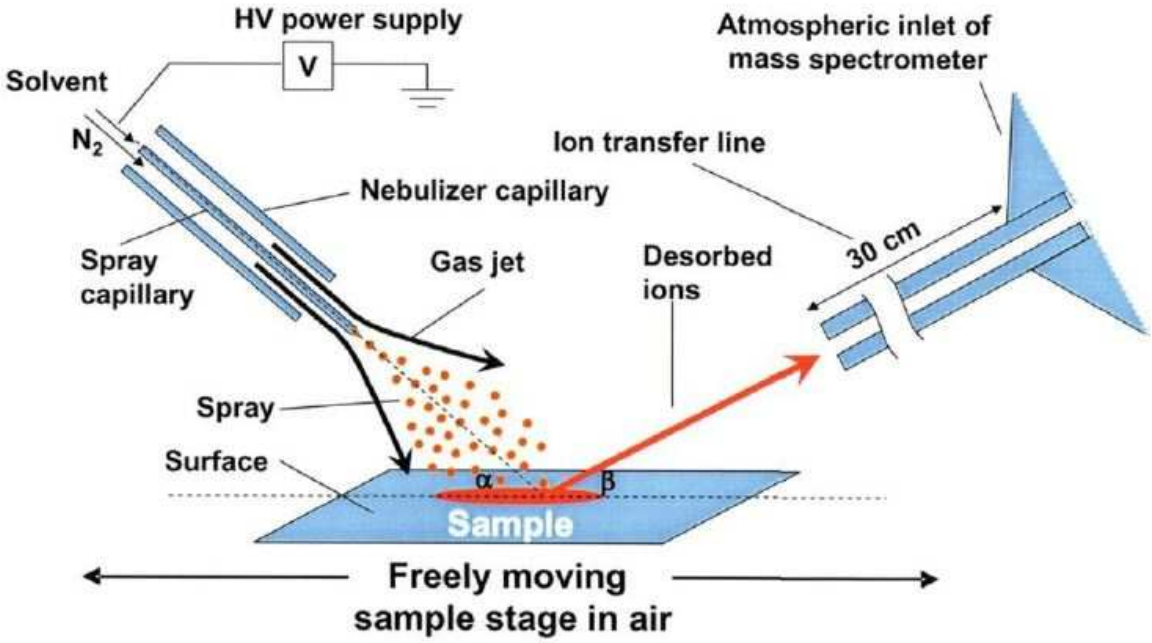


Figure 1.6. Schematic representation of the ionisation process in DESI. A pneumatically assisted electrospray is directed at the sample surface. The molecules are desorbed from the sample surface and are protonated or deprotonated and then transferred to the mass spectrometer through a capillary. Reproduced with permission [78].

A key feature of DESI is that it allows keeping the sample outside of the instrument. Only ions are allowed to enter the *m/z* analyser via a capillary vacuum sweeper [81]. Contrary to SIMS, the application of this technique causes no damage to metallic or polymeric surfaces. However, physical ablation of soft sample material might occur [82]. Another advantage of this method is that no matrix is needed, and sample preparation is straightforward [79].
DESI is considered as a 'soft' ionization technique, as little analyte fragmentation occurs during ionization. It is characterized by a spatial resolution of around 50-100 μm and an upper mass range detection limit of around 2 kDa [57].

Following its introduction in 2004 [78], the main application area of DESI has been cancer research focusing on histological and drug distribution applications. It has been tested to replace frozen section histology for margin control in cancer resection surgery [83-85]. Several studies focusing on different cancer types have also been published. Eberlin *et al.* used DESI to study prostate cancer [86] and reported novel tissue cancer biomarkers. Dill *et al.* applied this technique to study human bladder cancer [87] using supervised multivariate analysis-driven pixel-wise classification for DESI data. Afterwards, similar data treatment for projects

targeted on the DESI analysis was applied in multiple studies focusing on various human cancers, including colorectal adenocarcinoma [88], seminoma [89], and gliomas [90].
Studies focusing on the lipid composition in human cancers analysed by DESI included breast cancer [91], ovarian cancer [92], and brain tumours [90, 93], to list just a few.

1.7.4 DESI vs MALDI – selection of the technique

DESI is a common choice for drug, metabolite and lipid imaging, while MALDI is the most widespread high vacuum imaging technique for protein, peptide and lipids [94]. Therefore, projects focusing on alternations in lipid metabolism could use any of these techniques [93, 95-98]. As an untargeted approach was used for all fresh-frozen samples analysed during this PhD, and all of them were analysed at 100µm x 100µm pixel size, DESI was a sensible choice. MALDI, on the other hand, could be applied for experiments requiring high-resolution data.
The FFPE project could also use any of these two techniques. FFPE specimens undergo excessive processing before being embedded in wax, and formalin fixation is known to cause cross-linking of proteins [99, 100]. Therefore, the project described in *Chapter 5* aimed to instead focus on small metabolites and lipids mass range. Moreover, no results have been published for FFPE samples analysed by DESI-MSI, contrary to the MALDI approach [101, 102]. It was interesting to see if DESI-MSI could be applied for tissue classification in FFPE specimens.
Other practical considerations were also taken into account. DESI does not require the use of the matrix. Therefore, sample preparation is very straightforward. Moreover, it is a non-destructive technique, allowing for a histopathology assessment of the same tissue section that was used for imaging.

1.8 Data analysis strategies for the MSI data

MSI studies generate profiles containing signals corresponding to hundreds of biomolecules originating from the analysed tissue samples. The ultimate aim of these studies is to find *m/z* features of biological significance specific to a spatial region, e.g. those that are discriminative for a tumour region compared to a healthy region. The data analysis strategy plays an essential role in the whole process as it allows the correlation of biochemical and histological features. Different chemo-informatics strategies have been developed to achieve this aim, based on three

main steps: 1) pre-processing of raw imaging data, 2) unsupervised analysis as a preliminary step for data examination and 3) supervised classification.

1.8.1 Pre-processing raw data

Typically, the following pre-processing steps are involved:

- normalisation
- peak picking
- peak matching/spectra recalibration.

Incorporating these steps allows making the MSI data representing different samples comparable and manageable while simultaneously accounting for bioanalytical complexities common to all MSI studies.

1.8.1.1 Spectral normalisation

The normalisation, defined as scaling each spectrum to a factor (usually dataset-specific), is critical for pre-processing steps. It allows better inter-comparison of intensities between different spectra. Total ion count (TIC) normalisation, the most often recommended approach, is the most popular method [103]. It relies on the fact that the sum of all intensities (the TIC value) is calculated, and then all *m/z* intensities are divided by it. However, this method's disadvantage is that a single outlier m/z peak's intensity can significantly influence it. Another approach – median normalisation – has been suggested as an alternative to TIC. It has also been demonstrated to be the least likely to be compromised by biological variability [103].

1.8.1.2 Peak picking

Peak picking refers to selecting m/z values corresponding to high and relevant peaks and is a crucial step in pre-processing. It aims to reduce the number of *m/z* values by filtering out values corresponding to noise-related signals. When comparing lists of relevant peaks from different datasets, *m/z* values selected for the same peak but in different spectra can vary slightly. As a result, peak alignment is fundamental to correct these small *m/z* shifts, as instrument calibration cannot be used to counterbalance this issue [104]. Multiple peak alignment algorithms are based on hierarchical clustering [105] and an average spectrum to find the common peaks [106].

1.8.2 Multivariate statistical analysis

Multivariate statistical analysis can be applied to establish correlations in high dimensional datasets. In the MSI data analysis context, by treating each pixel as an independent sample, multivariate techniques can be used to visualise their output as score plots in the spatial domain. Multivariate techniques most often used for imaging datasets can be divided into two different groups: unsupervised and supervised.

1.8.2.1 Unsupervised approach

Principal component analysis (PCA) is the most commonly used unsupervised technique in analysing imaging data. It can be applied to spectra from specific regions of interest and the entire tissue section without prior knowledge. This unsupervised approach is often used to briefly inspect the data before implementing more advanced data analysis algorithms. PCA is described as a dimensionality reduction technique, which involves reducing the number of variables with minimal loss of information to reveal the data set's general distribution. This is achieved by establishing relationships between samples (pixels) and variables (*m/z* features) by finding the linear combination of variables with the highest variance. Linear orthogonal transformation is used to help to maximise the variance resulting in an ordered set of principal components, where the first principal component (PC1) contains the maximum variance in the data, followed by PC2, which is orthogonal to PC1 and so on [107]. In the majority of cases, the first three principal components usually explain most of the variation. The PCA results are usually reported in the form of loadings (represents the contribution of *m/z* peaks to the principal components), score plots (coordinates of pixels in the transformed axes), and variance (from each principal component) [107].

k-means clustering is another unsupervised analysis technique that incorporates the data's spatial component into a clustering analysis. This approach is a powerful method for a quick assessment of the spatial-chemical organisation of MSI datasets. It involves segmenting the imaging dataset into a pre-defined number, (k), of clusters, generally based on the Euclidian distances between the mass spectra [108]. The clustering algorithm is initialised by the selection of k-locations in the data space (k-means). Each pixel's mass spectrum is assigned to the cluster with the closest mean. Next, each cluster's centroid is defined as the new mean, and the whole process is repeated until convergence is achieved. Notably, the results are highly dependent on the number of clusters and the radius of the pixel groups clustered together [108].

Hierarchical clustering analysis (HCA) is an unsupervised approach where the pixels that share similar MS profiles (similar *m/z peaks* with similar intensity) are grouped first. Then pixels with different MS profiles are separated. The data is presented in the form of a dendrogram or a hierarchical tree using a 'linkage' calculation and a similarity metric (often Euclidian distance in imaging). Hierarchical clustering is extensively used for gene expression analysis [109], but it has also been applied to MSI-based cancer studies [110]. The advantage of hierarchical clustering is the ability to investigate subgroups within the data interactively. However, it is best suited for comparatively small numbers of 49 individual data points. Otherwise, visualising large datasets gets challenging quickly and causes the removal of the interactive nature by overcrowding the dendrogram [109].

Other unsupervised strategies, such as independent component analysis, probabilistic latent semantic analysis, and non-negative matrix factorisation, have also been reported for analysing MSI datasets [107].

1.8.2.2 Supervised approach

Discriminant analysis (DA) is the most widely used supervised strategy that considers the data's group membership. In linear discriminant analysis (LDA), PCA is usually employed first to reduce the data's dimensionality. Afterwards, LDA is applied, and it finds linear combinations in *m/z* features that best explain the group membership. This is achieved by maximising inter-class variance (e.g. cancer versus normal) while simultaneously reducing intra-class variability. The resulting discriminant function can be applied for classifying samples or patients by building a test set and a training set. The number of correct classifications then determines the quality of the model. However, employing PCA before DA is based on the assumption that the variance between different groups is a significant source of the data variance [111].
Additionally, LDA cannot be directly applied when the number of variables exceeds the number of samples. This is a prevalent situation in MSI datasets [111, 112]. On the other hand, partial least squares (PLS), which combines PCA and multiple regression features, is described as a regression technique. PLS extracts latent variables as linear combinations of the original explanatory variables so that most of their association with the response variable is explained [113].

Maximum margin criterion discriminant analysis (MMC-DA) is another example of a supervised multivariate technique that has been adapted for lipidomics-based MSI studies as

recursive maximum margin criterion (RMMC-DA) [114]. The RMMC feature extractor can effectively be applied to extract the most discriminatory features by maximising the variance between and within classes after dimensionality reduction. Moreover, unlike LDA, it does not suffer from the small sample size problem. Its feature extracting capacity does not depend on the non-singularity of the within-class scatter matrix [114]. Class-specific *m*/*z* signatures are derived based on weighted combinations of *m*/*z* feature patterns. The average margin between classes is increased by maximising the distance between their most outer data points, thus improving class separation. By avoiding selecting an optimal number of principal components before applying LDA, the RMMC-DA approach also avoids the model over-or under-fitting [114].

Many other multivariate algorithms, such as those incorporating random forest, neuronal networks, and support vector machines, to name just a few, have been used to study MSI data. The main feature shared by all of these strategies is building a statistical model to separate the training set, which can be used to classify the unknown [115].

1.8.3 Univariate analysis

In MSI, univariate tests are performed to compare spectral profiles from different cohorts (e.g. tumour versus normal) peak by the peak to identify potential biomarkers. In most MSI-based cancer studies, the mean spectra of histologically defined regions for each patient are extracted and used for univariate analysis [116]. Statistical tests, such as analysis of variance (ANOVA) and Student's t-test, assess the likelihood that the sample groups subjected to investigation are different. Univariate statistical tests are individually applied to each MS signal/feature. The t-test might be the most commonly known statistical test, but it is often not applicable to clinical data [107]. For multi-class systems, ANOVA is more suitable. A p-value, which is a measure of the probability that the observed result is caused by chance, results from such tests. A p-value below 0.05 is considered statistically significant (corresponding to a false positive rate of 5%) [107].

Hypothesis

Pathology is involved in 70% of all diagnoses made in the NHS. Examination of stained tissue sections is considered a gold standard for establishing cancer diagnosis and patient management. However, the whole process can be time-consuming, and it is not free from the subjective evaluation of tissue specimens. As a result, some alternative techniques could be implemented to accompany the whole process.

DESI-MSI technique has the potential to deliver a substantial shift in current histopathology services. It supplements traditional morphological analysis with objective lipidomic profiling. The spatially resolved nature of MSI imaging combined with multivariate statistical tools holds the potential for more accurate and effective cancer diagnosis and prognosis.

Therefore, the project's overall hypothesis is that the DESI-MSI protocol can be further tested and adjusted to make it more compatible with a standard histology workflow.

Aims and objectives of the project

There is a clear need for a system that can provide pathologists with a tissue diagnosis based on biochemical features to accompany the traditional morphology-based approach. DESI-MSI has been proven to be able to generate specific metabolomic profiles to allow tissue classification. This tissue identification system may serve pathologists as a complementary tool to the traditional histopathology procedure, allowing accurate tissue characterisation to guide cancer diagnosis better. This project aims to test if DESI-MSI could be applied in conditions that are routinely used in histopathology laboratories to make it more pathology-friendly.

To achieve this aim, the specific objectives are:

- To assess the impact of Optimal cutting temperature compound (OCT), which is routinely used in histopathology laboratories, on the data acquired by DESI-MSI
- To assess and optimise pre-DESI-MSI protocol to enable successful metabolomic analysis of FFPE tissue samples
- To apply DESI-MSI to an FFPE human colorectal samples, including colorectal tissue microarray (TMA)
- Proof of concept for the application of the DESI-MSI technique for the analysis of fresh-frozen colorectal samples
- Proof of concept for the application of the DESI-MSI methodology for the analysis of cirrhotic liver diseases samples

Chapter 2
Materials and Methods

2.1 Patient recruitment

2.1.1 Fresh-frozen colorectal samples

Single colorectal tissue samples utilised in the subsequent studies were obtained from patients undergoing resectional surgery within Imperial College Healthcare NHS Trust. Full ethical approval was obtained from the institutional review board at Imperial College Healthcare NHS Trust (REC reference number 14/EE-0024). Patients undergoing surgery for the colorectal disease were prospectively recruited before surgical intervention. Suitable patients were recruited from the operating list based on the review of preoperative imaging and histology. Patients were supplied with verbal and written information about the study and subsequently provided written informed consent. All patients over the age of 18 years undergoing surgical resection for benign or malignant colorectal disease who were able and willing to provide informed consent were eligible for inclusion into the study. Patients who had undergone neoadjuvant therapy were included, and this was documented in the corresponding clinical dataset. Additional information included gender, age, medical comorbidities, drug history and primary pathology, including tumour location, histological subtype, grade, stage, presence of EMVI and gene mutations.

2.1.2 Fresh-frozen liver samples

The first project, which involved the use of human liver samples, was looking into the effects of the OCT embedding on the DESI-MSI (*Chapter 3*). The tissue used in those experiments was identified as macroscopically normal liver taken from a patient who underwent resection for colorectal adenocarcinoma metastases to the liver. Written consent for the use of this tissue was taken from the patient prior to surgery. Ethical approval was obtained from the South East London National Research Ethics Committee (Study ID 11/LO/1686).

The second project where human liver samples were used was the DESI-MSI of cirrhotic liver disease (*Chapter 6*). Single liver tissue samples utilised in the subsequent study were obtained from patients undergoing resectional surgery within the University Hospitals Birmingham NHS Foundation Trust. Full ethical approval was obtained from the institutional review board at University Hospitals Birmingham NHS Foundation Trust (REC reference numbers: Immune Regulation: 06/Q2702/61, Cellular Trafficking: 06/Q2708/11). Patients undergoing surgery for liver disease were prospectively recruited before surgical intervention. Suitable patients were

recruited from the operating list based on the review of preoperative imaging and histology. Patients were supplied with verbal and written information about the study and subsequently provided written informed consent. All patients over the age of 18 years undergoing surgical resection for benign or malignant liver disease who were able and willing to provide informed consent were eligible for inclusion into the study. Details like gender, age, medical comorbidities, drug history and primary pathology were documented in the corresponding clinical dataset.

2.2 FFPE tissue samples

2.2.1 FFPE colorectal samples

Some of the samples were part of the same study as the fresh-frozen colorectal samples and were covered by the same ethics (REC reference number 14/EE-0024).

A human colorectal carcinoma FFPE sample was obtained from the Histopathology Department at St. Mary's Hospital, London, UK.

Colorectal Tissue microarray (TMA) slides were commercially available and were purchased from Biomax.us (https://www.biomax.us/).

2.2.2 FFPE mouse liver sample

FFPE mouse liver sample was obtained from the Histopathology Department at St. Mary's Hospital, London, UK.

2.3 Collection and storage of fresh-frozen tissue samples

Fresh (within 3 hours post-surgery) colorectal tissue specimens were collected from the operating theatre and sent to the same hospital's pathology laboratory. A histopathologist assessed the tumour size, and if possible, the samples were transferred on for research. The type of tissue available for the project was mainly dependent on the tumour's size and the extent of healthy tissue in the surgical specimen. For example, following a complete local excision for colorectal cancer, the extent of healthy colon tissue in the sample might have been too small to provide a sample safely. Therefore, it was not always possible to consistently obtain samples

of both tumour and healthy tissue from each patient. After tissue collection, fresh samples were snap-frozen and then stored at -80 °C [117].
The same workflow was in place during the collection of fresh-frozen liver samples.

2.4 Collection and processing of FFPE tissue samples

Tissue specimens harvested during surgical procedures were either transferred to 10% neutral buffered formalin prefilled pots straight away or had to be dissected by a histopathologist first. In order to allow appropriate fixation, samples were kept in this fixative for at least 24 hours. After this time, specimens were subjected to automated processing, which involved dehydration in graded ethanol (70% to 100%) and clearing with xylene. All the steps of this process are listed in Table 2.1.

Stage	Solution name	Time (hours)
1	Formalin	1:00
2	Formalin	1:00
3	70% Alcohol	1:00
4	Alcohol	1:00
5	Alcohol	1:00
6	Alcohol	1:00
7	Alcohol	1:00
8	Xylene	0:45
9	Xylene	0:45
10	Xylene	0:30
11	Paraffin	2:00
12	Paraffin	1:30
13	Paraffin	1:00
14	Paraffin	0:05

Table 2.1. List of stages of automated tissue samples processing. All FFPE samples were subjected to automated processing runs, consisting of many steps during which various solvents were used. The length of each phase is also listed.

Once the automated processing run was completed, tissue samples were embedded in wax. This allowed for the formation of tissue wax blocks that could be mounted on a microtome for section cutting.

2.5 General DESI-MSI workflow

The general workflow for the DESI-MSI sample analysis is shown in Figure 2.1.

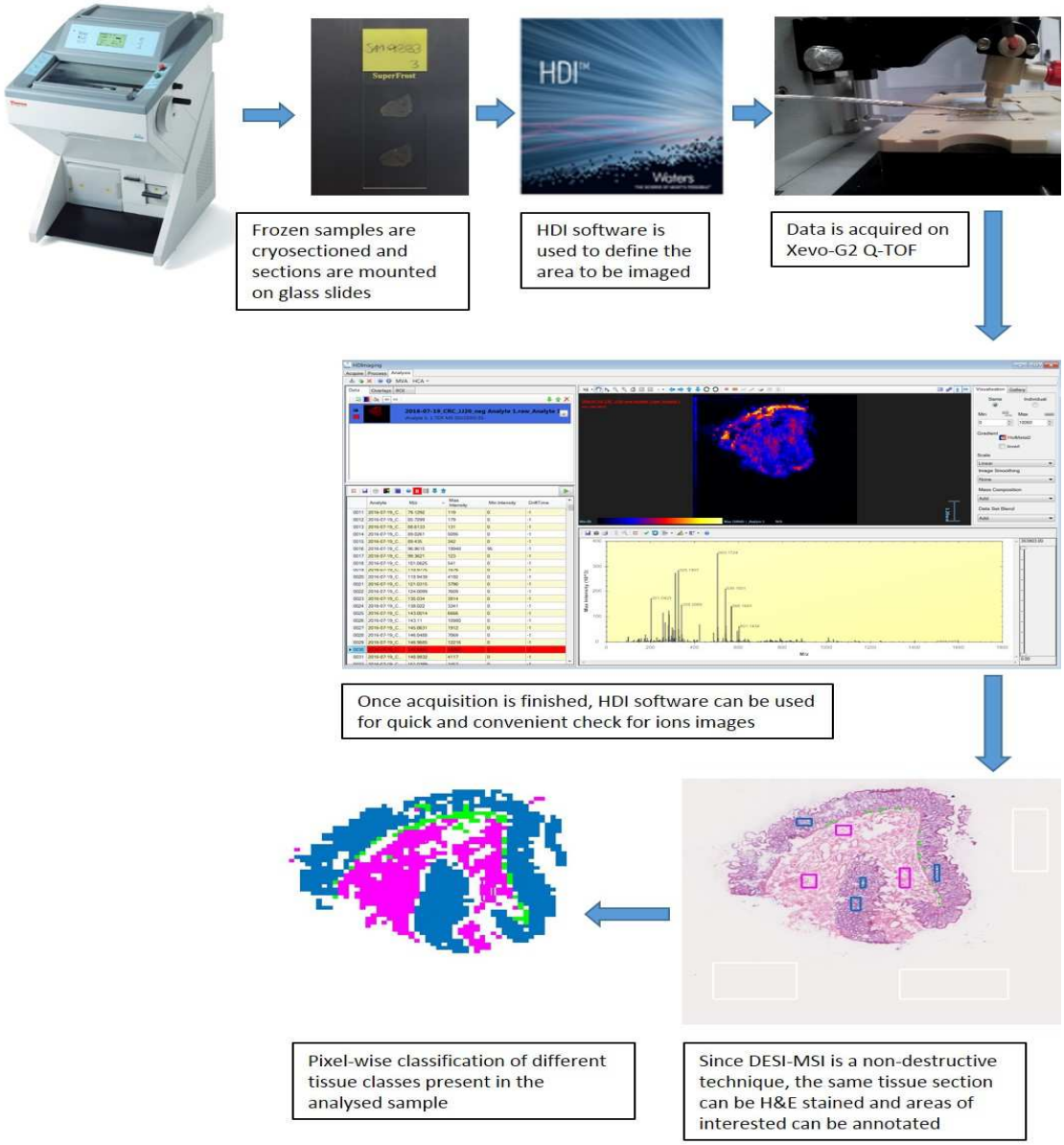


Figure 2.1. General DESI-MSI workflow. This particular chart reflects the steps related to samples analysed on Xevo-G2 QToF. HDI software is only compatible with this instrument; different software was used in conjunction with Exactive. Otherwise, all the steps were precisely the same for both mass spectrometers used to acquire the data. The same workflow was applied to both fresh-frozen and FFPE tissue samples.

2.6 Automated DESI-MSI

A small batch of colorectal samples was analysed as a part of a project run by Dr Emrys Jones from Waters, Wilmslow. The samples were analysed on Xevo-G2 QToF.

The workflow was the same as described in *Chapter* 2.5; however, the significant difference was related to the fact that the automated approach allowed for queuing up to 200 slides. Details are shown in Figure 2.2.

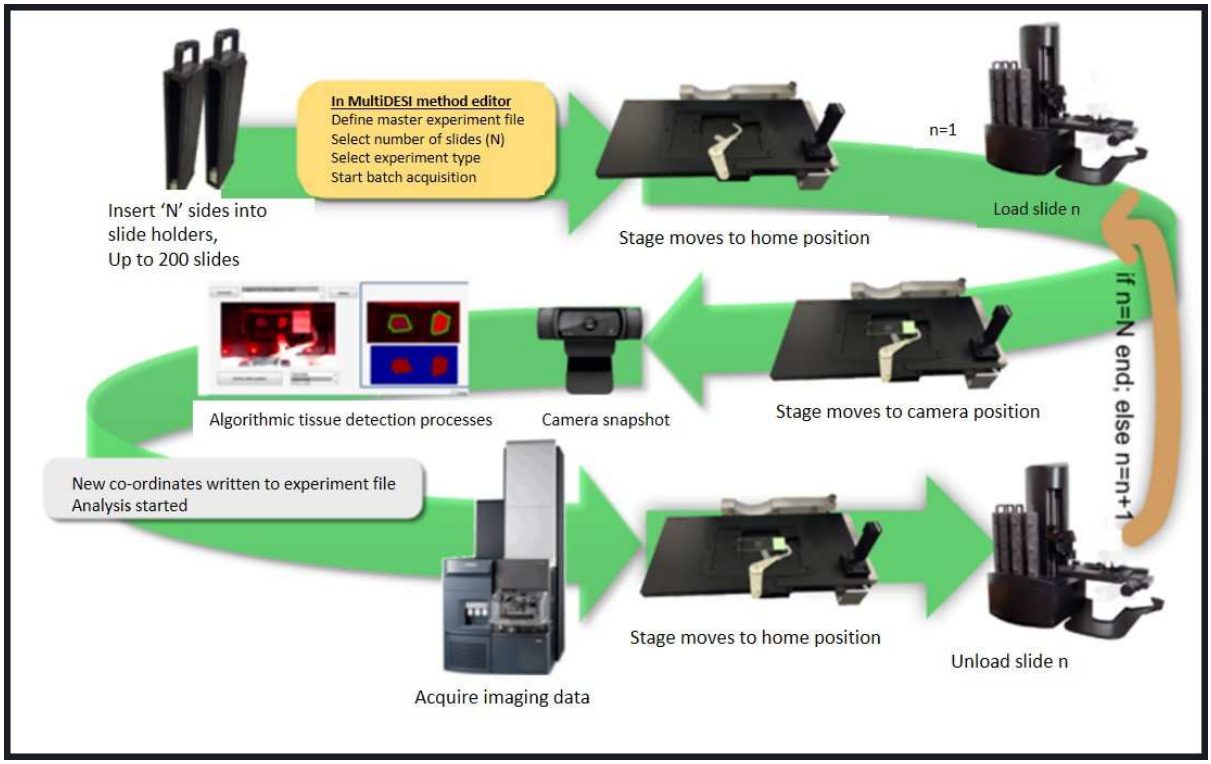


Figure 2.2. Automated DESI-MSI. The standard DESI-MSI experiments required setting up the runs one by one. In the case of Xevo-G2 QToF, a maximum of two slides could be queued. Here, the user could load up to 200 slides which would then be automatically loaded one by one for mass spectrometric analysis. A camera was used to take an image of each glass slide which was then used for the algorithmic tissue detection process.

2.7 Sample preparation

2.7.1 Fresh-frozen samples

In the case of fresh-frozen samples, sample preparation is relatively straightforward.

DESI-MS analysis of biological tissue is usually performed on tissue sections mounted on a non-conductive surface such as glass slides [79]. In order to enable cryosectioning of frozen samples, they must be mounted to a chuck first. Diagnostic histopathology laboratories routinely use OCT (optimal temperature cutting compound) for this purpose. OCT is a water-soluble blend of glycols and resins that provides a convenient specimen matrix for cryosectioning. However, this gel-like matrix is known to interfere with the mass spectra collected from tissue samples and introduce polymers in the positive ionisation mode. Therefore, it is not compatible with mass spectrometry applications. The effects of OCT on mass spectrometry analysis have been further analysed and described in *Chapter 3*. Liver samples used for that project were divided into two groups, and OCT and distilled water were used for embedding the specimens from each group, respectively.

Otherwise, all fresh-frozen tissue samples were embedded in distilled water (Figure 2.3A).

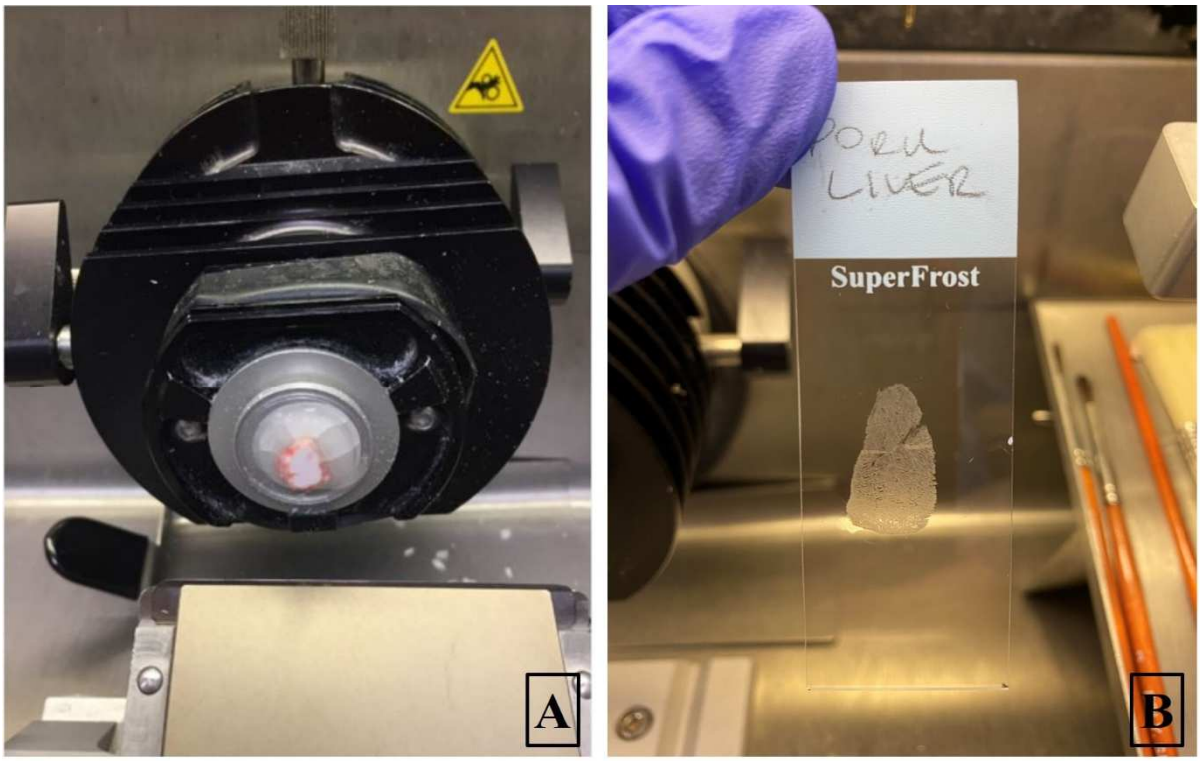


Figure 2.3. Cryosectioning of fresh-frozen tissue samples. Liver sample embedded in distilled water and mounted on a chuck (A). 10µm thick tissue section mounted on a super frost glass slide (B).

Frozen samples were cryosectioned at 10 µm thickness using an FSE cryostat (Thermo Fisher Scientific Inc, Waltham, MA, USA). Tissue sections were thaw mounted onto histology glass slides (SuperFrost® Plus, Thermo Fisher Scientific, USA) as shown in Figure 2.3B. Crysectioning protocol was the same for both OCT and water embedded samples. Tissue sections were kept frozen throughout the procedure, and slides were stored in a -80°C freezer before DESI-MSI analysis.

2.7.2 FFPE samples

Tissue sections for FFPE sample blocks were prepared using a ME+ microtome (Thermo Fisher Scientific Inc, Waltham, MA, USA). Details of this procedure are shown in Figure 2.4. Sections were either 5 or 10 µm thick.

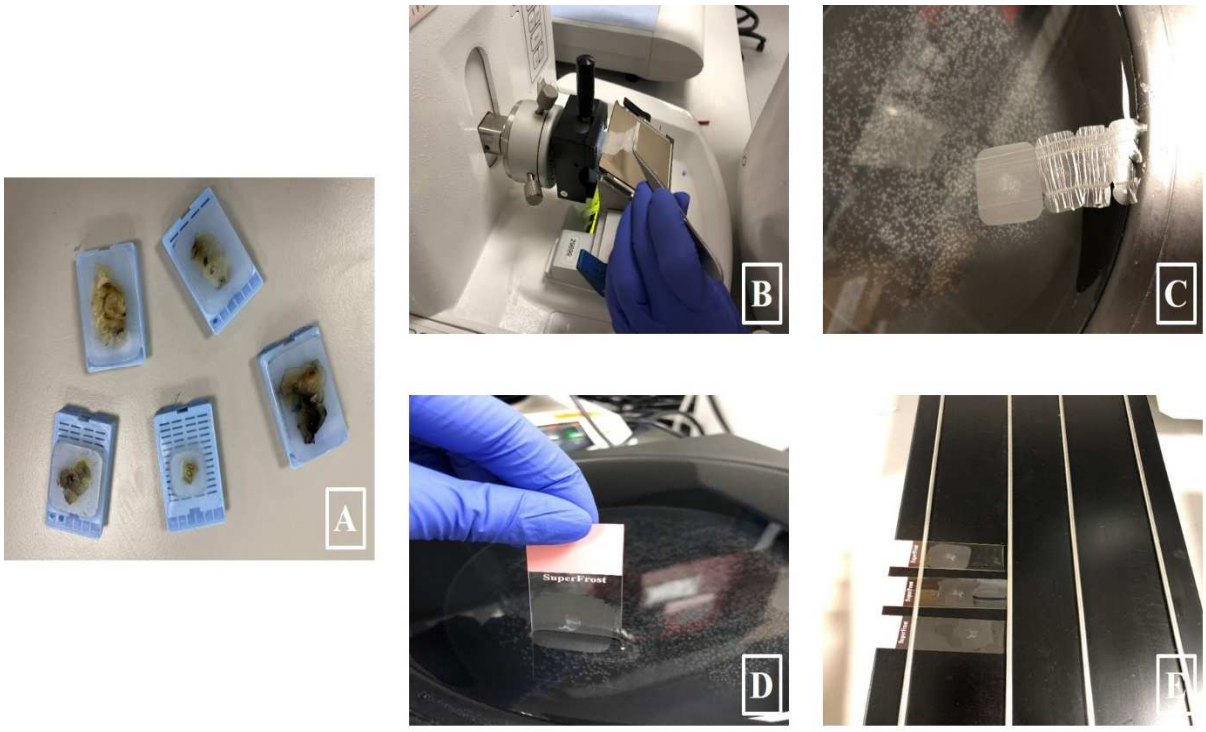


Figure 2.4. The process of preparation of tissue sections from the FFPE blocks. Tissue samples were embedded in wax to form blocks that could be mounted on a microtome (A). Next, those blocks were briefly cooled on a cooling plate, and one by one, mounted on the microtome. Each block had to be trimmed first to remove the excess of wax covering the sample. Once this was achieved, tissue sections could be cut (B). Once a good quality tissue section was prepared, it was placed on a water bath (water temperature 45°-50°) to allow stretching. (C). Then the section was picked up on a glass slide (D) and placed on a hot plate (~50°C) to evaporate water leftovers (E). In the end, the slides were baked out in the oven (80°) for at least 30 mins to ensure the tissue sections have firmly adhered to the slides.

Colorectal TMA slides were ordered directly from Biomax.us (https://www.biomax.us/) and were shipped cut at either 5µm or 10µm thick.

Before DESI-MSI analysis, FFPE slides were baked out for 1 hr in a 60 °C warming oven and then washed in xylene (Histology grade, Sigma-Aldrich, St Louis, MO, USA) twice for 8 minutes [101]. The samples were left to dry at room temperature for 24 hours prior to DESI-MSI measurements.

Some other pre-DESI protocols were also tested during the optimisation process – details are listed in Materials and methods, *Chapter 4.2*.

2.8 DESI-MSI analysis of fresh-frozen samples

2.8.1 Instrument set up

All fresh-frozen samples used in projects which are part of this dissertation were analysed on a Thermo Exactive Orbitrap instrument.

A home-built 3D XYZ integrated linear stage shown in Figure 2.5A was used in conjunction with a home-built DESI sprayer (Figure 2.5B). Before the analysis, slides with tissue sections were mounted with double-sided tape onto the aluminium stage. The collection of ions was performed using a home-built inlet capillary.

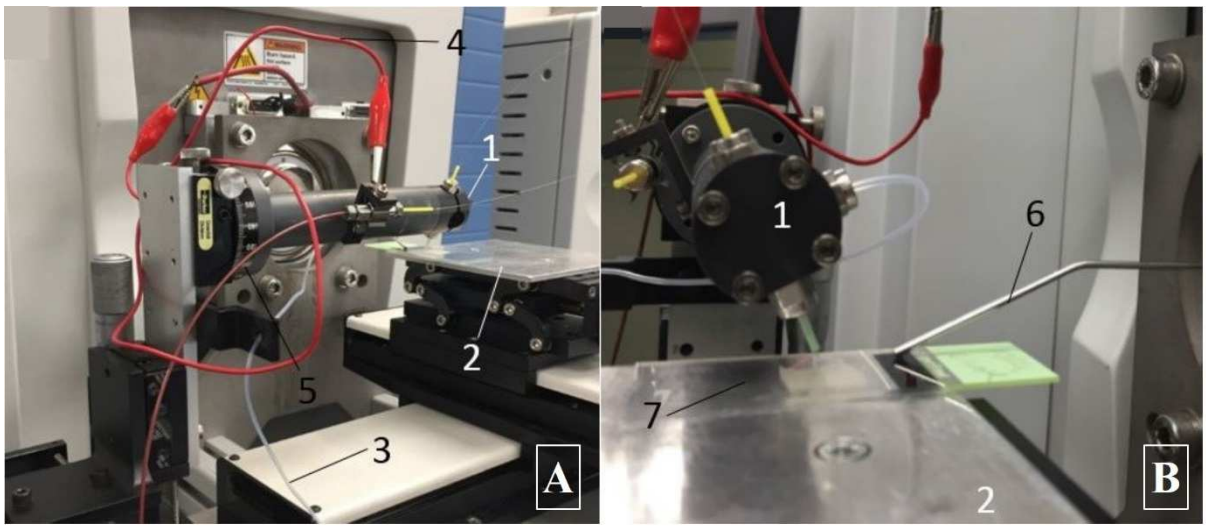


Figure 2.5. Home-built DESI sprayer mounted on the Thermo Exactive Orbitrap instrument. Secured DESI sprayer (1); sample table, movable in *x* and *y*-direction (2); gas supply (3); high voltage supply (4); solvent supply (5); inlet capillary (6); tissue section on glass slide mounted onto the stage using the double-sided tape (7).

2.8.2 Home built sprayer

A schematic representation of the home-built sprayer is shown in Figure 2.6. This sprayer was built using a stainless-steel Swagelok T element as the main body and two fused silica capillaries which were held in place by sleeves and ferrules. For solvent delivery, a 50 µm inner diameter (ID), 150 µm outer diameter (OD) fused silica capillary was used. It was inserted into an approximately 1.5 cm long 250 µm ID, 363 µm OD fused silica capillary to deliver the nebulising gas. The two capillaries were oriented with the inner one, which was touching the outer capillary at its top edge when the sprayer was directed towards the sample (Figure 2.6) at an angle up to 90°. This setup was crucial as it resulted in an elliptical spray point facing the mass spectrometer's inlet capillary and an increased ion yield [118].

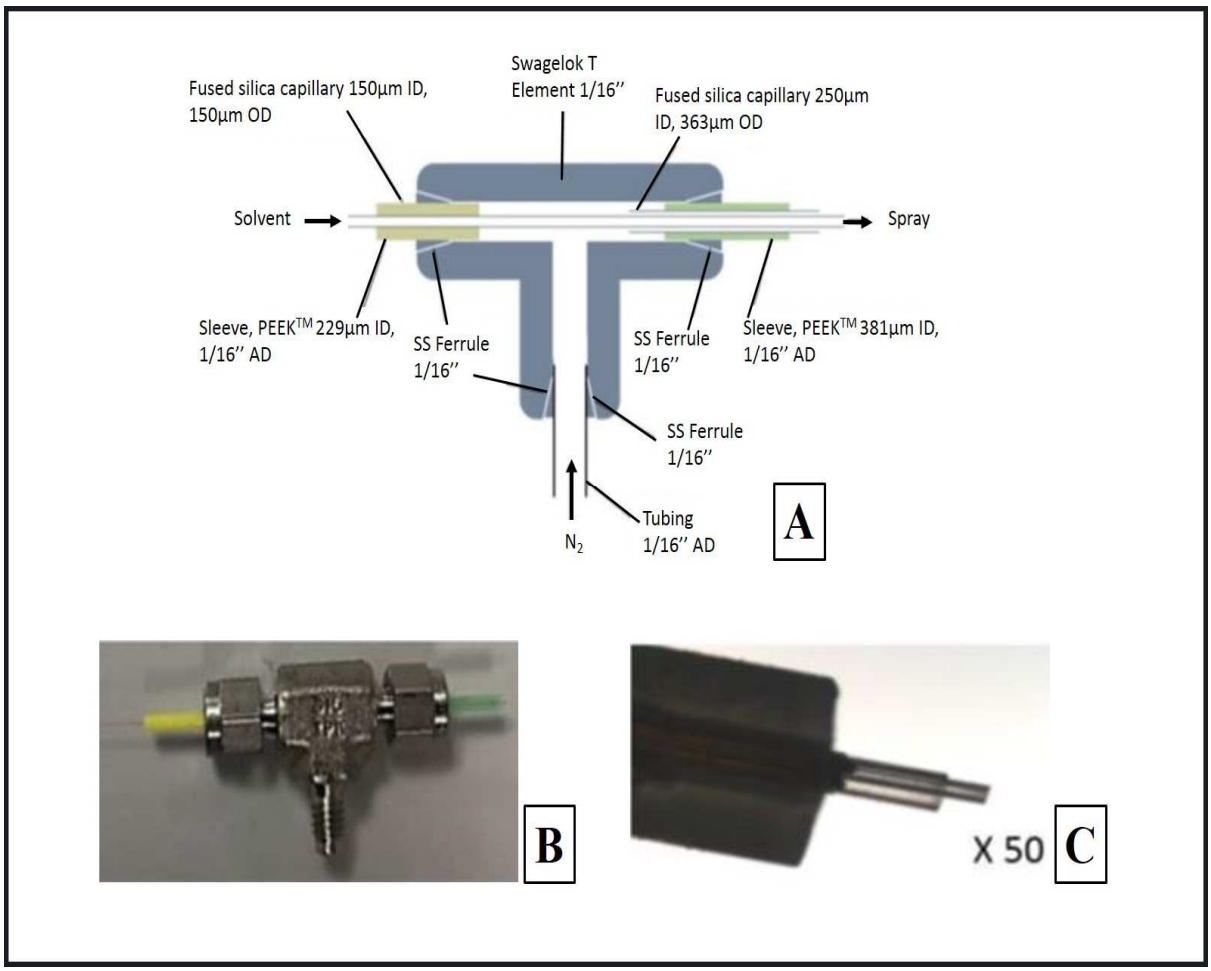


Figure 2.6. Home built DESI sprayer. Schematic representation of home-built DESI sprayer (A), a photograph of a sprayer (B) and a microscopic image of the DESI sprayer (50-fold magnification) (C). ID – inner diameter; OD – outer diameter.

2.8.3 DESI-MSI parameters for the analysis of fresh-frozen tissue samples

DESI-MSI analysis of fresh-frozen samples was performed in negative ion mode using a high-resolution orbital trapping Exactive mass spectrometer (ThermoScientific Inc., Bremen, Germany). The instrumental parameters are listed in Table 2.2. Spatial resolution was always set at 100 µm. Methanol: water, 95:5, v/v (Sigma-Aldrich, St. Lewis, Missouri, USA) was used as the electrospray solvent at a flow rate of 1.5 µl/min. Zero-grade nitrogen was used as the nebulising gas at a pressure of 7 bar. The height distance between the DESI sprayer and the sample surface was set to 2 mm. The distance between the sprayer and the inlet capillary of the mass spectrometer was set to 14 mm. The distance between the sample surface and the inlet capillary of the instrument was less than 1 mm. The angle between the sprayer tip and the sample surface was set at 75°. The angle between the inlet capillary and the sample surface was set to 10°. These parameters were selected based on the reproducibility study by Abbassi-Ghadi *et al*. [119].

Regions of interest were selected with the help of Moving stage controlling software.

Parameters	Settings of Thermo Exactive instrument
Polarity	Negative
Max injection time	1000 ms
Micro-scans	1
Mass resolution	100000
Mass range	150-2000 *m/z*
Capillary temperature	250°C
Capillary voltage	-50 V
Tube lens voltage	-150 V
Skimmer voltage	-40 V
Spray voltage	4.5 kV
Skimmer voltage	-40 V

Table 2.2. Thermo Exactive Orbitrap settings used for DESI-MSI of fresh-frozen samples.

2.9 DESI-MSI analysis of FFPE samples

2.9.1 Instrument set up

All FFPE samples were analysed using a Xevo G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA) coupled to a 2D sample stage (Prosolia Inc., Indianapolis, IN, USA), as shown in Figure 2.7A. FFPE slides were mounted onto the commercial stage for the analysis, and the collection of ions was performed using a custom-built inlet capillary heated to 500 °C (Figure 2.7B).

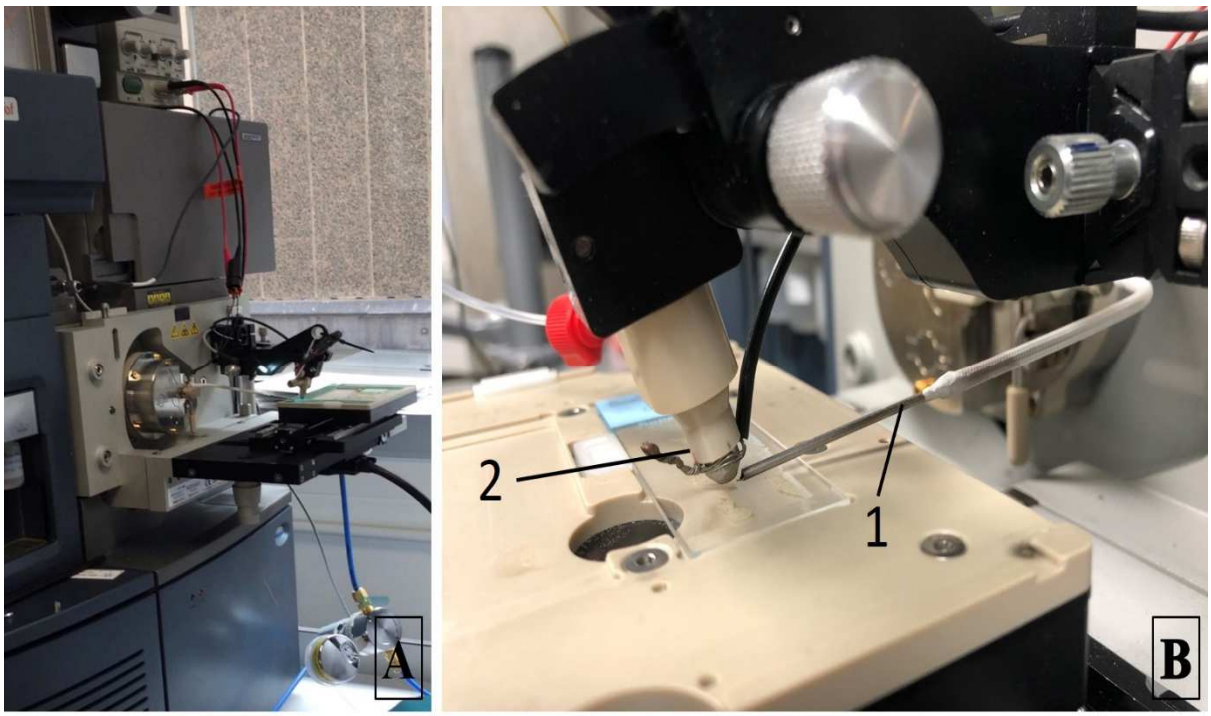


Figure 2.7. DESI-MSI set up on a Xevo G2-XS QToF mass spectrometer. Stage with a novel DESI sprayer mounted to the instrument (A); redesigned, more robust DESI sprayer (1); heated capillary (2) (B).

2.9.2 Novel DESI sprayer

A novel re-designed DESI sprayer (Waters, Milford, MA, USA) described by Tillner *et al.,* [120] was used for analysing FFPE samples. It consisted of a 400 µm aperture, stainless steel gas nozzle and a 360 µm OD and 20 µm ID TaperTip™ emitter. The emitter was secured by a stainless-steel emitter guide which slides into a stainless steel positioning disc with radially arranged holes for the nebulising gas flow (Figure 2.8) [120].

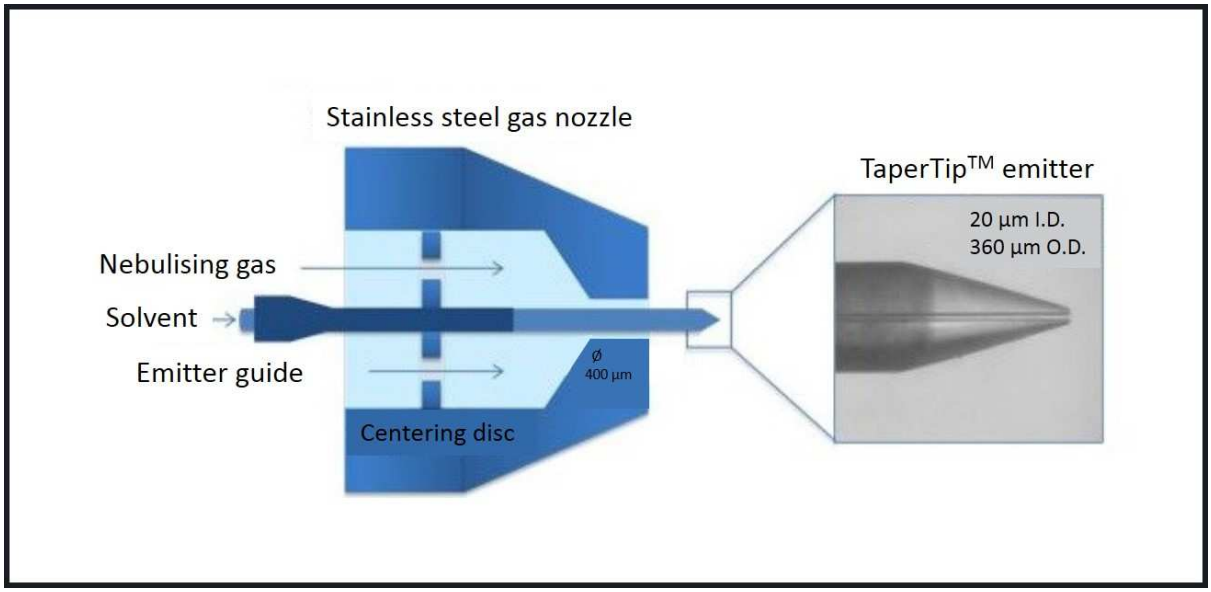


Figure 2.8. Schematic representation of the TaperTip™ emitter, gas nozzle and the emitter guide. Image reproduced under the terms of the Creative Commons Attributions International License 4.0 [120]. ID – inner diameter; OD – outer diameter.

2.9.3 DESI-MSI settings for the analysis of FFPE samples

All DESI-MSI experiments involving FFPE samples were performed on a Xevo G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA) coupled to a 2D sample stage (Prosolia Inc., Indianapolis, IN, USA). These samples were analysed in both positive and negative ionisation modes. A custom-built inlet mass spectrometry capillary heated at 500°C was used for all experimental measurements. A nanoAcquity binary solvent manager (Waters Corporation) was coupled to the DESI sprayer for DESI solvent delivery, which consisted of 95:5 methanol (Sigma-Aldrich): water (Thermo Fisher Scientific Inc) at a flow rate of 1.5 µl/min. Zero-grade nitrogen was used as the nebulising gas at an inlet pressure of 7 bar. The height distance between the DESI sprayer and the sample surface and between the sprayer and the heated capillary was set to 2 mm and 10 mm, respectively. The distance between the sample surface and the inlet capillary of the mass spectrometer was less than 1 mm. The angle between the sprayer tip and the sample surface was set at 75°. The angle between the inlet capillary and the sample surface was set to 10°. 100 μm^2 pixel size and 100 µm step size were set up for imaging analysis at a scan rate of 1 scan/second. All imaging analysis was performed using 50 to 1500 mass range.

Regions of interest were selected with High Definition Imaging software (HDI, v 1.4, Waters Corporation). The data collection required a combination of HDI and MassLynx version 4.1 software (Waters, Milford, MA, USA).

Parameters	Xevo G2-XS QToF (Waters Corporation, Milford, MA)	
Scan time	1 sec	
Scan mode	Sensitivity	
Mass analyser	TOF	
Mass range	50-1500 *m/z*	
Ionisation mode	Negative	Positive
Sampling cone voltage	-40 V	40 V
Source offset	-80 V	80 V
Source temperature	120°C	120°C
Spray voltage	4.5 kV	4.5 kV

Table 2.3. Xevo G2-XS QToF parameters used for DESI-MSI of FFPE samples.

2.10 The rationale behind using two different types of mass spectrometers

All fresh-frozen samples were analysed on a Thermo Exactive Orbitrap instrument. Orbitrap technology is known for higher resolution resulting in greater spectral quality compared to ToF. Therefore, the starting quality of the raw data acquired on Orbitrap is superior. As a result, all the data processing has less impact on the final results contributing to more reproducible data. This aspect was of great importance for long-term DESI-MSI studies of colorectal and cirrhotic liver samples.

ToF technology used in the laboratory, on the other hand, could be described as a commercial platform. It relied on more standardised parts, out of which the sprayer was especially crucial, and the whole setup was considered more robust. As the ultimate goal of DESI-MSI analysis of FFPE samples was to make this technique as histology-friendly as possible, it was clear that all the experiments should be run using a commercially available platform rather than a home-built setup.

2.11 DESI-MSI acquisition workflow

The acquisition pattern for DESI-MSI is shown in Figure 2.9. First, a slide was mounted onto the stage, and the area of interest was selected. Next, the automated DESI sprayer performed horizontal line scans over the tissue surface by moving at a predetermined speed to generate a unique mass spectrum (one scan) for each pixel (at a spatial resolution). After completing the analysis, ion images of the area sampled were constructed to show the spatial distribution of the intensity of selected ions. In this manner, a chemical image of particular lipid species present in the tissue section could be constructed.

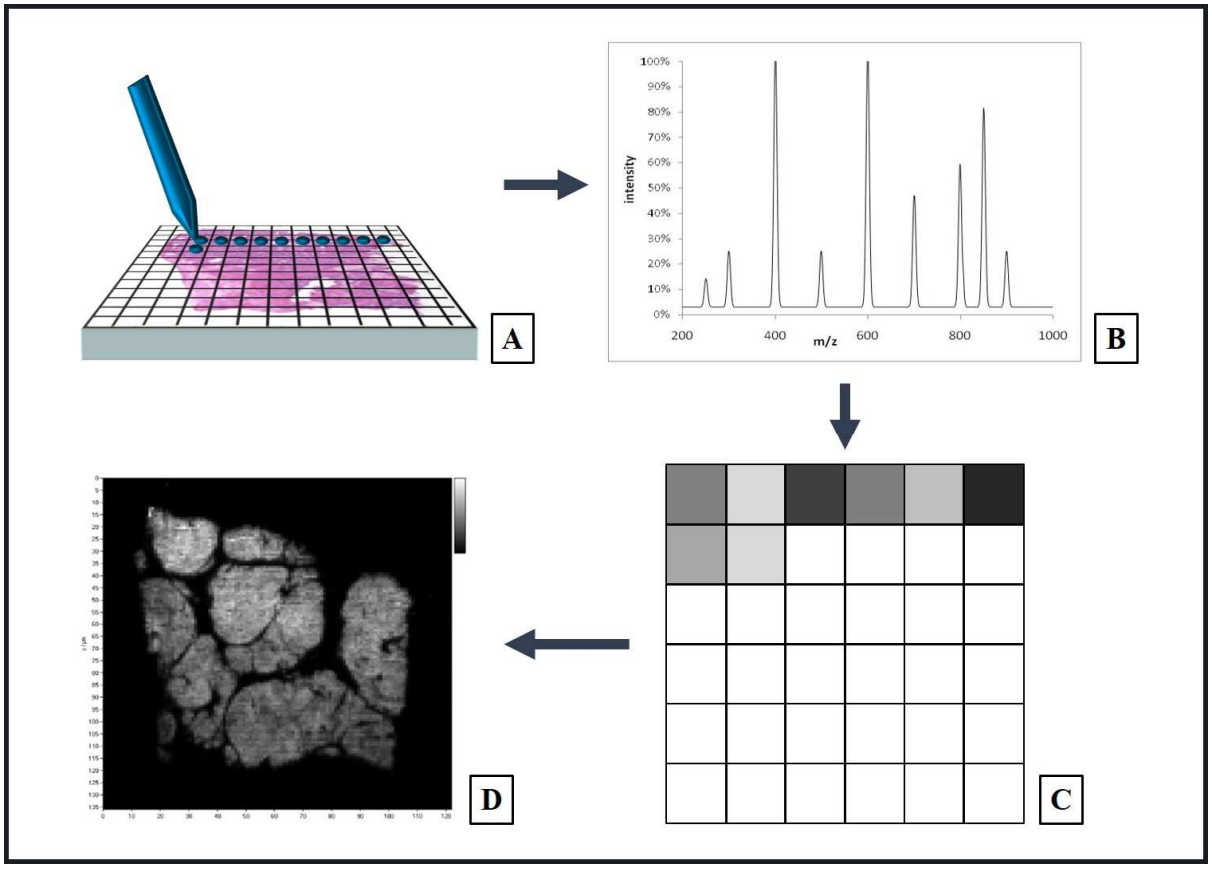


Figure 2.9. Schematic representation of DESI-MSI data acquisition workflow. The tissue section mounted on a glass slide was continuously analysed using horizontal line scans until the whole area of interest was covered (A). A unique mass spectrum was generated for each pixel (B). Ion images of the area sampled were constructed to show the spatial distribution of selected ions' intensity to generate a chemical image of species present in the tissue section (C, D).

2.12 Length of DESI-MSI measurements

All the data were acquired at 100μm pixel size, 1 scan per second.

The total length of the DESI-MSI runs depended on the size of a particular tissue section.

Chapter 3

The size of the tissue sections used for the analysis was around 10 mm x 6 mm. Acquisition time was about 5 – 5.5 hours. Therefore two samples were analysed per day. All the data for this chapter were acquired within one week.

Chapter 4

As this work focused on optimising pre-DESI-MSI protocol, various tissue sections were tested and acquisition times varied.

Chapter 5

Fresh-frozen colorectal tissue sections were around 7 mm x 5 mm, allowing for DESI-MSI experiments to be completed within 3 – 4 hours. Usually, three samples were run per day.

Data acquisition for each of the colorectal TMA slides took around 12 hours.

Chapter 6

Fresh-frozen samples from the cirrhotic liver cohort were the largest of all fresh-frozen tissue samples analysed during this PhD. The average size of liver tissue sections was about 12 mm x 8 mm, resulting in DESI-MSI analysis lasting between 10 to almost 12 hours. Therefore, these samples were only measured overnight.

2.13 Post-DESI-MSI H&E staining of the slides

Following DESI-MSI analysis, the tissue sections underwent H&E staining to allow for a histological tissue classification by a pathologist.

H&E is the most widely used histological stain as it is a simple way to demonstrate a high number of different tissue structures. The hematoxylin component stains the cell nuclei blue-black, allowing good intranuclear detail. Eosin allows for staining most connective tissue fibres and cell cytoplasm in varying intensities and shades of red, orange, and pink [55].

The details of the H&E staining protocol are listed in Table 2.4.

Step	Time
Stain nuclei in Harris hematoxylin	1 minute
Rinse (immerse in a dish with running water)	10 seconds
Differentiate in 1% acid alcohol	2-3 dips
Rinse (immerse in a dish with running water)	10 seconds
"blue" in Scotts water	20 seconds
Rinse (immerse in a dish with running water)	10 seconds
Stain in 1%eosin	45 seconds
Rinse (immerse in a dish with running water)	10 seconds
Dehydrate in 99% alcohol x3	20-30 dips
Clear in xylene x2	20-30 dips

Table 2.4. H&E staining protocol. Since frozen sections can be fragile, all the steps were done by hand to reduce the total time of the procedure.

After the H&E staining, the slides were cover-slipped to prevent tissue sections from dehydration and allow long-term and safe storage.

2.14 Slide scanning

All the H&E stained slides were scanned with a Hamamatsu NanoZoomer slide scanner. Digitalised images were produced at x40 resolution.

2.15 MS/MS analysis

MS/MS analysis was done as a part of the experimental work in *Chapter 6* to confirm the identity of the most abundant ions present in cirrhotic liver diseases samples.

Ten randomly selected slides (stored at -80°), each with two tissue sections, were used to prepare tissue extractions. Care was taken to ensure that these slides represented various liver diseases. A blade was used to scrape tissue sections into an Eppendorf tube. Next, 300µl of methanol was added, and the tube was centrifuged for 15 minutes at 12000 RPM.

A collision-induced dissociation feature on the LTQ-Orbitrap Discovery mass spectrometer (Thermo Scientific, Bremen, Germany) was employed, and tissue extractions were directly injected using a commercial source. Collision energy values between 30-45 eV were used to obtain the characteristic fragmentation spectra for the top correlating molecular species. Fragmentation spectra were only obtained when the ions of interest gave a sufficiently high signal.

2.16 Data processing

Following DESI-MSI data acquisition on Exactive instruments, an imzML converter (version 1.0.5) [121] was used to combine the series of raw files for each imaging dataset. The following imzML files were then imported into a MATLAB 2014a (http://www.mathworks.co.uk) environment. An in-house bioinformatics platform developed especially for interrogation of MSI data sets was used for data processing and feature extraction [114].
Imaging data obtained using Xevo G2-XS QToF were analysed directly in HDImaging, or concatenated raw files were centroided and lock mass corrected with the predominant raffinose peaks: 503.42 $[M-H]^-$, 539.89 $[M+Cl]^-$ in negative ion mode and 527.43 $[M+Na]^+$ and 543.43 $[M+K]^+$ in positive mode with MassLynx version 4.1. This procedure aimed to 'lock' on to peaks with known *m/z* to correct for calibration drift with time. Raffinose was found to be ideal for fatty acid and phospholipid mass spectrometric analysis because raffinose *m/z* peaks are located between the fatty acid (*m/z* 200-400) and the phospholipid (*m/z* 700-900) *m/z* ranges.

After lock mass correction, it was possible to achieve mass accuracies of less than 5 ppm (Table 2.5) for low mass peaks, such as arachidonic acid (exact mass $[M-H]^-$ of 303.2322), as well as for higher masses, for example for phosphoinositol (PI) 38:4 (exact mass $[M-H]^-$ of 885.5499).

Ionisation mode	Exact mass	Observed mass (uncorrected spectrum)	ppm_{uncorr}	Observed mass (corrected spectrum)	ppm_{corr}
Negative	303.2322 (arachidonic acid)	303.2255	22	303.2325	1
	885.5499 (PI 38:4)	885.5262	27	885.5468	3
	766.5392 (PE 38:4)	766.5186	27	766.5359	4
Positive	782.5670 (PC34:1 $[M+Na]^+$)	782.6082	53	782.5642	3
	824.5566 (PC 36:2 $[M+K]^+$)	824.6021	55	824.5554	1

Table 2.5. Table of mass accuracy for m/z peaks detected in pork liver in negative and positive ionization modes showing observed masses before and after lock mass correction. ppm – parts per million (mass error).

2.16.1 Removal of solvent/ noise-related peaks

Following the peak picking of spectra, solvent related peaks were removed from the data.

As DESI-MSI detects *m/z* species in the order of thousands, solvent and noise-related peak subtraction feature is essential to ensure that the extracted profiles are of genuine biological relevance.

2.16.2 Image alignment

Since each MSI dataset was composed of numerous pixels consisting of unique mass spectra, precise co-registration of the optical H&E stained image with the DESI-MSI image was essential for the effective extraction of morphological and biochemical features. For example, colorectal samples could be classified as healthy or tumour, and further histological discrimination was possible within each of these two types. In the healthy samples batch, tissue classes like mucosa, submucosa, and muscle were most commonly seen.

The image alignment procedure is shown in Figure 2.10. It was performed using an in-house developed affine image transformation (translations, rotation, and scaling) algorithm. Each histological image was co-registered to the DESI-MSI ion image. H&E optical image was used to select areas of interest (various tissue classes) which allowed tissue-specific mass spectra extraction (Figure 2.10), which were subsequently used for multivariate analyses. The image

alignment process was repeated for each sample to extract mass spectra of different tissue subtypes, build a database, and allow further statistical analysis [114].

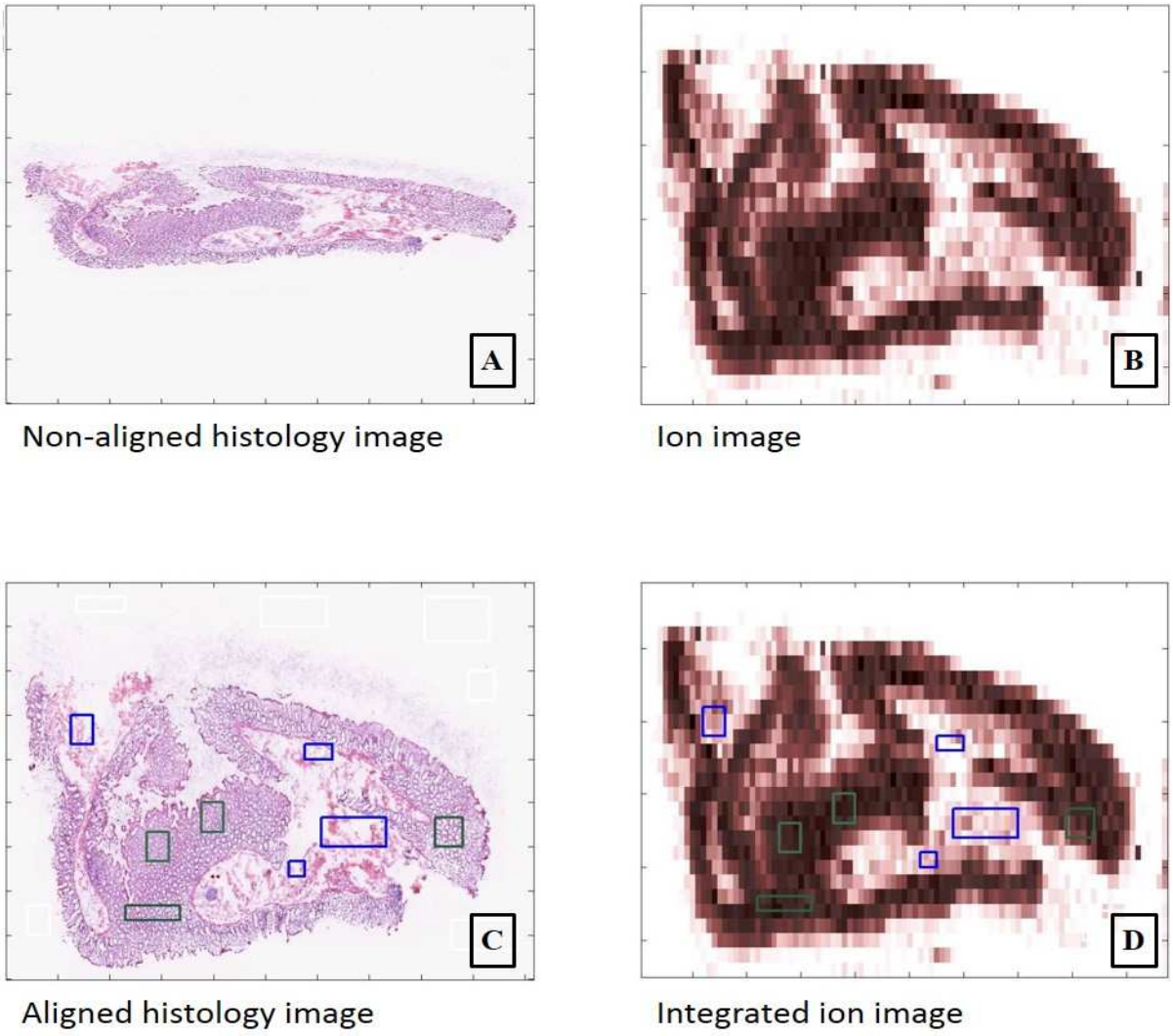


Figure 2.10. Image alignment using the in-house MATLAB platform for tissue-specific pixel labelling. Non-aligned optical H&E image (A); DESI Total Ion Current (TIC) image (B); aligned optical H&E image with annotated regions (C); MSI showing tissue-specific pixel labelling (D); Respective mass spectra of specific tissue types extracted from the ion image (E).

2.16.3 Mass range selection

All the data acquired on Exactive Orbitrap instruments covered 150-1500 *m/z*. These mass spectrometers were used for the analysis of fresh-frozen colorectal and liver samples. As future tissue classification in all fresh-frozen tissue samples relied on changes in lipids composition, 600-1000 *m/z* range was used for further statistical analysis.

Xevo G2-XS QToF was used for the DESI-MSI measurements of FFPE tissue samples, and the data were recorded in 50-1500 *m/z* range. In this case, uni- and multivariate statistics were performed on the *m/z* range 100 to 1000. It was hypothesized that the molecular content of the FFPE samples would be very different to fresh-frozen ones. Therefore, potential changes in the molecular content allowing successful tissue classification would not only (if at all) rely on lipids, and a much broader mass range allowed to cover alterations in other molecular species, e.g. metabolites. The FFPE samples were analysed in negative and positive ionization modes, as it was hypothesized that more chemical information would be captured in the positive mode.

In both cases, subsequent data pre-processing of raw data was performed on the selected *m/z* ranges of interest.

2.16.4 Peak alignment

Although both Exactive Orbitrap and Xevo QToF instruments were calibrated daily, non-linear peak shifts were quite common. To compensate for those shifts, an in-house developed peak alignment algorithm was used (dynamic programming, 10ppm matched) before performing multiple sample comparisons.

2.16.5 Normalisation

The aligned data were TIC normalised to account for differences in signal intensities of molecular species. These intensity variations could have arisen due to instrument variations and differences in cellular density in heterogeneous tissues. The same overall sum intensity of peaks was allocated for a given mass range by imposing TIC normalisation, irrespective of the sample being analysed. Therefore, the only difference would be the relative abundance of the spectral features found within that mass range.

2.16.6 Data averaging

From the selected regions/pixels of interest, the mean average spectra were calculated for each tissue type for each patient and were treated as single data units.

2.16.7 Multivariate statistical analysis

The processed data were subjected to supervised and unsupervised analysis strategies to reduce their complexity and identify spectral features of discriminatory power. Multivariate statistical methods allowed finding both similarities and differences between samples in each particular dataset. In unsupervised strategies, such as PCA, the classification was based on data analysis without the user providing sample classes. This method provided a relative measure of the relationship between samples and was used to identify the most significant variance without the influence of external factors [107].

In the case of supervised strategies such as recursive maximum margin criterion (RMMC), the user could select specific sample characteristics, which in turn could be used to classify different characteristics in samples [114].

2.16.8 Principal component analysis

In MSI, PCA can be applied to the extracted mass spectra from selected regions or the entire tissue section [107]. The averaged DESI-MSI profiles extracted from different patients were analysed by PCA in order to identify trends between various tissue classes. This separation was based on the lipidomic profiles of the samples without any prior histological information. The PCA results were presented in score plots and percentages of variance for a given principal component. In most cases, the majority of variation was explained by the first three components. The closer the points were on the scores plot, the closer the similarity in the data.

2.16.9 Recursive maximum margin criterion

RMMC is considered ideal for MSI datasets as it maximises the inter-class variability while minimising intra-class variability [114]. Therefore, it was used as the primary supervised data analysis strategy in order to generate tissue classification models. RMMC was applied to derive components for discriminating spectra by taking the tissue samples' histological class assignment into account. The final reduced set of discriminating components was equal to the number of classes minus one.

2.16.10 Cross-validation

In order to determine the predictive quality of models based on the RMMC approach, cross-validation was performed. As an external validation set was not available for this project, training and test sets were composed of the acquired data. The discriminating models were validated using leave-one-out cross-validation [122]. DESI-MSI data from one patient were excluded at a time from the sample set; this allowed for a new model to be calculated using the remaining data. The withheld data from the excluded patient were subsequently projected into the calculated model and classified using Mahalanobis distance-based classifier. This process was repeated until all patients' data were sequentially excluded, models were rebuilt, and the classification of excluded data was performed. The classification results were presented in the form of "confusion matrices" (Figure 2.11). A confusion matrix was used throughout this thesis to demonstrate how well the predicted classifications correspond with the actual classifications. From the confusion matrix, sensitivity, specificity, and overall classification accuracy of the models were calculated:

- Classification accuracy = (TP + TN) / (TP + FP + FN + TN)
- Sensitivity (true positive rate) = TP / (TP + FN)
- Specificity (true negative rate) = TN / (TN + FP)

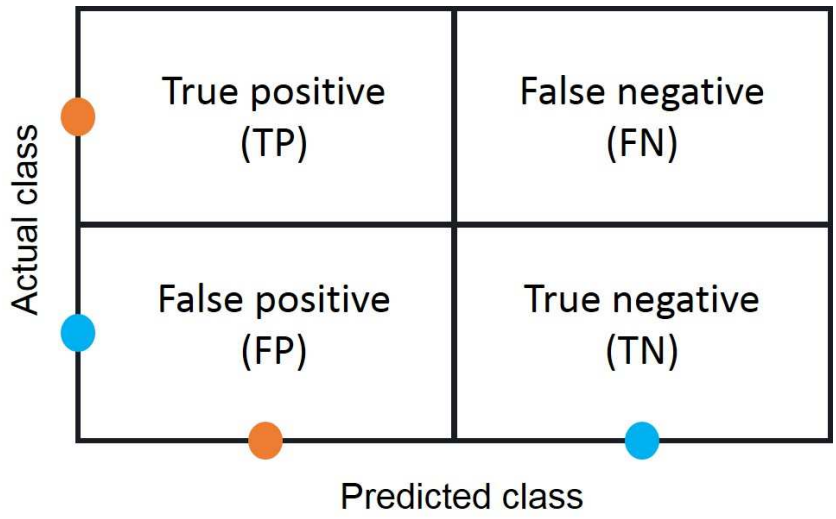


Figure 2.11. Example of a confusion matrix. TP – true positive; FP – false positive; FN – false negative; TN – true negative.

Chapter 3

Impact of OCT embedding on DESI-MSI

3.1 Introduction

3.1.1 Advantages of DESI-MSI

DESI-MSI was developed in 2004 [78] and allowed to analyse samples at ambient temperature and pressure. One of its advantages is the fact that it requires minimal sample preparation and is a non-destructive technique. Therefore, the same tissue sections can be H&E stained afterwards to allow for a histopathology assessment. Since its introduction, DESI-MSI relied on fresh-frozen samples from which so-called frozen sections were cut [78].

3.1.2 Properties of OCT

On the other hand, as part of routine diagnostic histopathology, when a fresh-frozen specimen is received, it is embedded in OCT. At room temperature, OCT is a clear, viscous gel and consists of polyvinyl alcohol, polyethene glycol and other non-reactive chemicals [123-125]. Once frozen, it becomes solid and is used as a specimen matrix to assist with cryosectioning [126, 127].

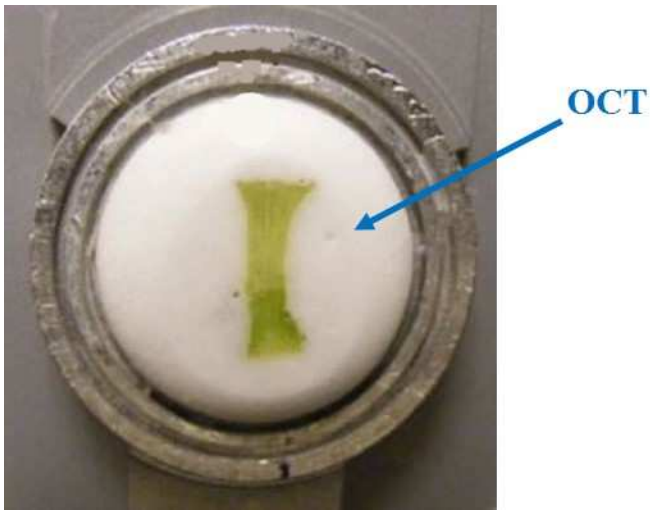


Figure 3.1. An example of a specimen embedded in OCT. Crysectioninig is performed at a minimum -15°C, and at this temperature, OCT turns into a white, solid matrix.

3.1.3 OCT and its use in MSI

However, from a mass spectrometry point of view, tissue samples embedded in OCT are not fully compatible with the imaging analysis. OCT has been documented to cause strong ion suppression when used to embed specimens analysed via MALDI imaging [128, 129] and also causes the detection of a series of polymer peaks at $m/z > 1100$ [129]. Other MALDI-based

studies employed alternative embedding media, such as gelatin [129-131] and carboxymethylcellulose [132-134]. Another study used a modified preparation of OCT to embed mouse lung before MALDI [135]. However, most previous DESI studies have typically relied on distilled water to attach samples to a chuck [92, 136, 137]. This approach has never resulted in the presence of polymers in either positive or negative mode data. As a result, it seems to be the most commonly used sample preparation protocol for DESI-MSI analysis [91, 92, 138].

3.1.4 Assessment of applicability of OCT to DESI-MSI

The translation of DESI from a research tool into one employed for clinical use requires it to be successfully integrated into conventional diagnostic histopathology workflows. The ability to use OCT to prepare frozen sections for analysis via DESI-MSI is a significant step towards this aim. This study compares the impact of the two sample preparation techniques – embedding in water and OCT – on the imaging data quality.

Liver tissue samples embedded in water and OCT were analysed by DESI-MSI in both positive and negative ion modes to assess the practicalities of using OCT as part of a routine DESI and histological workflow.

3.2 Materials and methods

3.2.1 Clinical specimens and sample collection

The tissue used in this experiment was identified as macroscopically normal liver taken from a patient who underwent resection for colorectal adenocarcinoma metastases to the liver. Written consent for the use of this tissue was taken from the patient before surgery. Ethical approval was obtained from the South East London National Research Ethics Committee (Study ID 11/LO/1686).

3.2.2 Sample preparation

Fresh-frozen samples were prepared as described in *Chapter 2.7.1*.

Five liver pieces were cut from the resected tissue. Each of them was fixed to a cryostat chuck with half-frozen distilled water. 10 µm thick frozen sections were then cut at a temperature of -18°C. Each frozen section was thaw mounted onto a SuperFrost Plus microscope slide (Thermo Scientific).

After sections had been cut from each tissue piece, the ice was washed off with room temperature sterile water. The tissue pieces were then embedded in OCT (Surgipath Frozen Section Compound, FSC 22, Leica), and further sections were cut under the same conditions. Thus, for each of the liver pieces (labelled from LIV3 through to LIV7), 4 sections were cut and analysed, i.e. one OCT-embedded and one H_2O-embedded section in each ionisation mode.

Slides were stored at -80°C, and before DESI-MSI, they were allowed to warm to room temperature. After the MS analysis, the slides were H&E stained to permit histological evaluation.

3.2.3 DESI-MSI analysis

DESI-MS was performed using a home-built DESI ion source coupled to a high-resolution orbital trapping mass spectrometer (Exactive, ThermoScientific, GmbH) controlled by XCalibur 2.1 software. The home-built source was operated at a spatial resolution of 100 µm using nitrogen at a pressure of 7 bar, with a voltage of 4.5 kV. A solvent mixture of 95/5 methanol/water was used at a flow rate of 1.5 µL / min. Mass analysis was performed in both negative and positive ion modes over an m/z range of 150-1500.
Full details of the DESI-MSI setup for this batch of samples can be found in *Chapter 2.8*.

Parameters	**Settings of Thermo Exactive instrument**	
Polarity	Negative	Positive
Max injection time	1000 ms	1000 ms
Micro-scans	1	1
Mass resolution	100000	100000
Mass range	150-1500 *m/z*	150-1500 *m/z*
Capillary temperature	250°C	250°C
Capillary voltage	-50 V	50 V
Tube lens voltage	-150 V	150 V
Skimmer voltage	-40 V	40 V
Spray voltage	4.5 kV	4.5 kV
Skimmer voltage	-40 V	40 V

Table 3.1. Thermo Exactive Orbitrap settings used for DESI-MSI of fresh-frozen liver samples.

3.2.4 Data analysis

Data for this project were analysed with Dr James McKenzie's help.

Data for each of the ten samples were converted into imzML format and imported into Matlab using an in-house developed data processing pipeline. The comparison of the positive and negative mode sections was performed using an average mass spectrum determined over all pixels (i.e. tissue and background regions). Four regions of 16 pixels each were annotated in each sample, and these annotated pixels were subsequently used for classification purposes.

Each positive mode OCT-embedded section was analysed individually. 13 individual polymeric distributions were identified across the five sections, each having spacings of 44.026Da corresponding to the repeat unit C_2H_4O. For each of the identified polymeric distributions, the charge was determined by inspecting the spacing between isotopologue peaks. For each distribution centred around the most intense polymeric peak, a linear series with spacings of 44.026/z was created and assessed against the experimental MS spectrum. Peaks were matched within ±10 ppm, and a single continuous distribution from the most intense was determined. For each of these peaks within the distribution, isotopologue peaks were determined. Thus, the spectral profile for each polymer consisted of a single continuous distribution of isotopic clusters. The *m/z* values identified by this method were removed.

Initial comparisons between the OCT- and H_2O-embedded sections for positive and negative mode data were performed using the average spectra over each sample. For each ion mode, a common *m/z* vector representative over all samples was determined, and individual samples were aligned to this with a maximum permitted peak shift of ±10 ppm. Each spectrum was normalised to its total intensity before principal components analysis (PCA).

For the classification of OCT-embedded samples, a cross-validating scheme was employed. Each OCT-embedded sample (i.e. the test set) was classified against a model determined using its paired H_2O-embedded sample and all other OCT-embedded samples (i.e. the training set). The training set samples were aligned, TIC normalised and logarithmically transformed. The test set was subsequently aligned to the training set and normalised and transformed. A maximum margin criterion linear discriminants analysis (MMC+LDA) model was determined using the training set, and the 64 annotated test spectra were classified against this model, using

a 5-nearest neighbours classifier. This approach was repeated for each OCT-embedded sample, with the results presented in confusion matrices.

Univariate analysis was performed using the Kruskal Wallis test. False discovery rate correction was performed using the Benjamini-Hochberg-Yekutieli method with $\alpha = 0.05$.

3.3 Results

After data acquisition was completed, the raw data were inspected by hand to assess its quality. Liver embedded in distilled water was treated as a baseline as this sample preparation method has been known not to introduce any contamination [92, 119, 138], providing the samples were adequately handled before cryosectioning. Figure 3.2 represents averaged spectra recorded in the negative mode for a liver sample embedded in water and OCT. The spectral composition was almost identical for these two samples, and no polymer has been seen in the negative mode data for any of the liver samples embedded in OCT.

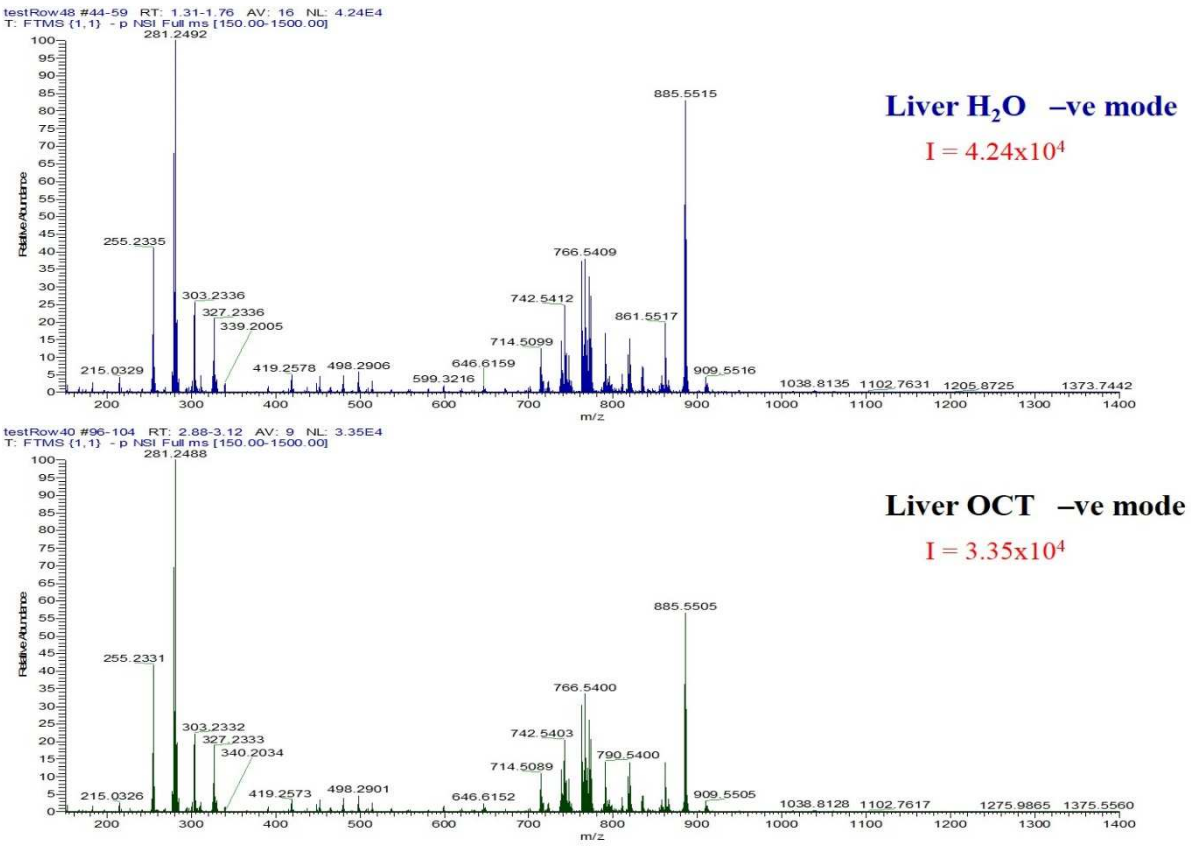


Figure 3.2. Comparison of averaged spectra recorded in negative ionisation mode for different embedding media (top – liver embedded in distilled H_2O; bottom – liver in OCT).

Similarly, the positive mode raw data were carefully checked, as it was expected that the potential polymer peaks would be visible in the positive ionisation mode. An averaged spectrum for the liver embedded in the water looked like a perfect example of what should be expected in terms of good quality data (Figure 3.3). Phospholipid mass range signals were very intense, but otherwise, no contamination was detected. However, when the same liver sample was cryosectioned using OCT, the acquired data started to look very different and very intense polymer peaks started to appear. There were times when those peaks were only present in the mass range from *m/z* 1100 onwards (Figure 3.3), but also, they could be present in the phospholipids range, causing a significant ion suppression (Figure 3.4).

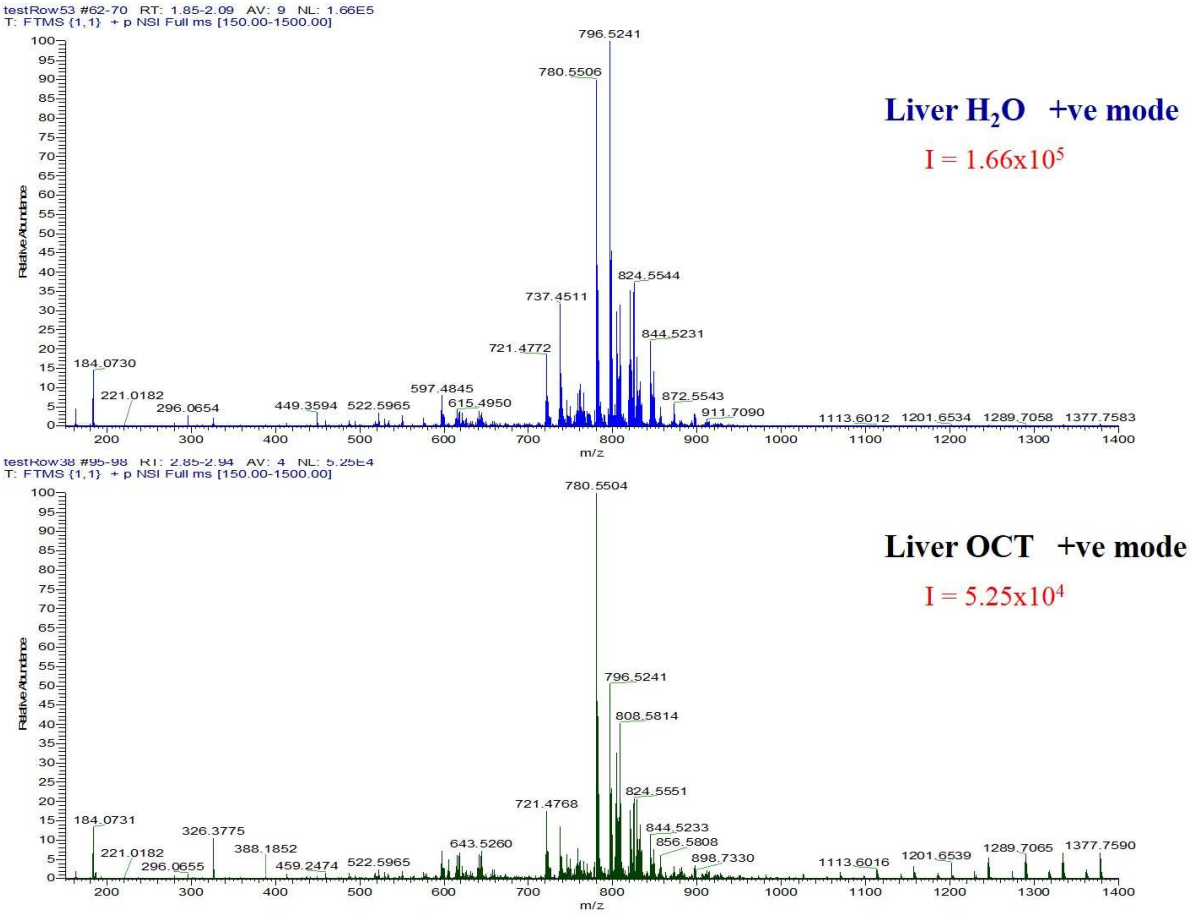


Figure 3.3. Comparison of averaged spectra recorded in positive ionisation mode for different embedding media (top – liver embedded in distilled H_2O; bottom – liver in OCT). Polymer starts to appear from *m/z* 1100 onwards.

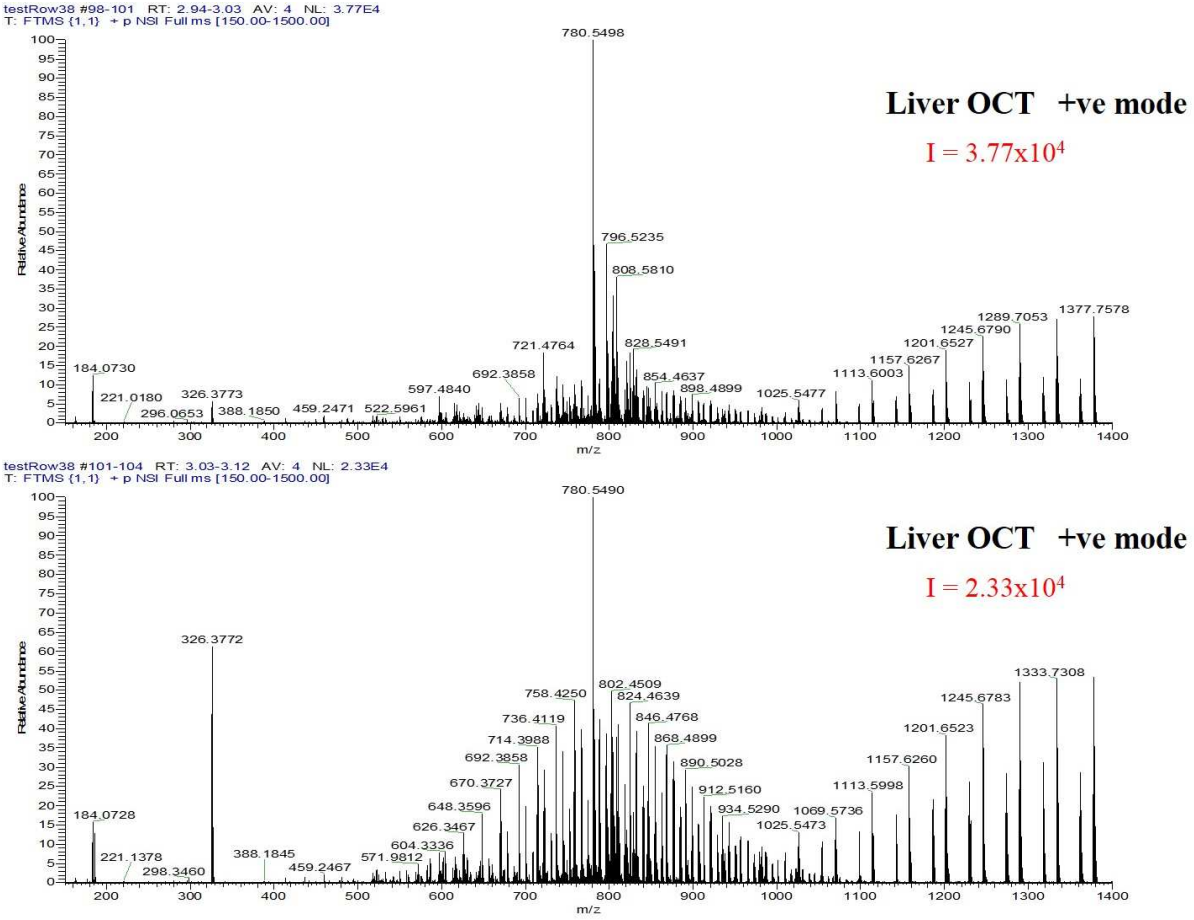


Figure 3.4. Examples of spectra recorded in positive mode for a liver sample embedded in OCT. Polymer peaks are very intense and cause ion suppression for signals of interest (phospholipids mass range) from the analysed tissue section.

One of the most common contaminants reported in metabolomic workflows are surfactants like polyethene glycol (PEG) that are often introduced during sample preparation [139, 140]. These contaminants can be significantly disruptive as they cause ion suppression and interfere with the target ions of interest (e.g. phospholipids – Figure 3.4) [141, 142].

In this study, all the samples have initially been taken from the same patients' liver. Therefore, all of them have been handled the same way. The only source of the polymer was OCT embedding.

3.3.1 Identification and removal of polymeric interference

Figure 3.5 shows the average spectra for a single liver sample (LIV3) analysed in negative and positive ion modes (similar plots for the other liver sections are shown in Appendix A). The difference in the H_2O-embedded and OCT-embedded spectra was clear for positive mode data, with extensive polymeric profiles dominating at m/z > 1000 and to a less noticeable extent

around the phospholipid region (600 < *m/z* < 1000). Additionally, a large peak at m/z = 326.378 was found in all other OCT-embedded spectra. There were no visible polymeric distributions in the negative mode spectra.

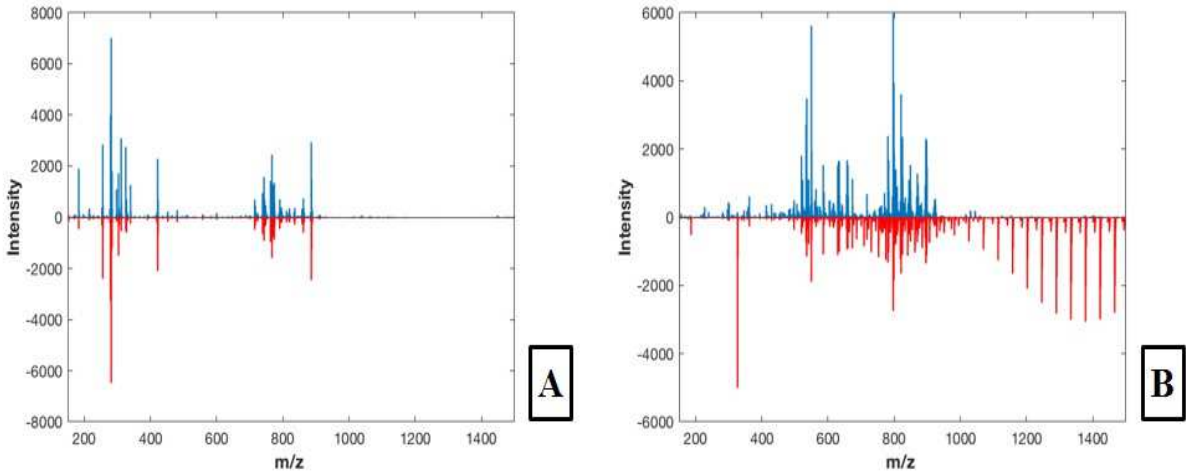


Figure 3.5. Average spectra for LIV3 analysed in the negative (A) and positive (B) ion modes. The H_2O-embedded spectra are shown in blue, and below, in red, are the OCT-embedded spectra.

The polymers in OCT have a monomeric unit of C_2H_4O with a mass of 44.026 Da.

A linear series of *m/z* values (±44.026/z) centred around the identified *m/z* value was created and matched (±10 ppm) against the mean spectrum for each sample. From the most intense peak within the distribution, only those peaks found in a continuous distribution were considered, i.e. there are no gaps in the polymer peaks identified in the spectrum. The isotopologue peaks for each of these were also identified based on the distance expected due to the charge state.

13 polymeric distributions were observed by eye across the five positive modes OCT-embedded samples and were removed by hand (along with the peak clusters at *m/z* = 326.378, 327.380, 328.384). Further details can be found in Appendix B and C. This was not performed for the negative mode sections. Figure 3.6 is similar to that shown in Figure 3.5B, but without the polymeric signals in the OCT-embedded sample. Similar plots for the other four samples are shown in the Supplementary Information, Appendix D. There was an apparent visual similarity between the two spectra, although the intensity was noticeably reduced in the OCT-embedded samples. Whilst there may have been residual traces of polymeric signal remaining

with these processed samples, it was clear that a significant part of the interference has been removed. The identification and removal of intense polymeric signals were important when normalising spectra according to their total intensity. The polymeric signal distribution across the tissue and background is shown in Figure 3.7; it demonstrates that the problem of polymeric contamination cannot be ignored simply by using 'tissue only' pixels.

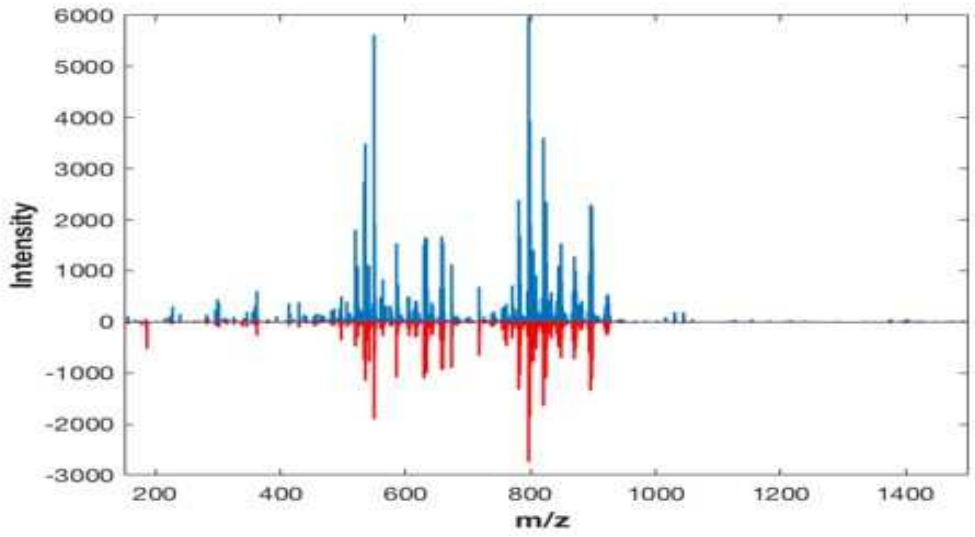


Figure 3.6. Average spectra for LIV3 analysed in positive mode, following the removal of the identified polymer-associated peaks. The H_2O-embedded spectra are shown in blue, and the OCT-embedded spectra in red.

Whilst most of the polymeric signal was confined to the background of each MS image, there was considerable 'contamination' of the tissue signal. Ion image of the polymeric peaks is shown below (Figure 3.7). This demonstrates that the polymer was not purely confined to the background and was abundant throughout the tissue itself. As such, removal of the polymeric signal was required even when excluding background pixels.

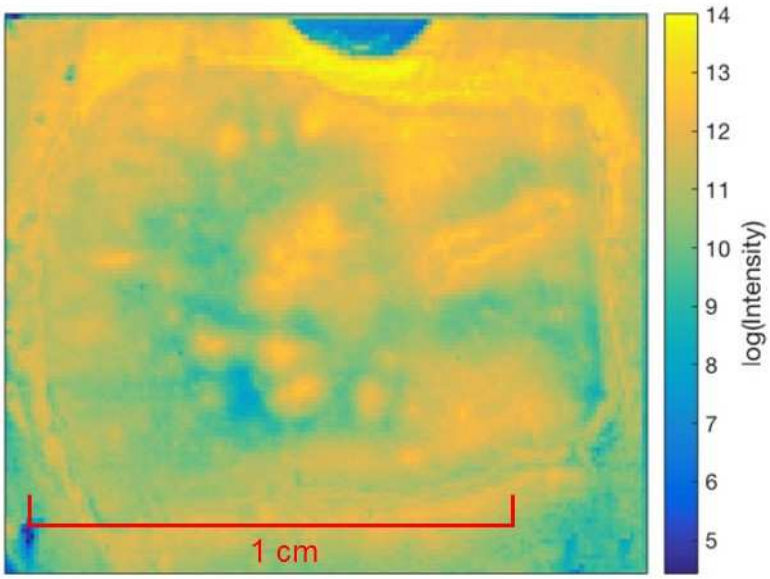


Figure 3.7. Summed and logged ion image composed of all polymeric signals identified in the OCT-embedded LIV3 positive mode liver section.

Following polymer removal, the various *m/z* vectors across the 10 negative and 10 positive mode sections were aligned and processed as described previously. Principal components analysis was performed on both ion modes. Four clusters dominated the PC1 loadings from the positive mode data. Those peaks were identified as surfactant molecules and eliminated from the data (further details can be found in Appendix E). The PCA scores (Figure 3.8A) showed that whilst some sections differ depending on the embedding medium, others (especially LIV3) showed relatively inconsequential differences between H_2O and OCT samples.

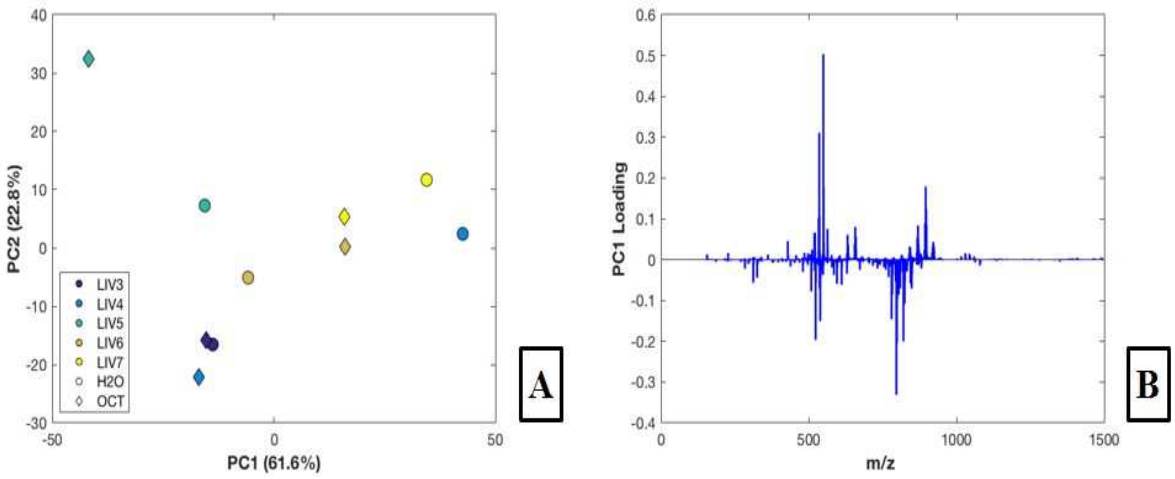


Figure 3.8. PCA of positive mode data (after removal of 4 surfactant clusters) (A). PC1 loadings representing these samples (B). Various colours in the PCA plot represent different samples, while O refers to samples embedded in H_2O and ◊ to samples embedded in OCT.

The negative mode data was different in one of the liver samples (LIV7) and showed a marked decline in *m/z* 279.233 and 281.248 but increased in *m/z* 215.030 (Figure 3.9). This difference has previously been observed in samples left too long at room temperature. However, this issue aside, the PCA was broadly in line with the positive mode data in that there was no apparent trend between OCT- and H_2O-embedded sections. Whilst it would be incorrect to say that OCT embedding did not affect the spectral profile, the scores plots in Figures 3.8A and 3.9A demonstrate that the effect was no more significant than the variation seen across distinct parts of the same tissue.

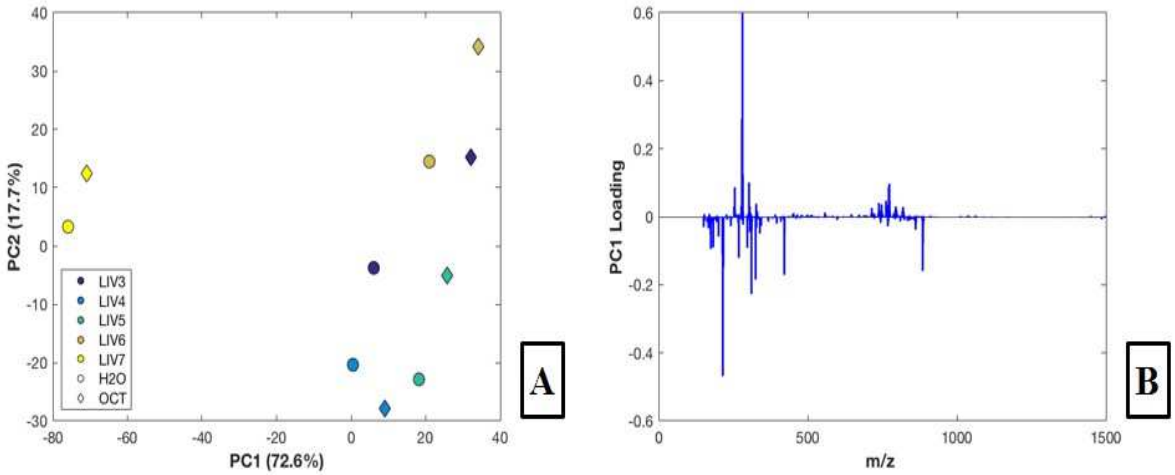


Figure 3.9. PCA of negative mode data. The difference in the LIV7 samples was attributed to the sample being left at room temperature for longer than the others (A). PC1 loadings representing samples analysed in the negative mode (B). Various colours in the PCA plot represent different samples. O refers to samples embedded in H_2O and ◊ to samples embedded in OCT.

3.3.2 Classification

In order to determine the similarity between OCT- and H_2O-embedded spectra from the same liver section, supervised classification was performed. For each OCT-embedded sample (i.e. the test set), supervised classification was determined using all other OCT-embedded samples and the H_2O-embedded sample matching the test set sample. Thus, spectra from each OCT-embedded sample were classified with respect to other OCT-embedded spectra or spectra from the equivalent H_2O-embedded sample.

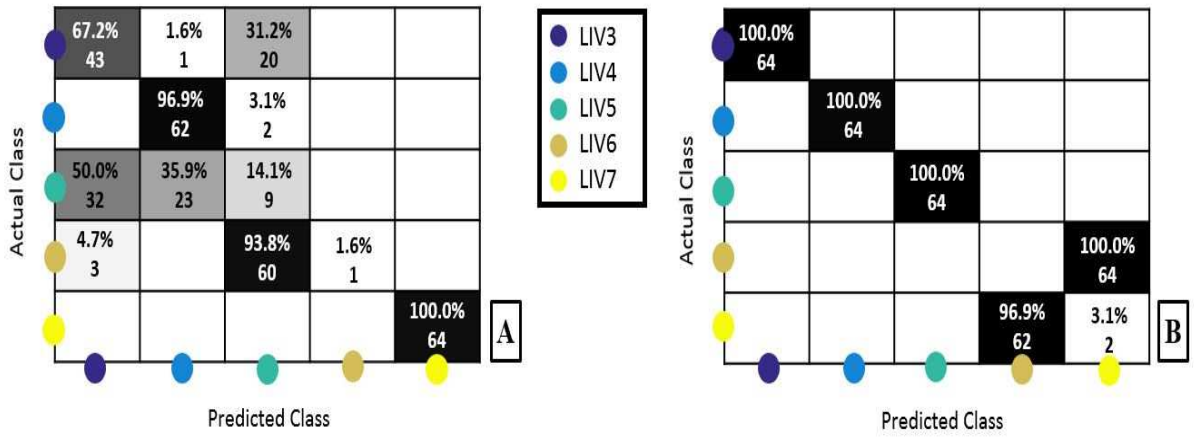


Figure 3.10. Confusion matrices showing the classification performance in the negative (A) and positive (B) ion modes.

The confusion matrices for negative and positive mode classifications are shown in Figures 3.10A and 3.10B, respectively. The negative mode samples exhibited poor classification, demonstrating that OCT-embedded samples were classified as other OCT-embedded samples rather than as the corresponding H_2O-embedded sample. Only the LIV7 sample obtained perfect classification, although this was due to the effect discussed previously (see Figure 3.9A). In contrast, the positive mode data (Figure 3.10B) exhibited perfect classification for 3 of the 5 sections, with LIV6 and LIV7 being almost perfectly swapped.

The confusion matrix in Figure 3.10B can be placed in context when compared to that shown in Figure 3.11. This confusion matrix shows the results of the same classification analysis but performed using the unprocessed positive mode data, i.e. containing all previously removed polymeric interference (the surfactant-like peaks identified in the Supplementary Information have been removed). The difference in the two confusion matrices demonstrates the value of removing the polymeric interference from the positive mode samples.

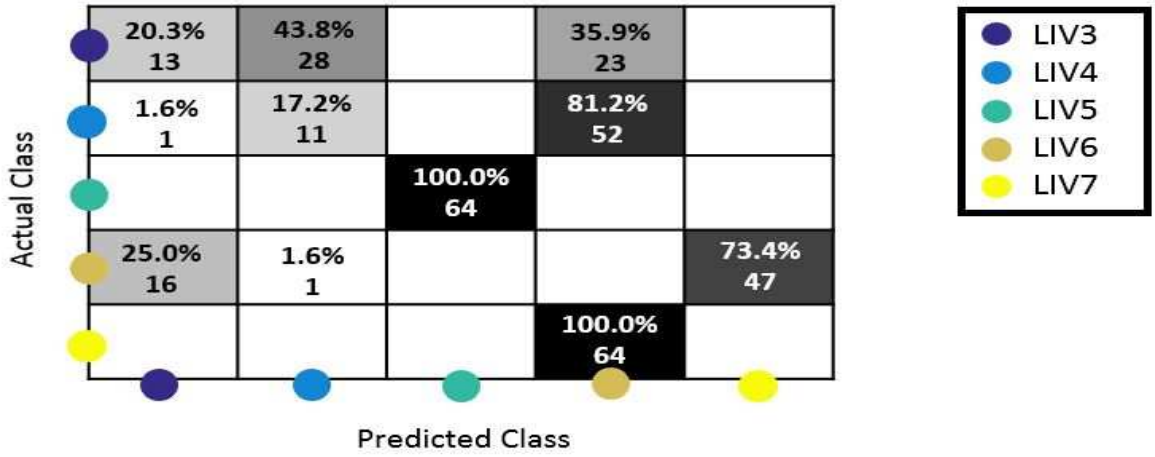


Figure 3.11. Confusion matrix showing the classification performance in the 'raw' positive mode data containing all polymer-identified peaks.

The similarity, however, between Figures 3.10A and 3.11 leads to an inference that the negative mode data was, in some way, affected by OCT. The polymeric peaks in the positive mode sample were pronounced and of considerable, yet variable, magnitude. Their presence also interfered significantly with the normalisation of spectra. Because such an extensive interference was not overt in the negative mode spectra, it suggested that OCT's effect was more subtle. However, the misclassification may also be partly due to the high similarity of the liver tissue sections. The negative mode data failed to distinguish their differences adequately regardless of the embedding medium.

3.3.3 Ion suppression

Univariate analysis between H_2O- and OCT-embedded spectra was performed to determine differences between the two embedding media over the ten negative mode sections. Variables with H_2O/OCT $\log_2$ fold changes greater than 0 with q less than 0.001 were identified; Figure 3.12 shows the boxplot of one such variable (split according to file and embedding medium). 4 of the 5 sections demonstrated lower median intensities in OCT compared to H_2O spectra, with LIV4 in contrast to this. A range of other variables is shown in the Supplementary Information, Appendix F, which shows similar ion suppression evidence. However, the evidence for the suppression of ions was not always consistent across 5 tissue pieces. The PCA scores plot in Figure 3.9A suggests that the differences between OCT- and H_2O-embedded sections were not always consistent.

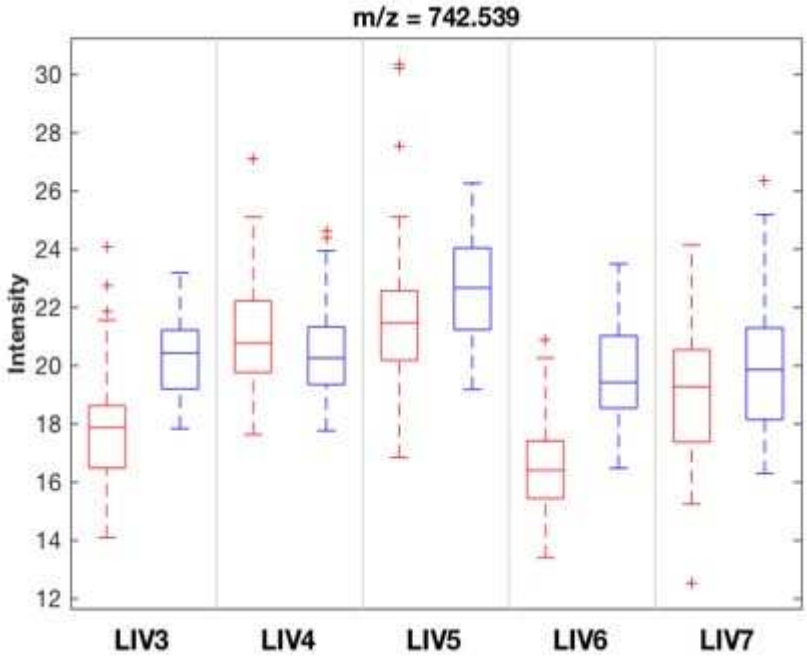


Figure 3.12. Boxplot showing the intensity of *m/z* 742.539, which has q < 0.001 and a positive H_2O/OCT log_2 fold change. OCT intensities are depicted in red and H_2O intensities in blue.

While these figures demonstrate possible ion suppression caused by OCT as an embedding medium, it may not be the only cause that leads to poor classification results. All of the tissue used in this study was obtained from the same patient and is thus expected to be homogeneous. Further investigation of the use of OCT as an embedding medium is required to determine if the use of OCT masks tissue-to-tissue or subject-to-subject variabilities.

3.4 Discussion

The data generated as part of this project suggest that it is possible to employ OCT as an embedding medium for histopathological and mass spectrometric analyses. Whilst some of the reduction in intensities may be attributable to polymer-induced ion suppression, there are no clear polymeric signals causing differences in the negative mode data. In positive mode data, the interferences due to OCT are more overt but can be negated by removing the regular peaks of the various polymeric distributions.

The study was designed using tissue from the same patient and is thus expected to be broadly homogeneous across the 5 tissue pieces. The variation seen over the two embedding media

was broadly of the same magnitude as the inter-tissue variation. The inability to correctly classify according to tissue section should be considered in light of the similarity of the liver sections and similarity in spectral profiles from OCT- and H_2O-embedded sections.

In order to further probe any potential ion suppression effects, a wider-reaching experiment is proposed. This would encompass tissue samples acquired from a range of subjects with a range of tissue types. Analysing such samples embedded in both media would enable the classification according to both tissue type and patient. Any potential interference could thus be placed in the context of inter-patient and inter-tissue type variation and hence permit a greater understanding of the effect of OCT in DESI.

Chapter 4

Optimisation of pre-DESI protocol for the analysis of FFPE samples

4.1 Introduction

4.1.1 Tissue fixation strategies

After fresh human tissue is collected by biopsy or during surgery for further pathological examination, it is typically very delicate and easily damaged or distorted. To allow the preparation of thin tissue sections that can be assessed by a histopathologist, these specimens need to be chemically preserved first. Two main strategies employed to allow tissue fixation is freezing tissue samples straight after removing them from a body or immersing them in 10% neutral buffered formalin or another fixative. However, only a limited number of defined clinical situations requires tissue specimens to be preserved by freezing. These mainly include only intraoperative consultations or muscle and renal biopsies [143]. Therefore, most tissue samples undergo an alternative treatment that allows the preparation of paraffin-embedded (FFPE) specimens. The whole process is summarised in Figure 4.1.

4.1.2 Advantages of FFPE samples

Although preservation as frozen tissue has long been regarded as the gold standard for samples that were to be used in research, in recent years, lots of efforts were put into making FFPE blocks accessible for this application. The recruitment of cryopreserved tissue samples in sufficient numbers allowing for robust study designs has always been challenging. Simultaneously, FFPE specimens represent an extensive collection of clinically annotated samples, mostly with known outcomes. Moreover, wax blocks can be stored at room temperature for many years and are considered an essential resource for historical studies in medicine [99]. Although FFPE samples offer an attractive alternative for the retrospective analysis of pathological processes, their use in research requires lots of optimisation due to how these tissue specimens are fixed and further processed.

4.1.3 Why is the use of FFPE samples in research a challenge?

Fixation in buffered formalin or another liquid fixing agent is only the first step of the procedure, leading to diagnosis based on histopathological examination. After this, tissue specimens are immersed in graded alcohols to remove the water before they can be infiltrated with wax. However, an intermediate solvent that is fully miscible with alcohol and paraffin wax must be used before this is possible. This solvent displaces alcohol in tissues, then it, in turn, is displaced by molten paraffin wax [100, 143]. All these steps forming a routinely used

specimen processing protocol contribute to the fact that using FFPE samples by widely applied mass spectrometry platforms can be challenging [99, 100].

4.1.4 Formalin fixation

Formalin fixation allows tissues preservation by forming methylene bridge cross-linkages between proteins, and extensive washes in alcohol and xylene contribute to the loss of tissue samples' chemical content. In recent years, several approaches have been reported to attempt to reverse formalin fixation, a process that is referred to as antigen retrieval [52, 144, 145]. This process typically relies on applying high-temperature treatment altogether with the use of a buffer solution. These steps are undertaken to return the tissue to its native state as closely as possible by reversing the protein cross-linking [52]. The development of effective and reproducible methods, allowing for antigen retrieval on FFPE samples, has contributed significantly to the standardisation of IHC protocols. As a result, it has recently allowed for proteomic and genomic analysis of a vast collection of archival clinical samples [146]. Studies were reporting the successful investigation of proteins from FFPE specimens [99, 145, 147, 148] by LC-MS/MS and metabolite imaging by MALDI [101, 102]. This project aims to test various pre-experimental treatments that would allow further DESI-MSI analysis of FFPE tissue samples.

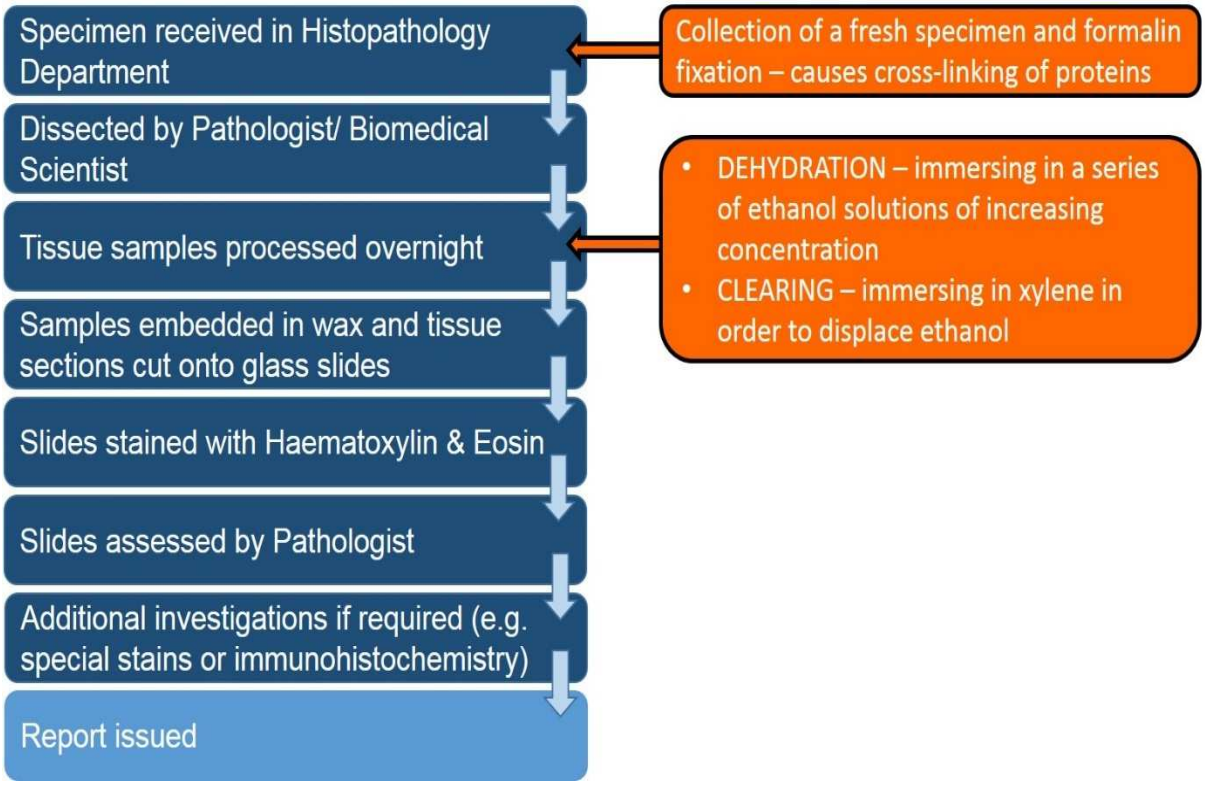


Figure 4.1. Typical workflow applied in routine histopathology laboratories for processing fresh tissue specimens fixed in formalin. Since many steps are needed in order to produce slides that can be assessed by a histopathologist, it can take many days before the final diagnosis is made. Steps contributing to making the use of FFPE samples by mass spectrometry platforms challenging are also highlighted.

4.2 Materials and methods

4.2.1 Tissue samples used for the optimisation

Since pork liver is very homogenous and gives a good signal intensity, all the initial tests were done using this readily available tissue type. It was purchased from a butcher and then divided into multiple pieces, each measuring about 1.5 cm x 1.5 cm x 1.5 cm.

In each case, 10µm thick tissue sections were prepared using a Thermo Fisher cryostat.

Mouse liver and human colorectal carcinoma FFPE samples were obtained from the Histopathology Department at St. Mary's Hospital, London, UK.

10 µm thick tissue sections were prepared using a microtome (Thermo Fisher Scientific Inc, Waltham, MA, USA), the sections were hydrated in a water bath before mounting onto

SuperFrost® Plus Glass slides (Thermo Fisher Scientific Inc) and incubating at 60°C for 30 minutes. The slides were then stored at room temperature until further analysis.

4.2.2 Pre-experimental sample preparation

Various pieces were subjected to different pre-experimental treatments, including the following:

- Fresh-frozen sample – it was snap-frozen straight away in a -80° freezer. It was then treated as a baseline in terms of the signal intensity and, even more importantly, the spectra's composition.
- Formalin-fixed (FF) samples – straight after dissection, fresh pork liver pieces were fixed in 10% Neutral Buffered Formalin for 24 hours. 60 ml prefilled pots were ordered from CellPath (catalogue number BAF-6000-08A). Formalin fixation is routinely used in Cellular Pathology laboratories as a part of the diagnostic protocol. Next, the samples were subjected to the following conditions:
    - Sucrose was tested as an alternative stabilising method. A FF pork liver sample was rinsed with distilled water and then soaked overnight in a sucrose solution (30% w/v) in distilled water. BioXtra ≥ 99.5% Sucrose was obtained from Sigma-Aldrich (St. Louis, MO, USA), catalogue number S7903. The following day, the sample was removed from the solution and rinsed with distilled water, and frozen at -80°C before cryosectioning.

      Another sample was subjected to the same treatment but was embedded in gelatin (10% w/v; catalogue number 4078; Merck, Darmstadt, Germany). Before this step, gelatin was heated to 50°C for about 10 minutes until a homogeneous solution was achieved. The embedded sample was frozen at -80°C before cryosectioning.
    - Glycerol, which should have similar protection properties as sucrose, was also tested. It was supplied by Sigma-Aldrich (St. Louis, MO, USA), catalogue number 49781. Two concentrations were tested – 0.1% and 5%. Tissue samples were soaked in solutions overnight and then were embedded in 10% gelatin and frozen at -80°C.
    - Urea was tested as an antigen retrieval treatment. Urea powder bioreagent for molecular biology was supplied by Sigma-Aldrich (St. Louis, MO, USA), catalogue number U5378. Three concentrations were tested: 1M, 5M and 10M.

FF pork liver was heat-treated at 100^0C for 4 minutes in urea solutions and sectioned onto glass slides.

- Sodium metabisulfite (Molecular Dimensions, catalogue number TRC-S667060) was tested as an antigen retrieval treatment. FF pork liver was heat-treated with sodium metabisulfite aqueous solution at either 15 or 40 % (w/v) at 95^0C for 2 minutes. The pork liver was then sectioned as above.
- Denator, The Stabilizor™ T1, was used for the experiments requiring heat treatment of pork liver tissue samples.

Four different solvents were also tested to assess their abilities to remove paraffin from tissue sections. These were as follows:

- Pentane (Chromasolv, Sigma Aldrich, St. Louis, MO, USA)
- Hexane (SupraSov, Merck KGaA, Darmstadt, Germany)
- Chloroform (≥99.8%, Sigma Aldrich, St. Louis, MO, USA)
- Xylene (Histological grade, Sigma Aldrich, St. Louis, MO, USA)

Before DESI-MSI analysis, the slides were washed for 8 minutes twice in each of the solvents listed above. Next, the samples were left to dry at room temperature for 24 hours.

4.2.3 Strategies for method optimisation

FF samples were subjected to various treatments to assess their effectiveness in assisting with future DESI-MSI analysis.

One of the challenges was the fact that FF tissues tended to become very brittle when frozen and therefore were hard to section. This issue is usually overcome by embedding samples. Therefore, sucrose and glycerol were tested as alternative stabilising agents. This set of experiments was performed to address the issues with cryosectioning samples prior to DESI-MSI.

Secondly, tissue fixation in formalin was known to result in the formation of methylene bridge cross-linkages between proteins. In order to attempt to reverse formalin fixation during a process referred to as antigen retrieval, several approaches were tested. These included heat treatment and neutralisation of formaldehyde with sodium metabisulfite and urea.

In the first phase, all the experiments described above were performed on FF samples that were not embedded in wax. If any of the tested treatments were successful, they would be applied to

FF samples that had been put in the form of wax blocks. That would require the successful removal of wax prior to repeating the experiments.

4.2.4 DESI-MSI analysis

All DESI-MSI experiments were performed on a Xevo G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA) coupled to a 2D sample stage (Prosolia Inc., Indianapolis, IN, USA).

100 μm^2 pixel size and 100 μm step size were set up for imaging analysis at a scan rate of 1 scan/second. Methanol: water, 95:5, v/v (Sigma-Aldrich, St. Lewis, Missouri, USA) was used as the electrospray solvent at a flow rate of 1.5 μl/min. Zero-grade nitrogen was used as the nebulising gas at a pressure of 7 bar. All imaging analysis was performed using 50 to 1500 mass range.

Full details of the DESI-MSI setup can be found in *Chapter 2.9*.

Regions of interest were selected with High Definition Imaging software (HDI, v 1.4, Waters Corporation) for analysis in positive and negative ion modes.

Parameters	**Xevo G2-XS QToF (Waters Corporation, Milford, MA)**	
Scan time	1 sec	
Scan mode	Sensitivity	
Mass analyser	TOF	
Mass range	50-1500 *m/z*	
Ionisation mode	Negative	Positive
Sampling cone voltage	-40 V	40 V
Source offset	-80 V	80 V
Source temperature	120°C	120°C
Spray voltage	4.5 kV	4.5 kV

Table 4.1. Xevo G2-XS QToF parameters used for the optimisation of the DESI-MSI protocol for the analysis of FFPE tissue samples.

Following DESI-MSI, the analysed tissue sections were stained with H&E, and optical images were acquired.

4.2.5 Data analysis

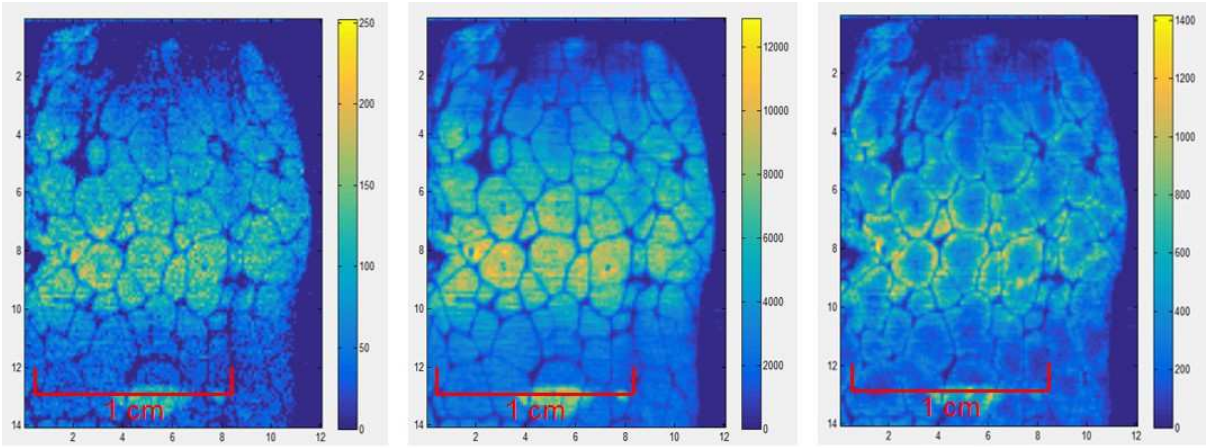
Imaging raw files (.DAT) were uploaded into a custom-build imaging toolbox in MatLab (R2014a, MathWorks, Natick, MA, USA) environment and followed by the optical images' upload.

Methodological details of data analyses are described in *Chapter 2.13*.

4.3 Results

4.3.1 Fresh-frozen pork liver in OCT

Conditions that were the closest to those that have routinely been employed in diagnostic laboratories when dealing with fresh-frozen samples were tested first. Fresh pork liver was embedded in OCT, sectioned and imaged in negative ion mode without any previous sample treatment. Clear morphological structure (Figure 4.2) was observed, including for ions with signal intensities as low as 250 counts, *e.g. m/z* 726.5.



m/z 726.54 *m/z* 766.53 *m/z* 742.51

Figure 4.2. Comparison of DESI-MSI results in negative ion mode for morphological structures for different signal intensities for fresh pork liver embedded in OCT. *m/z* 726.54 – PE(P-36:2); *m/z* 766.53 – PE(38:4); *m/z* 742.51 – PE(36:2).

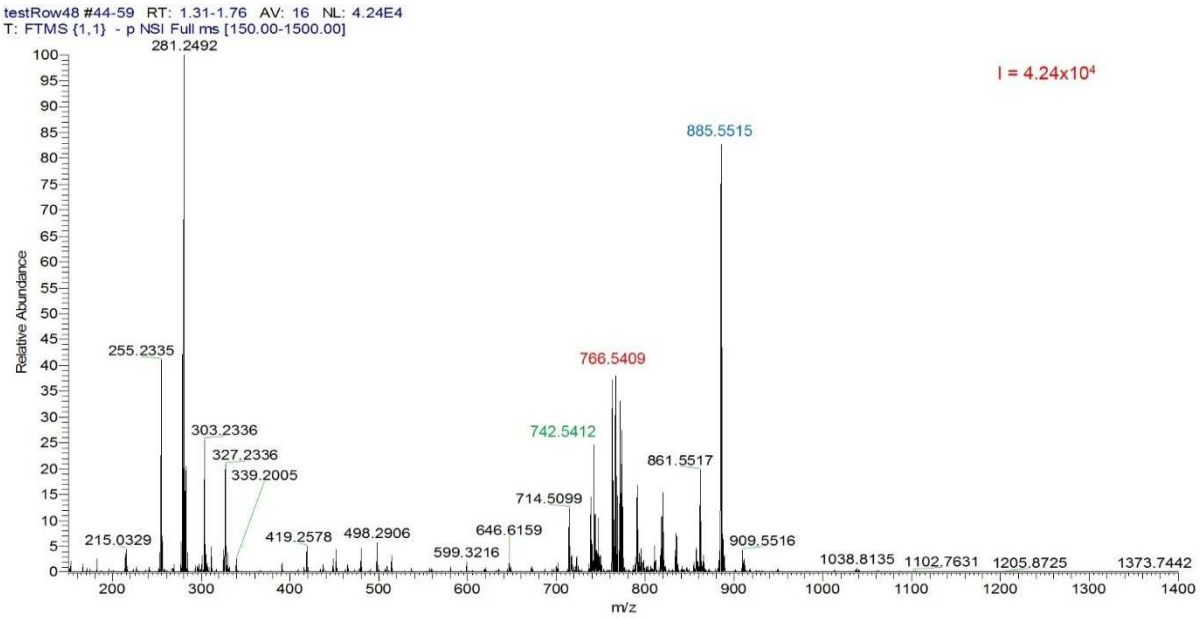


Figure 4.3. An example of an averaged mass spectrum acquired in negative ionisation mode for fresh pork liver embedded in OCT.

As can be seen in Figure 4.3, the right intensity spectrum was recorded, demonstrating a good background: phospholipids peak ratio. It was treated as a baseline to assess further attempts to optimise the pre-DESI-MSI sample preparation protocol to analyse FFPE tissue samples. The three most intense peaks related to pork liver: *m/z* 742.54, 766.54 and 885.55, have been highlighted and used as signature peaks for future experiments.

4.3.2 Removing excess formalin by rinsing the sample in water

Firstly, the FF pork liver tissue sample was washed in LC-MS grade water to remove the excess formalin present in the tissue. It was then analysed by DESI-MSI, and the results were compared with a fresh-frozen liver sample. It was observed that the signal intensity was higher (1.19×10^4) for the fresh-frozen pork liver than for the FF pork liver (7.23×10^3) (Figure 4.4). Also, different phospholipid species were observed in each sample treatment (Figure 4.4). No morphological features could be discerned despite the high signal intensity for the FF pork liver (Figure 4.5).

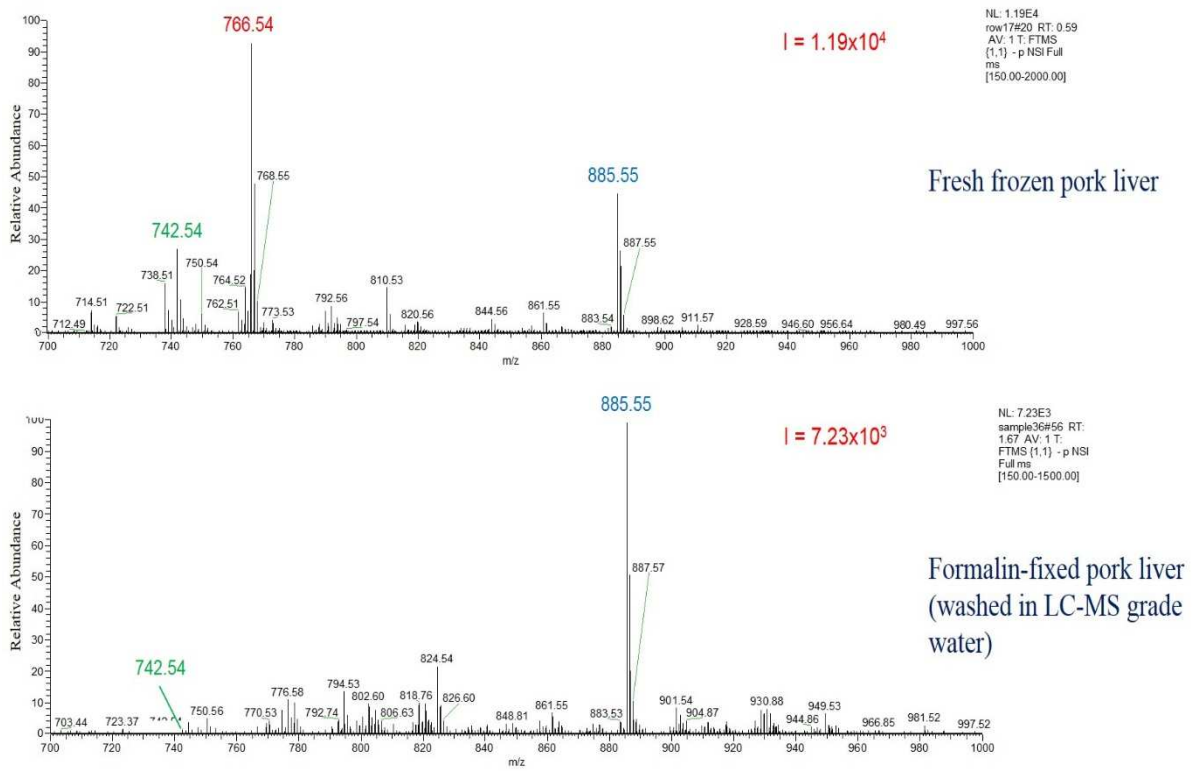


Figure 4.4. Comparison of spectra intensity in negative ion mode for fresh-frozen pork liver and FF pork liver washed in LC-MS grade water.

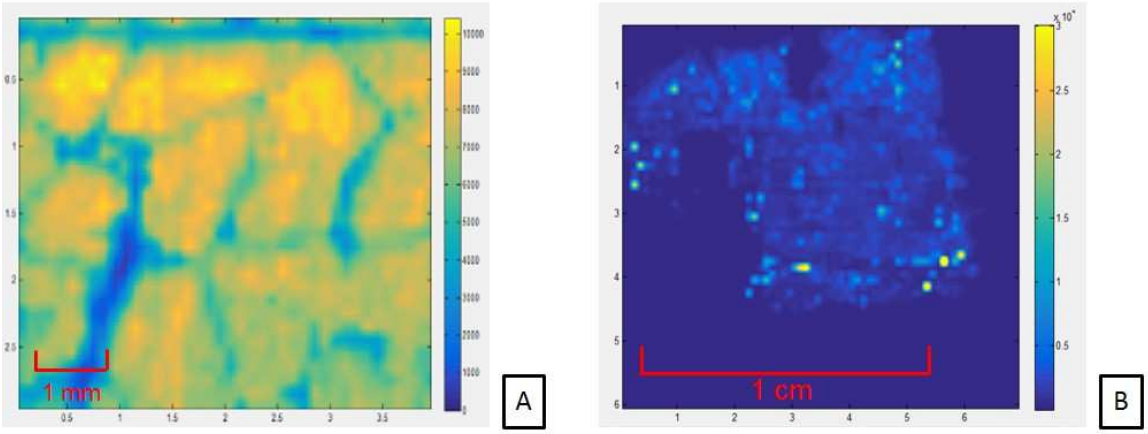


Figure 4.5. Comparison of DESI-MSI results in negative ion mode for fresh-frozen pork liver (A) and FF pork liver washed in LC-MS grade water (B). Ion images of phosphoinositol (PI) 38:4 (*m/z* 885.5499).

4.3.3 Sucrose as an alternative stabilising agent

A significant challenge in making DESI more histologically friendly is making it accessible to analyse FF tissue. FF tissues tend to become very brittle when frozen and therefore are hard to

section. This issue is usually overcome by embedding samples. However, as discussed in Chapter 3, embedding can significantly impact mass spectra and lipid species conservation. It was reported that a sucrose infiltration step significantly contributed to the improvement of the quality of tissue sections prepared from whole insects [149]. Therefore, sucrose cryoprotection was investigated as an alternative stabilising method. In order to test this, FF pork liver was rinsed and then soaked overnight in a 25% sucrose solution. It was later removed, rinsed again, and frozen for sectioning. 10% gelatin was used as an embedding medium, as it has been reported to help with cryosectioning of some other fragile tissue samples like crab's organs and brain [131]. The tissue section was analysed in both positive and negative ion modes.

Comparison of spectra from these sections with spectra from FF, OCT embedded sample showed that spectral intensity was slightly better but still not as good as for fresh-frozen tissue. There were also differences in spectral composition, which could probably be attributed to shifted adduct formation. Sucrose was clearly visible in positive ($[M+H]^+$, $[M+Na]^+$) ion spectra. No morphological structure of the analysed sample could be seen (Figure 4.6A).

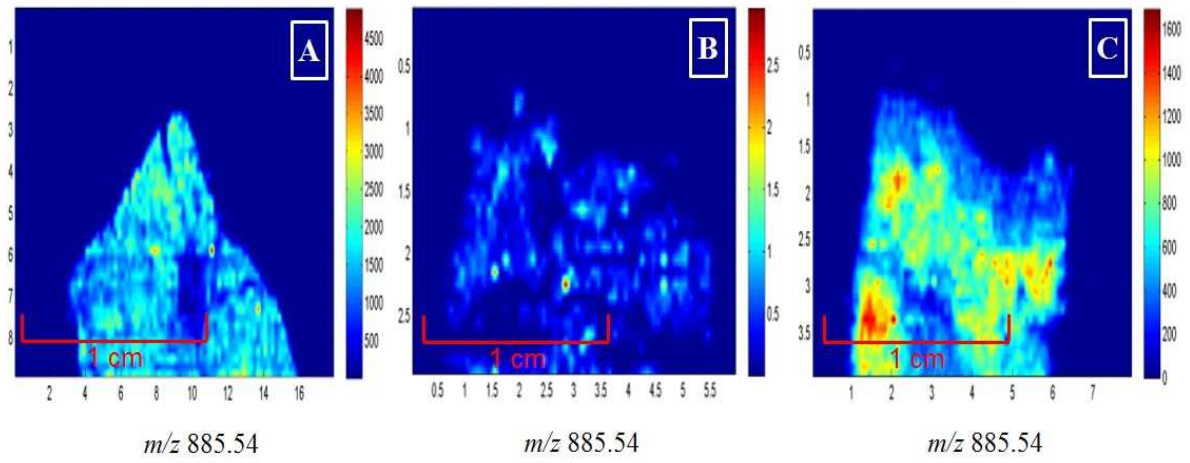


Figure 4.6. MSI in negative ion mode of FF pork liver treated with Sucrose 25% (A) or Glycerol (B - 0.1%; C - 5%). Ion images of phosphoinositol (PI) 38:4 (*m/z* 885.5499).

4.3.4 Glycerol as an alternative cryoprotection agent

As the next step, glycerol was tested as an alternative cryoprotection agent. It was hypothesised that it would show similar protection properties as sucrose. In addition to this, glycerol was thought to be more compatible with formaldehyde scavenging protocols as it did not contain an aldehyde group.

Two different glycerol concentrations - 0.1% and 5% - were used as a pre-embedding treatment for the FF pork liver samples, which were then embedded in 10% gelatin and analysed in both negative and positive ionisation modes.

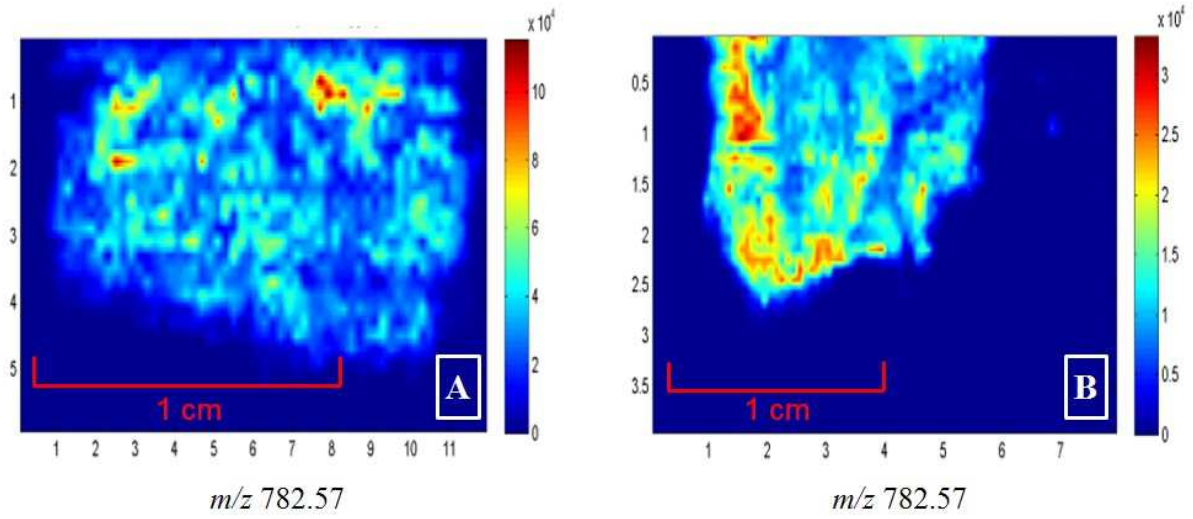


Figure 4.7. MSI in positive ion mode of FF pork liver soaked in Glycerol (A - 0.1%; B - 5%). Ion images for *m/z* 782.5770 (PC 34:1[M+Na]$^+$).

As shown in Figures 4.6 and 4.7, lower ionisation intensity was observed for the cryoprotectant 0.1% glycerol compared to 5% glycerol and 25% sucrose in both ionisation modes.

Figures 4.8 and 4.9 summarise the outcome of testing the following two media – 30% sucrose + 10% gelatin and 5% glycerol + 10% gelatin. The tissue morphology was best preserved for the fresh pork liver when a mix of glycerol + gelatin was used; however, some structures could also be seen for the other medium. In both cases, FF pork liver tissue sections did not show any morphological features. The one treated with glycerol + gelatin appeared very destroyed and mainly blown away from the glass slide (Figure 4.8).

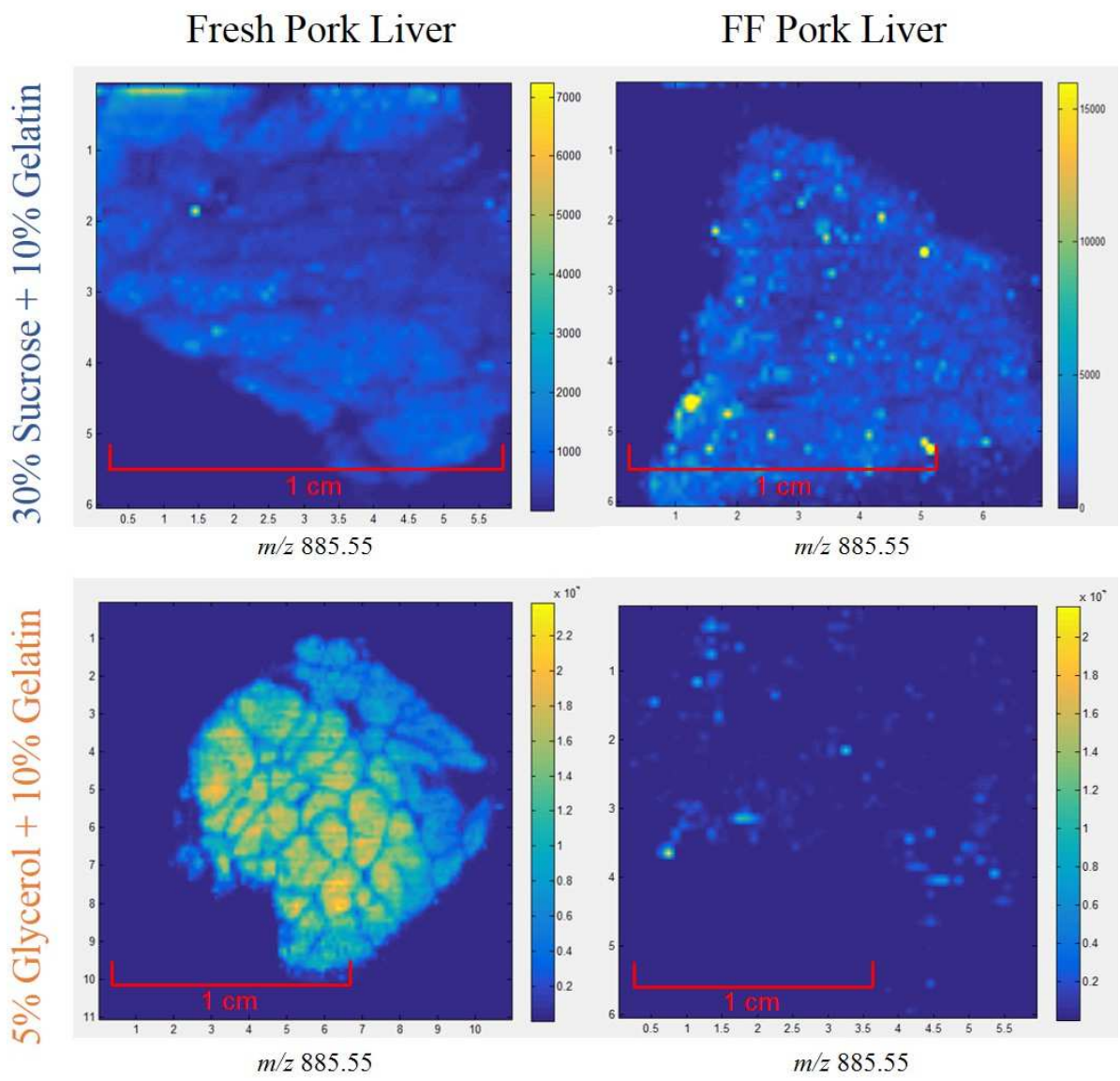


Figure 4.8. Comparison of post-DESI morphological structures of analysed samples for both fresh and FF pork liver embedded in gelatin with either 30% sucrose (top row) or 5% glycerol (bottom row). Ion images of phosphoinositol (PI) 38:4 (*m/z* 885.5499).

Figure 4.9 shows that averaged spectra recorded for the fresh pork liver looked very similar after both treatments. Signal intensity was slightly higher for the sample treated with glycerol, but otherwise, the spectral composition and the intensities of the signatures peaks were almost the same. However, different phospholipid species were seen in the case of FF liver samples. Significantly, the data recorded for the piece treated with glycerol showed many more peaks than in the fresh liver samples.

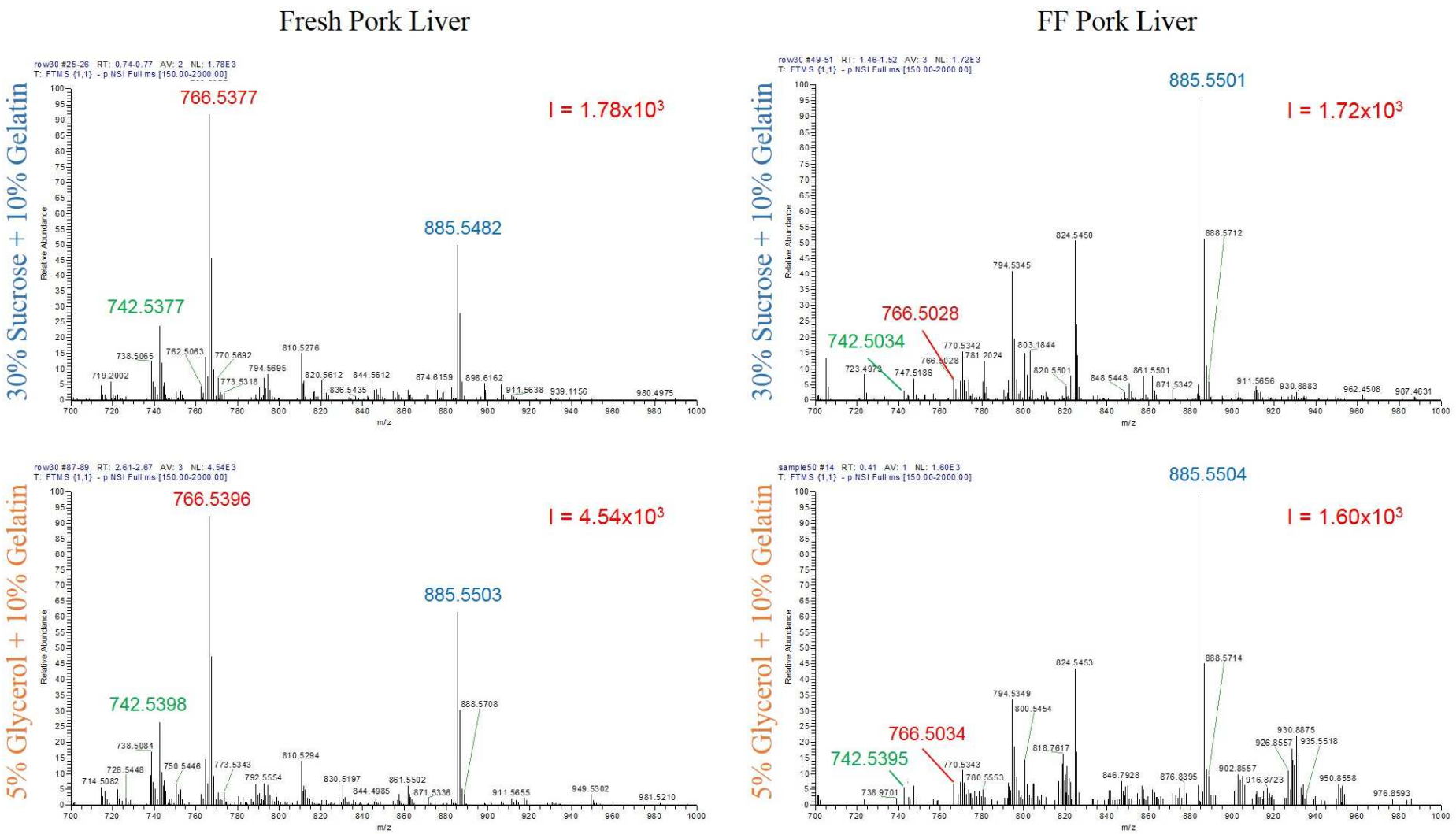


Figure 4.9. Comparison of DESI-MSI results for both fresh and FF pork liver embedded in gelatin with either 5% glycerol or 30% sucrose. Averaged spectra for fresh and FF pork. All the results shown are for the negative ionisation mode data.

4.3.5 Formalin and glycerol tissue sample fixation

Next, fresh pork liver was fixed either in formalin with 0.1% glycerol or formalin with 5% glycerol for 72 hours at room temperature. The tissues were embedded in 10% gelatin and OCT and sectioned onto glass slides. However, when the tissue section was analysed in negative ion mode, morphological structures were barely discernable, despite a relatively high signal intensity (Figure 4.10).

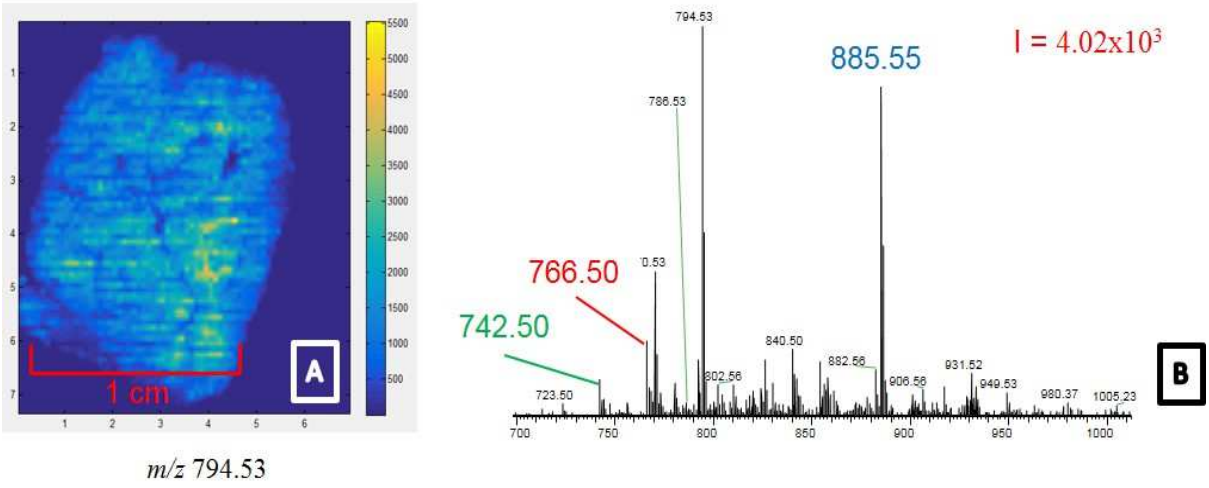


Figure 4.10. DESI-MSI analysis of pork liver fixed in formalin + 5% glycerol embedded in OCT. Signal intensity in negative ion mode (A), and averaged spectrum focusing on the phospholipids mass range (B). Ion image for *m/z* 794.535.

4.3.6 Heat treatment as an antigen retrieval method

It has been reported that the adaptation of heat-induced antigen retrieval methods has significantly contributed to the development of FFPE proteomics [150-152].

Therefore, experiments were performed using heat treatment and formalin scavengers (*e.g.* urea) to reverse formalin action. FF pork liver was heat-treated at 100^0C for 4 minutes in water or 10M urea solution and sectioned onto glass slides for DESI-MSI analysis in negative ion mode. It was observed that the FF liver sample heat-treated in the water had a higher signal for fatty acids ($1.24x10^5$) by two orders of magnitude when compared to fresh-frozen pork liver (Figure 4.11). This effect was also observed when comparing FF pork liver and heat-treated FF pork liver. Again, an increase by order of magnitude was detected, while morphological features were beginning to be discernable (Figure 4.12).

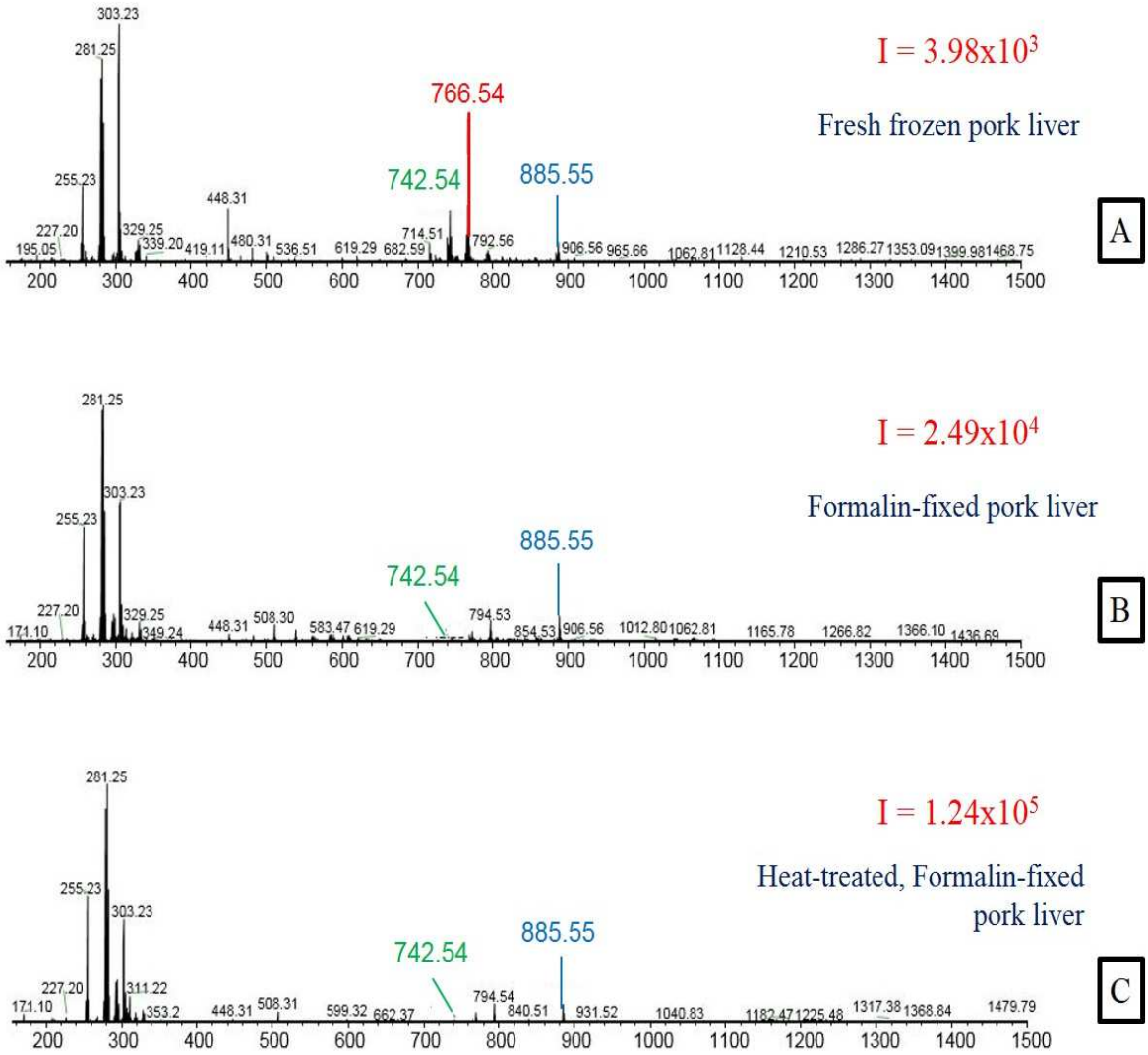


Figure 4.11. Comparison of spectra intensity in negative ion mode for fresh-frozen pork liver (A), FF pork liver (B), heat-treated, FF pork liver (C).

As shown in Figure 4.11, the spectral composition for FF and heat-treated FF samples was very similar, with only two signature peaks, *m/z* 742.54 and 885.55, being detected. The most intense peak present in the fresh pork liver sample – *m/z* 766.54 – was not seen.

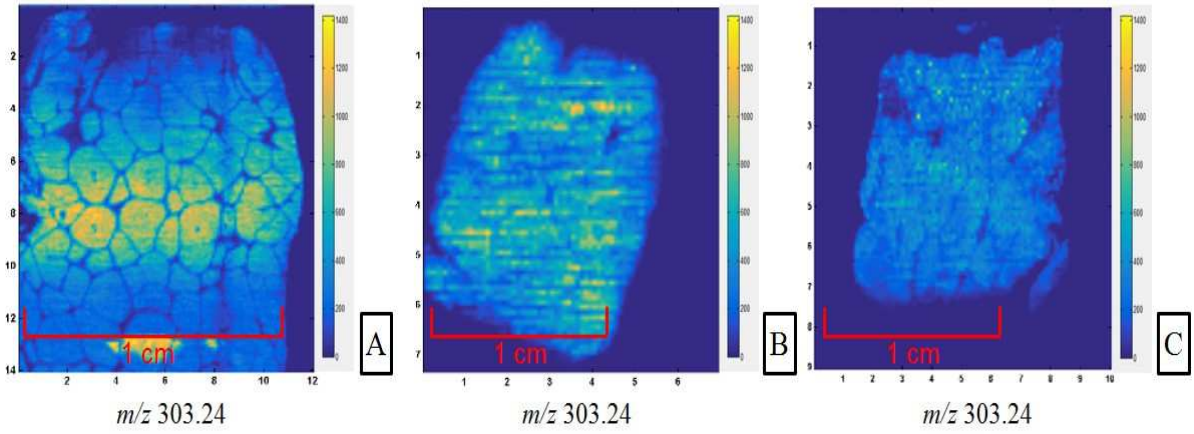


Figure 4.12. Comparison of DESI-MSI in negative ion mode for fresh-frozen pork liver (A), FF pork liver (B) and heat-treated FF pork liver (C). Ion images of arachidonic acid (20:4; *m/z* 303.24).

After these experiments, the critical observation was that urea/ heat-treated FF pork liver could not be sectioned as the tissue became too soft after-treatment. Therefore, further work needed to be performed using other formalin scavengers and other heat-treatment conditions to understand if formalin's action can be effectively reversed in an attempt to obtain morphological information from FF tissue.

4.3.7 Sodium metabisulfite as antigen retrieval treatment

Further experiments were performed regarding the removal of formalin from FF tissues. Since the neutralisation of formaldehyde with sodium metabisulfite has been reported [153, 154], it was tested as a formalin removal reagent in the FF pork liver. Two concentrations of sodium metabisulfite were tested (15 and 40 % (w/v)). FF pork liver was heat-treated with sodium metabisulfite aqueous solution at either 15 or 40 % at 95^{0}C for two minutes. After this, the tissue block was embedded in OCT, sectioned and analysed by DESI-MSI in negative ion mode.

It was observed that the signal intensity for the phospholipid mass range for samples heat-treated with sodium metabisulfite was high. In contrast, the fatty acid: phospholipid ratio was low, as observed in other heat treatment experiments. No morphological features were detected (Figure 4.13).

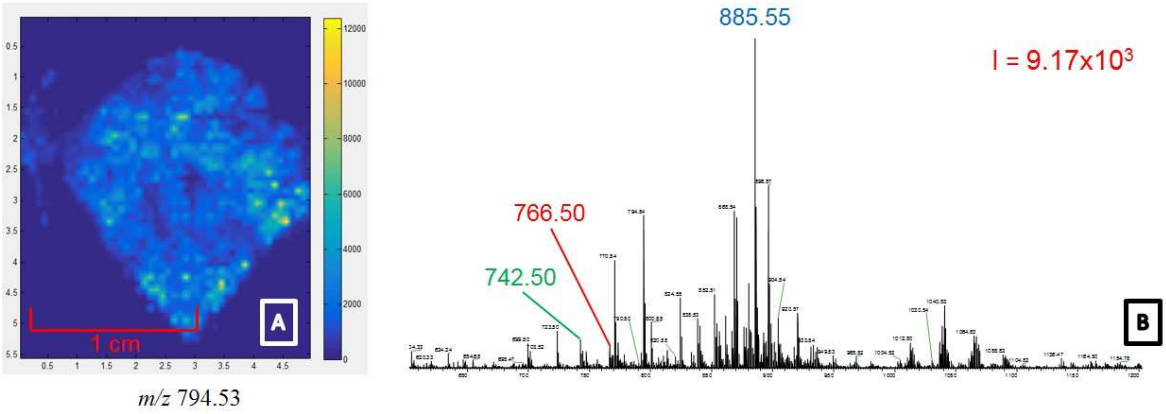


Figure 4.13. Spectra intensity in negative ion mode for FF pork liver heat-treated with 15% (w/v) sodium metabisulfite (B). Ion image for *m/z* 794.535 (A).

This phenomenon was surprising as before analyses, and extensive morphological features were visible to the naked eye; however, in the DESI-MSI image, no morphological features were detected. In order to try and understand this, the tissue sections were examined under the microscope. Before analysis, it was observed that morphological features could be seen in the tissue section, but after analysis, these seem to have been destroyed (Figure 4.14).

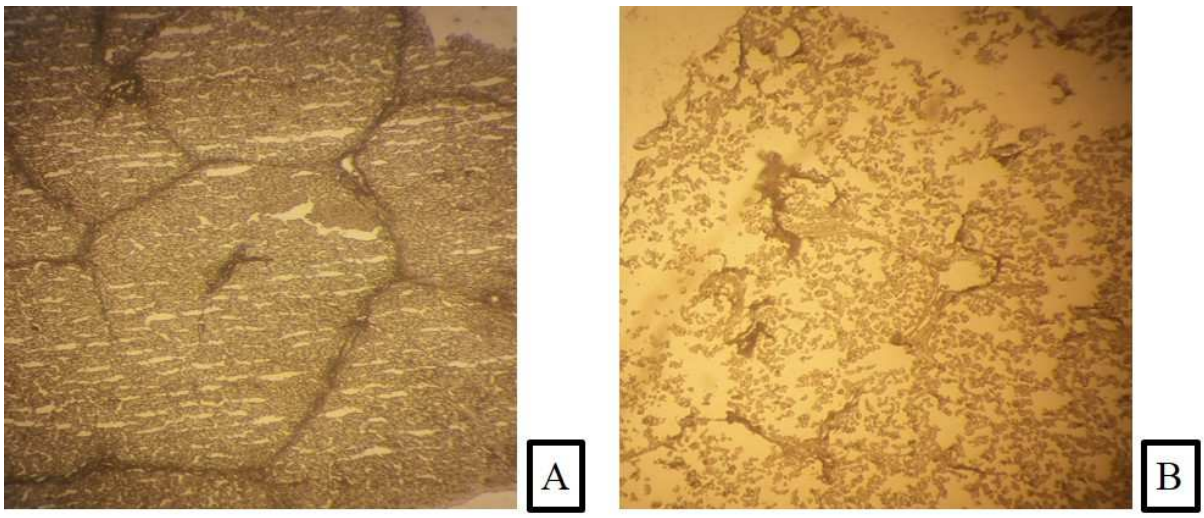


Figure 4.14. Microscope images for FF pork liver heat-treated with sodium metabisulfite. Before (A) and post (B) DESI-MSI analyses.

In order to overcome this, thicker tissue sections (20µm) were prepared and analysed by DESI-MSI. However, the same phenomenon was observed. The effect of the gas flow (7 bar gas pressure) was also examined by analysing a sample with only 3 bar gas pressure, and it was noted that morphological features were much more apparent after the experiment.

4.3.8 Various numbers of heat-treatment combined with different glass slides types as another attempt to reverse the effects of formalin fixation

It was hypothesised that the tissue section did not adhere to the glass slide as firmly as frozen sections do. This might be because formalin cross-linking of proteins was not reversed to a sufficient extent to promote enough ionic interactions between the proteins and the glass slide. To improve these ionic interactions, the time of heat-treatment was increased (heat-treatment cycles), urea was used as a formalin scavenger, and microscope adhesion slides were also tested (Table 4.2).

Formalin scavenger	Concentrations	Heat-treatment (Denator)	Microcope slides
Water	-	1 cycle	SF
Sodium metabisulfite	15%, 40%	1, 3 cycles	SF,SFP
Urea	1M, 5M, 10M	1, 3, 5, 10 cycles	SF,SFP, PL, SFPU, SFG

Legend: SF-SuperFrost; SFP-SuperFrostPlus; PL-Poly-Lysine; SFPU-SuperFrostPlus Ultra; SFG-SuperFrost Gold

Table 4.2. Summary of the formalin scavengers and heat-treatment times used for FF pork liver tissue to reverse formalin protein cross-linking bonds.

It was observed that when the number of heat-treatment cycles was increased for the treatment of FF pork liver, the spectra showed a difference in the ratio of fatty acids: phospholipids (Figure 4.15). However, no morphological features were detected, and post-DESI-MSI analysis, the tissue section was still destroyed even when using adhesion slides SuperFrost Plus (Figure 4.16).

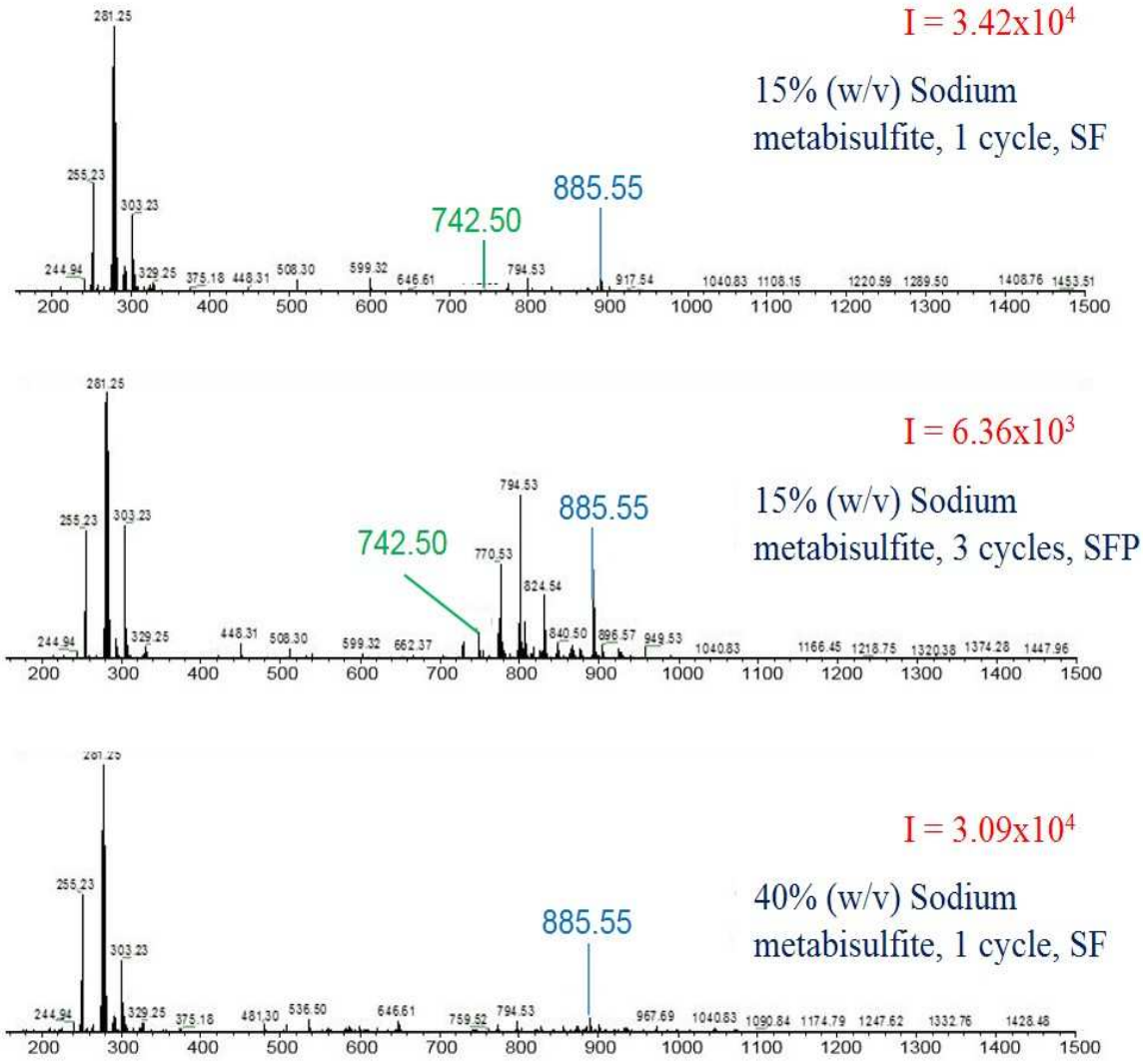


Figure 4.15. Comparison of spectral intensity in negative ion mode for FF pork liver using sodium metabisulfite (15% and 40% (w/v) and heat-treatment (1 and 3 treatment cycles) as formalin scavengers. SF-SuperFrost glass slides; SFP-SuperFrostPlus glass slides.

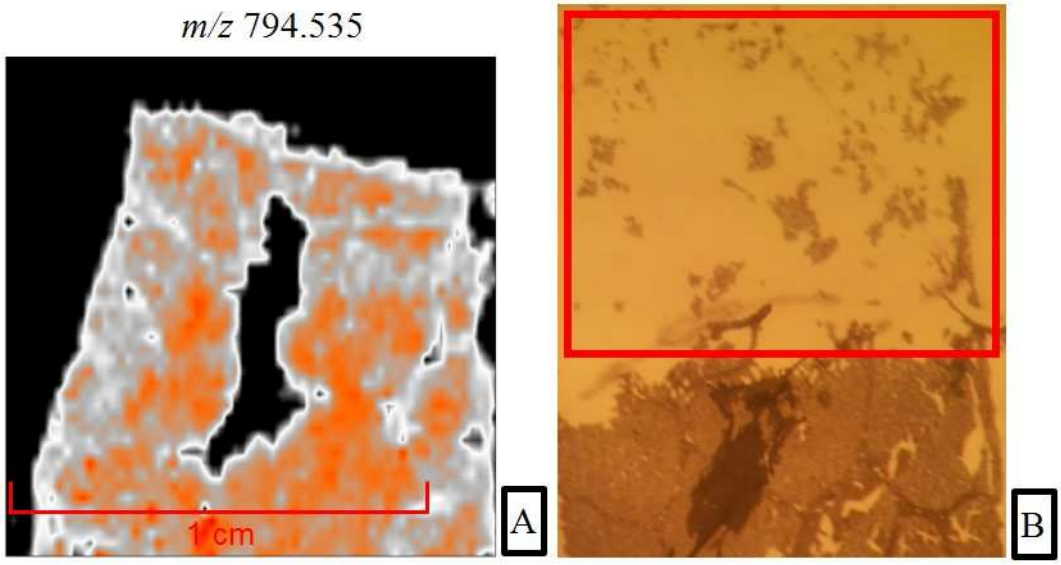


Figure 4.16. FF pork liver heat-treated for 3 cycles with 15% (w/v) sodium metabisulfite after DESI-MSI analysis in negative ion mode. Ion image (*m/z* 794.535) (A); microscope image of tissue after DESI-MSI (red square) (B).

4.3.9 Urea as a formalin scavenger

As no morphological features were detected so far, urea was tested as a formalin scavenger. Different concentrations were prepared: 1M, 5M and 10M and various heat-treatment cycles and adhesion slides were also applied. It was observed that urea seemed to have more effective formalin reversal properties than sodium metabisulfite as the fatty acid: phospholipid ratio was improved with all urea concentrations tested (Figure 4.17). That could be because urea is a chaotropic agent, and such a molecule can disrupt the hydrogen bonding [155, 156]. Regarding the number of cycles tested, it appeared that 3 and 5 cycles were the most effective to reverse the cross-link bonds of FF tissue proteins.

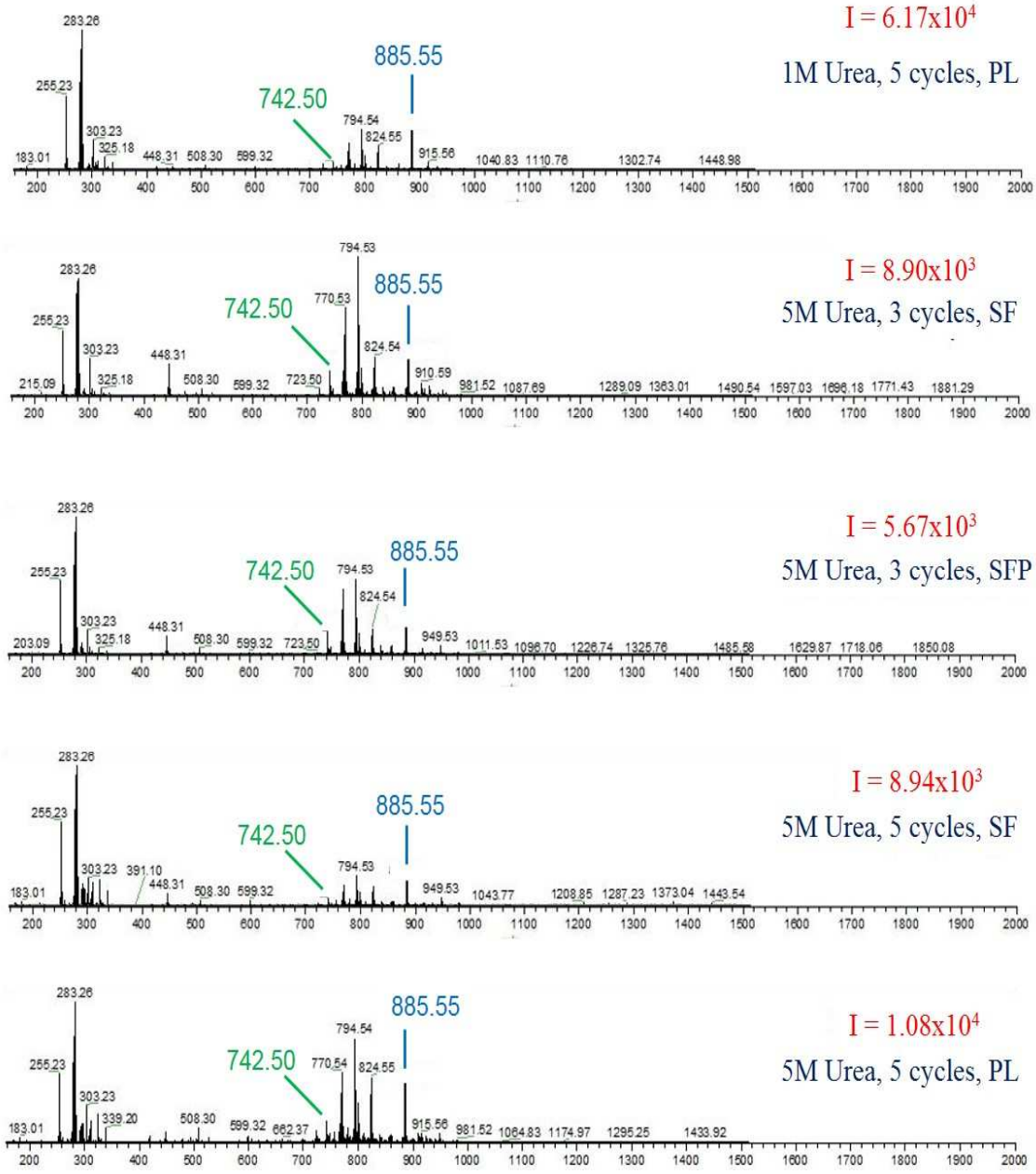


Figure 4.17. Comparison of spectra intensity for FF pork liver using urea and heat-treatment as formalin scavengers. Different concentrations of urea, as well as various heat treatments, were tested**.** Data acquired in negative ionisation mode. Types of glass slides used: SF-SuperFrost; SFP-SuperFrostPlus; PL-Poly-Lysine.

The comparison between standard histology slides (SuperFrost) and adhesion slides (SuperFrostPlus, Poly-Lysine, SuperFrostPlus Ultra and SuperFrost Gold) showed no improvement in the adhesion of the FF pork liver tissue sections (Figure 4.18).

Figure 4.18. Microscope image of FF pork liver post-DESI-MSI analysis using different types of glass slides. 5M urea, 3 heat-treatment cycles, SuperFrost slide (A); 5M urea, 3 heat-treatment cycles, SuperFrostPlus slide (B); 5M urea, 5 heat-treatment cycles, Poly-Lysine slide (C).

No morphological features were discernable when using sodium metabisulfite or urea, except when urea at 5M concentration and three heat-treatment cycles were applied (Figure 4.19). However, method development needed to be performed to further overcome tissue destruction and improve tissue morphological feature detection.

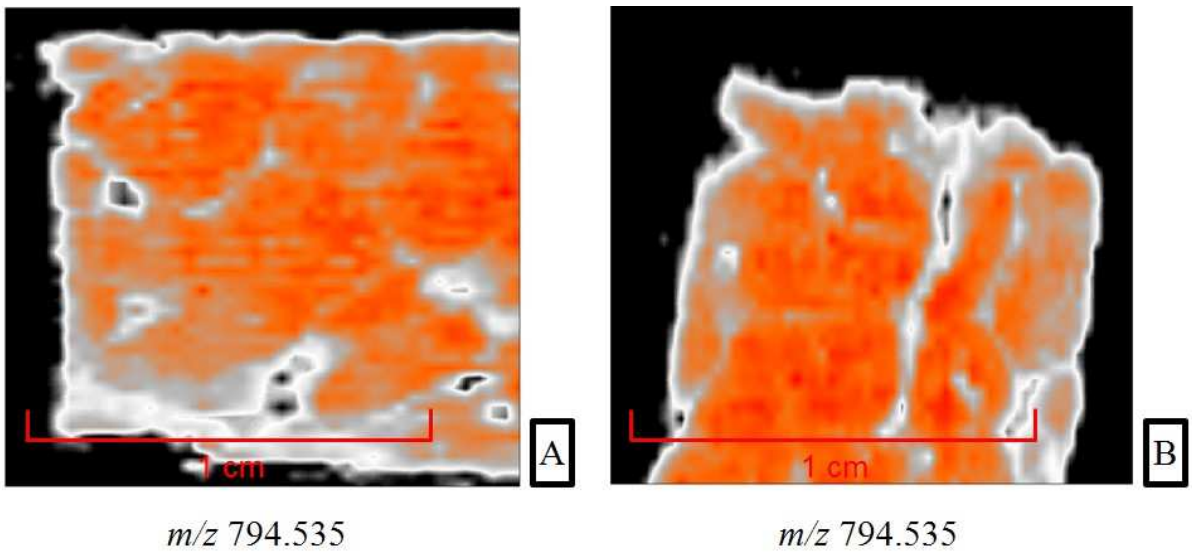


Figure 4.19. Ion images for FF pork liver heat-treated for 3 cycles with 5M Urea after DESI-MSI analysis in negative ion mode. Ion image using SuperFrost slide (A) and SuperFrostPlus slide (B).

4.3.10 Conclusions from the experiments performed so far

Up to this point, none of the pre-experimental treatments tested gave satisfactory results. Not only spectra composition and signal intensity were taken into account, but also the ability to detect morphological features from the analysed tissue sections. Even if it were possible to acquire sufficient chemical information during DESI-MSI measurements, post-experimental quality of tissue sections would be equally important. In order to allow for further data analysis, all tissue sections needed to be H&E stained and assessed by a qualified histopathologist. Without this input and the ability to match chemical information to various tissue classes, the DESI technique would lose most of its potential. Therefore, it was clear that further optimisation was required if FFPE samples were to be compatible with DESI-MSI.

A new paper was published around this time that claimed that the *in situ* imaging of metabolites from FFPE human tissue samples using MALDI-MSI was successful [101]. The ionisation process is not the same when we compare MALDI-MSI and DESI-MSI, and the first technique requires the use of a matrix, which results in the creation of matrix ions artefacts. Nevertheless, it was hypothesised that the same pre-experimental treatment could be tested and applied to DESI-MSI.

4.3.11 A new approach to histologically friendly DESI-MSI

The pre-DESI protocol for MALDI-MSI analysis of FFPE human tissue samples was very straightforward. Tissue sections were prepared using a microtome, then incubated for 1h at 70°C. The only additional step after the incubation was deparaffinisation in xylene 2 x 8 minutes. After this, the slides were allowed to air-dry for 24 hours before the matrix application and MALDI-MSI measurements [101].

It was hypothesised that the same treatment might be useful for DESI-MSI of FFPE tissue samples' metabolic content.

4.3.12 Deparaffinization of FFPE tissue sections using different solvents

Four different solvents were tested during this experiment – Pentane, Hexane, Chloroform, and Xylene. Tissue sections were prepared as described in *Chapter 2.7.2*. Prior to DESI-MSI, each slide was washed 2 x 8 minutes in one of the solvents mentioned above. Then, the slides were left to dry at room temperature for 24 hours.

As can be seen in Figure 4.20, after the standard DESI-MSI experiments, the section that was deparaffinised in xylene was intact, and morphological features were very well preserved. The

most destroyed section was rinsed in hexane, while both pentane and chloroform gave comparable results.

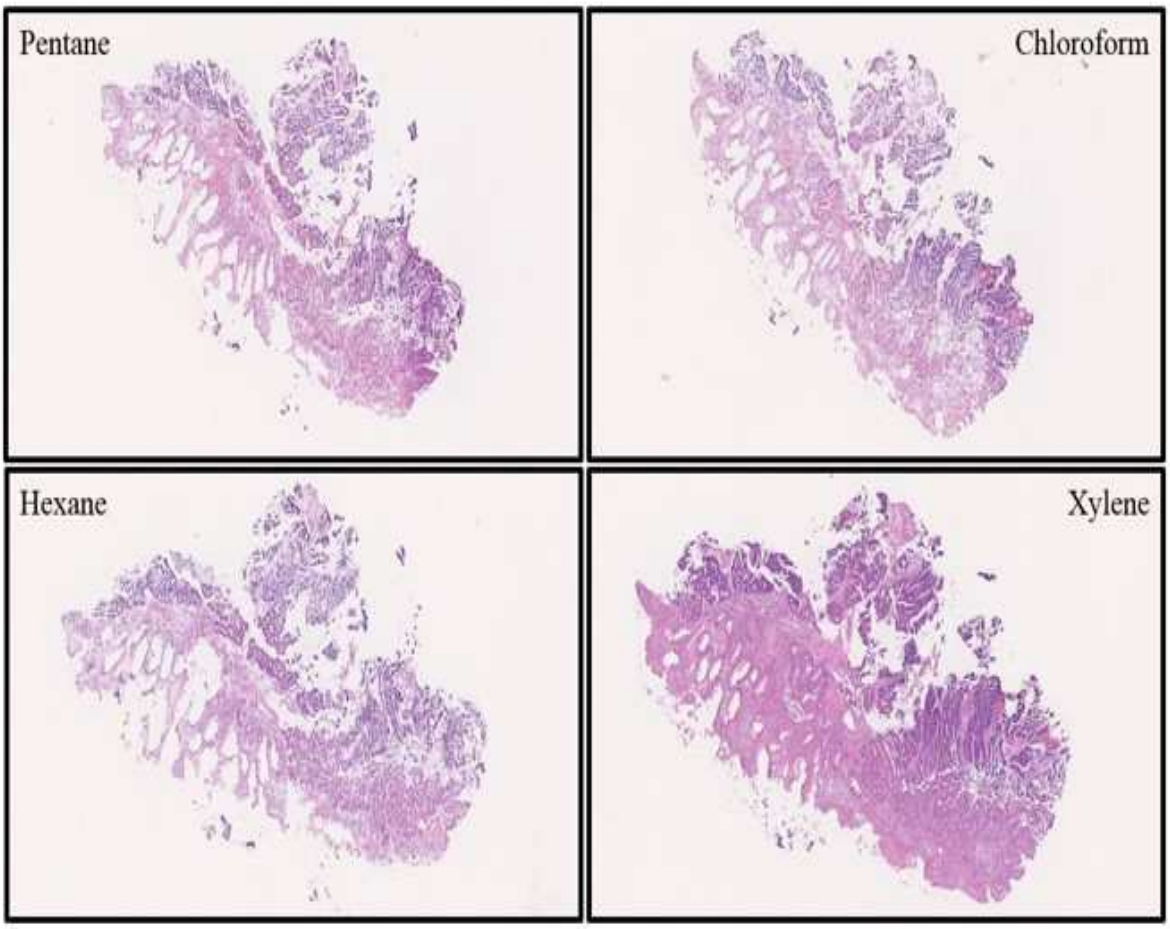


Figure 4.20. Quality of sections prepared from the same sample after DESI-MSI. The slides were subjected to deparaffinisation using four different solvents before standard DESI-MSI measurements. After the analysis, the slides were H&E stained to assess the quality of the analysed tissue sections. Optical images represent four tissue sections cut from the same FFPE block but rinsed in different solvents before the DESI-MSI measurements.

As expected, the quality of the ion images followed the same pattern. As can be seen in Figure 4.21, the positive mode data appeared to contain much more information.

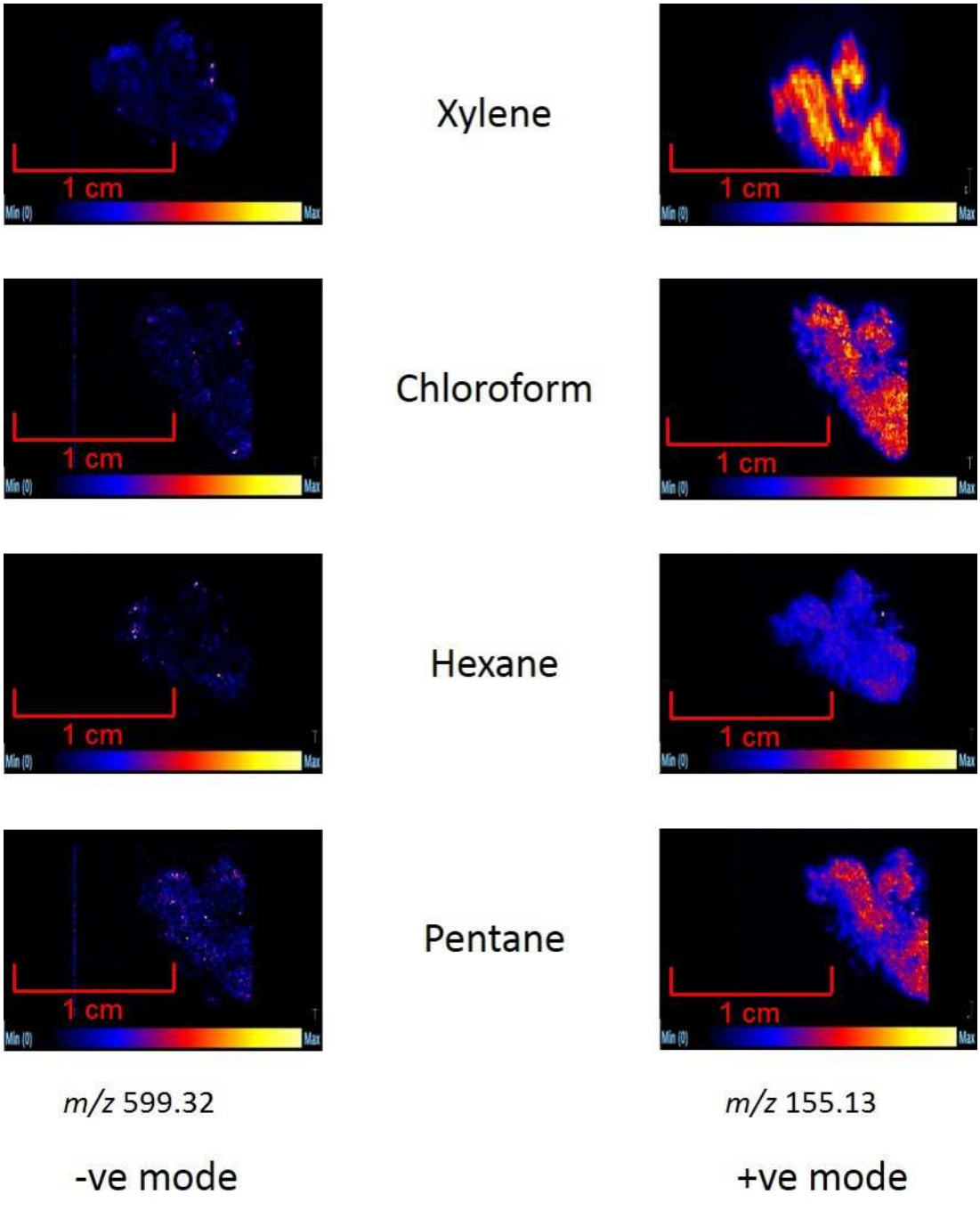
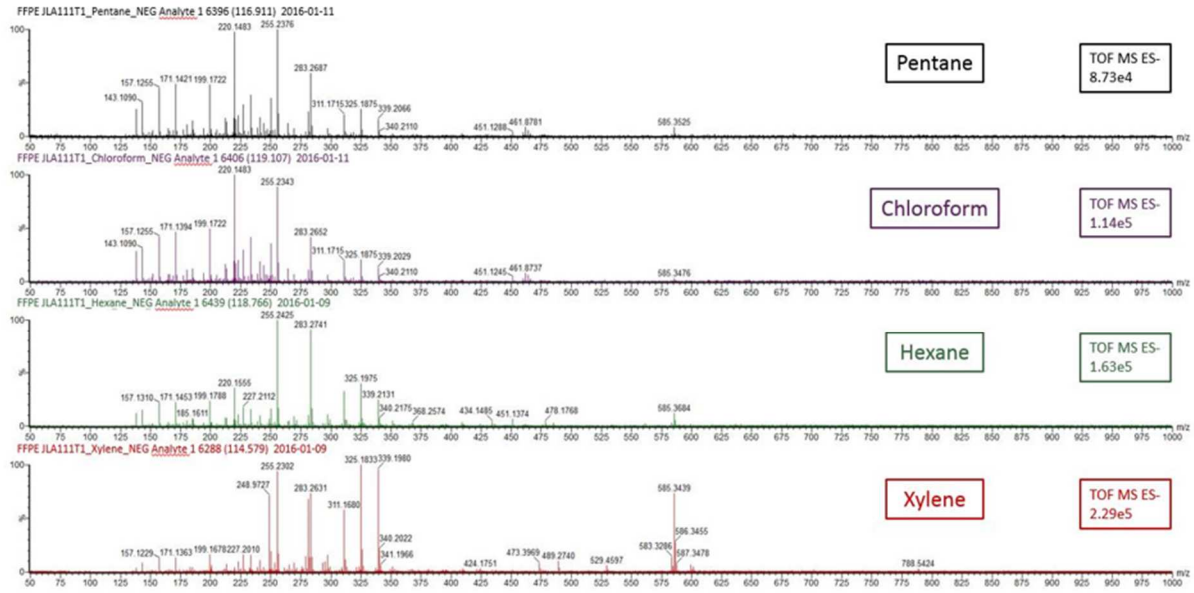


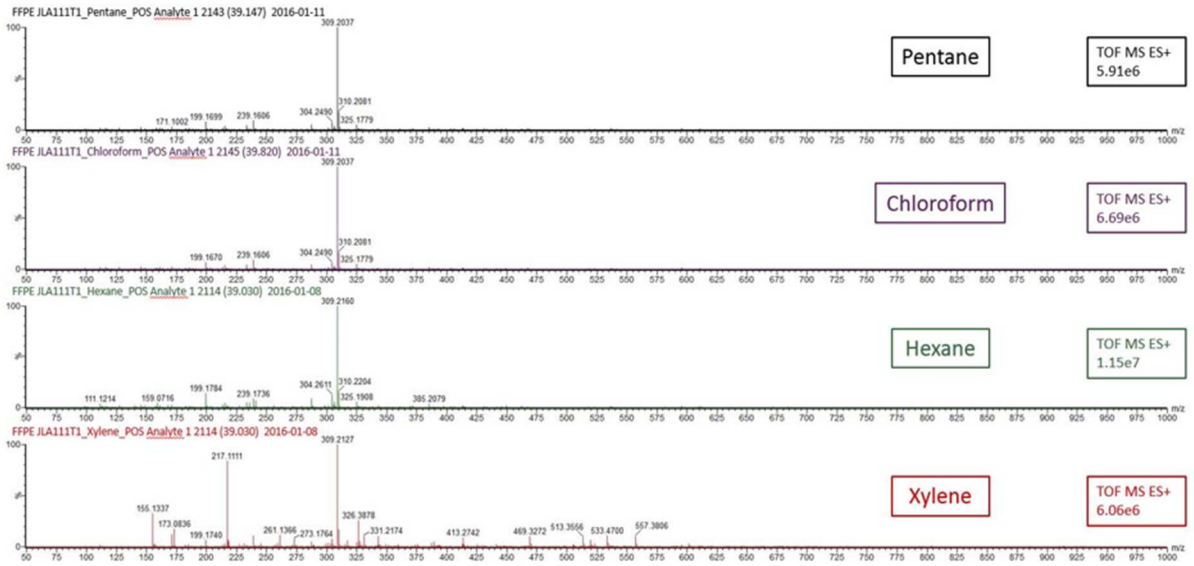
Figure 4.21. Ion images representing the data acquired in both ionisation modes for four tested solvents. The distribution of *m/z* 599.32 was visualized for the negative mode and *m/z* 155.13 for the positive mode data. The putative ID of LPI (18:0) was assigned to *m/z* 599.32. These two ions were chosen by hand in the HDI software.

Different tissue classes were distinguishable in the tissue section washed in xylene and to a certain extent also for the sections rinsed in chloroform and pentane.

The negative mode data did not look promising, contrary to the observed spectra composition (Figure 4.22).



A – negative mode



B – positive mode

Figure 4.22. Averaged mass spectra acquired after tissue sections were washed in four different solvents as a part of testing the new pre-DESI-MSI protocol for FFPE tissue samples.

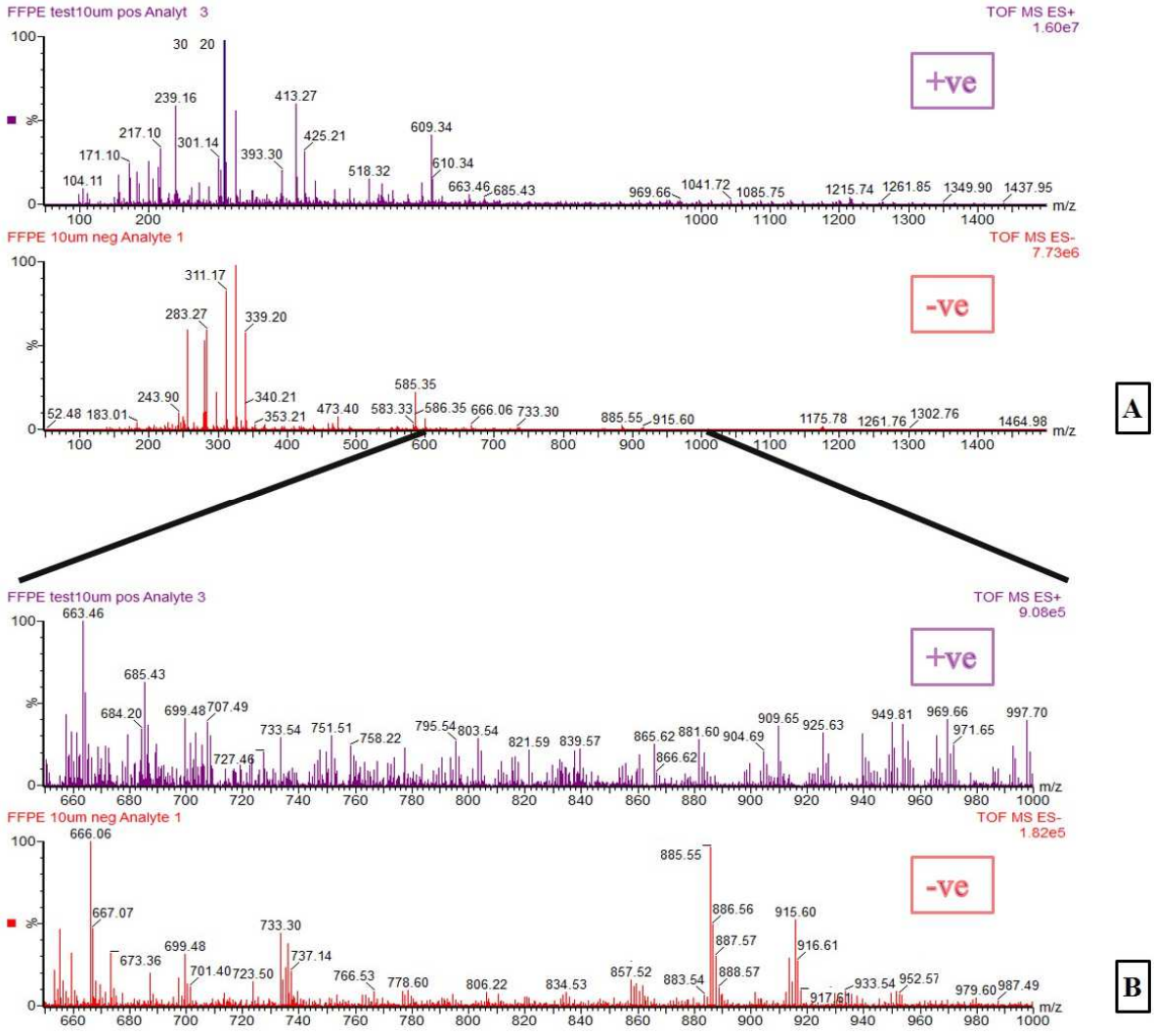


Figure 4.23. Averaged mass spectra for the colorectal sample rinsed in xylene 2 x 8 minutes before the standard DESI-MSI analysis was conducted. Data were acquired in both positive and negative ionisation modes (A). Zoom in on the phospholipid mass range (B).

It was observed that spectral composition for FFPE samples was very different from those acquired for fresh-frozen specimens. There were hardly any peaks present in the phospholipid mass range, which was thought to result from extensive washes in alcohol and xylene during sample processing. Still, an averaged spectrum for the tissue section washed in xylene seemed to contain the most information for the negative mode measurements (Figure 4.22A). As for the positive mode data, results recorded for samples washed in pentane, chloroform, and hexane all looked very similar, with hardly any metabolic information present in analysed sections (Figure 4.22B). Therefore, these results confirmed that pre-DESI washes in xylene contributed to the best conservation of the sample's chemical composition out of all tested

solvents. A zoomed-in view of the phospholipid mass range for the tissue section rinsed in xylene is shown in Figure 4.23B.

4.3.13 Preliminary DESI-MSI results for the FFPE sample washed in xylene

The tissue section rinsed in xylene was subjected to the post-DESI analysis. Two tissue classes present in the sample were annotated with the help of a qualified histopathologist. This was a first attempt to assess if an FFPE sample's chemical content might be enough to allow successful tissue classes classification.

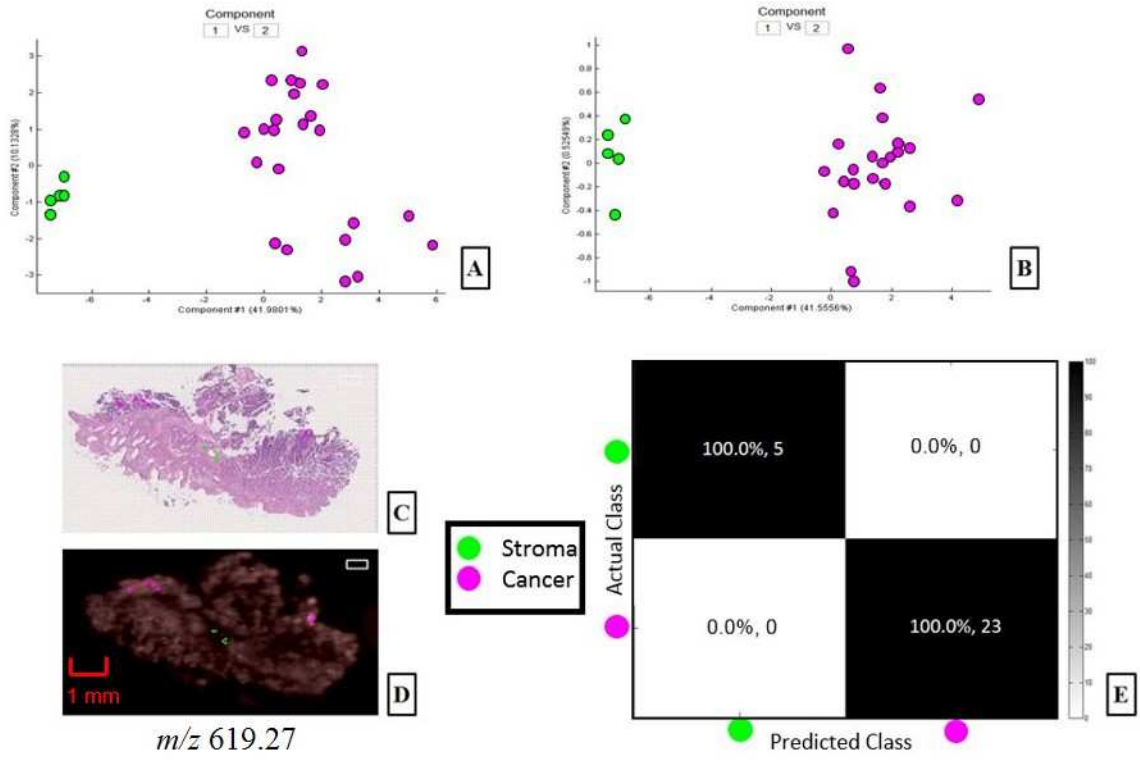


Figure 4.24. Statistical analysis of FFPE colorectal sample washed in xylene and analysed in negative ion mode. Showing PCA (A) and MMC+LDA (B) plots, cross-validation accuracy for the prediction of stroma and cancer (E) H&E-stained analysed tissue section (C) and ion image of *m/z* 619.27 (D).

It is worth mentioning that the histopathologist only managed to select five pixels representing stroma for this particular sample. Ideally, to build a reliable and robust model, a higher number of pixels would be used. A similar number of pixels would be annotated for each tissue class present in each particular tissue section. However, a clear trend was still seen when an unsupervised analysis was attempted (Figure 4.24A) for the negative mode data. Leave-one-

out cross-validation confirmed a 100% accuracy in classifying stroma and cancerous tissue (Figure 4.24E).

Similarly, the positive mode data also looked promising. An adequate tissue separation was achieved with the supervised analysis, as observed on the MMC plot (Figure 4.25), with a cross-validation accuracy of over 85%.

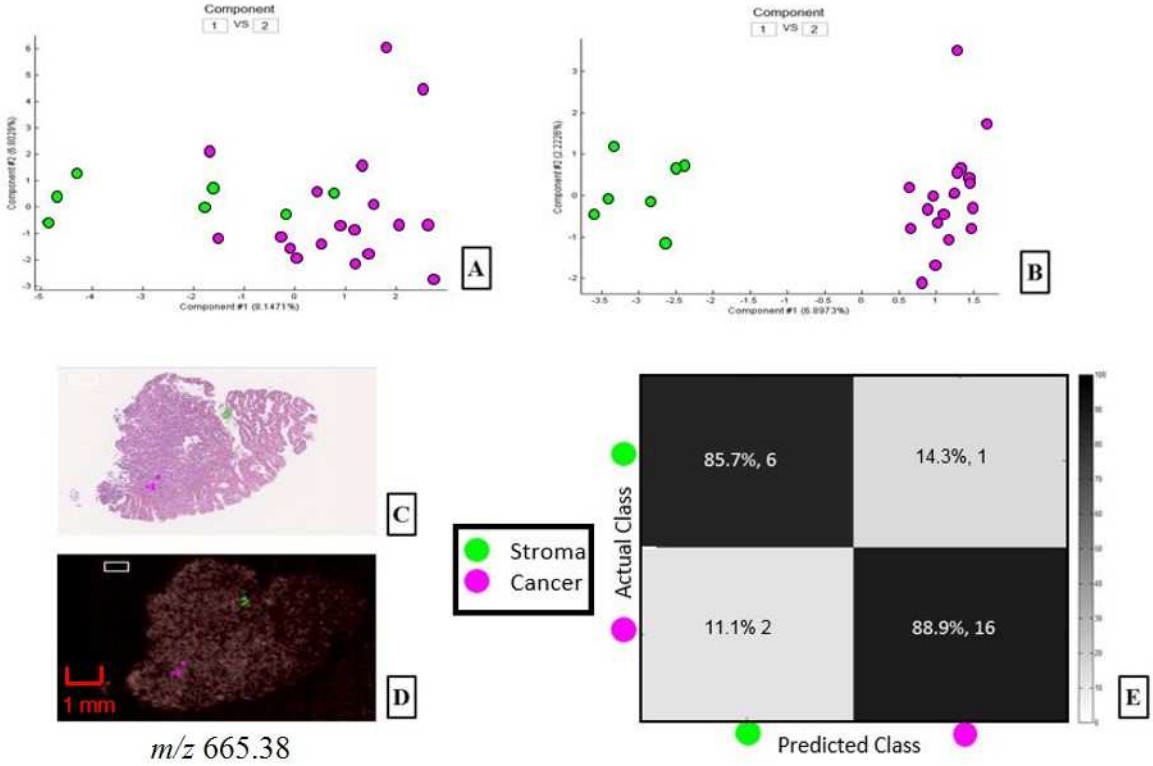


Figure 4.25. Statistical analysis of FFPE colorectal sample washed in xylene and analysed in positive ionisation mode. Showing PCA (A) and MMC+LDA (B) plots, cross-validation accuracy for two tissue classes present in the analysed sample (E) and H&E image of the analysed tissue section (C) and ion image of *m/z* 665.38 (D).

4.3.14 Statistical analysis of a 10µm thick FFPE mouse liver sample

As mentioned above, one potential drawback of the procedure for processing tissue samples was that while performing the ethanol gradient washes and the xylene clearing steps, molecules of interest could also be washed from the tissue sections. Therefore, it has been investigated whether enough metabolite information could be obtained from FFPE tissue sections using DESI-MSI to discriminate between different tissue types. Hence, a mouse liver sample with basophilic and acidophilic nodules was prepared according to the standard processing protocol. Before the DESI-MSI experiment, the slides were washed twice with xylene for 8 minutes in

order to remove paraffin, allowed to dry for 24 hours and analysed in both positive and negative ion mode.

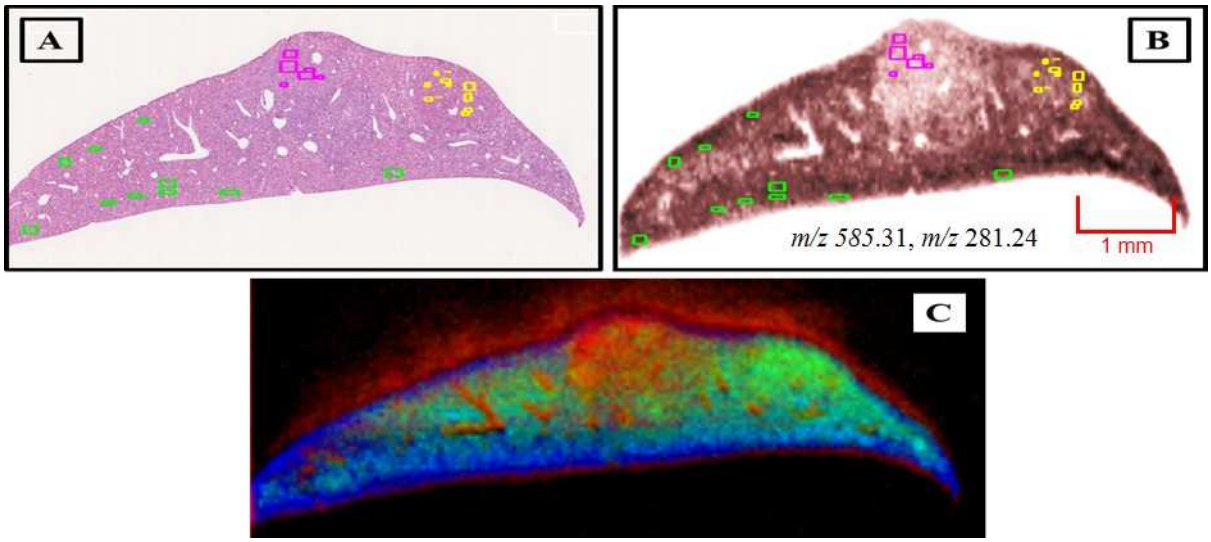


Figure 4.26. FFPE mouse liver sample with three tissue types used for data acquisition in negative ionisation mode. H&E stained digitalised image with tissue annotations: healthy liver tissue is marked in green, acidophilic nodule 1 – pink, basophilic nodule 2 – yellow (A); integrated ion image (B); overlaid RGB image of PCA components (C).

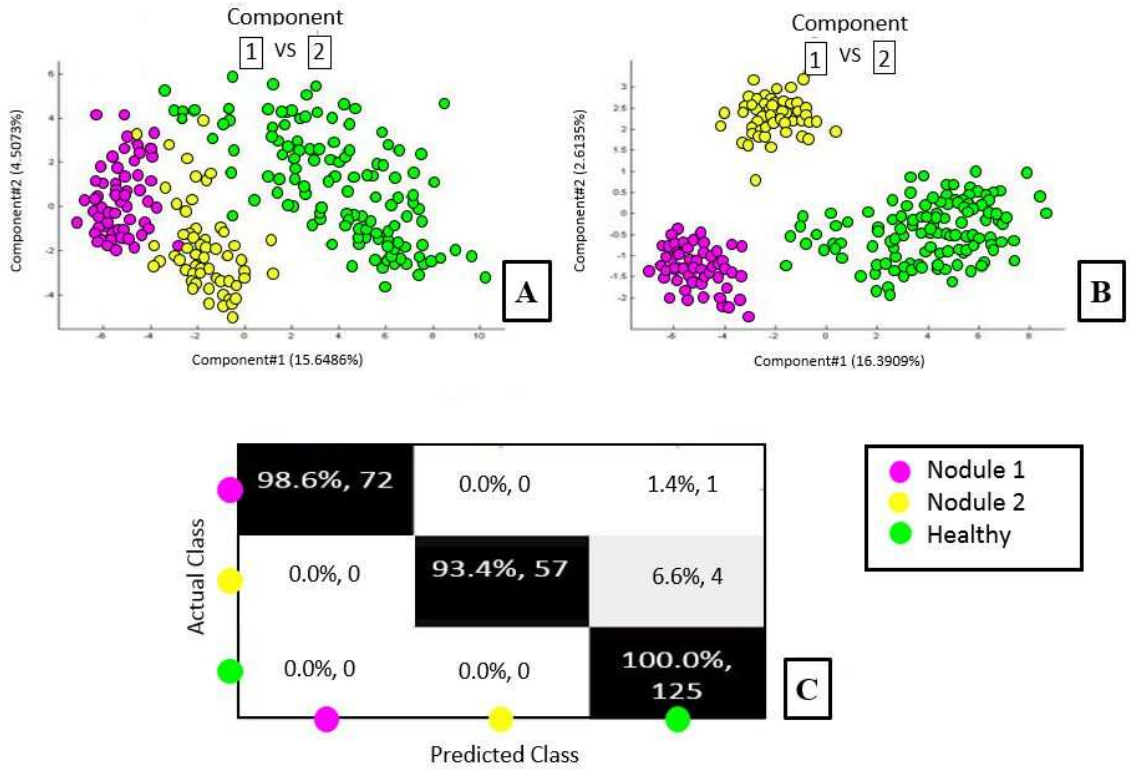


Figure 4.27. Statistical analysis of the results for a mouse liver sample analysed in negative ion mode. PCA analysis of different components (A); supervised analysis MMC+LDA (B); leave-one-out Mahalanobis cross-validation (C).

Statistical analysis was performed using PCA and MMC, and it was observed that differentiation between tissue types could be accomplished using FFPE samples in both positive and negative ionisation mode.

The analysed tissue section with histopathological annotations is shown in Figure 4.26A. There were three different tissue types present in this sample: healthy liver tissue, basophilic nodule and acidophilic nodule. When statistical analysis in negative ion mode was performed, it was possible to separate the nodules, and healthy liver tissue in the first component of the PCA, and separation between the nodules was achieved in the second component of the MMC, with a cross-validation accuracy of over 93% (Figure 4.27).

The tissue section analysed in positive ion mode is shown in Figure 4.28A. Again, three tissue classes were annotated with a qualified histopathologist's help to allow further statistical analysis. These three different tissue types were quite clearly visible on the overlaid RGB image (Figure 4.28C).

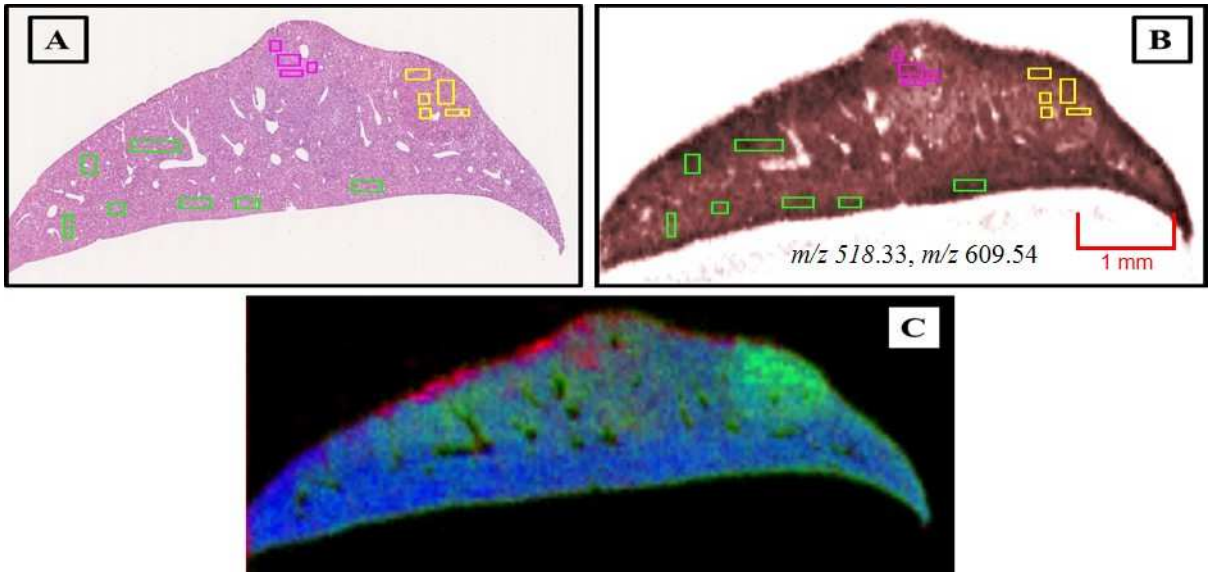


Figure 4.28. Mouse liver sample with three tissue types used for data acquisition in positive ionisation mode. H&E stained digitalised image with tissue annotations: healthy liver tissue is marked in green, acidophilic nodule 1 – pink, basophilic nodule 2 – yellow (A); integrated ion image (B); overlaid RGB image of PCA components (C).

In the positive ion mode, tissue separation could also be achieved, with a cross-validation accuracy of over 94% for the acidophilic nodule and healthy liver tissue and 70% for the basophilic nodule (Figure 4.29).

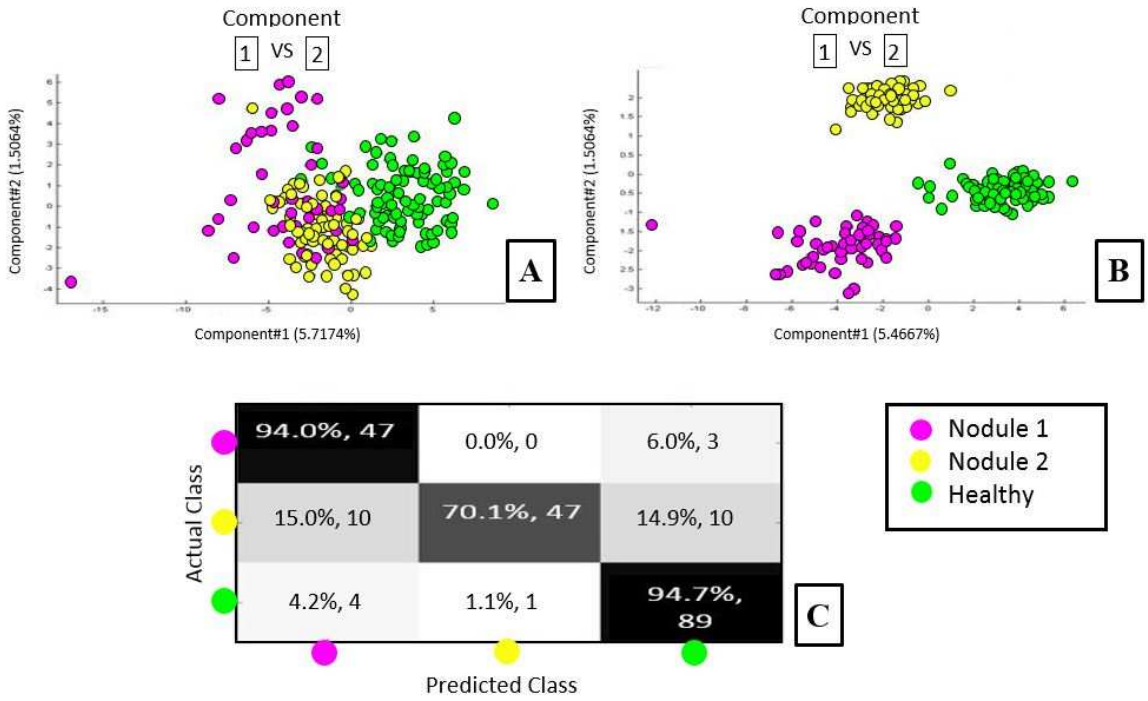


Figure 4.29. Statistical analysis results for the FFPE mouse liver sample analysed in positive ion mode. PCA analysis of different components (A); supervised analysis MMC+LDA (B); leave-one-out Mahalanobis cross-validation (C).

4.3.15 Statistical analysis of a 5µm thick FFPE mouse liver sample

The optimal thickness of fresh-frozen sections prepared to be analysed by DESI-MSI has been evaluated to be 10-14µm. Therefore, all the experiments performed so far were relaying on 10µm thick tissue sections. However, diagnostic histopathology laboratories do not cut sections at that thickness as a standard. If they are dealing with frozen samples, sections are cut at 4-5µm thick. Therefore, as a part of ongoing efforts to make DESI-MSI more histology-friendly, additional experiments were run to assess whether FFPE samples cut at 4µm would still contain enough metabolomic information to allow tissue classification.

The same mouse liver sample that was used for experiments described in *Chapter 4.3.13* was cryosectioned at 5µm thick. Before DESI-MSI, the slide was washed in xylene 2 x 8 minutes, and the rest of the protocol remained precisely the same, including the instrument set-up. The data were recorded in both ionisation modes.

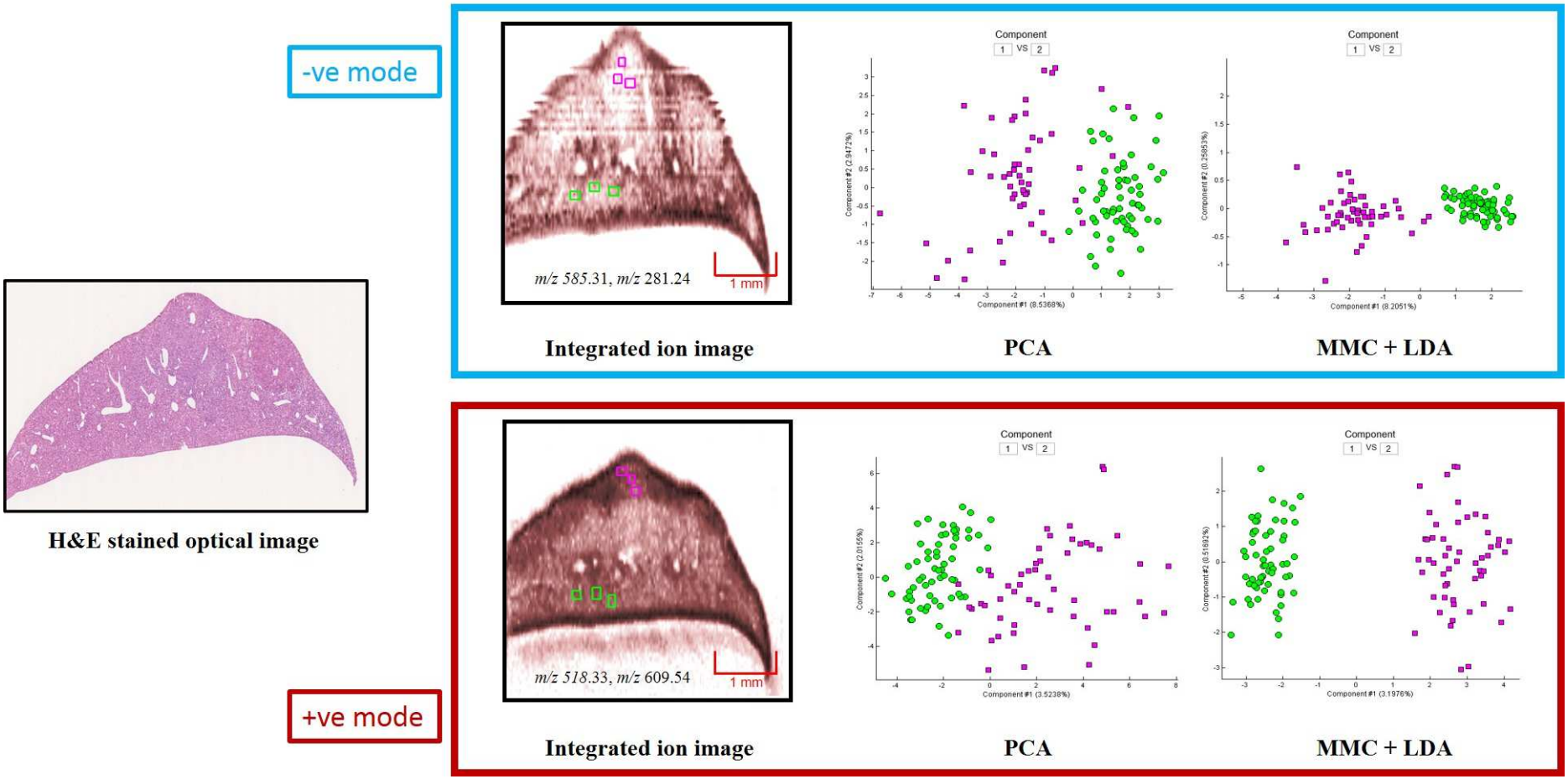


Figure 4.30. The FFPE mouse liver sample cut at 5µm washed in xylene and analysed by DESI-MSI in both ionisation modes. The sample contained an acidophilic nodule (pink) and normal liver tissue (green) that were annotated with a histopathologist's help. A perfect separation was achieved for both ionization modes data even when an unsupervised analysis (PCA) was attempted.

A clear trend was observed even during an unsupervised PCA, with two tissue classes separating from each other for both negative and positive mode data (Figure 4.30). Therefore, this experiment has proven that even tissue sections cut at 5µm thick might contain enough metabolic information to classify different tissue types. However, contrary to tissue classification in fresh-frozen samples, it did not rely entirely on phospholipids mass range for the FFPE tissue specimens. Due to extensive washes in various solvents performed during the fixation phase, FFPE samples' metabolic content was understandably very different. Therefore, the whole mass range (100 – 2000) was taken into account when tissue classification was attempted. Some phospholipid species were still present when the acquired data were examined, but the low mass range seemed to be especially rich. As a result, it was hypothesized that small metabolites could mainly contribute to successful tissue differentiation after DESI-MSI. An example of an averaged mass spectra for these sections is shown in Appendix G.

4.4 Discussion

There are numerous reasons why many efforts have been put into making FFPE tissue samples available for various mass spectrometry platforms. Not only most of the samples taken as a part of the diagnostic process are fixed in formalin and embedded in wax, but also the clinical information for those samples is readily accessible. Often, some additional tests like IHC and special stains are also available for those cases.

Therefore, this project aimed to optimise a pre-DESI-MSI protocol to allow the metabolic analysis of FFPE tissue samples.

First, a sucrose infiltration step and embedding in gelatin have been tested. It has been reported that these steps could significantly improve the quality of cryosections prepared from whole *Drosophila* [149]. It has, however, been observed that the signal intensity was not comparable to fresh-frozen samples, and no morphological structure of the analysed sections could be seen. The study using gelatin to assist with cryosectioning *Drosophila* reported that one of the difficulties they have been experiencing was related to the fact that each section consisted of heterogeneous tissue types. As the external cuticle was very hard, and the internal structures were much softer and surrounded by the fat body, each tissue class required different cryosectioning conditions. Also, the hydrophobic external cuticle was not compatible with a hydrophilic embedding medium. Therefore complete coverage by the medium used to assist with the cryosectioning was not possible [149]. While tissue samples fixed in formalin were often very brittle and challenging to cryosection, they did not pose the same issues as those

mentioned above. The majority of times, at least two different tissue types were present in each sample, but they were not as significantly different as in *Drosophila.* Therefore, this approach did not seem to be compatible with the requirements of this project.

Secondly, an attempt was made to assess whether glycerol might be more suitable as a cryoprotection agent. It would likely show similar protection properties as sucrose but would not ionise as readily in negative ion mode, thus being less likely to cause ion suppression. Moreover, glycerol was thought to be more compatible with formaldehyde scavenging protocols as, unlike sucrose, it did not contain an aldehyde group. However, the signal intensity was also not satisfactory. The lowest ionisation was observed for the sample treated with 0.1% glycerol. On the other hand, the one treated with 5% glycerol appeared to be completely destroyed after the DESI measurements with hardly any tissue remaining on the glass slide.

Next, the heat-induced antigen retrieval method was taken into consideration. Many studies reported that satisfactory protein extraction protocols had been established based on the heat-induced antigen retrieval technique, which was widely applied in IHC for FFPE tissue sections [150-152, 154]. It was hypothesised that a similar pre-experimental approach could be helpful for this project's aims, even though proteins were not the primary molecules of interest. Heat treatment was the first approach out of those tested so far, which resulted in the morphology of analysed sections becoming discernable. However, after these experiments, the most crucial observation was that urea/ heat-treated FF pork liver was tough to section as the tissue became too soft after the treatment. Therefore, this approach was also not compatible with the intended use.

Since some papers reported that sodium metabisulfite could be used to neutralise formaldehyde [153, 154], this approach was also tested. Recorded signal intensity was surprisingly high; however, no morphological features could be seen in the analysed tissue sections. Pictures of pork liver sections mounted on glass slides were taken before and after DESI measurements. Prior to DESI-MSI, the tissue sections' morphological features could be clearly seen; therefore, it was hypothesised that the frozen section might not have adequately adhered to SuperFrost slides that were used as a standard. That led to further experiments being carried out using various types of glass slides (SuperFrost, SuperFrostPlus, Poly-Lysine, SuperFrostPlus Ultra, and SuperFrost Gold). They were combined with different formalin scavengers used at various concentrations (sodium metabisulfite, urea) with varying numbers of heat-treatment cycles (details are summarised in Table 4.2). However, none of these pre-experimental treatments resulted in satisfactory results. Urea seemed to have more effective formalin reversal properties than sodium metabisulfite as the fatty acid: phospholipid ratio was improved with all urea

concentrations tested. However, post-DESI tissue sections' quality was still unacceptable, and no morphological structures could be seen.

It was then decided to assess the recently published protocol, which claimed to allow for the successful analysis of FFPE tissue sections by MALDI-MSI [101]. Since it has already been tested for the MSI application, it was the most suitable approach to DESI-MSI.

After some initial tests, it was established that reasonably simple preparation steps were enough to obtain some metabolic information for FFPE tissue sections. Simultaneously, the sections were still intact after DESI-MSI measurements allowing for a detailed histological examination. As expected, the spectral composition for FFPE samples was very different from those acquired for fresh-frozen specimens. Peak intensities were significantly reduced above 600 *m/z* in FFPE samples. However, it was tested and confirmed that despite extensive washes in gradients of alcohol and xylene, the remaining metabolic content was sufficient to allow for reasonable tissue classification within the analysed FFPE tissue sections.

Moreover, even 5µm thick tissue sections were tested, and the results of an unsupervised analysis were very encouraging. These preliminary data have proven that it was possible to conduct metabolite measurements from FFPE tissue samples by DESI-MSI. Interestingly, as it has already been stated, the use of MSI platforms for FFPE peptide and protein content required many preparation steps [157]. In contrast, a straightforward workflow consisting only of a deparaffinisation step was enough to measure FFPE tissue samples' metabolic content by both MALDI-MSI [101] and DESI-MSI.

The work carried out and described in this Chapter was used as a foundation and was further expanded in *Chapter 5*.

Chapter 5

DESI-MSI of fresh-frozen and FFPE colorectal tissue samples

5.1 Introduction

5.1.1 Colorectal cancer incidence rates

Colorectal cancer (CRC) is one of the leading causes of cancer-related deaths worldwide. It is considered the third most common cause of cancer death, after lung and breast cancer in women, and after lung and prostate cancer in men [1, 158, 159]. In England, CRC incidence rates increased by 14 per cent for women and 33 per cent for men between 1971 and 2009. Around 6,000 women and 7,000 men died from CRC in England in 2010, a rate of 13 deaths per 100,000 and 21 per 100,000, respectively [160]. It has been estimated that 85% of colorectal cases occur in people who are more than 55 years old [161]. However, many studies have recently reported an increasing incidence rate in patients less than 40 years old [162-167].

5.1.2 Aetiology

The molecular mechanisms still require further investigation, even though significant efforts have been undertaken to understand the pathogenesis of CRC. Its aetiology has been linked to multiple factors, which include genetic mutations [168, 169], inflammatory processes [170, 171], diet [172-175], and more recently, gut microbiota [176-179]. Cigarette smoking [180, 181] and alcohol consumption [182, 183] are also associated with an increased risk of developing CRC. It was established that more than 95% of colorectal cancers are sporadic and the individuals affected are not genetically predisposed to the disease [184].

Research studies have shown that the gut microbiota is linked to CRC development as a primary driver of inflammation in the colon [185]. One mechanism through which microbes can influence disease development is thought to be through microbe-driven intestinal inflammation. However, only a slight increase in CRC incidence has been observed in individuals with chronic inflammation due to inflammatory bowel disease (IBD), including ulcerative colitis and Crohn's disease. Moreover, since genetic abnormalities are hallmarks of all cancers, it remains unclear how a deregulated immune response can lead to genetic lesions predisposing to CRC [186, 187].

5.1.3 Diagnosis

A CRC diagnosis is made through manual histological evaluation of biopsies taken from the tumour and surrounding tumour regions. This evaluation acts as one of the key prognostic factors for most cancers, providing information on the tumour type, grade and any morphological abnormalities. This information is then used to decide on the treatment for a

particular patient. CRC develops through many different stages, and these can be classified as either benign or malignant (Figure 5.1). Endoscopy is routinely used for the identification of gastrointestinal polyps and tumours [188]. However, early diagnosis is essential and may prevent the formation of tumours. Therefore, the aim is to identify and remove polyps when they are still at the benign stage. In this case, cancerous cells will only be defined within the polyps, and as a result development of the tumour can be prevented in the future.

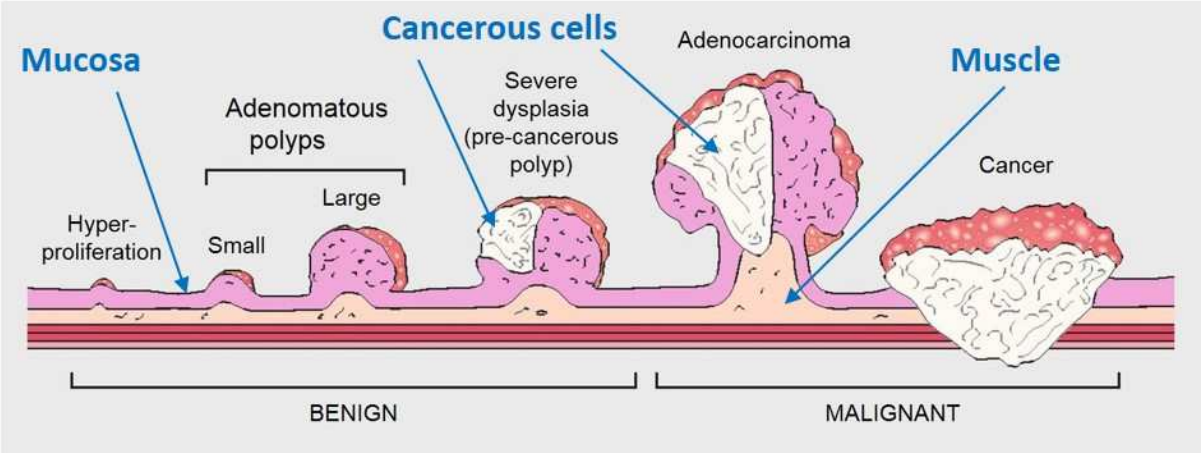


Figure 5.1. Progression from colorectal polyp to cancer. Adapted from [189]. The whole process is divided into two phases: benign and malignant. In the benign stage, cancerous cells are confined with polyps, and the tumour did not grow beyond mucosa, the inner layer of the colon or rectum. Therefore, if a polyp is removed during a routine gastroscopy, the development of the tumour is prevented. In the malignant stage, cancerous cells are invading through the layer of the underlying muscle.

In each case, removed polyps are sent for histology assessment, and Tumour Nodes Metastases (TNM) staging is used to describe the invasiveness and spread of cancer (Supplementary information, Appendix H). In the TNM system, the T refers to the size and extent of the primary tumour, the N refers to the number of nearby lymph nodes with cancer, and the M refers to whether cancer has metastasized [188].

5.1.4 Histopathological assessment

Traditional histological evaluation of post-operative tissue sections is the standard method that has been used over the decades to enable clinicians to characterise tumours. Whilst this approach is relatively simple and powerful, it has some significant drawbacks. Manual examinations are very time consuming and hence limit the speed and throughput of diagnosis. Another aspect that should also be mentioned is an inter-observer error [35-37]. There is a

broad spectrum in cancer morphology; therefore, many tumours can be atypical or lack morphological features crucial for future diagnosis [33]. Moreover, the advanced form of CRC's low survival outcome has prompted the need for reliable predictive and prognostic markers.

5.1.5 Prognostic markers for the CRC

Recent advances in the molecular subtypes of colorectal cancer, methylation of DNA in CRC, micro-RNA biogenesis, and their involvement in CRC have resulted in many new colorectal biomarkers [190-192]. Such biomarkers may be used for earlier diagnosis (microsatellite instability, APC mutation, methylation biomarkers) [192], selection of 'personalised' therapy (KRAS, BRAF) [191], and prognosis of CRC (KRAS, BRAF, and potentially miR-21, miR-92a, miR-200, miR-320e) [193, 194]. Several marker classes have been evaluated for their CRC screening use and have shown potential in early phase biomarker studies. Importantly, non-invasive biomarkers derived from biological fluids (blood- or stool-based markers), due to their easy accessibility, could be considered practical tools for CRC detection and monitoring [195]. Unfortunately, most of the identified biomarkers failed in the validation studies due to the lack of consistency between biomarker panels. This failure highlights a significant obstacle to the development of robust CRC biomarkers [190]. Therefore, it is clear that alternative methods are required for accurate and timely diagnosis and characterisation of tumour grade and stage [196]. As a result, further research and new modalities involving automated diagnosis are needed to overcome these challenges to lessen the traditional pathology-based tissue diagnosis burden.

5.1.6 DESI-MSI as a tool to complement traditional diagnosis

DESI-MSI has been used in many projects aiming to characterise various types of cancer, including colorectal cancer [83, 84, 88, 90-92]. This project also aims to apply the DESI-MSI to study colorectal cancer and use fresh-frozen tissue sections as has been done so far and also FFPE tissue samples. The reasons why FFPE samples are a very desirable material to use in research studies have been discussed in *Chapter 4.1*.

5.1.7 Use of FFPE samples in MS-omics studies

Currently, fresh-frozen samples are still preferred over FFPE in tissue-based DESI-MSI and MALDI-MSI studies. This is mainly due to concerns related to the effect of chemical processing on the molecular content of tissue samples, molecular delocalization resulting from

washing steps, and the risk that paraffin might cause ion suppression [197, 198]. However, several separation studies have been performed in recent years, demonstrating that unlocking FF proteins was possible. Ahram *et al*. [199] reported that they could extract proteins from ethanol-fixed and paraffin-embedded tissue and analyse them by HPLC MS/MS. Shi *et al.* [150, 151] showed evidence that obtaining a good overlap between proteins identified from frozen and FFPE sections was possible. Some encouraging results were also published by Crockett *et al.* [147] and Prieto *et al.* [200] regarding the identification of proteins from enzymatically digested FFPE tissues by LC-MS/MS.

There was also much effort put into the application of MALDI-MSI to FFPE samples. Ronci *et al.* [201] described a methodological investigation of protein unlocking protocols to allow MALDI-MSI experiments using FFPE tissue sections. Many protocols have been published for MALDI-MSI of peptides [202-205] and N-glycans [202, 206] from FFPE specimens.

Moreover, it has been reported that metabolite measurements from homogenised FFPE tissue samples using liquid-based mass spectrometry were performed [207, 208]. It is important to stress that liquid samples cannot provide information on molecular spatial distribution. This aspect is particularly essential as tissues are heterogeneous and highly complex systems [101]. The liquid-based MS approach requires significant amounts of tissue to allow for satisfactory metabolite yields to be used for analyses. Additionally, it has been established that tissue homogenization leads to loss of specificity and therefore contributes to reducing significant results [101]. As an analysis of FFPE samples is very desirable from a biological and clinical context, it has been hypothesized that given the ability to measure metabolites from FFPE samples by LC-MS, it should also be possible to do so in MSI.

When this work was carried out, it was the first study of *in situ* DESI-MSI analysis of FFPE tissue samples.

These efforts are undertaken in order to make DESI-MSI as histology-friendly technique as possible.

5.2 Materials and methods

5.2.1 Fresh-frozen vs FFPE – two types of samples used

Two types of human colorectal tissue samples were used for experiments performed for this project's needs: fresh-frozen samples and FFPE tissue samples (including tissue microarray).

5.2.2 Clinical specimens and sample collection

5.2.2.1 Fresh-frozen colorectal tissue samples

Approval for the study was obtained from the institutional ethics review committee and Imperial College Healthcare Tissue Bank (REC reference number 14/EE-0024). Tissue samples for this study were collected from patients who were undergoing colorectal surgery at Imperial College Healthcare NHS trust. Further details can be found in *Chapter 2.1.1*.

After retrieving a surgical specimen, it was first sent for routine histopathological examination, where suitable samples were taken and snap-frozen for DESI-MSI.

Patient characteristics for this cohort were recorded and are summarised in Table 5.1.

	Number of samples	Number of patients	Age range
Normal	27	27	33-69
Adenoma	8	8	43-55
Cancer	12	12	35-71
TOTAL	47	47	

Table 5.1. Histological characteristics of fresh-frozen colorectal tissue samples.

5.2.2.2 FFPE colorectal tissue samples

All the details of FFPE colorectal tissue samples were described in *Chapter 2.2.1*.

5.2.3 Sample preparation

Fresh-frozen samples were prepared as described in *Chapter 2.7.1*.

FFPE samples were prepared as described in *Chapter 2.7.2*.

5.2.4 DESI-MSI analysis of the colorectal tissue samples

5.2.4.1 Fresh-frozen samples

Fresh-frozen human colorectal samples were analysed in a random order in negative ion mode using a high-resolution orbital trapping mass spectrometer (Exactive, ThermoScientific, GmbH) controlled by XCalibur 2.1 software. Full details of the DESI-MSI setup for this batch of samples can be found in *Chapter 2.8*.

Parameters	Settings of Thermo Exactive instrument
Polarity	Negative
Max injection time	1000 ms
Micro-scans	1
Mass resolution	100000
Mass range	150-2000 *m/z*
Capillary temperature	250°C
Capillary voltage	-50 V
Tube lens voltage	-150 V
Skimmer voltage	-40 V
Spray voltage	4.5 kV
Skimmer voltage	-40 V

Table 5.2. Thermo Exactive Orbitrap settings used for DESI-MSI of fresh-frozen colorectal samples.

5.2.4.2 FFPE samples

FFPE colorectal samples were analysed using a Xevo G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA) coupled to a 2D sample stage (Prosolia Inc., Indianapolis, IN, USA). HDI software was used to select the regions of interest on the glass slides. Full details of the DESI-MSI setup for the FFPE samples is described in *Chapter 2.9*.

Parameters	Xevo G2-XS QToF (Waters Corporation, Milford, MA)	
Scan time	1 sec	
Scan mode	Sensitivity	
Mass analyser	TOF	
Mass range	50-1500 *m/z*	
Ionisation mode	Negative	Positive
Sampling cone voltage	-40 V	40 V
Source offset	-80 V	80 V
Source temperature	120°C	120°C
Spray voltage	4.5 kV	4.5 kV

Table 5.3. Xevo G2-XS QToF parameters used for the DESI-MSI protocol for the analysis of FFPE colorectal tissue samples.

5.2.5 Histological assessment

As DESI-MSI is a soft ionisation technique, the analysed tissue sections can be stained with H&E for histopathological evaluation. The DESI-MSI image and histology image of the same tissue section can then be used for tissue-specific data extraction. The stained slides were digitally scanned at high resolution, using a Nanozoomer 2.0-HT digital slide scanner (Hamamatsu, Japan) and analysed by an in-house histopathologist blinded to the results of DESI-MSI.

5.2.6 Data analysis

5.2.6.1 Fresh-frozen samples

Methodological details of data analyses are described in *Chapter 2.13*.

5.2.6.2 FFPE samples

The raw data files were lock-mass corrected to molecules present endogenously in the samples: FA(18:1) with an *m/z* of 281.2486 and hexadecanedioic acid with an *m/z* of 287.2217 for negative and positive ion modes, respectively, and background subtracted. The purpose of this procedure was to “lock” on to a single peak to correct for calibration drift with time. Once the “lock” onto the chosen peak's theoretical mass has been done, the spectrum was recalibrated to that theoretical mass, improving the mass accuracy at all points. Imaging raw files were

uploaded into a custom-build imaging toolbox in MatLab (R2016a, MathWorks, Natick, MA, USA) environment. The H&E stained optical image was co-registered to the MS image. Next, areas of the tissue section corresponding to different tissue types (such as mucosa, mucin, muscle, and tumour) were annotated with a qualified histopathologist. The annotated regions were used for uni- and multivariate statistics and as the training set to predict the sample's un-annotated pixels. Uni- and multivariate statistics were performed on the *m/z* range 100 to 1000. Multivariate analysis was done using median fold change normalised data which was subsequently log-transformed. PCA and MMC-LDA were attempted for the multivariate differentiation of annotated spectra. Univariate ANOVA was performed with data normalised in the same manner but which was not log-transformed to identify significantly different *m/z* variables. Pixel prediction was tested using leave-one-spectrum-out cross-validated MMC models with a 3-nearest neighbour classifier.

5.3 Results

5.3.1 Fresh-frozen colorectal tissue samples

A small set of 47 human colorectal tissue samples was analysed. It consisted of 27 healthy, 8 adenoma, and 12 tumour samples and the tissue sections were measured in negative ionization mode. The data were recorded in 150-1500 *m/z*, but the tissue classification attempts were based only on the phospholipids mass range. It has been established that PCs usually do not ionise well in negative ion mode, but they are highly intense in positive ion mode and therefore dominate the spectra. As a result, it is challenging to achieve a good intensity for other lipid classes, contributing to less diversity of the spectra from the positive ionization mode compared to the negative mode. For this reason, the majority of previous studies have focused on negative ion mode for lipid characterisation and identification of different tissue types by DESI-MSI [91, 138].

In order to compare and link the chemical composition with the morphological features of the analysed samples, optical images were aligned with ion images and examined by a qualified histopathologist for a precise selection of morphological regions of interest.

In most healthy colorectal samples, three tissue classes were present – mucosa, submucosa, and muscle. Figure 5.2 shows different metabolic profiles for these tissue classes. DESI-MSI generated intense fatty acids and phospholipids signals in negative ionisation mode. By comparing these tissue classes' averaged spectra, it was possible to observe changes in their

molecular composition from one class to another, with the most apparent differences in the phospholipids mass range.

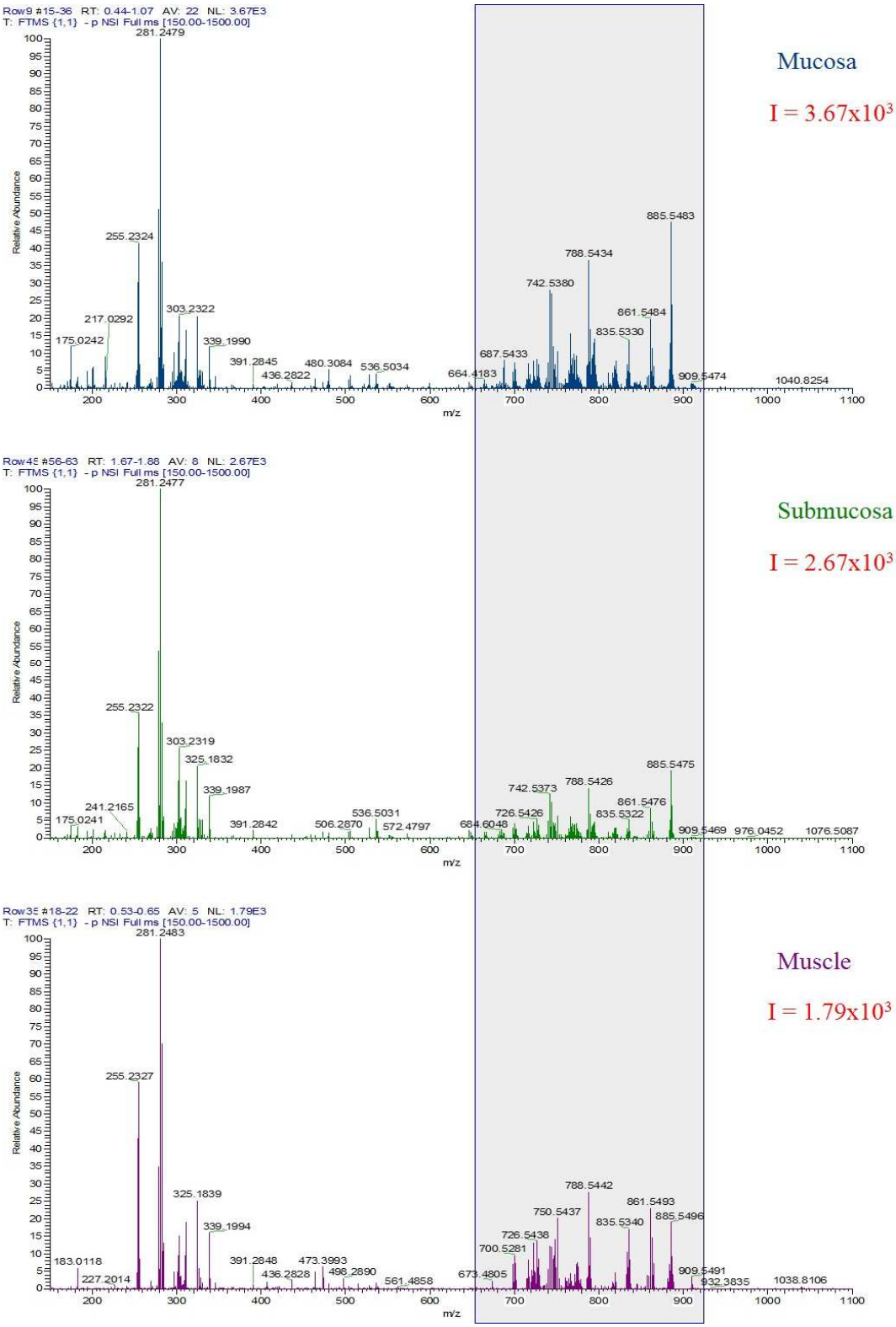


Figure 5.2. Averaged mass spectra acquired in negative ion mode from the different tissue types within a fresh-frozen, healthy colorectal tissue sample. Mass ranges corresponding to phospholipids are highlighted.

5.3.1.1 Prediction of tissue types in individual fresh-frozen samples

First, a single healthy colorectal sample was subjected to multivariate statistical analysis. A histopathologist annotated three tissue types with the smallest number of pixels assigned to muscle. That was because there was only a thin layer of that tissue between mucosa and submucosa (Figure 5.3E).

PCA was then applied to identify overall similarities/differences in lipid composition between different tissue regions in all sample surfaces without any prior information in the analysis. The resulting principal component scores for every pixel in the tissue section were visualized in a colour-coded manner with the colour intensity proportional to a given pixel score. Based on the PCA plot, there was a clear separation between the mucosa and submucosa. Muscle-related data points seemed to scatter a bit over these two tissue classes, but that could be explained by their biological composition. Since submucosa comprises a layer of dense irregular connective tissue that supports the mucosa and joins it to the muscular layer and bulk of overlying smooth muscle, it was expected that it would be similar to the muscle itself. The first component explained more than 75% of the variation in the negative mode data (Figure 5.3F). The tissue classes' spectral profiles were distinct, which was also demonstrated in the MMC plot (Figure 5.3G). The leave-one-out cross-validation results are presented in Figure 5.3H. The accuracy of over 98% was achieved for mucosa and submucosa and 94% for muscle.

Next, another healthy colorectal sample was subjected to statistical analysis, but this time four tissue classes were annotated. There were not too many pixels available for lymphoid tissue, but still, the overlaid RGB image of PCA components showed an apparent prediction of all four tissue classes (Figure 5.4C). The only two tissue types that were not distinct on the PCA plot were submucosa and muscle (Figure 5.4D). However, supervised analysis results were very clearly showing three groups representing mucosa, submucosa, and muscle. Two subgroups present for lymphoid tissue most likely resulted from a small number of annotated pixels (Figure 5.4E).

Lastly, a single colorectal tumour sample was annotated and analysed. Most cancerous samples were classified as an entire tumour, but the one shown in Figure 5.5A had two tissue types. A clear trend was seen after the unsupervised analysis. Leave-one-out cross-validation results confirmed a 100% accuracy for both tumour and muscle (Figure 5.5).

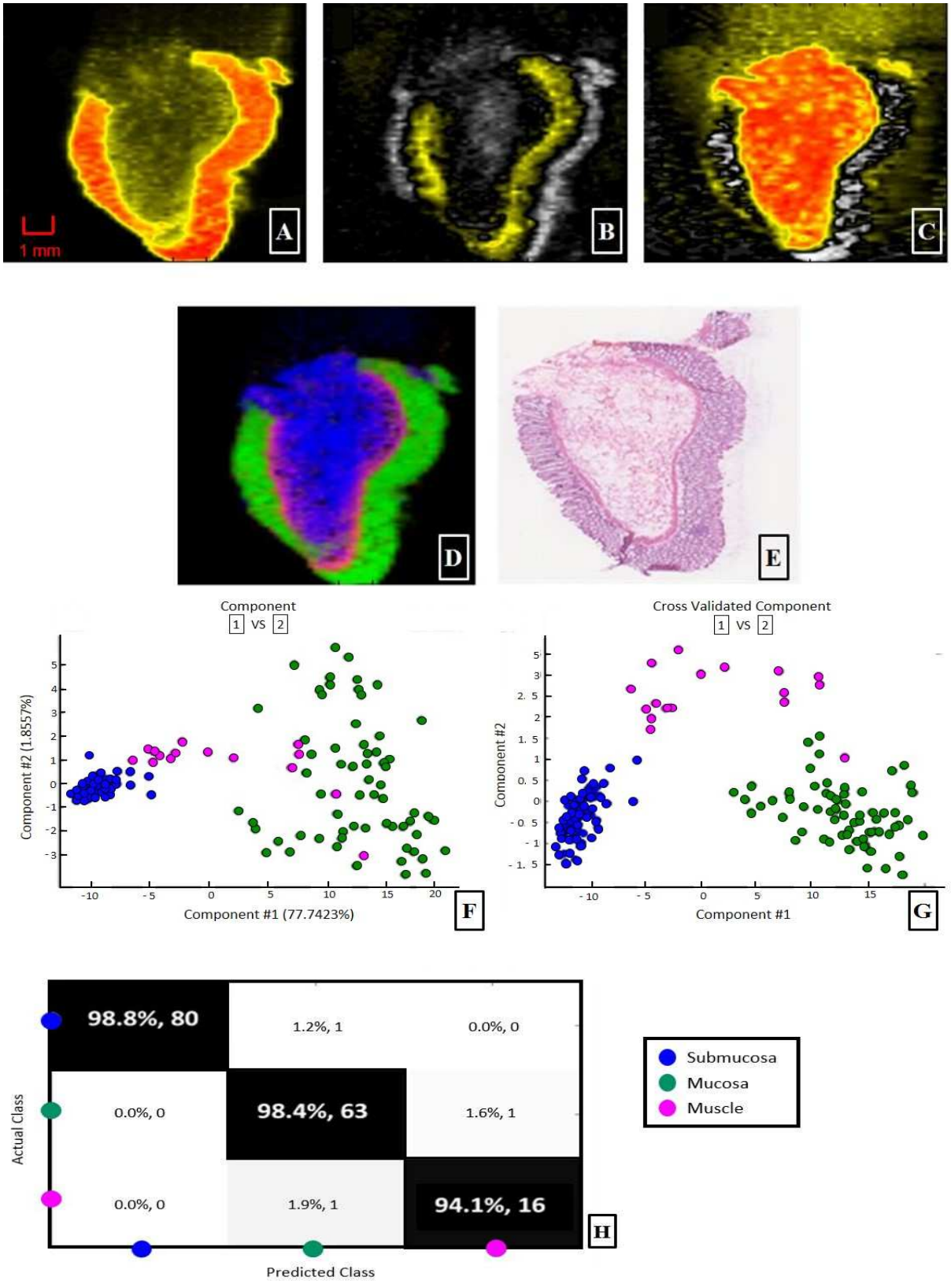


Figure 5.3. A single, fresh-frozen, healthy colorectal sample analysed in negative ion mode. PCA component images for the different tissue types identified (A-C); overlaid RGB image of PCA components (green – mucosa; pink – muscle; blue – submucosa) (D); an optical image of the H&E stained section (E) PCA analysis of different components (F); supervised analysis MMC+LDA (G); leave-one-out Mahalanobis cross-validation results (H).

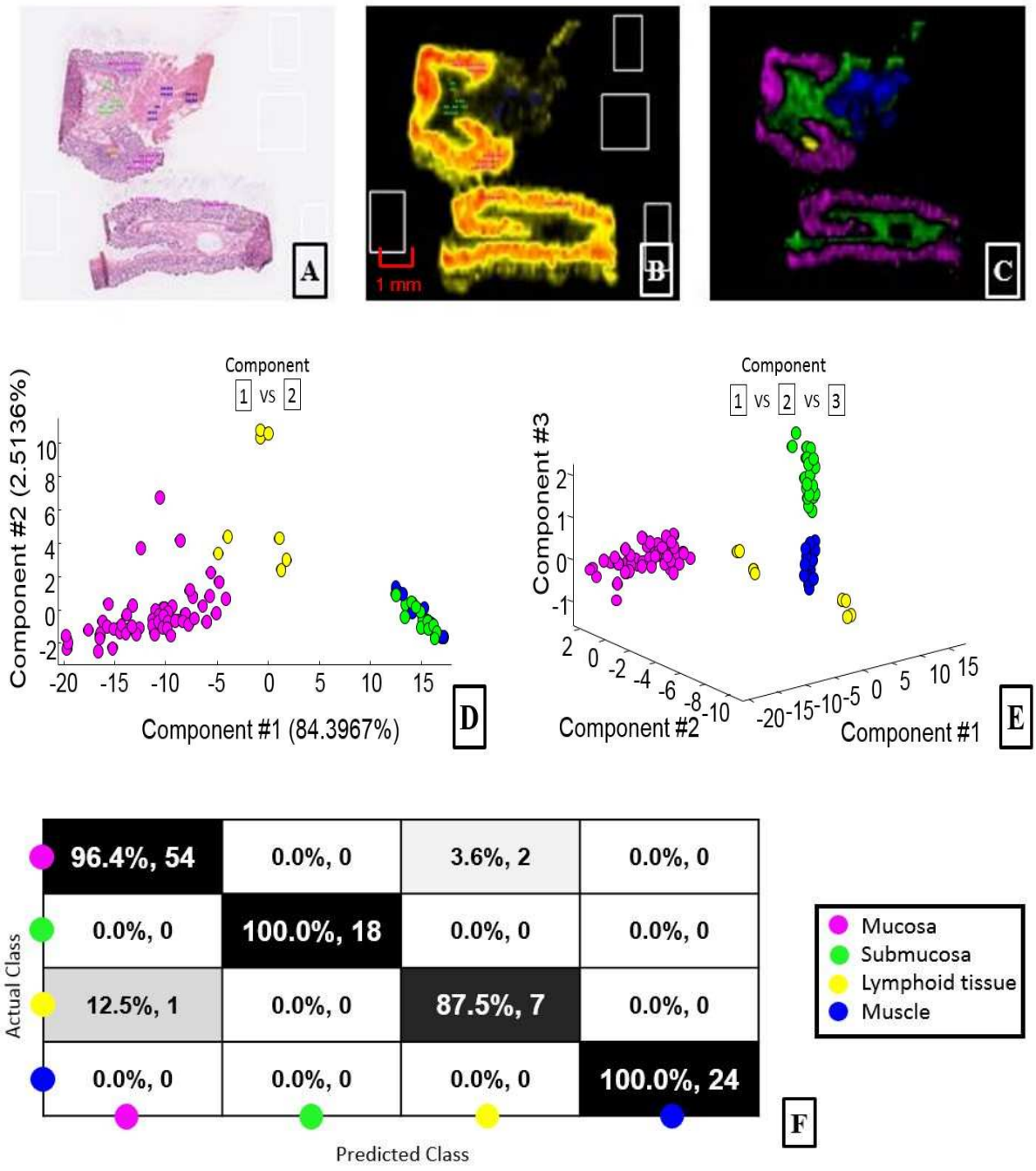


Figure 5.4. A single, fresh-frozen, healthy colorectal sample with four tissue classes analysed in negative ion mode. An optical image of the H&E stained section with histopathological annotations (A) first principal component image (B); overlaid RGB image of PCA components (pink – mucosa; green – submucosa; yellow – lymphoid, blue - muscle) (C); PCA analysis of different components (D); supervised analysis MMC+LDA (E); confusion matrix showing the classification performance (F).

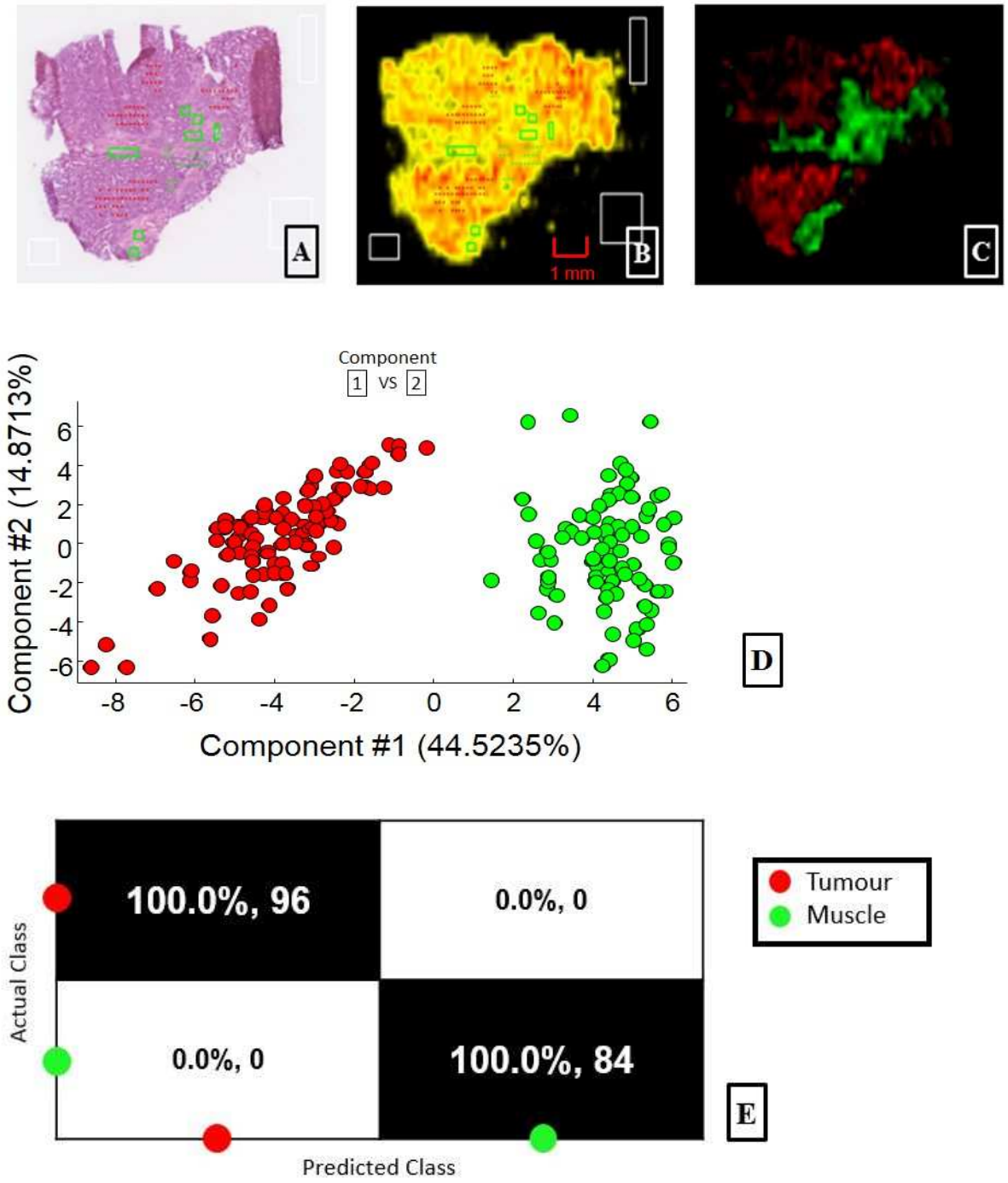


Figure 5.5. Results for a single, fresh-frozen tumour sample analysed in negative ionization mode. H&E stained tissue section with annotations for different tissue classes (red – tumour and green – muscle) (A); PC1, which already revealed the structure of analysed tissue section (B); MMC component image (C); unsupervised principal analysis plot showed a clear separation between tumour and muscle (D); leave-one-out cross-validation results (E).

5.3.1.2 Normal versus cancerous fresh-frozen colorectal tissue samples

The most obvious question that arose after the statistical analysis of individual samples was how different the surrounding healthy tissue was from a tumour. Figure 5.6 represents averaged spectra from these two tissue types as a visual representation of collected data from one randomly chosen sample. It was observed that most of the time, the signal intensity was significantly lower for cancerous samples. As colorectal cancer originates from the mucosa, all samples described as normal ones were focusing on their healthy mucosa component. It did not make much sense to try and compare cancerous tissues with submucosa or muscle.

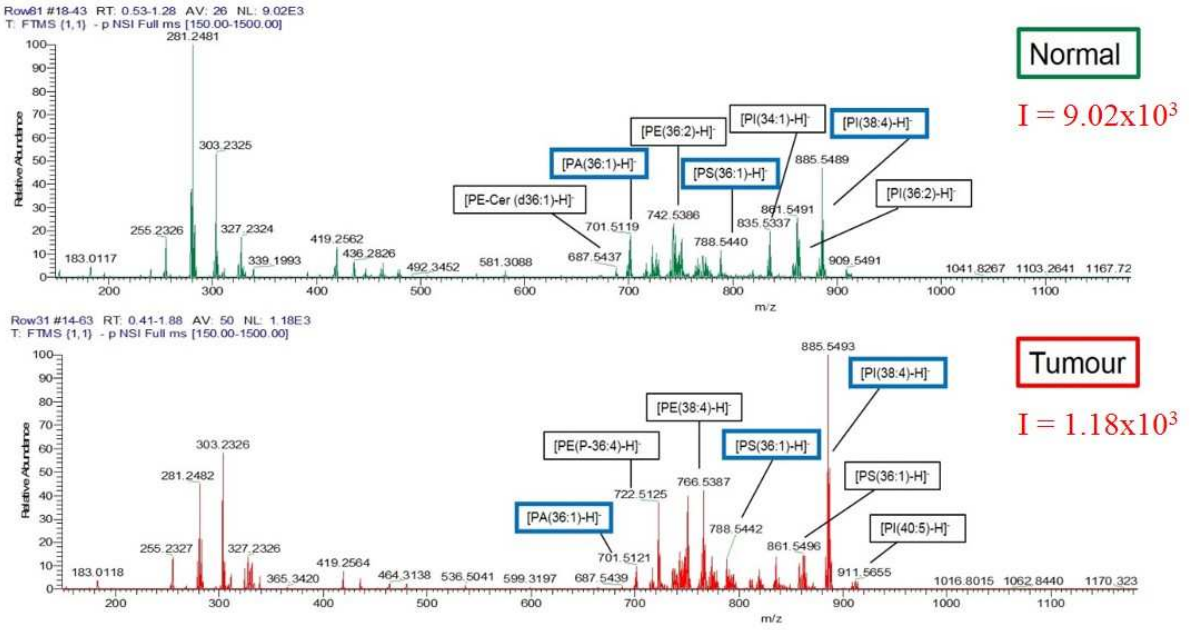


Figure 5.6. Averaged mass spectra acquired in negative ion mode from fresh-frozen healthy and colorectal tumour tissue samples. Some putative IDs have been assigned, and highlighted are those which were shared between these two tissue types.

One of the essential advantages of DESI-MSI is that, straight after the data acquisition is completed, it is possible to visualise the distribution of metabolites within analysed tissue sections. Figure 5.7 displays examples of DESI images representing four endogenous molecules for normal colorectal and colorectal tumour tissue sections, alongside the corresponding H&E stained images of the same tissue samples. Colorectal tumour samples were more homogeneous, generally presenting just one tissue type.

These four *m/z* values were selected by hand in the HDI software. First, the healthy colorectal sample was subjected to manual inspection in order to be able to identify some *m/z* values which were specific to individual tissue classes, e.g. mucosa, submucosa or muscle. Next, four molecules were selected, out of which three were only present in healthy mucosa, while *m/z* 878.55 seemed to be evenly distributed within the whole tissue section (Figure 5.7A). Next, the same molecules were visualized in the colorectal tumour sample.

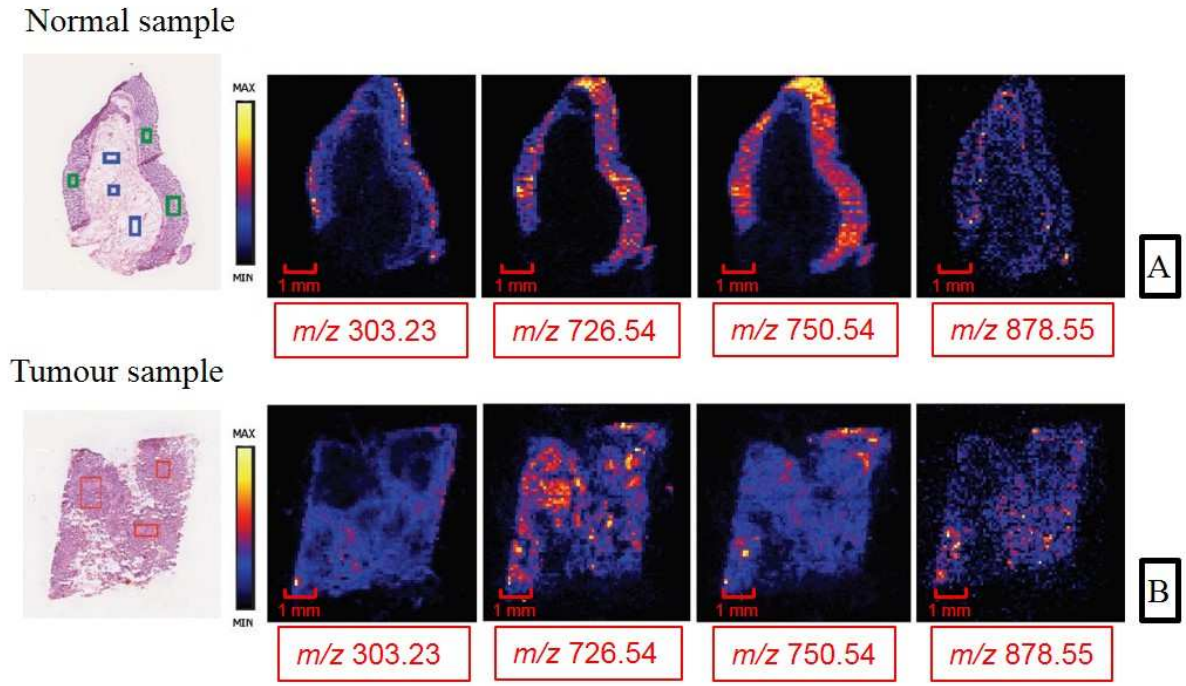


Figure 5.7. Molecule distribution in fresh-frozen normal and colorectal tumour samples. Healthy colorectal sample (A) and colorectal tumour sample (B) analysed in negative ionisation mode. The same molecules could be found in these two tissue types, and one of the advantages of this technique is that their distribution could also be visualised in the analysed tissue sections. Putative IDs: *m/z* 303.23 – FA(20:4); *m/z* 726.54 – PE(P-36:2); *m/z* 750.54 – PE(P-38:4); *m/z* 878.55 – PE(44:8).

Next, an attempt was made to identify some molecules contributing to the separation of these two tissue types (Figure 5.8). *m/z* 701.54 and 772.55 were more abundant in a healthy mucosa, while *m/z* 887.5725 was in a tumour. Putative IDs were $[PA(36:1)-H]^-$, $[PS(P-36:1)-H]^-$ and $[PI(38:3)-H]^-$, respectively.

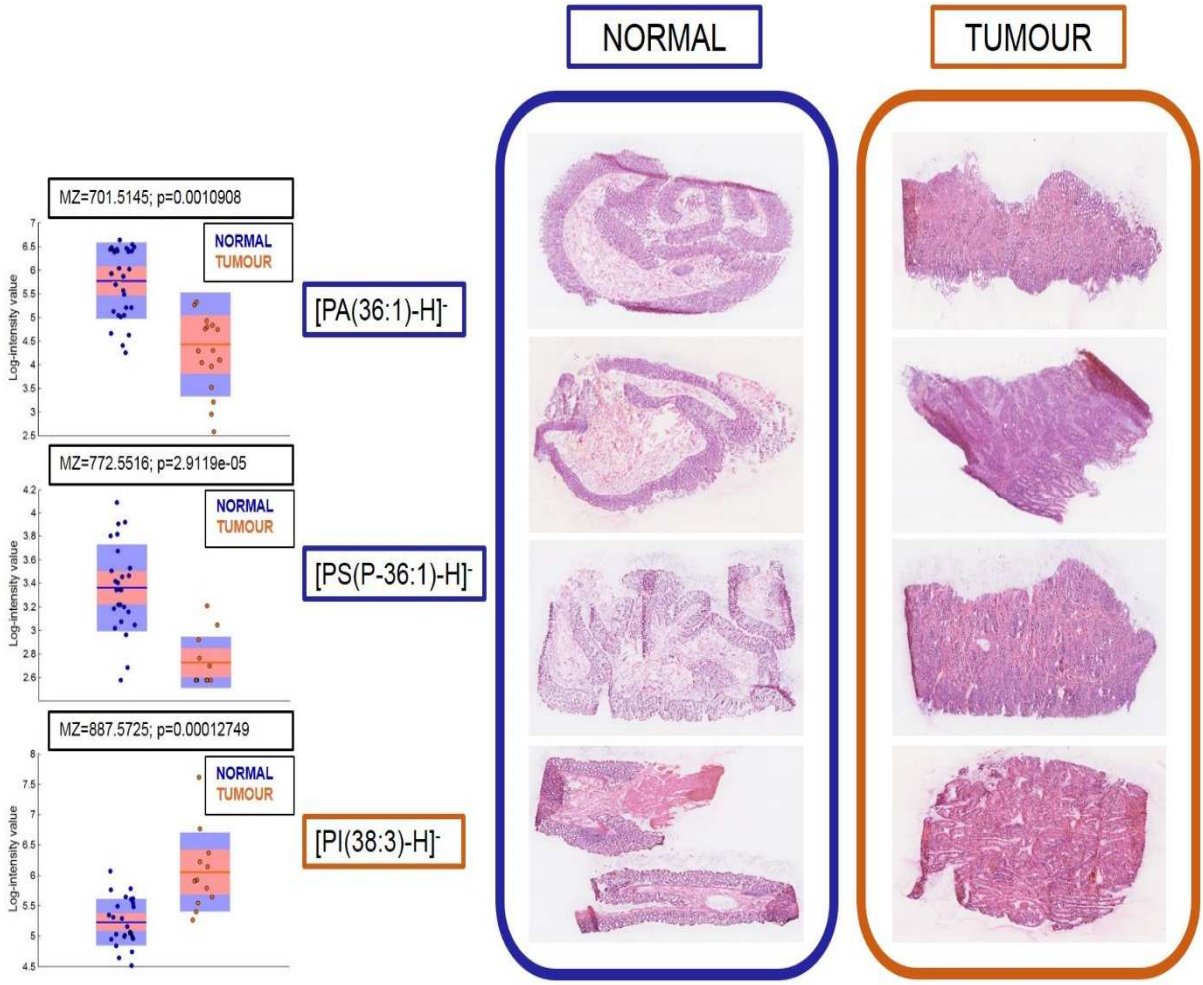


Figure 5.8. The Kruskal-Wallis ANOVA test results were used to identify molecules contributing to discrimination between fresh-frozen healthy and cancerous tissue samples. Four randomly chosen healthy and matching cancerous samples were used to show the distribution of the identified molecules. *m/z* 701.54 and *m/z* 772.55 were more abundant in a healthy mucosa, while *m/z* 887.5725 was in a tumour.

Four patient-matched healthy versus tumour pairs were randomly chosen to demonstrate the distribution of these identified molecules. As shown in Figure 5.9, *m/z* 701.5145, which was more abundant in healthy samples, was present in all normal tissue sections and was especially intense in the mucosa. A histopathologist first described the only tumour sample in which *m/z* 701.5145 was also present as cancer. However, when blindly rechecking the tissue sections, he commented that there were a couple of areas in that particular sample where healthy-looking mucosa was present. This can serve as an explanation of why that molecule was present in that cancerous sample.

On the other hand, [PI(38:3)-H]$^-$ which was identified as more abundant in the tumour, was found to be rather intense in healthy colorectal mucosa as well (Figure 5.10).

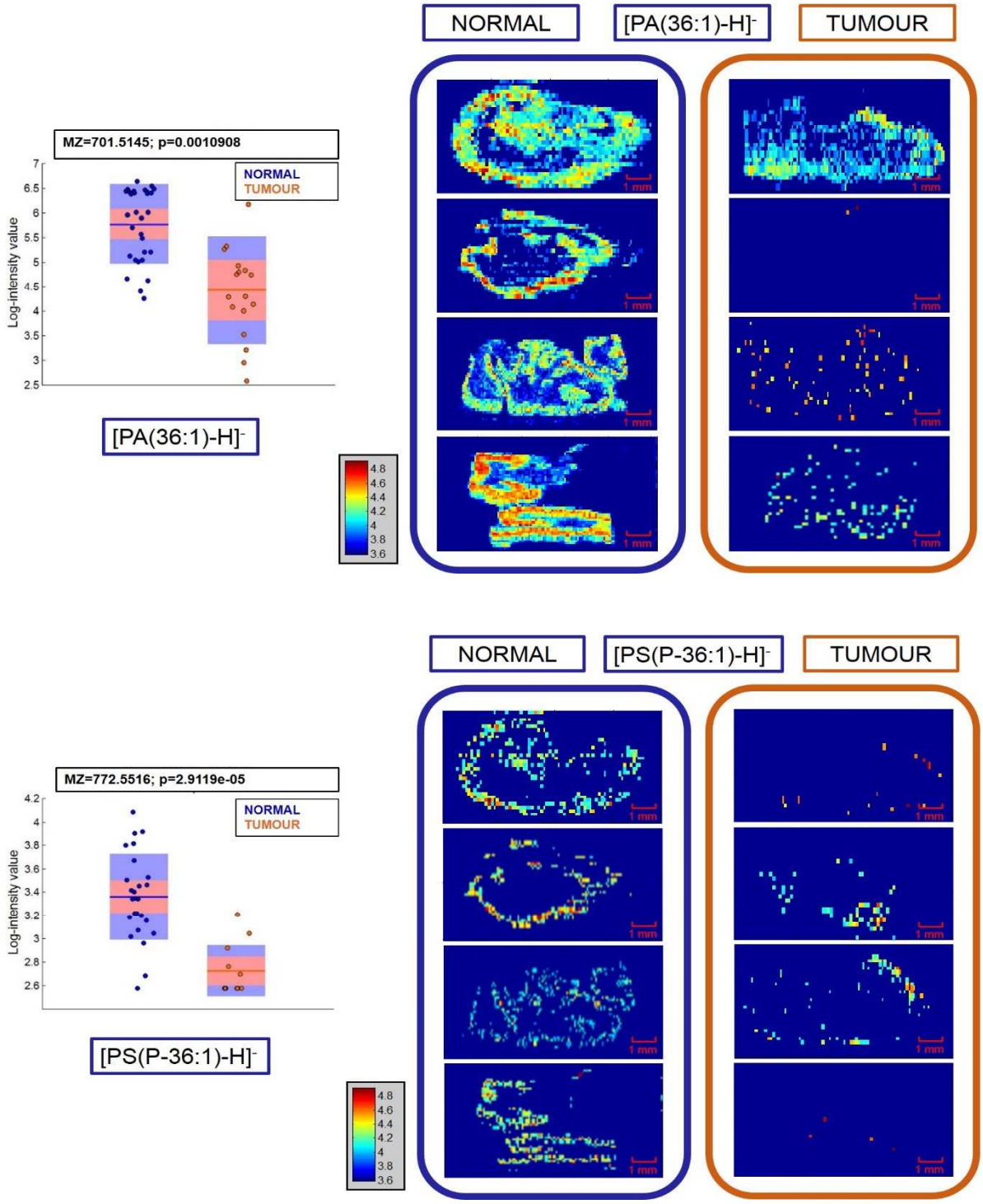


Figure 5.9. An example of molecules that were more abundant in fresh-frozen, healthy colorectal samples. Their distribution is also shown in both healthy and colorectal tumour samples.

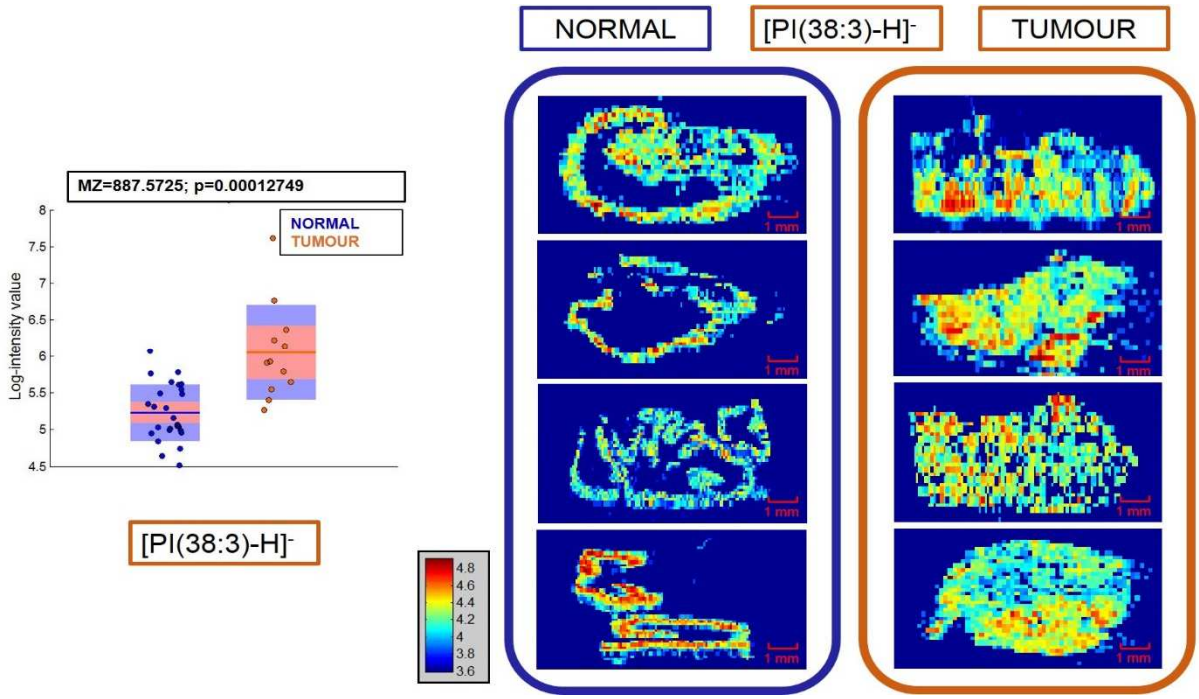


Figure 5.10. *m/z* 887.5725 with a putative ID [PI(38:3)-H]$^-$ was identified as more abundant in colorectal tumour samples. However, it was found in healthy colorectal mucosa too.

5.3.1.3 Prediction of tissue classes in multiple fresh-frozen colorectal tissue samples

Next, 27 healthy colorectal samples were analysed. Three tissue classes were annotated – mucosa, muscle and submucosa. As shown in Figure 5.11A, the unsupervised analysis revealed that muscle and submucosa seemed somewhat similar. In contrast, the metabolomic composition of mucosa seemed to be completely different. However, since submucosa comprises a layer of dense irregular connective tissue that supports the mucosa and joins it to the muscular layer and bulk of overlying smooth muscle, it was expected that it would be similar to the muscle itself. A good separation of these three tissue classes was achieved in the supervised analysis (Figure 5.11B). The leave-one-out cross-validation results confirmed it with an accuracy of 96% for mucosa and over 76% for both submucosa and muscle (Figure 5.11C). Each data point represents a single sample meaning that all the collected information for each particular specimen needed to be averaged first.

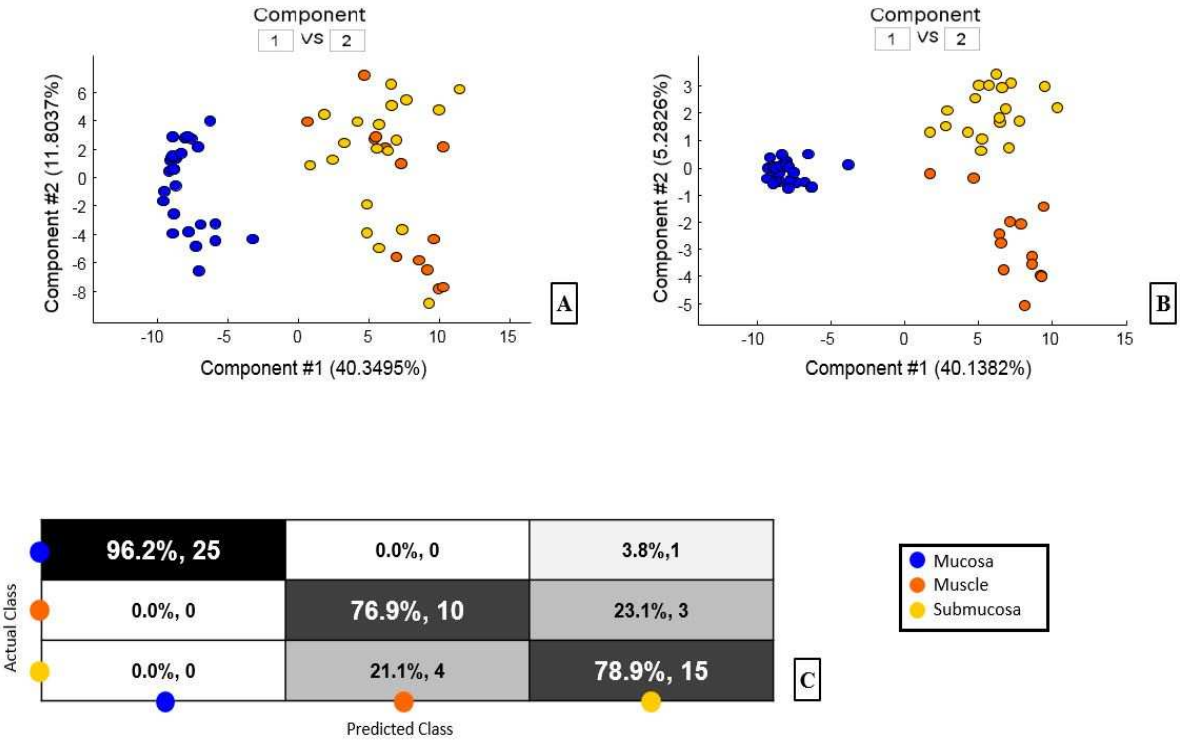


Figure 5.11. Statistical analysis of 27 fresh-frozen healthy colorectal samples. PCA plot (A); MMC+LDA plot (B); leave-one-out cross-validation matrix (C).

The next step was to build a model for normal, adenoma and tumour samples. Only pixels representing mucosa were used to represent the healthy samples. This was as per the histopathologist's advice, as colorectal cancer originates from the mucosa.

The PCA plot did not show a clear differentiation, especially data points matching adenoma samples seemed to be scattered over normal and tumour samples (Figure 5.12A). However, when a supervised analysis was performed, different sample types were clearly classified (Figure 5.12B). An accuracy of 85% was achieved for healthy, 83% for tumour and 25% for adenoma samples when a leave-one-out cross-validation was performed.

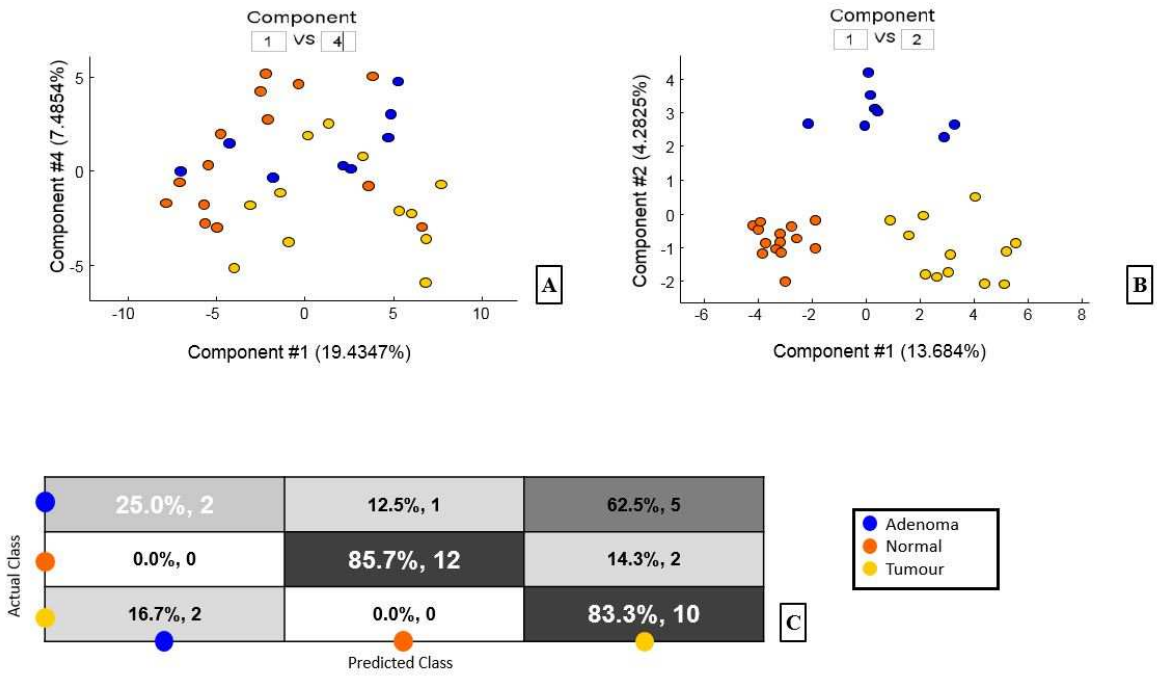


Figure 5.12. Multiple sample analysis results for 14 healthy, 8 adenoma and 12 colorectal tumour samples (fresh-frozen). PCA plot (A); MMC+LDA plot (B); leave-one-out cross-validation matrix with an accuracy of 85.7% for healthy and 83.3% for tumour samples (C).

Adenoma samples were mainly polyps removed during a colonoscopy. Even though a colonoscopy is routinely used, the number of available adenoma samples was deficient when this analysis was run. In order to build a robust model, the number of analysed samples needs to be much higher and balanced for different tissue types included in a study. Therefore, these results can only be treated as preliminary ones. As five out of eight adenoma tissue samples were classified as a tumour, it would be interesting to see patients' clinical history and how the disease has progressed. It might be that DESI-MSI could capture some metabolomic information indicating the potential progression of the disease.

Lastly, an attempt was made to classify only healthy and cancerous samples.

As shown in Figure 5.13, a clear trend was observed when the unsupervised analysis was performed for these two tissue classes. This data was also complemented by a leave-one-out cross-validation plot which confirmed 88.5% accuracy for normal samples and 91.7% for tumours.

The three samples with labels ‘healthy’ that were classified as tumours were identified and assessed by a histopathologist again. It was confirmed that there was no reason to consider

them as cancerous from a histology perspective. It might have happened that those samples were taken very close to the centre of the tumour. Despite tissues themselves looking 'healthy', it might be that their metabolic content suggested the opposite.

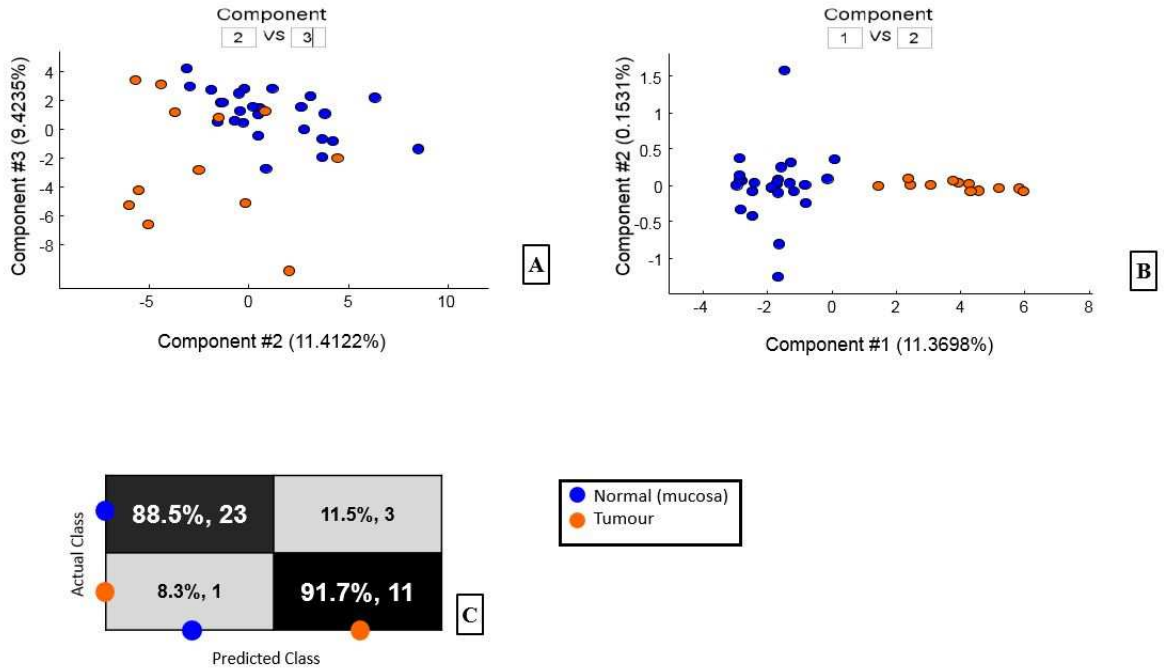


Figure 5.13. Multiple sample analysis results for 26 healthy and 12 colorectal tumour samples (fresh-frozen). PCA plot (A); MMC+LDA plot (B); leave-one-out cross-validation matrix with an accuracy of 88.5% for healthy and 91.7% for tumour samples (C).

5.3.2 High-throughput profiling of fresh-frozen colorectal samples

5.3.2.1 Introduction

The current procedure of analysing tissue sections by DESI-MSI in our lab allows setting up measurements for only one sample at a time. This is because queuing more than one slide means that every tissue section would be exposed to the room temperature during the previous samples' analysis. When tissue specimens are removed from bodies, every care is taken to reduce the time they spend exposed to ambient conditions to avoid degradation of their components. Therefore, unless a reliable system allowing to control the temperature is in place, only one slide can be mounted on a DESI stage and needs to be placed under the nitrogen flow straight away.

At the same time, these days, diagnostic laboratories more and more rely on automation. Introducing new instruments that can perform designated tasks with the minimum input from staff members speeds up the whole process and eliminates human errors. Therefore, it is evident that if DESI-MSI is meant to be introduced to routine diagnostic use, it has to be as user-friendly as possible. That includes the ability to queue the slides awaiting measurements. This feature would also allow using the instrument time much more efficiently as multiple samples could be analysed, for example, overnight rather than just one as it is at the moment.
As a result, multiple efforts have been undertaken to design a new high throughput DESI-MSI platform allowing safe queuing of multiple glass slides awaiting measurements. Dr Emrys Jones from Waters, Wilmslow, UK, has been working on a similar project in collaboration with Professor Takats group from Imperial College London. The preliminary results presented below form a part of this collaboration.
All the initial tests assessing the stability of the DESI-MS setup over time were done by Dr Jones. After running some experiments using pork liver, he concluded that the experimental DESI-MS system could operate at high performance for extended periods. It has been observed that intra and inter tissue spectra were comparable, albeit for a simple test system of the food-grade liver. Therefore, based on the study results, the method was adapted for clinical samples as well.

5.3.2.2 Materials and methods

Fresh-frozen colorectal tissue samples

24 human colorectal samples (12 patient-matched healthy vs tumour pairs) were randomly selected from the colorectal cohort described in *Chapter 2.1.1*.

5.3.2.3 Sample preparation

Fresh-frozen samples were prepared as described in *Chapter 2.7.1*.

5.3.2.4 DESI-MSI instrument setup

Samples were measured in Waters laboratory in Wilmslow. This batch of fresh-frozen colorectal samples was analysed using a Xevo G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA) coupled to a 2D sample stage (Prosolia Inc., Indianapolis, IN, USA). HDI software was used to select the regions of interest on the glass slides. Contrary to all other DESI-MSI experiments performed for projects described in this Thesis, these samples

were measured at a resolution of 50µm. Full details of the DESI-MSI setup for this batch of samples is described in *Chapter 2.9*.

Parameters	Xevo G2-XS QToF (Waters Corporation, Milford, MA)
Scan time	10 scans/ sec
Scan mode	Sensitivity
Mass analyser	TOF
Mass range	50-2000 *m/z*
Ionisation mode	Negative
Sampling cone voltage	-40 V
Source offset	-80 V
Source temperature	120°C
Spray voltage	4.5 kV

Table 5.4. Xevo G2-XS QToF parameters used for the high throughput automated DESI-MSI protocol to analyse fresh-frozen colorectal tissue samples.

The schematic of high-throughput automated DESI is described in *Chapter 2.6*.

5.3.2.5 Data analysis

At this stage, only three ion RGB overlay images were prepared using HDI software.

5.3.2.6 Results and discussion

These are only preliminary results, but they look very promising. In order to be able to make any firm conclusions, the acquired data need to be subjected to proper statistical analysis. Once a qualified histopathologist annotates tissue sections, the usual post-DESI-MSI data analysis workflow will be applied to this dataset.

After completion of 17 hours of DESI-MSI automated run, data for each sample were looked at separately using HDI software. At this stage, the main aim was to assess if it was possible to distinguish different tissue types within each tissue section. In order to do that, three ion RGB images were created.

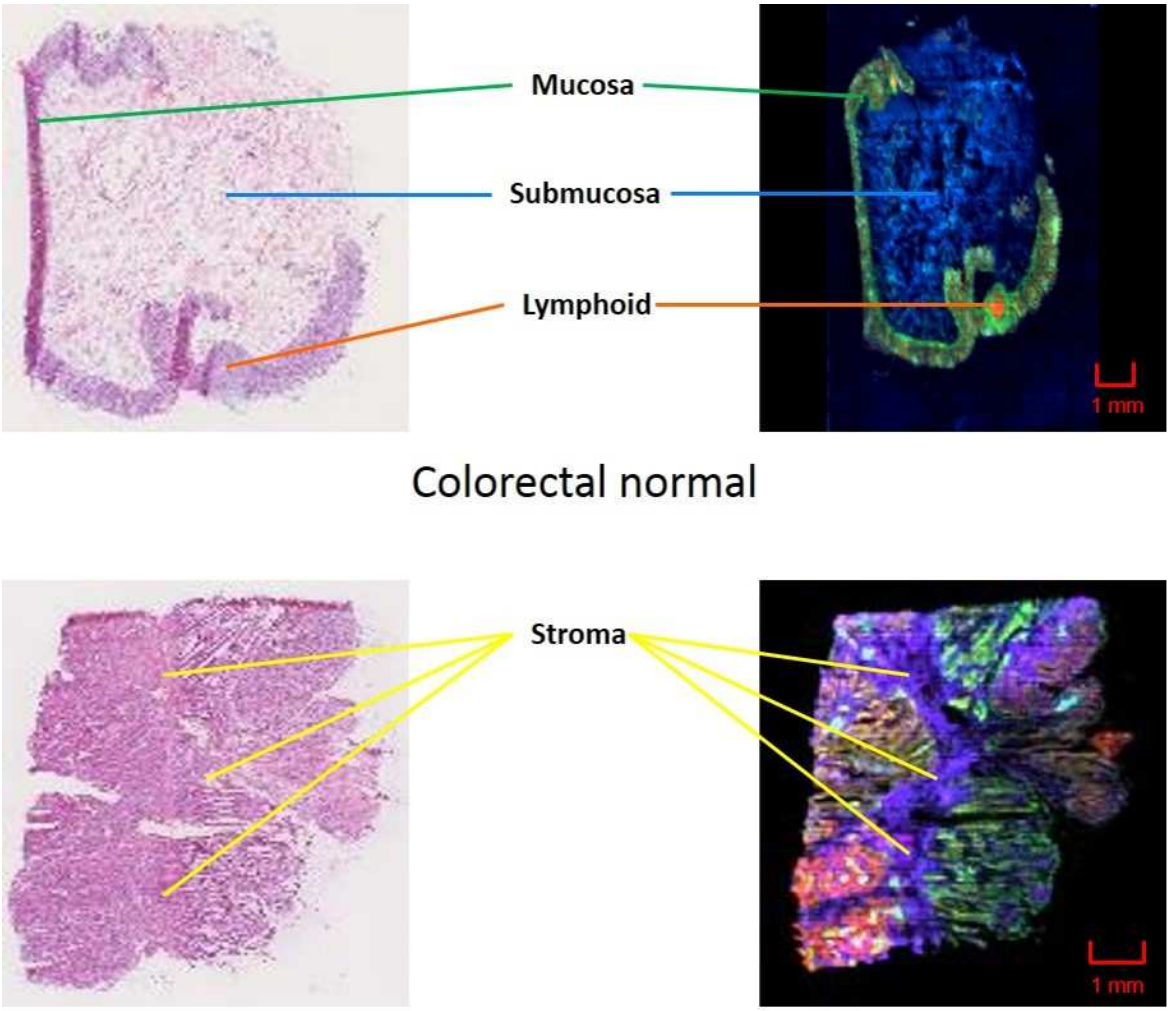


Figure 5.14. Optical images of H&E stained tissue sections (left) and matching RGB image (right). Fresh-frozen healthy and cancerous samples from the same patient were used here as an example. Matching tissue classes are annotated in both H&E and RGB images. RGB images are composed of the following ions: Colorectal normal: green – 738.50 (PE 36:4), blue – 885.54 (PI 38:4), orange – 790.53 (PE 40:6); Colorectal tumour: green – 747.51 (PG 34:1), blue – 773.53 (PG 36:2), orange – 861.54 (PI 36:2).

As it can be seen in Figure 5.14, data acquired at 50µm x 50µm pixel size allowed for the collection of detailed information. Not only mucosa (green) and submucosa (blue) could be seen in the image representing a healthy sample, but also lymphoid tissue (orange) could be visualised. The cancerous sample's RGB image contained even more information and revealed a few different tissue types present in that tissue section. One of them was stroma which is highlighted in yellow. That colorectal tumour sample would be particularly interesting to be reviewed by a histopathologist as, based on its metabolic content, it seems to contain three or

four tissue types. It would be exciting to verify if the same statement could be made from the histopathology perspective.

These two tissue sections were imaged as the third and the fourth out of the whole batch consisting of twenty-four specimens. However, equally encouraging results were seen for all other samples, including those run towards the end. In every case, the RGB image showed an apparent separation of various tissue classes present in each section (Figure 5.15). This observation compared with averaged spectra recorded for the 1st and the 21st sample (Figure 5.16) concluded that tissue sections were stable from the biological perspective. The spectral composition was very similar, and the signal intensity was consistent throughout the run.

As mentioned earlier, all the above are preliminary conclusions as this dataset needs to undergo the usual post-DESI data analysis workflow. The first step would be to get the tissue sections annotated with a qualified histopathologist's help, as all further statistical analysis relies on assigned labels representing various tissue classes.

However, based on Dr Jones's initial tests and the first results from the batch of clinical samples, it seemed that the current DESI-MS system could operate at high performance for extended periods. This is a crucial step towards making the DESI-MSI setup more histology-friendly. Moreover, running whole studies as a single batch rather than over days or weeks should provide more comparable data and contribute to better use of the instrument time.

DESI-MS imaging allows tissues to be classified based on their chemical composition, and multiple levels of information can be obtained quickly from the same analysis with no sample preparation. The relative simplicity of DESI-MS as a molecular imaging platform lends itself to routine use, and improvements in performance and robustness make this even more feasible.

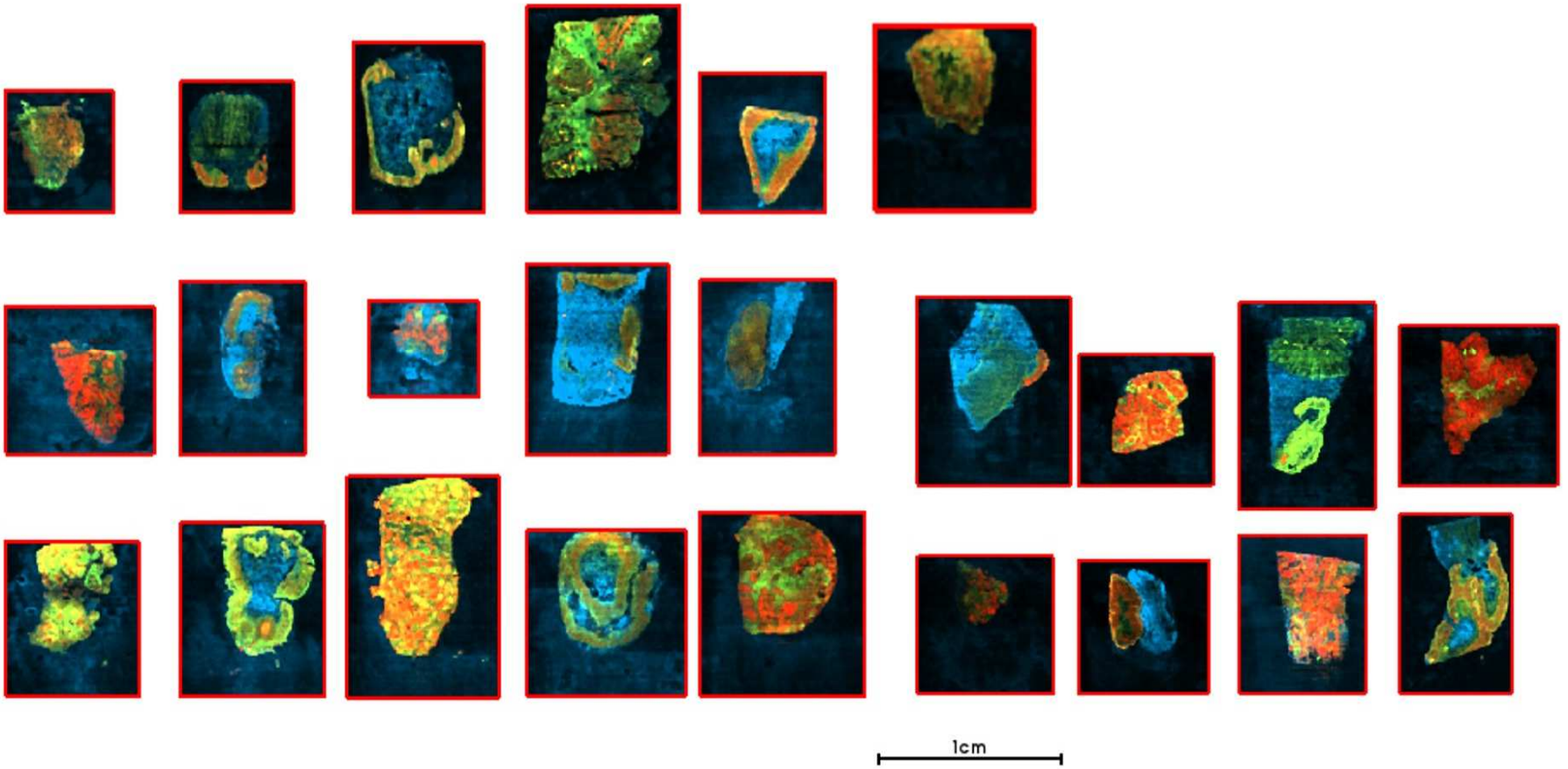


Figure 5.15. RGB images representing all 24 fresh-frozen tissue samples run as one batch during the 17 hours of automated DESI-MSI analysis. Different tissue classes were distinguishable in each tissue section and were represented by different colours in RGB images: green – 747.51 (PG 34:1), blue – 835.53 (PI 34:1), orange – 790.53 (PE 40:6).

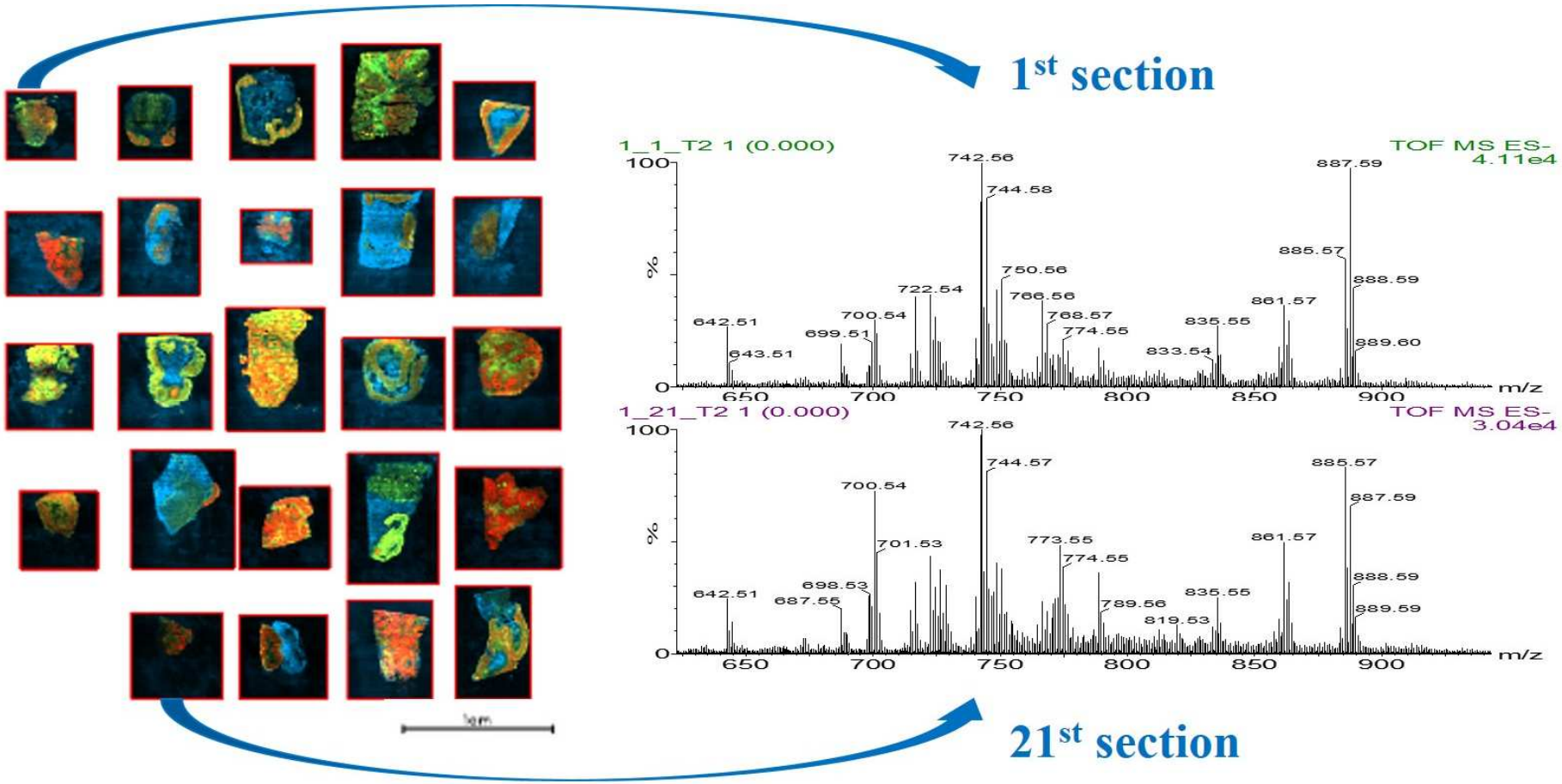


Figure 5.16. Comparison of mass spectra acquired for the first and the twenty-first fresh-frozen colorectal sample analysed during the automated DESI-MS imaging run. Signal intensity and spectral composition were consistent throughout the 17 hours long run. Different tissue classes were distinguishable in each tissue section and were represented by different colours in RGB images: green – 747.51 (PG 34:1), blue – 835.53 (PI 34:1), orange – 790.53 (PE 40:6).

5.3.3 Colorectal samples: Fresh-frozen versus FFPE

5.3.3.1 Introduction

Before commencing to FFPE colorectal samples analysis, a small cohort consisting of fresh-frozen and matched FF tissues was subjected to DESI-MSI. This study was performed to make a quick and simple comparison of metabolic content of precisely the same tissues, which were first processed as fresh-frozen samples, and then fixed in formalin and treated as FFPE.

5.3.3.2 Materials and methods

4 human colorectal samples (2 patient-matched healthy versus tumour pairs) were randomly selected from the colorectal cohort described in *Chapter 2.1.1*.
First, the fresh-frozen samples were prepared, as described in *Chapter 2.7.1*. After preparing all required frozen sections, the same piece of tissue was fixed in formalin for 24 hours and processed as described in *Chapter 2.4*. FFPE samples were then prepared, as described in *Chapter 2.7.2*. 10μm thick sections were cut for both fresh-frozen and FFPE samples.

Fresh-frozen human colorectal samples were analysed in negative ion mode using a high-resolution orbital trapping mass spectrometer (Exactive, ThermoScientific, GmbH).
Full details of the DESI-MSI setup for this batch of samples can be found in *Chapter 2.8*.

FFPE colorectal samples were analysed in negative mode using a Xevo G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA)
Full details of the DESI-MSI setup for the FFPE samples is described in *Chapter 2.9*.

5.3.3.3 Data analysis

This small cohort of colorectal samples was analysed with the help of Dr Paolo Inglese. The same workflow was in place, as described in *Chapter 2.13*. R package was used to perform data analysis.

5.3.3.4 Results

It is important to stress that this small cohort of samples was only run to assess the chemical composition of the same tissues treated as fresh-frozen and then embedded in the wax as FFPE. Since two types of mass spectrometers were used to run DESI-MSI, apparent differences came from the instruments themselves. Orbitrap is known to have excellent mass accuracy, while ToFs have high sensitivity. That was one reason why FFPE samples were analysed on the latter.

Their metabolomic content was expected to be severely impacted by all the washes taking place during sample processing [99, 100]. Other reasons were purely practical and included instrument availability and sturdiness. As described in *Chapter 2.9.*, the sprayer used on Waters ToF mass spectrometer was more robust and easier to optimise, which was an important factor when dealing with a novel application of DESI-MSI.

First, the acquired data were compared in terms of spectral composition. In the fresh-frozen tissue sections, most tissue-related peaks were present in the phospholipid mass range. On the other hand, completely different molecular species were seen in the FFPE imaging data. This was expected and already seen in the results described in *Chapter 4*. The average peak intensity for each sample is presented in Figure 5.17.

Next, two unsupervised analyses – k-means and PCA – were used to see if tissue discrimination was possible, especially in FFPE samples. The objective of k-means is to group similar data points and discover underlying patterns. In order to do so, k-means looks for a fixed number of clusters in a dataset. A cluster refers to a collection of data points that are grouped because they share certain similarities [209]. On the other hand, PCA might be the most popular multivariate statistical technique and functions as a linear transformation of the original variables. It is described as an unsupervised pattern recognition method, and as such no *a priori* knowledge about the dataset is used. PCA projects multivariate data by reducing the number of dimensions in a dataset whilst retaining the necessary information [210].

Results of these two unsupervised statistical techniques are presented in Figure 5.18. Matching H&E-stained tissue sections with histopathological annotations are shown as well as a reference. In every tissue sample, its metabolic content specific to each tissue type was preserved and allowed discrimination of different tissue classes represented by various colours in k-means and PCA images. The colour-coded classification was especially apparent in healthy samples as, by default, those were presented with at least two tissue types.

Sample Fresh-frozen_368_normal contained two tissue types: mucosa and submucosa. However, k-means and PCA images were mainly composed of one colour matching the shape of the mucosa. There was a bit of background surrounding it, which was thought to come from the submucosa part. That particular tissue section was quite destroyed after DESI-MSI, and only the mucosa part was well preserved on the glass once the measurements were completed. DESI-MSI is a non-destructive technique, but tissue sections can be destroyed by the solvent and gas flow on rare occasions if they do not adhere to the slide properly. A consecutive slide was H&E-stained in the case of this particular sample. All the remaining fresh-frozen and FFPE normal colorectal samples were annotated with three labels: mucosa, submucosa, and

muscle/lymphoid tissue. The best statistical results were achieved for the sample FFPE_369_normal, where the k-means and PCA images were composed of three colours, perfectly matching histological annotations (Figure 5.18).
Interestingly, in colorectal tumour samples (both fresh-frozen and FFPE), all the k-means images mainly contained one colour. That classification was correct as those entire tissue sections were classified as tumours. However, PCA images for the samples FFPE_368_tumour and FFPE_369_tumour were composed of a couple of colours. Therefore, it would be interesting to have those sections re-checked by a histopathologist more in detail. Potentially, there could be two tissue types present in tumour samples, too.

It is essential to mention that the H&E images representing the same sample after two different treatments (fresh-frozen or FFPE) do not match. After cryosectioning, each sample had to be fully thawed (so its 'shape' was no longer sustained) and then transferred to a formalin pot where it was loosely floating for 24 hours. As a result, it was impossible to ensure that the same orientation would be maintained when the sample was embedded in wax. Therefore, tissue sections were cut from different sides of a sample resulting in different morphology features seen on the H&E images.

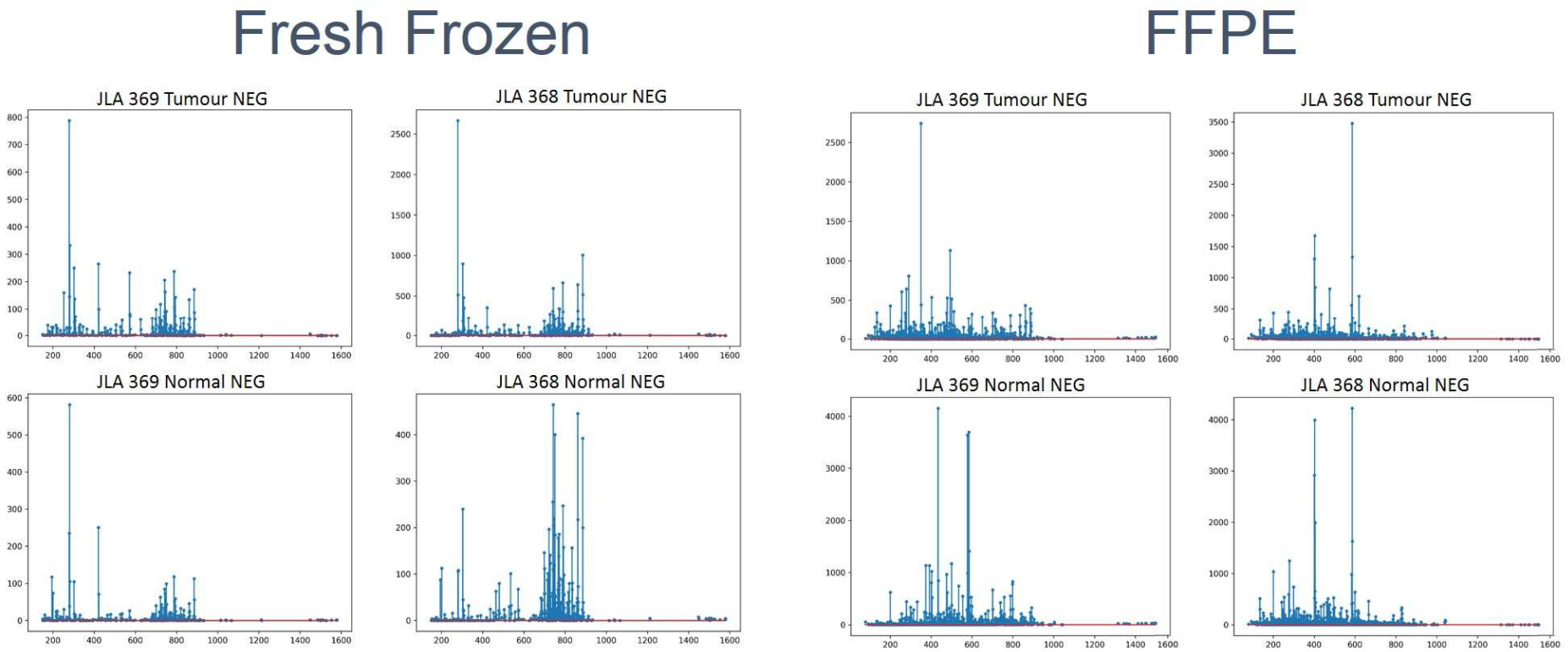


Figure 5.17. Spectra comparison of fresh-frozen vs FFPE colorectal tissue samples analysed in negative mode by DESI-MSI. Average peak intensity is shown for four randomly chosen pairs of fresh-frozen colorectal samples and matching FFPE samples.

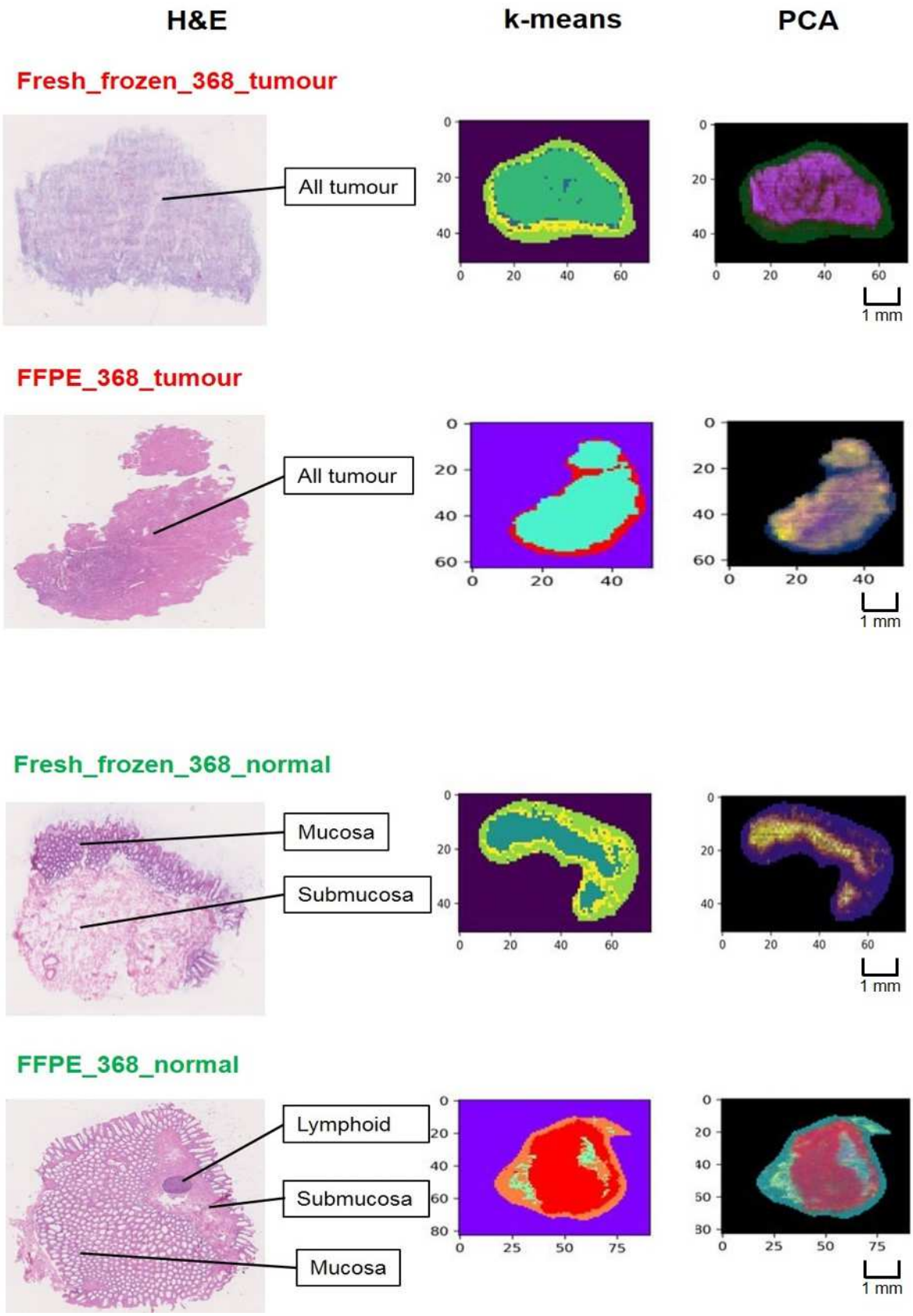


Figure 5.18. k-means and PCA results for a cohort of the same colorectal samples analysed as fresh-frozen first and then as FFPE. Matching H&E-stained optical images are also shown. Different colours in k-means and PCA images correspond to specific tissue classes.

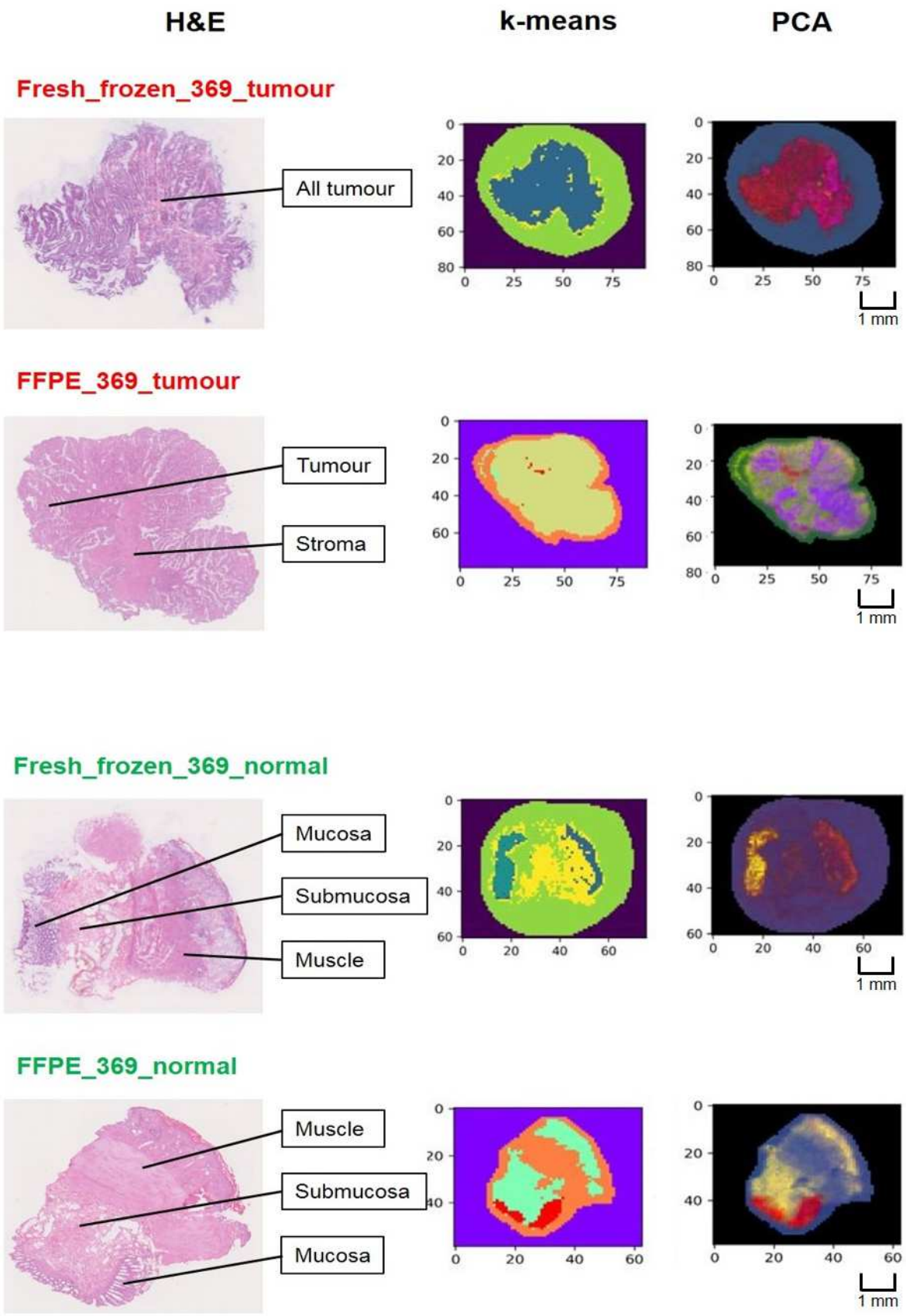


Figure 5.18. k-means and PCA results for a cohort of the same colorectal samples analysed as fresh-frozen first and then as FFPE – continued. In each case, four regions were set in the k-means analysis.

5.3.3.5 Discussion

The results obtained from the analysis of DESI-MSI data acquired for fresh-frozen and matching FFPE colorectal samples agreed with those presented in *Chapter 4*.

Despite extensive washes, the remaining metabolic content of tissue samples was sufficient to allow tissue discrimination. Interesting results were obtained when the same samples were first analysed as fresh-frozen and then as an FFPE, as the PCA images representing FFPE tumours revealed exciting composition. As expected, molecular species seen in fresh-frozen samples were very different from those recorded for FFPE tissue sections. In comparison, the results presented in *Chapter 5.3.1* relied on the changes in phospholipid composition. At the same time, in the case of FFPE samples, the whole mass range was taken into account for the tissue classification. Although some lipids were still present, the 600-900 mass range was noticeably less affluent in the data acquired for FFPE samples. However, the low mass range seemed to contain much information and molecules such as small metabolites were thought to play a crucial role in post-DESI-MSI tissue classification.

After this experiment, it was decided to focus on the novel application of DESI-MSI as a tool that could be used for molecular mapping of FFPE tissue samples.

5.3.4 Analysis of FFPE colorectal samples by DESI-MSI

5.3.4.1 Analysis of a single FFPE colorectal sample - methodology

A human colorectal carcinoma FFPE sample containing non-tumour and tumour tissues was obtained from the Histopathology Department at St. Mary's Hospital, London, UK.

The sample was sectioned into 10 μm thickness with the help of a microtome (Thermo Fisher Scientific Inc, Waltham, MA, USA), hydrated in the water bath until the sections were fully stretched. After this, the tissue sections were mounted onto SuperFrost® Plus Glass slides (Thermo Fisher Scientific Inc) and incubated at 60°C for 30 minutes. The samples were then reserved at room temperature until further analysis.

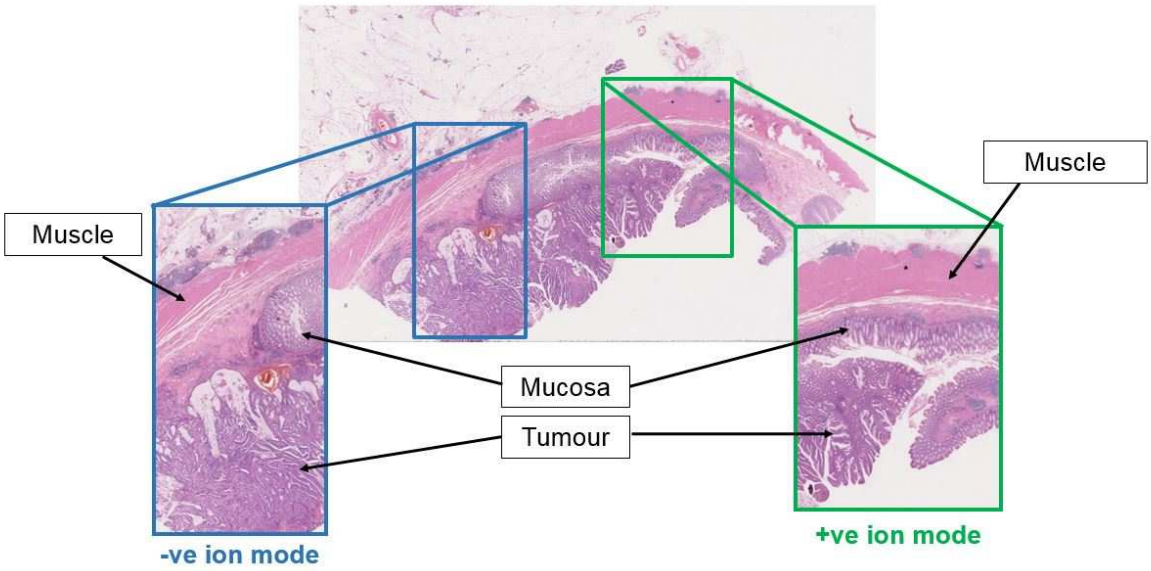


Figure 5.19. H&E stained optical image of the human colorectal FFPE sample analysed by DESI-MSI. Two different areas were selected for negative and positive ionization modes.

Before DESI-MSI analysis, the samples were washed in xylene (histology grade, Sigma-Aldrich, St Louis, MO, USA) for 8 minutes twice. Then, the slides were left to dry at room temperature for 24 hours.

Two 'regions of interest' per sample were selected with High Definition Imaging software (HDI, v1.4, Waters Corporation) for analysis in positive and negative ionisation modes (Figure 5.19).

5.3.4.2 Prediction of tissue types in a single FFPE sample

After DESI-MSI analysis in negative ion mode, the biochemical content of the samples was assessed. There were few spectral peaks compared with non-FFPE sections (Figure 5.20); however, several metabolites remained in the sample after the ethanol and xylene washes utilised during the FFPE sample preparation. It was noticed that the phospholipid mass range had fewer peaks than in fresh-frozen specimens. However, the lower mass range (50-500 *m/z*) contained tissue-specific information, which is in agreement with previous findings [101].

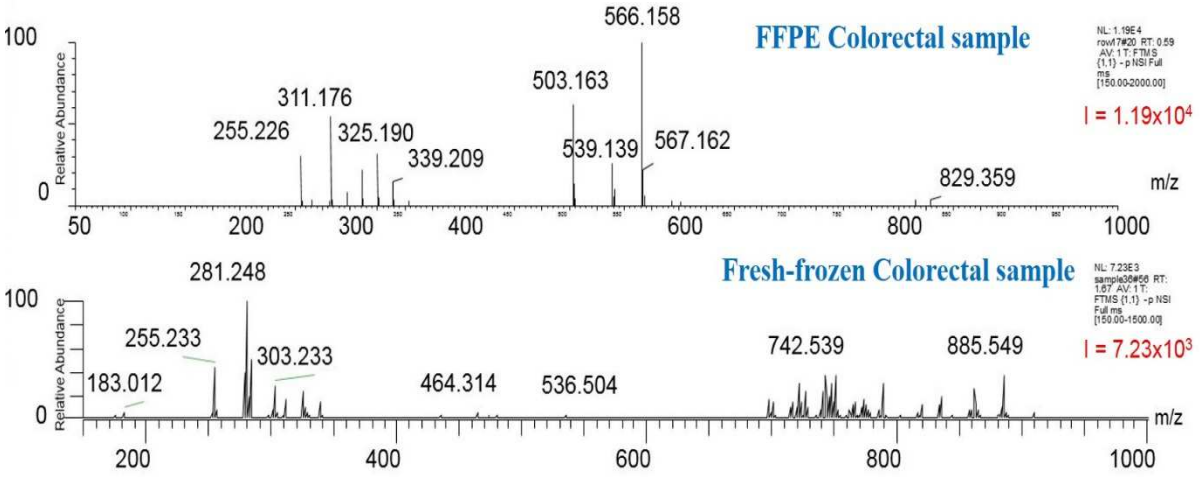


Figure 5.20. Averaged spectra representing normal colorectal samples analysed as an FFPE (top) and fresh-frozen (bottom) by DESI-MSI in negative ionization mode. Noticeably fewer molecular species were seen in the case of the FFPE sample.

As shown in Figure 5.21, several *m/z* ions were localised in specific regions of the tissue sample, suggesting that metabolites' spatial distributions were preserved during FFPE sample preparation. While the preservation of spatial distributions has been previously reported in MALDI-imaged FFPE samples [101], this is the first reported instance when using DESI-MSI. It was possible to observe that *m/z* 619.27 was substantially more abundant in the sample's mucosal region (Figure 5.21A), while *m/z* 650.24 – in the mucin (Figure 5.21B). When these two ions were overlaid as an RGB image (Figure 5.21C), both regions could be compared to the H&E stained areas from the same tissue section's optical image (Figure 5.21D).

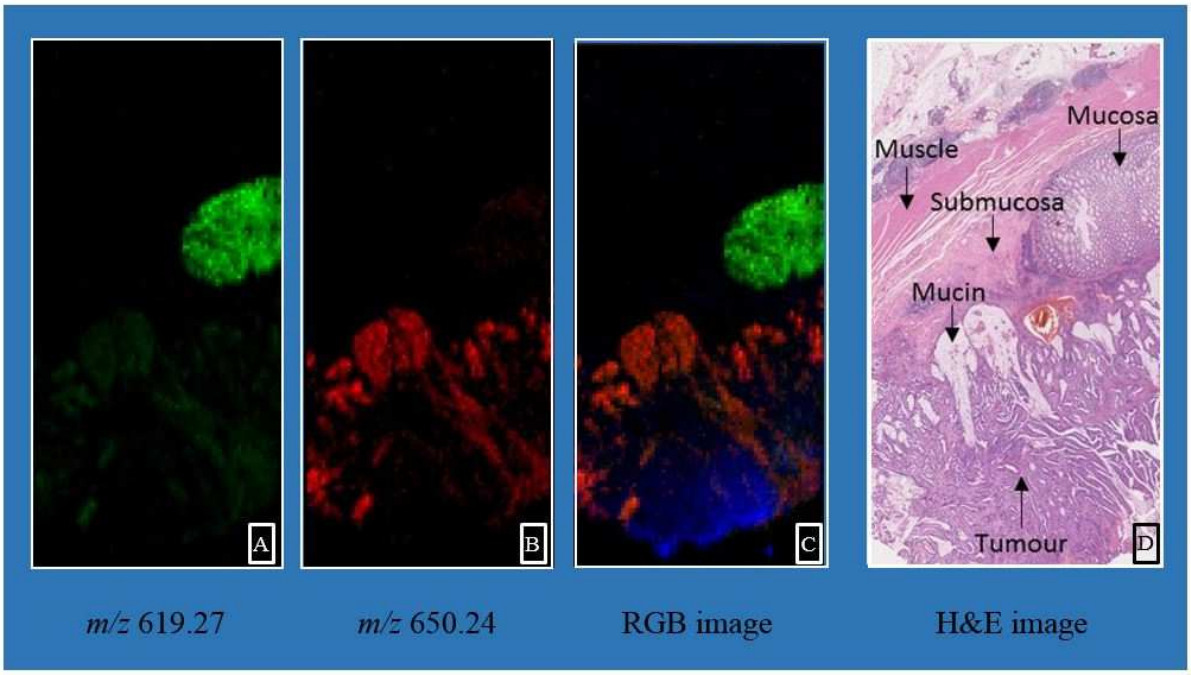


Figure 5.21. DESI-MSI analysis of the human colorectal FFPE sample in negative ionisation mode. Single ion distributions where mucosa (*m/z* 619.27) (A) and mucin (*m/z* 650.24) (B) could be detected. RGB image of overlaid *m/z* ions (*m/z* 619.27 – green, mucosa; 650.24 – red, mucin; 400.89 – blue, tumour) (C). The annotated H&E stained optical image is also shown (D).

Differentiation of the various tissue types was performed using multivariate PCA and MMC methods. PCA scores and loadings plots reflected this normalised and log-transformed data. The ellipses drawn in PCA and MMC scores plots (Figure 5.22A and Figure 5.22B) represented the 100% confidence interval. The PCA scores plot showed a clear separation between the four tissue types in the first two PCs, similar to that achieved using MMC. The cross-validated results from leave-one-pixel-out prediction are shown in the confusion matrix with an overall classification result of 100% for four tissue classes (Figure 5.22D).

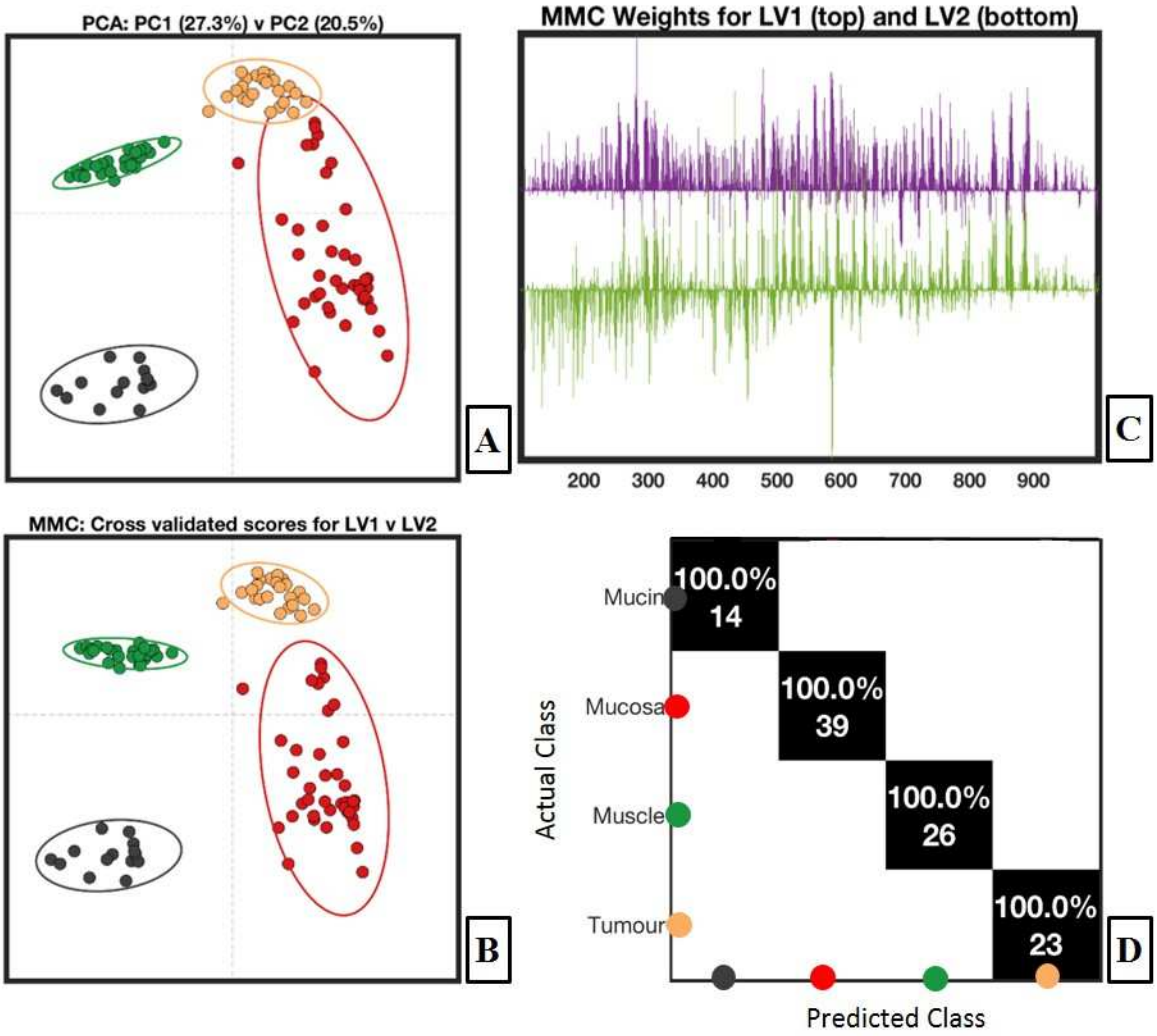


Figure 5.22. Multivariate statistical analysis of a single human colorectal FFPE sample analysed in negative ionisation mode. PCA (A) and MMC+LDA (B) of mucosa, muscle, mucin, and tumour show a clear separation across the first two components. The leave-one-spectrum-out cross-validated MMC scores plot shows a clear separation between the four tissue types. The weights demonstrate that the separation was directed by a few ions (C). The cross-validated prediction accuracy is also shown (D). The mucosa is represented by red, muscle – green, mucin – black, and tumour – yellow.

The MMC-LDA weights identified several discriminatory ions (Figure 5.22B and Appendix I). Those ions were next inspected by hand to select some *m/z* values that were more abundant in various tissue classes present in the analysed sample. Four such ions, all with Kruskal Wallis ANOVA $p < 0.001$, are shown in Figure 5.23. Box plots showed this normalised data, with the grey box representing the 25th and 75th percentiles. The thick black line denoted the median value.

The intensities of individual data points were shown on the axes with a random x-axis positional variation. Putative identifications for these ions were attempted by reference

to the Metlin database (https://metlin.scripps.edu/index.php) using a 30ppm *m/z* deviation. However, there were no possible identifications for some of the ions, which suggested that MS/MS analysis was necessary to confirm their identity. The ions *m/z* 350.1100 and *m/z* 650.2400 (more abundant in mucin) were given putative identifications of N-acetyl-7-O-acetylneuraminic acid and Neu5Acα2-6GalNacα-Thr, respectively. The identification of molecules containing sugar moieties in FFPE tissue samples has previously been reported [101, 206], including sialic acid containing-molecules in mucin and colorectal cancer [211, 212].

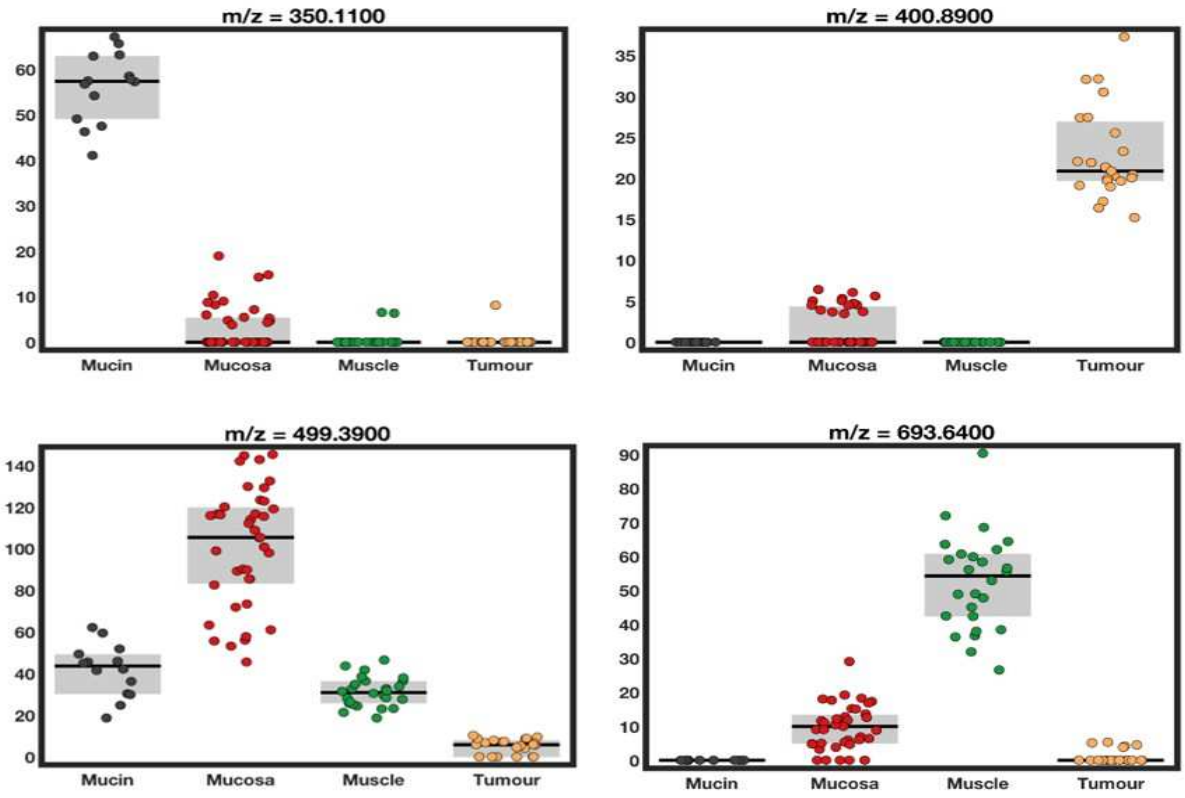


Figure 5.23. Box plots showing examples of ions detected in a human colorectal FFPE sample analysed by DESI-MSI in negative ion mode. *m/z* 350.11 was found to be more abundant in the mucin, *m/z* 400.89 in the tumour, *m/z* 499.39 in the mucosa, and *m/z* 693.64 in the muscle. All ions have Kruskal Wallis ANOVA $p < 0.001$.

The annotated spectra from the previous MMC-LDA analysis were used for the classification of all pixels from the analysed sample. All unambiguously classified pixels were used for PCA analysis. The RGB classification images were generated from a supervised MMC analysis. The histopathology-annotated pixels were normalised and transformed. Then they were used as a training set with which an MMC model was derived. All pixels were classified against this

model in one class-against-all approach, whereby probabilities of a spectrum/pixel belonging to each class were determined. For the negative mode image (containing four tissue classes), these probabilities were expressed in two images as either the red, green or blue channels of an image (Figure 5.24).

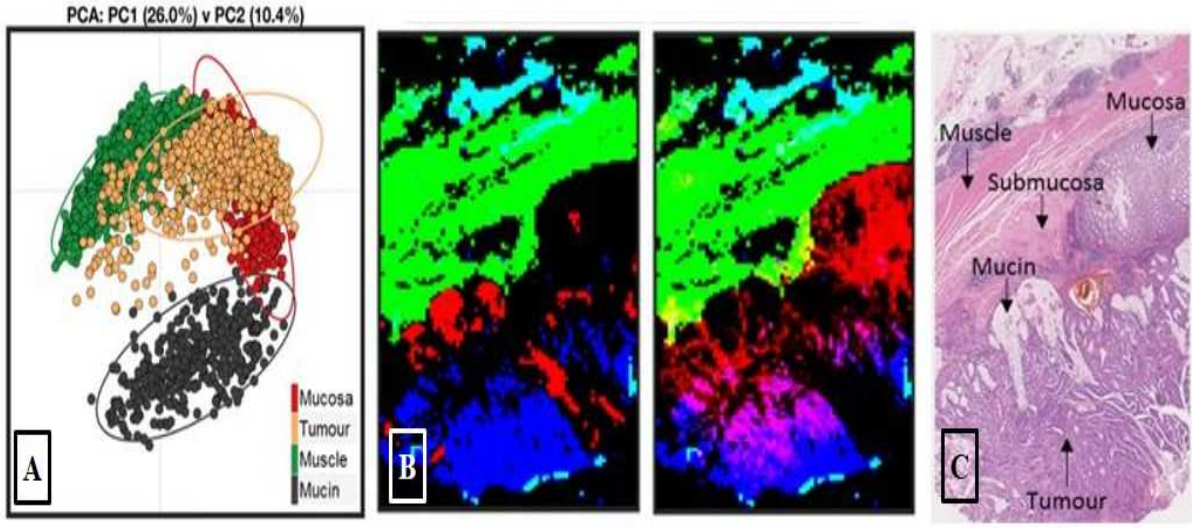


Figure 5.24. Multivariate statistical analysis of a single human colorectal FFPE sample for all pixels. PCA of muscle, mucin, mucosa and tumour shows a separation across the first two components (A). The MMC components of RGB images are also shown (B). H&E stained matching tissue sections with annotations of various tissue classes.

Another area of the same sample (see Figure 5.19) was analysed in positive ionisation mode to determine if tissue differentiation could also be observed using DESI-MSI. The same parameters as described for negative mode analysis were employed. Also, the same statistical approach was used for tissue classification using positive ionisation data. The clustering of the different tissue types was observed from the PCA, where muscle tissue could be separated from the other tissue types (Figure 5.25). In the supervised analysis and subsequent cross-validated predictions, tumour tissue was well separated from muscle and mucosa in the first and second components (Figure 5.25). From the PCA classification using all the pixels performed as described for negative ion mode, clustering of the different tissue types was attained. The RGB classification images obtained from the supervised MMC analysis were in significant agreement with the H&E optical image (Figure 5.27). For the positive mode image (containing three tissue classes), these probabilities were expressed as either the red, green or blue channels of an image. Examples of ions identified as being discriminatory are shown in boxplots in Figure 5.26 and Appendix J.

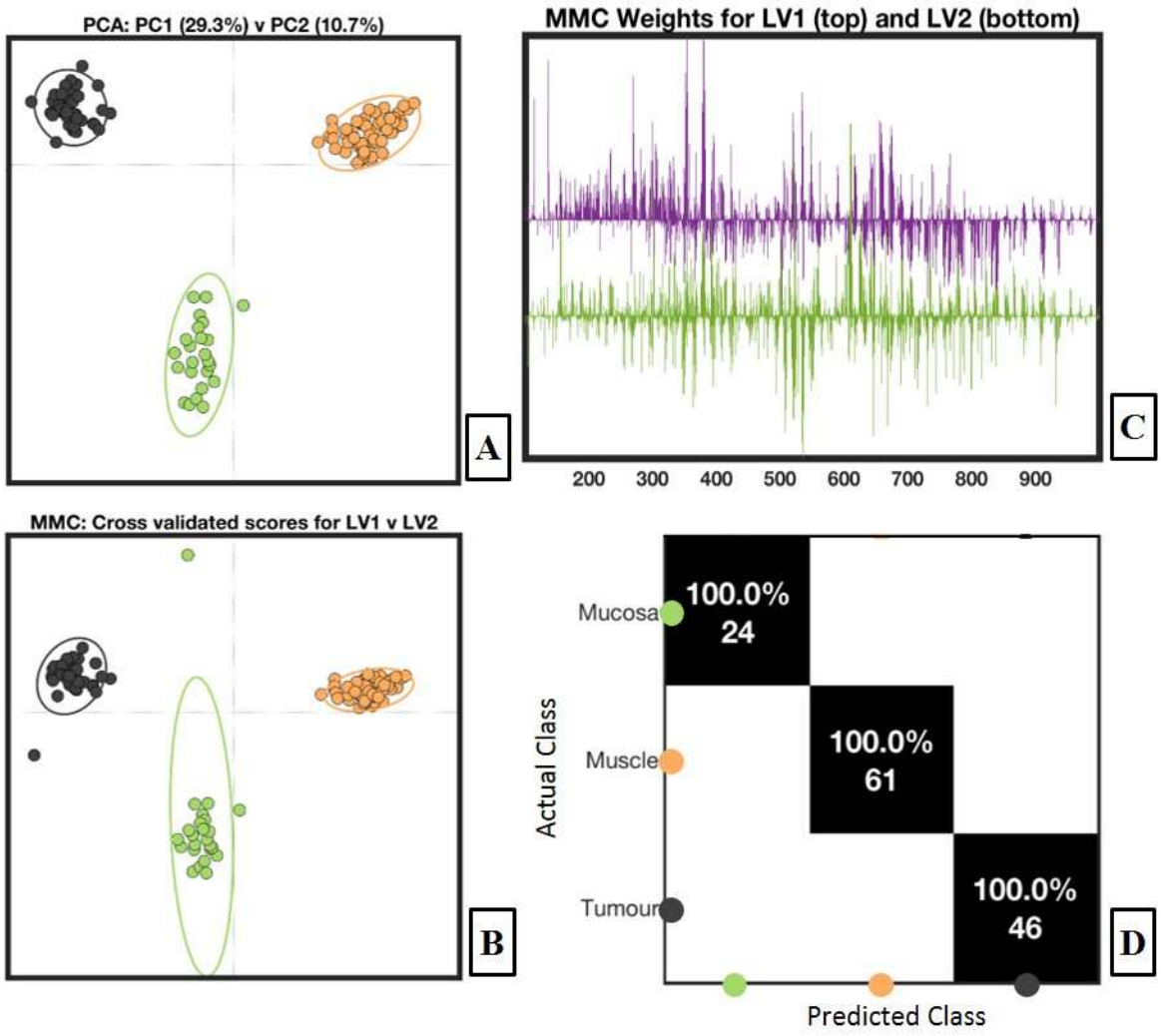


Figure 5.25. The human colorectal FFPE sample analysed by positive mode DESI-MSI. Clear separation between the three tissue types was visible in the PCA (A) and leave-one-spectrum-out cross-validated MMC+LDA (B) scores plots. The weights demonstrated that a larger quantity of ions than observed in the negative mode was causing the between-group separation (C). The cross-validation accuracy of the multiple annotated tissue classes is also shown (D). The mucosa is represented by green, muscle – yellow, and tumour – black.

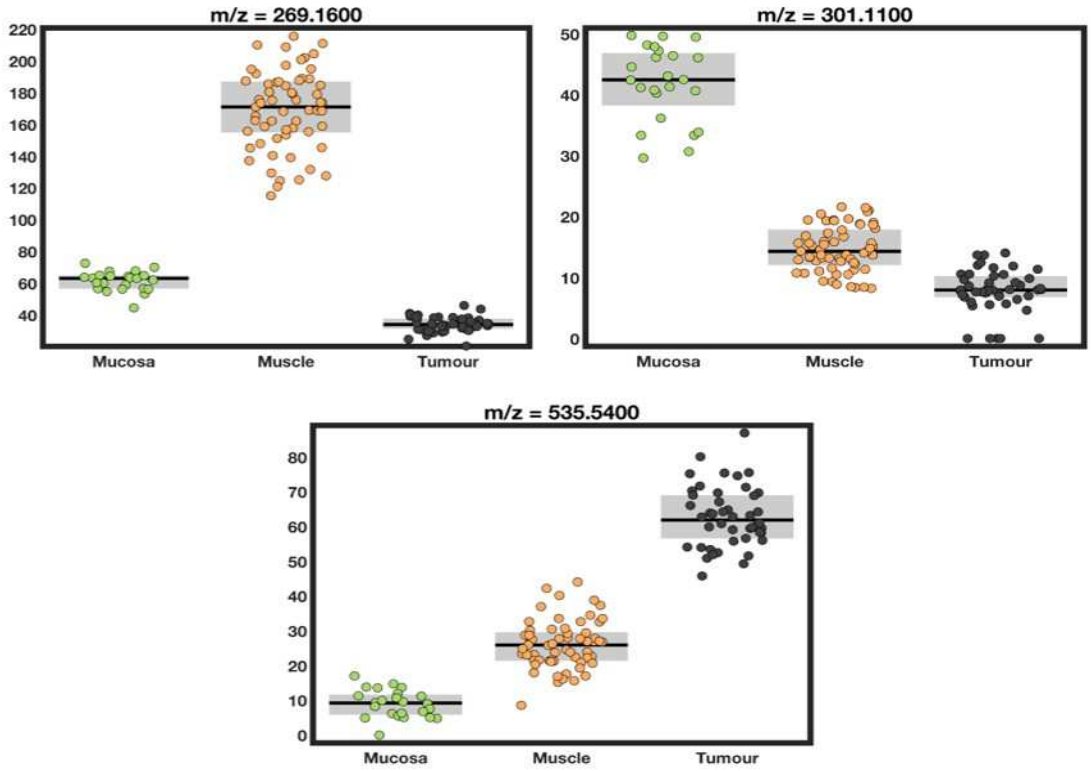


Figure 5.26. Box plots of ions detected in a human colorectal FFPE sample analysed by DESI-MSI in positive ion mode. The MMC-LDA weights identified several discriminatory ions, which were then inspected by hand to choose more abundant molecules in each tissue type. *m/z* 269.16 was found to be more abundant in the muscle, *m/z* 301.11 in the mucosa, and *m/z* 535.54 in the tumour. All ions have Kruskal Wallis ANOVA $p < 0.001$.

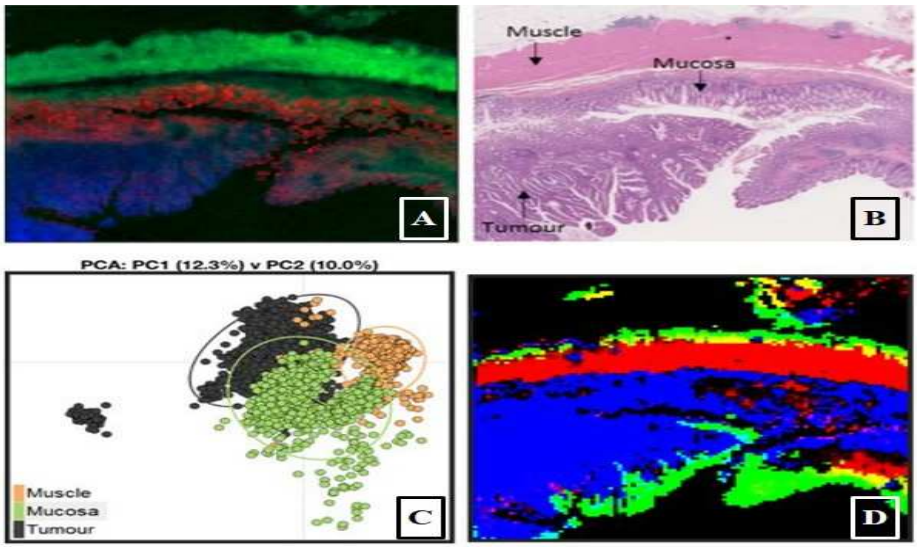


Figure 5.27. DESI-MSI analysis of a human colorectal FFPE sample in positive ionisation mode. RGB image of overlaid *m/z* ions (*m/z* 269.16 – green, muscle; 665.38 – red, mucosa; 349.12 – blue, tumour) (A). H&E stained optical image of tissue sample analysed (B). PCA of all the pixels (C) and MMC components RGB image is also shown (D).

5.3.4.3 DESI-MSI of a human colorectal tissue microarray (TMA)

5.3.4.3.1 Introduction

The tissue microarray was first described by Kononen in 1998 [213] and represented a high-throughput technology allowing the assessment of histology-based laboratory tests. Tissue microarrays are paraffin blocks that contain many small representative tissue samples from hundreds of different cases. The construction of a TMA block is shown schematically in Figure 5.28. Core tissue biopsies are taken from individual FFPE donor blocks and then precisely arrayed into a new recipient block [214, 215].

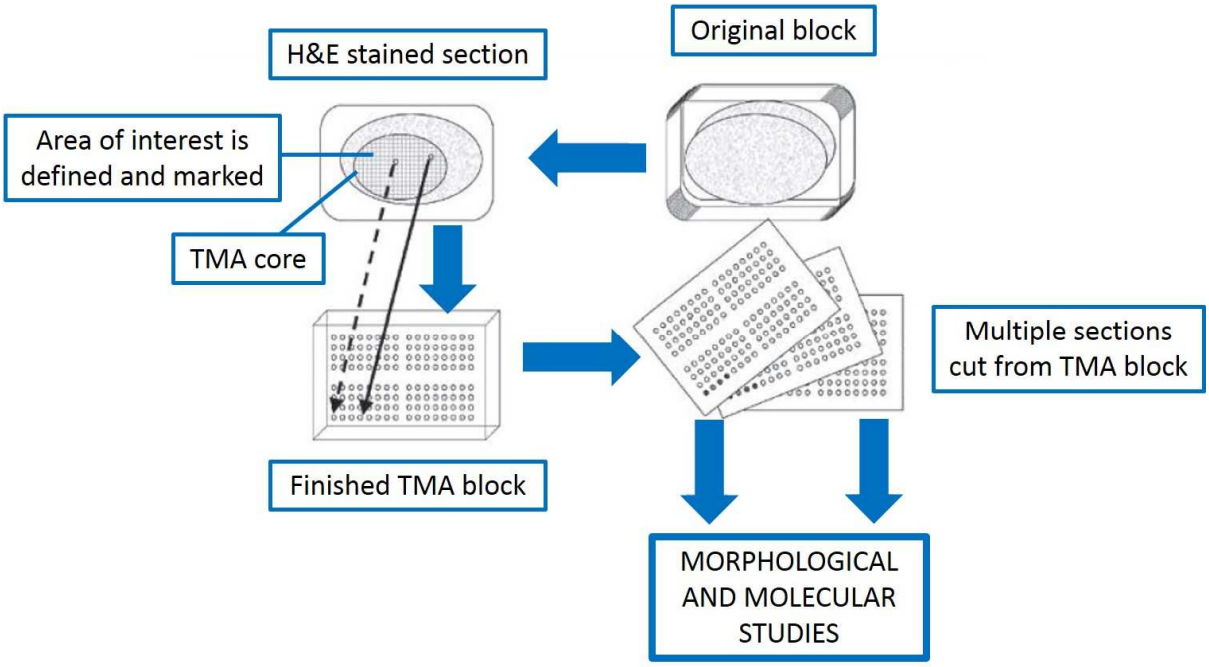


Figure 5.28. Schematic of the process of constructing a tissue microarray. Adapted from [216].

From a mass spectrometry point of view, the ability to analyse FFPE TMAs would be a significant advancement allowing for rapid measurements of hundreds of tissue cores. Therefore, this project aimed to assess if TMA sections were compatible with DESI-MSI requirements.

5.3.4.3.2 Materials and methods

Two thicknesses – 5µm and 10µm – human colorectal tissue microarray slides were purchased from Biomax.us (https://www.biomax.us/).

The layout of the TMA is presented in Figure 5.29. The total number of cores was 54, and samples from 18 patients were used to construct this TMA. The diameter of each core was 1.5 mm.

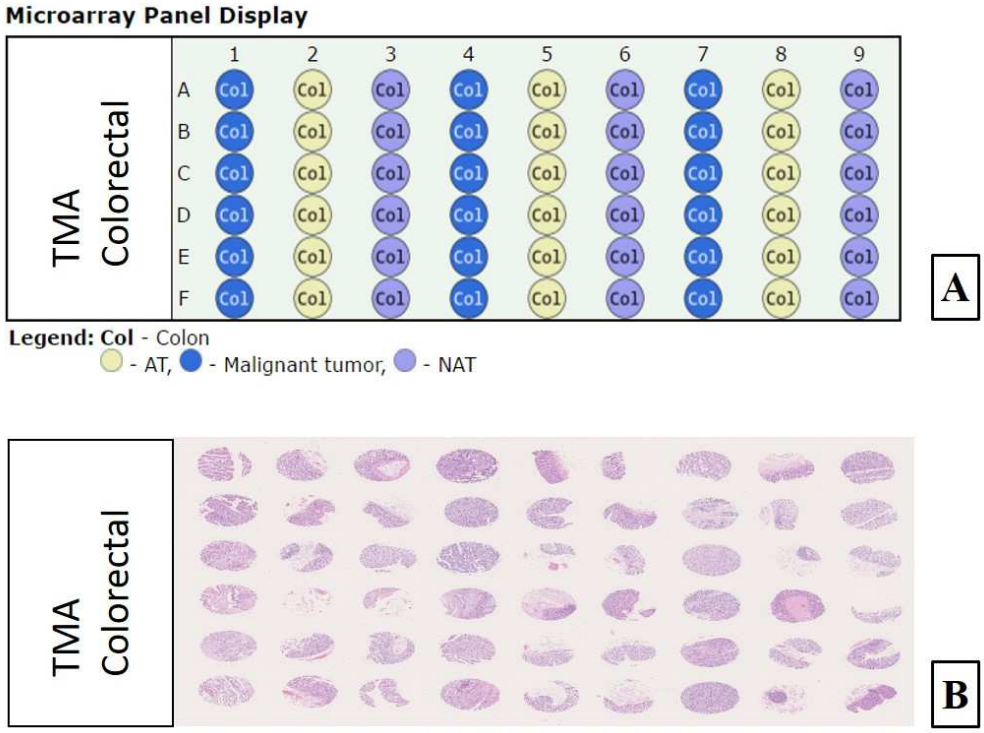


Figure 5.29. The layout of the human colorectal FFPE TMA. Microarray panel display: AT – adjacent tissue, NAT – normal adjacent tissue (A); H&E stained digitalised image of the TMA slide (B).

FFPE TMA slides were then prepared as described in *Chapter 2.7.2*.

FFPE TMA colorectal slides were analysed in negative and positive mode using a Xevo G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA)

Full details of the DESI-MSI setup for the FFPE samples is described in *Chapter 2.9*.

5.3.4.3.3 Data analysis

Full details can be found in *Chapter 2.13* and *Chapter 5.2.6.2*.

5.3.4.3.4 Results

First, a TMA section cut at 10µm was analysed in the negative ionization mode. It is important to stress that the analysis time for 54 samples was 12 hours.

Four tissue types were annotated with a histopathologist's help: mucosa, submucosa, lymphoid tissue, and tumour. The unsupervised analysis revealed a clear trend in the grouping of those tissue classes. The data points clustered together during the supervised approach, where a clear trend was seen. That observation was confirmed by the results of a leave-one-out cross-validation test where an accuracy of over 90% was achieved for mucosa and tumour (Figure 5.30).

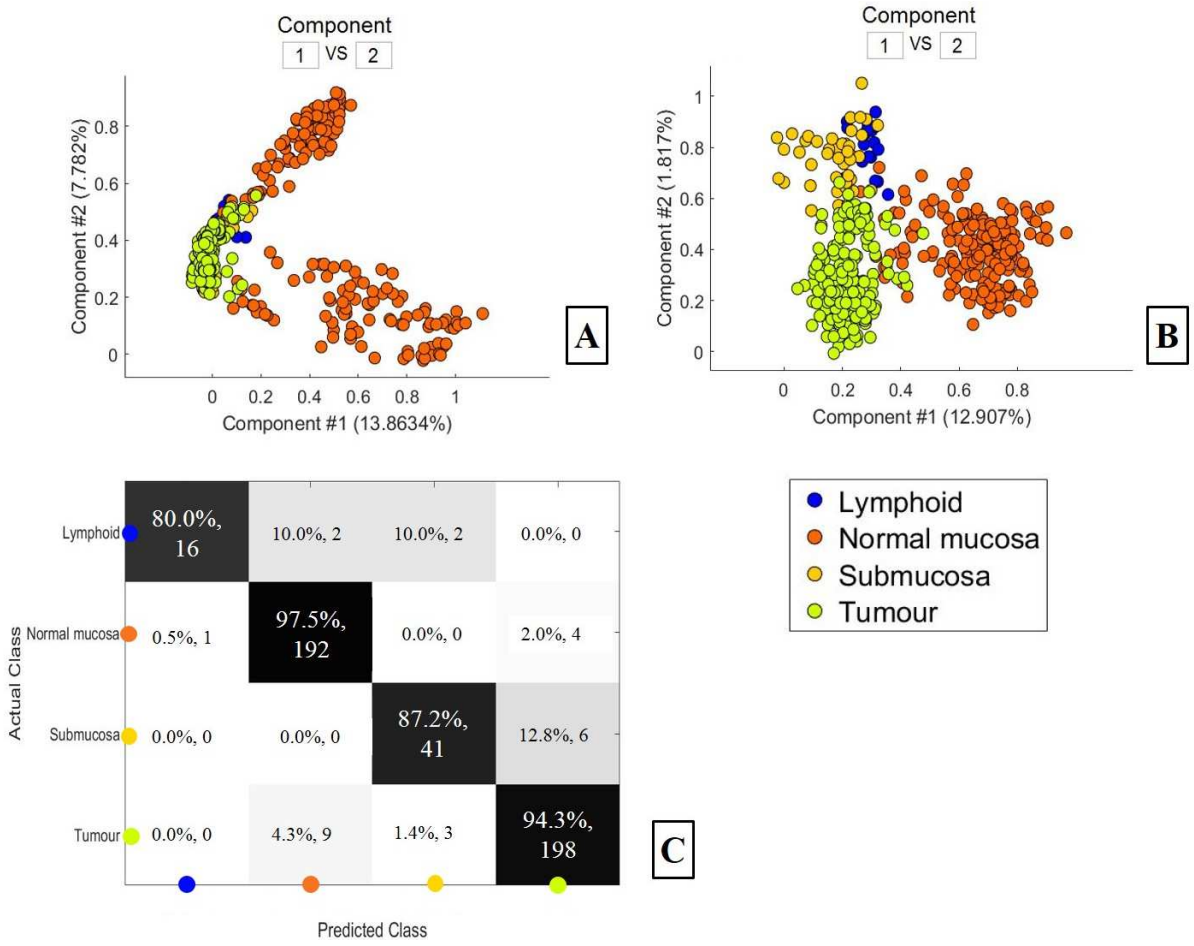


Figure 5.30. 10µm thick colorectal TMA analysed by DESI-MSI in negative ionisation mode. Four tissue labels were annotated for this slide: lymphoid, mucosa, submucosa, and tumour. Results of an unsupervised analysis (A); a plot representing the outcome of supervised analysis (B); leave-one-out cross-validation results with Mahalanobis classifier (C).

Similarly, the 10µm thick TMA section was subjected to the DESI-MSI followed by labelling the tissue classes. This time only two cores were seen to contain lymphoid tissue, and as a result, this tissue type was excluded from further statistical analysis.

PCA plot represents the results of an unsupervised analysis where data points labelled as mucosa and tumour formed well-defined clusters. An apparent trend was seen in the supervised analysis, where three groups of data points were observed. An accuracy of over 90% was achieved in the leave-one-out cross-validation (Figure 5.31).

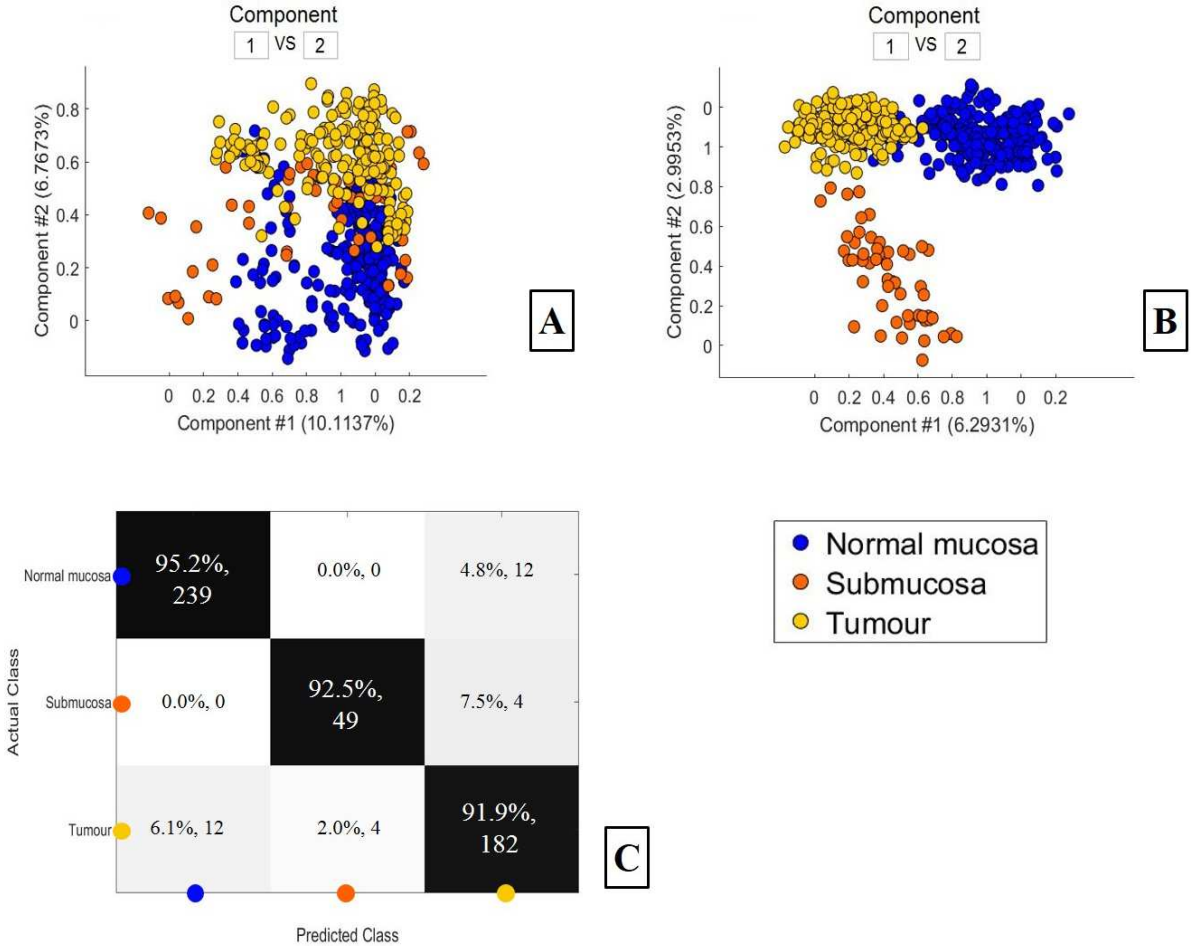


Figure 5.31. 10µm thick colorectal TMA analysed by DESI-MSI in positive ionisation mode. PCA (A); maximum margin criteria analysis (B); leave-one-out cross-validation matrix using Mahalanobis as a classifier (C).

The results obtained for negative and positive mode data confirmed that tissue classification was possible for FF samples. Moreover, it was achievable for tiny, 1.5 mm diameter tissue cores used to construct a TMA. It was a significant finding which, without any doubt, would

play a crucial role in the whole process of making DESI-MSI more compatible with routine diagnostic use.

As has already been mentioned in *Chapter 4.3.14*, routine histopathology laboratories do not cut 10µm thick tissue sections as a standard. Therefore, 5µm sections were requested for the same TMA and were subjected to DESI-MSI. All the experimental conditions were the same as in the case of 10µm slides.

The statistical results for the slide analysed in the negative ionization mode are shown in Figure 5.32. An apparent trend was seen even when an unsupervised analysis was attempted. The highest accuracy was achieved for mucosa, and it reached over 98%. Correct tissue prediction was equally high for submucosa (88,4%) and tumour (89,4%).

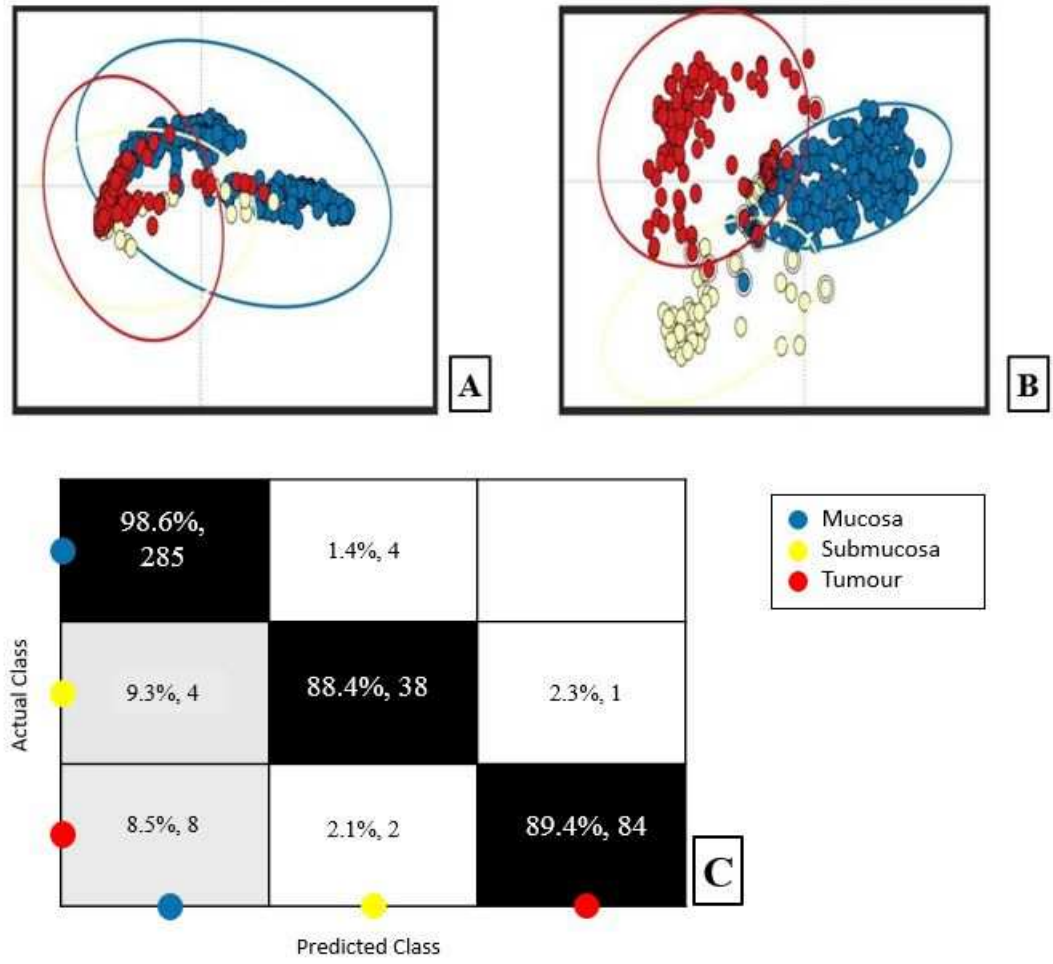


Figure 5.32. 5µm thick colorectal TMA analysed by DESI-MSI in negative ionisation mode. PCA (A); RMMC analysis (B) cross-validated with the respective leave-one-patient-out cross-validation using Mahalanobis as a classifier (C).

Statistical results for the positive mode data were even more impressive (Figure 5.33). This time, it was possible to label four tissue classes, and again, a clear separation was seen in the PCA plot. With the cross-validated RMMC, it was possible to separate all four classes with an overall accuracy of over 95%.

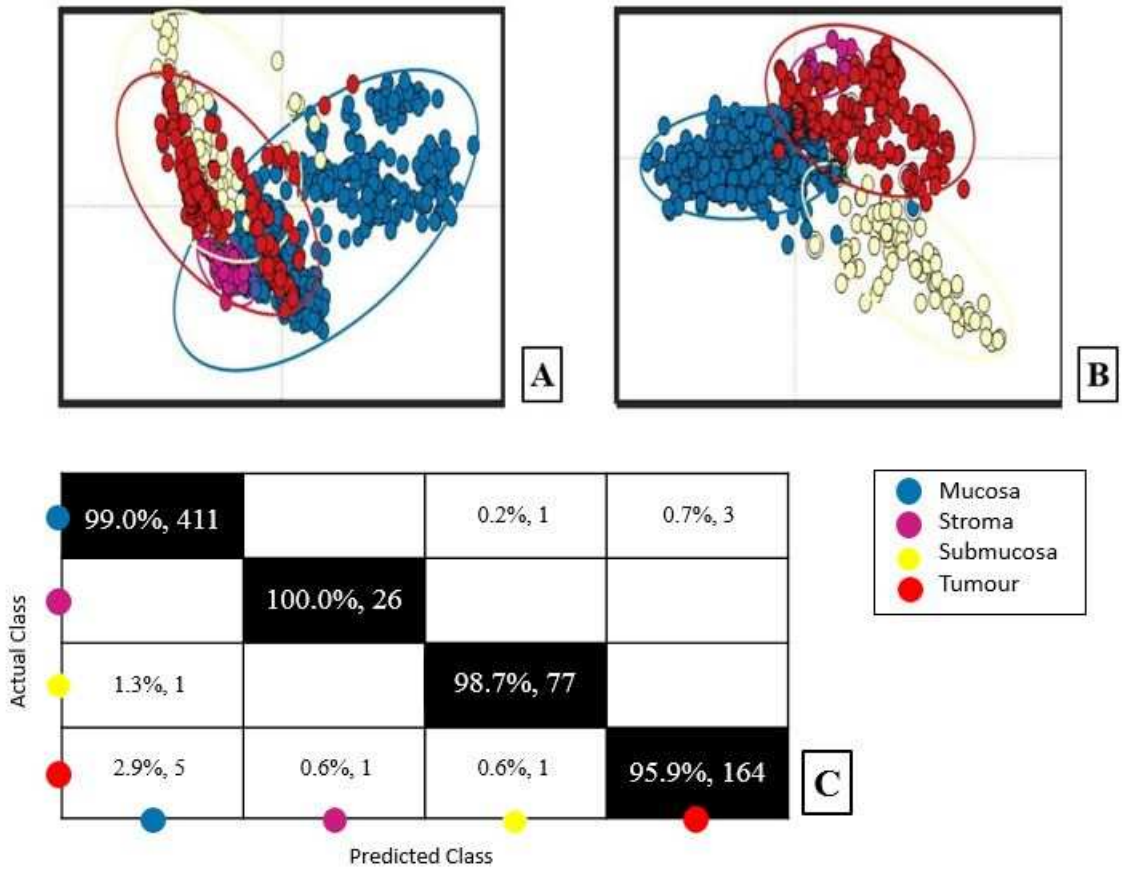


Figure 5.33. 5µm thick colorectal TMA analysed by DESI-MSI in positive ionisation mode. Four tissue classes were assigned to this slide and were used for the statistical analysis. Results of an unsupervised PCA (A); supervised RMMC (B) cross-validated with the respective leave-one-patient-out cross-validation using Mahalanobis as a classifier (C).

Overall, higher accuracy was achieved in the leave-one-patient-out cross-validations for the 5 and 10µm samples analysed in the positive ionization mode. It was related to the fact that the row data recorded in the positive mode seemed to contain more molecular species than the negative mode one. This effect was also seen during the optimisation experiments described in *Chapter 4*.

5.4 Discussion

In the last two decades, imaging mass spectrometry has been playing an essential role in histological research. One of the essential milestones allowing that was the development and application of ambient ionisation methods, such as DESI-MSI.

DESI was first reported in 2004 [78], and since then, it has been demonstrated for many potential applications, mainly involving biological tissue analysis [136, 137, 217]. To date, most studies involving DESI-MSI have relied on the use of fresh-frozen tissue samples.

Although there have already been studies using fresh-frozen samples that have successfully used DESI-MSI to study colorectal cancer further [88, 218], the first part of this Chapter also focused on the same approach. However, the ultimate goal was to apply this technique to FFPE samples for the very first time.

The data collected for the fresh-frozen colorectal tissue samples revealed characteristic molecular differences that allowed the classification of various tissue types within single samples. Multivariate pattern recognition methods such as PCA and MMC+LDA have been shown to provide user-independent data in individual samples and for a larger cohort. Tissue-specific molecular ion signatures allowed to achieve an accuracy of 96% for mucosa and over 75% for muscle and submucosa in a leave-one-out cross-validation for a cohort of 27 normal samples.

Similarly, 88% and 91% accuracy was achieved for normal and tumour samples, respectively, when a batch of 38 samples was subjected to statistical analysis. These results correspond to those obtained by other similar projects [88, 114], confirming that DESI-MSI can successfully analyse the metabolic content of fresh-frozen colorectal samples to allow tissue differentiation. As this has been well established by now, the goal is to make this technique as automated as possible. Ideally, queueing samples would be the desired option, allowing for better use of the instrument time and reducing staff members' input.

MSI offers the possibility to chemically map morphological regions of interest, contributing to the development of next-generation prognostic biomarkers. Moreover, it provides the opportunity to work on automated approaches for tissues classification based entirely on their molecular ion patterns [219]. With this idea in mind, a new high throughput DESI-MSI platform was used to analyse 24 fresh-frozen colorectal samples. After completing 17 hours of the automated run, the collected preliminary data suggested that the new set-up was stable. RGB images showed an apparent separation of various tissue classes present within analysed

tissue sections for the samples queued at the beginning of the run and the very last ones. Without any doubt, this was an essential first step towards making DESI-MSI a more histology-friendly technique. This project will continue as high-throughput, automated DESI is a must if it is meant to be introduced into the routine diagnostic services.

As it has already been mentioned since its introduction, DESI required the use of frozen sections. However, many reasons make FFPE tissues a desirable material as these samples are considered the gold standard for histopathological analysis. As a result, vast archival collections of FFPE tissue samples with long-term clinical data are readily available.
This first proof-of-principle study examined whether it was possible to conduct metabolite measurements from a single FFPE sample and multiple FFPE tissue cores embedded in one block (tissue microarray) by DESI-MSI. The only similar paper available when this was carried out was by Buck *et al.* [21] and looked into MALDI-MSI's suitability for global metabolite analysis from FFPE tissue samples. However, it is crucial to stress again that these two MSI techniques require different sample preparation, and as a result, the obtained results are not entirely comparable. MALDI-MSI requires the use of the matrix, which is coated on the surface of the sample. On the other hand, one of the benefits of DESI is that it combines the advantages of electrospray ionization and other desorption ionization methods with no need for matrix application. As a result, the sample is only subjected to minimal damage, and the matrix's complicated ionization interference can be avoided [220].

It was observed that peaks of metabolites were detected in the low mass range – *m/z* 50 – 400. Simultaneously, a significant reduction of peak intensities was seen above the mass range of *m/z* 600. This observation was correlated with the expectation that tissues processing consisting of multiple washes in gradients of alcohol and xylene might remove lipids and was also reported by Buck *et al.* [101].
However, despite this, several lipids were detected and classified as significant features in all three tissue classes – mucosa, muscle, and tumour – in a single human colorectal sample analysed in *Chapter 5.3.4.1* (Supplementary information, Appendix I&J). This could be due to some cellular components such as lipids, nucleic acids, and carbohydrates, which might not have been directly fixed by formalin, but could be trapped in the network of insoluble cross-links between proteins [221]. A Fourier Transform Infrared (FTIR) spectroscopy imaging study on prostate FFPE tissue sections hypothesized that although methylene chains of free, unbound tissue lipids were removed from the tissues, solvent-resistant lipids remained present in FFPE

tissue samples. Those lipids were thought to be locked into protein-lipid complex matrices, predominantly in the membranes [222].
Buck *et al.* [101] reported that the data collected for fresh-frozen and FFPE colorectal tumour samples were compared, and 1465 and 1470 peaks were found in those two sample types, respectively. This observation indicated that the sample preparation did not seem to impact the number of detected peaks. Moreover, 1226 peaks overlapped between the fresh-frozen and FFPE samples leading to the conclusion that the detected metabolite content was similar. A similar comparison could be made using the data collected for this project as a part of future work.

In the case of both the single FFPE colorectal sample and the TMA, the samples' remaining metabolomic content was sufficient to allow for tissue classification. At least 80% accuracy was achieved in the leave-one-out cross-validation for the TMA data collected in the positive and negative modes. It has also been proven that the tissue sections cut at 5 µm instead of 10 µm were still compatible with DESI-MSI. Groseclose *et al.* [146] published results confirming that different lung cancer histologies could be distinguished and individual cancer types subclassified based on MALDI-MSI analysis of a single TMA. The core diameter was 1 mm, so it was smaller than the cores used in this project. However, in the colorectal TMA case, it was impossible to classify tumour samples according to cancer grade. It could be because the TMA used by Groseclose *et al.*contained more tissue needle core biopsies taken from a higher number of patients. The TMA used for this project had samples from only 18 patients, out of whom the majority had grade 2 colorectal cancer (n=14). Only one patient was diagnosed with grade 3, and three patients had grade 1. Aerni *et al.* [223] also used MALDI-MSI to analyse human kidney TMA. The study described a high-throughput workflow and reported that over 700 peptides were detected. Peptide profiles from fresh-frozen tissue sections were similar to those from FFPE samples. They also reported that the sections as thin as 5 µm were suitable for analysis. This is in line with the results obtained for the 5 µm thick colorectal TMA tissue sections described in this Chapter.

To summarize, this study has demonstrated a high degree of chemical metabolite conservation in FFPE tissue samples. The ability to perform MSI of metabolites is the starting point towards investigating metabolic pathways in FFPE tissues. This advance could play a significant role in the determination of the exact location of metabolic disturbances. Furthermore, the future use of TMAs can allow hundreds of molecules' measurements and, therefore, can help generate

pathways from thousands of patients. As FFPE samples have been archived for many years in hospitals worldwide, this presents an excellent source of information on multiple diseases and patients' outcomes to various treatments. This project and obtained results represent a significant technical advance and can play an essential role in diagnosis and predicting a patient outcome in the future

Chapter 6

DESI-MSI of cirrhotic liver tissue samples

6.1 Introduction

6.1.1 Liver cirrhosis and its aetiology

In recent years, liver cirrhosis has been reported as an increasing cause of mortality in more developed countries, being the fourth most common death cause in central Europe and 14th worldwide [224]. Cirrhosis is defined as the end-stage of chronic liver disease and results from different liver injury mechanisms leading to necroinflammation and fibrogenesis [224].

Most of the time, cirrhosis's aetiology can be identified by the patient's history combined with serologic tests and histologic evaluation. Hepatitis C and alcoholic liver disease (ALD) are the most common causes in the Western world, while Hepatitis B prevails in sub-Saharan Africa and most parts of Asia [225]. Other liver diseases that can lead to cirrhosis include Autoimmune hepatitis (AIH), Non-alcoholic steatohepatitis (NASH), Primary biliary cholangitis (PBC) and Primary sclerosing cholangitis (PSC) [225].

The transition from chronic liver disease to cirrhosis involves a few phases, including inflammation, activation of hepatic stellate cells with ensuing fibrogenesis, angiogenesis, and lastly, parenchymal extinction lesions caused by vascular occlusion [226] Majority of chronic liver diseases are asymptomatic and unsuspected until cirrhosis with clinical decompensation occurs [224].

6.1.2 Diagnosis of cirrhotic liver diseases and its challenges

Ultrasonography, magnetic resonance imaging (MRI), and computerized tomography (CT) are not sensitive enough to detect cirrhosis, and final diagnosis still relies on histology [225]. As a result, histological analysis of ultrasound-guided liver biopsy is still considered a gold standard in modern clinical practise [227, 228].

However, diagnosis can still be challenging when conditions can exhibit features of two different diseases. These conditions are commonly designated as overlap syndromes. They include descriptions of consecutive or simultaneous AIH and PBC, consecutive or simultaneous AIH and PSC, and very rarely, PBC and PSC [229].

6.1.3 An overlapping syndrome: AIH and PBC

This overlap syndrome is not a rare condition. 2-19% of patients with typical PBC hallmarks also show AIH features [230, 231]. The overlapping features include clinical symptoms, biochemical tests, immunological findings, and histologic features. However, it should be noted that it is still controversial whether these overlap syndromes are distinct entities or

variants of the major autoimmune liver diseases, and standardization of diagnostic criteria and terminology is lacking [232, 233].

6.1.4 An overlapping syndrome: AIH and PSC

AIH-PSC overlap syndrome most often occurs among young/ pediatric people [234]. PSC differs from AIH and PBC in that males represent 2/3 of patients in most populations [232]. Concurrent IBD is common in patients with this overlap syndrome [235]; however, its absence does not rule out the diagnosis [236]. Similarly, as with AIH and PBC, the overlapping features include clinical symptoms, biochemical tests, immunological findings, and histologic features [229, 237].

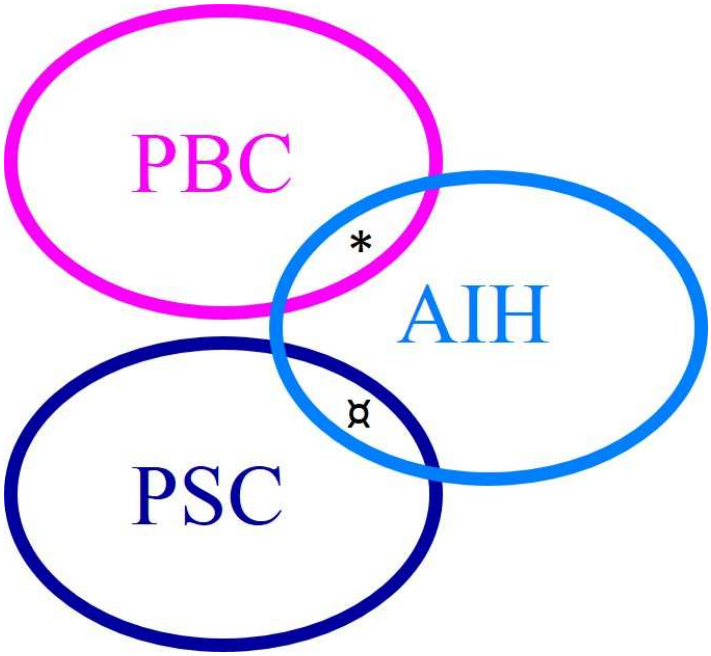


Figure 6.1. Relationship between the clinical expressions of AIH, PBC, and PSC. Depending on diagnostic criteria, 2-19% of patients with PBC * and 7-14% with PSC ¤ have been reported to have overlapping features with those of AIH [232].

6.1.5 An overlapping syndrome: PBC and PSC

The overlap between these two conditions has only been described in a few patients. Most of the time, these two conditions can be distinguished. However, liver histology may be similar [232, 238].

6.1.6 MSI and cirrhotic liver diseases

So far, no results for MSI projects looking into cirrhotic liver diseases have been published, except for a paper reporting MALDI-MSI application to study Non-alcoholic fatty liver disease (NAFLD) samples [239]. The authors concluded that the lipid composition of steatotic and non-steatotic tissue samples was highly distinct, implying that spatial context was essential for understanding lipid accumulation mechanisms in NAFLD.

Therefore, this project aimed to apply DESI-MSI to study metabolomic signatures of tissue samples representing various cirrhotic liver diseases. It was hypothesized that the molecular content differences might allow successful liver disease classification, primarily focusing on those with the so-called overlap syndrome.

6.2 Materials and methods

6.2.1 Clinical specimens and sample collection

Full ethical approval was obtained from the institutional review board at University Hospitals Birmingham NHS Foundation Trust (REC reference numbers: Immune Regulation: 06/Q2702/61, Cellular Trafficking: 06/Q2708/11). Further details can be found in *Chapter 2.1.2*.

Patient characteristics for this cohort were recorded and are summarised in Table 6.1.

Diagnosis	Number of samples per gender		Age range		Total number of samples
	Female	Male	Female	Male	
AIH	3	0	35-67	-	3
ALD	3	7	33-59	48-67	10
HBV	1	8	35	37-69	9
HCV	2	5	54-65	45-56	7
NASH	3	4	53-57	45-56	7
PBC	9	1	33-72	51	10
PSC	4	6	22-67	18-74	10
Resection	4	6	43-71	57-80	10

Table 6.1. Characteristics of the liver samples.

If a patient had liver cancer and if that cancer was under specific cm, then the patient could have a liver resection surgery (= part of the liver removed). In that resected liver, there was always a normal (healthy) area and the tumour area. The samples described as 'Resection' were taken from the normal area and were believed to be a healthy part of this organ.

6.2.2 Sample preparation

Fresh-frozen samples were prepared as described in *Chapter 2.7.1*.

6.2.3 DESI-MSI analysis of the liver tissue samples

Fresh-frozen human liver samples were analysed in a random order in negative ion mode using a high-resolution orbital trapping mass spectrometer (Exactive, ThermoScientific, GmbH) controlled by XCalibur 2.1 software. Full details of the DESI-MSI setup for this batch of samples can be found in *Chapter 2.8*.

Parameters	Settings of Thermo Exactive instrument
Polarity	Negative
Max injection time	1000 ms
Micro-scans	1
Mass resolution	100000
Mass range	150-2000 *m/z*
Capillary temperature	250°C
Capillary voltage	-50 V
Tube lens voltage	-150 V
Skimmer voltage	-40 V
Spray voltage	4.5 kV
Skimmer voltage	-40 V

Table 6.2. Thermo Exactive Orbitrap settings used for DESI-MSI of fresh-frozen liver samples.

6.2.4 Histological assessment

As DESI-MSI is a soft ionisation technique, the analysed tissue sections could be stained with H&E for histopathological evaluation. The stained slides were digitally scanned at high resolution, using a Nanozoomer 2.0-HT digital slide scanner (Hamamatsu, Japan) and analysed by an in-house histopathologist blinded to DESI-MSI results.

6.2.5 Data analysis

Methodological details of data analyses are described in *Chapter 2.13*.

6.3 Results

6.3.1 Histopathological presentation

From a histopathology point of view, the presentation of all cirrhotic liver H&E-stained tissue sections is very similar. Histologically, cirrhosis is characterised by diffuse nodular regeneration surrounded by dense fibrotic septa with subsequent extinction of parenchymal and liver structures collapse. Together these changes are causing pronounced distortion of hepatic vascular architecture [224]. Some H&E-stained tissue sections representing four liver diseases are shown in Figure 6.2, with examples of nodules marked in blue and darker staining fibrotic tissue filling spaces between them.

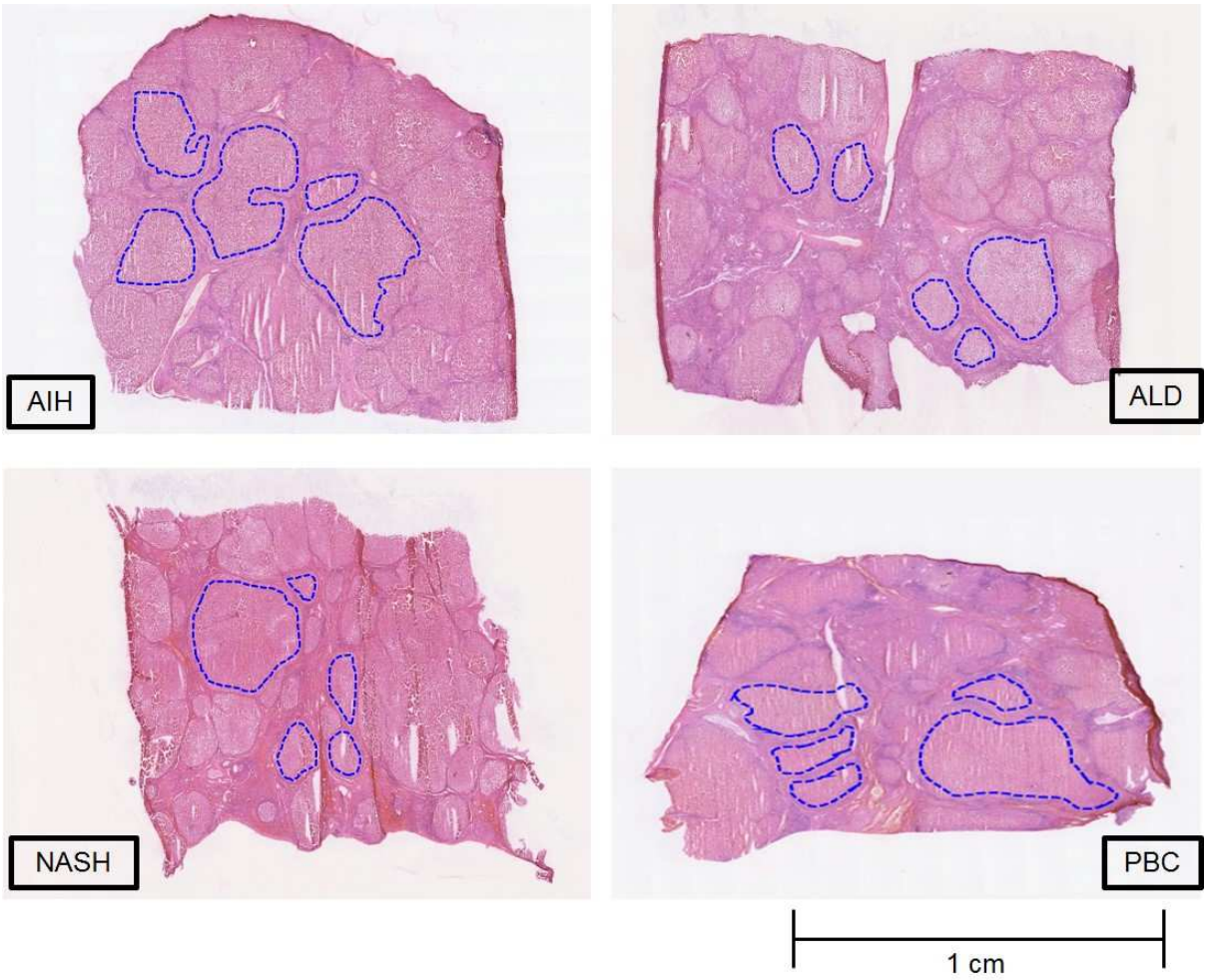


Figure 6.2. H&E-stained liver tissue sections representing various cirrhotic liver diseases. Nodules are marked in blue, and the darker pink/ purple staining tissue in between is the fibrotic tissue.

Distortion of hepatic vascular architecture leads to increased resistance to portal blood flow, causing portal hypertension and hepatic synthetic dysfunction. Fibrosis describes the replacement of injured tissue by a collagenous scar. Cirrhosis is an advanced stage of liver fibrosis and is accompanied by distortion of the hepatic vasculature [225]. Clinically, cirrhosis has been described as an end-stage disease, invariably leading to death unless liver transplantation is done. Screening for oesophageal varices and hepatocellular carcinoma have been regarded as the only preventive strategies [224].

All the samples have been annotated with the help of a qualified histopathologist. At this stage, annotations only focused on two tissue types: nodules and fibrotic tissue. As the latter was

thought to be an outcome of each disease, only nodules' metabolomic content was taken into account for classification purposes.

6.3.2 DESI-MSI data acquisition

All the liver samples received as a part of this project were around 12mm x 9mm x 7mm large. That resulted in individual DESI-MSI runs taking between 9 – 11 hours. For that reason, all the samples were only analysed in the negative ionization mode. All the runs were randomised to avoid a batch effect.

MSI generated intense phospholipids signals, and the statistical analysis focused on the changes in the phospholipid mass range to attempt tissue/ diseases differentiation. Figure 6.3 represents averaged spectra for randomly chosen sections representing all cirrhotic liver diseases included in this project. By comparing those spectra, it was possible to observe specific ion changes in their relative abundances from one liver disease to another. It was hypothesized that those variations were significant enough to allow diseases classification.

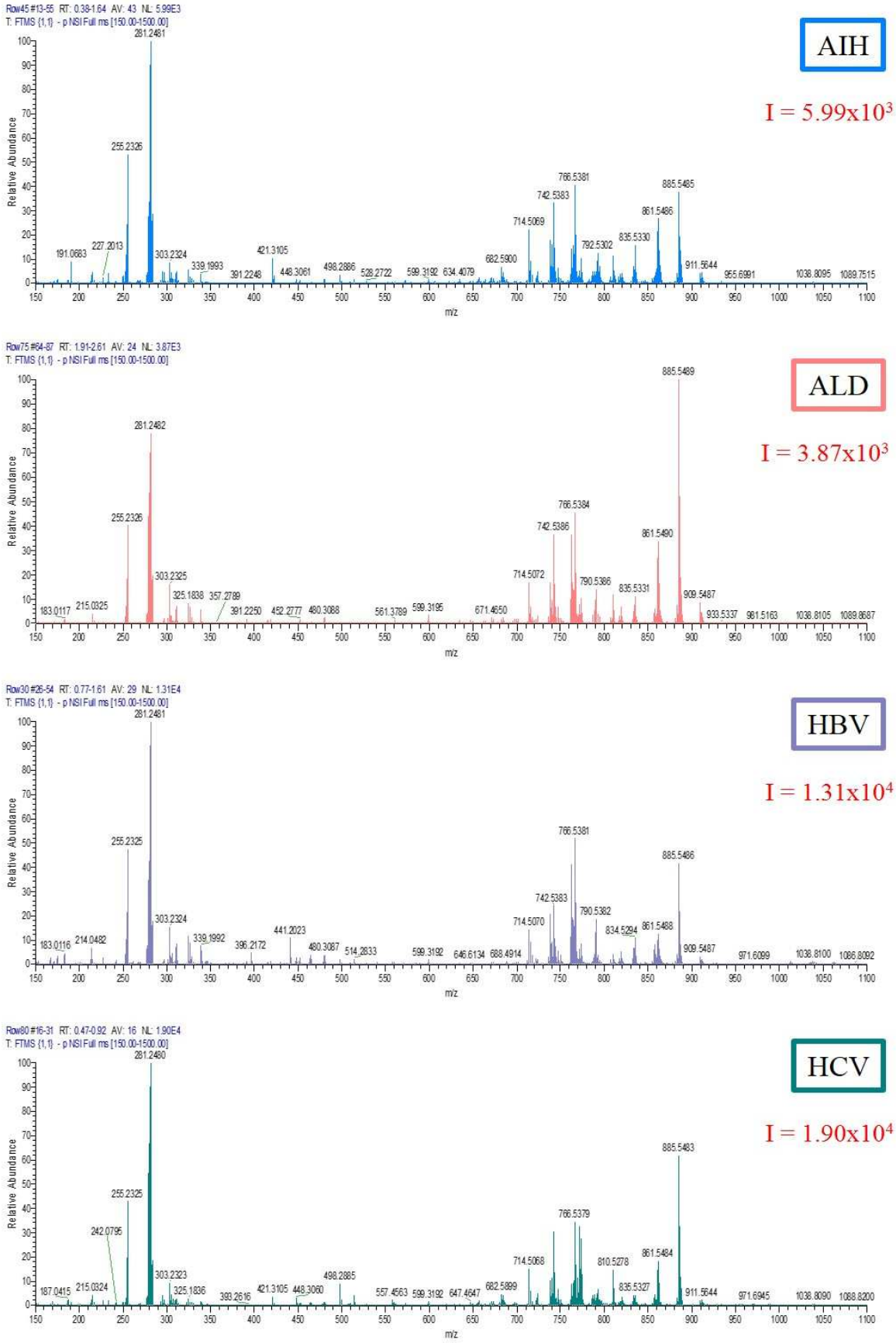


Figure 6.3. Averaged mass spectra representing different cirrhotic liver diseases. DESI-MSI data acquired in negative ionization mode.

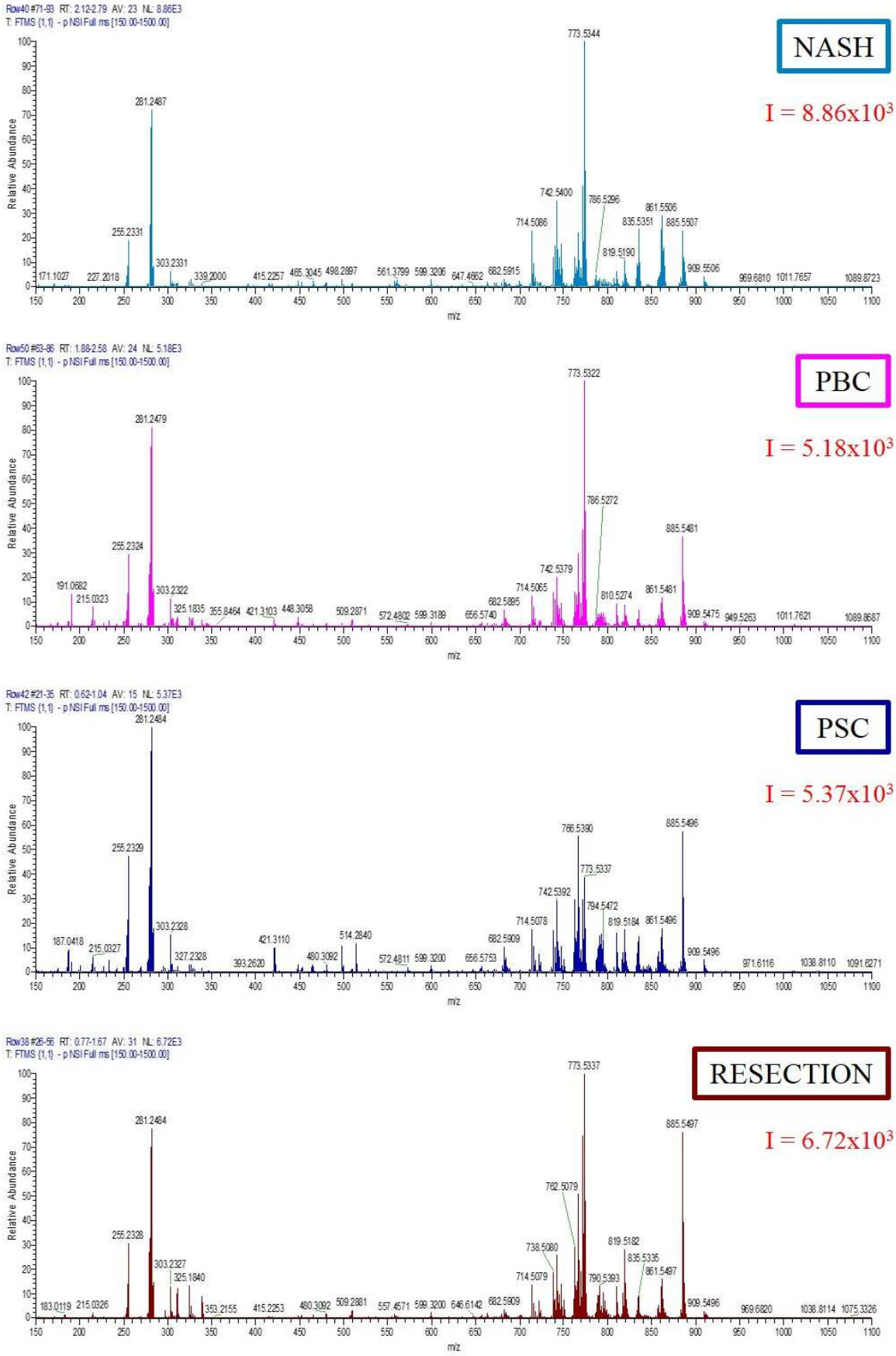


Figure 6.3. Averaged mass spectra representing different cirrhotic liver diseases - continued.

6.3.3 Prediction of tissue types in individual samples

The first step was to determine the ability of DESI-MSI to distinguish between different tissue classes in individual samples based on the lipidomic profiles. Multivariate statistical analysis was performed in each sample.

First of all, tissue types in an AIH sample and their spatial distribution were determined by an independent histopathologist. Nodules and fibrotic tissue were the two main types of tissue identified based on the corresponding histological image. Based on this information, a small number of representative pixels per tissue type was selected from the integrated MS ion image to build a sample-specific RMMC model used to classify all the MS ion image pixels. The different tissue types were classified with the PCA approach, as shown in the PCA plot (Figure 6.4). This data was complemented by a leave-one-out cross-validation test that confirmed 100% accuracy in classifying all the pixels from the MS ion image (Figure 6.4).

All other samples were subjected to the same data analysis, and results for PBC, ALD, and NASH samples are shown in Figures 6.4 and 6.5.

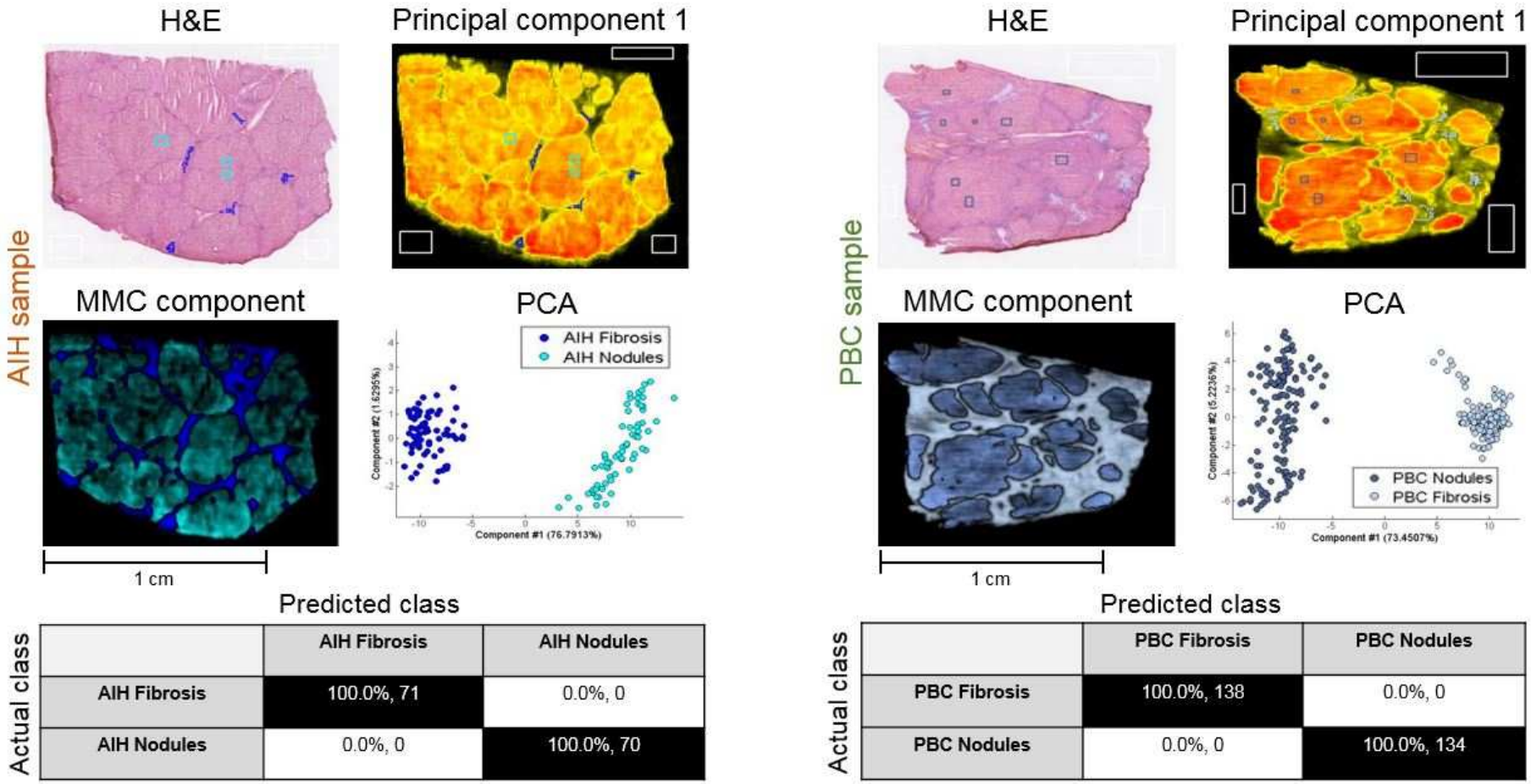


Figure 6.4. Results for individual AIH and PBC samples. Post-DESI-MSI, the same tissue section was H&E stained and annotated. PC1 revealed the histological structure of the analysed sections. An unsupervised principal analysis plot showed a clear separation between fibrotic tissue and nodules in the case of both AIH and PBC samples. MMC component images and confusion matrices showing classification performances are also presented.

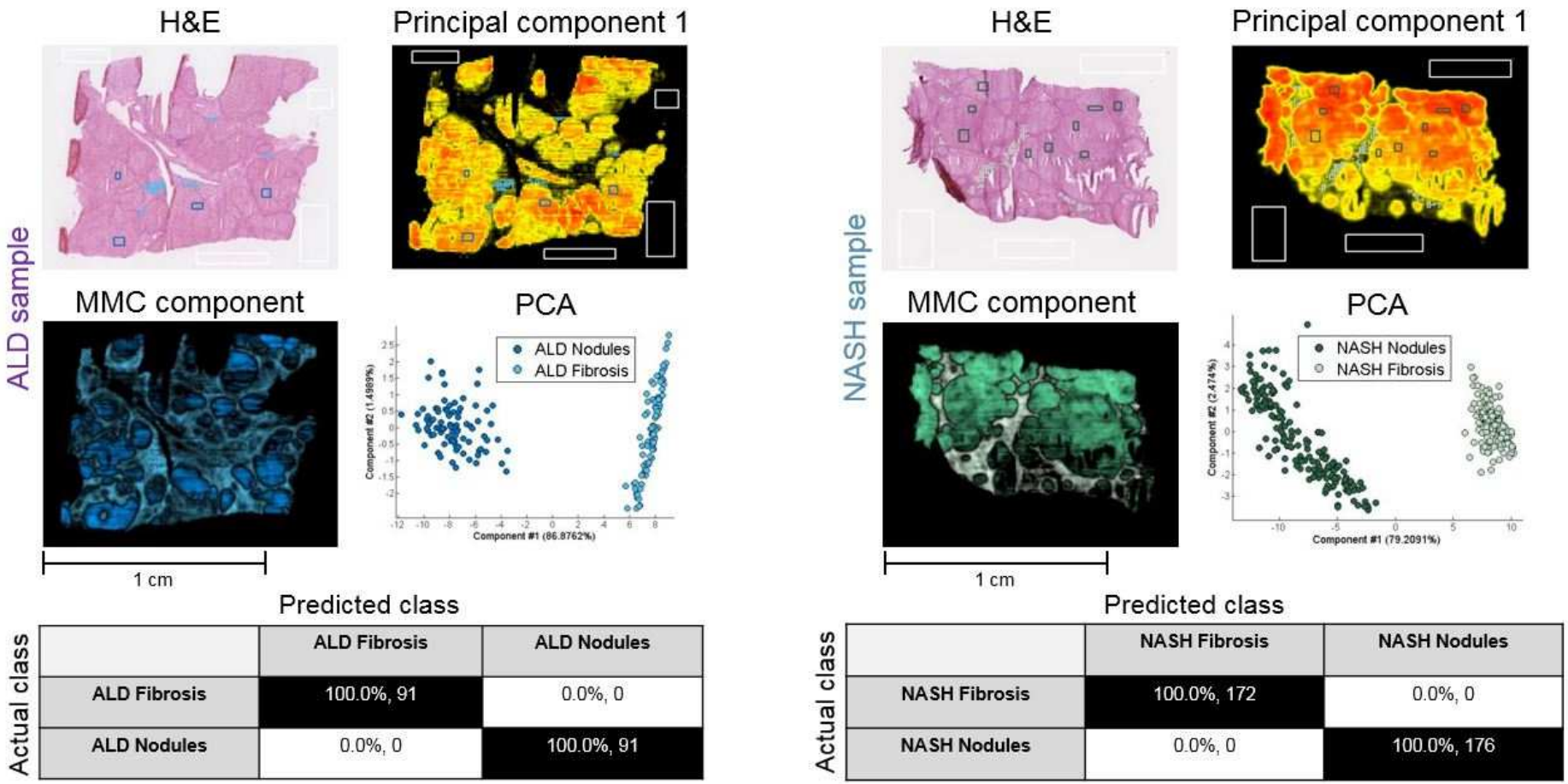


Figure 6.5. Results for individual ALD and NASH liver samples. Again, a 100% accuracy was achieved when tissue classification was attempted.

The same workflow was applied to all the remaining cirrhotic liver samples included in this study.

6.3.4 Prediction of tissue types in multiple cirrhotic liver diseases samples

Further data analysis was performed in a MATLAB environment using the imaging toolbox to compare multiple cirrhotic liver tissue samples; resections were excluded.

The unsupervised PCA was performed, and a clear separation was achieved between fibrotic tissue and the nodules. This was confirmed by a supervised analysis and the results of leave-one-out cross-validation, where an accuracy of 100% was achieved for both tissue classes (Figure 6.6).

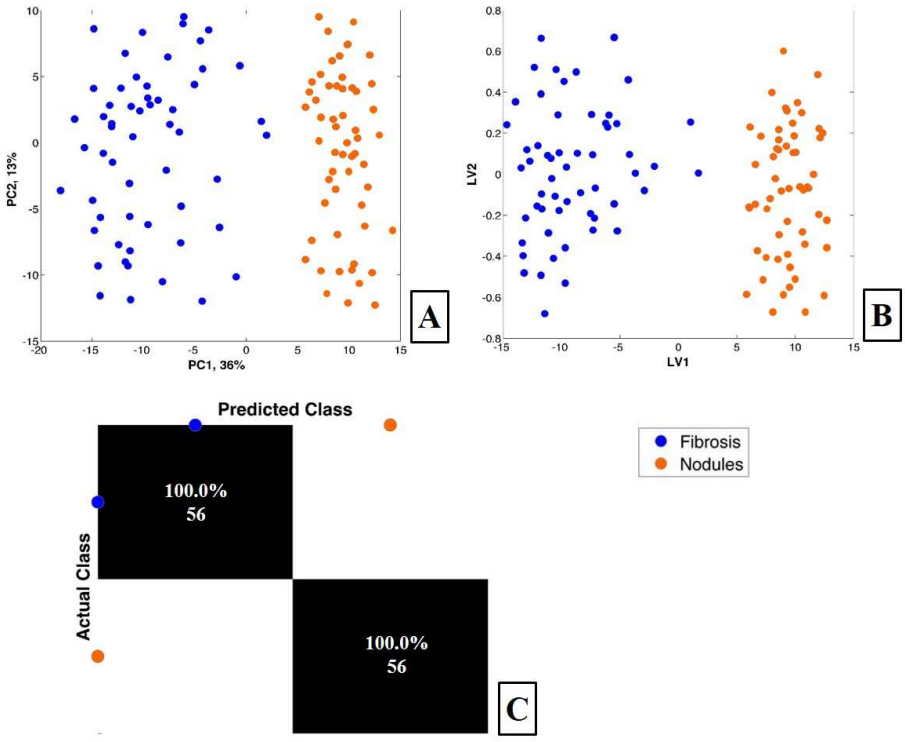


Figure 6.6. Results for all 56 diseased samples – comparison of the chemical content of nodules and fibrotic tissue. PCA plot showed good separation between these two tissue classes (A), which was then confirmed by the maximum margin criteria analysis (B). Accuracy of 100% was achieved for both nodules and fibrotic tissue in a leave-one-out cross-validation (C).

Some molecules which contributed to the discrimination of the nodules and fibrotic tissue were putatively identified using Metlin https://metlin.scripps.edu/metabo_search_alt2.php (Figure 6.7). PE(P-38:4) was found to be more abundant in fibrosis, and three PIs molecular species –

PI(36:2), PI(38:3), and PI(40:6) were more abundant in nodules. The difference between these two tissue classes appeared to be more evident in the case of the PIs.

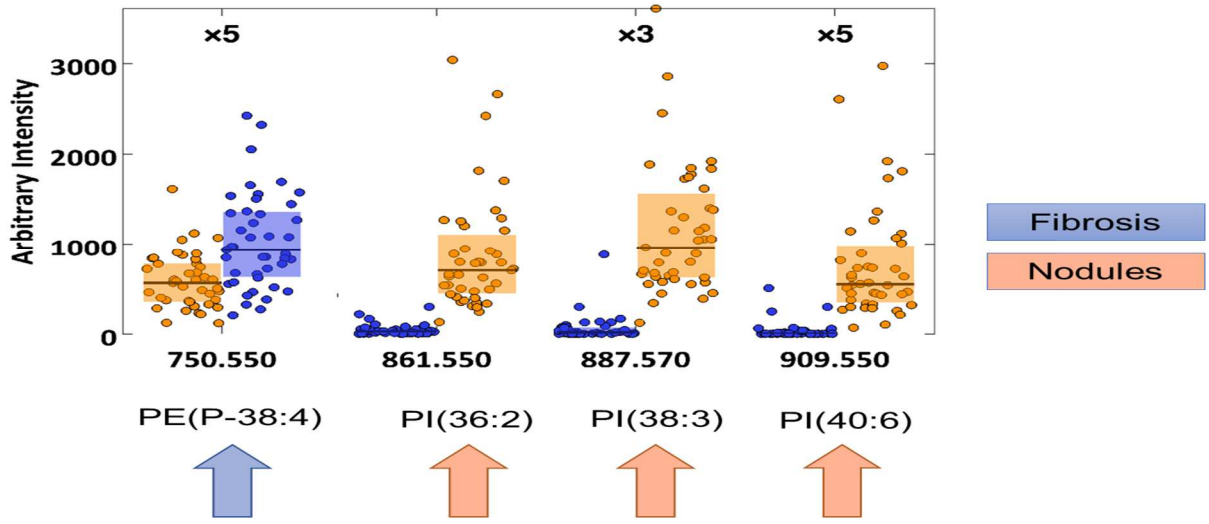


Figure 6.7. Molecules contributing to the separation of nodules and fibrotic tissue. PE(P-38:4) was found to be more abundant in fibrosis, while three different PIs molecular species were more abundant in nodules. Some intensities have been zoomed in order to put them on the same scale.

6.3.5 Liver disease prediction based on DESI-MSI data

The same approach was used to analyse the differences in the nodules' chemical composition from all the cirrhotic liver diseases samples. However, for this particular analysis, each data point represented ten averaged spectra for visualisation purposes. Figure 6.8 shows some trends in the grouping of the pixels representing the same disease. However, it also appeared that there was a slight effect for a couple of PBC/PSC samples, most likely due to the low number of samples per disease. The PCA plot suggested that whilst the difference between fibrotic tissue and the nodules was large, there was no significant difference between nodules of the various cirrhotic liver diseases. Leave-one-out cross-validation results did not show any trend as all, with the majority of samples being misclassified.

However, since various testing options are available (e.g. biochemical tests, immunological findings, histopathological features), the real challenge lies in classifying liver diseases with the so-called overlapping syndrome.

It was then decided that the acquired data should not be used to separate an individual disease from the pool of all the conditions mixed all together. Instead, an attempt was made to try and distinguish AIH and PBC and AIH and PSC.

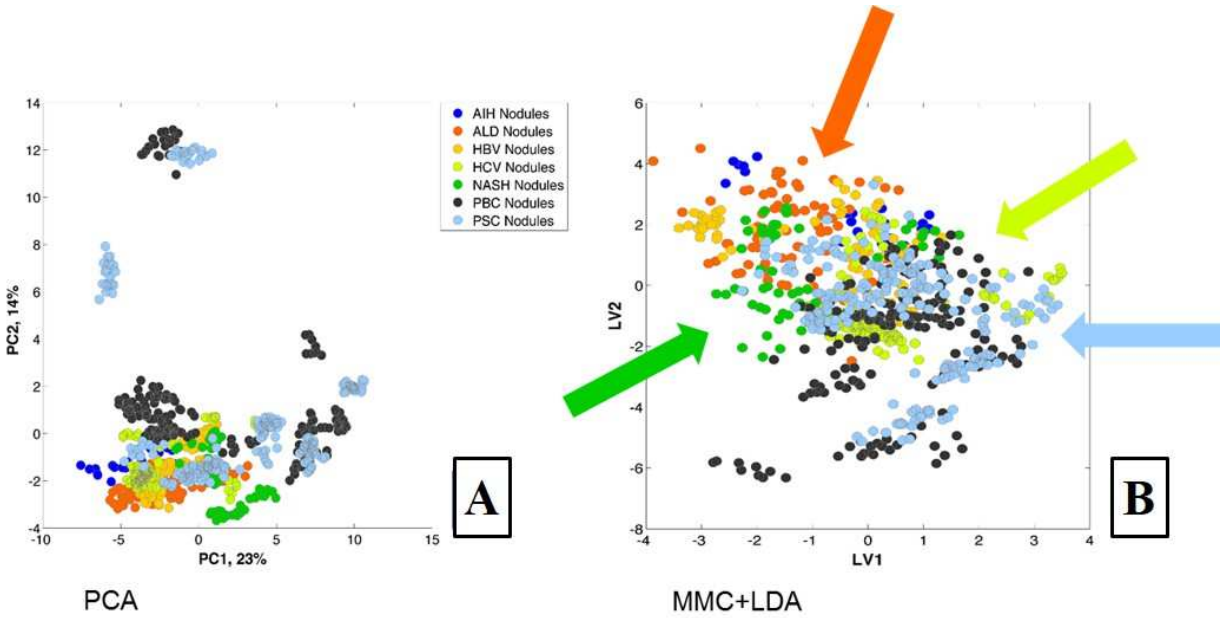


Figure 6.8. An attempt of cirrhotic liver disease classification based on metabolomic content analysed by DESI-MSI. Some trends in the grouping of the pixels from the same cirrhotic diseases could be seen in the PCA plot; however, it was evident that the results were suffering from the low number of samples included (A). The supervised analysis also revealed some trends, which are highlighted by the arrows (B).

One data point per sample was used, meaning that all the acquired data per sample were averaged and represented as a single point.

Based on the collected metabolomics data, the separation was possible between neither AIH and PBC nor AIH and PSC. The statistical analysis not only suffered from a low number of samples, especially in the case of AIH but also from some batch effect. As a result, no trend was seen in the unsupervised PCA or even supervised MMC+LDA analysis.

Next, an attempt was made to use all annotated pixels per sample, and obtained results were promising. This was discussed with a bioinformatician who highlighted the fact that that was only because the model was overfitted. Overfitting a model is described as a condition where a statistical model describes the random error in the data rather than the relationships between observed variables. This problem often occurs when the model is too complicated or complex. An overfitted model will fail to replicate in future samples, creating considerable uncertainty about the finding's scientific merit [240]. Some of these results are presented in the

Supplementary Information (Appendix K). However, those results must not be treated as a reliable statistical model.

Therefore, it has been suggested that another approach to look into this set of data should be tested.

6.3.6 Colocalization features as a new statistical approach

All of the results presented from this point onwards were prepared with extensive help from Dr Paolo Inglese.

The significant advantage of this alternative approach described by Inglese *et al.* is the fact that it allows the classification of the analysed sample using the acquired MSI data and at the same time preserves the spatial information which is associated with the mass spectra. This method relies on ion colocalization features and is less sensitive to spectral intensities' variations induced by the batch effect [218].

6.3.7 Data pre-processing

The same workflow, as described by Inglese *et al.* [218], was applied to the whole dataset.

All the raw data were converted into the imzML (centroided mode) format before their *m/z* values were corrected using a single point recalibration (palmitic acid – 255.2330 *m/z*, $[M-H]^-$). A search window of 10ppm was used, which was compatible with the theoretical mass spectrometer instrument error.

The recalibrated samples were pre-processed independently through the following workflow:

- Total ion count (TIC) scaling intensity normalization
- Peak matching
- Log transformation

Next, Sputnik, an R package, was used to removing noise peaks.

Sputnik provides a series of filters that are used to select informative and meaningful peaks based on the probability of their spatial distributions. Moreover, it provides a correlation-based filter, an estimation of split peaks, and a series of tests based on complete spatial randomness [241]. Its application to both DESI-MSI and MALDI-MSI datasets has already been successfully tested and described by Inglese *at el.* [241].

The PCA was run within each tissue section before and after noise-related peak removal. The results are presented in Figures 6.9 and 6.10.

Significant variability was discovered within each tissue section, and obviously, it was not related to histological components present in the analysed samples (Figure 6.9). Therefore, it was apparent that this set of data had to be subjected to batch effect correction before any further statistical analysis.

There were a couple of reasons why an apparent variability was seen between the samples and within them. Due to the size of tissue sections, the whole process of data acquisition took months. In the meantime, a new sprayer had to be built and optimised. During this time, the instrument was shared between different users who also analysed various tissue samples types. Every care was taken to ensure the instrument was in good shape before commencing with the DESI-MSI measurements. However, some signal fluctuations were inevitable, especially during such a long time needed to measure the whole batch of these samples.

6.3.8 Colocalization features extraction

The MSI images were composed of tissue-related and tissue-unrelated pixels. Therefore, to filter and only use tissue-related pixels to calculate the colocalization features, the *'kmeans2'* method from the Sputnik package was applied to determine the region of interest (ROI) representing the part of the image which was occupied by tissue. k-means was applied to entire MSI images allowing the extraction of four clusters, which were then used to allow a more refined MS image segmentation into tissue-related and unrelated areas [218].

Only tissue-related ROI pixels were used to calculate Spearman's rank correlations between each pair of matched peaks within each MS image.

The Spearman's rank correlations were calculated on a fixed number of randomly selected pixels belonging to the tissue-related ROI to accommodate the fact that the size of the tissue-related ROI could vary between MS images [218].

None of the spatially related intensity transformations was applied to the ion images before calculating the correlations [218].

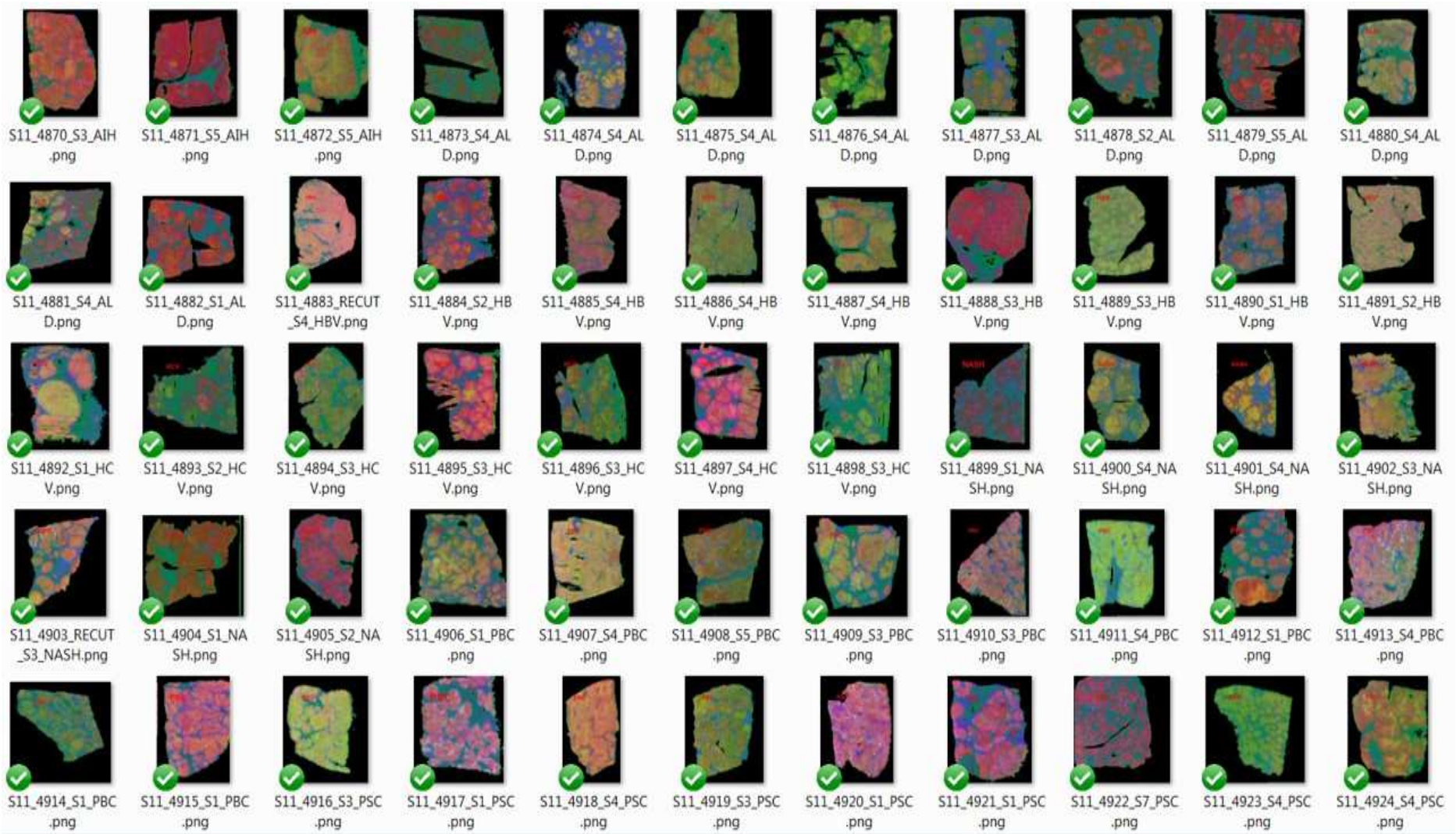


Figure 6.9. PCA representation of the analysed liver tissue sections before the batch effect correction. Various colours present within one tissue section represent a high variability, which does not reflect the sample's histopathological composition.

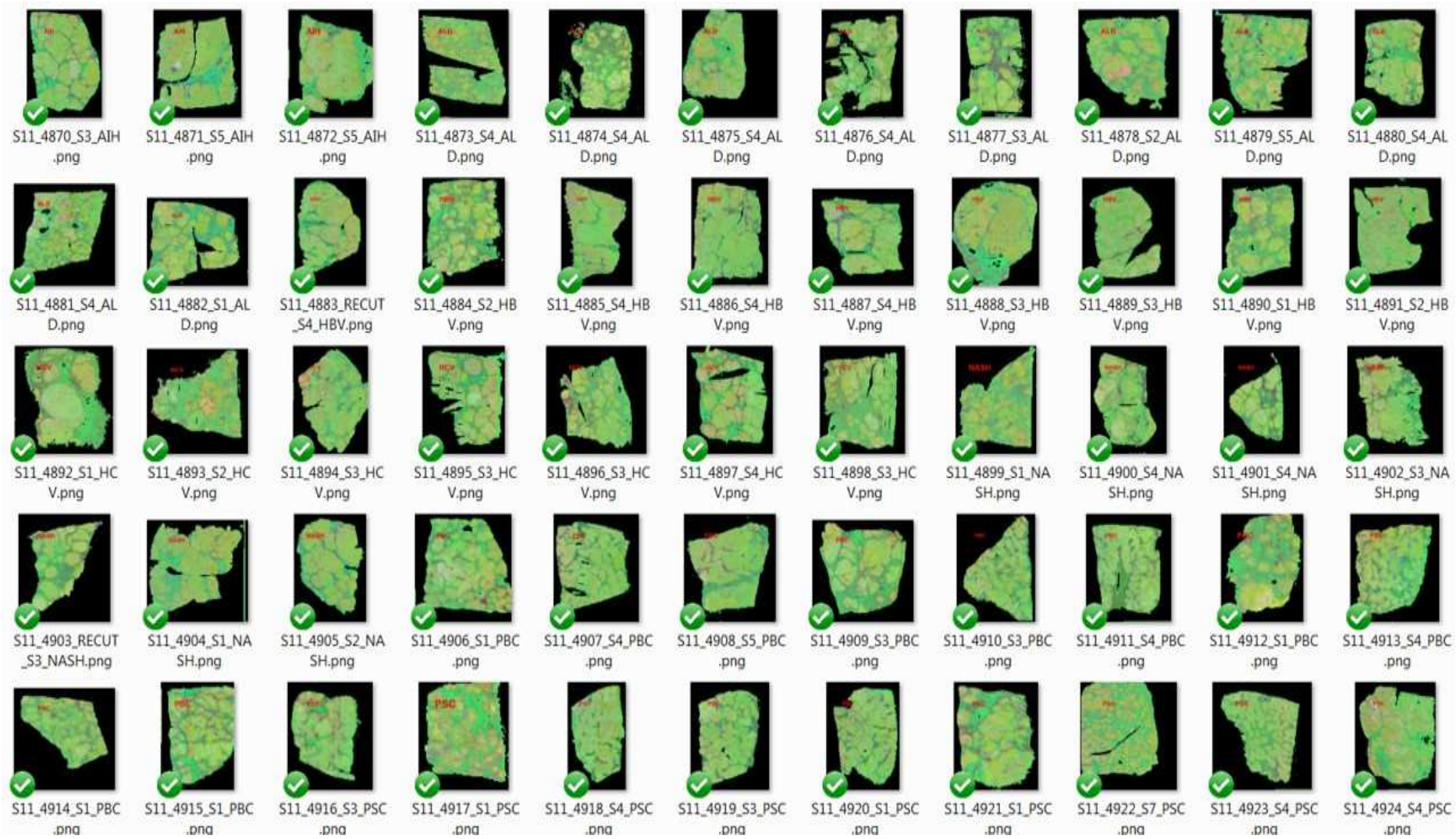


Figure 6.10. PCA of the same tissue section after removal of the batch effect. A significant improvement in the data quality was achieved; the acquired data related to and represented the samples' metabolomic content and histological details.

Instead of trying to classify individual cirrhotic liver diseases, they were grouped into three classes based on similar aetiology:

- metabolic (ALD and NASH)
- cholangiopathy (PBC and PSC)
- hepatitis (AIH, HBV and HCV).

6.3.9 Colocalization features and their visualization

It has been hypothesized that different diseases were represented by different pathways locally. Therefore, those different pathways could be expressed in the acquired mass spectrometry data because some ions were co-localised in specific regions. They were potentially part of the same biochemical mechanism. In a different disease, the same ions were not co-localised as the exact mechanism was not occurring.

A multiple pairwise univariate Kruskal–Wallis test was used to determine the colocalization features that were significantly different between the three classes of liver diseases (metabolic, cholangiopathy and hepatitis) [218].

That allowed us to identify the essential correlations and see which correlations were more important to discriminate between three classes of liver diseases. The univariate plot of the values of the correlations is shown in Figure 6.11.

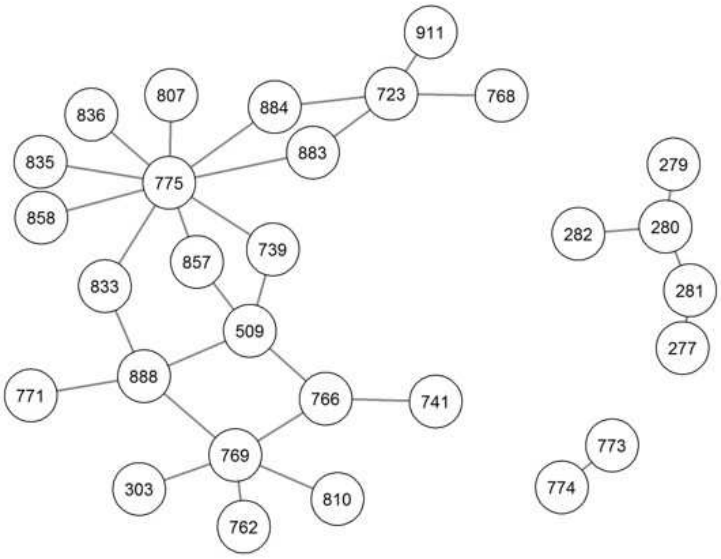


Figure 6.11. Correlations pairs found among the top-100 in 80% (=30) of the leave-one-out cross-validation rounds.

Next, it was identified which pairs of ions were assigned to either cholangiopathy, metabolic or hepatitis group. The results are presented in Figure 6.12. Each ion pair is assigned to either cholangiopathy, metabolic, or hepatitis group.

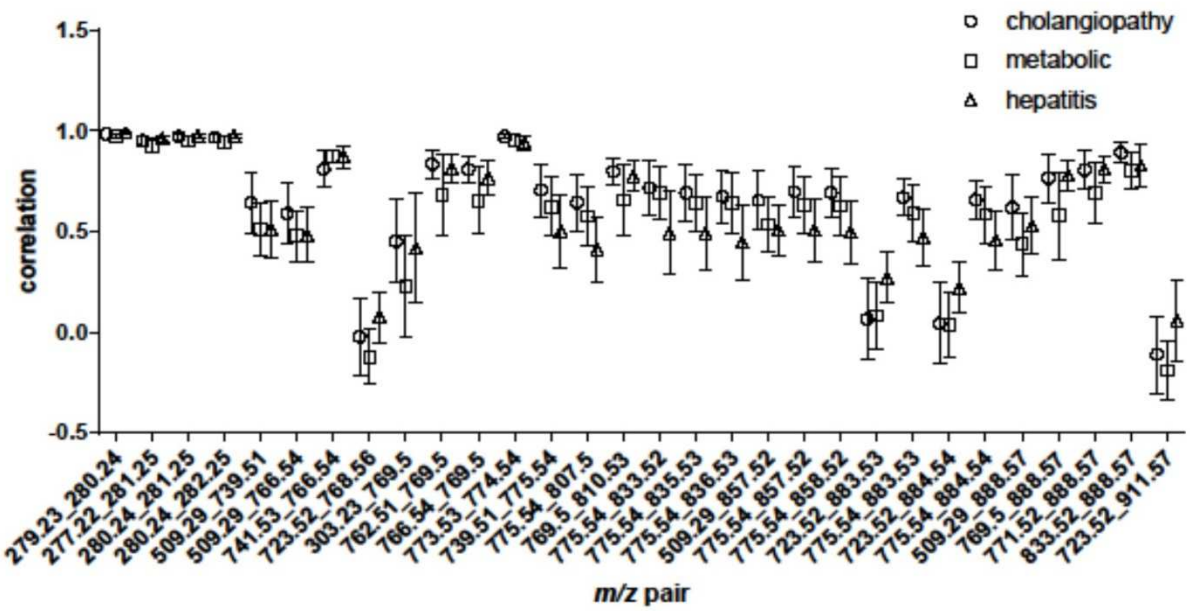


Figure 6.12. Values of correlations found among the top-100 in 80% (=30) of the leave-one-out cross-validation rounds.

The data presented in Figure 6.11 and Figure 6.13 were used to identify the top-10 correlations, and the *mz* values are presented in Supplementary Information, Appendix L.

To assign IDs to those ions, MS/MS was performed (Details described in *Chapter 2.13*), and the results are shown in Table 6.3.

Mz	Annotation	Accurate mass	ppm Error
277.2173	FA(18:3) [M-H]-	277.2173	0
280.2363	C13 isotope of FA(18:2) [M-H]-	280.2358	1.784211725
281.2486	FA(18:1) [M-H]-	281.2486	0
282.2519	C13 isotope of FA(18:1) [M-H]-	282.25142	1.700611462
723.5163	PA(18:0_20:4) [M-H]-	723.497	26.67599175
741.5268	C13 isotope of PE(18:1/18:2) [M-H]-	741.52642	0.512456454
742.5391	PE(18:2_18:0) [M-H]- and PE(18:1/18:1)	742.5392	-0.134673025
766.5389	PE(18:0_20:4) [M-H]-	766.5392	-0.391369417
769.5023	PG(18:2/18:2) [M-H]-	769.5025	-0.259908188
775.5387	PG(18:0_18:1) [M-H]-	775.5495	-13.92561016
807.502	PI(16:1_16:0) [M-H]-	807.5029	-1.114547081
810.5289	PS(18:0_20:4) [M-H]-	810.5291	-0.246752399
883.5338	PI(18:1_20:4) [M-H]-	883.5342	-0.45272724
888.5696	PS(22:3_22:4) [M-H]-	888.576	-7.202535292

Table 6.3. Annotations the ions forming the top-10 correlations.

The spatial localization of the ions listed in Figure 6.11 could be investigated by plotting the relative abundance images of the particular spectral peaks. These results are presented in Figure 6.13, and the obtained images confirmed that the selected ions were localized only in tissue-related areas. One randomly chosen sample was used to represent each group of liver diseases. Ion correlations were assigned various colours, and then their spatial distribution was visualised within the sample representing metabolic, hepatitis, and cholangiopathy group.

In the case of every ion group, the highest intensities were observed within the nodules.

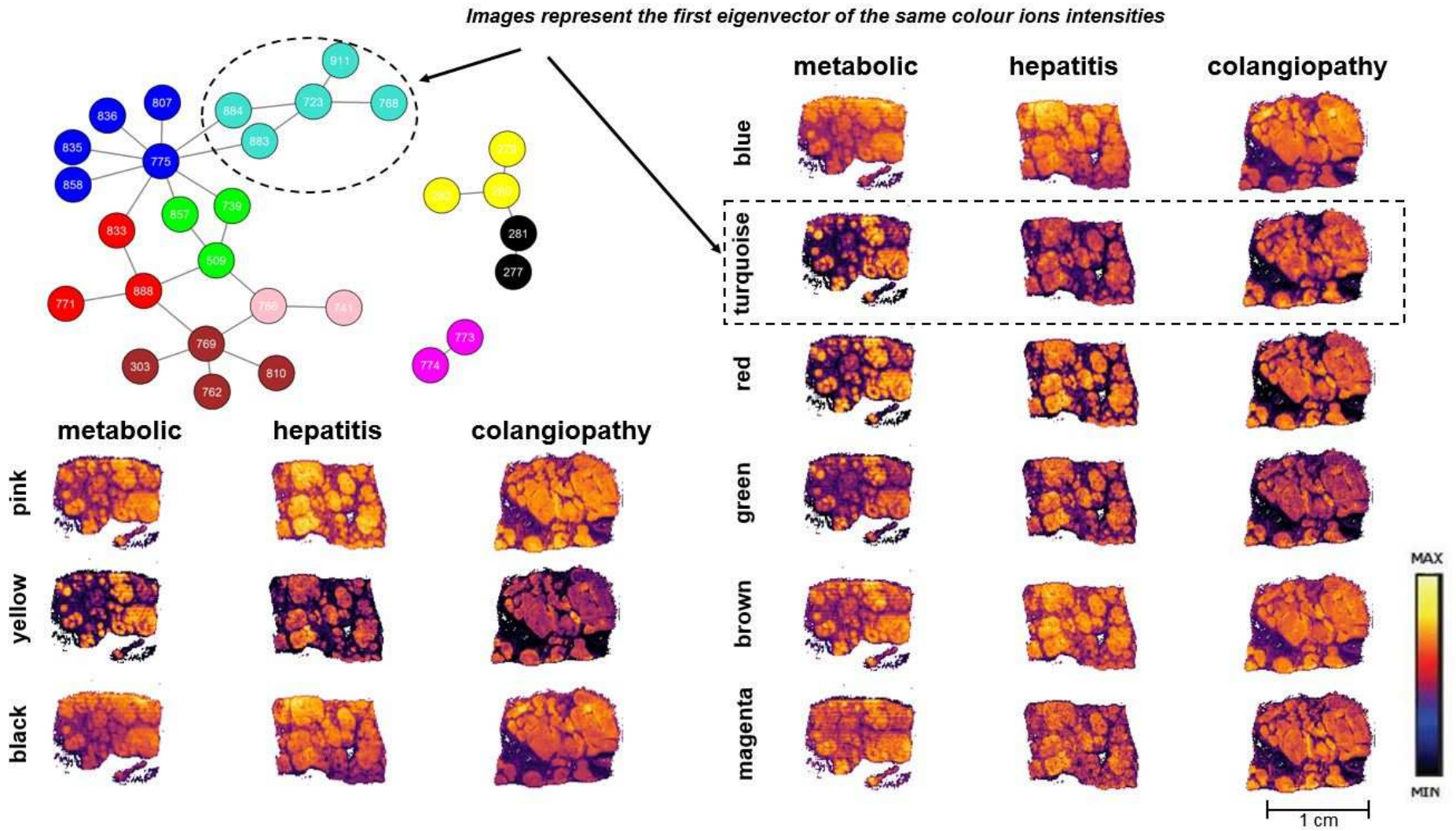


Figure 6.13. The spatial localization of ions forming the ten most significant correlations. Groups of ions were colour-coded, and their spatial distribution within one sample from each group was visualized. In the case of each ion correlation, they were mainly present in nodules of the analysed samples.

6.4 Discussion

Since its introduction in 2004 [78], DESI-MSI was applied multiple times to successfully analyse tissue samples. It was reported that this technique had been used to discriminate tumour grades and subtypes, as well as to identify tumour margins in various types of human cancers [83, 85, 90, 92]. However, DESI-MSI has never been applied to analyse cirrhotic liver diseases. This project aimed to employ this technique to look into the metabolomic content of various liver diseases – AIH, ALD, HBV, HCV, NASH, PBC, and PSC – to assess if this additional information could be helpful in the process of making a final diagnosis. The diseases with the so-called overlapping syndrome – AIH/PBC and AIH/PSC – were particularly interesting as the overlapping features, including clinical symptoms, biochemical tests, immunological findings, and histologic features, make diagnosis challenging.

First, using pixel-wise analysis of the DESI data confirmed that the successful discrimination of two types of tissue classes – nodules and fibrosis – was possible, both in individual samples and the batch of all analysed specimens. However, this information had no prognostic value as fibrotic tissue can be described as the disease's outcome. Therefore, it is the molecular content of nodules that remains interesting and can reveal some useful metabolomic information.

Next, an attempt was made to build a model for AIH/PBC and AIH/PSC. At this point, it became apparent that the chemical content of nodules representing these diseases was not distinctive enough to allow classification. A deficient number of AIH samples (n=3) was a significant problem, but the models were also suffering from batch effects. All the statistical analysis performed so far was done in a Matlab environment using the home-built toolbox. Each sample was represented by one averaged (median) data point.

Unsupervised and supervised statistical modelling approaches are commonly used to identify relationships between ion signatures and properties of the sample of interest. However, the large number of collected mass spectra per sample makes it challenging to invent a model that can capture a single property preserving the sample's molecular information [218]. Therefore, it was suggested that the two most important properties of MS imaging – their spatial distributions and molecular heterogeneity – were not adequately preserved while using one averaged data point per sample.

It was hypothesized that the classification of individual diseases would be too challenging, at least at that stage of the project. Based on the disease aetiology, samples were assigned to either metabolomic, hepatitis or cholangiopathy group. An alternative approach described by Inglese *et al.* [218] was then implemented. It relied on the fact that the observed spatial patterns

distributions of the ion peak intensities reflected the analysed samples' molecular properties. The raw data were pre-processed again as a part of this new workflow. An unsupervised analysis was run for each sample and revealed a significant variability within each analysed tissue section. The Sputnik package for *R* was used to filter tissue-unrelated peaks, and it allowed an enormous improvement in the data quality. Ion correlations were identified and visualised in randomly chosen analysed tissue sections.

In the meantime, a second batch of the liver sample has been received and analysed. Therefore, this project is ongoing, and the final version of the results is yet to be presented. Detailed histological annotations will also be required to achieve a meaningful correlation between the MS imaging data and the histopathological presentation of the analysed samples.

However, the most important achievement so far is that two independent statistical workflows have been tested. The obtained results were carefully discussed, and the future approach has already been agreed on.

Chapter 7

Overall discussion and future directions

7.1 Overall discussion

For over 100 years, the histopathological analysis of cytology, biopsy or resection specimens has been the final step in the process of diagnosing multiple diseases, including cancer [242]. The final pathology report plays a crucial role in guiding clinicians in finding the optimal treatment and builds the framework for selecting new markers and new therapies for a patient in the future. In recent years, standard clinical care is continuously becoming more complex, and as a result, diagnostic pathology workup is also more complex and extensive [243]. For example, in the 1960s, the guidelines for preparing lung cancer specimens recommended only H&E and periodic acid-Schiff stain. In contrast, techniques like IHC and molecular analyses are essential these days and play a vital role in acquiring additional essential information [243]. Moreover, the increasing complexity in pathology workup resulted in a significant increase – by more than 60% over the last decade – in the number of tissue blocks per patient and the number of required spare slides per block [56]. However, this enormous increase in workup is not the only issue that pathology services are facing.

Despite being considered a gold standard in making a diagnosis, histopathological investigations can be time-consuming. Depending on the required stains, it can take days before the final pathology report is issued. Additionally, an examination of the stained slides is subjected to intra-observer error [35-37]. Therefore, it is evident that some additional techniques are required to complement making a diagnosis.

For several decades, MS has been the focus of technology development and imaging application. DESI-MSI was developed in 2004 [78] and since then has been used in multiple projects aiming to characterise various types of cancer [83, 90, 92, 244, 245]. As DESI-MSI is a non-destructive technique, histological features remain intact during the analysis of tissue sections. Therefore, the distribution of multiple analytes can be correlated with clinical and histological features [65]. Moreover, the analysis is performed at ambient conditions, sample preparation is straightforward, and the data collection is user-independent [78, 79]. However, there are also some limitations, especially in the sample preparation step. Since its introduction, DESI-MSI has always relied on frozen sections [79], which had to be prepared without OCT [128, 129]. At the same time, OCT is routinely used as an embedding medium in the histology diagnostic laboratories when frozen samples are cryosectioned.

Therefore, this project aimed to study how to make DESI-MSI more compatible with the standard histopathological workflow. Making this technique as histology friendly as possible is crucial if it is meant to be implemented into the routine pathology portfolio.

Chapter 3 describes the impact of OCT embedding on the DESI-MSI data. Integration of sample preparation workflows to meet different techniques' requirements is crucial if we aim to make those techniques compatible with each other. Since all work involving the preparation of frozen samples in the diagnostic laboratories relies on OCT, the practicalities of using this embedding medium in DESI-MSI measurements have been tested.

OCT has been documented to cause strong ion suppression when used to embed specimens analysed via MALDI imaging [128, 129] and also causes the detection of series of polymer peaks at $m/z > 1100$ [129]. Therefore, five pieces have been cut from a resected liver sample, cryosectioned embedded in distilled H_2O (fully compatible with DESI-MSI), and then cryosectioned again in OCT. Acquired data allowed to focus purely on assessing the impact of the later embedding medium on the DESI-MSI.

The difference in the H_2O-embedded and OCT-embedded spectra was clear for positive mode data, with extensive polymeric profiles dominating at $m/z > 1000$ and to a less noticeable extent around the phospholipid region ($600 < m/z < 1000$). There were no visible polymeric distributions in the negative mode spectra. Following the polymer removal from the positive mode data, supervised classification was performed to determine the similarity between OCT- and H_2O-embedded spectra from the same liver section. While a good classification was achieved for the positive mode data, the negative mode samples demonstrated poor results. Those results highlighted the value of removing the polymeric interference in samples analysed in the positive mode. They demonstrated that despite no evident polymeric distributions seen in the negative mode data, most likely subtle OCT effect was still present and impacted the classification.

Obtained results suggested that it was possible to employ OCT as an embedding medium for histopathological and mass spectrometric analyses. There were no clear polymeric signals causing differences in the negative mode data, but some reduction in intensities might be attributable to polymer-induced ion suppression. In positive mode data, the interferences due to OCT were more overt but could be negated by removing the regular peaks of the various polymeric distributions. However, studies involving various tissue types collected from multiple objects are needed to further assess the effects of potential ion suppression.

Chapter 4 focuses on optimising the pre-DESI-MSI sample preparation protocol to allow the analysis of FFPE tissue samples. Over the past three decades, many efforts have been put into reversing formalin-fixation, thus allowing the use of FFPE blocks in research [146]. Some projects focused on protein identification [145, 147, 148, 150, 199], while others reported

successful metabolite imaging [101, 102]. Despite testing various pre-experimental conditions, only the protocol published by Buck *et al.* [101] allowed to obtain metabolic information for FFPE tissue sections by DESI-MSI. The spectral composition for FFPE samples was very different from those acquired for fresh-frozen specimens. It was observed that peak intensities were significantly reduced above 600 *m/z*. However, it was proven that the metabolic content remaining in the analysed sections was sufficient to allow for the successful classification of different tissue classes. These observations were in agreement with the results published by Buck *et al.* [101].

Moreover, as a part of ongoing efforts to make DESI-MSI as histology-friendly as possible, the same experiments were performed using 5µm thick FFPE tissue sections (10µm thick sections are not routinely cut in diagnostic laboratories). Again, good classification results were achieved, confirming that even 5µm thick sections were suitable to DESI-MSI.

The results obtained in *Chapter 4* were further tested and applied to FFPE colorectal samples analysed as a part of another project described in *Chapter 5*.

Results presented in the first part of *Chapter 5* are for the fresh-frozen samples. Recent MS imaging projects reported that specific lipids show excellent histological specificity. Therefore, those lipids can be used to recognise various tissue classes unambiguously [154-156]. This observation has been widely confirmed by DESI-MSI, particularly for cancer-related applications [86, 246]. Although this project did not aim to identify lipids IDs, including potential biomarkers, it still relied on the phospholipids mass range for tissue classification. The investigated colorectal samples featured many histologically different tissue types, including mucosa, submucosa, muscle, tumour, and adenoma. Those various tissue classes were found to yield different DESI spectra.

As a result, tissue-specific molecular ion signatures allowed to achieve an accuracy of 96% for mucosa and over 75% for muscle and submucosa in a leave-one-out cross-validation for a cohort of 27 normal samples. Similarly, 88% and 91% accuracy was achieved for normal and tumour samples, respectively, when a batch of 38 samples was subjected to statistical analysis. These results agree with Veselkov *et al.* [114], who reported a highly accurate localized biochemical profiling of various tissue types. The workflow described by him allowed tissue-specific molecular ion pattern extraction leading to an accurate chemical reconstruction of histopathologically different regions within analysed colorectal samples.

The weakest reproducibility of spectral patterns was seen in adenocarcinoma samples and was reflected by the highest misclassification rate. A similar observation was reported by Gerbig *et*

al. [88]. This phenomenon could be associated with necrosis and ischemia of the tumour. Enormously accelerated membrane lipid biosynthesis in proliferating tumour cells could also contribute to a noticeably wider variability of the acquired DESI data, as well as the different levels of de-differentiation of the analysed tumour samples [88].
In general, the ion maps obtained for fresh-frozen colorectal samples followed the histological patterns, providing a mean for histology-level tissue identification. Multivariate pattern recognition methods such as PCA and MMC+LDA could provide user-independent data in individual samples and larger cohorts of samples.

The results of the experiments described in the second part of *Chapter 5* aim to demonstrate that FFPE tissue samples are compatible with DESI-MSI, and their molecular content is sufficient to allow tissue discrimination.
The data presented here are based on a single TMA which was constructed using samples from only 18 patients. The collected data proved that tissue classification was possible but not at the level of tumour grade discrimination. This could be because most patients were diagnosed with grade 2 colorectal cancer (14 samples) and only one and three with grades 3 and 1, respectively. As a result, the sample number was deficient, and the number of samples representing different tumour grades was very unbalanced. Groseclose *et al.* reported that they distinguished different lung cancer histologies and individual cancer types [146]. Similar results were obtained by Morita *et al.*, who detected specific signals for each status of gastric cancer differentiation [247]. These two projects, however, focused on proteomic MALDI data. It is crucial to keep in mind that a complete lack of matrix deposition and the ability to perform the analysis at atmospheric pressure are two significant advantages of DESI analysis vs MALDI-MSI [220], but at the same time data generated by these two techniques are not entirely comparable. Nevertheless, it would be desirable to analyse more TMAs with a higher number of colorectal cancer subtypes to validate if similar results could be obtained with DESI-MSI.
It is important to stress that one of the main advantages of DESI-MSI is the visualization of the metabolites detected in samples while maintaining spatial integrity. This ability proves essential for tissue samples measurements that are thought to contain a heterogeneous distribution of cell types. Buck *et al.* employed MALDI-MSI to compare molecular content between fresh-frozen and FFPE samples and reported a similar metabolite spatial localization in analysed tissue sections [101]. That study was very similar to the one described here. It also reported that the collected spatially resolved mass spectra allowed successful discrimination of metabolic content between normal and diseased colorectal samples.

Moreover, TMAs were subjected to MALDI-MSI, and the acquired data were used to determine the metabolic 'fingerprints' of renal cell carcinoma and renal oncocytoma [101]. The ability to correctly separate samples representing those two diseases based on their metabolite signatures plus the high throughput analysis of large patients cohorts demonstrated the advantages of MALDI-MSI as a potential diagnostic tool [101]. Therefore, it is hypothesized that DESI-MSI should have similar potential.

Without any doubt, the ability to image metabolites should be considered a starting point in investigating metabolic pathways in FFPE tissue samples. In the future, it should lead to distinguishing the exact location of metabolic disturbances. Additionally, there is a continuous improvement in the mass resolution of MSI instruments, which may increase the chances of detecting low abundance metabolites. As a result, it should be possible to annotate detected peaks more accurately. Measurements of hundreds of molecules from FFPE blocks will likely be possible in the foreseeable future, helping generate pathways from thousands of patients. The ability to analyse TMAs will only speed up that process [101]. Therefore, the results presented in *Chapter 5* are significant technical advancements and carry the potential to directly impact diagnosis and predict patients' outcomes in the future.

Chapter 6 assesses the application of DESI-MSI to the process of diagnosis of cirrhotic liver diseases. 56 fresh-frozen liver samples representing seven liver diseases were subjected to the standard DESI-MSI and data analysis workflow. The first approach focused on the classification of individual diseases; however, that was not successful. Next, an attempt was made to distinguish AIH from PBC and AIH from PSC, as an overlapping syndrome characterises these two pairs of conditions. Since achieved results also were not satisfactory, it was hypothesized that the standard data analysis workflow might not be suitable for this dataset. As the median spectrum calculated over pixels representing each sample was used, it was feared that this approach failed to preserve spatial distributions and molecular heterogeneity of each MS image. This observation was reported by Inglese *et al.* [218]. As a result, an alternative statistical method described by the same author was implemented. It was based on ion colocalization features and not only preserved the spatial information but was also less sensitive to potential batch effects [218]. At this stage, all the samples were divided into three groups to reflect each disease's aetiology. Top correlations of ions were identified, and their distribution within analysed tissue sections was visualized.

It is possible that using this approach, some biochemical interactions that are distinguishing the three classes of cirrhotic liver diseases can be captured. The colocalization patterns can potentially be used for data-driven hypothesis generation, suggesting possible local molecular mechanisms characterizing the samples of interest. By visualizing the most significantly correlated ions, we can identify the specimen areas that are particularly interesting in future experiments.

Although DESI-MSI is not likely to replace routine histology techniques, it has the potential to be implemented into the current diagnostic workflow. The ultimate goal is to make DESI-MSI as histology-compatible as possible and the results presented in this thesis have proven that this is possible. The ability to use OCT as an embedding matrix for DESI-MSI means that all required tissue sections can easily be prepared as a part of the standard histological protocol. Those slides can then be subjected to histological and MS assessment, and the obtained results will complement each other and contribute to a quicker, more accurate diagnosis.
Moreover, the preliminary data suggest that high-throughput DESI-MSI runs are possible, and the quality of collected data is not compromised even during a 17-hours-long analysis. Without any doubt, further efforts should be undertaken to work on DESI-MSI automation, so the whole procedure is time-efficient and requires minimal input from staff members.
Additionally, collected data support the statement that DESI-MSI can be used to successfully analyse FFPE samples, including TMAs. The future use of TMAs can allow measurements of hundreds of molecules and can help generate pathways from thousands of patients. As FFPE samples have been archived for many years in hospitals worldwide, this presents an excellent source of information on multiple diseases and patients' outcomes to various treatments. Therefore, obtained results represent a significant technical advance and can play an essential role in diagnosis and predicting a patient outcome in the future.

7.2 Consideration for future studies

7.2.1 Quality control

Without any doubt, one of the limitations of the projects described in this thesis is the lack of proper quality control (QC). QC of the identified biomolecule peak intensities is crucial for reducing extreme biological outliers and process-based sources of variation. Without this step, statistical results can be biased.

When the data used in this thesis were acquired, we did not have any validated QC for DESI-MSI analysis. As a part of ongoing efforts to make up for this, the instrument was calibrated every day, and pork liver was used as a provisional QC. The reason why this tissue type was chosen was its availability and relatively homogenous structure. After calibration, a row in negative and positive ionization mode was recorded using a fresh-frozen pork liver tissue section. These data were then used to assess the instrument's shape, e.g. signal intensity, and, if needed, were treated as a 'baseline' for the rest of the data acquired on that particular day.
A small pork liver tissue section is placed on every slide alongside the sample to be analysed these days. This way, some test data are recorded under precisely the same conditions before moving onto the analysis of the sample. This approach has already proven to be very useful when statistical analysis of multiple samples is undertaken.

7.2.2 Batch effects

Another limitation that needs to be taken into account is so-called batch effects.
The process of identifying differential effects between sample classes (e.g. healthy versus diseased) in high-throughput biological data is crucial for providing potential drug targets as well as developing gene signatures for prognosis/ diagnosis. These effects are usually identified by statistical testing for relevant biological features (such as proteins or genes). This process requires identifying accurate signals against a complex backdrop of intra- and intersample variation [248]. However, it is essential to remember that some sources of variation are unrelated to intra- and intersample class differences. These are termed 'batch effects' and can arise from different experiment times and conditions, reagents, instrument users and instruments themselves [248, 249]. As a result, the presence of these batch effects can statistical bias models. Biased datasets, which consist of multiple MSI samples, often exhibit systematic variations of the mass spectral intensities of analysed subjects. Importantly, these variations are not a product of samples' inner properties but a direct consequence of the specific experimental conditions during the data acquisition. For this reason, normalisation techniques are necessary to reduce the variations in spectral intensities caused by possible batch effects [218]. As a more general overview, existing approaches to minimise batch effects include an appropriate study design (randomisation in the order of samples analysis, technical replicates), outlier detection (excluding extreme data, which allows increasing sensitivity for finding differences between classes), and data transformation [249].
Every possible care was taken to minimise potential batch effects when analysing tissue samples by DESI-MSI. Samples representing different conditions (normal versus cancer or

various cirrhotic liver diseases) were always analyzed in randomised order. However, it was appreciated that factors like fluctuations in signal intensity or even conditions in the laboratory (temperature, humidity) could contribute to batch effects. Due to the high number of samples (n=66) and lengthy analysis time per sample (on average 10-11 hours), the data acquired and presented in *Chapter 6* were the most likely to suffer from batch effects.

7.2.3 Data acquisition using a home-built sprayer

All frozen samples analysed as a part of this Thesis were measured using a home-built sprayer mounted to Thermo Exactive Orbitrap instrument. Therefore, the results presented in *Chapter 5* (fresh-frozen colorectal samples) and *Chapters 3* and *6* (fresh-frozen liver specimens) were based on the Swagelok T element sprayer. It is unavoidable to encounter some minor differences in the sprayer design depending on the person building it. Also, every possible care was taken when optimising the sprayer, but again, there was no formal written protocol to follow to ensure this step was performed as per the SOP. Lastly, geometric setup and environmental factors such as the temperature can directly impact the sprayer's performance. As a result, the sprayer setup needs to be closely controlled to obtain reliable data [118]. Therefore, continuous optimisation of the sprayer before DESI-MSI analysis was crucial to ensure the quality of collected data was up to the expected standards. Before commencing with tissue samples analysis, a couple of lines were recorded in both positive and negative modes using a pork liver tissue section. That material was relatively homogenous and therefore has been used as quality control.

Nevertheless, the optimisation procedure is currently non-standardised and relies on continuous adjustments to the sprayer before DESI-MSI analysis. Although normalisation was incorporated during data-processing to account for these variations, there is a high possibility that low-intensity peaks were eliminated due to ion suppression. Without any doubt, if the DESI-MSI platform is to be introduced as a part of a routine diagnostic portfolio, a more robust approach with a commercially available sprayer is a must. As such, the use of a redesigned sprayer described in *Chapter 2.9.2* should be considered for all future studies.

7.2.4 Expansion of the DESI-MSI database

An adequate DESI-MSI database is needed in order to be able to build reliable models for diseased versus normal tissue samples.

Larger cohorts of patient samples would need to be analysed to validate the findings from *Chapter 3*. In order to further probe any potential ion suppression effects, a wider-reaching experiment is proposed. This would encompass tissue samples acquired from a range of subjects with a range of tissue types. Analysing such samples embedded in H_2O and OCT would enable the classification according to tissue type and patient. Any potential interference could thus be placed in the context of inter-patient and inter-tissue type variation.

A limited number of fresh-frozen colorectal samples was analysed and included in *Chapter 5.3.1*. Ideally, potential future work will focus more on samples histologically classified as adenoma. Moreover, access to complete patients' metadata, including survival data, is essential.

In *Chapter 5* only one slide per patient was analysed. Going forward, DESI-MSI of TMA slides where multiple sections from the same patient should be considered "replication analysis" for future studies.

A very limited number of cirrhotic liver samples was available per each disease when the project described in *Chapter 6* was running. In a later course, additional samples were provided by the collaborators to make the groups representing various diseases more balanced.

7.2.5 DESI-MSI as a routinely used tool in clinical settings

The goal is to make DESI-MSI compatible with the requirements of clinical laboratory settings. The findings reported in this thesis regarding FFPE samples' analysis need to be further validated using established techniques like LC-MS or NMR. However, the preliminary results presented in *Chapter 4* and *Chapter 5* look very promising and bring DESI-MSI one step closer to applying this technique in histology-level cancer diagnostics. The molecular picture of the tissue section obtained by DESI-MSI makes this technique ideal for diagnostic tools to complement the current histology-based approach. Furthermore, this molecular information brings new insights about the systems lipid biochemistry.

The future of DESI-MSI ultimately lies in developing an automated platform of appropriate size to be placed in hospitals providing spatial data to aid diagnosis. In a clinical scenario, a biopsy may be taken, and instead of following the traditional histopathology processing protocol, clinicians may assign it to DESI-MSI analysis. Since this is a non-destructive technique, analysed tissue slides could be kept for histological interpretation and validation. Challenges that would need to be overcome include the mass spectrometer's size, the time required for image acquisition, and a validated lipidomic profiles database. Additional

considerations include the cost of mass spectrometer instrument implementation (instrument affordability, staff training, and required lab space).

7.3 Future directions

7.3.1 Automation of DESI-MSI as a step towards routine pathology testing

Future efforts should focus on making DESI-MSI as compatible with routine histopathology workflow as possible. That means that the following aspects covered in this thesis will require further investigation:

- OCT as an embedding medium vs MSI

Although results presented in *Chapter 3* indicate that OCT can assist with cryosectioning fresh-frozen samples as long as the data are adequately treated afterwards, a study involving a higher number of samples is required. Tissue specimens should represent various tissue types and be collected from multiple subjects. Acquiring data for such samples embedded in H_2O and OCT will enable the classification according to tissue type and patient. As a result, any potential interference could be placed in the context of the inter-tissue type and inter-patient variation. Proven compatibility of OCT with DESI-MSI would mean that there is no need to adjust the existing histopathology protocols in the future.

- High-throughput DESI-MSI

Only preliminary data for a relatively small number of samples were presented in *Chapter 5.3.2*. Again, the future continuation of this study needs to employ more tissue samples with clinically relevant metadata. The collection of the DESI data for various tissue types is essential to test the acquired data for tissue classification and build a sufficient database for future reference. That would allow fully automated tissue sections analysis where a dedicated software could compare and match the metabolomic content of an analysed specimen with the database.

- DESI of FFPE tissue samples

The results presented in *Chapter 5.3.4* indicate that metabolomic analysis of FFPE samples by DESI-MSI is possible. Despite the concerns that the routine histology procedures employed to allow the fixation of tissue specimens and further processing, it was proven that the molecular content of those samples was sufficient to allow for tissue classification. The findings reported in this thesis need to be further validated by a well-established analytical technique like LC-MS. That would allow minimising any bias from a single DESI-MSI platform. The next stage

would involve analysing a larger cohort of independent patient samples subjected to DESI-MSI and LC-MS.

7.3.2 DESI-MSI for cirrhotic liver diseases

The results presented in *Chapter 6* are for a batch of 66 liver specimens. However, as the total number of samples representing some diseases was deficient (e.g. only three AIH samples), the second batch of 69 specimens was added to this project. Those additional 69 samples have been analysed by DESI-MSI, but the analysis took place about 18-24 months after the first batch. An attempt was made to try and merge the DESI data so all the samples could be subjected to statistical analysis as one dataset. However, it was then discovered that there were apparent differences in the quality of acquired data. This was mainly due to the ongoing efforts to improve the signal quality, and as a result, a new sprayer with a slightly altered design has been used to analyse the second batch.

Nevertheless, the two batches' variability was too significant to try and treat them as one dataset. Therefore, it has been suggested that the whole project should be moved to a different instrument (Xevo QToF from Waters). This is the mass spectrometer that uses a novel and more reliable sprayer, as described in *Chapter 2.9.2.* Also, to take advantage of re-analysis of all the liver samples, the measurements will use the lower resolution to capture much more detailed information. A new, much more detailed histological validation will be required, too.

7.3.3 Cross comparison of colorectal samples lipid profiles with REIMS platform

Although the work carried out as part of this thesis focused on the practicalities of DESI-MSI application to routine histopathology settings, the collected data could be used to study changes in colorectal cancer lipidomics. There is a batch of about 120 fresh-frozen colorectal samples for which both DESI-MSI data and REIMS data are readily available.

Rapid evaporative ionisation mass spectrometry (REIMS), contrary to DESI-MSI, is a non-imaging technique that allows near real-time characterisation of human tissue *in vivo* or *in situ* via the analysis of the aerosol released during surgical interventions [250-252]. Similarly to DESI-MSI, the resulting mass spectrometric profiles are mainly generated from membrane phospholipids and are characteristic of the type of tissue under investigation [253, 254]. Although a cross-platform normalisation method for DESI-MSI and REIMS data has already been proposed [255], this kind of comparison has never been done so far.

The main aim of the cross-comparison between the DESI-MSI and the REIMS data for colorectal cancer data would be to elucidate essential diagnostic ions and validate the lipidomic changes in colorectal cancer. The validation of the "lipidomic fingerprint" would improve the REIMS technique's probability of becoming a routine, providing a real-time tissue characterisation tool during surgery, and enhancing the implementation of DESI-MSI in diagnostic pathology.

Comparative analysis between these two platforms would contribute to strengthening our understanding of lipidomic dysregulation in colorectal cancer.

7.4 Conclusions

The general aim of this PhD was to look into the DESI-MSI technique from the histopathology's perspective. As the ultimate goal is to introduce DESI-MSI to diagnostic laboratories, it needs to be further optimised and tested so it can potentially be implemented into routine histopathology investigations. Therefore, the first chapter looks into the compatibility of OCT embedding medium and DESI-MSI. Chapter two describes the optimisation process of utilising DESI-MSI to analyse FFPE tissue samples. Lastly, chapters three and four present the results of DESI-MSI analysis for both fresh-frozen and FFPE colorectal samples, as well as for some cirrhotic liver diseases.

The results presented in this thesis suggest that DESI-MSI can be considered a histology-friendly analytical technique. It can be applied to analyse tissue samples embedded in OCT, providing that the collected data is subjected to a polymer removal step first. Moreover, the DESI-MSI protocol can be adjusted to allow for a successful analysis of FFPE samples. Without any doubt, this is a significant achievement, as the majority of samples processed by diagnostic histopathology laboratories are fixed in formalin. Additionally, as other studies have claimed, DESI-MSI is an efficient approach to look into alternations in phospholipids biochemistry in colorectal cancer and cirrhotic liver diseases.

As the ultimate goal is not to replace the current histopathology approach but add DESI-MSI as a complementary technique, it is evident that it needs to be compatible with the standard pathology protocols. The data collected during my PhD prove that DESI-MSI has the potential to be implemented alongside the routine histopathology workflow.

References

1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2019.* CA Cancer J Clin, 2019. **69**(1): p. 7-34.
2. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2020.* CA Cancer J Clin, 2020. **70**(1): p. 7-30.
3. Torre, L.A., et al., *Global cancer statistics, 2012.* CA Cancer J Clin, 2015. **65**(2): p. 87-108.
4. Kamangar, F.D., G. M.; Anderson, W. F., *Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world.* J Clin Oncol, 2006. **24**(14): p. 2137-2150.
5. Cooper, G.M., *The Cell: A Molecular Approach. 2nd edition.* Sinauer Associates, 2000.
6. Hanahan, D.W., R. A., *The hallmarks of cancer.* Cell, 2000. **100**: p. 57-70.
7. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
8. Harris, C.C., *p53 tumor suppressor gene: from the basic research laboratory to the clinic - an abridged historical perspective.* Carcinogenesis, 1996. **17**: p. 1187-1198.
9. Leine, A.J., *p53, the cellular gatekeeper for growth and division.* Cell, 1997. **88**: p. 323–331.
10. Bouck, N.S., V.; Hsu, S. C., *How tumors become angiogenic.* Adv Cancer Res, 1996. **69**: p. 135-174.
11. Folkman, J., *Tumor angiogenesis.* In Cancer Medicine, 1997: p. 181-204.
12. Fedi, P.T., S. R.; Aaronson, S. A., *Growth factors.* In Cancer Medicine, 1997: p. 41-64.
13. Veikkola, T.A., K., *VEGFs, receptors and angiogenesis.* Semin Cancer Biol, 1999. **9**: p. 211-220.
14. Bull, H.A.B., P. M.; Dowd, P. M., *Src-related protein tyrosine kinases are physically associated with the surface antigen CD36 in human dermal microvascular endothelial cells.* FEBS Lett, 1994. **351**: p. 41-44.
15. Blasco, M.A., *Telomeres and human disease: ageing, cancer and beyond.* Nat Rev Genet, 2005. **6**: p. 611-622.
16. Shay, J.W.W., W. E., *Hayflick, his limit, and cellular ageing.* . Nat Rev Mol Cell Biol, 2000. **1**: p. 72-76.
17. Berx, G.v.R., F., *Involvement of members of the cadherin superfamily in cancer* Cold Spring Harb Perspect Biol., 2009. **1**(6): p. 1-27.
18. Cavallaro, U.C., G., *Cell adhesion and signalling by cadherins and Ig-CAMs in cancer* Nat Rev Cancer, 2004. **4**: p. 118-132.
19. Fahy, E., et al., *A comprehensive classification system for lipids.* J Lipid Res, 2005. **46**(5): p. 839-61.
20. Fahy, E., et al., *Update of the LIPID MAPS comprehensive classification system for lipids.* J Lipid Res, 2009. **50 Suppl**: p. S9-14.
21. Fahy, E., et al., *Lipid classification, structures and tools.* Biochim Biophys Acta, 2011. **1811**(11): p. 637-47.
22. Sandra, K. and P. Sandra, *Lipidomics from an analytical perspective.* Curr Opin Chem Biol, 2013. **17**(5): p. 847-53.
23. Wishart, D.S., et al., *HMDB: a knowledgebase for the human metabolome.* Nucleic Acids Res, 2009. **37**(Database issue): p. D603-10.
24. Zhao, Y.Y.C., X.I.; Lin, R.C., *Lipidomics applications for discovering biomarkers of diseases in clinical chemistry.* . International Review of Cell and Molecular Biology, 2014. **313**: p. 1-26.
25. Santos, C.R. and A. Schulze, *Lipid metabolism in cancer.* FEBS J, 2012. **279**(15): p. 2610-23.
26. Tennant, D.A.D., R. V.; Gottlieb, E., *Targeting metabolic transformation for cancer therapy.* Nat Rev Cancer, 2010. **10**(4): p. 267-277.

27. Cairns, R.A.H., I. S.; Mak, T. W., *Regulation of cancer cell metabolism.* Nat Rev Cancer, 2011. **11**: p. 85-95.
28. Warburg, O., *On the origin of cancer cells.* Science, 1956. **123**: p. 309-314.
29. Vander Heiden, M.G.C., L. C.; Thompson, C. B., *Understanding the Warburg effect: the metabolic requirements of cell proliferation.* Science, 2009. **324**: p. 1029-1033.
30. Menendez, J.A.L., R., *Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis* Nat Rev Cancer, 2007. **7**: p. 763-777.
31. Zhang, F. and G. Du, *Dysregulated lipid metabolism in cancer.* World J Biol Chem, 2012. **3**(8): p. 167-74.
32. Cardiff, R.D.G., J.P.; Miller, J.W.; Axelrod, D.E.; Borowsky, A.D., *Histopathology as a Predictive Biomarker Strengths and Limitations.* The Journal of Nutrition, 2006. **136**: p. 2673S–2675S.
33. Ramaswamy, S., et al., *Multiclass cancer diagnosis using tumor gene expression signatures.* Proc Natl Acad Sci U S A, 2001. **98**(26): p. 15149-54.
34. Galant, C.M., J.; Sibille, C.; Docquier, P. L.; Delloye, C., *Current limitations to the histopathological diagnosis of some frequently encountered bone tumours.* Acta Orthop. Belg., 2008. **74**: p. 1-6.
35. Grønbæk, K.C., P. B.; Hamilton-Dutoit, S.; Federspiel, B. H.; Hage, E.; Jensen, O. J.; Vyberg, M., *Interobserver variation in interpretation of serial liver biopsies from patients with chronic hepatitis C.* Journal of Viral Hepatitis, 2002. **9**: p. 443–449.
36. Rousselet, M.C., et al., *Sources of variability in histological scoring of chronic viral hepatitis.* Hepatology, 2005. **41**(2): p. 257-64.
37. Goldin, R.D.G., J. G.; Burt, A. D.; Dhillon, P. A.; Hubscher, S.; Wyatt, J.; Patel, N., *Intra-observer and inter-observer variation in the histopathological assessment of chronic viral hepatitis.* Journal of Hepatology, 1996. **25**: p. 649-654.
38. Hillman, H., *Limitations of clinical and biological histology.* Med Hypotheses, 2000. **54**(4): p. 553-64.
39. Muskett, D., *From specimen to slide. In: Orchard, G. & Nation, B. Histopathology.* Oxford University Press, 2012: p. 79-126.
40. Cook, D.J., *Fixation and fixatives. In: Cook, D.J. Cellular pathology. An introduction to techniques and applications. 2nd edition.* Scion Publishing Ltd, 2006: p. 9-20.
41. Hopwood, D., *Fixation and fixatives. In: Bancroft, J.D. & Gamble, M. Theory and practice of histological techniques. 6th edition.* Elsevier Health Sciences, 2007: p. 63-84.
42. Kiernan, J.A., *Formaldehyde, formalin, paraformaldehyde and glutaraldehyde: what they are and what they do.* Microscopy Today, 2000. **00-1**: p. 8-12.
43. Buesa, R.J., *Histology without formalin?* Methods in Pathology, 2008. **12**: p. 387-396.
44. Thavarajah, R.M., V. K.; Elizabeth, J.; Krishnamohan Rao, U.; Ranganathan, K., *Chemical and physical basics of routine formaldehyde fixation.* Journal of Oral and Maxillofacial Pathology, 2012. **16**(3): p. 400-405.
45. Abe, M., Takahashi, M., Horiuchi, K. & Nagano, A., *The changes in crosslink contents in tissue after formalin fixation.* Analytical Biochemistry, 2003. **318**: p. 118-123.
46. Schulz, G., Crooijmans, H. J. A., German, M., Scheffler, K., Müller-Gerbl, M. & Müller, B. , *Three-dimensional strain fields in human brain resulting from formalin fixation.* Journal of Neuroscience Methods, 2011. **202**: p. 17-27.
47. Chen, C.H., Hsu, M. Y., Jiang, R. S., Wu, S. H., Chen. F. J. & Liu, S. A., *Shrinkage of head and neck cancer specimens after formalin fixation.* Journal of the Chinese Medical Association, 2012. **75**: p. 109-113.
48. Cox. M. L., S., C. L., Luster, C. N., Stewart, Z. S., Korytko, P. J., Khan, K. N. M., Paulauskis, J. D. & Dunstan, R. W., *Assessment of fixatives, fixation, and tissue processing on morphology and RNA integrity.* Experimental and Molecular Pathology, 2006. **80**: p. 183-191.
49. Milcheva, R., Janega, P., Celec, P., Russev, R. & Babál, P., *Alcohol based fixatives provide excellent tissue morphology, protein immunoreactivity and RNA integrity in parrafin embedded tissue specimens.* Acta Histochemica, http://dx.doi.org/10.1016/j.acthis.2012.08.002, 2012.
50. Gillespie, J.W., Best, C. J. M., Bichsel, V. E., Cole, K. A., Greenhut, S. F., Hewitt, S. M., Ahram, M., Gathright, Y. B., Merino, M. J., Strausberg, R. L., Epstein, J. I., Hamilton, S. R.,

Gannot, G., Baibakova, G. V., Calvert, V. S., Flaig, M. J., Chuaqui, R. F., Herring, J. C., Pfeifer, J., Petricoin, E. F., Linehan, W. M., Duray, P. H., Bova, G. S. & Emmert-Buck, M. R., *Evaluation of non-formalin tissue fixation for molecular profiling studies.* . American Journal of Pathology, 2002. **160**(2): p. 449-457.
51. Scalia, C.R., et al., *Antigen masking during fixation and embedding, dissected.* J Histochem Cytochem, 2017. **65**(1): p. 5-20.
52. Sompuram, S.R.V., K.; Messana, E.; Bogen, S. A., *A molecular mechanism of formalin fixation and antigen retrieval.* American Journal of Clinical Pathology, 2004. **121**: p. 190-199.
53. O'Leary, T.J., et al., *Protein fixation and antigen retrieval: chemical studies.* Biotech Histochem, 2009. **84**(5): p. 217-21.
54. Singer, S.J.,; Nicolson, G. L., *The fluid mosaic model of the structure of cell membranes* Science, 1972. **17**: p. 720-731.
55. Suvarna, S.K.L., C.; Bancroft, J. D., *Bancroft's Theory and practice of histological techniques.* 2013(Seventh Edition).
56. Warth, A., et al., *Individualized medicine and demographic change as determining workload factors in pathology: quo vadis?* Virchows Arch, 2016. **468**(1): p. 101-8.
57. Arentz, G., et al., *Applications of mass spectrometry imaging to cancer.* Adv Cancer Res, 2017. **134**: p. 27-66.
58. Watson, T.S., D., *Introduction to mass spectrometry. Instrumentation, applications and strategies for data interpretation. 4th edition.* John Wiley & Sons, Ltd, 2007.
59. Paul, W.S., H., *Z. Naturforsch, A.* 1953. **8**: p. 448.
60. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis.* . Anal Chem, 2000. **78**: p. 5994-6002.
61. Guilhaus, M., *Principles and instrumentation for TOF-MS.* J Mass Spectrom, 1995. **30**: p. 1519-1532.
62. Wollnik, H., *TOF-MS.* Mass Spectrom Rev, 1993. **12**: p. 89-114.
63. Mamyrin, B.A., *Time-of-flight mass spectrometry (concepts, achievements, and prospects).* . International Journal of Mass Spectrometry, 2001. **206**(3): p. 251-266.
64. Wiseman, J.M.I., D. R.; Zhu, Y.; Kissinger, C. B.; Manicke, N. E.; Kissinger, P. T.; Cooks, R. G., *Desorption electrospray ionization mass spectrometry: imaging drugs and metabolites in tissues.* . PNAS, 2008. **105**(47): p. 18120-18125.
65. Schwamborn, K. and R.M. Caprioli, *Molecular imaging by mass spectrometry--looking beyond classical histology.* Nat Rev Cancer, 2010. **10**(9): p. 639-46.
66. Prideaux, B. and M. Stoeckli, *Mass spectrometry imaging for drug distribution studies.* J Proteomics, 2012. **75**(16): p. 4999-5013.
67. Chaurand, P., *Imaging mass spectrometry of thin tissue sections: a decade of collective efforts.* J Proteomics, 2012. **75**(16): p. 4883-4892.
68. Vickerman, J.C., *Molecular imaging and depth profiling by mass spectrometry--SIMS, MALDI or DESI?* Analyst, 2011. **136**(11): p. 2199-217.
69. Vaysse, P.M., et al., *Mass spectrometry imaging for clinical research - latest developments, applications, and current limitations.* Analyst, 2017. **142**(15): p. 2690-2712.
70. Fletcher, J.S.L., N. P.; Vickerman, J. C., *Developments in molecular SIMS depth profiling and 3D imaging of biological systems using polyatomic primary ions* Mass Spectrom Rev, 2011. **30**(1): p. 142-174.
71. Bodzon-Kulakowska, A.S., P., *Imaging mass spectrometry: Instrumentation, applications, and combination with other visualization techniques* Mass Spectrom Rev, 2016. **35**(1): p. 147-169.
72. Karas, M.B., D.; Bahr, U.; Hillenkamp F., *Matrix-Assisted Laser Desorption Ionization Mass Spectrometry.* Int J Mass Spec Ion Proc, 1987. **78**: p. 53-68.
73. Karas, M.H.F., *Laser desorption of proteins with molecular masses exceeding 10,000 Daltons.* Anal Chem, 1988. **60**(20): p. 2299-2301.
74. Karas, M.B., U.; Giessman, U., *MALDI-MS.* Mass Spectrom Rev, 1991. **10**: p. 335-357.
75. Hillenkamp, F.K., M,; Beavis, R. C.; Chait, B. T., *Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers.* Anal Chem, 1991. **63**(24): p. 1193A-1203A.
76. Goodwin, R.J.A., *Sample preparation for mass spectrometry imaging: small mistakes can lead to big consequences.* J Proteomics, 2012. **75**(16): p. 4893-4911.

77. Kaletas, B.K.v.d.W., I. M.; Stauber, J.; Dekker, L. J.; Güzel, C.; Kros, J. M.; Luider, T. M.; Heeren, R. M. A., *Sample preparation issues for tissue imaging by imaging MS.* Proteomics, 2009. **9**(10): p. 2622-2633.
78. Takats, Z.W., J. M.; Gologan, B.; Cooks, R. G., *Mass spectrometry sampling under ambient conditions with desorption electrospray ionization.* Science, 2004. **306**: p. 471 - 473.
79. Takats, Z., J.M. Wiseman, and R.G. Cooks, *Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology.* J Mass Spectrom, 2005. **40**(10): p. 1261-75.
80. Fenn, J.B.M.M.M., C.K.; Wong, S.F.; Whitehouse, C.M., *Electrospray ionization – principles and practice.* Mass Spectrometry Reviews, 1990. **9**(1): p. 37 - 70.
81. Cooks, R.G.O., Z.; Takats, Z.; Wiseman, J. M., *Ambient mass spectrometry.* Science, 2006. **311**(5767): p. 1566-1570.
82. Venter, A.S., P. E.; Cooks, R. G., *Droplet dynamics and ionization mechanisms in desorption electrospray ionization mass spectrometry.* Anal Chem, 2006. **78**(24): p. 8549–8555.
83. Calligaris, D.C., D.; Liu, X.; Norton, I.; Thompson, C. J.; Richardson, A. L.; Golshan, M.; Easterling, M. L.; Santagata, S.; Dillon, D. A.; Jolesz, F. A.; Agar, N. Y. R., *Application of desorption electrospray ionization mass spectrometry imaging in breast cancer margin analysis.* PNAS, 2014. **111**(42): p. 15184 - 15189.
84. Eberlin, L.S., et al., *Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors.* Proc Natl Acad Sci U S A, 2013. **110**(5): p. 1611-6.
85. Eberlin, L.S.T., R. J.; Zhang, J.; Longacre, T. A.; Berry, G. J.; Bingham, D. B.; Norton, J. A.; Zare, R. N.; Poultsides, G. A., *Molecular assessment of surgical-resection margins of gastric cancer by mass-spectrometric imaging.* PNAS, 2014. **111**(7): p. 2436–2441.
86. Eberlin, L.S.D., A. L.; Costa, A. B.; Ifa, D. R.; Cheng, L.; Masterson, T.; Koch, M.; Ratliff, T. L.; Cooks, R. G., *Cholesterol sulfate imaging in human prostate cancer tissue by Desorption Electrospray Ionization Mass Spectrometry.* Anal Chem, 2010. **82**: p. 3430–3434.
87. Dill, A.L.E., L. S.; Costa, A. B.; Zheng, C.; Ifa, D. R.; Cheng, L., *Multivariate statistical identification of human bladder carcinomas using ambient ionization imaging mass spectrometry* Chemistry - A European Journal, 2011. **17**(10): p. 2897-2902.
88. Gerbig, S.G., O.; Balog, J.; Denes, J.; Baranyai, Z.; Zarand, A.; Raso, E.; Timar J.; Takats, Z., *Analysis of colorectal adenocarcinoma tissue by desorption electrospray ionization MSI.* Anal Bioanal Chem, 2012. **403**: p. 2315-2325.
89. Masterson, T.A.D., A. L.; Eberlin, L. S.; Mattarozzi, M.; Cheng, L.; Beck, S. D. W.; Bianchi, F.; Cooks, R. G., *Distinctive glycerophospholipid profiles of human seminoma and adjacent normal tissues by desorption electrospray ionization imaging mass spectrometry* J Am Soc Mass Spectrom, 2011. **22**(8): p. 1326-1333.
90. Eberlin, L.S., et al., *Classifying human brain tumors by lipid imaging with mass spectrometry.* Cancer Res, 2012. **72**(3): p. 645-54.
91. Guenther, S., et al., *Spatially resolved metabolic phenotyping of breast cancer by desorption electrospray ionization mass spectrometry.* Cancer Res, 2015. **75**(9): p. 1828-37.
92. Doria, M.L., et al., *Epithelial ovarian carcinoma diagnosis by desorption electrospray ionization mass spectrometry imaging.* Sci Rep, 2016. **6**: p. 39219.
93. Jarmusch, A.K.P., V.; Baird, Z.; Hattab, E. M.; Cohen-Gadol, A. A.; Cooks, R. G., *Lipid and metabolite profiles of human brain tumors by desorption electrospray ionization-MS.* PNAS, 2016. **113**(6): p. 1486-1491.
94. Watrous, J.D., T. Alexandrov, and P.C. Dorrestein, *The evolving field of imaging mass spectrometry and its impact on future biological research.* J Mass Spectrom, 2011. **46**(2): p. 209-22.
95. Blanksby, S.J. and T.W. Mitchell, *Advances in mass spectrometry for lipidomics.* Annu Rev Anal Chem (Palo Alto Calif), 2010. **3**: p. 433-65.
96. Jackson, S.N. and A.S. Woods, *Direct profiling of tissue lipids by MALDI-TOFMS.* J Chromatogr B Analyt Technol Biomed Life Sci, 2009. **877**(26): p. 2822-9.
97. Murphy, R.C., J.A. Hankin, and R.M. Barkley, *Imaging of lipid species by MALDI mass spectrometry.* J Lipid Res, 2009. **50 Suppl**: p. S317-22.

98. Eberlin, L.S., et al., *Desorption electrospray ionization mass spectrometry for lipid characterization and biological tissue imaging.* Biochim Biophys Acta, 2011. **1811**(11): p. 946-60.
99. Lai, Z.W.W., J.; Nilse, L.; Costa§, F.; Keller, E.; Tholen, M; Kizhakkedathu, J. N.; Biniossek, M.; Bronsert, P.; Schilling, O., *Formalin-fixed, paraffin-embedded tissues (FFPE) as a robust source for the profiling of native and protease-generated protein amino termin.* Technological Innovation and Resources, 2016. **15**: p. 2203-2213.
100. Fowler, C.B., T.J. O'Leary, and J.T. Mason, *Toward improving the proteomic analysis of formalin-fixed, paraffin-embedded tissue.* Expert Rev Proteomics, 2013. **10**(4): p. 389-400.
101. Buck, A., et al., *High-resolution MALDI-FT-ICR MS imaging for the analysis of metabolites from formalin-fixed, paraffin-embedded clinical tissue samples.* J Pathol, 2015. **237**(1): p. 123-32.
102. Buck, A., et al., *How Suitable is Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight for Metabolite Imaging from Clinical Formalin-Fixed and Paraffin-Embedded Tissue Samples in Comparison to Matrix-Assisted Laser Desorption/Ionization-Fourier Transform Ion Cyclotron Resonance Mass Spectrometry?* Anal Chem, 2016. **88**(10): p. 5281-9.
103. Deininger, S.O.C., D. S.; Paape, R.; Becker, M.; Pineau, C.; Rauser, S.; Walch, A.; Wolski, E., *Normalization in MALDI-TOF imaging datasets of proteins: practical considerations.* Analytical and Bioanalytical Chemistry, 2011. **401**(1): p. 167-181.
104. Alexandrov, T., *MALDI imaging mass spectrometry: statistical data analysis and current computational challenges.* BMC Bioinformatics, 2012. **13**(16): p. S11.
105. Tibshirani, R.H., T.; Narasimhan, B.; Soltys, S.; Shi, G.; Koong, A.; Le, Q. T., *Sample classification from protein mass spectrometry, by 'peak probability contrasts'.* Bioinformatics, 2004. **20**(17): p. 3034-3044.
106. Morris, J.S.C., K. R.; Koomen, J.; Baggerly, K. A.; Kobayashi, R., *Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum.* Bioinformatics, 2005. **21**(9): p. 1764-1775.
107. Jones, E.A., et al., *Imaging mass spectrometry statistical analysis.* J Proteomics, 2012. **75**(16): p. 4962-4989.
108. Alexandrov, T.K., J. H., *Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering.* Bioinformatics, 2011. **27**(13): p. i230-238.
109. Scherf, U.R., D. T.; Waltham, M.; Smith, L. H.; Lee, J. K.; Tanabe, L.; Kohn, K. W.; Reinhold, W. C.; Myers, T. G.; Andrews, D. T.; Scudiero, D. A.; Eisen, M. B.; Sausville, E. A.; Pommier, Y.; Botstein, D.; Brown, P. O.; Weinstein, J. N., *A gene expression database for the molecular pharmacology of cancer.* Nat Genet, 2000. **24**(3): p. 236-244.
110. Schwartz, S.A.W., R. J.; Johnson, M. D.; Toms, S. A.; Caprioli, R. M., *Protein profiling in brain tumors using mass spectrometry: Feasibility of a new technique for the analysis of protein expression.* Clinical Cancer Research, 2004. **10**(3): p. 981-987.
111. Luo, D.D., C.; Huang, H., *Linear Discriminant Analysis: New Formulations and Overfit Analysis.* 25th AAAI Conference on Artificial Intelligence, 2011: p. 417-422.
112. Yuan, H.L.L., Y.; Yang, L. N.; Luo, H. W.; Tang, Y. Y., *Spectral-Spatial Linear Discriminant Analysis for Hyperspectral Image Classification.* 2013 Ieee International Conference on Cybernetics (Cybconf), 2013.
113. Barker, M.R., W., *Partial least squares for discrimination.* Journal of Chemometrics, 2003. **17**(3): p. 166-173.
114. Veselkov, K.A.M., R.; Strittmatter, N.; Goldin, R.; Kinross, J.; Speller, A. V. M.; Abramov, T.; Jones, E. A.; Darzi, A.; Holmes, E.; Nicholson, J. K.; Takats, Z., *Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer.* PNAS, 2014. **111**(3): p. 1216–1221.
115. Hilario, M.K., A.; Pellegrini, C.; Muller, M., *Processing and classification of protein mass spectra.* Mass Spectrom Rev, 2006. **25**(3): p. 409-449.
116. Balluff, B.E., M.; Kowarsch, A.; Rauser, S.; Meding, S.; Schuhmacher, C.; Feith, M.; Herrmann, K.; Rocken, C.; Schmid, R. M.; Hofler, H.; Walch, A.; Ebert, M. P., *Classification of HER2/neu Status in Gastric Cancer Using a Breast-Cancer Derived Proteome Classifier.* J Proteome Res, 2010. **9**: p. 6317-6322.

117. Mager, S.R., et al., *Standard operating procedure for the collection of fresh frozen tissue samples.* Eur J Cancer, 2007. **43**(5): p. 828-34.
118. Tillner, J., et al., *Investigation of the Impact of Desorption Electrospray Ionization Sprayer Geometry on Its Performance in Imaging of Biological Tissue.* Anal Chem, 2016. **88**(9): p. 4808-16.
119. Abbassi-Ghadi, N.J., E. A.; Veselkov, K. A.; Huang, J.; Kumar, S.; Strittmatter, N.; Golf, O.; Kudo, H.; Goldin, R. D.; Hanna, G. B.; Takats, Z., *Repeatability and reproducibility of desorption electrospray ionization-mass spectrometry (DESI-MS) for the imaging analysis of human cancer tissue: a gateway for clinical applications.* Analytical Methods, 2015. **7**(1): p. 71-80.
120. Tillner, J., et al., *Faster, More Reproducible DESI-MS for Biological Tissue Imaging.* J Am Soc Mass Spectrom, 2017. **28**(10): p. 2090-2098.
121. Schramm, T., et al., *imzML--a common data format for the flexible exchange and processing of mass spectrometry imaging data.* J Proteomics, 2012. **75**(16): p. 5106-5110.
122. Molinaro, A.M., R. Simon, and R.M. Pfeiffer, *Prediction error estimation: a comparison of resampling methods.* Bioinformatics, 2005. **21**(15): p. 3301-7.
123. Jones, M.L.L.J., M.; Bancroft, J. D.; Gamble, M., *Theory and Practice of Histological Techniques.* 2008: p. 135-160.
124. Lim, J., et al., *Fresh-frozen, optimal cutting temperature (OCT) compound-embedded bone marrow aspirates: a reliable resource for morphological, immunohistochemical and molecular examinations.* Int J Lab Hematol, 2010. **32**(1 Pt 1): p. e34-9.
125. Weston, L.A. and A.B. Hummon, *Comparative LC-MS/MS analysis of optimal cutting temperature (OCT) compound removal for the study of mammalian proteomes.* Analyst, 2013. **138**(21): p. 6380-4.
126. Fischer, A.H., et al., *Cutting sections of paraffin-embedded tissues.* CSH Protoc, 2008. **2008**: p. pdb prot4987.
127. Turbett, G.R.S., L. N., *The use of optimal cutting temperature compound can inhibit amplification by polymerase chain reaction.* Diagnostic Molecular Pathology: the American Journal of Surgical Pathology, 1997. **6**(5): p. 298-303.
128. Schwartz, S.A., M.L. Reyzer, and R.M. Caprioli, *Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation.* J Mass Spectrom, 2003. **38**(7): p. 699-708.
129. Strohalm, M., et al., *Poly[N-(2-hydroxypropyl)methacrylamide]-based tissue-embedding medium compatible with MALDI mass spectrometry imaging experiments.* Anal Chem, 2011. **83**(13): p. 5458-62.
130. Altelaar, A.F.M.v.M., J.; Jimenez, C. R.; Heeren, R. M. A.; Piersma, S. R., *Direct molecular imaging of Lymnaea stagnalis nervous tissue at subcellular spatial resolution by mass spectrometry.* Anal Chem, 2005. **77**: p. 735-741.
131. DeKeyser, S.S., et al., *Imaging mass spectrometry of neuropeptides in decapod crustacean neuronal tissues.* J Proteome Res, 2007. **6**(5): p. 1782-91.
132. Stoeckli, M., et al., *Imaging of a beta-peptide distribution in whole-body mice sections by MALDI mass spectrometry.* J Am Soc Mass Spectrom, 2007. **18**(11): p. 1921-4.
133. Stoeckli, M.S., D.; Schweitzer, A., *Compound and metabolite distribution measured by MALDI mass spectrometric imaging in whole-body tissue sections.* International Journal of Mass Spectrometry, 2007. **260**(2-3): p. 195-202.
134. Zaima, N., et al., *Application of imaging mass spectrometry for the analysis of Oryza sativa rice.* Rapid Commun Mass Spectrom, 2010. **24**(18): p. 2723-9.
135. Berry, K.A., et al., *MALDI imaging MS of phospholipids in the mouse lung.* J Lipid Res, 2011. **52**(8): p. 1551-60.
136. Manicke, N.E., et al., *High-resolution tissue imaging on an orbitrap mass spectrometer by desorption electrospray ionization mass spectrometry.* J Mass Spectrom, 2010. **45**(2): p. 223-6.
137. Wiseman, J.M., et al., *Tissue imaging at atmospheric pressure using desorption electrospray ionization (DESI) mass spectrometry.* Angew Chem Int Ed Engl, 2006. **45**(43): p. 7188-92.

138. Abbassi-Ghadi, N., et al., *Imaging of Esophageal Lymph Node Metastases by Desorption Electrospray Ionization Mass Spectrometry.* Cancer Res, 2016. **76**(19): p. 5647-5656.
139. Weaver, R. and R.J. Riley, *Identification and reduction of ion suppression effects on pharmacokinetic parameters by polyethylene glycol 400.* Rapid Commun Mass Spectrom, 2006. **20**(17): p. 2559-64.
140. Tong, H.B., D.; Tabei, K.; Siegel, M. M., *Automated Data Massaging, Interpretation, and E-Mailing Modules for High Throughput Open Access Mass Spectrometry.* J Am Soc Mass Spectrom, 1999. **10**: p. 1174–1187.
141. Furey, A., et al., *Ion suppression; a critical review on causes, evaluation, prevention and applications.* Talanta, 2013. **115**: p. 104-22.
142. Annesley, T.M., *Ion suppression in mass spectrometry.* Clinical Chemistry, 2004. **47**(9): p. 1041–1044.
143. Gaffney, E.F., et al., *Factors that drive the increasing use of FFPE tissue in basic and translational cancer research.* Biotech Histochem, 2018: p. 1-14.
144. Jiang, X.F., S.; Tian, R.; Ye, M.; Zou, H., *Development of efficient protein extraction methods for shotgun proteome analysis of formalin-fixed tissues.* Journal of Proteome Research 2007. **6**: p. 1038-1047.
145. Hwang, S.I., et al., *Direct cancer tissue proteomics: a method to identify candidate cancer biomarkers from formalin-fixed paraffin-embedded archival tissues.* Oncogene, 2007. **26**(1): p. 65-76.
146. Groseclose, M.R., et al., *High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry.* Proteomics, 2008. **8**(18): p. 3715-24.
147. Crockett, D.K., et al., *Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS.* Lab Invest, 2005. **85**(11): p. 1405-15.
148. Hood, B.L., et al., *Proteomic analysis of formalin-fixed prostate cancer tissue.* Mol Cell Proteomics, 2005. **4**(11): p. 1741-53.
149. Niehoff, A.C., et al., *Analysis of Drosophila lipids by matrix-assisted laser desorption/ionization mass spectrometric imaging.* Anal Chem, 2014. **86**(22): p. 11086-92.
150. Shi, S.R.K., C.; Balgley, B. M.; Lee, C.; Taylor, C. R., *Protein extraction from formalin-fixed, paraffin-embedded tissue sections quality evaluation by mass spectrometry.* Journal of Histochemistry and Cytochemistry, 2006. **54**(6): p. 739-743.
151. Shi, S.R., et al., *Complete solubilization of formalin-fixed, paraffin-embedded tissue may improve proteomic studies.* Proteomics Clin Appl, 2013. **7**(3-4): p. 264-72.
152. Ikeda, K.M., T.; Kanoh, T.; et al., *Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections.* Journal of Histochemistry and Cytochemistry, 1998. **46**(3): p. 397-403.
153. Warner, C.R.D., D. H.; Fitzgerald, M. C.; Joe Jr, F. L.; Diachenko, G. W., *Determination of free and reversibly bound sulphite in foods by reverse-phase, ion-pairing high-performance liquid chromatography.* Food Additives & Contaminants, 1990. **7**(5): p. 575-581.
154. Tanca, A., D. Pagnozzi, and M.F. Addis, *Setting proteins free: progresses and achievements in proteomics of formalin-fixed, paraffin-embedded tissues.* Proteomics Clin Appl, 2012. **6**(1-2): p. 7-21.
155. Bhaganna, P., et al., *Hydrophobic substances induce water stress in microbial cells.* Microb Biotechnol, 2010. **3**(6): p. 701-16.
156. Salvi, G.D.L.R., P.; Vendruscolo, M., *Effective interactions between chaotropic agents and proteins.* PROTEINS: Structure, Function, and Bioinformatics, 2005. **61**: p. 492-499.
157. Casadonte, R. and R.M. Caprioli, *Proteomic analysis of formalin-fixed paraffin-embedded tissue by MALDI imaging mass spectrometry.* Nat Protoc, 2011. **6**(11): p. 1695-709.
158. Siegel, R., C. Desantis, and A. Jemal, *Colorectal cancer statistics, 2014.* CA Cancer J Clin, 2014. **64**(2): p. 104-17.
159. Siegel, R.L., et al., *Colorectal cancer statistics, 2017.* CA Cancer J Clin, 2017. **67**(3): p. 177-193.

160. Statistics, O.f.N., *Cancer diagnoses and age-standardised incidence rates for all types of cancer by age, sex and region including breast, prostate, lung and colorectal cancer.* Cancer registration statistics, England: 2017, 2019. **1-16**.
161. force., P.U.S.t., *Screening for colorectal cancer: U.S. preventive services task force recommendation statement.* Ann Intern Med, 2008. **149**: p. 627–637.
162. Ahnen, D.J., et al., *The increasing incidence of young-onset colorectal cancer: a call to action.* Mayo Clin Proc, 2014. **89**(2): p. 216-24.
163. Campos, F.G., *Colorectal cancer in young adults: A difficult challenge.* World J Gastroenterol, 2017. **23**(28): p. 5041-5044.
164. Connell, L.C., et al., *The rising incidence of younger patients with colorectal cancer: questions about screening, biology, and treatment.* Curr Treat Options Oncol, 2017. **18**(4): p. 23.
165. Inra, J.A. and S. Syngal, *Colorectal cancer in young adults.* Dig Dis Sci, 2015. **60**(3): p. 722-33.
166. Mauri, G., et al., *Early-onset colorectal cancer in young individuals.* Mol Oncol, 2019. **13**(2): p. 109-131.
167. Patel, S.G. and D.J. Ahnen, *Colorectal cancer in the young.* Curr Gastroenterol Rep, 2018. **20**(4): p. 15.
168. Fearon, E.R., *Molecular genetics of colorectal cancer.* Annu Rev Pathol, 2011. **6**: p. 479-507.
169. Ma, H., et al., *Pathology and genetics of hereditary colorectal cancer.* Pathology, 2018. **50**(1): p. 49-59.
170. Ballinger, A.B. and C. Anggiansah, *Colorectal cancer.* BMJ, 2007. **335**(7622): p. 715-8.
171. Cunningham, D.A., W.; Lenz, H. J.; Lynch, H. T.; Minsk, B.; Nordlinger, B.; Starling, N., *Colorectal cancer.* Lancet, 2010. **375**: p. 1030-1047.
172. Baena Ruiz, R. and P. Salinas Hernandez, *Diet and cancer: risk factors and epidemiological evidence.* Maturitas, 2014. **77**(3): p. 202-8.
173. Baena, R. and P. Salinas, *Diet and colorectal cancer.* Maturitas, 2015. **80**(3): p. 258-64.
174. Bardou, M., A.N. Barkun, and M. Martel, *Obesity and colorectal cancer.* Gut, 2013. **62**(6): p. 933-47.
175. Mehta, R.S., et al., *Dietary Patterns and Risk of Colorectal Cancer: Analysis by Tumor Location and Molecular Subtypes.* Gastroenterology, 2017. **152**(8): p. 1944-1953 e1.
176. Bultman, S.J., *The microbiome and its potential as a cancer preventive intervention.* Semin Oncol, 2016. **43**(1): p. 97-106.
177. Gagniere, J., et al., *Gut microbiota imbalance and colorectal cancer.* World J Gastroenterol, 2016. **22**(2): p. 501-18.
178. Goodman, B. and H. Gardner, *The microbiome and cancer.* J Pathol, 2018. **244**(5): p. 667-676.
179. Raza, M.H., et al., *Microbiota in cancer development and treatment.* J Cancer Res Clin Oncol, 2019. **145**(1): p. 49-63.
180. Hansen, R.D., et al., *Effects of smoking and antioxidant micronutrients on risk of colorectal cancer.* Clin Gastroenterol Hepatol, 2013. **11**(4): p. 406-15 e3.
181. Leufkens, A.M., et al., *Cigarette smoking and colorectal cancer risk in the European Prospective Investigation into Cancer and Nutrition study.* Clin Gastroenterol Hepatol, 2011. **9**(2): p. 137-44.
182. Fedirko, V., et al., *Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies.* Ann Oncol, 2011. **22**(9): p. 1958-72.
183. Klarich, D.S., S.M. Brasser, and M.Y. Hong, *Moderate Alcohol Consumption and Colorectal Cancer Risk.* Alcohol Clin Exp Res, 2015. **39**(8): p. 1280-91.
184. Watson, A.J. and P.D. Collins, *Colon cancer: a civilization disorder.* Dig Dis, 2011. **29**(2): p. 222-8.
185. Schwabe, R.F. and C. Jobin, *The microbiome and cancer.* Nat Rev Cancer, 2013. **13**(11): p. 800-12.
186. Eaden, J.A.A., K.R; Mayberry, J.F., *The risk of colorectal cancer in ulcerative colitis a meta-analysis.* Gut, 2001. **48**: p. 526–535.
187. Irrazabal, T., et al., *The multifaceted role of the intestinal microbiota in colon cancer.* Mol Cell, 2014. **54**(2): p. 309-20.

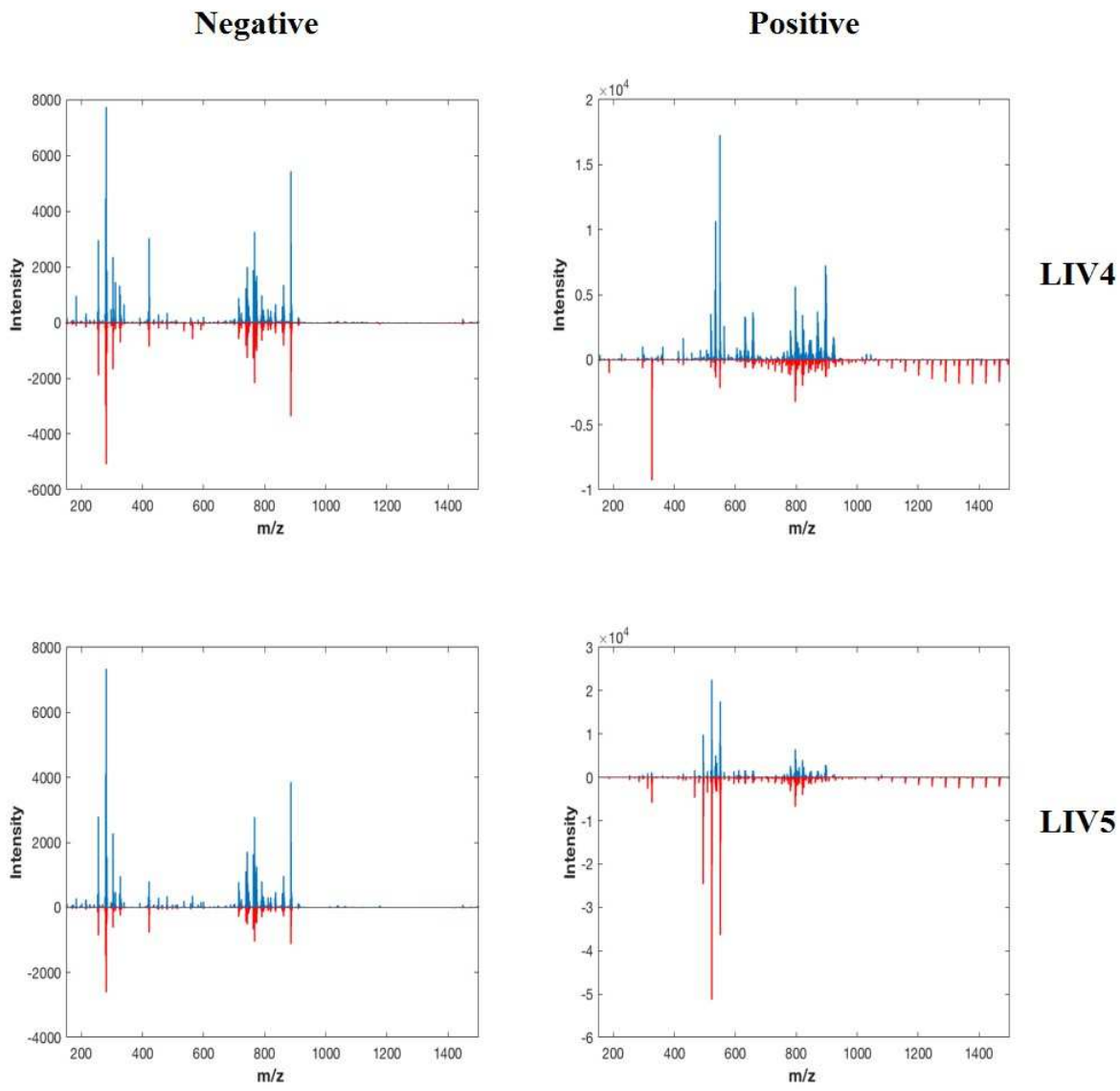
188. Kudo, S.K., H.; Nakajima, T.; Tamura, S.; Nakajo, K., *Endoscopic diagnosis and treatment of early colorectal cancer.* World J. Surg, 1997. **21**: p. 694-701.
189. Thrumurthy, S.G.T., S. S.; Gilbert, C. E.; Ross, P.; Haji, A., *Colorectal adenocarcinoma: risks, prevention and diagnosis.* BMJ, 2016. **354**: p. 1-12.
190. Aghagolzadeh, P. and R. Radpour, *New trends in molecular and cellular biomarker discovery for colorectal cancer.* World J Gastroenterol, 2016. **22**(25): p. 5678-93.
191. Cuyle, P.J. and H. Prenen, *Current and future biomarkers in the treatment of colorectal cancer.* Acta Clin Belg, 2017. **72**(2): p. 103-115.
192. Mahasneh, A., F. Al-Shaheri, and E. Jamal, *Molecular biomarkers for an early diagnosis, effective treatment and prognosis of colorectal cancer: Current updates.* Exp Mol Pathol, 2017. **102**(3): p. 475-483.
193. George, B. and S. Kopetz, *Predictive and prognostic markers in colorectal cancer.* Curr Oncol Rep, 2011. **13**(3): p. 206-15.
194. Luo, H.Y. and R.H. Xu, *Predictive and prognostic biomarkers with therapeutic targets in advanced colorectal cancer.* World J Gastroenterol, 2014. **20**(14): p. 3858-74.
195. Yiu, A.J.Y., C.Y., *Biomarkers in colorectal cancer.* Anticancer Research, 2016. **36**: p. 1093-1102.
196. Cawkwell, L.G., S.; Murgatroyd, H.; Sutherland, F.; Haine, L.; Longfellow, M.; O'Loughlin, S.; *et al.*, *Choice of management strategy for colorectalcancer based on a diagnostic immunohistochemicaltest for defective mismatch repair.* Gut, 1999. **45**: p. 409-415.
197. Norris, J.L. and R.M. Caprioli, *Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research.* Chem Rev, 2013. **113**(4): p. 2309-42.
198. Nilsson, A., et al., *Mass spectrometry imaging in drug development.* Anal Chem, 2015. **87**(3): p. 1437-55.
199. Ahram, M.F., M. J.; Gillespie, J. W.; Duray, P. H.; Linehan, W. M.; Ornstein, D. K.; Niu, S.; Zhao, Y.; Petricoin III, E. F.; Emmert-Buck, M. R., *Evaluation of ethanol-fixed, paraffin-embedded tissues for proteomic applications.* Proteomics, 2003. **3**: p. 413–421
200. Prieto, D.A.H., B. L.; Darfler, M. M.; Guiel, T. G.; Lucas, D. A.; Conrads, T. P.; Veenstra, T. D.; Krizman, D. B., *Liquid tissue: Proteomic profiling of formalin-fixed tissues.* BioTechniques, 2005. **38**: p. 32-35.
201. Ronci, M., et al., *Protein unlocking procedures of formalin-fixed paraffin-embedded tissues: application to MALDI-TOF imaging MS investigations.* Proteomics, 2008. **8**(18): p. 3702-14.
202. Heijs, B., et al., *Multimodal Mass Spectrometry Imaging of N-Glycans and Proteins from the Same Tissue Section.* Anal Chem, 2016. **88**(15): p. 7745-53.
203. Oetjen, J., et al., *An approach to optimize sample preparation for MALDI imaging MS of FFPE sections using fractional factorial design of experiments.* Anal Bioanal Chem, 2016. **408**(24): p. 6729-40.
204. Diehl, H.C.B., B.; Elm, D.; Trede, M.; Ahrens, M.; Eisenacher, M.; Marcus, K.; Meyer, H. E.; Henkel, C., *The challenge of on-tissue digestion for MALDI MSI- a comparison of different protocols to improve imaging experiments.* Anal Bioanal Chem, 2015. **407**: p. 2223–2243.
205. Beine, B., Diehl, H. C.; Meyer, H. E.; Henkel, C., *Tissue MALDI Mass Spectrometry Imaging (MALDI MSI) of Peptides* Proteomics in Systems Biology, 2015: p. 129-150.
206. Holst, S., et al., *Linkage-Specific in Situ Sialic Acid Derivatization for N-Glycan Mass Spectrometry Imaging of Formalin-Fixed Paraffin-Embedded Tissues.* Anal Chem, 2016. **88**(11): p. 5904-13.
207. Kelly, A.D., et al., *Metabolomic profiling from formalin-fixed, paraffin-embedded tumor tissue using targeted LC/MS/MS: application in sarcoma.* PLoS One, 2011. **6**(10): p. e25357.
208. Yuan, M., et al., *A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue.* Nat Protoc, 2012. **7**(5): p. 872-81.
209. Alashwal, H.E.H., M.; Crouse, J. J.; Abdalla, A.; Moustafa, A. A., *The Application of Unsupervised Clustering Methods to Alzheimer's Disease.* Frontiers in Computational Neuroscience, 2019. **13**: p. 1-9.

210. Abdi, H.,; Williams, L. J., *Principal component analysis.* Wiley Interdisciplinary Reviews: Computational Statistics, 2010. **2**(4): p. 433-459.
211. Julien, S., P.A. Videira, and P. Delannoy, *Sialyl-tn in cancer: (how) did we miss the target?* Biomolecules, 2012. **2**(4): p. 435-66.
212. Shen, Y.T., J.; Kohla, G.; Schauer, R., *Regulation of sialic acid O-acetylation in human colon mucosa.* Biological Chemistry, 2004. **385**: p. 145-152.
213. Kononen, J.B., L.; Kallioniemi, A.; Bärlund, M.; Schraml, P.; Leighton, S.; Torhorst, J.; Mihatsch, M. J.; Sauter, G.; Kallioniemi, O. P., *Tissue microarrays for high-throughput molecular profiling of tumor specimens.* Nature Medicine, 1998. **4**(7): p. 844-847.
214. Horvath, L. and S. Henshall, *The application of tissue microarrays to cancer research.* Pathology, 2001. **33**(2): p. 125-9.
215. Simon, R.M., M.; Sauter, G., *Tissue microarrays.* BioTechniques, 2004. **36**: p. 98-105.
216. Jawhar, N.M.T., *Tissue Microarray A rapidly evolving diagnostic and research tool.* Ann Saudi Med 2009. **29**(2): p. 123-127.
217. Thunig, J., S.H. Hansen, and C. Janfelt, *Analysis of secondary plant metabolites by indirect desorption electrospray ionization imaging mass spectrometry.* Anal Chem, 2011. **83**(9): p. 3256-9.
218. Inglese, P., et al., *Colocalization Features for Classification of Tumors Using Desorption Electrospray Ionization Mass Spectrometry Imaging.* Anal Chem, 2019. **91**(10): p. 6530-6540.
219. Kwak, J.T.H., S. M.; Sinha, S.; Bhargava, R., *Multimodal microscopy for automated histologic analysis of prostate cancer.* BMC Cancer, 2011. **11**(62): p. 1-16.
220. Yukihiro, Y. and N. Zaima, *Application of Mass Spectrometry Imaging for Visualizing Food Components.* Foods, 2020. **9**(5).
221. Renshaw, S., *'Immunochemical staining techniques', Immunohistochemistry: Methods Express.* Bloxham, UK: Scion Publishing Ltd, 2007: p. 45-96.
222. Hughes, C.G., L.; Brown, M.; Clarke, N. W.; Gardner, P., *Assessment of paraffin removal from prostate FFPE sections using transmission mode FTIR-FPA imaging.* Anal. Methods, 2014. **6**(4): p. 1028-1035.
223. Aerni, H.R., D.S. Cornett, and R.M. Caprioli, *High-throughput profiling of formalin-fixed paraffin-embedded tissue using parallel electrophoresis and matrix-assisted laser desorption ionization mass spectrometry.* Anal Chem, 2009. **81**(17): p. 7490-5.
224. Tsochatzis, E.A.B., J.; Burroughs, A. K., *Liver cirrhosis.* The Lancet, 2014. **383**(9930): p. 1749-1761.
225. Schuppan, D.A., N. H. , *Liver cirrhosis.* Lancet, 2008. **371**(9615): p. 838–851.
226. Wanless, I.R.W., F.; Blendis, L. M.; Greig, P.; Heathcote, E. J.; Levy, G., *Hepatic and portal vein thrombosis in cirrhosis: possible role in development of parenchymal extinction and portal hypertension.* Hepatology, 1995. **21**(5): p. 1238-1247.
227. Boyd, A.C., O.; Chauhan, A.; Webb, G.J., *Medical liver biopsy: background, indications, procedure and histopathology.* Frontline Gastroenterology, 2020. **11**: p. 40-47.
228. Nalbantoglu, I.L. and E.M. Brunt, *Role of liver biopsy in nonalcoholic fatty liver disease.* World J Gastroenterol, 2014. **20**(27): p. 9026-37.
229. Silveira, M.G.L., K.D., *Overlap syndromes with autoimmune hepatitis in chronic cholestatic liver diseases.* Expert Review of Gastroenterology & Hepatology, 2007. **1**(232): p. 329-340.
230. Chazouillères, O.W., D. Serfaty, L.; Montembault, S.; Rosmorduc, O.; Poupon, R., *Primary biliary cirrhosis-autoimmune hepatitis overlap syndrome: clinical features and response to therapy* Hepatology, 1998. **28**: p. 296-301.
231. Czaja, A.J., *Frequency and nature of the variant syndromes of autoimmune liver disease.* Hepatology, 1998. **28**: p. 360-365.
232. Boberg, K.M., et al., *Overlap syndromes: the International Autoimmune Hepatitis Group (IAIHG) position statement on a controversial issue.* J Hepatol, 2011. **54**(2): p. 374-85.
233. Beuers, U., *Hepatic overlap syndromes.* J Hepatol, 2005. **42 Suppl**(1): p. S93-9.
234. Durazzo, M., et al., *Overlap syndromes of autoimmune hepatitis: an open question.* Dig Dis Sci, 2013. **58**(2): p. 344-8.
235. Perdigoto, R.C., H. A.; Czaja, A. J., *Frequency and significance of chronic ulcerative colitis in severe corticosteroid-treated autoimmune hepatitis* J Hepatol, 1992. **14**: p. 325-331.

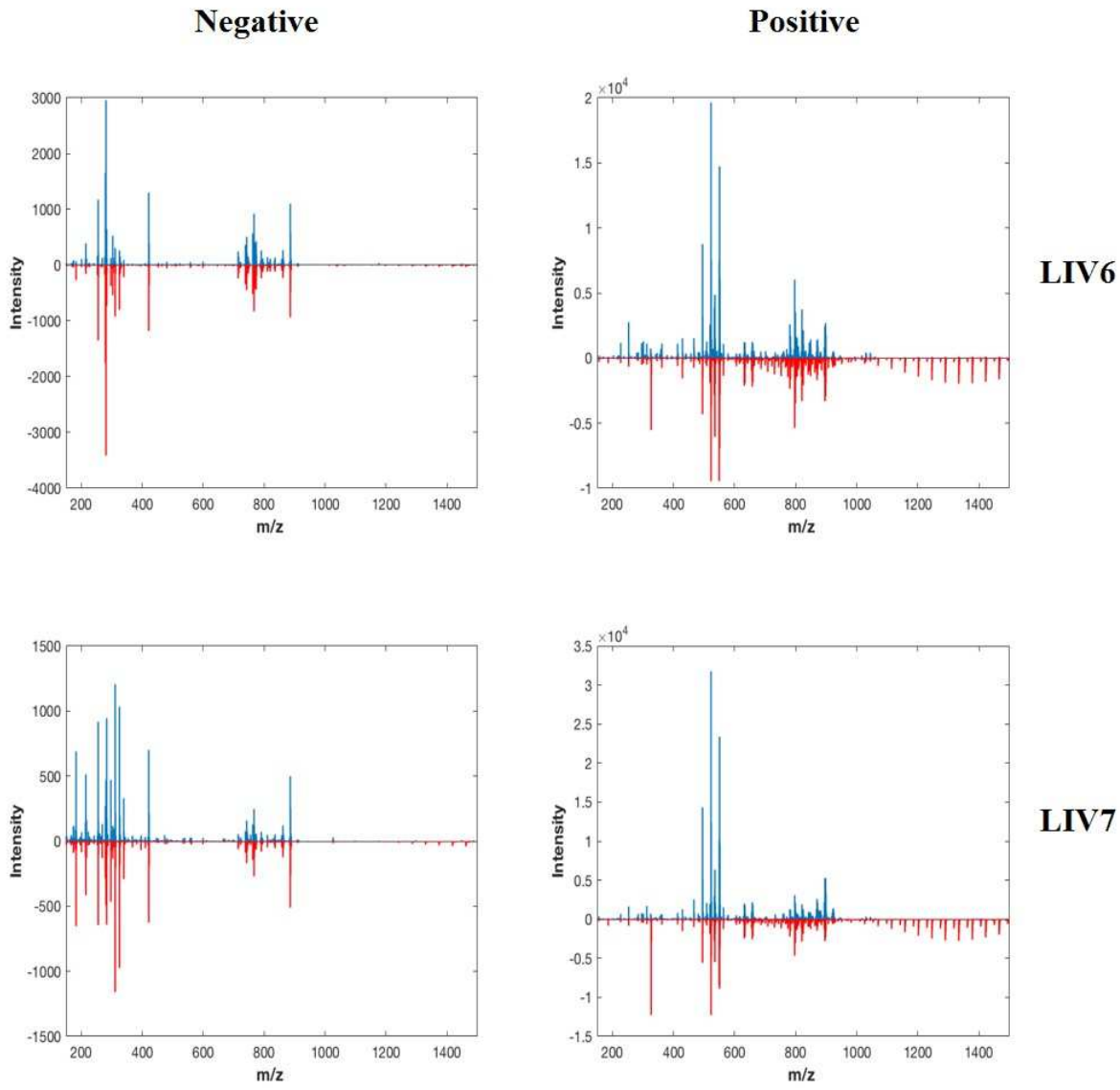
236. Abdalian, R.D., P.; Jhaveri, K.; Haider, M.; Guindi, M.; Heathcote, J. E., *Prevalence of sclerosing cholangitis in adults with autoimmune hepatitis: evaluating the role of routine magnetic resonance imaging.* Hepatology, 2008. **47**(3): p. 949-957.
237. Czaja, A.J., *The overlap syndromes of autoimmune hepatitis.* Dig Dis Sci, 2013. **58**(2): p. 326-43.
238. Wiesner, R.H.L., N. F.; Ludwig, J.; Dickson, E. R., *Comparison of the clinicopathologic features of primary sclerosing cholangitis and primary biliary cirrhosis* Gastroenterology, 1985. **88**: p. 108-114.
239. Scupakova, K., et al., *Spatial Systems Lipidomics Reveals Nonalcoholic Fatty Liver Disease Heterogeneity.* Anal Chem, 2018. **90**(8): p. 5130-5138.
240. Babyak, M.A., *What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models.* Psychosomatic Medicine, 2004. **66**(3): p. 411-421.
241. Inglese, P., et al., *SPUTNIK: an R package for filtering of spatially related peaks in mass spectrometry imaging data.* Bioinformatics, 2019. **35**(1): p. 178-180.
242. Lakhani, S.R.A., A., *Microarray and histopathological analysis of tumours: the future and the past?* Nature Reviews Cancer, 2001. **1**(2): p. 151-157.
243. Schwamborn, K., *The importance of histology and pathology in mass spectrometry imaging.* Adv Cancer Res, 2017. **134**: p. 1-26.
244. Dill, A.L.E., L. S.; Zheng, C.; Costa, A. B.; Ifa, D. R.; Zheng, L.; Masterson, T. A.; Koch, M. O.; Vitek, O.; Cooks, R. G., *Multivariate statistical differentiation of renal cell carcinomas based on lipidomic analysis by ambient ionization imaging mass spectrometry.* Anal Bioanal Chem, 2010. **398**: p. 2969 - 2978.
245. Dobrzynska, I.S.-P., B.; Sulkowski, S.; Figaszewski, Z., *Changes in electric charge and phospholipids composition in human colorectal cancer cells.* Molecular and Cellular Biochemistry, 2005. **276**(1): p. 113-119.
246. Eberlin, L.S., et al., *Discrimination of human astrocytoma subtypes by lipid analysis using desorption electrospray ionization imaging mass spectrometry.* Angew Chem Int Ed Engl, 2010. **49**(34): p. 5953-6.
247. Morita, Y., et al., *Imaging mass spectrometry of gastric carcinoma in formalin-fixed paraffin-embedded tissue microarray.* Cancer Sci, 2010. **101**(1): p. 267-73.
248. Goh, W.W.B., W. Wang, and L. Wong, *Why Batch Effects Matter in Omics Data, and How to Avoid Them.* Trends Biotechnol, 2017. **35**(6): p. 498-507.
249. Balluff, B., et al., *Batch Effects in MALDI Mass Spectrometry Imaging.* J Am Soc Mass Spectrom, 2021. **32**(3): p. 628-635.
250. Balog, J.S., T.; Schaefer, K. C.; Denes, J.; Lopata, A.; Godorhazy, L.; Szalay, D.; Balogh, L.; Sasi-Szabo, L.; Toth, M.; Takats, Z., *Identification of biological tissues by Rapid Evaporative Ionization Mass Spectrometry.* Anal Chem, 2010. **82**: p. 7343–7350.
251. Balog, J.S.-S., L.; Kinross, J.; Lewis, M. R.; Muirhead, L. J.; Veselkov, K.; Mirnezami, R.; Dezso, B.; Damjanovich, L.; Darzi, A.; Nicholson, J. K.; Takats, Z., *Intraoperative tissue identification using Rapid Evaporative Ionization Mass Spectrometry.* Sci Transl Med, 2013. **5**(194): p. 1-11.
252. Phelps, D.L., et al., *The surgical intelligent knife distinguishes normal, borderline and malignant gynaecological tissues using rapid evaporative ionisation mass spectrometry (REIMS).* Br J Cancer, 2018. **118**(10): p. 1349-1358.
253. St John, E.R., et al., *Intraoperative tissue identification by mass spectrometric technologies.* TrAC Trends in Analytical Chemistry, 2016. **85**: p. 2-9.
254. Takats, Z., N. Strittmatter, and J.S. McKenzie, *Ambient mass spectrometry in cancer research.* Adv Cancer Res, 2017. **134**: p. 231-256.
255. Golf, O.M., L. J.; Speller, A.; Balog, J.; Abbassi-Ghadi, N.; Kumar, S.; Mroz, A.; Veselkov, K.; Takats, Z., *XMS: Cross-Platform Normalization Method for Multimodal Mass Spectrometric Tissue Profiling.* J. Am. Soc. Mass Spectrom., 2015. **26**: p. 44-54.

Supplementary Information

Appendix A: Average spectra for each of the other 4 liver sections, showing negative and positive ion modes. H_2O-embedded spectra are shown in blue, and those from OCT-embedded samples are in red.



Appendix A: Average spectra for each of the other 4 liver sections, showing negative and positive ion modes. H_2O-embedded spectra are shown in blue, and those from OCT-embedded samples are in red – continued.

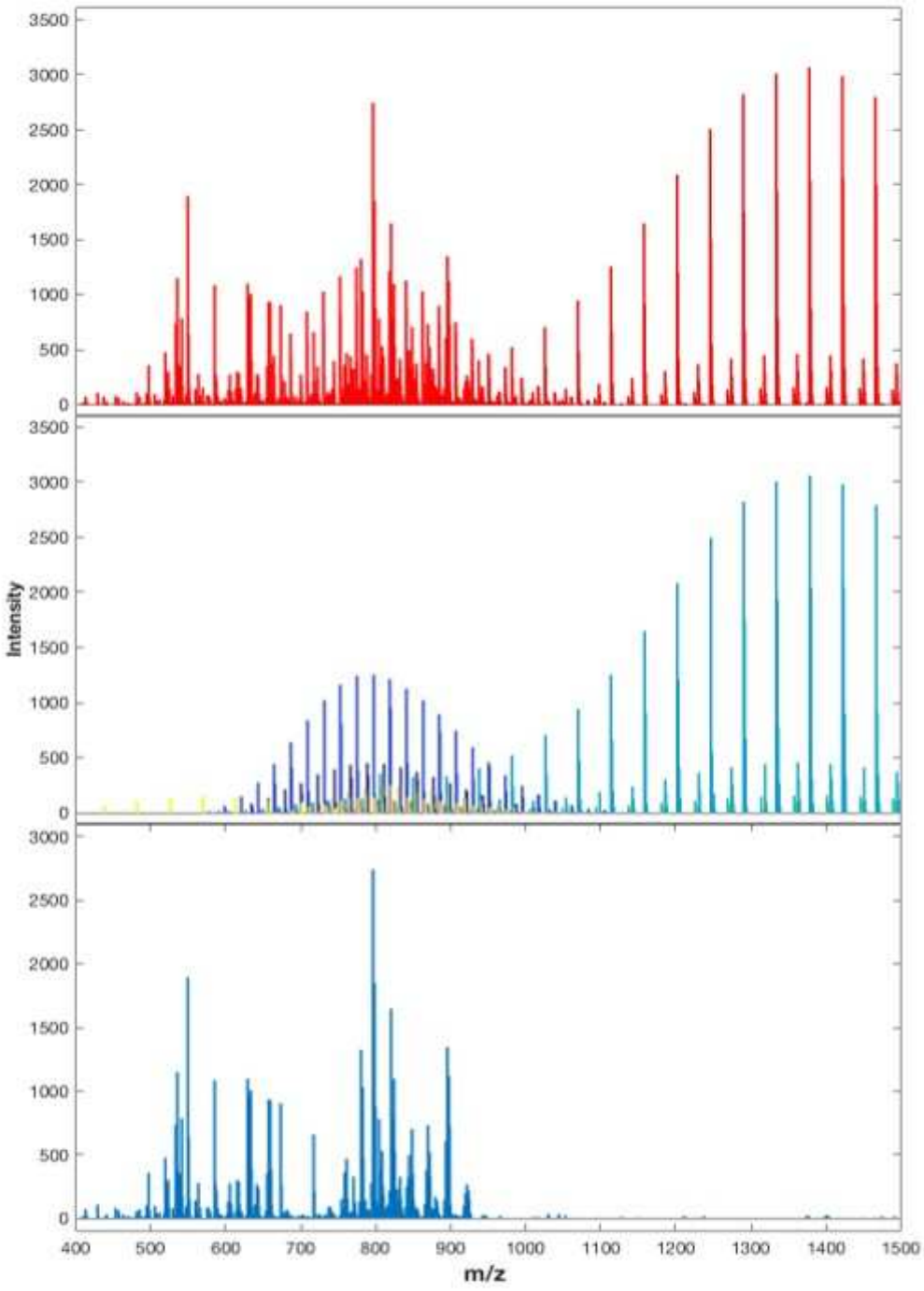


Appendix B: Information about each of the polymeric distributions identified in the 5 positive mode OCT-embedded samples. The polymers in OCT have a monomeric unit of C_2H_4O with a mass of 44.026 Da. The 13 distributions identified in the 5 positive mode OCT-embedded samples are shown below. The charge of each distribution was determined by inspection of the M+1 isotopologue.
A linear series of *m/z* values (±44.026/z) centred around the identified *m/z* value was created and matched (±10 ppm) against the mean spectrum for each sample. From the most intense peak within the distribution, only those peaks found in a continuous distribution were considered, i.e. there are no gaps in the polymer peaks identified in the spectrum. The isotopologue peaks for each of these were also identified based on the distance expected due to the charge state.

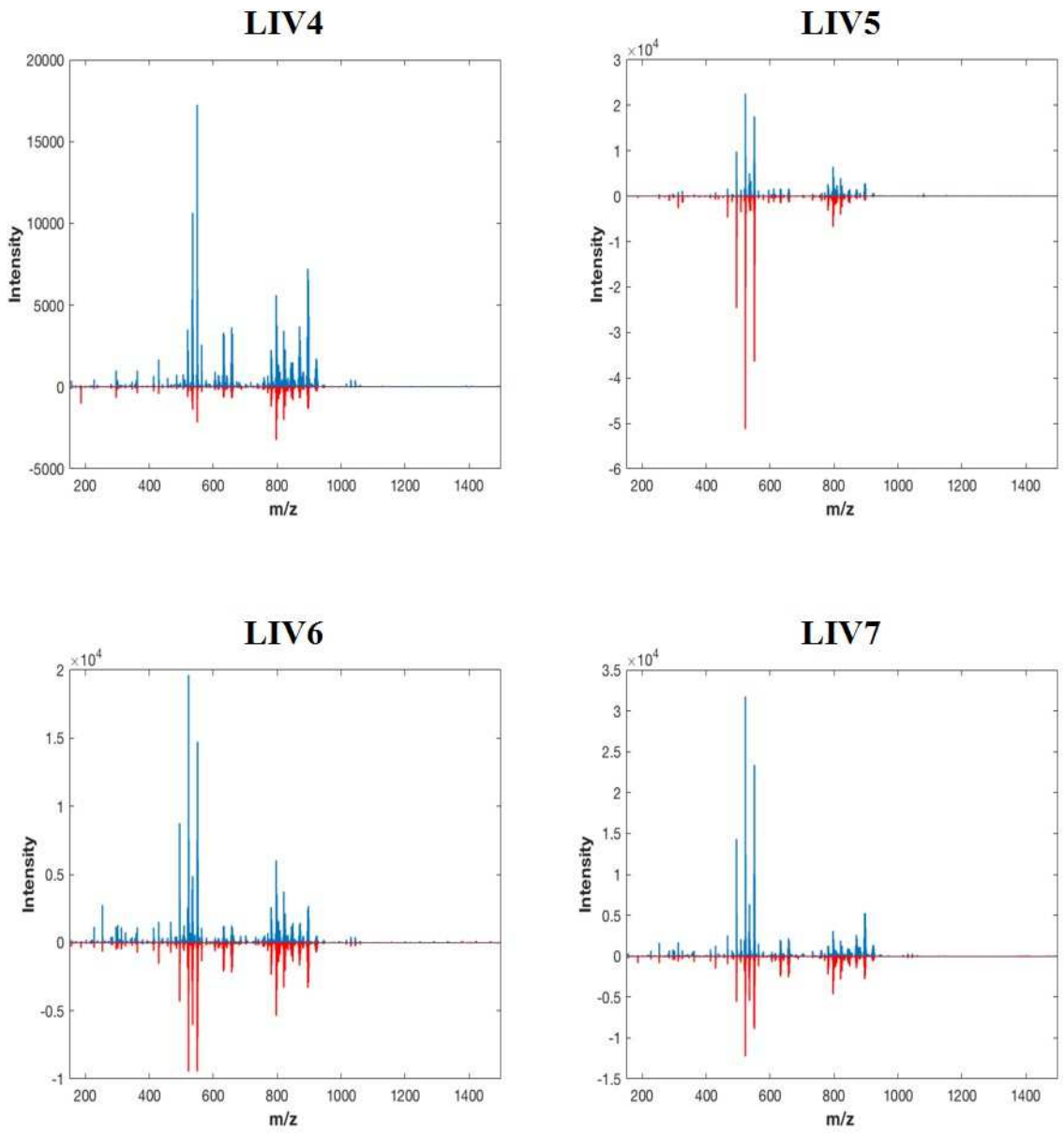
m	*z*
569.313	1
586.657	3
700.375	2
708.360	2
780.439	2
829.974	2
841.524	2
844.000	2
1289.705	1
1333.733	1
1361.784	1
1400.854	1
1489.068	1

Appendix C: The three panels in the plot show, from top to bottom: the mean spectrum of the raw data; the peaks identified as polymeric; the mean spectrum with the polymeric peaks removed.

Appendix C shows the removal of the various polymeric peaks from the positive mode LIV3 sample that was embedded with OCT. Average spectra for the 4 other OCT-embedded positive mode samples, with polymeric profiles removed, are shown below (Appendix D). The comparison was made between the paired H_2O-embedded liver section. For all but LIV5, the average signal intensities of the OCT-embedded samples are much lower than for the H_20-embedded samples.

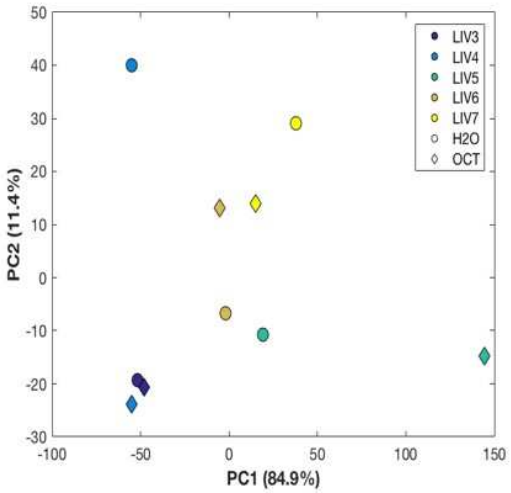


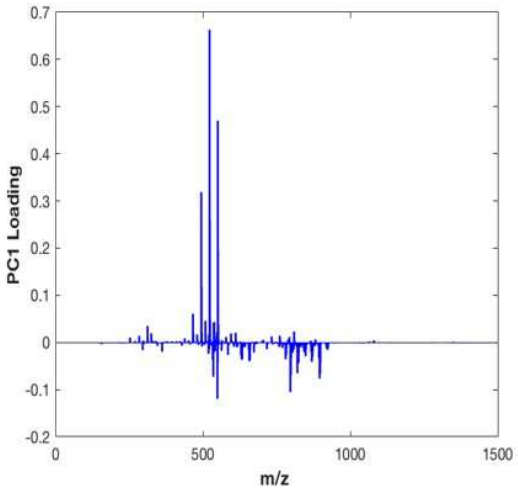
Appendix D: Average spectra, following polymeric signal removal, of the remaining liver sections. The H_2O-embedded mean spectrum is shown in blue, and the OCT-embedded equivalent is shown in red.



Appendix E: PCA scores and loadings plots for positive mode data.

The loadings of PC1 using positive mode data revealed the presence of 4 peak clusters with elevated intensities in 3 of the 5 liver samples (regardless of embedding medium). The PCA scores and loadings plots revealed these clusters with *m/z* ratios of 466.534, 494.566, 522.597 and 550.628. The most intense of these (522.597) was likely to be a surfactant with the chemical formula of $C_{36}H_{75}N$, with the other 3 clusters differing by $\pm[C_2H_4]n$. These peak clusters were subsequently removed from all positive mode spectra.

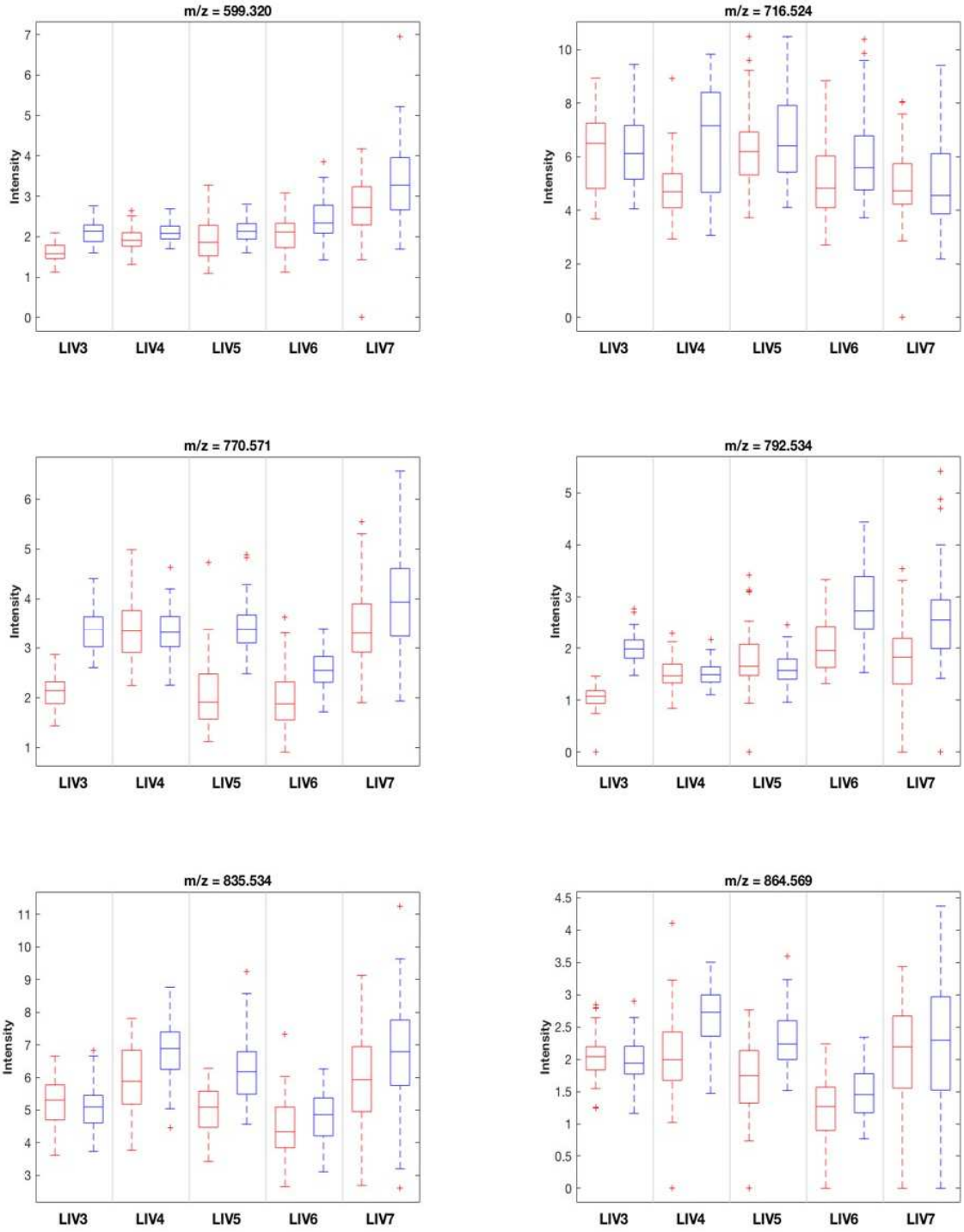




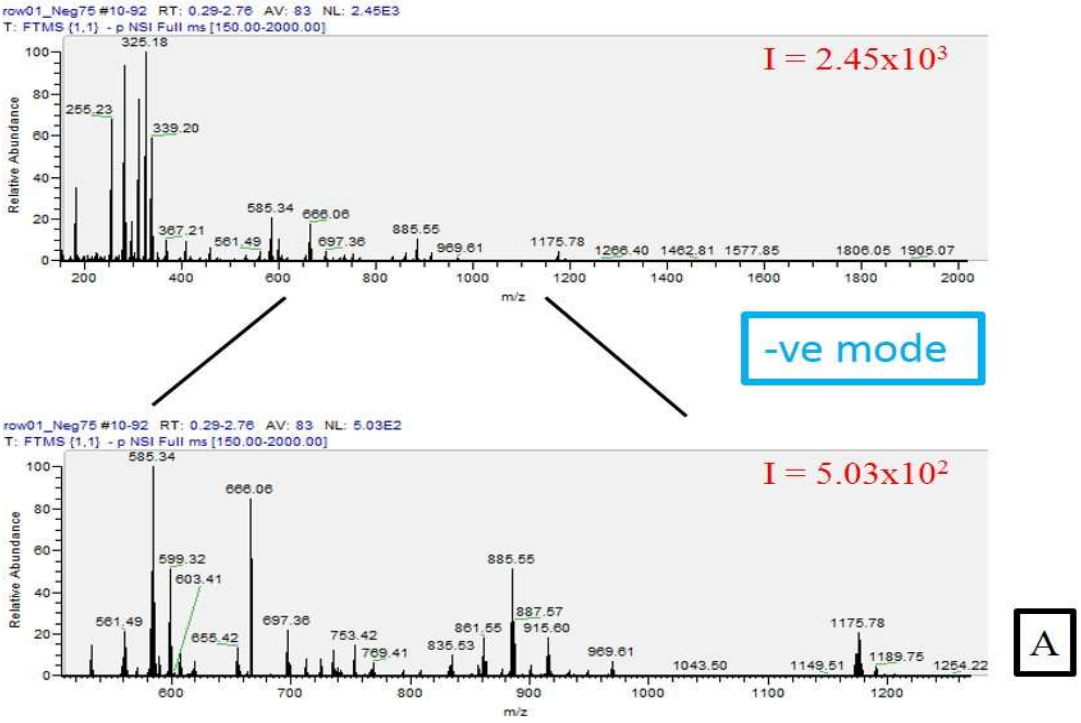
Appendix F: Boxplots showing ions at lower intensities in OCT-embedded samples. OCT intensities are depicted in red, and H_2O intensities are shown in blue.

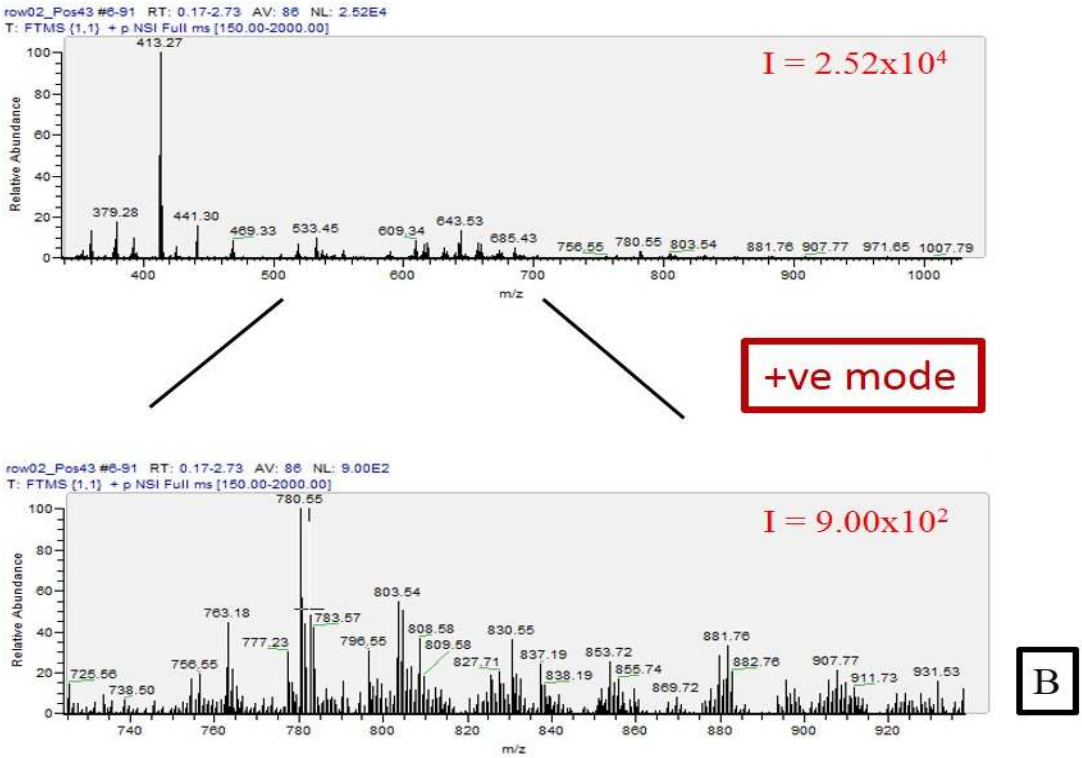
Boxplots below show the intensities of specific ions, which were at lower intensities in the OCT-embedded spectra compared to the H_2O-embedded spectra. All variables have $q < 0.001$ and $\log_2$ H_2O/OCT fold changes > 0. Few of these variables demonstrated a consistent trend across the 5 paired sections.

Of the 6 boxplots shown below, only that with *m/z* 599.320 shows lower median intensities across all 5 OCT-embedded sections compared to H_20-embedding. The remaining sections showed some sections with reduced OCT intensities and others with elevated intensities.



Appendix G. Averaged spectra recorded for the 5µm thick FFPE mouse liver sample. Zoom in on the phospholipids mass range is also shown. -ve mode data (A); +ve mode data (B).





Appendix H: TNM staging classification of the colon and rectal cancer.

AJCC stage	TNM stage	TNM stage criteria for colorectal cancer	Dukes
Stage 0	Tis N0 M0	Tis: Tumour confined to mucosa; cancer-*in-situ*	-
Stage I	T1 N0 M0	T1: Tumour invades *submucosa*	A
Stage I	T2 N0 M0	T2: Tumour invades *muscularis propria*	A
Stage II - A	T3 N0 M0	T3: Tumour invades *subserosa* or beyond (without other organs involved)	B
Stage II - B	T4 N0 M0	T4: Tumour invades adjacent organs or perforates the visceral peritoneum	B
Stage III - A	T1-2 N1 M0	N1: Metastasis to 1 to 3 regional lymph nodes. T1 or T2.	C
Stage III - B	T3-4 N1 M0	N1: Metastasis to 1 to 3 regional lymph nodes. T3 or T4.	C
Stage III - C	any T, N2 M0	N2: Metastasis to 4 or more regional lymph nodes. Any T.	C
Stage IV	any T, any N, M1	M1: Distant metastases present. Any T, any N.	-

Appendix I: Significant features detected for the different tissue types (mucin, mucosa, muscle and tumour) identified in a human colorectal FFPE sample analysed by DESI-MSI in negative ion mode. All ions have ANOVA Benjamini-Hochberg-Yekutieli-corrected $p < 0.001$ and a mean intensity threshold of 10.

			Mean Intensity			
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucin**	**Mucosa**	**Muscle**	**Tumour**
261.11	Mucin	<0.001	49.6392	32.6435	29.6322	8.7663
262.06	Mucin	<0.001	18.1007	0.3108	0.0000	0.0000
290.06	Mucin	<0.001	27.4443	0.7274	0.0000	0.0000
308.07	Mucin	<0.001	20.2305	0.3169	0.0000	0.0000
332.08	Mucin	<0.001	18.2404	0.5235	0.0000	0.0000
346.10	Mucin	2.22E-15	10.0384	2.5882	0.0000	0.0000
350.11	Mucin	<0.001	56.0621	3.2286	0.4959	0.3535
374.10	Mucin	<0.001	15.6223	7.6469	0.0000	0.0000
392.12	Mucin	<0.001	74.2800	14.7128	0.0000	0.0000
410.17	Mucin	<0.001	20.6282	0.6827	1.1400	0.6607
434.14	Mucin	<0.001	84.9303	64.2900	0.0000	0.0000
452.15	Mucin	<0.001	26.9229	0.1921	0.0000	0.0000
468.20	Mucin	<0.001	15.0576	2.4336	0.6318	8.1888
470.20	Mucin	<0.001	10.6845	1.5376	0.2763	5.5811
494.18	Mucin	<0.001	41.8908	4.1135	0.0000	0.0000
509.25	Mucin	<0.001	29.0356	5.1881	1.9678	5.9776
510.21	Mucin	<0.001	10.3986	0.3967	0.0000	0.3604
511.26	Mucin	<0.001	29.5445	2.2259	0.6841	3.4775
512.20	Mucin	<0.001	12.8239	0.4009	0.0000	0.2154
533.22	Mucin	<0.001	17.2730	0.2001	0.0000	0.2048
535.26	Mucin	<0.001	33.0786	5.6667	0.4104	0.5325
536.20	Mucin	<0.001	38.1305	11.7273	0.0000	0.0000
537.42	Mucin	3.46E-06	13.1670	8.2928	2.5291	5.3688
551.22	Mucin	<0.001	45.5694	0.2162	0.0000	0.4320
552.21	Mucin	<0.001	14.3132	1.9877	0.0000	0.0000
553.25	Mucin	<0.001	48.9778	2.1710	1.6604	0.2244
554.23	Mucin	<0.001	16.3659	3.8881	0.2384	0.3906
575.24	Mucin	<0.001	22.2097	5.4860	0.0000	0.0000
577.25	Mucin	<0.001	49.8710	30.4554	0.0000	1.7674
593.26	Mucin	<0.001	65.5335	5.0029	0.5575	0.1880
594.24	Mucin	<0.001	21.1493	5.6570	0.0000	0.1700
595.27	Mucin	<0.001	74.5279	5.4468	0.2896	0.1920
596.26	Mucin	<0.001	18.9341	6.5879	0.0000	0.0000
608.21	Mucin	<0.001	10.0952	1.2666	0.0000	0.3213
617.25	Mucin	4.08E-13	15.1051	12.2399	0.2581	0.0000
623.28	Mucin	<0.001	10.5642	0.4017	0.0000	0.2140
635.27	Mucin	<0.001	45.6562	12.1460	0.4898	0.2496

			Mean Intensity			
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucin**	**Mucosa**	**Muscle**	**Tumour**
637.28	Mucin	<0.001	60.5516	12.2584	0.0000	0.2070
650.24	Mucin	<0.001	28.9000	0.3257	0.0000	0.0000
655.30	Mucin	<0.001	14.8722	0.4019	0.0000	0.0000
679.32	Mucin	<0.001	18.5642	6.0281	0.0000	0.0000
715.35	Mucin	<0.001	11.6294	0.4224	0.0000	0.4820
756.38	Mucin	<0.001	11.4788	0.5940	0.0000	0.0000
763.73	Mucin	<0.001	28.2928	13.4069	4.3143	0.9056
793.35	Mucin	<0.001	16.2688	0.0000	0.0000	0.4069
796.39	Mucin	2.14E-14	10.8581	3.8743	0.0000	0.1925
798.41	Mucin	<0.001	22.0366	9.6287	0.0000	0.0000
829.86	Mucin	<0.001	13.6411	0.8463	0.0000	0.0000
831.88	Mucin	<0.001	12.1671	0.3463	0.0000	0.0000
855.89	Mucin	<0.001	26.2023	1.1143	0.0000	0.0000
857.91	Mucin	<0.001	43.6223	1.7922	0.4834	0.3711
881.91	Mucin	<0.001	11.9385	0.3996	0.0000	0.0000
883.93	Mucin	<0.001	24.6605	0.5072	0.0000	0.0000
171.05	Mucosa	<0.001	20.7370	83.8191	64.4134	68.6883
187.05	Mucosa	1.12E-12	33.7141	42.2529	24.8077	24.4132
199.09	Mucosa	2.37E-09	3.9920	15.1525	9.4201	12.6066
214.01	Mucosa	8.11E-08	10.9706	15.0913	7.5984	13.9348
215.09	Mucosa	3.89E-10	11.9224	23.2959	12.4378	14.5806
234.12	Mucosa	<0.001	5.3572	51.1614	34.1025	35.0106
237.08	Mucosa	4.77E-15	1.1034	14.8536	7.8138	12.6848
251.12	Mucosa	7.45E-14	8.1095	29.0719	21.4556	27.6355
253.18	Mucosa	<0.001	96.6506	327.1840	120.3742	259.7761
255.19	Mucosa	<0.001	965.2612	2641.6609	1332.3804	2288.2710
255.24	Mucosa	<0.001	0.4373	12.3837	0.0000	4.5978
267.20	Mucosa	<0.001	15.9716	54.8567	23.5786	46.6682
269.00	Mucosa	1.74E-08	7.8799	13.2313	4.7319	10.9055
269.18	Mucosa	<0.001	10.4062	34.8015	9.8648	18.6559
269.21	Mucosa	<0.001	35.3463	103.0920	47.0651	84.7898
271.20	Mucosa	<0.001	7.2245	24.2620	7.8569	13.5727
277.19	Mucosa	<0.001	17.9389	43.8409	15.0523	35.0046
279.20	Mucosa	<0.001	357.0331	1154.3913	382.4971	939.5096
281.21	Mucosa	<0.001	2813.9493	8454.7073	3410.8704	5988.5829
281.26	Mucosa	<0.001	4.8754	95.9261	9.5069	43.4436
281.35	Mucosa	<0.001	0.9227	30.4091	1.0457	10.7206
282.22	Mucosa	<0.001	555.1404	1754.2433	685.5910	1237.8379
283.23	Mucosa	<0.001	640.2842	1227.9497	724.4103	1046.8690
293.18	Mucosa	<0.001	29.4666	60.9281	14.3177	30.2557
294.22	Mucosa	<0.001	2.8900	14.6925	0.2754	11.0915
295.20	Mucosa	<0.001	103.8801	288.1704	75.3792	150.5407
295.24	Mucosa	<0.001	7.2117	31.7659	11.4320	19.8868

			Mean Intensity			
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucin**	**Mucosa**	**Muscle**	**Tumour**
297.21	Mucosa	<0.001	157.6942	495.6380	141.7942	227.3538
299.24	Mucosa	<0.001	6.3959	22.1960	5.3660	11.1008
303.21	Mucosa	<0.001	13.8877	54.7680	22.1001	47.7418
305.22	Mucosa	<0.001	12.4700	44.3242	19.1506	41.9861
307.24	Mucosa	<0.001	15.4943	49.3584	20.2829	36.8434
309.18	Mucosa	<0.001	14.4861	26.1235	8.0767	13.2693
309.25	Mucosa	<0.001	59.4476	195.9591	68.1830	123.2435
311.20	Mucosa	<0.001	43.1289	84.1923	20.8605	34.7262
311.27	Mucosa	<0.001	21.4937	39.3963	20.3373	26.8775
323.24	Mucosa	<0.001	2.1249	12.8500	2.3537	6.5448
325.25	Mucosa	<0.001	2.6996	14.2089	1.5817	7.1814
327.24	Mucosa	<0.001	22.0396	44.3912	19.1436	41.8819
329.24	Mucosa	<0.001	22.9078	53.6712	15.5879	34.4359
331.24	Mucosa	<0.001	10.3959	27.5611	11.5834	24.2653
337.29	Mucosa	<0.001	4.4569	22.4074	7.9000	11.7023
349.22	Mucosa	4.92E-10	68.7986	134.2598	125.2112	120.9186
357.21	Mucosa	1.02E-05	3.4227	11.3137	5.6215	10.6764
363.24	Mucosa	4.81E-09	17.4649	32.9271	25.6732	28.5140
365.26	Mucosa	1.75E-08	15.4930	28.7931	21.2385	25.6764
366.22	Mucosa	1.29E-14	48.4542	76.9615	45.4825	71.0759
379.27	Mucosa	7.68E-10	6.9105	14.3455	5.9249	12.5401
389.09	Mucosa	9.88E-15	19.4337	54.3404	28.3879	51.3070
415.27	Mucosa	<0.001	15.8807	41.3307	16.0666	5.8632
437.30	Mucosa	1.90E-14	3.6763	16.6335	5.8841	7.1562
471.36	Mucosa	<0.001	5.4801	14.9578	4.0517	0.1714
476.16	Mucosa	<0.001	33.8218	68.2179	0.0000	0.0000
476.30	Mucosa	<0.001	3.1897	37.3479	0.3270	3.2697
477.42	Mucosa	<0.001	6.1197	24.2693	4.4121	12.8361
478.32	Mucosa	<0.001	16.2634	66.4183	1.3891	23.7075
479.39	Mucosa	2.02E-10	7.4699	28.8170	9.2029	12.3732
480.33	Mucosa	<0.001	1.8623	27.2947	0.5981	2.1642
485.38	Mucosa	<0.001	4.5611	18.5618	10.2902	0.3675
497.38	Mucosa	<0.001	4.1666	14.5059	2.8952	0.3767
499.39	Mucosa	<0.001	41.6937	101.0125	31.2551	5.0053
500.31	Mucosa	<0.001	20.0551	69.0505	15.5497	21.7180
501.42	Mucosa	<0.001	6.8277	28.6854	11.2846	12.1142
508.34	Mucosa	<0.001	1.2699	14.6335	0.9573	1.4546
522.37	Mucosa	<0.001	1.7887	13.4447	0.3040	3.7108
523.40	Mucosa	<0.001	7.2824	16.3374	2.2427	0.5239
524.34	Mucosa	<0.001	7.8652	27.0066	6.8645	9.7310
525.42	Mucosa	<0.001	38.9692	94.6635	26.1222	4.0130
526.33	Mucosa	7.77E-16	3.5582	12.2853	1.5915	5.2001
533.49	Mucosa	<0.001	14.2884	65.9176	26.6598	56.6914

			Mean Intensity			
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucin**	**Mucosa**	**Muscle**	**Tumour**
539.44	Mucosa	<0.001	5.6213	18.3934	5.5834	0.6049
541.42	Mucosa	4.04E-14	3.9301	15.2109	6.7591	3.6295
553.46	Mucosa	2.66E-15	3.3652	11.3613	1.7554	0.1897
555.48	Mucosa	4.69E-14	2.1975	13.7432	2.8704	5.0776
557.50	Mucosa	<0.001	12.0066	78.1760	22.3569	62.8556
559.52	Mucosa	<0.001	64.6509	328.8802	113.7260	243.4245
565.46	Mucosa	1.11E-16	9.3620	21.7113	5.3637	7.3982
571.53	Mucosa	<0.001	11.3691	32.9844	4.3879	18.2623
573.54	Mucosa	<0.001	3.2314	27.3071	4.3570	19.4514
575.54	Mucosa	<0.001	8.4412	19.9815	3.6247	12.0891
581.51	Mucosa	<0.001	0.0000	10.5135	0.5428	7.1545
583.52	Mucosa	<0.001	21.4511	124.1522	27.4835	83.9191
584.53	Mucosa	<0.001	4.9380	45.5852	10.7599	32.4382
585.54	Mucosa	<0.001	93.5042	498.0412	133.7663	297.4372
587.56	Mucosa	<0.001	37.8869	151.1592	49.7465	102.6002
589.58	Mucosa	1.22E-15	8.8626	21.0122	11.0186	15.8336
591.43	Mucosa	<0.001	4.1780	21.6867	5.0051	16.5712
597.37	Mucosa	1.42E-13	7.9098	16.1773	4.3907	14.4350
599.38	Mucosa	<0.001	65.1101	162.9795	50.7805	87.2919
599.57	Mucosa	<0.001	3.3158	34.2755	3.2518	16.5641
600.39	Mucosa	<0.001	22.1010	48.1535	14.6680	25.7137
601.55	Mucosa	<0.001	11.1660	50.9645	9.8502	24.3909
607.44	Mucosa	<0.001	7.3029	32.9813	9.4521	28.8467
609.58	Mucosa	4.44E-16	0.5048	10.9977	2.2524	5.5067
611.59	Mucosa	<0.001	2.6177	17.6598	6.7600	8.2064
613.51	Mucosa	5.91E-07	2.7586	11.1640	3.9848	9.6495
613.59	Mucosa	<0.001	3.0726	21.3230	3.9770	10.7032
615.53	Mucosa	<0.001	2.2833	13.0278	1.0483	9.6560
617.54	Mucosa	<0.001	0.9928	11.0344	0.4756	4.4155
619.27	Mucosa	<0.001	32.3848	75.1322	0.0000	0.0000
619.47	Mucosa	7.05E-08	5.9923	10.6497	2.1248	8.2138
625.58	Mucosa	<0.001	2.9684	21.3315	5.7633	7.9569
627.60	Mucosa	1.67E-15	3.0809	21.0808	15.3916	18.4307
635.60	Mucosa	5.55E-16	0.9005	10.1388	1.5789	4.1633
637.62	Mucosa	5.55E-16	2.9363	14.5153	8.2879	3.3057
639.57	Mucosa	<0.001	3.5957	14.6951	3.8541	0.6506
641.58	Mucosa	<0.001	4.0998	22.5924	11.6597	10.0905
651.60	Mucosa	<0.001	5.4338	28.6601	9.0555	14.0916
653.55	Mucosa	<0.001	13.0950	40.0580	19.2132	17.1470
653.62	Mucosa	1.89E-15	4.6698	14.9543	5.1438	4.8169
654.56	Mucosa	<0.001	1.8242	16.5620	6.0239	6.3619
655.58	Mucosa	<0.001	6.2731	31.2269	14.5833	9.8876
667.61	Mucosa	<0.001	11.2565	31.0186	17.8866	12.2258

			Mean Intensity			
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucin**	**Mucosa**	**Muscle**	**Tumour**
669.55	Mucosa	<0.001	19.8618	57.1365	22.1539	11.4256
671.57	Mucosa	4.20E-11	0.5636	10.5117	3.7501	7.2242
681.62	Mucosa	<0.001	5.8018	16.4022	2.7518	1.9238
683.43	Mucosa	1.11E-13	12.4065	23.1261	7.1704	19.7045
683.61	Mucosa	<0.001	5.6541	19.1726	5.8909	5.8520
691.55	Mucosa	1.75E-11	1.7322	13.1012	2.9248	4.7272
697.45	Mucosa	<0.001	34.9672	68.1407	17.6882	48.7779
697.60	Mucosa	1.11E-16	6.0076	22.9671	7.1878	9.9633
699.61	Mucosa	<0.001	2.0622	10.9939	0.4797	3.3629
701.62	Mucosa	<0.001	11.2016	29.6575	4.6401	11.7572
713.46	Mucosa	<0.001	13.3543	34.4365	8.1393	30.4012
715.65	Mucosa	1.40E-05	2.6503	12.2091	8.7162	6.3132
716.66	Mucosa	2.47E-13	0.6853	13.9690	13.0842	5.4426
737.70	Mucosa	<0.001	14.2990	23.0523	6.5910	0.9162
742.65	Mucosa	6.42E-13	0.9142	14.1006	0.0000	2.4912
753.54	Mucosa	<0.001	19.5411	50.7872	5.3679	42.9982
765.75	Mucosa	<0.001	7.4104	14.1194	1.4166	1.1376
788.68	Mucosa	4.44E-16	14.8678	34.8362	17.4794	11.7501
789.70	Mucosa	3.65E-11	7.3427	17.9947	7.2090	6.3374
829.76	Mucosa	1.11E-16	1.0652	10.1197	8.8823	1.5393
833.75	Mucosa	7.40E-09	13.7543	18.9769	5.5632	11.9404
835.68	Mucosa	<0.001	35.5249	53.4038	4.2137	27.8072
837.84	Mucosa	<0.001	1.1178	18.2709	5.5171	15.6208
838.42	Mucosa	1.08E-11	10.9957	14.6336	0.0000	0.0000
840.43	Mucosa	1.50E-12	19.3361	25.3375	0.0000	0.9322
857.67	Mucosa	4.74E-12	7.4384	10.8432	0.3217	4.9514
859.69	Mucosa	5.13E-11	9.3901	10.2375	0.5545	6.5946
861.72	Mucosa	<0.001	37.1048	52.3319	4.0718	25.9020
861.86	Mucosa	<0.001	2.5247	17.5993	4.0635	14.5425
863.73	Mucosa	<0.001	36.0447	57.0051	6.1385	20.1274
863.87	Mucosa	<0.001	4.1143	52.6035	17.3436	37.4886
883.70	Mucosa	1.52E-13	13.9848	18.6666	1.4028	12.5369
885.71	Mucosa	<0.001	99.3745	124.4126	25.4397	41.9162
887.88	Mucosa	<0.001	0.8020	18.2639	2.6657	12.3377
889.90	Mucosa	<0.001	4.9857	53.5006	13.3272	31.4064
905.92	Mucosa	<0.001	2.1923	13.0605	1.9333	3.8102
911.76	Mucosa	3.21E-11	3.3564	11.7706	1.0012	7.5388
931.94	Mucosa	7.87E-10	1.5689	11.3929	5.0096	5.5604
933.77	Mucosa	1.44E-07	7.0417	13.7891	5.6807	6.3803
949.73	Mucosa	1.56E-13	5.6518	12.1561	2.2333	3.2931
193.04	Muscle	3.00E-15	3.1283	16.7794	28.7307	6.1892
323.20	Muscle	2.62E-10	25.5466	48.9429	60.5363	53.7645
329.17	Muscle	3.89E-06	2.1115	9.2834	13.4467	12.8899

			Mean Intensity			
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucin**	**Mucosa**	**Muscle**	**Tumour**
417.29	Muscle	1.17E-11	9.5631	23.5996	24.2373	12.7222
427.30	Muscle	<0.001	1.1729	4.4963	24.6641	8.2271
451.36	Muscle	7.21E-13	4.3755	19.6634	31.9467	28.7697
452.32	Muscle	1.12E-14	0.6104	14.6243	20.7457	14.9073
665.60	Muscle	7.77E-16	4.7120	15.8218	22.0434	6.7370
691.64	Muscle	<0.001	0.4390	7.9962	37.5879	2.2068
693.64	Muscle	<0.001	0.0000	9.7650	52.4946	1.0081
695.65	Muscle	1.80E-07	8.9041	14.1354	17.1584	5.9796
717.65	Muscle	<0.001	1.7982	28.0009	50.8552	13.7299
719.67	Muscle	<0.001	0.0000	22.0769	84.4764	7.6814
733.66	Muscle	2.28E-09	0.0000	9.6823	16.5949	7.5631
777.73	Muscle	5.23E-09	2.6028	8.2824	10.6612	0.8885
787.69	Muscle	<0.001	1.2085	2.9170	13.4209	0.9185
802.71	Muscle	2.63E-14	0.6213	6.3520	13.0757	2.0694
803.74	Muscle	<0.001	4.8460	17.2666	18.5133	1.6093
805.76	Muscle	1.67E-15	3.3180	11.8685	18.0948	4.1691
831.78	Muscle	<0.001	2.5965	5.9083	16.4947	1.0754
995.98	Muscle	1.11E-15	0.0000	8.0260	16.3280	1.7696
997.00	Muscle	2.86E-08	1.1608	5.3630	11.1047	1.4779
999.01	Muscle	2.22E-16	1.2301	3.5352	10.8100	0.4113
107.99	Tumour	2.08E-13	2.0708	12.7660	16.6131	21.3201
120.97	Tumour	<0.001	16.3474	44.8724	68.9917	81.7916
123.95	Tumour	<0.001	14.5966	18.3736	10.3937	35.2461
133.99	Tumour	<0.001	1.3001	15.8246	18.6971	61.8238
136.98	Tumour	<0.001	2.1748	9.5640	10.1495	24.9910
137.96	Tumour	<0.001	30.9291	88.1579	117.1125	148.0447
143.99	Tumour	1.15E-07	1.4879	7.2625	5.3216	11.0161
144.91	Tumour	<0.001	3.3288	11.4367	4.0522	30.6426
146.91	Tumour	<0.001	14.6345	13.9043	18.5522	32.0973
150.99	Tumour	2.20E-10	4.3730	9.7202	12.6455	15.7784
151.98	Tumour	<0.001	2.1266	18.8036	21.4426	25.2086
153.00	Tumour	1.85E-06	0.5166	6.4772	4.9162	11.0248
157.03	Tumour	8.87E-10	4.3066	12.9420	11.8705	15.9802
161.01	Tumour	<0.001	0.5294	4.2712	15.4614	23.0567
165.96	Tumour	<0.001	0.8294	8.4738	18.7426	20.9791
166.00	Tumour	2.03E-13	0.0000	10.4511	11.3971	13.9056
171.09	Tumour	6.44E-10	21.4959	32.3624	42.2416	45.7950
173.95	Tumour	<0.001	9.2583	11.5278	16.1414	28.4333
177.04	Tumour	3.55E-14	0.5052	8.3792	6.8496	13.5851
180.02	Tumour	4.10E-14	8.0402	14.9707	16.9759	23.7013
182.95	Tumour	<0.001	30.7884	85.0504	120.3443	212.2578
185.11	Tumour	2.33E-11	6.3119	13.1312	16.1520	19.9799
188.01	Tumour	<0.001	0.0000	0.9143	1.0102	10.8861

			Mean Intensity			
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucin**	**Mucosa**	**Muscle**	**Tumour**
188.90	Tumour	<0.001	9.9594	28.0751	21.2000	83.6333
189.03	Tumour	<0.001	6.0370	5.7915	8.8349	32.6086
190.90	Tumour	<0.001	5.1012	20.4753	15.9133	56.6382
191.02	Tumour	<0.001	0.0000	0.5983	0.0000	10.9830
192.90	Tumour	<0.001	0.4612	2.2252	0.5720	10.5758
196.97	Tumour	<0.001	3.4359	16.0304	25.2880	40.1085
199.12	Tumour	<0.001	20.2499	39.6783	42.8201	73.9717
200.13	Tumour	<0.001	0.0000	3.8689	2.7120	10.6313
200.99	Tumour	<0.001	25.9771	31.2687	31.9776	71.2820
213.14	Tumour	<0.001	6.4939	15.8471	18.4999	38.7649
222.98	Tumour	<0.001	10.7947	21.5028	26.5022	33.0810
223.06	Tumour	1.58E-09	2.9849	11.7199	6.5358	13.1121
225.04	Tumour	6.19E-09	2.1748	7.1133	8.4431	15.0175
227.02	Tumour	1.85E-06	3.4070	4.4211	6.2250	10.1234
227.16	Tumour	<0.001	37.2034	109.0372	60.0828	123.0558
233.11	Tumour	<0.001	24.2926	72.9371	43.3213	88.6588
235.13	Tumour	6.07E-14	0.0000	11.2193	5.9836	11.2508
235.98	Tumour	<0.001	0.0000	1.8877	1.8703	10.7474
239.05	Tumour	<0.001	11.4393	27.4575	33.1744	46.0537
239.09	Tumour	3.05E-14	2.0925	6.2794	9.6190	13.3225
241.18	Tumour	<0.001	24.1805	59.0552	39.0962	73.8801
246.87	Tumour	3.50E-12	0.0000	3.9893	5.6120	10.5018
248.13	Tumour	<0.001	0.6503	10.4571	12.2188	18.7561
249.00	Tumour	<0.001	7.5174	9.8917	7.0870	50.5243
250.11	Tumour	<0.001	32.0714	74.7505	74.5324	101.6549
253.07	Tumour	4.28E-06	2.0355	7.9845	8.7834	12.0291
256.90	Tumour	<0.001	2.0473	6.2339	6.2033	16.8165
258.90	Tumour	4.44E-16	0.0000	3.9623	3.9550	11.1728
265.11	Tumour	<0.001	24.5436	33.0967	31.4019	118.0745
266.12	Tumour	1.33E-15	0.9794	3.6386	1.8955	15.6481
267.13	Tumour	2.02E-08	2.2463	8.7453	6.8718	16.3433
273.90	Tumour	<0.001	2.8785	9.8416	7.0117	35.6297
275.89	Tumour	<0.001	1.2462	6.4193	1.8496	21.9074
283.13	Tumour	2.81E-10	1.9894	9.7162	6.2030	14.8704
293.15	Tumour	5.34E-08	8.7431	6.8894	4.9402	24.5316
297.14	Tumour	<0.001	20.5701	18.4653	16.2953	46.6507
299.20	Tumour	4.88E-10	6.4124	12.0527	4.7468	20.1547
300.94	Tumour	<0.001	5.2205	12.7357	9.4824	55.9851
302.94	Tumour	<0.001	2.2574	8.4507	6.2164	38.1112
303.16	Tumour	1.20E-14	1.0238	9.3609	2.4370	13.5374
309.15	Tumour	1.28E-07	4.2427	6.3507	5.0040	24.5132
311.14	Tumour	4.35E-12	58.7001	60.8920	46.7269	159.3133
312.16	Tumour	5.41E-11	13.3459	12.6462	8.0110	31.1661

			Mean Intensity			
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucin**	**Mucosa**	**Muscle**	**Tumour**
313.17	Tumour	7.92E-12	9.7101	14.2063	11.4279	29.0202
315.17	Tumour	9.21E-15	3.4542	7.2923	3.2465	15.8776
321.19	Tumour	4.03E-07	1.8793	4.7913	3.5981	10.0624
325.16	Tumour	4.01E-10	94.2639	86.6160	71.2032	197.9468
337.22	Tumour	8.13E-06	8.5225	18.9234	17.6451	23.0480
339.18	Tumour	1.49E-08	53.3103	50.6702	47.4940	89.9682
340.19	Tumour	4.22E-12	28.2326	36.0031	29.0730	47.4547
341.21	Tumour	1.97E-10	22.1196	34.2044	23.6810	39.9894
343.20	Tumour	1.17E-12	14.4568	40.2466	24.1023	41.2438
352.24	Tumour	1.49E-07	4.3121	11.6067	9.9112	12.8572
364.20	Tumour	1.41E-13	3.6136	9.9673	1.8093	10.2440
367.24	Tumour	<0.001	38.4929	55.3324	42.3900	70.8608
377.26	Tumour	1.11E-16	4.4970	17.2400	3.6906	19.4981
381.27	Tumour	1.11E-07	11.7663	14.2864	4.7783	16.7145
384.20	Tumour	<0.001	0.0000	3.9450	0.4252	13.8573
391.28	Tumour	7.69E-13	10.6390	19.1078	10.4796	24.6353
393.26	Tumour	<0.001	61.6246	92.8869	51.6284	115.5951
395.28	Tumour	2.46E-10	19.5045	21.2432	17.3129	28.5215
397.24	Tumour	6.33E-07	4.2639	4.9835	1.6390	13.3670
400.89	Tumour	<0.001	0.0000	1.7110	0.0000	23.0143
402.89	Tumour	<0.001	0.0000	3.4883	0.5416	28.3538
404.89	Tumour	<0.001	0.0000	0.6454	0.0000	13.7683
405.26	Tumour	4.63E-06	4.1692	9.9237	4.6543	10.1602
407.28	Tumour	4.26E-06	3.8226	7.0379	4.7185	10.3331
409.31	Tumour	1.15E-06	5.7011	10.4454	3.5242	11.3614
419.28	Tumour	9.44E-09	5.6798	16.7063	10.0310	19.4099
419.33	Tumour	7.77E-15	0.5551	7.5976	4.8884	13.7181
421.23	Tumour	<0.001	2.2210	2.8473	1.7183	12.8740
421.34	Tumour	3.86E-13	2.1216	7.5218	3.7691	14.0534
423.20	Tumour	1.76E-13	1.2819	7.4596	5.9353	13.8816
425.31	Tumour	<0.001	3.7735	13.1587	22.1512	27.8318
428.25	Tumour	<0.001	0.6304	7.3841	2.8480	21.8204
429.30	Tumour	2.00E-15	2.2786	11.8784	15.7437	19.5207
433.34	Tumour	<0.001	1.6360	10.2954	5.4223	15.7472
435.36	Tumour	4.39E-11	2.2101	14.8340	7.0053	16.8050
439.30	Tumour	1.17E-05	3.0116	9.3799	5.8972	11.8870
441.30	Tumour	5.64E-11	1.2949	3.8696	1.8486	12.3397
447.36	Tumour	<0.001	2.3665	11.7118	4.5999	18.9851
449.37	Tumour	9.99E-16	3.7284	14.2440	5.1554	17.7015
455.34	Tumour	<0.001	0.4787	18.7062	35.0899	38.9537
461.38	Tumour	<0.001	3.6584	17.6440	13.0260	28.6657
463.39	Tumour	3.05E-12	1.7134	7.0838	2.0522	10.7023
467.34	Tumour	<0.001	6.0239	18.6416	11.5027	41.9147

			Mean Intensity			
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucin**	**Mucosa**	**Muscle**	**Tumour**
469.22	Tumour	<0.001	2.3510	10.5841	5.1112	23.6610
475.41	Tumour	<0.001	7.0563	23.6776	11.8615	30.6830
485.10	Tumour	5.02E-13	0.5452	2.3420	6.6624	10.7126
489.42	Tumour	2.89E-15	2.5710	8.8749	3.4626	15.6657
491.21	Tumour	9.67E-11	11.9130	6.7416	5.1607	13.8131
493.22	Tumour	<0.001	33.0243	44.1398	13.6021	81.8253
495.25	Tumour	<0.001	17.7395	36.1171	15.3036	64.3277
497.26	Tumour	<0.001	1.1470	9.3176	1.4713	20.0450
503.44	Tumour	<0.001	5.0176	11.3305	4.0613	20.4479
505.45	Tumour	<0.001	0.0000	7.4538	1.1258	12.3286
517.46	Tumour	<0.001	2.8338	7.5279	1.6798	13.5905
531.48	Tumour	<0.001	8.3783	30.8879	11.7917	33.2074
545.50	Tumour	<0.001	3.4668	12.7407	3.5767	16.5040
583.38	Tumour	<0.001	3.2194	23.0968	28.9477	43.1608
585.39	Tumour	<0.001	22.3115	86.3138	120.9338	331.9856
586.26	Tumour	<0.001	0.0000	2.1502	0.7866	10.3041
586.40	Tumour	<0.001	5.4807	28.4442	44.2110	122.7781
587.41	Tumour	<0.001	2.2529	11.2969	10.3515	28.3678
601.41	Tumour	<0.001	12.4840	34.1274	16.8850	41.3195
615.42	Tumour	5.55E-16	1.5243	9.5740	6.0409	16.5078
629.43	Tumour	<0.001	2.7783	16.1950	10.7987	26.5287
664.51	Tumour	4.37E-11	1.8680	10.9086	3.8265	12.5513
699.47	Tumour	1.44E-09	7.0852	10.1171	0.8347	10.4059
711.48	Tumour	2.66E-15	4.5989	10.9128	1.4365	15.8572
714.46	Tumour	3.33E-16	3.0610	11.0807	1.0694	11.4779
725.50	Tumour	<0.001	18.4963	37.2563	4.2487	40.3553
727.62	Tumour	<0.001	4.5964	14.4444	2.2478	17.3977
735.45	Tumour	<0.001	0.4785	12.5115	5.7290	17.0056
739.52	Tumour	<0.001	10.8856	24.5390	0.6521	25.4074
741.53	Tumour	<0.001	5.9324	22.3579	2.6024	26.3171
755.56	Tumour	<0.001	2.8160	15.5479	1.3428	16.3764
763.49	Tumour	<0.001	1.6588	7.4147	0.5868	13.4163
769.54	Tumour	<0.001	8.7007	28.7520	3.0259	30.9160
770.54	Tumour	<0.001	1.3317	10.8863	0.0000	11.9782
771.60	Tumour	<0.001	1.1960	9.4088	1.2942	13.3466
781.63	Tumour	3.95E-14	2.2855	9.8381	0.2166	11.6567
791.53	Tumour	1.11E-16	1.0843	10.5462	1.3594	15.7011
797.58	Tumour	6.33E-15	1.1982	8.3295	2.2655	13.9364

Appendix J: Significant features detected for the different tissue types (mucosa, muscle and tumour) present in a human colorectal FFPE sample analysed by DESI-MSI in positive ion mode. All ions have ANOVA Benjamini-Hochberg-Yekutieli-corrected $p < 0.001$ and a mean intensity threshold of 10.

			Mean Intensity		
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucosa**	**Muscle**	**Tumour**
129.03	Mucosa	<0.001	96.5310	94.5009	74.1204
134.99	Mucosa	3.58E-13	13.6244	8.5223	9.2772
151.11	Mucosa	7.37E-05	16.9944	12.4333	16.1137
155.11	Mucosa	<0.001	1243.0841	821.9628	471.8559
156.12	Mucosa	<0.001	118.6423	76.5337	43.0947
164.91	Mucosa	7.01E-13	13.5674	9.5101	7.9080
173.07	Mucosa	<0.001	788.3578	626.2211	600.8798
177.08	Mucosa	4.85E-07	15.0210	10.9995	13.7932
181.08	Mucosa	7.53E-12	24.4279	24.2328	18.6053
191.10	Mucosa	1.32E-05	75.9828	75.2665	67.4742
217.11	Mucosa	<0.001	1238.0599	941.4287	926.5596
217.16	Mucosa	5.68E-13	18.7419	9.2875	16.5293
219.14	Mucosa	1.14E-06	18.3002	14.3448	13.3354
223.14	Mucosa	8.96E-07	47.5054	47.3125	39.9247
231.13	Mucosa	5.11E-15	61.8190	48.5335	47.3156
233.17	Mucosa	9.88E-15	20.7362	19.7991	13.2892
234.94	Mucosa	<0.001	10.9301	10.2137	4.4079
245.11	Mucosa	2.35E-09	66.8677	55.7950	59.3041
297.27	Mucosa	1.07E-05	43.6905	41.6270	36.7663
299.26	Mucosa	2.58E-11	36.8533	28.7215	29.9033
301.11	Mucosa	<0.001	41.8816	14.6671	7.8816
313.28	Mucosa	1.80E-11	86.5049	69.2275	73.9301
325.27	Mucosa	<0.001	351.1586	289.7278	226.7982
325.32	Mucosa	7.17E-06	11.4340	7.2130	10.1040
326.28	Mucosa	<0.001	64.9301	51.8285	37.0020
329.29	Mucosa	7.74E-07	45.4712	38.4242	44.7362
333.21	Mucosa	<0.001	73.6947	43.3757	42.4282
334.20	Mucosa	<0.001	10.4931	3.0513	3.0688
337.16	Mucosa	<0.001	51.4469	33.4194	41.3911
343.27	Mucosa	2.25E-06	129.9917	122.0842	114.5953
351.26	Mucosa	2.84E-10	16.0827	8.6124	8.6691
355.34	Mucosa	6.66E-09	51.2673	41.3549	43.5186
357.32	Mucosa	<0.001	15.7703	6.0877	11.2565
360.42	Mucosa	<0.001	13.8798	4.7720	6.1370
363.28	Mucosa	2.08E-12	11.3232	4.0858	5.6905
365.20	Mucosa	2.45E-08	42.7523	34.2185	39.7815
375.31	Mucosa	<0.001	94.7102	89.3500	65.4780
376.32	Mucosa	1.11E-16	13.8455	13.2123	5.5031

			Mean Intensity		
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucosa**	**Muscle**	**Tumour**
387.31	Mucosa	1.91E-10	33.2241	23.7383	27.6251
391.32	Mucosa	<0.001	95.9292	89.8397	56.9621
392.36	Mucosa	7.02E-13	18.8738	17.0011	9.3969
394.38	Mucosa	1.08E-14	148.7453	124.0462	132.3320
401.34	Mucosa	<0.001	37.9989	27.9304	20.0983
403.36	Mucosa	<0.001	15.5190	12.3838	3.2796
413.35	Mucosa	<0.001	1543.9173	971.2763	1422.4869
425.38	Mucosa	<0.001	85.9441	44.8957	70.8266
427.37	Mucosa	6.84E-14	74.9778	66.8662	52.0673
429.41	Mucosa	1.33E-14	27.1098	17.1696	14.2693
453.35	Mucosa	7.05E-05	10.8629	5.5804	6.2480
467.22	Mucosa	4.32E-06	194.1853	177.2392	180.5576
469.43	Mucosa	3.85E-13	148.6337	115.5462	139.7444
475.44	Mucosa	7.11E-07	27.6672	20.3055	24.4293
481.38	Mucosa	6.06E-08	22.3815	15.5631	15.6746
487.23	Mucosa	3.11E-10	11.7533	8.2320	11.7441
500.40	Mucosa	<0.001	13.7109	0.3375	9.5167
501.42	Mucosa	<0.001	11.9454	3.3471	1.1083
507.38	Mucosa	<0.001	20.9668	11.3407	18.2676
513.47	Mucosa	<0.001	104.4613	68.2533	75.7589
517.51	Mucosa	2.68E-07	10.8974	5.6590	9.4371
522.25	Mucosa	<0.001	49.6094	0.0845	18.4487
522.39	Mucosa	<0.001	11.4882	1.2679	9.4708
523.27	Mucosa	<0.001	13.2015	0.8547	4.6707
555.33	Mucosa	5.08E-14	14.7239	10.1946	14.6647
557.32	Mucosa	7.72E-12	17.2660	10.6710	13.9412
557.52	Mucosa	<0.001	241.8381	152.2428	184.6095
595.59	Mucosa	4.29E-07	114.6206	102.3475	105.5221
596.60	Mucosa	8.30E-06	47.0746	41.3131	39.3579
601.56	Mucosa	<0.001	369.2567	215.4552	291.5509
604.65	Mucosa	1.11E-15	14.4083	5.2619	2.6657
607.51	Mucosa	<0.001	43.7505	23.9539	20.5099
608.50	Mucosa	<0.001	31.5644	7.3725	7.6206
609.51	Mucosa	<0.001	973.4119	255.7398	201.2460
615.33	Mucosa	4.02E-05	58.8143	53.4689	58.1347
617.32	Mucosa	8.80E-05	30.6490	27.1075	29.8406
623.51	Mucosa	<0.001	41.2854	18.4717	16.8176
625.52	Mucosa	<0.001	97.5977	38.3164	41.5784
629.38	Mucosa	<0.001	26.4294	13.3875	20.3015
631.41	Mucosa	<0.001	18.7521	8.0980	15.0880
645.61	Mucosa	<0.001	432.8940	234.6000	337.0116
657.59	Mucosa	2.21E-11	14.7609	6.2291	6.4290
661.60	Mucosa	2.09E-10	38.3686	28.6583	35.6863

			Mean Intensity		
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucosa**	**Muscle**	**Tumour**
665.38	Mucosa	<0.001	14.7389	0.0000	1.8726
689.66	Mucosa	<0.001	401.5676	218.3764	317.0176
699.72	Mucosa	<0.001	62.9545	41.0705	62.2475
703.47	Mucosa	<0.001	48.3879	23.6225	45.1892
704.44	Mucosa	<0.001	32.8400	15.3540	29.9626
705.65	Mucosa	9.49E-12	37.9617	25.8950	29.6724
733.71	Mucosa	<0.001	288.5857	142.2296	220.2000
749.70	Mucosa	3.03E-11	30.5087	21.8012	21.5251
775.75	Mucosa	<0.001	12.5195	4.7454	11.5555
777.49	Mucosa	<0.001	25.0470	7.2745	20.4457
777.76	Mucosa	<0.001	179.0262	90.3928	131.1411
779.77	Mucosa	1.81E-07	20.2104	15.9551	13.7776
793.75	Mucosa	8.40E-06	20.7766	16.1169	14.3122
805.84	Mucosa	4.22E-15	11.3687	4.3080	10.9938
821.81	Mucosa	<0.001	95.9879	48.2173	71.8193
865.86	Mucosa	<0.001	47.3257	24.6975	35.5635
880.06	Mucosa	1.03E-10	18.1513	13.4944	10.2341
882.07	Mucosa	2.83E-08	21.6951	19.1752	14.4627
908.11	Mucosa	1.66E-07	12.5123	9.6802	6.4549
104.08	Muscle	<0.001	29.2308	58.7161	39.7071
111.09	Muscle	1.13E-11	24.5212	28.2767	22.7673
114.04	Muscle	<0.001	7.7089	28.7916	0.6498
114.06	Muscle	2.18E-10	8.6729	10.9407	7.6072
136.06	Muscle	<0.001	42.9787	118.1483	16.8193
149.01	Muscle	5.71E-08	19.5759	22.4512	17.8217
157.07	Muscle	<0.001	115.4353	141.3258	106.6851
158.08	Muscle	1.26E-07	14.4529	16.5274	12.6733
163.08	Muscle	9.89E-09	14.7635	14.9901	11.2765
165.08	Muscle	1.63E-10	19.4020	21.2646	15.8327
169.97	Muscle	<0.001	11.8681	17.8311	12.9928
171.10	Muscle	1.51E-10	533.9635	543.6932	469.8940
178.15	Muscle	<0.001	19.0991	26.7098	14.1515
182.18	Muscle	3.50E-09	37.2378	46.9135	43.2188
183.09	Muscle	5.74E-12	35.5522	42.4085	32.8195
185.11	Muscle	1.14E-12	87.7831	98.4655	83.6845
187.09	Muscle	<0.001	32.2874	45.9108	29.6087
191.03	Muscle	3.87E-14	8.0263	10.7254	4.6121
192.17	Muscle	<0.001	8.9797	14.9743	9.1950
195.15	Muscle	4.45E-11	51.3616	58.4802	47.1478
197.08	Muscle	<0.001	30.6515	40.5936	25.8535
197.11	Muscle	<0.001	13.7553	17.1318	10.5749
199.09	Muscle	6.87E-11	30.0694	35.2976	27.6145
201.12	Muscle	<0.001	37.6380	48.4641	34.9258

			Mean Intensity		
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucosa**	**Muscle**	**Tumour**
202.18	Muscle	<0.001	10.6147	13.7528	4.5467
203.10	Muscle	5.61E-10	15.4868	18.1101	11.6438
205.09	Muscle	1.92E-07	46.7073	49.4078	42.3744
207.10	Muscle	6.39E-09	26.3243	29.5267	24.3772
207.97	Muscle	<0.001	11.7768	12.1229	4.9160
209.12	Muscle	2.93E-07	46.9353	50.8145	44.1676
212.24	Muscle	<0.001	7.2935	11.0121	4.3086
213.15	Muscle	1.59E-04	51.9441	55.2470	50.3543
215.13	Muscle	<0.001	129.8685	197.2412	118.6095
226.20	Muscle	<0.001	2.6315	10.1306	2.7882
228.15	Muscle	<0.001	7.3895	11.9932	3.2390
228.18	Muscle	<0.001	2.8735	11.4956	2.8101
233.12	Muscle	<0.001	141.0864	163.6599	78.7432
236.08	Muscle	2.59E-07	31.9600	36.1029	29.7965
241.09	Muscle	6.88E-10	68.7838	76.9001	67.9177
241.19	Muscle	2.17E-12	30.1510	37.2426	27.4449
243.14	Muscle	<0.001	45.8898	62.8773	44.8063
245.15	Muscle	<0.001	193.2380	228.2726	157.7408
254.24	Muscle	<0.001	4.1048	16.1292	6.7604
255.15	Muscle	<0.001	14.0164	24.9323	12.5648
255.21	Muscle	<0.001	41.4759	61.9399	35.6426
256.22	Muscle	<0.001	3.0764	15.7218	5.8877
257.15	Muscle	<0.001	55.9386	67.6918	49.9363
259.17	Muscle	<0.001	179.4790	193.9965	154.4756
261.15	Muscle	1.44E-15	149.7832	166.3862	141.4424
265.13	Muscle	1.34E-09	22.2044	29.4393	24.1264
267.19	Muscle	1.59E-06	20.1127	22.9847	18.4606
268.26	Muscle	<0.001	11.6108	27.6922	16.8201
269.16	Muscle	<0.001	61.5451	169.3704	34.2699
269.23	Muscle	<0.001	57.0241	78.2693	53.0396
271.18	Muscle	<0.001	72.9756	95.7139	64.5695
273.19	Muscle	<0.001	314.4038	415.6983	255.2252
277.24	Muscle	6.56E-06	8.7651	11.1691	9.0257
279.19	Muscle	3.27E-10	17.7221	22.4776	17.3979
281.17	Muscle	<0.001	14.2792	27.7110	8.2800
282.28	Muscle	<0.001	1.6848	10.7694	6.0448
283.25	Muscle	<0.001	42.8997	55.3060	38.1655
285.17	Muscle	<0.001	71.9232	95.5757	56.4749
287.04	Muscle	2.23E-07	8.1820	12.1680	8.5710
287.18	Muscle	<0.001	115.7149	155.5222	91.4496
287.25	Muscle	1.53E-14	48.4920	52.5697	41.8401
297.24	Muscle	<0.001	9.8917	11.8937	1.0966
299.22	Muscle	<0.001	42.1513	115.4064	38.6617

			Mean Intensity		
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucosa**	**Muscle**	**Tumour**
301.18	Muscle	<0.001	663.8152	766.0373	587.7617
303.22	Muscle	1.01E-13	81.0047	87.6982	69.5196
304.29	Muscle	<0.001	74.3811	82.2645	65.9814
304.34	Muscle	<0.001	375.7918	570.2889	269.5636
306.12	Muscle	1.57E-10	13.6989	17.3317	11.6897
306.35	Muscle	<0.001	10.6633	15.6162	6.7599
306.98	Muscle	<0.001	61.8290	70.5168	57.6486
307.26	Muscle	1.71E-12	27.6601	36.0530	28.1325
308.98	Muscle	<0.001	62.3454	70.8755	57.8107
309.25	Muscle	2.62E-09	8951.9052	9136.9363	8585.4671
309.34	Muscle	1.31E-14	169.9625	186.4241	157.8859
309.46	Muscle	1.90E-06	54.6871	56.3306	47.7534
309.52	Muscle	5.75E-05	10.7822	11.1275	9.0720
310.98	Muscle	2.66E-06	21.3657	24.0007	20.2131
311.26	Muscle	6.34E-08	228.1350	234.0632	216.9544
311.30	Muscle	<0.001	60.8556	67.0999	51.1053
315.22	Muscle	<0.001	126.0164	167.9910	107.5537
323.27	Muscle	<0.001	10.8935	18.6308	9.1661
326.42	Muscle	<0.001	376.5159	518.4649	291.5519
327.22	Muscle	1.22E-15	36.5633	46.4925	36.9385
327.43	Muscle	<0.001	93.0191	131.2712	73.8388
329.06	Muscle	<0.001	22.3760	32.3156	20.5868
331.08	Muscle	4.33E-15	7.6306	11.5873	7.3428
331.31	Muscle	2.18E-11	4.8051	12.9863	10.8875
332.38	Muscle	<0.001	107.2064	143.9252	71.6245
335.18	Muscle	1.78E-08	46.7846	48.7985	41.4445
337.29	Muscle	1.80E-05	24.2730	30.3676	28.1678
339.31	Muscle	1.47E-14	5.8689	11.1227	4.9827
342.43	Muscle	1.37E-14	9.5300	11.9895	7.3835
349.04	Muscle	<0.001	203.8897	308.8370	185.0637
349.33	Muscle	2.14E-14	34.5261	74.1706	64.1400
351.04	Muscle	<0.001	196.6705	301.3522	180.5607
351.31	Muscle	<0.001	107.0732	184.9386	35.0468
353.06	Muscle	<0.001	64.6889	101.1830	61.1562
353.32	Muscle	<0.001	177.6014	302.5913	40.1497
357.36	Muscle	<0.001	3.2612	13.4568	0.1034
365.33	Muscle	<0.001	22.7508	33.0822	24.8741
367.35	Muscle	<0.001	23.5597	31.3013	10.6543
377.33	Muscle	<0.001	623.0954	666.8446	195.8956
379.35	Muscle	<0.001	805.1844	1382.5351	216.4895
381.36	Muscle	<0.001	65.0891	99.8185	19.4512
393.34	Muscle	<0.001	183.9436	196.0130	92.1038
395.36	Muscle	<0.001	150.8968	200.8707	81.1840

			Mean Intensity		
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucosa**	**Muscle**	**Tumour**
405.31	Muscle	3.93E-13	9.4862	16.2397	9.6163
407.32	Muscle	9.01E-08	36.4485	36.7310	29.7634
407.39	Muscle	<0.001	31.9400	34.7076	17.6434
409.34	Muscle	<0.001	88.3990	116.5540	79.4343
411.18	Muscle	1.83E-08	14.6445	16.2744	12.7921
411.35	Muscle	7.99E-15	27.4713	36.6524	25.4298
422.34	Muscle	<0.001	4.0992	16.4677	1.7403
423.33	Muscle	<0.001	37.9902	49.6033	33.6352
445.28	Muscle	1.19E-04	15.7205	18.3803	16.1111
462.25	Muscle	1.39E-06	10.7853	13.0058	9.5506
479.45	Muscle	1.49E-13	25.8274	39.4877	35.6909
491.53	Muscle	1.16E-07	4.3382	12.9807	7.1664
493.47	Muscle	<0.001	7.6068	22.3489	20.7699
495.49	Muscle	5.03E-07	10.0810	18.4116	13.7730
505.48	Muscle	3.03E-13	8.1465	14.4714	9.7834
505.55	Muscle	5.33E-11	12.1519	44.9923	16.4727
508.50	Muscle	<0.001	3.9640	18.1261	17.8577
517.56	Muscle	1.09E-11	0.4756	12.7981	1.8837
519.50	Muscle	<0.001	4.6733	17.3545	6.5654
519.57	Muscle	1.29E-11	32.7157	184.2411	48.5431
520.58	Muscle	8.86E-12	11.6795	65.3178	16.6659
523.50	Muscle	<0.001	11.2210	36.2688	23.0278
531.52	Muscle	1.01E-11	7.4843	12.8120	6.6998
533.52	Muscle	<0.001	13.9007	36.5213	19.9783
533.59	Muscle	4.47E-12	40.7566	289.1739	66.0725
534.60	Muscle	2.14E-12	14.9099	102.1852	23.0756
536.31	Muscle	1.14E-12	36.7351	44.1879	35.6799
545.54	Muscle	<0.001	12.2798	28.7267	17.6612
547.54	Muscle	2.05E-10	26.6718	33.4408	23.6370
547.62	Muscle	5.61E-12	4.8509	42.1181	8.4849
548.65	Muscle	4.06E-14	5.2330	18.6160	5.1525
549.58	Muscle	<0.001	11.7792	30.8988	19.6042
561.64	Muscle	<0.001	5.8683	13.9063	2.3263
587.63	Muscle	5.77E-05	15.8279	18.7211	13.9961
589.65	Muscle	<0.001	22.5902	33.2209	14.9419
591.66	Muscle	4.80E-08	10.5494	14.2814	10.5024
611.37	Muscle	1.65E-04	19.6569	23.4063	21.7186
615.67	Muscle	<0.001	58.7205	80.8544	37.4460
617.69	Muscle	<0.001	64.1524	93.5387	52.4009
619.71	Muscle	4.54E-11	6.3217	12.9549	8.4242
629.68	Muscle	1.71E-08	41.1833	53.1840	42.9281
630.67	Muscle	9.28E-08	9.4835	14.6304	9.4062
632.69	Muscle	1.19E-13	24.8678	36.2979	30.0485

			Mean Intensity		
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucosa**	**Muscle**	**Tumour**
634.70	Muscle	3.11E-08	10.6066	16.2895	12.6062
635.40	Muscle	3.89E-12	6.1314	10.6558	5.5499
639.69	Muscle	9.10E-14	14.9454	16.1280	6.6220
641.70	Muscle	<0.001	65.9011	81.3319	36.4905
643.72	Muscle	<0.001	90.1011	131.0901	72.1512
647.69	Muscle	<0.001	19.6276	45.8633	20.0659
655.69	Muscle	<0.001	30.3501	52.8675	12.7418
657.70	Muscle	<0.001	68.1357	120.4985	40.1450
659.72	Muscle	<0.001	52.3919	91.8009	42.2445
661.67	Muscle	<0.001	9.3152	21.0168	9.3259
671.69	Muscle	<0.001	25.0348	45.2040	16.2195
673.71	Muscle	<0.001	46.2047	95.1639	32.6703
675.72	Muscle	<0.001	27.2903	48.3400	26.7610
676.74	Muscle	<0.001	6.7544	14.0668	5.8397
677.69	Muscle	1.03E-13	3.8634	10.0240	3.9623
683.74	Muscle	8.97E-14	3.9149	11.6065	5.5810
684.42	Muscle	2.85E-12	24.3274	31.8470	27.3145
685.71	Muscle	2.61E-04	21.4854	24.4524	19.5210
686.42	Muscle	2.61E-06	14.2218	17.5142	14.1609
687.70	Muscle	1.11E-15	31.0903	44.6322	33.0410
691.73	Muscle	4.65E-13	18.5623	30.8993	24.3880
705.74	Muscle	<0.001	10.3174	21.1869	10.1116
707.74	Muscle	3.03E-05	10.5274	11.8357	7.4869
737.74	Muscle	<0.001	6.4406	11.7169	1.6603
761.85	Muscle	5.50E-13	5.1902	10.2608	4.1170
763.78	Muscle	<0.001	2.4742	10.3374	1.2600
765.79	Muscle	<0.001	0.4190	10.8865	0.0000
130.14	Tumour	<0.001	14.4507	20.5514	24.2283
143.10	Tumour	<0.001	11.6125	12.9666	22.0081
294.31	Tumour	<0.001	6.6039	8.9045	18.9191
301.25	Tumour	<0.001	82.3463	68.3562	100.2242
309.12	Tumour	3.02E-06	11.3165	14.1748	15.4235
317.24	Tumour	2.43E-04	145.5891	146.8738	160.4403
319.27	Tumour	<0.001	11.7475	15.0308	31.3312
321.28	Tumour	<0.001	26.1982	30.4553	49.0590
327.30	Tumour	<0.001	280.9902	229.1627	353.1605
335.27	Tumour	<0.001	1.0405	2.6849	11.5020
345.29	Tumour	6.66E-11	17.5325	27.2036	34.1220
347.31	Tumour	4.44E-15	29.0779	53.9394	64.4117
349.12	Tumour	<0.001	95.6712	61.8203	121.8416
359.26	Tumour	7.74E-05	154.5746	147.3970	161.6539
363.31	Tumour	3.33E-16	15.6558	36.8516	42.5435
366.41	Tumour	<0.001	26.3186	34.0081	74.9526

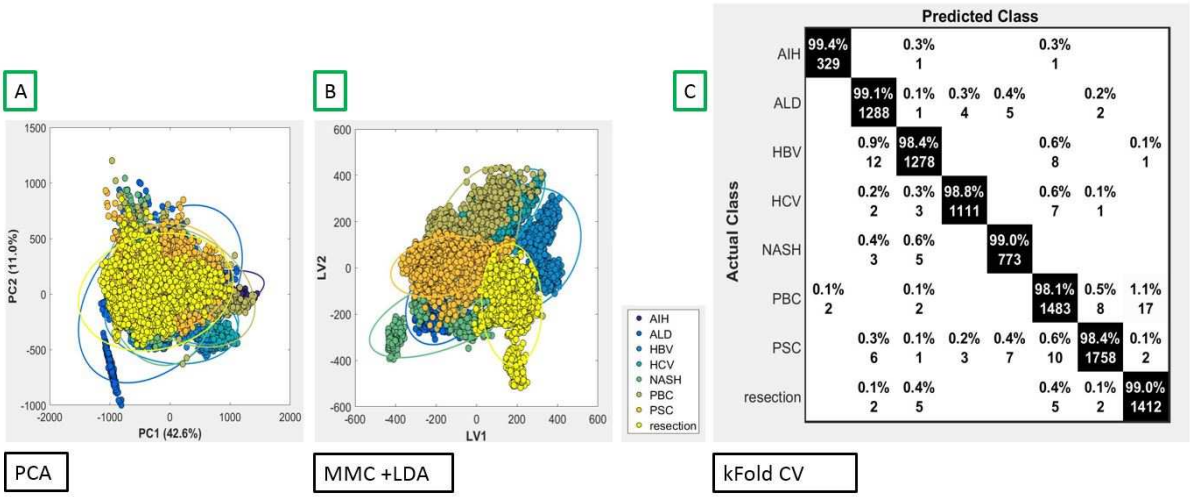
			Mean Intensity		
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucosa**	**Muscle**	**Tumour**
373.29	Tumour	6.98E-05	113.3058	104.7790	114.2254
380.43	Tumour	<0.001	6.5197	7.9495	13.2751
381.33	Tumour	<0.001	1.8492	1.6608	10.7237
383.28	Tumour	4.21E-13	22.5222	19.1138	26.4918
385.27	Tumour	1.63E-14	91.6901	83.2655	97.7835
385.31	Tumour	<0.001	12.0293	7.1900	15.3868
386.23	Tumour	1.18E-04	15.3652	14.7207	17.3632
389.32	Tumour	1.12E-10	71.2817	69.6677	85.8852
392.44	Tumour	<0.001	4.1082	5.9676	12.6329
393.37	Tumour	<0.001	464.6936	350.0794	475.2672
394.45	Tumour	<0.001	37.0779	52.1857	81.9863
397.42	Tumour	<0.001	5.3809	9.3861	13.3502
403.31	Tumour	2.30E-06	24.6032	24.0423	28.8911
417.33	Tumour	1.04E-10	27.0276	23.3157	31.3473
421.33	Tumour	2.46E-07	36.1960	31.3053	37.9182
421.42	Tumour	<0.001	39.5386	29.5875	45.3284
422.50	Tumour	<0.001	12.4185	19.4488	27.8373
425.30	Tumour	<0.001	716.9868	621.2831	821.8096
425.46	Tumour	2.37E-12	6.0054	12.8519	13.6597
427.32	Tumour	2.20E-12	32.9563	27.5470	37.8804
429.30	Tumour	<0.001	1.2641	0.5069	10.7788
429.34	Tumour	3.81E-14	30.5884	22.9262	31.7027
431.37	Tumour	1.11E-16	31.5288	23.1933	33.0328
435.44	Tumour	1.56E-05	14.0276	11.6544	15.7046
441.39	Tumour	<0.001	165.3521	137.9443	201.5919
447.44	Tumour	<0.001	25.2982	21.8527	34.6889
449.46	Tumour	5.40E-13	18.3291	16.0089	23.1418
450.97	Tumour	3.11E-15	33.4949	33.8208	41.7514
452.97	Tumour	<0.001	61.6300	63.8922	80.4472
454.97	Tumour	<0.001	50.5287	50.9659	63.8164
455.42	Tumour	7.97E-12	17.8398	17.8691	28.7759
456.97	Tumour	5.52E-10	22.7553	23.1776	28.0301
459.38	Tumour	3.02E-13	19.3170	17.2085	24.5464
465.43	Tumour	1.03E-09	10.3462	14.9832	20.5497
485.23	Tumour	1.11E-16	36.6951	27.7625	37.3440
502.42	Tumour	<0.001	23.1341	7.2674	35.7283
503.45	Tumour	<0.001	10.5455	4.8712	14.7975
505.38	Tumour	2.01E-04	8.3271	7.2480	10.4198
507.49	Tumour	<0.001	19.8542	56.6427	60.5140
518.45	Tumour	<0.001	46.1126	65.3551	74.1496
519.46	Tumour	3.62E-08	34.1829	29.9891	37.4096
521.51	Tumour	<0.001	7.9207	21.7867	36.1729
524.41	Tumour	1.44E-08	44.6417	36.6593	45.6709

			Mean Intensity		
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucosa**	**Muscle**	**Tumour**
527.32	Tumour	<0.001	0.0000	4.0650	13.5861
535.54	Tumour	<0.001	9.0378	26.1004	62.8525
544.48	Tumour	1.15E-10	13.4430	14.3930	19.9301
546.50	Tumour	1.52E-08	20.6877	28.2089	29.9066
548.42	Tumour	5.99E-06	12.7629	9.9628	13.8432
551.54	Tumour	7.44E-15	8.7083	15.7921	24.5315
553.61	Tumour	4.25E-05	27.1284	31.1142	32.3693
559.29	Tumour	<0.001	27.0965	19.4986	31.2744
561.56	Tumour	4.47E-10	27.4318	20.8106	27.8364
563.62	Tumour	2.18E-07	20.5088	24.4323	31.1383
579.60	Tumour	7.48E-06	12.2843	11.7556	17.0723
605.66	Tumour	1.11E-16	50.5758	43.9588	62.2733
607.65	Tumour	6.46E-14	4.4930	6.6121	14.6696
633.36	Tumour	1.55E-12	12.2504	6.9009	12.6299
635.47	Tumour	<0.001	9.3587	4.1833	12.7977
644.59	Tumour	<0.001	8.2200	11.5125	17.6035
645.50	Tumour	<0.001	0.6985	10.0741	23.8512
658.61	Tumour	<0.001	11.1195	19.0257	29.1135
659.64	Tumour	2.71E-10	17.2100	14.8836	23.7566
661.74	Tumour	9.98E-12	10.1363	14.2502	16.9189
665.70	Tumour	1.09E-10	8.7124	11.9619	18.6763
667.74	Tumour	<0.001	0.8595	10.0154	20.8470
671.63	Tumour	2.42E-04	15.2609	12.7026	16.6537
673.64	Tumour	<0.001	27.3781	26.6234	46.4774
675.65	Tumour	4.44E-10	6.4754	5.9550	13.2668
686.66	Tumour	7.51E-10	16.8282	17.2290	23.0445
688.64	Tumour	<0.001	10.3901	18.3476	29.0334
689.37	Tumour	3.54E-09	70.0045	62.8985	74.3017
693.76	Tumour	1.58E-14	9.9988	13.4563	21.0294
700.72	Tumour	<0.001	37.0459	30.0449	50.5394
701.76	Tumour	3.26E-06	31.7362	29.1148	35.4495
710.62	Tumour	<0.001	5.9940	11.2756	18.7671
713.73	Tumour	2.28E-13	5.9588	9.7842	16.4007
715.75	Tumour	<0.001	11.1879	23.2927	27.3927
716.69	Tumour	<0.001	7.9424	10.0128	20.4596
717.81	Tumour	<0.001	22.6090	21.7413	33.6655
725.79	Tumour	<0.001	49.5058	32.6064	52.5545
727.81	Tumour	<0.001	13.9807	11.5562	21.1581
728.76	Tumour	<0.001	20.3539	13.2206	25.4186
729.77	Tumour	6.93E-07	12.3925	9.8387	16.6630
730.48	Tumour	<0.001	0.1920	2.5525	21.0917
731.76	Tumour	4.95E-06	7.6881	9.3343	13.9552
743.59	Tumour	<0.001	0.0000	5.0085	12.4391

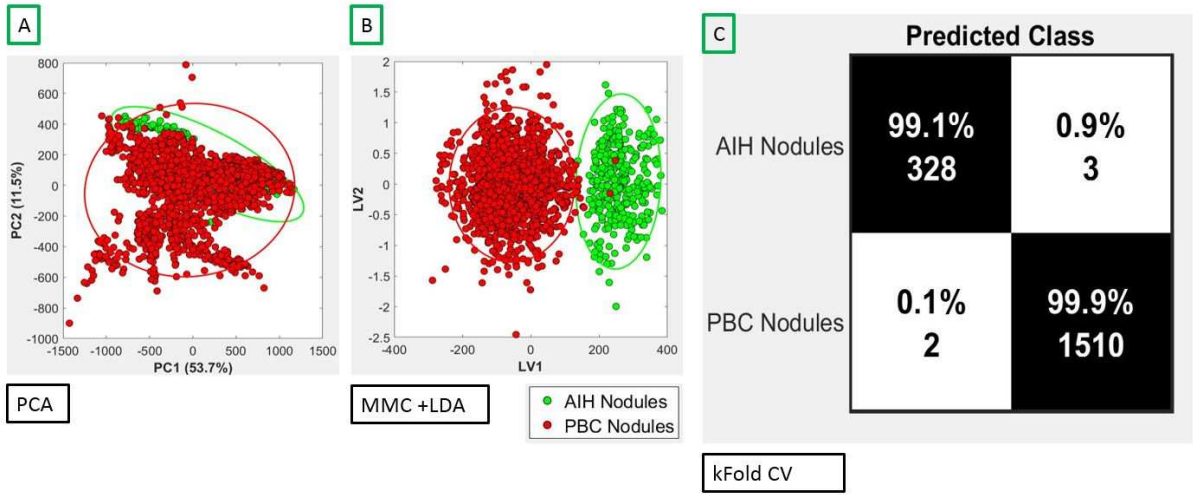
			Mean Intensity		
m/z	**Tissue Diff**	**ANOVA *p*Val**	**Mucosa**	**Muscle**	**Tumour**
743.73	Tumour	<0.001	9.2227	4.7759	15.9473
754.79	Tumour	<0.001	3.3563	1.6432	10.0682
755.80	Tumour	3.32E-09	8.7348	4.4339	11.1437
758.47	Tumour	1.11E-16	30.5179	24.5671	35.0195
759.59	Tumour	<0.001	0.0000	2.9281	10.5058
760.47	Tumour	2.03E-12	17.7275	14.1562	20.2619
763.43	Tumour	<0.001	82.0287	48.0112	99.4602
771.63	Tumour	<0.001	0.3686	2.1253	12.7127
781.62	Tumour	<0.001	0.0000	4.1346	11.6733
782.83	Tumour	2.22E-16	59.0139	50.3376	79.4222
783.84	Tumour	<0.001	47.5599	32.6991	56.9588
785.88	Tumour	6.07E-05	8.4226	7.0783	11.0215
798.84	Tumour	<0.001	2.7388	9.3391	15.8075
803.82	Tumour	<0.001	43.9410	13.8874	50.1498
808.86	Tumour	1.07E-10	20.7257	17.0865	26.1917
810.88	Tumour	4.44E-16	6.7527	4.8603	12.5383
811.89	Tumour	1.69E-14	6.7410	3.8082	10.4021
815.68	Tumour	<0.001	0.0000	1.2960	12.9692
830.85	Tumour	<0.001	7.9287	7.1243	24.6102
831.89	Tumour	<0.001	13.6190	7.3671	23.4631
832.54	Tumour	<0.001	12.2989	6.3745	13.8960
832.88	Tumour	6.66E-16	13.4207	7.6014	16.7202
837.49	Tumour	<0.001	36.1160	15.3963	43.6659
837.69	Tumour	<0.001	0.0000	0.5230	10.4112
859.92	Tumour	4.43E-14	5.9904	3.3910	10.4426
909.91	Tumour	<0.001	20.9972	13.4286	22.4608
910.04	Tumour	<0.001	17.5229	15.9735	26.9277
911.55	Tumour	<0.001	11.9337	3.4029	13.1856
931.88	Tumour	<0.001	0.8254	10.4316	12.6036
936.07	Tumour	<0.001	9.2400	6.8359	19.9407
938.09	Tumour	9.39E-05	8.1748	8.4729	11.5360

Appendix K. The results obtained using all annotated pixels per sample. This approach led to building models which were hugely overfitted; therefore, the results presented below are not correct.

A model was built for all analysed samples (cirrhotic liver diseases and resections, n=66). All annotated pixels were used. Some grouping was seen on the PCA plot (A); this trend was even more apparent when the supervised analysis was performed (B). An accuracy of over 95% was achieved for each disease in the K-fold cross-validation (C).



An attempt was made to built a model for AIH and PBC. A clear grouping was seen in both unsupervised (A) and supervised approached (B). Results of the cross-validation revealed an accuracy of over 99% for both AIH and PBC.



Appendix L. The *m/z* values representing the top-10 correlations found in all analysed liver samples.

mz1	*mz2*
277.2173	281.2486
769.5023	888.5696
775.5387	883.5338
766.5389	769.5023
280.2363	281.2486
280.2363	282.2519
723.5163	807.502
775.5387	807.502
741.5268	742.5391
769.5023	810.5289

List of conference presentations and publications

Oral presentations:

1. DESI-MSI-based diagnostics of cirrhotic liver diseases. ASMS 2016, San Antonio, Texas, USA

2. Metabolic phenotyping of colorectal tissue samples by Desorption Electrospray Ionization Mass Spectrometry Imaging. MSACL 2016 EU, Salzburg, Austria
    - MSACL 2016 EU Young Investigator Travel Grant

3. DESI-MSI-based diagnostics of cirrhotic liver diseases. OurCon IV 2016, Ustron, Poland

4. Metabolic Phenotyping of Fresh-Frozen and Formalin-Fixed, Paraffin-Embedded Colorectal Tissue Samples Using DESI-MSI. MSACL 2018 US, Palm Springs, USA
    - MSACL 2018 US Young Investigator Travel Grant

Poster presentations:

1. DESI-MSI-based diagnostics of cirrhotic liver diseases. Precision Medicine 2016, London, UK

2. DESI-MSI-based diagnostics of cirrhotic liver diseases. BMSS_BMRC MSI Symposium 2016, Sheffield, UK

3. Metabolic phenotyping of cirrhotic liver samples by desorption electrospray ionisation mass spectrometry imaging (DESI-MSI). MSACL 2017 EU, Salzburg, Austria
    - MSACL 2017 EU Young Investigator Travel Grant
    - Best Poster Award

4. DESI-MSI as a tool for molecular mapping of colorectal tissue samples. ASMS 2018, San Diego, USA

Publications:

1. Metabolic Phenotyping of Colorectal Tissue Samples by DESI-MSI for Clinical Research. Authors: **Anna Mroz**, Luisa Doria, Emmanuelle Claude, and Zoltan Takats, Waters Technical Brief, 2018

2. BASIS: High-performance bioinformatics platform for processing of large-scale mass spectrometry imaging data in chemically augmented histology. Authors: Kirill A.

Veselkov, Jonathan Sleeman, Emmanuelle Claude, Hans Vissers, Dieter Galea, **Anna Mroz**, Ivan Laponogov, Mark Towers, Robert Tonge, Reza Mirnezami, Zoltan Takats, Jeremy K. Nicholson, James I Langridge, *Scientific Reports,* 8(1), DOI: 10.1038/s41598-018-22499-z, 2018

3. Network analysis of mass spectrometry imaging data from colorectal cancer identifies key metabolites common to metastatic development. Authors: Paolo Inglese, Nicole Strittmatter, Luisa Doria, **Anna Mroz**, Abigail Speller, Liam Poynter, Andreas Dannhorn, Hiromi Kudo, Reza Mirnezami, Robert D. Goldin, Jeremy K. Nicholson, Zoltan Takats, Robert C. Glen, https://doi.org/10.1101/230052 , 2018

4. Variational autoencoders for tissue heterogeneity exploration from (almost) no pre-processed mass spectrometry imaging data. Authors: Paolo Inglese, James Alexander, **Anna Mroz**, Zoltan Takats, Robert C. Glen, arXiv:1708.07012, 2017

5. Deep learning and 3D-DESI imaging reveal the hidden metabolic heterogeneity of cancer. Authors: Paolo Inglese, James S. McKenzie, **Anna Mroz**, James Kinross, Kirill A. Veselkov, Elaine Holmes, Zoltan Takats, Jeremy K. Nicholson, Robert C. Glen, *Chemical Science,* 8(1), 2017

6. The immunoreceptor NKG2D promotes tumour growth in a model of hepatocellular carcinoma. Authors: Sam Sheppard, Joana Guedes, **Anna Mroz**, Anastasia-Maria Zavitsanou, Stephen M. Rothery, Panagiotis Angelopoulos, Robert D. Goldin, Nadia Guerra, *Nature Communications,* 8, 2017

7. Epithelial ovarian carcinoma diagnosis by desorption electrospray ionisation mass spectrometry imaging. Authors: Luisa Doria, James S. McKenzie, **Anna Mroz**, David Phelps, Abigail Speller, Francesca Rosini, Nicole Strittmatter, Ottmar Golf, Kirill A. Veselkov, Robert Brown, Sadaf Ghaem-Maghami, Zoltan Takats, *Scientific Reports,* 6, 2016

8. Targeted chemical analysis of the colon cancer microbiome using desorption electrospray ionisation mass spectrometry imaging (DESI-MSI). Authors: James Alexander, **Anna Mroz**, Alvaro Perdones-Monteiro, Louise Francesca Gildea, Simon J.S. Cameron, Frankie Bolt, Jeremy K. Nicholson, Julian Marchesi, Zoltan Takats, James Kinross, Julian P. Teare, *United European Gastroenterology Journal,* 4(55), 2016

9. 91 Mass Spectrometry Imaging (MSI) of Microbiome-Metabolome Interactions in Colorectal Cancer. Authors: James Alexander, Alasdair Scott, **Anna Mroz**, Alvaro Perdones-Monteiro, James S. McKenzie, Douglas N. Rees, Abigail Speller, Kirill A. Veselkov, James Kinross, Zoltan Takats, Julian Marchesi, Julian P. Teare, *Gastroenterology,* 150(4), 2016

10. XMS: Cross-Platform Normalization Method for Multimodal Mass Spectrometric Tissue Profiling. Authors: Ottmar Golf, Laura J. Muirhead, Abigail Speller, Júlia Balog, Nima Abbassi-Ghadi, Sacheen Kumar, **Anna Mroz**, Kirill Veselkov, Zoltan Takats, *J. Am. Soc. Mass Spectrom.*, 26, 2015

11. Benchmark datasets for 3D MALDI- and DESI-imaging mass spectrometry. Authors: Janina Oetjen, Kirill A. Veselkov, Jeramie D. Watrous, James S. McKenzie, Michael Becker, Lena Hauberg-Lotte, Jan Hendrik Kobarg, Nicole Strittmatter, **Anna Mroz**, Pieter C. Dorrestein, Zoltan Takats, Theodore Alexandrov, *GigaScience,* 4(1), 2015

12. Rapid Evaporative Ionization Mass Spectrometry Imaging Platform for Direct Mapping from Bulk Tissue and Bacterial Growth Media. Authors: Ottmar Golf, Nicole Strittmatter, Tamas Karancsi, Steven D. Pringle, Abigail Speller, **Anna Mroz**, James Kinross, Nima Abbassi-Ghadi, Emrys A. Jones, Zoltan Takats, *Analytical Chemistry,* DOI: 10.1021/ac5046752, 2015

13. PTU-121 Llt1 Is Upregulated in Hepatocellular Carcinoma and Inhibits Nk Cell Cytotoxicity. Authors: Naveenta Kumar, **Anna Mroz**, Mina Aletrari, Robert D. Goldin, Marco A. Purbhoo, *Gut,* 63, 2014

14. Assessing the impact of OCT embedding medium for desorption electrospray ionisation mass spectrometry imaging. Authors: James S. McKenzie, **Anna Mroz**, Renata Soares, Abigail Speller, Zoltan Takats. *In preparation*

15. Analysis of human colorectal carcinoma FFPE samples by desorption electrospray ionisation mass spectrometry imaging (DESI-MSI). Authors: Renata Soares, **Anna Mroz**, James S. McKenzie, Francesca Rosini, Jocelyn Tillner, Rob Goldin, Zoltan Takats. *In preparation*

16. Desorption electrospray ionisation mass spectrometry imaging for the classification of estrogen receptor status in breast cancer. Authors: Dipa Kumari Gurung, Paolo Inglese, James S. McKenzie, Francesca Rosini, Edward St-John, **Anna Mroz**, Hiromi Kudo, Rathi Ramakrishnan, Kirill Veselkov, Robert Glen, Jeremy K. Nicholson, Zoltan Takats. *In preparation*

17. Desorption electrospray ionisation for targeted mass spectrometry imaging. Authors: Andreas Dannhorn, Maria Luisa Doria, James McKenzie, Emrys Jones, Renata Soares, **Anna Mroz**, Jocelyn Tillner, John G. Swales, Richard J.A. Goodwin, Sadaf Ghaem-Maghami, Zoltan Takats, *Chemical Communications*, *In preparation*

Copyright and Permissions

Figure 1.1

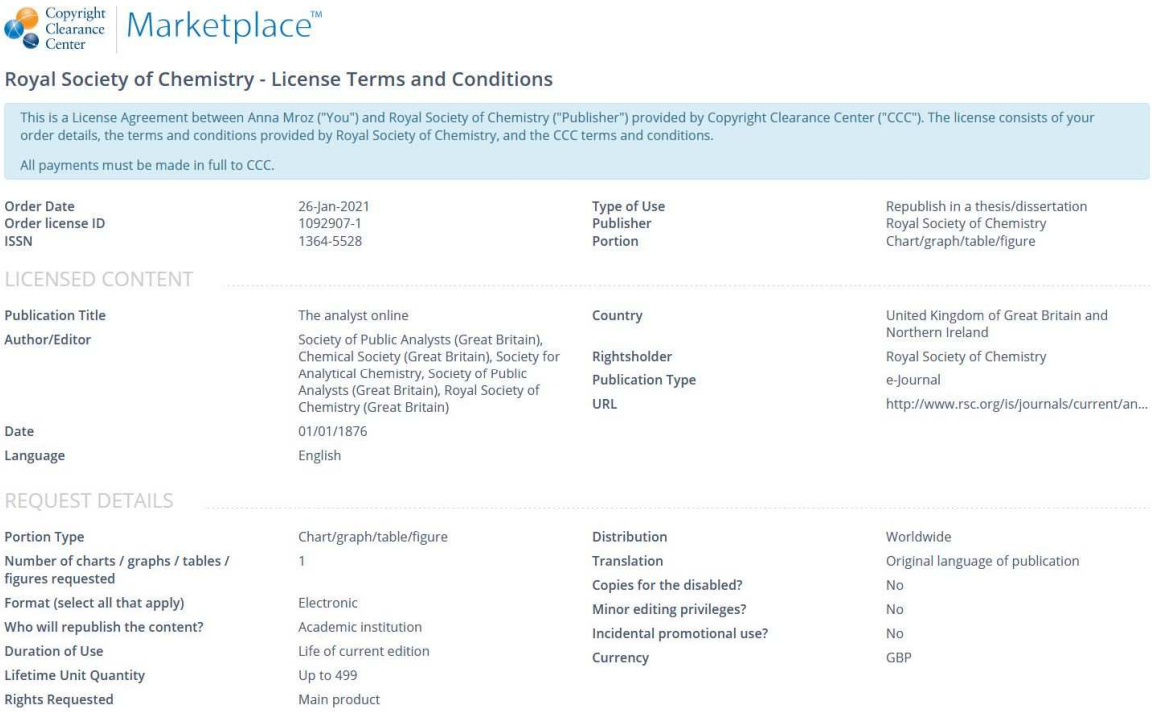
ELSEVIER LICENSE
TERMS AND CONDITIONS

Jan 25, 2021

This Agreement between Ms. Anna Mroz ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4832121397750
License date	May 18, 2020
Licensed Content Publisher	Elsevier
Licensed Content Publication	Cell
Licensed Content Title	Hallmarks of Cancer: The Next Generation
Licensed Content Author	Douglas Hanahan,Robert A. Weinberg
Licensed Content Date	Mar 4, 2011
Licensed Content Volume	144

Figure 1.4



Copyright Clearance Center | Marketplace™

Royal Society of Chemistry - License Terms and Conditions

This is a License Agreement between Anna Mroz ("You") and Royal Society of Chemistry ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Royal Society of Chemistry, and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	26-Jan-2021	Type of Use	Republish in a thesis/dissertation
Order license ID	1092907-1	Publisher	Royal Society of Chemistry
ISSN	1364-5528	Portion	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	The analyst online	Country	United Kingdom of Great Britain and Northern Ireland
Author/Editor	Society of Public Analysts (Great Britain), Chemical Society (Great Britain), Society for Analytical Chemistry, Society of Public Analysts (Great Britain), Royal Society of Chemistry (Great Britain)	Rightsholder	Royal Society of Chemistry
		Publication Type	e-Journal
		URL	http://www.rsc.org/is/journals/current/an...
Date	01/01/1876		
Language	English		

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	Distribution	Worldwide
Number of charts / graphs / tables / figures requested	1	Translation	Original language of publication
		Copies for the disabled?	No
Format (select all that apply)	Electronic	Minor editing privileges?	No
Who will republish the content?	Academic institution	Incidental promotional use?	No
Duration of Use	Life of current edition	Currency	GBP
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

Figure 1.6

THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE LICENSE TERMS AND CONDITIONS

Jan 25, 2021

This Agreement between Ms. Anna Mroz ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") consists of your license details and the terms and conditions provided by The American Association for the Advancement of Science and Copyright Clearance Center.

License Number	4825901468252
License date	May 11, 2020
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science
Licensed Content Title	Mass Spectrometry Sampling Under Ambient Conditions with Desorption Electrospray Ionization
Licensed Content Author	Zoltán Takáts,Justin M. Wiseman,Bogdan Gologan,R. Graham Cooks
Licensed Content Date	Oct 15, 2004