# Developing Laser Assisted - Rapid Evaporative Ionisation Mass Spectrometry (LA-REIMS) as a platform for the direct analysis of faecal metabolites in gastrointestinal health and disease

A thesis by

Petra Paizs

Submitted for the degree of Doctor of Philosophy (PhD)

Imperial College London

Department of Metabolism, Digestion, and Reproduction

2022

*For my family and friends who have supported me throughout this journey.*

## Abstract

Faecal metabolomics presents an avenue for the non-invasive study of the gastrointestinal (GI) tract, the gut microbiome, and their interactions with environmental factors (1).

This thesis serves as a first step towards developing Laser Assisted - Rapid Evaporative Ionisation Mass Spectrometry (LA-REIMS) as a platform for faecal metabolomics analysis in GI disease. Faecal LA-REIMS optimisation work was carried out and yielded improved signal-to-noise ratios and a decrease in % carry-over between samples. Further, pilot work on faecal water content in LA-REIMS analysis demonstrates the importance of correcting for variable water content. A novel application of LA-REIMS imaging for the real-time mapping of metabolites in whole fresh and frozen faecal samples was developed.

The results of the optimisation work were implemented in two large scale clinical datasets. In a cohort of 234 patients (control n=131, adenoma n=77, colorectal cancer n=26), LA-REIMS demonstrates the potential for accurate direct from faecal sample adenoma and colorectal cancer (CRC) detection; this could reduce the burden of unnecessary colonoscopy and optimise cancer referral pathways. In a cohort of 169 patients (Crohn's disease n=60, Ulcerative Colitis n=60, and control n= 49), faecal LA-REIMS demonstrates moderate-to-high diagnostic accuracy for direct from sample inflammatory bowel disease (IBD) detection. Faecal LA-REIMS analysis has the potential to optimise IBD diagnostic pathways and could reduce the burden of unnecessary endoscopic, radiological, histological and laboratory investigations. This thesis presents the first study utilising optimised LA-REIMS faecal metabolomics analysis in CRC and IBD with several tentatively identified features significantly different between disease and control states. Patient-related modifiers of the faecal metabolome were investigated in both the CRC and IBD datasets.

Optimised faecal LA-REIMS analysis has potential as a rapid, preparation-free screening and diagnostic technique for CRC and IBD.

---

## Copyright declaration



## Statement of originality

I certify that this thesis and the research to which it refers are the product of my own 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.

Petra Paizs

## Acknowledgements

My foremost gratitude goes to my supervisors, Prof. Zoltan Takats, Mr James Kinross, and Dr James Alexander.

Zoltan, I cannot thank you enough for giving me so many opportunities during this PhD. Thank you for always teaching me, for encouraging me to delve into the world of analytical chemistry and ultimately for helping me become the scientist I am today. Thank you for giving me so much of your time and patience, and first and foremost, for giving me the opportunity to grow as a person both inside and outside of the lab.

James K, your encouragement and enthusiasm for science and translational medicine are non-comparable. Thank you for being a constant force for new developments and areas of interest. Thank you for your support and for all the time and energy you put into this work and my development.

James A, thank you very much for all your help and support during this thesis. Thank you for providing me with opportunities to expand my research within the field of gastroenterology and for giving advice and support whenever I needed it.

Furthermore, I would like to thank my post-doctoral mentors, Dr Lauren Ford and Dr Simon Cameron, for helping me throughout this period.

Lauren, thank you for teaching me. Not only did you support me with the science of this thesis, but you also encouraged me to become an independent scientist. Thank you for lifting me up. You were a true inspiration and role model for me during this time.

Simon, thank you for teaching me the fundamentals. I couldn't have done this one without you! Thank you for all the patience and time you took to educate me.

A special thank you goes to Maria Sani and Stefania Maneta. Thank you for always keeping your door open and supporting me with your knowledge as well as friendship. Thank you for making the long hours of the PhD so much more joyful and enjoyable.

I want to thank Alvaro Perdones-Montero. Thank you for supporting me with my work, for being so patient, and for being a wonderful friend. You are greatly missed and will always be remembered.

I would also like to thank Dr Yuchen Xiang and Dr James Mckenzie. Thank you for always being there when I needed help and for supporting me throughout this.

I would like to thank all the students and staff at Imperial College London, including other PhD students and everyone else who helped me during my PhD.

I also thank our colleagues from the University Hospitals Coventry & Warwickshire, especially Prof Ramesh Arasadnam and Dr Monika Widlak, for their generous contribution of samples as well as the staff at Imperial College London involved in the IBD project, including Dr Sam T Powles, Dr Leo Chong, Dr Horace Williams, Dr Lucy Hicks and Prof Timothy Orchard.

I would like to thank the NIH British Research Council for funding this thesis as well as Waters Corporation for providing consumables and instruments.

# Table of Contents

# List of Figures

# List of Tables

## List of Appendices

# 1. Introduction

## 1.1. Metabolomics

Metabolomics has emerged as a novel, rapidly growing technology and is defined as "the systematic identification and quantification of small molecule metabolic products (metabolites) of a biological system" (2). An array of complex biochemical reactions are executed in cells and involve diverse metabolites to carry out functions, including, amongst others, biosynthesis and the regulation of gene expression (3,4). The metabolome is representative of the complex underlying biochemical pathways and provides an insight into the interactions between genetic and environmental factors (5).

Metabolomics belongs to the family of "omics" sciences, which includes genomics, transcriptomics, and proteomics; each investigates the genome, transcriptome, and proteome, respectively (6). The diverse and dynamic metabolome encompasses the study of the downstream products of genomic, transcriptomic, and proteomic processes (7). Multi-omics research attempts to study the flow of information, from the original causative agent to the functional consequences or interactions (8).

The overall aim of omics sciences is the identification, characterisation, and quantification of biological molecules (9). The benefit of studying metabolomics, as opposed to other omics sciences, is that the metabolome is the endpoint of the biochemical "omics" cascade, that is, the phenotypic information expressed and modulated by all other layers (10). The metabolome is a desirable tool for the study of diseases as it is predictive of the actual phenotype (11).

Typically, the metabolome is investigated in specific diseases to identify potential markers of disease onset and progress or to investigate biochemical processes and pathways altered in disease states (12). Understanding changes in the metabolome with specific diseases demonstrates potential for the discovery of biomarkers (13). Furthermore, the metabolome can indicate early biochemical changes in metabolic pathways, and therefore it's study

demonstrates potential for the identification of early disease biomarkers (14). Additionally, metabolomics may give an insight into the mechanistic drivers of specific diseases and how these can be used for therapeutic purposes (15).

### 1.2. Metabolomics in systems biology and precision medicine

Biomedical research has traditionally been pursued using a reductionist approach (16). In reductionism, a compounded and multiplex problem is solved by dividing it into smaller parts and studying the subunits in isolation (17) . Reductionism has limited applicability for the study of complex and multi-factorial chronic disorders such as cancer (16). The aetiology of chronic disorders is multi-fold, with a combination of genetic and environmental factors driving disease onset and progression (18). One of the main limitations of a reductionist approach in the study of complex disorders is the potential loss of information of the whole integrated system (19).

The systems biology approach, where rather than dividing a complex problem into its components, a holistic approach is taken to study the complex interactions and dynamics of a network, offers a valuable alternative approach (20). Systems biology utilises highly specialised mathematical and computational tools to analyse data collected via global measurement technologies such as genomics, metabolomics, or proteomics aiming to understand the complexity of biology (21). Systems medicine is “the application of systems biology to human physiology in a clinical context” (17). Metabolomics research is key to systems biology or medicine, as it represents integrative information across multiple functional levels and links genetic information with environmental factors (22).

Precision medicine considers individual variability in environment, lifestyle, and genetics for disease prevention and treatment (23). Metabolomics, offering an insight into the interplay between genetic and environmental factors, is suitable to investigate these phenomena. The direct characterisation of metabolic phenotypes in health and disease enables precision medicine through the identification of metabolic profiles in disease states, the discovery of therapeutic targets as well as the discovery of potential biomarkers for therapeutic stratification (2). However, the difficulty of translating metabolomics research findings into

clinical practice needs to be highlighted and the fact that no single metabolomics biomarker has to date been approved for use in precision medicine.

### 1.3. Untargeted-discovery *vs* targeted-validation metabolomics

Two main analytical approaches are traditionally utilised in metabolomics, that is, untargeted and targeted analysis. Untargeted metabolomics is a hypothesis-generating technique for the global scanning of the metabolome and extensive measurement of metabolites present in each sample (24), this includes non-targeted profiling, metabolic fingerprinting, and metabolic foot printing (25). Untargeted metabolic profiling is the measurement of metabolites in a way that reflects their response to stimuli (26). On the other hand, the investigation of patterns of metabolites that change in response to stimuli is known as metabolic fingerprinting (27). Moreover, metabolic foot printing is the study of extracellular metabolites (28).

In targeted metabolomics pre-defined sets of metabolites are measured with high precision and accuracy, usually performed as a way of validating untargeted analysis (24). In the targeted approach, a quantitative analysis is applied to specific metabolites or perturbations in metabolic pathways, also known as targeted metabolic profiling (16,29).

### 1.4. Lipidomics is a sub-field of metabolomics

Lipidomics, a sub-field of metabolomics, is “the systems-based study of the structure and function of the entire set of lipids (the lipidome)” (30). Lipids are multi-functional metabolites that play essential roles in biological processes and structures, including cellular barriers, membrane matrices, signalling, and energy depots (31). Lipids are an integral part of human physiology; it is therefore not surprising that aberrations in lipid metabolism are seen in the pathogenesis of a multitude of diseases (31,32).

## 1.5. Faecal metabolomics and lipidomics

Metabolomics is used for the analysis of various biospecimens, including tissues, tissue extracts or biofluids such as plasma, serum, urine, etc. Each of the specific biofluids offers information on the organs, tissues, and cells from which they originate. To investigate the gastrointestinal (GI) tract, faecal metabolomics presents an avenue for non-invasive study (33). Faecal metabolomics can be utilised in the study of the GI tract, the gut microbiome, and the interplay between them and environmental factors, such as diet (33)

A key distinction between the GI metabolome and the faecal metabolome needs to be made. When investigating the GI metabolome, tissue samples taken along the intestinal lining are most representative of the actual phenotype (34). However, most studies focus on the analysis of faeces, mostly because these samples can be obtained in a non-invasive fashion. It is crucial to recognise that the metabolite composition of faeces may differ from the metabolite composition of the different areas of the intestine (35,36). Therefore, conclusions from studies investigating the GI system and the gut microbiome deducted from faecal sample analysis must be interpreted cautiously.

Faeces contains an abundance of molecules that are the end products of nutrient ingestion, digestion, secretion and absorption by the GI microbiota and the host (33). Faeces consists of 63-85% water, depending on the amount of fibre intake (37). Faecal dry weight is a complex mixture of bacterial biomass (25-54%), exfoliated colonic epithelial cells, undigested diet residues, macromolecules, and metabolites (37).

Faecal metabolomics is a potential useful tool for studying GI health and disease. Reasons for this include the proximity of faeces to the colorectal mucosa ,where metabolic processes and subsequent metabolites produced provide signatures of GI health (33). Further, cells are being constantly shed from the intestinal lining into faeces, and colonocyte exfoliation is significantly altered in different pathological conditions (38). Furthermore, the cells contain lipids that are key to GI biochemistry and more conserved across the microbiota than for instance proteins (39).

Additionally, faecal metabolomics allows the study of the gut microbiome, which aids not only in maintaining homeostasis but also has a large number of symbiotic metabolic functions with the host (40). The gut microbiome is defined as "the totality of all bacteria, archaea, bacteriophages, eukaryotic virus, and fungi coexisting in the GI tract" (41). The gut microbiota has important roles in host immunity, protection against pathogen overgrowth, food digestion, host-cell proliferation and vascularisation, and the regulation of intestinal endocrine functions, neurologic signalling, and bone density (42). Furthermore, it is involved in energy biogenesis, the biosynthesis of vitamins and neurotransmitters, the metabolism of bile salts, and the elimination of exogenous toxins (43).

Intestinal microbiota dysbiosis, an imbalance in the composition of intestinal microbes, is associated with diseases ranging from GI to neurologic, respiratory, metabolic, hepatic, and cardiovascular illnesses (43). Metabolomics and the microbiota are closely related, the gut microbiome is involved in a number of metabolic functions, including vitamin and short-chain fatty acid (SCFA) production, amino acid synthesis, bile acid transformation, hydrolysis and the fermentation of non-digestible substrates (44,45). The metabolome is described as "the functional readout of the gut microbiome, and its interaction with host and other environmental factors" (46).

Overall, the advantages of studying the faecal lipidome and metabolome include among others a non-invasive sampling nature, no need for qualified personnel for collection of samples, a strong correlation existing between the organ of pathogenesis and the biofluid, as well as the gut microbiota (47). Furthermore, it allows the study of interactions between dietary components, host and the microbiota, which has been established to play a role in intestinal diseases such as colorectal cancer (CRC) and inflammatory bowel disease (IBD) (47,48). The disadvantages include a relatively variable metabolome due to heavy influences of food and water intake, a high chemical shift variability and the extensive sample preparation and extraction needed with traditional metabolomic techniques (49).

### 1.6. Analytical techniques for faecal metabolomics research

### 1.6.1. Background

Several advanced analytical chemistry techniques exist that allow the investigation of the faecal metabolome and lipidome (50); the two most utilised platforms are Mass spectrometry (MS) and Nuclear Magnetic Resonance (NMR) spectroscopy.

### 1.6.2. Nuclear magnetic resonance (NMR) spectroscopy analysis of faeces

NMR spectroscopy of faeces generates a comprehensive biochemical profile of low-molecular-weight metabolites of high abundance (51). Though complex, these NMR spectral profiles can be readily obtained and through the use of data reduction and chemometric techniques analysed and interpreted (51). NMR spectroscopy is a well-established platform for metabolic profiling and fingerprinting of faeces (52). NMR can simultaneously provide structural and quantitative data on metabolites with varying biochemistry (53,54). Next to being able to operate in a largely untargeted fashion, it can analyse biofluids in a non-destructive manner (55,56). It is a robust and reproducible technique for investigating the faecal metabolome (57,58). The major weakness of NMR analysis is the lack of sensitivity of the analytical technique (59).

### 1.6.3. Mass spectrometry (MS) analysis of faeces

MS is a sensitive and specific approach for the measurement of a large number of metabolites (60). In MS analysis, differences in the relative abundance of features of interest can be investigated between study groups (59). Comparatively, MS has the advantage of superior sensitivity over NMR (59). However, MS signal is affected by the sample preparation utilised, ion suppression, and matrix effects (59). Different MS approaches can be used in combination, as various ionisation techniques and mass analysers are available to maximise the number of metabolites detected (52). A variety of analytical techniques exist which are frequently coupled to MS, here the focus will be on liquid chromatography (LC), gas chromatography (GC), capillary electrophoresis (CE), and direct infusion/injection (DI), due to their potential applicability to the study of GI pathology (61) .

#### *1.6.3.1. Liquid Chromatography-Mass Spectrometry (LC-MS)*

MS is often coupled to chromatographic techniques for complex samples to be separated before sample inlet into the MS (61). In chromatography, complex mixtures are introduced into a column that contains solid support coated with a stationary phase (with specific chemical properties) through which a mobile phase (liquid in LC) of changeable composition passes (62). The sample introduced is initially adsorbed onto the stationary phase followed by changes in retention by the stationary phase sequentially as the solvent composition is altered (62,63). The sequence of elution of the components of a sample is a function of the preference of particular compounds to be associated with either the stationary phase or the mobile phase as the composition of the solvent changes (61). Several chromatographic approaches can be used in LC-MS to target molecules with varying chemical properties. Normal phase LC contains a stationary phase that is polar and uses non-polar solvents as the mobile phase. Here, compounds elute progressively from least to most polar (64). Hydrophilic liquid interaction chromatography (HILIC) utilises an aqueous mobile phase (62) and allows for the analysis of a broad range of compounds (61). Another type of LC is reverse phase, where compounds elute in the order of their polarity but in the reverse order to that of normal phase chromatography, with the highest polarity compounds eluting first (62). Samples containing non-polar compounds will adsorb preferentially to the reversed phase stationary phase (non-polar) coated on the solid support (65). Compounds are eluted progressively, using a gradient of the composition of the mobile phase, from aqueous to organic (62).

One of the main benefits of LC-MS is that metabolites do not have to be in the gaseous phase for analysis. Therefore, no sample derivatisation is required and sample preparation requires less time when compared to GC-MS (66). Furthermore, due to coupling to chromatography, reduced ion suppression and higher detection limits can be achieved (61). On the other hand, limitations of the technology include increased analytical equipment required and long analytical runs (when compared to ambient MS) (61).

#### *1.6.3.2. Gas Chromatography-Mass Spectrometry (GC-MS)*

GC-MS is a highly sensitive technique widely used in the analysis of faecal metabolites (50,67–69) that provides high-chromatographic resolution and potential quantification of compounds using internal standards (61). However, only volatile molecules can be analysed, this necessitates the derivatisation of non-volatile molecules for GC-MS analysis (70). One major concern is the necessity for samples to be thoroughly dried before analysis, as derivatisation techniques are highly sensitive to water content (61). Long sample pre-processing times, due to derivatisation, pose a problem for its use in the clinical setting for rapid diagnostics (69). Nevertheless, due to the large volume of volatile compounds present in the GI tract, it is a widely used technique in faecal metabolomics, typically targeting important components of the metabolome (e.g. SCFAs) (71).

#### *1.6.3.3. Capillary Electrophoresis -Mass Spectroscopy (CE-MS)*

CE-MS applies a high voltage electrical current to a sample in order to achieve liquid separation (61). CE-MS demonstrates potential for its application to the analysis of highly complex biological samples, such as for instance faecal metabolomics, because of its capabilities to provide highly targeted analysis (72). However, a downside of the technique is the considerable length of run times (61). To date, CE-MS has not been used for metabolomics analysis of GI diseases. However, it demonstrated utility in numerous biomarker discovery studies and clinical applications (73).

#### *1.6.3.4. Direct Infusion/injection – Mass Spectroscopy (DI-MS)*

In DI-MS, samples are directly injected into the ionisation source of the MS (61). This results in a higher sample throughput (reduced analysis time) with minimal sample preparation in metabolic profiling because the chromatographic step is eliminated (74). One of the major limitations of DI-MS is the relatively strong matrix effect (74). A number of studies using DI-MS for the metabolic profiling of microbial metabolism have been reported (75,76). DI-MS

demonstrates potential as a tool for GI metabolomics research (61). However, to date, no published data exist on the application of DI-MS to faecal metabolomics research.

### 1.6.4. Summary of analytical techniques

The analytical techniques described in section *1.6.3. Mass spectrometry (MS) analysis of faeces* have limitations for their use in faecal metabolomics and lipidomics analysis. MS techniques such as LC-MS and GC-MS require time-intensive sample preparation (77). CE-MS and DI-MS demonstrate potential in their applicability to faecal metabolomics studies, but further research is needed to optimise and standardise the techniques before biological interpretation can be attempted. Additionally, NMR-based techniques lack in sensitivity (52).

## 1.7. Mass Spectrometry (MS) based metabolomics workflow

The MS-based workflow consists of: experimental design, sample preparation, MS analysis, data analysis and metabolite identification, see Figure 1 (78).

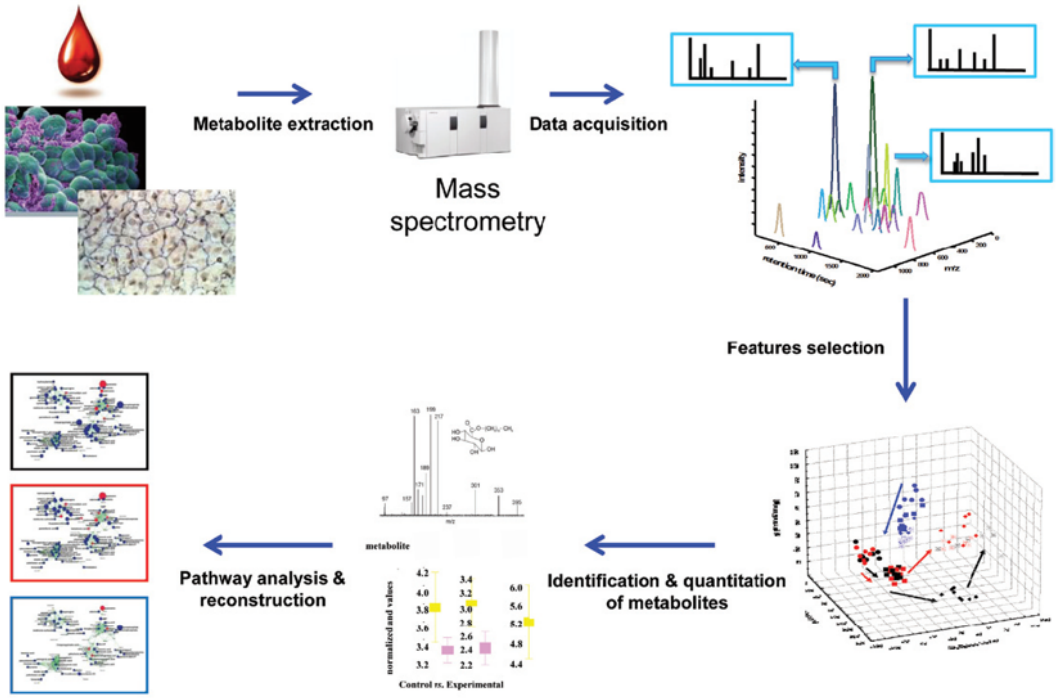


Figure 1: Workflow of MS-based metabolomics study. Adapted with permission from Zhang et al 2012 (78); Copyright 2021 Analyst see section *Appendix 1: Permission to reprint.*

### 1.7.1. Biological question formulation and experimental design

Well-designed studies with careful sampling, sample preparation and analysis are essential. The study should ensure that the analytical data derived from the samples allows the answering of the proposed biological question reproducibly (79).

### 1.7.2. Sample collection/preparation and chromatographic separation

A crucial step when designing a metabolomics study is the collection and storage of samples. Sample preparation and extraction protocols/procedures should be considered along with the sample type used (blood, faeces, urine, etc,), the polarity of compounds in the sample, and the type of study to be carried out (80). It is recommended to keep biological samples at low temperature, preferably -80°C, and to avoid freeze-thaw cycles (81). Chromatographic separation techniques for MS have been described in detail in section *1.6.3. Mass Spectrometry (MS) analysis of faeces.*

### 1.7.3. MS analysis

The first step in typical MS-based analysis is sample introduction into the MS (most commonly after chromatographic separation). This is followed by the ionisation process, of which the most common types are electrospray ionisation, chemical ionisation, and electron ionisation (62). Ions enter the analyser, where separation based on their mass-to-charge *(m/z)* ratio occurs followed by detection. The detector counts the number of ions at each m/z, producing a mass spectrum (61).

### 1.7.4. MS data analysis steps

MS analysis is typically followed by data analysis in the form of data pre-treatment, analysis using a variety of statistical tools, and finally metabolite identification and biological interpretation (66).

#### *1.7.4.1 Data pre-treatment*

Raw data obtained through MS analysis requires pre-processing, as in addition to analyte ions of interest, background ions and noise are produced (81). Filtering noise, or performing a baseline correction of the raw data, facilitates peak detection and potentially reduces the detection of false positives (82).

Unwanted variation in raw MS data originate from experimental (e.g. human error or within instrument variation) (83) or biological factors (e.g. concentration of biofluid analysed) (84). To address these factors, normalisation is frequently utilised to identify and remove observed and unobserved sources of variation (83). The choice of normalisation technique is a complex one, and is data- and experimental- dependent (85). A common technique is total ion count (TIC) normalisation, whereby, all features in a sample are divided by the total number of ions observed (86).

Other techniques involve scaling methods, which are data pre-treatment approaches whereby variables are divided by a scaling factor (85). A variety of techniques are used for scaling purposes, including data dispersion techniques (uses standard deviation as scaling factor) e.g. pareto scaling, and the size measure techniques (uses, for instance, the mean as scaling factor) e.g. level scaling (87).

Transformations, such as for instance log transformation, are non-linear conversions of data, these are applied to correct for heteroscedasticity or to make skewed distributions (more) symmetric (87). Heteroscedasticity describes a phenomenon where “the variability of a variable is unequal across the range of values of a second variable that predicts it” (88). Another key step in data pre-treatment is peak detection, whereby filters with predetermined parameters are used to select peaks in the MS profile for further analysis (89).

### *1.7.4.2. Data analysis*

Due to many variables and complexity of metabolomics data sets, multivariate analysis, where all variables are considered at once, is commonly performed. Multivariate tools include exploratory approaches based on unsupervised techniques as well as supervised models (90). In unsupervised analysis, data is analysed without considering experimental groups or expected outcome; this first step in metabolomics data analysis allows for the observation of natural groupings of variations in the sample set (91). A supervised technique makes use of prior knowledge of study groupings to drive the analysis (91). Metabolic patterns in relation to variables of interest such as for instance disease phenotype, can be studied using supervised methods (24).

For supervised multivariate analysis, the model validity and ability to handle new observations reliably need to be checked; procedures used for this include, cross-validation and permutation tests (91). Differences observed in metabolomics data may not be solely due to the biochemical changes in the observed systems but rather due to factors such as sample handling, or for instance changes in freeze/thaw cycles. It is, therefore, crucial to test and validate identified variables for potential biological underpinnings (61).

Univariate analysis, where only one variable is analysed at a particular time, include tests such as for instance the *t*-test (92). One of the problems of univariate analysis is that changes which are only jointly significant may be overlooked (92). By selecting those metabolites that appear to be significant in both univariate and multivariate analysis, there is an increased chance that they can be validated in separate experiments (61).

### *1.7.4.3. Metabolite identification and biological interpretation in untargeted metabolomics*

Metabolite identification is a significant bottleneck in MS-based analysis and will be considered next, specifically in the context of known non-novel, known novel metabolites, as well as unknown metabolites. Guidelines have been established for the standard of metabolite identification, these classify compound identification based on confidence in the identity with grades ranging from 1-4 (93,94).To achieve grade one characterisation, that is

confident identification, for known non-novel metabolites, Sumner et al (95) proposes as the minimum reporting standard at least two independent and orthogonal data points relative to an authentic compound analysed under identical experimental conditions. This could be for instance the comparison of samples analysed to authentic compounds with regards to retention time and the mass spectrum or alternatively accurate mass and tandem MS (95) under identical conditions. The importance of identical experimental conditions is highlighted by retention time shifts from assay to assay due among others analysis conditions (temperature shifts), type of column, column dimension, degradation of the column, and contaminations (96). Therefore, it is evident that experimental conditions must be identical in comparative assessments.

Tandem MS is commonly used by comparing the fragmentation patterns of authentic compounds of the putatively identified feature of interest with the fragmentation pattern of the sample (97). Furthermore, tandem MS can allow for the identification of fragments which can give more information about the compound of interest such as head groups of lipids, or the length of fatty acid chains in phospholipid species (98). The absolute structure is still not possible with MS/MS alone as some parameters would remain unknown such as the position of unsaturation of a fatty acid chain (98).

The use of literature reports for authentic samples from external laboratories yields only grade 2 identification, that is a putative annotation (94). Once the statistical analysis has generated a list of potentially biologically important metabolite ions, the *m/z* values with accurate mass or tandem MS fragmentation patterns can be matched against public metabolite databases such as METLIN (https://metlin.scripps.edu/index.php), the Human Metabolite Database (http://www.hmdb.ca), SciFinder (http://www.scifinder.cas.org) and the Lipidmaps database (http:// www.lipidmaps/org). However, these databases are often incomplete (99).

For the novel identifications of metabolites, a more in-depth approach is recommended. Here, evidence for full stereochemical structural identification should be sought, which involves among others extraction, isolation, and purification, followed by elemental analysis,

accurate mass measurement, isotope distribution, tandem MS fragmentation patterns, as well as NMR analysis (95).

The identification of unknown metabolites that are not present in databases is a challenging and time-consuming task. For a full structural analysis of a previously unknown molecule, usually, multiple techniques are required such as MS, NMR and the ultimate gold standard of x-ray crystal structure elucidation (100). Workflows in this area follow on from natural product chemistry such as fractionation, purification of target analytes, high-resolution accurate mass MS, and 1D and 2D NMR for structure determination (101). However, fractionation and purification of target analytes are more difficult with metabolites from human samples such as in faecal metabolomics as original biomass is usually available at low volumes. Further, it is a time-intensive and costly process, bioactivity is lost as compounds degrade and mixtures of synergistic compounds are separated, thus introducing challenges for the identification of specific compounds (102). The identification of unknown metabolites is not always possible due to for instance lack of time or other resources. Unknown metabolites can often still be differentiated based on m/z values and/or chromatographic features such as retention time and reporting of these is recommended (95).

## 1.8. Challenges in faecal metabolomics

### 1.8.1. Complexity and dynamic nature of metabolomics - need for standardisation

Metabolomics presents a significant analytical challenge as it aims to measure molecules with diverse biochemical properties, such as complex lipids and hydrophilic molecules (2). In addition to the complexity and diversity of the faecal metabolome, it is also a dynamic system which is affected by various stimuli including genetic, environmental, as well as host-microbial factors (103).

Minor stimuli have the potential to cause major shifts in metabolic profiles. Factors that may contribute to variation in a given network include amongst others, diet, stress, disease, age, gender, epigenetics, and the gut microbiota (103). Next to inter-individual variation in the

faecal metabolome, considerable intra-individual variation also exists, depending on for instance changes in eating patterns, bowel activity, and medication use (33).

Metabolomics research necessitates the standardisation of analytical workflows and techniques to maintain consistency, reduce variation between subjects, and optimise information recovery (14). Despite the clear need for standardisation in the field, studies to date often present various and sometimes contradictory strategies for sample preparation, sample storage, analytical platforms and data analysis (104). The variation between research studies has led to a major bottleneck, characterised by a lack of reproducibility and difficulty in establishing common phenomena, networks and pathways (105).

One of the main needs for optimisation and standardisation in metabolomics is in the development of in-vitro diagnostics (IVD) instruments and point-of-care tests and methods (106). In order for metabolomics to be translated into clinical practice, specifically precision medicine, standardised protocols, careful optimisation and appropriate validation techniques are essential (107).

#### *1.8.1.1. Collection of samples and storage before analysis*

Faeces is more sensitive to differences in study design than other body fluids such as urine or blood (108). For instance, faeces is susceptible to post-collection metabolite changes caused by microbial fermentation (33). In current faecal metabolomics studies, large variations exist in the collection methods of samples, ranging from home to laboratory and clinical site collection, which can potentially introduce a bias. To allow for the analysis of the metabolome in its native biological context, *in vivo* sample collection in the form of endoscopic collection, has been proposed (109). However, this has traditionally been avoided due to the invasive nature of the procedure as well as the costs associated (109).

Faecal samples are mostly collected after bowel emptying in an *ex vivo* setting, where the sample is exposed to the ambient environment before analysis (33). However, this allows for *ex vivo* bacterial metabolism, which can significantly alter the faecal metabolome, from the

time of passage up to the point of analysis. Furthermore, in most settings, it is not feasible to analyse samples straight after bowel emptying.

Couch et al., 2013 (109) demonstrated in a study investigating faecal volatile organic compound (VOC) metabolomes in pairwise matched endoscopic (*in-vivo*) and home collected samples (*ex vivo*) that whilst the two collection techniques were nearly identical in their specific VOC composition, global changes in the relative metabolite abundances existed. The results invite further studies evaluating methods of faecal collection and preservation, especially in the context of other analytical techniques. Overall, the collection methodology should be kept constant across study groups and throughout the experiment.

A lack of agreement exists on the optimum storage conditions for faecal samples in human metabolomics analysis. Most studies aim to store faeces as soon as possible after bowel emptying in a range from -80°C to -20°C before analysis to maximise metabolic stability and minimise sample degradation (110–112). Lotfield et al., 2016 (113) demonstrated that faecal samples collected with additive, such as 95% ethanol, are suitable for metabolomics analysis. Further, it was suggested that samples stored with 95% ethanol yield similar results after four days of storage at ambient temperature or shortly after collection (113). Gratton et al., 2016 (114) illustrated that crude faecal samples metabolic profiles differed over time at different storage temperatures, with branched-chain and aromatic amino acids increased in faecal samples undergoing freeze-thaw cycles later on(114).

On the other hand, De Spiegeleer et al (108), demonstrated in an ultra-performance liquid chromatography (UPLC) MS-based storage condition study of faeces, that the polar metabolome was affect more by storage conditions than the lipidome. The study concluded with a recommendation of a storage time no longer than 18 weeks at -80°C for faecal lipidomics and a storage time no longer than 8 weeks at -80°C for faecal polar metabolomics. Furthermore, a review in faecal metabolomics by Karu et al., 2018 (33) recommends that if faecal samples cannot be processed and analysed straight away, they should be frozen and stored at -80°C. The samples should be homogenised and aliquoted before freezing, to minimize freeze-thaw cycles (33).

#### *1.8.1.2. Homogenisation*

Faeces is a highly heterogeneous biofluid consisting of bacteria, food remains, nutrients such as lipids, fibres, and non-digestible elements (61,115) To eliminate the bias of a heterogeneous sample, homogenisation should be carried out before freezing and should be kept constant throughout an experiment (33). A variety of techniques exist for homogenisation, ranging from simple stirring using a spatula to using a mixer.

#### *1.8.1.3. Faecal sample analysis – analytical platforms*

To date, no single technique exists that covers the entire diversity of the metabolome (103). The decision as to which platform to use is based on the research question driving the study, the nature of the sample, and a careful evaluation of the benefits versus limitations of each analytical platform. A major consideration is the extent to which a particular technique can accurately reflect the biological faecal composition (111). Further details of the different analytical platforms are described in section *1.6. Analytical techniques for faecal metabolomics research.*

#### *1.8.1.4. Water content of samples/ Intestinal transit times/ Stool consistency*

The water content of faeces fluctuates based on inter-individual, intra-individual, and diurnal variations (the range of water content in human faeces is 60%-85%) (116). This variation, if not accounted for may introduce bias (33). A direct or indirect approach is often used to correct for water content of faecal samples. A correction for between sample faecal water content variation is carried out in the indirect approach, here, normalisation of data in the pre-processing step is utilised (33).

In the direct approach, the exact water content is measured or alternatively, all water is removed/dried from the sample through for example lyophilisation (33). The benefits of lyophilisation include reproducibility and prevention of bacterial growth, whereas the limitations include loss of metabolites and sample volume, change in metabolic profile, and extensive sample preparation which is time-consuming (108).

### 1.8.2. Bioinformatics challenge – From MS features to metabolite identities

A major challenge in metabolomics is the vast amount of data generated during an experiment, most of which is left unexplored and uninterpreted. To maximise experimental output and help answer the research question posed, in-depth data evaluation is essential (117). Furthermore, the identification and validation of potential biomarkers demonstrate a key potential of metabolomics research (103). However, lack of automation and standardisation in annotation and assignments of identities to biomarkers exist. A key problem is a limited number of comprehensive MS-based databases available. Moreover, in essence, it is difficult to compare data between laboratories due to among others, differences in analytical platforms, columns, gradients, and collision energies used for instance in MS experiments (103). Overall, these bioinformatics challenges pose the major bottleneck in current metabolomics research.

## 1.9. Ambient mass spectrometry techniques

### 1.9.1. Background

Ambient ionisation MS is an emerging technique in which ions are generated at atmospheric pressure with little-to-no sample preparation (118). Ambient MS, a high-throughput alternative to more traditional hyphenated methods, allows samples to be assessed with little-to-no need for sample preparation (119). Further, ambient MS allows for the rapid and direct analysis of liquids, gases, and solids, both *in situ* and *in vivo*, as well as MS imaging (118).

Traditional hyphenated-MS approaches allow for the development of robust quantitation methods (120). However, they necessitate lengthy sample preparation times with associated risks of metabolite alterations or loss during the process (118). This is of importance in clinical application and rapid diagnostics, as time is of critical importance in those fields. It is well established that the metabolome is constantly changing and that many biological questions can only be answered with samples being analysed in near real-time (118). This poses a

challenge for hyphenated-MS techniques and is where the benefit of ambient MS lies. Ambient MS techniques have wide applicability in clinical metabolomics, the study of microorganisms, food omics, and agriculture (118). The focus of this thesis is the clinical application of ambient MS; the other areas although promising are out of the scope of this work.

### 1.9.2. Common ambient MS techniques in faecal metabolomics research

A variety of ambient techniques exist, a brief introduction will be given to the main techniques applicable to clinical faecal metabolomics research. The main techniques described include desorption electrospray ionisation (DESI) (121), direct analysis in real-time (DART) (122) and rapid evaporative ionisation mass spectrometry (REIMS) (123). It is important to highlight that to the best of the candidate's knowledge, no published data on DESI or DART faecal metabolomics analysis exists. REIMS has been previously applied to faecal metabolomics by our research group and will be discussed in more detail in section *1.9.3. Rapid evaporative ionisation mass spectrometry (REIMS).*

#### *1.9.2.1. Desorption Electrospray Ionisation MS (DESI-MS)*

Ambient MS was first described by Takats et al., in 2004 with the introduction of DESI-MS (121). DESI-MS is performed at atmospheric pressure with minimal sample preparation (121). Briefly, a spray mixture of charged micro-droplets is directed onto the sample surface, the analyte interacts with the solvent, and subsequent desorption into gas phase occurs as well as ionisation (124). In the next step, the ionised compounds are moved to the mass spectrometer inlet via an atmospheric pressure ion-transfer line capillary and are detected (125).

DESI-MS has many benefits which include a time-efficient high-throughput setup, minimally destructive nature, and the ability to analyse specimens in a variety of different sample forms (124,125). Challenges related to reproducibility problems, measurements automation, and

effects of the sample matrix on the efficiency of ionisation and droplet formation are disadvantages of DESI-MS (125).

In addition to allowing for direct analysis, DESI-MS can be used for imaging biological tissues and surfaces. When first developed, DESI-MS imaging (DESI-MSI) allowed for a spatial distribution of molecules with a lateral resolution of ≈250 µm (126). Nowadays, studies indicate lateral resolutions that reach to up to 10 µm, however the typical resolution is 100 µm (127).

DESI-MSI can be applied to the study of biological tissues (128). DESI-MSI and DESI-MS have been applied successfully in biomedical research. Tissue analysis by DESI-MSI can provide exogenous drug and/or drug metabolite distributions, as in the analysis of dried blood samples (DBS) for quantitative evaluation of xenobiotics (129). Further, DESI-MS is a promising method for molecular pathology applications. DESI-MSI research as a cancer diagnostic tool and profiling technique of different tissue types has been carried out in breast cancer (130), prostate cancer (131), liver adenocarcinoma (132), bladder cancer (133), gastric cancer (134), colon and rectal cancer (135), ovarian (136) and brain cancer (137). Although DESI-MS has been used in the analysis of gastric and colorectal tissue, to date no published data exist on its use in the analysis of faeces.

#### *1.9.2.2. Direct Analysis in real time (DART) MS*

DART-MS is a direct sampling plasma ionisation technique utilised for the high-throughput analysis of solids, liquids, or gases (138). Heated gas (metastable species, typically $H_2$ or $N_2$) is used to thermally desorb the sample, followed by ionisation through proton and charge transfer mechanisms (118).

DART-MS has been used to vaporise and ionise a large variety of samples, including among others, drugs, plant metabolites, peptide, and oligosaccharides from a variety of surfaces including glass, paper, and thin-layer chromatography (TLC) plates (62). DART-MS has been applied to DBS for newborn screening of phenylketonuria (122) and for pharmacokinetic/toxicokinetic research (138). DART MS is an easy to use and rapid method,

however, it is only useful for small molecule analysis. Further, it cannot be combined with chromatography because it is a surface ionisation method (62). To the best of the candidate's understanding, DART-MS has not been used in the investigation of the faecal metabolome to date.

### 1.9.3. Rapid evaporative ionisation mass spectrometry (REIMS)

#### *1.9.3.1. REIMS development/ iKnife*

REIMS is an ambient ionisation technique, which allows for the rapid thermal evaporation of the sample of interest and releases ions in the gaseous form in the process (139). REIMS, first described in by our research group in 2009, is coupled to instruments used in surgery, such as electrosurgical diathermy (intelligent Knife or iKnife) (123). REIMS allows for *in situ* chemical analysis (139). Since the first description of REIMS, it has proven to be a valuable tool for real-time monitoring, identification, and discrimination of tissue types based on lipid profiles (118). REIMS has since been used in the analysis of lipid profiles of tissues for colorectal adenomas and cancer (140), ovarian cancer(136), breast cancer (141), cervical cancer (142), as well as lung cancer, brain, liver, and gastric cancer (143). Additionally, REIMS was developed as an MS imaging platform for the analysis of biological samples (144).

#### *1.9.3.2. REIMS in microbiology*

In studies previously published by our group on microorganisms, REIMS was shown to provide accurate speciation of bacteria and yeast directly from colonies (144,145). The initial work of REIMS on microbiology samples was performed using a hand-held bipolar probe allowing REIMS to analyse microbial cultures directly.

Further work by our group automated the setup with a customised platform for high-throughput REIMS analysis by using a TECAN Freedom EVO instrument integrated with a Pickolo system (SciRobotics, Israel) (146). This platform requires minimal user input and minimises user-related errors. To test the accuracy of the setup, 375 different clinical

bacterial isolates were analysed using the hand-held bipolar probe REIMS and the high-throughput setup; the bipolar probe REIMS showed a speciation accuracy of 96.3% and the high-throughput REIMS platform demonstrates an accuracy of 93.9% (146).

Furthermore, REIMS has been utilised for the metabolic phenotyping of the important Cystic fibrosis pathogen *Pseudomonas aeruginosa* (147). In a different study published by our group, REIMS was able to accurately identify 153 clinical Candida isolates to species level with bipolar REIMS and high-throughput REIMS platforms demonstrating 96% and 100% classification accuracy to species level, respectively (148).

#### *1.9.3.3. Ex-vivo and in-vivo tissue analysis of the gastrointestinal tract using REIMS*

REIMS has previously been utilised for the investigation of the GI tract, our group demonstrated that REIMS can provide rapid metabolic phenotyping of rectal and colonic cancers intra-operatively (143). REIMS demonstrated a sensitivity of 98% and a specificity of 97.9% for colonic and rectal cancer during surgical interventions in the operating theatre (143). Further work (140) demonstrated in a study of twenty-eight patients that REIMS is capable of an overall diagnostic accuracy of 94.4 % for the detection of cancer versus adenoma tissue. Cancer samples demonstrated a significant increase in long-chain phosphatidylserines and bacterial phosphatidylglycerols, normal adjacent mucosa showed raised plasmalogens and triacylglycerols, and adenomas demonstrated an increase in ceramides (140). Furthermore, to allow *in vivo* analysis of the GI lumen, REIMS was integrated with an endoscopic polypectomy snare; endoscopic REIMS was able to differentiate cancer, adenoma, and normal tissue *in vivo* (140).

### 1.9.4. High-throughput Laser Assisted (LA) – REIMS in faecal metabolomics analysis

Our group reported the first application of the high-throughput REIMS platform to the analysis of human faecal sample (149). REIMS was capable of detecting bile acids and phospholipids of possible bacterial origin (149). In the faecal REIMS analysis of patients undergoing bariatric surgery, REIMS demonstrated predicted separation between pre- and

post-surgery specimens (149). It is important to highlight that the samples were lyophilised to control variation in faecal water content.

Further, the potential of faecal REIMS in a clinical CRC dataset was first investigated between April 2018 and May 2018. The results of the pilot work demonstrated the feasibility of a LA-REIMS based model for the screening of potential biomarkers for colorectal adenomas and cancer (150). However, it was concluded that careful optimisation and method development was necessary for future faecal LA-REIMS metabolomics studies. This study was part of a BSc project (by the candidate) and has not been published. Therefore, to allow for an optimised more reproducible high-throughput LA-REIMS setup for the analysis of faecal metabolites, the laser and instrumentation optimisation study outlined in Chapter 4 was carried out.

Since the adoption of the high-throughput liquid handling platform for analysis, there has also been a change from the use of monopolar probes penetrating the sample (radiofrequency (RF) electrical current), to the use of a 10.6 μM carbon dioxide ($CO_2$) laser lens assembly system (FELS-25A, OmniGuide, USA) (151). The current optimised high-throughput LA-REIMS setup for faecal samples and water extract metabolomics and lipidomics analysis is depicted in Figure 2. During faecal sample and water extract analysis, the $CO_2$ laser beam is fired onto the sample with aerosol aspirated into a Xevo G2-XS Quadrupole Time-of-Flight (Q-ToF) MS (Water Corporation) for analysis (Figure 2A) (146,151).

A light stage Pickolo platform (SciRobotics, Israel)) for viewing agar plates (Figure 2B) via an inbuilt camera as well as a cooling plate (Figure 2C) for the analysis of well plates have been fitted. The Q-ToF mass analyser demonstrates high-mass resolution and mass ranges; it offers fast scanning capabilities and mass accuracy (according to the developer) in the order of 5 ppm. Q-ToF instruments are suitable for obtaining metabolite fragmentation data; MS/MS experiments (152).

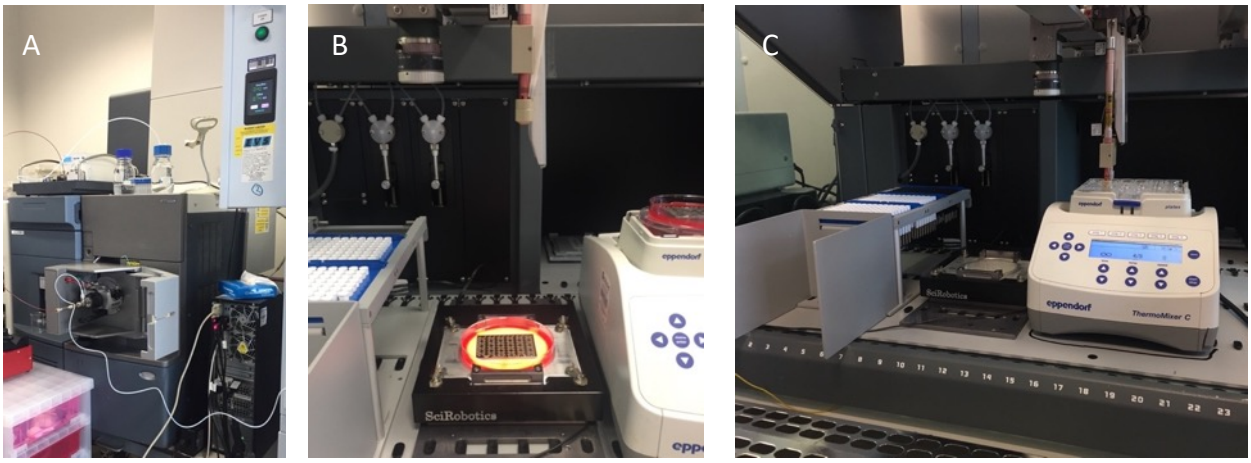



Figure 2: High-throughput LA-REIMS setup for faecal biospecimen and faecal water analysis with A Xevo G2-XS QToF MS B Pickolo visualisation platform for faecal water analysis C Cooling plate set to 4°C, $CO_2$ laser, and well plate for faecal sample analysis

Overall, laser-based sampling is a promising method (151), for instance, there is no need for contact to be made to the sampled surface using the laser, and the spatial resolution may be more easily defined. Genangeli et al 2019 (153) compared the diathermy unit to a $CO_2$ laser handpiece. The laser-generated an improved signal-to-noise ratio when compared with the diathermic knife (153). Further, the focused laser beam minimises tissue damage (153).

Besides, Van Meulebroek et al 2020 (120), demonstrated the use of LA-REIMS with an OpoletteTM HE2940 pump laser (OPOTEK, LLC), with the analysis of mimicked crude faeces samples type 2 diabetes or euglycemic study participants. Significant separation based on disease classification was achieved with an accuracy of 90.5% (120). It is important to note, that faecal samples were lyophilised and re-suspended in fixed volumes of water before LA-

REIMS analysis to control for variability in the water content of the sample. Furthermore, no optimisation of laser or instrumentation settings was carried out for the REIMS setup.

## 1.10. Colorectal cancer

### 1.10.1. Background

CRC is the 4th most common cancer in the UK, accounting for 11% of all new cancer cases with incidence steadily rising (154,155). Globally, men are more likely to be diagnosed with CRC than women, with the age-standardised incidence in males being 23.6, and in females being 16.3 (per 100,000) (156). Furthermore, CRC is 3–4 times more common in developed than in developing nations (157). Genetic and environmental factors can increase the likelihood of developing CRC (158), these will be considered in more detail as non-modifiable and modifiable risk factors, respectively.

CRC arises through three distinct carcinogenic pathways: the adenoma-carcinoma sequence, the serrated pathway, and the inflammatory pathway (159). Of the different types of CRCs that exist, around 90% originate from epithelial cells of the colorectal lining and are termed adenocarcinomas (160). The adenoma-carcinoma sequence is characterised by genetic or epigenetic mutations occurring sequentially, with the formation of a pre-cancerous adenoma, which may evolve into cancer (161).

Around 60-65% of CRC cases arise sporadically, via acquired somatic genetic and epigenetic changes mostly due to modifiable risk factors (162). The other 35-40% of CRC cases are due to inherited CRC susceptibility (159). For instance, studies have reported that individuals with one affected first-degree relative have, on average, a two-fold increased risk of CRC compared with individuals with no family history (163–165). Only approximately 5% of CRC cases are due to hereditary cancer syndromes such as hereditary nonpolyposis colorectal cancer (Lynch syndrome) or familial adenomatous polyposis (162).

CRC carcinogenesis encompasses three major genetic and epigenetic aberrations: chromosomal instability (CIN), CpG island methylator phenotype (CIMP), and microsatellite instability (MSI) (159). These aberrations accumulate in CRC pathogenesis with 85% CIN, 20 % CIMP-positive, and 15% MSI-high status seen in sporadic CRC (166).

Cancer cells harbour distinctive characteristics that can contribute to tumour development. A major hallmarks of cancer cells is enhanced proliferation, characterised by unique metabolic re-programming, through which cancer cells obtain enough energy for new biomass synthesis (167). Metabolic pathways are actively re-programmed, leading to increased glycolysis, glutaminolysis, and fatty acid (FA) synthesis, a hallmark of cancer (168).

Cancer cells can produce energy in a nutrient-deficient environment. Otto Warburg (169) first demonstrated the preference of cancer cells for glycolysis rather than oxidative phosphorylations (OXPHOs) in the presence of oxygen, effectively establishing aberrant glucose metabolism as one of the hallmarks of cancer. In a shift towards the so-called reversed Warburg effect, it was established that cancer cells have unique metabolic features and environments, with some synthesizing ATP utilising OXPHOs (170). Furthermore, glutamine dependency and dysregulated lipid metabolism is recognized as another metabolic phenotype in many different cancers, including CRC (168).

Non-modifiable risk factors include chronic IBD (157). IBD, a chronic inflammatory condition in the colon, results in the abnormal release of among others, growth cytokines, excess blood flow, and metabolic free radicals; these factors increase CRC risk (157). Additionally, children who survived malignancies and who received treatment in the form of abdominal radiation are at increased risk of CRC as adults (157).

Non-modifiable risk factors also include ethnicity, with for instance in the United States African Americans and Native Americans having the highest incidence rates of CRC and suffering lower survival rates (171). Hispanic Americans demonstrate the same rates and survival for CRC as do white Americans (157). Evidence indicates the pattern seen in terms of

ethnic disparities, may be driven by exposure to modifiable risk factors and medical care (172).

Further to this, other non-modifiable risk factors have been identified including having cystic fibrosis (173), a cholecystectomy (174), or androgen deprivation therapy (175). The most significant risk factor for CRC is age, the disease is most common in people older than 50, with an estimated 90% of cases and deaths happening after this age (159,176). However, CRC rates have been shown to rise among people younger than 50 (159).

In addition to genetic non-modifiable risk factors, CRC risk is largely affected by factors related to lifestyle and diet choices (159); obesity is an established risk factor for CRC It must be noted, that is not known whether increased CRC risk by obesity-related inflammation is caused directly or secondarily through for instance insulin resistance; chronically increased insulin levels have been associated with CRC pathogenesis (177). The high levels of insulin found in CRC and obesity increase the bioavailability and expression of insulin-like growth factor (IGF1), which in turn, increases cell proliferation and reduces apoptosis (178).

Physical inactivity is an acknowledged risk factor for CRC, based on observational epidemiological evidence (159). Furthermore, evidence exists that diet changes may play an important factor in the primary prevention of CRC. Several foods, vitamins, but also dietary patterns have been studied in relation to CRC risk (179). The World Cancer Research Fund in association with the American Institute for Cancer Research published a revised report on the association of CRC and diet in 2018, concluding that strong evidence exists that consuming whole grains and foods containing dietary fibres, dairy products and calcium supplements decreases the risk of CRC and that consuming red and processed meat and approximately two or more alcoholic drinks per day increases the risk of CRC (179).

Ethanol intake in alcoholic drinks is associated with an increased risk of CRC, with the first metabolite of ethanol breakdown in the body, acetaldehyde, being carcinogenic to humans according to the International Agency for Research (180). Alcohol, along with its metabolites, causes epigenetic, biochemical, and immunological changes, which in turn lead to chronic inflammation and increased risk of CRC (181).

Furthermore, tobacco smoking is a risk factor for colorectal adenomas (182) and CRC (183). Besides, it has been associated with increased CRC-specific mortality (183–185) in previously cancer-free individuals. Genetic and epigenetic aberrations can be induced by carcinogenic molecules form cigarette smoke reaching the colorectal lining. (186).

Among environmental factors in the pathogenesis of CRC, the role of GI microbiome in CRC pathogenesis has been increasingly recognised (187). The GI microbiota is in constant crosstalk with the intestinal epithelium and changes in their relative abundance have been shown in intestinal and extra-intestinal diseases (188,189). Furthermore, CRC development is associated with increased "intestinal permeability", which is a result of perturbations in gut barrier functions (190). The host immune response is influenced by the gut microbiota and their metabolic products (190). CRC development has been strongly associated with intestinal dysbiosis (158).

### 1.10.2. Diagnosis and early detection – current techniques

The current screening, surveillance, and preventative measures for CRC are driven by the notion that adenomas are the precursors of adenocarcinomas (160).
Early detection and treatment of CRC is key, as is demonstrated by the net survival for CRC patients diagnosed at stage I disease being 98% for both men and women compared to 44% in men and 35% in women with stage IV disease (191).

The faecal immunochemical test (FIT) is the screening modality of choice in the UK Bowel Cancer Screening Program (192). FIT demonstrates significantly high sensitivities and specificities for the detection of CRC. In a meta-analysis of 19 studies, FIT demonstrated a pooled sensitivity of 79% (95% CI 0.69-0.89) and a pooled specificity of 94% (95% CI 0.92-0.95) for CRC detection (193). Further studies have demonstrated a FIT sensitivity of 73.8% and a specificity of 96.4% for CRC detection (194).

On the other hand, the sensitivity of FIT for detecting advanced adenomas is significantly lower (194), with results of multiple studies yielding large differences in sensitivities (6-56%)

(195–200). Overall, FIT appears to have high sensitivities and specificities for the detection of CRC but demonstrates a low to moderate sensitivity for adenoma detection. This poses a problem in terms of using FIT as an early detection tool for CRC, as preferably, a screening tool would identify the pre-cancerous stage of disease, thus allowing for intervention and prevention of the development of neoplasia. This highlights an important limitation of FIT and reinforces the need for improved screening tests for CRC.

### 1.10.3. Colorectal cancer – metabolomics as a disease diagnostic tool

Metabolomics may be used as a fingerprinting as well as a comprehensive metabolic profiling technique, to investigate underlying drivers and molecular networks of diseases, such as CRC (201). Cancer cells demonstrate distinctive characteristics that contribute to tumour formation and progression; these include changes in carbohydrate, amino acid metabolism, and lipid-related metabolic pathways (201). Metabolism has been shown to differ in normal and neoplastic tissues (202), with metabolic reprogramming being crucial for the proliferation of cancer cells (168). Changes in metabolism have been observed in CRC tumour cells, cancer stem cells, the tumour microenvironment, and host-microbiome interactions (168).

#### *1.10.3.1. Colorectal cancer – faecal metabolomics*

Metabolomics research has focused on a variety of different areas in the characterisation of CRC. To date, a key investigation has been into the identification of adenoma and CRC disease signature biomarkers when compared to non-CRC “healthy” control populations (203) . Here, the emphasis lies on identifying early disease CRC markers. The goal of metabolomics research is to gain a better understanding of CRC biology, to identify mechanistic drivers of carcinogenesis and the hallmark pathways/networks that could be used for early detection or therapeutic purposes.

The faecal metabolome is a highly complex matrix, containing dietary and endogenous substrates, along with common bacterial and mammalian metabolites (33). Dietary and endogenous components include amongst others, indigestible dietary fibre, proteins, choline,

vitamins, and primary bile acids. The common bacterial and mammalian metabolites include methylamines, secondary and tertiary bile acids, carboxylic acids, vitamins, short-chain fatty acids (SCFAs), amino acids, and volatile organic compounds (phenols, ammonia, esters, aldehydes) (204).

The close interlink between the gut microbiome and the faecal metabolome has been widely investigated in CRC (43). For instance, in a cohort of 786 participants, Zierer et al (46) established that the faecal metabolome largely reflects the gut microbial composition, explaining on average 67.7% (+/- 18.8%) of its variance.

Further, according to the faecal Human Metabolome Database (HMDB), intestinal bacterial byproducts account for a large proportion of the faecal metabolome (205). Whilst the analysis of the gut microbiome allows profiling of complex microbial communities, it does not provide data on their actual activity. The study of the gut microbiome cannot indicate the transcriptional activity of the genes or indicate whether microbes or alive or dead (206,207). On the other hand, faecal metabolomics provides a functional readout of microbial activity (208).

Faecal metabolomics in CRC research is frequently carried out through targeted analysis, focusing on abundant and well-characterised metabolites such as SCFAs, bile acids, as well as polar compounds such as amino acids, small organic acids, phenols, sugars, and medium-chain fatty acids (209). Several metabolic pathways altered in CRC have been identified to date, including carbohydrate and amino acid metabolisms, as well as lipid-related metabolic pathways have been implicated in CRC pathogenesis (47,210). A summary of the existing literature in faecal metabolomics studies in CRC to date, along with main findings, is listed in Table 1. The studies discussed are heterogenous and use a variety of sample collection, preparation, analytical techniques. However, several common themes of the CRC faecal metabolome have been identified. A strong microbiome-metabolome correlation exists in faeces from CRC patients (211–213). Further, the faecal metabolomics profiles of patients with CRC tend to differ from the healthy control population (214). On the other hand, fewer studies investigate the differences between adenoma and non-CRC healthy control faecal metabolic profiles (211). In terms of metabolite groups, dysregulation of bile acid metabolism,

changes in amino acid levels, sphingolipids, polyamines, and fatty acids are commonly observed between CRC and control populations(49).

In an untargeted faecal metabolomic study, Kim, et al 2020 (211) demonstrated that polyunsaturated fatty acids (PUFAs), secondary bile acids, and sphingolipids, were elevated in adenoma patients compared to the controls. Furthermore, changes in the relative intensity of PUFAs and sphingolipid were observed in the carcinoma group in addition to the adenoma group (211).

Table 1: Summary of current published research studies in CRC human faecal metabolomics

| Reference | Technique | Cases/controls | Increased levels in cases compared to controls | Decreased levels in cases compared to controls | Main findings and significant differences between cases and controls - no report of direction of increase/decrease |
|---|---|---|---|---|---|
| Bezabeh 2009 | 1H NMR | 523 human subjects (412 controls; 111 colorectal cancer patients) | NA | NA | Glutamic acid, valine,isoleucine, n-butyric acid, and lipids (non-specified) |
| Monleón 2009 | 1H NMR | 32 human subjects (11 controls; 21 colorectal cancer patients) | Proline, cysteine, Leucine | Butyrate, Acetate | |
| Lin 2016 | 1H NMR | 100 human subjects (32 controls; 68 colorectal cancer patients) | Succinate, proline, alanine, dimethylglycine, valine, glutamate, leuine, isoleucine, and lactate | Acetate, Butyrate, Propionate, Glucose, Glutamine | |
| Amiot 2015 | 1H NMR | 55 human subjects (22 controls; 33 advanced colorectal neoplasia patients (including both advanced adenomas and invasive CRC) | SCFAs (valerate, acetate, propionate and butyrate) | β-glucose glutamine glutamate | |
| Batty 2015 | Direct MS | 62 human subjects ( 31 low risk (FOBT positive, no findings of concern in colonoscopy); 31 high risk (FOBT positive, high grade adeoma or adenocarcinoma confirmed) | Tentatively identified hydrogen sulphide, dimethyl sulphide, dimethyl disulphide | NA | Multivariate SWIFT MS modelling shows 75% specificity and 72% sensitivity on FOBT positive samples |
| Weir 2013 | GC-MS | 21 human subjects (10 controls; 11 colorectal cancer patients) | Acetic acid (SCFA) , valeric acid (SCFA), isobutyric acid (SCFA), isovarleric acid (SCFA) | Butyrate (SCFA), oleic acid* (unsaturated fatty acid (UFA)) , linoleic acid* (UFA), elaidic acid* (UFA), glycerol (Polyol), monooleoylglycerol (polyol derivative), ursodeoxycholic acid (bile acid) | |
| Phua 2014 | GC-MS | 21 human subjects (10 controls; 11 colorectal cancer patients) | NA | Fructose, linoleic aicd, nicotinic acid | |
| Wang 2017 | GC-MS | 27 human subjects (12 controls; 15 colorectal cancer patients) | Acetic acid, valeric acid, butyric acid, isovaleric acid, glutamic acid, glycine, aspartic acid, leucine, glycerin, proline, serine, valine, phenylalanine, phenylacetic acid, propionic acid, cholesterol derivatives | Isobutyric acid, oleic acid, linoleic acid, elaidic acid, glycerin, monoacyl glycerol, ursodesoxycholic acid, myristic acid, pantothentic acid | |
| Goedert 2014 | GC-MS; LC-MS | 150 human subjects (102 controls; 48 colorectal cancer) | Faecal heme-related molecules (OR = 17–345), 18 peptides/amino acids (OR = 3–14), palmitoyl-sphingomyelin (OR = 14), mandelate (OR = 3) and p-hydroxy-benzaldehyde (OR = 4) | Acetaminophen metabolites (OR <0.1), tocopherols (OR = 0.3), sitostanol(OR= 0.2), 3-dehydrocarnitine (OR = 0.4), pterin (OR = 0.3), conjugated-linoleate-18-2N7 (OR = 0.2), N-2-furoyl-glycine (OR = 0.3)and p-aminobenzoate (PABA, OR = 0.2) | |
| Sinha 2016 | GC-MS; LC-MS | 131 human subjects (89 controls; 42 Colorectal cancer patients) | p-hydroxy-banzaldehyde, palmitoyl-sphingomyelin | p-aminobenzoate, conugated linoleate | |
| Cubiella 2018 | UPLC-MS | 129 human subjects (49 controls; 80 advanced neoplasia patients (40 advanced adenoms; 40 colorectal cancer) | PC(16:0/16:0), PC(32:1), PC(O-16:0/16:0), PE(16:0/18:1), PE(16:0/18:2), SM(d18:1/16:0), SM(d18:1/23:0), SM(d18:2/24:1) + SM(d18:1/24:0), SM(42:1), Cer(d18:1/24:1) + Cer(d18:2/24:0), Cer(42:1), SM(42:3), ChoE(16:0), ChoE(18:1), ChoE(18:2), ChoE(20:4) | Cer(d18:1/16:0) and TG(54:1) | |

| Reference | Technique | Cases/controls | Increased levels in cases compared to controls | Decreased levels in cases compared to controls | Main findings and significant differences between cases and controls - no report of direction of increase/decrease |
|---|---|---|---|---|---|
| Clos-Garcia 2020 | UHPLC-MS | 245 human subjects (77 controls; 168 colorectal neoplasia (66 advanced adenomas; 99 colorectal cancer patients) | NA | NA | Metabolites from cholesteryl esters (ChoE), sphingomyelins, glycerophosphatidylcholine |
| Kim 2020 | GC-MS | 72 human subjects (40 controls; 30 colorectal cancer patients) | Leucine, isoleucine, alanine, lysine, tyramine, ethanolamine (amino alcohol), phenol (aromatic alcohol), furoic acid (carboxylic acid), succinnic acid (caboxylic acid), oxalic acid (carboxylic acid), hexanoic acid, palimitic acid, oleic acid | Aminoisobutyric acid, butanoic acid | |
| Kim 2020 | UPLC-MS | 240 human subjects (102 controls; 102 advanced adenoms; and 36 colorectal cancer patients) | N-acetyl-cadaverine, bilirubin, linoleoyl ethanolamide, oleoyl ethanolamide, 3-hydroxypalmitate, myristoleate (14:1, n-5), palmitoleate (16:1, n-7), 1-linoleoyl-GPE (18:2), 1-pamitoyl-GPC(16:0), 1-palmitoyl-GPE(16:0), docosahexaenoate (DHA; 22:6, n-3), docosapentaenoate (n-3 DPA; 22:5, n-3), hexadecdienoate (16:2,n-6), 3b-hydroxy-5-cholenoic acid, deoxycholate, n-pamitoyl-sphinganine (d18:0/16:0), hexadecasphinganine (d16:0), hexadecasphingosine (d16:1), sphinganine, sphingosine, piperine, 3,7-dimethylureate | Carotene diol | Adenoma group displayes distinct pathway-level profiles in endocannabinoid, polyunsaturated fatty acid [PUFA], secondary bile acid metabolism, xanthine metabolism, and sphingolipid metabolism at a false-discovery rate (FDR) of 0.1 |
| Yachida 2019 | CE-MS | 406 human subjects (149 control; 45 multiple polypoid adnomas with low grade dysplasia; 20 S0 (intramucosal carcinoma (polypoid adenoma(s) with high-grade dysplasia), stage 0/pTis CRC); 80 SI/II (stage I/II); 68 SIII/IV (stage SIII/IV); 34 HS (control with a history of colorectal surgery)) | Deoxycholate, gychocholate, taurocholate, isoleucine, leucine, valine, phenylalanine, tyrosine, glycine, serine, isovalerate | NA | Metabolites listed demonstrate significant difference between any of control vs MP, control vs S0, control vs SI/II, control vs SIII/IV, etc. |
| Yang 2019 | GC-MS | 100 human subects (50 control; 50 colorectal cancer) | Polyamines (cadaverine, 1,4-Butanediamine), amino acids (Pro,Glu) and urea | Sugars (maltose, fuctose), sugar alcohols, amines (galactosamine) organic and fatty acids (glycerol, octdecanoic acid, hexanedioic acid, benezenepropanoic acid, linoleic acid, oleic acid) | |
| Chetwynd 2019 | NUHPLC | 58 human subjects (26 control; 7 CRC; 25 adenomatous) | Cholesterol oxidation product, porphyrin metabolite, octadecyl lysophosphatidic acid like metabolite, phytyl phosphate, keto bile acid structure, dinor prostaglandin | Androgen sulfate | |
| Song 2018 | GC-MS | 44 human subjects (28 controls; 26 CRC patients) | ω-6 polyunsaturated fatty acid ( linoleic acid (C18:2ω-6)) and monounsaturated fatty acid (oleic acid (C18:1ω-9)) | NA | |

Furthermore, in a targeted analysis of 105 metabolites using UPLC-MS, Cubiella et al 2018 (47), investigated the faecal metabolomic profile of 80 advanced neoplasia patients (40 advanced adenomas and 40 CRCs) and 49 controls. In the univariate comparison of advanced neoplasia (adenoma and cancer) vs control, differences were seen in the sphingolipid family (sphingomyelin (SM) and ceramides (Cer), Cholesteryl Esters (ChoE), Phosphatidylcholines (PC), Phosphatidylethanolamines (PE) and triglyceride (TG) metabolites (47).

Overall, faecal metabolomics analysis is a promising platform for the investigation of CRC. Studies to date have used more traditional hyphenated MS techniques for the investigation of possible disease pathways and mechanisms in CRC as well as biomarker discovery.
However, the translation of metabolomics to the clinical environment has lagged. This delay in the translational potential of the technique can be partially explained by some of the limitations of the existing analytical techniques, such as extensive sample preparation, cost, expertise needed for analysis. Ambient MS, which allows for the analysis of samples in near real-time with little to no sample preparation, is a potential alternative to more time-consuming methods. To date, no published faecal metabolomics study using ambient MS has been carried out for the analysis of adenomas and CRC.

### 1.10.4. Colorectal cancer – faecal lipidomics as a subset of metabolomics

Lipid metabolism plays a crucial role in cancer formation. Lipid dysregulation is a well-established hallmark of carcinogenesis (167). Aberrations in lipid metabolic pathways in CRC are commonly observed and include amongst others synthesis, desaturation, elongation and mitochondrial oxidation of the FAs (167). Lipids are a diverse group of compounds; according to LIPID MAPS (215), they can be classified into eight subgroups, FAs, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids and polyketides (216). Lipids are essential for many metabolic functions such as cellular signalling, bioenergetics, and membrane structure and function (167).

The study of lipids in faeces, faecal lipidomics, is an emerging field, with lipids making up to 15% of the faecal composition. The analysis of lipids has demonstrated applicability in the

assessment of intestinal function by offering a way of assessing gut physiology in a broad and non-invasive manner (217). To date, CRC studies have identified perturbations in bioactivate lipids, including fatty acid metabolism and sphingolipid metabolism, as described above. However, no single marker has seen the translation from bench to bedside.

## 1.11. Inflammatory Bowel Disease (IBD)

### 1.11.1. Background

IBD, including Crohn's disease (CD) and Ulcerative Colitis (UC) and inflammatory bowel disease unclassified (IBDU), is a heterogeneous disease group of the GI tract characterised by relapsing and remitting chronic inflammation often presenting with the varying phenotype (218). It is characterised by a combination of an impaired mucosal immune response to the gut microbiota in a genetically predisposed host (218). IBD generally tends to present itself in people in their 20s and 40s (219). Amongst the European countries, the UK has the highest age-standardised prevalence rate at 449.6 per 100 000 population (220). Further, the prevalence of IBD has increased substantially in many regions (including the UK) from 1990 to 2017. The age-standardised prevalence rate of IBD is higher in females than males (221,222).

IBD presents a challenging disorder in terms of understanding its pathogenesis and developing a cure. IBD imposes significant health and economic burdens and substantially reduces patients' quality of life (221). The symptoms of IBD include abdominal pain, bloody diarrhoea, fatigue, malaise, fever, weight loss, increased frequency of bowel motions, an urgency to open bowels, and occasionally, faecal incontinence (218).

While these symptoms have significant impacts on the patients' health and well-being, they also lead to disruptions of activities, including dietary restrictions and necessary lifestyle changes; IBD symptoms persist throughout life and produce disability and mortality (223). Further, patients with IBD experience periods of relapse and remission; the aim of treatment is both to induce and maintain remission (224). IBD is multifactorial, the aetiology (although incompletely understood) is thought to be influenced by a patient's genetic predisposition,

environment, gut microbiota and immune system (225). The genetic contribution of IBD can be observed in terms of familial aggregation of the disease, as exemplified by 5-23% of IBD patients having an affected first-degree relative (226). Furthermore, the British Twin Study of 1996 illustrated markedly increased concordance rates in identical (monozygotic) twins (CD:35% and UC:11%), providing proof of a genetic component to the underlying IBD pathology (227).

Our understanding of the genetic factors contributing to IBD pathophysiology has seen major advances over the past years. This has been made possible mostly by technological advances in DNA analysis and sequencing, multi-national databases as well as advancements in genetic testing which led to the development of genome-wide association studies (GWAS) which investigate single nucleotide polymorphisms (SNPs) (228). Large scale GWAS have identified over 200 genetic loci associated with IBD, some of which are shared with chronic autoimmune disorders (226).

Caspase recruitment domain gene 15 (CARD15), encodes the nucleotide-binding oligomerization domain-containing protein 2 (NOD2), which was the first identified susceptibility gene for CD in 2001 (229). CARD15 acts as a susceptibility allele, with three polymorphisms (R702W, G908R, and 1007fs) being more prevalent in IBD patients as compared with unaffected controls. NOD2 protein protects against bacteria (pathogen),by eliminating intracellular pathogens at the GI mucosa barrier. (230)

Furthermore, Hampe et al. 2007 (231) and Rioux et al.2007 (232) identified the autophagy 16-likone 1 gene (ATG16L1) as a new IBD locus through GWAS. ATG16L1 is a protein found in the intestinal epithelium and forms part of the autophagy complex which is involved in the clearance of intracellular pathogens (231) . The variant, Thr300Ala has a reduced capacity to capture bacteria and is strongly associated with the incidence of CD (233). Furthermore, the Interleukin-23 receptor (IL23R) gene encodes the subunit of the receptor IL23, a pro-inflammatory cytokine (230). It has been demonstrated that the variant Arg381Gln is associated with a decreased risk of developing CD (234).

Toll-like receptors (TLRs) play a role in the innate immunity and TLR mutations have been demonstrated in IBD (235). Furthermore, human leukocyte antigen (HLA) genes (particularly HLA class II), chief modulators of mammalian immunity, have demonstrated their role in genetic susceptibility to IBD. Multiple studies have reported associations with either UC or CD (236). Genetic research in IBD has offered novel information on the pathogenesis of IBD; however, our understanding of the disease is far from complete. The large number of susceptibility gene loci in IBD indicate that genetics are key in IBD pathophysiology. However, susceptibility loci discovered so far explain less than 25% of the heritability of either CD or UC (237).

Environmental factors play a key role in IBD pathogenesis (238). Key evidence for the impact of environmental factors on IBD is provided by migrant studies. Li et al.al, 2011 demonstrated that first-generation immigrants from low incidence countries continue to have low risk compared to the native population. However, for second-generation immigrants, the risk rises towards that of the native country (239). In a study of immigrants (from low incidence countries) to Canada, Benchiol, et al. 2015 demonstrated that younger age at the time of arrival to Canada increased the risk of IBD to that of the native country population. Further, children born in Canada to immigrants had the same incidence of IBD as native children (240). This suggests that exposure to the environment early in life is important in IBD pathogenesis.

Smoking is another risk factor for IBD (241). Interestingly, smoking is inversely associated with the risk of UC (242), whereas smoking poses an increased risk for CD, as well as worse outcome and an increased need for surgery (243). Vitamin D modulates intestinal inflammation and is increasingly being recognised as important in CD pathogenesis. Ananthakrishnan AN et al., 2012, demonstrated in a prospective study of 72,719 women in the USA Nurses' Health study, that higher levels of Vitamin D status were associated with reduced risk of CD (244). Other environmental factors influencing IBD include among others having had an appendectomy (245,246), breastfeeding (247), urban environment (air pollution (248) and hygiene hypothesis (249)), NSAIDs (250), stress (251), antibiotics (252), diet(253), and exercise (254).

The IBD gut microbiome is characterized by a relative lack of *Firmicutes, Bacteroidetes*, *Lactobacilli,* and *Bifidobacteria* (255,256). Furthermore, in CD, for instance, there is evidence for an increase in pathogenic adherent and invasive *E. coli* (257). Moreover, Manichanh et al (258) demonstrated a decrease in intestinal bacteria in the phylum *Firmicutes*, specifically the *Clostridium leptum* phylogenetic group, in CD.

Additionally, Sokol et al (259) found *F. prausnitzii* low in ileal mucose samples from patients that exhibited endoscopic recurrence six months after surgery for surgical resection for active UC disease (259). Mucosal immunology studies in IBD demonstrate that dysfunctions of innate and adaptive immune pathways contribute to IBD (260), characterised by immune responses to commensal bacteria and chronic intestinal inflammation (256). Studies in adaptive immune responses demonstrate two main mechanisms: CD being driven by a Th1 response, whereas UC is associated with a Th2 response (260). Mucosal innate immune responses are the focus of current immunological research into IBD (260).

### 1.11.2. Crohn's disease (CD) vs Ulcerative Colitis (UC)

CD can affect any part of the GI tract intermittently (most commonly, the terminal ileum or perianal region), is characterised by chronic transmural granulomatous inflammation and a tendency to form abscesses, fistulas and strictures (230). In contrast, UC is an inflammatory disorder often limited to the colonic mucosa and primarily affects the colon, extending continuously from the rectum to proximal parts of the large intestine (230). Furthermore, patients with IBD often suffer from extra-intestinal complications (261).

### 1.11.3. Inflammatory bowel disease unclassified (IBDU)

IBD-unclassified (IBD-U) (around 5–15% of IBD patients) occurs when "endoscopic and histological assessments cannot distinguish between Crohn's colitis and UC"; furthermore, if features are still indeterminate after colectomy histology, the pathology is referred to as indeterminate colitis (262). IBDU remains a poorly understood disease with no definitive histological or clinical features in either children or adults (263). Distinguishing between UC

and CD is important in terms of management and prognosis of the disease, as patients with CD have a higher risk of strictures and fistula formation that may lead to a more severe disease course as well as an increased risk of surgical intervention. A better understanding of the underlying pathophysiology of the distinct diseases of IBD is of pivotal importance for tailored clinical management and, ultimately patient outcome (263).

### 1.11.4. Diagnosis, assessing disease severity, and evaluating treatment response

#### *1.11.4.1. Current guideline in diagnosis, disease severity assessment and treatment response evaluation*

IBD diagnosis is based on the combination of clinical, laboratory, radiological, endoscopic, and histological criteria (264,265). In IBD, symptoms include diarrhoea, weight loss, abdominal pain, or cramping, bloating, lethargy, fevers, night sweats, anaemia, and constipation (262,263). Several disorders can mimic IBD clinically, including CRC, coeliac disease, endometriosis, ovarian cancer, and irritable bowel syndrome (IBS) (266).

Patients presenting with lower GI symptoms, red flag indicators such as anaemia, rectal masses, inflammatory markers in IBD and late-onset (>60 years of age) change in bowel habits warrant referral to secondary care for further investigation (267). Tests carried out are based on likely differential diagnosis, which may include full blood count, erythrocyte sedimentation rate, C-reactive protein, endomysial antibodies, tissue transglutaminase antibodies, as well as faecal calprotectin (267).Endoscopic investigation and histology results are the current gold standard for the diagnosis and assessment of IBD (268). In addition, endoscopy is also used for the determination of the extent, location, and severity of the disease, as well as to differentiate between CD and UC (224). However, it is invasive, time-consuming, expensive, and often not very well accepted by patients, especially in the paediatric population (269). Many patients suspected of having IBD will have functional disorders and negative findings on endoscopic investigations. Further, it is not without risk, although serious complications are rare (270).

IBD is a severe and progressive disease, quick referral filtering of patients with a high probability of IBD is vital (271). A need for the identification of biomarkers that can accurately identify IBD from other common aetiologies, distinguish between CD and UC, as well as establish disease activity and treatment response exist.

The Montreal classification (272) in adults and the Paris classification (273) in children have emerged as useful tool for ascribing define disease extent according to endoscopic or macroscopic features; they are useful for service delivery and research. IBD disease activity can be endoscopically determined or in the form of clinical and biochemical measures of disease severity (274). The therapeutic target of IBD has shifted in recent years from symptomatic clinical remission towards the resolution of disease assessed endoscopically, that is, complete mucosal healing to prevent structural gut wall damage (275).

Non-invasive assessments, such as clinical estimates, are used for disease monitoring; however these are often inaccurate (276). Often a combination of markers such as faecal calprotectin, endoscopy, cross-sectional imaging is used (224). To ensure tracking disease activity reliably during the disease, a need for improved objective non-invasive means of disease activity monitoring exists.

#### *1.11.4.2. Biological markers*

##### 1.11.4.2.1. Serological markers

The main systemic markers used today, including erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), anti-neutrophil cytoplasmic antibodies (ANCA) and anti-Saccharomyces cerevisiae antibodies (ASCA), have low diagnostic accuracy for intestinal inflammation (277,278). CRP and ESR are non-specific markers of inflammation that may be elevated in patients with IBD; however, these are not IBD specific (279). Further, up to 50% of patients with an active UC may show normal CRP levels (280).

#### 1.11.4.2.2. Faecal markers

Several faecal markers for IBD have been investigated, including lipocalin-2, lactoferrin and faecal calprotectin (FC) (281). One of the most widely and promising tests to date is FC, a calcium-binding protein (282). Activated neutrophils release FC at the site of inflammation; FC correlates with levels of inflammation (283). The protein appears stable at room temperature for up to seven days, is homogeneously distributed in faecal samples, and is resistant to digestion (282,284).

FC is approved by the Food and Drug Administration, however, clear guidance on which settings it should be used in is lacking (282). In the UK, NICE guidelines recommend FC testing only for differentiating between IBD in primary care, for patients with lower GI symptoms when referral for secondary care is considered, and cancer is not suspected (285). FC demonstrates excellent diagnostic accuracy for differentiating organic GI disorders (IBD, CRC, etc.) from functional disorder (IBS) (286).

A study of 602 patients at King's College London investigated faecal calprotectin as a tool for differentiating organic GI disorders from non-organic IBS. It was demonstrated that FC had 89% sensitivity and 79% specificity for detecting organic bowel disease, which included, amongst others CD, UC, cancer, diverticular disease, and coeliac disease (287). Further, in a meta-analysis of 2475 patients, FC demonstrated a sensitivity of 83% (95% CI, 81–84%) and specificity of 84% (95% CI, 82–85%) for distinguishing organic from non-organic disease (288).

FC increases after the use of non-steroidal anti-inflammatory drugs (NSAIDs), including aspirin (288). Further, it has been demonstrated that FC levels change with age, leading to a debate on using different cut off values for the adult and paediatric population (288). FC is a sensitive marker for distinguishing organic GI disorders from functional disorders, such as IBS (286). However, multiple diseases can increase FC, and therefore, its specificity for IBD is lower than desirable, FC is inflammation-and not disease-specific (289).

On the other hand, very high FC values (over 500 to 600 µg/mg) are very predictive of IBD or food infections. However, there are no fixed rules for calprotectin values or cut-off point for

IBD diagnosis (287). Therefore, a high faecal calprotectin is not used as a diagnosis but rather shifts the clinical decision making in favour of an endoscopic examination (288). Overall, the main disadvantage of FC is that it is inflammation and not IBD specific, and evidence is lacking for its ability to distinguish CD and UC (290). A need for the development of more accurate biomarkers for IBD diagnosis and the differentiation of CD versus UC exist.

### 1.11.5. Metabolomics and lipidomics in IBD

Metabolomics research has focused on a variety of different areas in the characterisation of IBD (281,291). To date, the most frequent investigation has been into the identification of IBD disease signature biomarkers when compared to non-IBD "healthy" control populations (291). Further work investigated metabolic differences related to disease subtype and activity, location of disease, with a few studies also monitoring treatment response in IBD patients (263).

#### *1.11.5.1. Metabolomics in IBD disease signature biomarker identification*

In IBD metabolomics research, a number of biofluids, including serum, plasma, urine, faeces, biopsies, isolated colonocytes and lymphocytes, have been investigated (291). Analysis of IBD serum or plasma samples indicate that both CD and UC are implicated in lipid and amino acid metabolism (292–295). In the urinary analysis of IBD, a number of studies differentiated between IBD and non-IBD controls based mainly on the differences in amino acids and amino acid-derived metabolite levels (292,293,296,297). Interestingly, lower levels of hippurate, a mammalian-microbial co-metabolite, were found in IBD patients as compared to non-IBD controls (292,293,296,297). Another interesting direction of IBD metabolomics research has been the investigation of breath metabolites as a potential diagnostic apparatus for the identification of disease-specific markers (298). For instance, Van Malderen et al. 2020 (299) utilises volatomics to differentiate IBS, IBD and healthy volunteers from breath with accuracies of 70%-100%.

#### *1.11.5.2. Faecal metabolomics*

GI microbiota dysbiosis is believed to play an important role in the pathogenesis of IBD (225). The dysbiosis is characterised by restricted biodiversity and temporal instability has been reported in IBD (300). The IBD gut microbiome exhibits a general decrease in taxonomic diversity relative to non-IBD controls, in addition to phylum-level decreases in *Firmicutes* and increases in *Proteobacteria* (301–303). Further, a close relationship between IBD-associated microbiota and the IBD faecal metabolome has been established. The faecal metabolome in IBD is characterised by a decrease in butyrate and a decrease in butyrate-producing bacteria (304). A summary of the current literature to date in faecal metabolomics research in IBD is provided in Table 2.

Several common themes were identified in studies described in Table 2: Faecal metabolomics, as a classification tool, demonstrates potential for differentiating IBD and control study populations, Table 2. On the other hand, it is less successful at discriminating CD and UC (305). Dysregulation of bile acid metabolism, changes in amino acid levels, sphingolipids, polyamines, and medium-chain fatty acids, as well as SCFAs, have been consistently identified in the comparative analysis of IBD vs health, see Table 2.

Table 2: Summary of current research studies in human faecal metabolomics in IBD

| Reference | Technique | Cases/controls | Increased levels in cases compared to controls | Decreased levels in cases compared to controls | Main findings and significant differences between cases and controls when no report of direction of increase/decrease |
|---|---|---|---|---|---|
| Marchesi 2007 | 1H NMR | 33 human subjects (13 controls; 10 UC; 10 CD) | IBD vs control: valine, lysine, leucine, isoleucine, alanine | Control vs IBD: butyrate, acetate, methylamine, trimethylamine | Good discrimanatory power between IBD and control |
| Le Gall 2011 | 1H NMR | 45 human subjects (22 controls; 10 IBS; 13 UC) | UC vs control: taurine, choline, glucose, cadaverine IBS vs control: choline, taurine, glucose, cadaverine | Control vs UC: 2-methylbutyrate Control vs IBS: 2-methylbutyrate, isovalerate, isobutyrate | Good classification for UC vs control |
| Bjerrum 2015 | 1H NMR | 113 human subjects (21 controls;48 UC; 44 CD) | Active CD vs control: isoleucine, leucine, valine, lysine, alanine, phenylalanine<br>Active UC vs controls: leucine, valine, lysine, alanine<br>Active UC vs inactive UC: isoleucine, leucine, valine, lysine, alanine, lactate, taurine<br>Active UC vs control: isoleucine, leucine, valine, lysine, alanine, lactate, taurine | Control vs active CD: Butyrate | Ability to discriminate between IBD and controls as well as IBD subtypes when patients on biologics followingsurgery were included; The metabolites holding differential power primarily belonged to a range of amino acids, microbiota-related short chain fatty acids, and lactate |
| Jansson 2009 | Direct MS | 36 human subjects (22 control; 8 predominantly colonic colitis (CCD); 6 predominantly ileal colitis (ICD)) | NA | NA | Able to discriminate ileal CD and colonic CD from health (control); Pathways with differentiating metabolites included those involved in the metabolism and or synthesis of amino acids, fatty acids, bile acids, and arachidonic acid |
| Walton 2013 | GC-MS | 87 human subjects (19 control; 22 CD; 20 UC; 26 IBS) | Control vs pre-treatment CD: butanoic acid; 1-propanol; propanoic acid, ethyl ester; butanoic acid, methyl ester; butanoic acid, ethyl ester; indole; 1-butanol; butanoic acid, 3-methyl; phenol<br>Control vs. post-treatment CD: propanoic acid, ethyl ester | NA | CD samples showed significant elevations in the concentrations of ester and alcohol derivates of short-chain fatty acids and indole compared with the other groups |
| Cauchi 2014 | GC-MS | 91 human subjects (20 control, 24 CD, 19 UC, 28 IBS) | NA | NA | Faecal sample analysis via GC-MS demonstrated a discrimination power between CD and healthy controls with an overall classification accuracy of 85 % (78 % specificity; 93 % sensitivity) |
| De Preter 2015 | GC-MS | 204 human subjects (40 controls; 83 patients with CD; 68 with UC; 13 with pouchitis) | CD vs (healthy control, pouchitis and UC): 1-ethyl3-methylbenzene, benzene acetaldehyde, phenol, 2-methyl propanal, carbon disulfide and 1-methoxy-4-methylbenzene<br>(UC and Pouchitis) vs (healthy control and UC): cyclohexane, 3-methyl butanal, and pyrrole | Healthy controls vs (CD, UC and pouchitis): Medium chain fatty acids pentanoate, hexanoate, heptanoate, octanoate and nonanoate; butanoic acid; protein fermentation metabolites (3-methyl1H-indole, p-cresol, dimethyl sulfide, 2-methyl butanoate, methyl propyl sulfide and methyl-2-propenyl disulfide)<br>(Healthy controls + pouchitis) vs CD: 2-pentyl furan, 3-carene and α-methylstyrene | Good classification of all IBD types |
| Ahmed 2016 | GC-MS | 328 human subjects (109 controls; 117 CD; 100 UC) | Active CD vs inactive CD + control: Heptanal, 1-octen-3-ol, 2-piperidinone, 6-methyl-2heptanone | Inactive CD +control vs active CD: methanethiol, 3-methyl-phenol, SCFAs and ester derivatives | NA |
| Santoru 2017 | GC-MS; LC-MS; 1H-NMR | 183 human subjects (51 control; 82 UC; 50 CD) | CD vs control: Glyceric acid, beta-alanine, L-alanine, phenylacetic acid, cadaveirnes, putrescine, phenylethylamine, phenylalanine, 4-Hydroxyphenylacetic acid, DG (18:0/22:2)<br>UC vs control: Tyramine, Phenylalanine, TMAO, NAPE (18:1/16:1/18:0), DG (16:0/18:2), DG (18:0/22:2), D-Glucose, Cadaverine, Beta-alanine, 4-hydroxyphenylacetic acid, 5-aminovaleric acid, L-Glycine | Control vs CD: 5beta-Coprostanol, methyladipic acid, nicotinic acid, and pathothenic acid, methylamine, 3-hydrobutyric acid, 2-hydroxy-3-methylvaleric acid, citric acid, hydrocinnamic acid, urobilinogen, urobilin, PS (22:2/18:0), PA (19:0/16:1), PC(16:0/3:1)<br>Control vs UC: 5beta-Coprostanol, methyladipic acid, nicotinic acid, pathothenic acid, sebacic acid, hydrocinammic acid, 3-hydroxybutyric acid, linoleic acid, trycarballyliic acid, methylamine, pyroglutamic acid, 2-hydroxy-3-methylvaleric acid, L-Glutamic acid, citric acid, PC(16:0/3:1), Urobilinogen, PS (22:2/18:0), Cer (18:1/22:0), PC (22:2/14:1) | Able to discriminate between IBD and health but not clearly between IBD subtypes; metabolites including biogenic amines, amino acids, lipids, were signifcantly increased in IBD, while others, such as two B group vitamins, were decreased in IBD compared to healthy subjects |

| Reference | Technique | Cases/controls | Increased levels in cases compared to controls | Decreased levels in cases compared to controls | Main findings and significant differences between cases and controls when no report of direction of increase/decrease |
|---|---|---|---|---|---|
| Jacobs 2016 | LC-MS | 90 human subjects (54 control (control = healthy siblings/parents of IBD subjects); 26 CD in remission; 10 UC in remission) | CD vs non-IBD control: Amino acid derivatives (phenylethylamine and N-acetylcadaverine) and bile acids (cholic acid, 7-ketodeoxycholic acid, chenodeoxycholic acid sulfate, and 3-sulfodeoxycholic acid) | Non-IBD control vs CD + UC: heme degradation product (stercobilin), a glutamate derivative (acetyl-glutamic acid), and a product of steroid catabolism (boldione) | NA |
| Kolho 2017 | LC-MS | 98 human subjects (29 controls; 69 IBD) | NA | NA | Faecal metabolites provide significant improvement in classification accuracy over serum metabolites, classification between subtypes; many faecal metabolic pathways affected in UC (none significantly in CD), including amino acid, sphingolipid and urea cycle |
| Franzosa 2019 | LC-MS | 155 human subjects (34 controls; 68 with CD; 53 with UC) | CD vs control: Sphingolipids; carboximidic acids and derivatives; biles acids, alcohols, and derivatives; organonitrogen compounds, cholesteryl esters, phenylacetamides, phosphatidylcholines, alpha-amino acids<br>UC vs control: Sphingolipids; carboximidic acids and derivatives; biles acids, alcohols, and derivatives; organonitrogen compounds, cholesteryl esters, phenylacetamides, phosphatidylcholines, alpha-amino acids | Control vs CD: long-chain fatty acids; fllavonoids; indoles and derivates; cinnamic acid and derivatives; triacylglycerols; tetrapyrroles and derivatives; triterpenoids; alkyl-phenylketones; brassinolides and derivatives: ergosterols and derivatives; quinolones and derivatives; vitamin D and derivatives; stigmastanes and derivatives; lactones and derivatives; beta-diketones; cholesterols and derivatives; phenylbenzodioxanes<br>Control vs UC: long-chain fatty acids; fllavonoids; cinnamic acid and derivatives; triacylglycerols; triterpenoids; alkyl-phenylketones; ergosterols and derivatives; vitamin D and derivatives; stigmastanes and derivatives; lactones and derivatives; cholesterols and derivatives; phenylbenzodioxanes | Able to discriminate between IBD and health, less so for IBD subtypes |
| Duboc 2012 | LC-MS | 71 human subjects (29 controls; 23 flare-up IBD; 19 remission IBD) | Active IBD flare vs (Remission IBD + control) = Proportion of conjugated bile acids as well as 3-OH-sulphated BA increased | Remission IBD + cotrol vs active IBD flare = Decrease in proportion of secondary bile acids | NA |
| El Hassani 2019 | GC-MS | 20 human subjects (10 controls; 5 UC; 5 CD) | NA | NA | Intention-to-diagnose cohort - A significant difference was demonstrated for faecal (p-value, area under the curve; 0.038, 0.73) volatile organic compound profiles between IBD and controls |
| Garner 2007 | GC-MS | 101 human subjects (30 controls, 18 UC; 31 *C. jejuni* , 22 *C. difficile* ) | UC vs control: Alkenes | Control vs UC: Nitrogen-containing compounds | Able to discriminate between different disease types |
| Lloyd-Price 2009 | LC-MS | 132 human subjects (27 controls, 38 UC, 67 CD) | IBD vs control: Polyunsaturated fatty acids, primary bile acids, putrescine, taurine and nicotinuric acid | Control vs IBD: Vitamins B3 and B5, SCFAs, indole-3-propionate and secondary bile acids | Longitudinal multi-omics study on IBD; metabolomics offered the best discrimination between health and IBD; lower diversity of metabolites in IBD; decreases in levels of vitamins B3 and B5, SCFAs, indole-3-propionate and secondary bile acids |

The first studies in IBD by Marchesi et al. (306) and Bjerrum et al. (307) demonstrated high amino acid and low SCFA levels in faecal extracts of IBD patients versus non-IBD controls using the $^{1}$H-NMR. Moreover, Le Gall et al. (308) showed a reduction in SCFA and branched-chain fatty acids (BCFA) levels in IBD using faecal $^{1}$H-NMR.

Alterations of sphingolipid metabolism have been observed in the faecal metabolome of IBD patients (335). Sphingolipids perform important signalling functions, which are important in cellular events, including intestinal cell survival, growth, differentiation, and apoptosis. (311). Intestinal sphingolipids have been shown to play a role in immunity and inflammatory disorders, and sphingolipid metabolism has been demonstrated to be disrupted in IBD (304,309,310).

Furthermore, Duboc et al., 2013 (300) demonstrated that the proportion of conjugated BAs increased in IBD patients compared with control and that the proportion of secondary BAs decreased in IBD patients, particularly during a flare. Higher levels of 3-OH-sulphated BA were found in active IBD compared to IBD remission and control (300). Overall, impaired BA metabolism, characterised by defective deconjugation, transformation and desulphation, was observed in IBD-associated dysbiosis (300).

Furthermore, higher levels of protein- or amino acid metabolites (e.g. indoles), ester and alcohol derivates of SCFA and medium-chain fatty acids have been identified in faeces from IBD patients versus control patients (295). Additionally, metabolites involved in the synthesis and metabolism of amino acids, fatty acids, bile acids and arachidonic acid were found to be significantly different between non-IBD control versus CD patients (112).

## 1.12. Summary

Faecal metabolomics poses a potential avenue for both clinical biomarker research and for gaining a better understanding of pathological mechanisms (291). However, current techniques are limited in their use as high-throughput screening tools for GI-related diseases due to extensive sample processing times or lack of sensitivity (61). Furthermore, optimisation of experimental study design and parameters are crucial for the successful

translation of scientific results from bench to bedside and are often lacking in metabolomics studies to date. Moreover, patient- as well as sample-related modifiers of the faecal metabolome need to be explored in the context of investigating differences between disease and healthy states as these may act as confounding factors in the analysis.

LA-REIMS demonstrates potential as a platform for the non-invasive high-throughput direct-from-sample investigation of the faecal metabolome in GI diseases (151). Furthermore, the need for minimal sample preparation allows the investigation of metabolomic processes and networks underlying pathogenesis in the sample's native form and potential use in rapid diagnostics. LA-REIMS provides an avenue for investigating the carcinogenic and inflammatory pathways in the GI tract. Investigations into REIMS in faecal metabolomics have been limited to proof-of-concept studies, with lack of optimisation, small study numbers, the use of monopolar probes instead of a laser for analysis and limiting the analysis to faecal sample mostly ionised in negative mode.

The approach taken here in optimising LA-REIMS as a tool for faecal metabolomics has involved the careful investigation of optimal laser and instrumentation parameters for faecal samples and faecal water analysis (2:1 ratio water: faeces). The optimised settings were then utilised in the analysis of a CRC patient and IBD patient cohort. The effect of variable faecal water content and the heterogeneity of feature of interest in faecal samples was investigated using LA-REIMS.

Here, an untargeted metabolic fingerprinting approach was taken to establish the diagnostic accuracy of REIMS for CRC and IBD. In CRC, LA-REIMS was also compared to FIT in terms of accuracy of adenoma and CRC detection. Validation of tentative metabolite identities was attempted using more traditional hyphenated MS techniques, such as LC-MS/MS, were utilised. Furthermore, major patient related modifiers of the LA-REIMS faecal metabolome were explored.

# 2. Hypothesis and aims

## 2.1. Hypothesis

Optimised LA-REIMS analysis of faecal biospecimens allows for direct metabolic fingerprinting and demonstrates the diagnostic potential for CRC and IBD.

## 2.2. Aims

1. To optimise LA-REIMS as an analytical tool for faecal metabolomics analysis

2. To investigate sample-related major modifiers of faecal LA-REIMS metabolomic analysis, including the effect of the water content of samples on analysis and the spatial distribution of important metabolites in whole faecal samples

3. To implement the optimised LA-REIMS settings in the analysis of a colorectal cancer patient cohort
   A. To determine the diagnostic accuracy of optimised LA-REIMS in
      a. Adenoma vs control
      b. Cancer vs control
      c. Disease (adenoma + cancer) vs control
      d. Adenoma *vs* cancer
   B. To compare the diagnostic accuracy of LA-REIMS for adenomas and CRC to the current screening technique used in the Bowel Cancer Screening Program, namely FIT for a-c
   C. To determine the metabolic signature of adenomas and CRC and identify discriminatory metabolites for both and validate these using LC-MS/MS analysis
   D. To investigate patient-related major modifiers of the LA-REIMS CRC faecal metabolomic analysis

4. To implement the optimised LA-REIMS settings in the analysis of inflammatory bowel disease (IBD) patient cohort
    A. To determine the diagnostic accuracy of optimised LA-REIMS for IBD (as compared to non-IBD controls)
    B. To determine the accuracy of optimised LA-REIMS for Crohn's disease vs Ulcerative Colitis (UC)
    C. To determine the metabolic signature of CD and UC and to identify discriminatory metabolites for both and validate these using LC-MS analysis and tandem mass spectrometry
    D. To investigate patient-related major modifiers of faecal LA-REIMS metabolomic analysis for IBD

# 3. Materials and Methods

## 3.1. Health and Safety

All experiments were carried out in line with Imperial College Health and Safety regulations and after the completion of relevant risk assessment for work involving biological material (Form Bio1, project assessment ID: IC-4987 v3). All samples related to this project were treated as Hazard Group II materials and handled in a containment level 2 facilities. LA-REIMS analysis was carried out in a custom-made class 2 biological safety cabinet to ensure appropriate safety procedures to minimize exposure to aerosolised material and ensure containment of the category 4 laser. Emergency and decontamination procedures were in place as well as occupational health monitoring. All substance classed as hazardous were handled in line with Control of Substances Hazardous to Health (COSHH) regulations.

## 3.2. Ethics

The research was conducted in accordance with the recommendations for physicians involved in research on human subjects adopted by the 18th World Medical Assembly, Helsinki 1964 and later revisions and in accordance with the guidelines laid down by the International Conference on Harmonisation for Good Clinical Practice (ICH GCP E6 guidelines). All participants provided signed consent to enter the study for research projects outlined.

## 3.3. Standard laboratory protocol

### 3.3.1. Faecal sample storage and homogenisation

All faecal samples utilised were homogenised before freezing by stirring using a spatula or mixer. Once faecal samples were obtained, they were stored at -80°C at all times apart from LA-REIMS analysis and freeze-thaw cycles were minimised.

### 3.3.2. LA-REIMS analysis

MS analysis was performed on a Waters Xevo G2-XS QToF mass spectrometer (Water Corporation, USA) connected to a REIMS atmospheric interface ion source (Water Corporation, USA). All analysis was carried out according to the instrument parameters outlined in Table 3.

Table 3: Instrument parameters for LA-REIMS analysis

| | |
|---|---|
| Source | REIMS |
| Mass analyser | Time of Flight |
| Scan time | 1000 ms |
| Scan mode | Sensitivity |
| Ion mode | Negative and Positive |
| Mass range | *m/z* 50 – 1200 (*m/z* varies in some experiments) |
| Sampling cone | 80 V |
| Source offset | 50 V |
| Source temperature | 100C |

The instrument was calibrated every day prior to a sampling run using 5 mM sodium formate solution (Sigma-Aldrich, UK) at a rate of 0.25 mL/min for up to 5 seconds via a syringe pump. Calibration was performed across the 50-1200 m/z range in both negative and positive ion detection mode using the manual Intellistart method through MassLynx (Waters, USA), with pass parameters of <3 ppm RMS residual mass and >34 reference peaks detected.

To account for mass drift, leucine enkephalin lock mass compound (Sigma-Aldrich, Dorset, UK) was infused throughout the duration of each experiment through a HPLC pump (Knauer, Northampton, UK). The instrumentation optimisation work yielded the following optimum flow rates which were implemented into future analyses: Faecal sample negative and positive

mode at a solvent flow rate of 0.25 mL/min. Faecal water negative mode at a rate of 0.3 mL/min and faecal water extract positive mode at a rate of 0.35 mL/min. The leucine enkephalin stock solution was made using HPLC grade 2-propanol (ROMIL Cambridge, UK) and stored at -20°C. A 10 nG/mL working solution was made fresh before the start of each experiment and stored in an insulated cool box during pumping to prevent degradation of the compound and discarded at the end of each day.

LA-REIMS analysis was automated via the use of a modified custom-made liquid handling and colony picker robot (Freedom EVO 75/TECAN, Switzerland). Further, a cooling rack for the analysis of well plates was fitted. The sample to be analysed were placed onto the light stage Pickolo platform (SciRobotics, Israel)) or the cooling platform for the 24- or 96- well plates and the script started through EVOware following Tecan initialisation. The robotic probe arm of the modified liquid handling robot (Freedom Evo 75, TECAN, Switzerland) was equipped with the 10.6 µM carbon dioxide laser lens assembly (FELS-25A, OmniGuide, USA). The Class 4 laser was retrofitted into the TECAN housing using a fibre optic beam delivery system routed through the robotic arm. The Class II biological safety cabinet was also modified to contain diffuse reflections by fitting a louvre cover over the viewing panel. The Omniguide module was connected to a helium gas cylinder, for cooling purposes, with a supply pressure of 2 bar during operation.

During LA-REIMS analysis, a $CO_2$ laser beam is passed from the OmniGuide FELS-25A® and is focussed in a modified Aesculight® lens cell into a spot size of ≈ 500 µM. A continuous burn is produced aimed at centre of each well with a laser set to fire at a depth of 2 mm set distance away from the sample, Figure 3. Scripts for running of the automated TECAN platform were written in Freedom EVOware software (Tecan Group Ltd., Switzerland) according to the requirements of each experiment by Dr Simon Cameron and Alvaro Perdones-Montero.

EVOware scripts were written to accommodate analysis from agar media, 24-well plates, 96-deep well plates and to automatically guide the laser fibre optic, activate the beam and start the acquisition in MassLynx. A wait period was inputted to allow vapour dispersal between sample replicates: Instrumentation optimisation identified 20 seconds as optimum time

between samples for faecal sample negative and positive mode and faecal water sample mode, for faecal water negative mode, 10 seconds were sufficient.

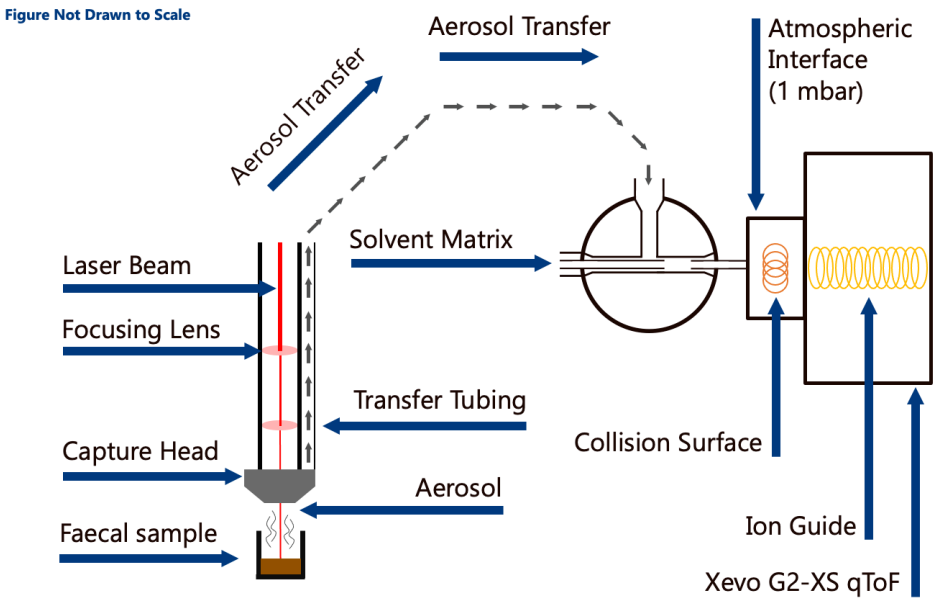



Figure 3: Overview of simplified high-throughput LA-REIMS setup for faecal metabolomics analysis. Adapted with permission from Cameron et al. 2019 (312); Copyright 2021 American Chemical Society see section *Appendix 1: Permission to reprint*.

### *3.3.2.1. LA-REIMS negative and positive mode analysis switch*

In the studies outlined in this thesis, positive and negative ion detection mode analysis were carried out. The order of analysis was randomised and samples in each well were stirred using a spatula between changing from negative to positive mode, and vice versa. Additionally, the laser lens was cleaned, and the tubing checked for any visible contaminants between ionisation mode and plate changes.

### *3.3.2.2. LA-REIMS cleaning protocol*

The REIMS atmospheric ion source interface included an isolated collision surface chamber, allowing removal of the inlet capillary and collision surface for cleaning without venting the

instrument vacuum. However, as part of a general maintenance routine and to clean possible sources of contamination, the instrument was vented once a week and the REIMS interface was removed to allow for in depth cleaning.

After removing source housing, the capillary was removed and placed into a measuring cylinder which was filled with a 1:1 ratio of 2-propanol (ROMIL, Cambridge, UK) and HPLC grade water (VWR, Leicestershire, UK). The StepWave ion guide was also removed and placed into a measuring cylinder and immersed in the solvent mixture described above. Both measuring cylinders were placed in ultrasonic bath and sonicated for 10 minutes to clean. The housing, collision surface and StepWave surroundings were also wiped clean using cotton swabs or Kimwipes (VWR,Leicestershire, UK). Once the StepWave and interface were replaced the instrument was pumped down again for use the following day.

Each day after calibration, a reduced cleaning protocol was implemented, by subjecting the T-piece and inlet capillary to the same procedure described above. The laser lens was also cleaned daily following each experiment by sonicating or wiping with cleaning solution as required. Further, the plastic laser lens cover was removed after each experiment and submerged in 2-propanol before placing it in ultrasonic bath for 10 minutes. As with the other methods, transfer tubing was replaced or flushed with 2-propanol to clean. The instrument inlet capillary and T piece were also submerged in a 1:1 ratio of 2-propanol (ROMIL, Cambridge, UK) and HPLC grade water (VWR, Leicestershire, UK) after each experiment to remove residue.

### 3.3.3. Data processing and statistical analysis

#### *3.3.3.1. Pre-processing*

Data files were generated as raw files in MassLynx® V4.1 (Waters Corporation) during sample analysis. The software allows for visualisation and manual inspection of spectra using the accurate mass measure function from the tool page. To view files, input files are loaded, automatic peak detection parameters set to automatic, with selected corresponding lock mass correction and background subtraction.

### *3.3.3.2. Peak picking*

#### 3.3.3.2.1. Binning

For further processing using binning, raw data as well as metadata were imported into Offline Model Builder (OMB) (Waters Corporation). For each file, sections of the total ion current (TIC) chromatogram, corresponding to each burn/analysis, were selected. In a typical experiment, each file contained one burn/sample analysed and was saved accordingly. Data was lock mass corrected, background subtracted, and re-binned to 0.1 Da over the *m/z* range 50-1200. Pre-processed data was exported from OMB as .csv data files.

#### 3.3.3.2.2. Peak picking

For further processing using peak picking, the in-house built peak picking package (313) developed by Alvaro Perdones Montero for $CO_2$ LA-REIMS spectra was utilised. The data was imported into RStudio v1.1.419 using raw data as well as metadata in .csv format. The metadata contained the file name, class, sample number and the start and end scan to be selected for analysis. The m/z range to be analysed was dictated by the selected range during analysis but could be amended later (using the output file). The main script (according to developer installation and running instructions) was carried out for each dataset individually. The main components of the pipeline included noise filtering, peak selection, and lock mass correction with an output file of the processed data provided as a .csv data file, see Figure 4. The noise filtering function allowed for the selection of one of three filtering options, i.e., no filtering, soft filtering, and hard filtering, ranging from least to most stringent denoising parameters. Furthermore, peak selection allowed for the choice between a classification approach or an on-off approach. The classification peak selection was utilised in the processing of data aimed at multivariate analysis and classification modelling and utilised no a-priori information about samples. The on-off peak selection approach was suitable for univariate analysis such as ANOVA, ROC analysis, etc. It creates a peaks profile based on subclasses and utilises a-priori information about samples.

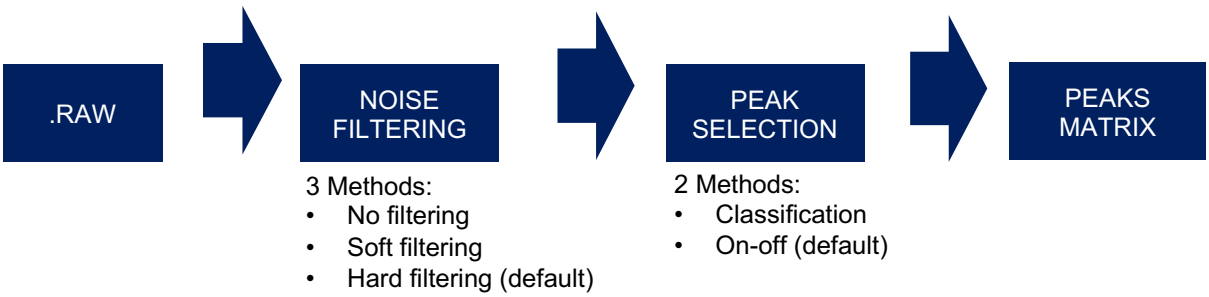



Figure 4: Peak picking pipeline (reproduced and amended from (314))

Peak picking is the preferential analytical peak selection technique to binning when biomarker discovery and bio-chemical information recovery is driving the research. This is particularly the case in LA-REIMS data, where binning has been a particular challenge due to the high prevalence of analytical interference in the data, masking some real peaks. The challenge of using binning in data with high degree of interference is the selection of real peaks among the noise for further analysis.

#### *3.3.3.3. Lock mass correction*

The lock mass compound of choice for LA-REIMS analysis is leucine enkephalin (Sigma-Aldrich, UK). Leucine enkephalin stock solution is made using HPLC grade 2-propanol (ROMIL Cambridge, UK) and stored at -20°C until analysis. For LA-REIMS analysis, a 0.1 nG/mL working solution is made fresh before the start of each experiment and stored in an insulated cool box during pumping to prevent degradation of the compound and discarded at the end of each day. Leucine enkephalin has a *m/z* of 554.2516 in negative ion detection mode and *m/z* 556.2775 in positive ion detection mode. Leucine enkephalin is an extensively studied pentapeptide that has become one of the most-commonly used reference compounds in MS (315). The *m/z* of leucine enkephalin is in the middle of the range of interest (*m/z* 50-1200) in LA-REIMS faecal metabolomics research. Therefore, it is a useful compound for the correction of mass drift in both the low mass (*m/z* 50-550) and complex lipid region (*m/z* 600-1200).

Further to utilising leucine enkephalin, potential alternative internal lock mass compounds were investigated. Due to the nature of these experiments being limited to only faecal sample analysis, an internal lock mass was sought, i.e., a compound that was present in high intensity in all faecal samples identified. The selection criteria were the presence of the compound/ion throughout all spectra from the existing faecal sample datasets and a relatively constant intensity deemed as x $10^4$ or higher. The list of internal lock mass compounds is provided in (Table 4). The features are present in the spectra of faecal samples throughout all datasets. Furthermore, Cameron et al 2018 (149), observed these features in an independent sample set using faecal LA-REIMS analysis and tentatively identified them.

Table 4: Features for internal lock mass correction with m/z peaks, tentative ID and ionic species in negative ion detection mode

| m/z | Tentative ID | Ionic Species |
|---|---|---|
| 241.216 | FA(15:0) | M-H |
| 255.232 | FA(16:0) | M-H |
| 279.233 | FA(18:2) | M-H |
| 281.248 | FA(18:1) | M-H |
| 283.263 | FA(18:0) | M-H |
| 297.244 | FA(18:1(OH)) | M-H |
| 299.258 | FA(18:0(OH)) | M-H |
| 339.326 | FA(22:0) | M-H |

It must be noted that the features of interest range from *m/z* 240-340 and therefore are best suited for the correction of the low mass range (*m/z* 50-550). For the analysis of the complex lipid *m/z* range, the internal lock mass compounds identified are less reliable. An attempt at identifying features present throughout all faecal samples in the complex lipid range, to allow more accurate re-calibration of this *m/z* region, was unsuccessful. Even though features are present in the *m/z* range 600-1000, no single feature was consistently observed throughout the spectral data. In the context of this work, for experiments with minimal variation in leucine enkephalin intensity, it was utilised for lock mass correction. However, for the IBD and

CRC analysis, where significant leucine enkephalin signal variation was observed, internal lock mass features where utilised.

#### *3.3.3.4. Data analysis – Metaboanalyst*

Further processing, either after binning or peak picking, was carried out in Metaboanalyst (5.0). CSW files were formatted according to Metaboanalyst specifications and subsequently uploaded into the statistical analysis pipeline. For all LA-REIMS spectral data, mean intensity data filtering, normalisation by sum, pareto scaling, and log transformation was carried out. In multivariate analysis, unsupervised Principal Component Analysis (PCA) was used as an initial test on all datasets, together with Partial Least Squares – Discriminant Analysis (PLS-DA) as a supervised approach, for distinguishing discrete disease cohorts or groupings of interest within the data.

PCA summarises the variability in the data by geometrical projection of variables onto lower dimensions called principal components, this allows high-dimensional data to be reduced to its main components (316). PCA measures the direction of largest variance (317). The output of a PCA plot may be used to identify groups of similar samples or signals and to identify trends of unexpected patterns in a dataset: The scores of a PCA allow the distribution of the samples to be easily displayed and evaluated. Relationships between variables can be studied by inspecting their contributions (loadings) to the model and potential outliers can be checked by looking for observations outside of the 95% confidence interval (Hotelling's T2 ellipse) (91).

Partial least-squares discriminant analysis (PLS-DA) analysis is a supervised methods in metabolomics. In PLS-DA, the co-variance between the variable of interest is maximised (133). It represents a measurement of the contribution of a feature to the discrimination between the different sample groups (24).

Multivariate statistics were utilised to build diagnostic models that could predict the diagnostic accuracy of LA-REIMS for patients with cancerous lesions and adenomatous lesions as well as UC and CD in the CRC and IBD cohorts, respectively. To evaluate patient-related

modifiers of the faecal metabolome, further predictive models were built based on baseline characteristics and demographic data. For supervised analysis validation purposes, Leave-one-patient-out cross-validation (LOOCV) with Q2 and R2 as performance measure and a maximum component of five was carried out in addition to permutation testing to test for model fitness. Q2 is an estimate of the predictive ability of the model whereas R2 is defined as a fraction of variance explained by a component.

Furthermore, to assess LA-REIMS performance, random forest algorithms (RFA) were used as a machine learning approach for sample classification purposes in MetaboAnalyst 4.0 online platform. In MetaboAnalyst 4.0, multivariate ROC curve based exploratory analysis (Explorer) tool was utilised for RFA classification. RFA classification is an approach which allows for the prediction of a test dataset, further to training on a known dataset and has been utilised in previous LA-REIMS diagnostic accuracy studies as a means of evaluating LA-REIMS performance (318). Monte-Carlo cross validation (MCCV) using balanced subsampling, a method previously tested and validated in a LA-REIMS diagnostic accuracy study published by our research group was used to create Receiver operating characteristic (ROC) curves (318). In MCCV, random splits of the data are made into training and validation groups (318). In each individual split, the model is built using the training data and accuracy is calculated utilising the validation data; further to this an average is calculated (318). Here, two thirds of the samples were used as training data and the one third left out was used as test dataset.

Univariate analysis techniques, including Analysis of Variance (ANOVA), along with Fisher's Least Significant Difference post-hoc test, were used for biomarker discovery investigations. Spectral mass features with a false discovery rate (FDR) corrected P value of $< 0.05$ were accepted as significant biomarkers. For analysis of two groups only, volcano plots displaying fold change against FDR corrected P value from t tests were utilised to aid in significant feature selection. Box plots were generated for all features of interest (i.e., demonstrating significant difference between cohorts). ANOVA and Volcano data were exported to Microsoft Excel in tabular form for further manual inspection, processing, and creation of charts. The biomarker analysis tool suite within Metaboanalyst was utilised for exploratory

univariate area under the receiver operator characteristic curve (AUROC) analysis of significant features.

#### *3.3.3.5. Metabolite Identification*

In LA-REIMS analysis metabolite identification, univariate methods (ANOVA, Volcano plots, etc.) were utilised to generate a list of potentially biologically important metabolite features. Accurate m/z were taken from the raw spectrum after background subtraction and lock mass correction (internal or external lock mass compound applied as discussed in section *3.3.3.3. Processing – Lock mass correction*). Particular attention was paid to the identification of real peaks versus noise. Real peaks were categorised as peaks that are 1:3 (noise: peak) in intensity ratio from the background noise level.

Furthermore, annotation of the *m/z* of interest was carried out based on their accurate mass, the isotopic distribution as well tandem MS based data (to be discussed in section *3.3.3.6. Tentative identifications – validation.* When performing ion annotation, special attention was paid to isotopes, adducts, and neutral loss fragments that may originate from the same compound of interest. The most common adducts in LA-REIMS negative mode ionisation include M-H, M+Cl, and M-$H_20$-H. In LA-REIMS positive mode they include, in addition to the commonly observed protonated molecules, sodium, potassium, and ammonium adducts.

After ion annotation, the monoisotopic masses were calculated based on adducts/isotopes (97). In ambient techniques such as LA-REIMS, no retention time information is acquired. In chromatography-coupled MS, such as LC-MS, peaks (multiple adducts) belonging to a specific molecule will co-elute at the same retention time. This information can be used to identify the most accurate monoisotopic mass with multiple statistical platforms available which can integrate ion annotation data.

In LA-REIMS analysis, visual inspection of the data is carried out with particular attention paid to isotope patterns and potential adducts visible from the raw spectrum. For instance for the M+Cl adduct, particular attention was paid for the natural isotope distribution pattern of chlorine, $^{35}Cl/^{37}Cl$ in a 3/1 ratio to be present (319). For an M-$H_20$-H adduct, a

neutral loss with the elimination of water (18 Daltons) in addition to de-protonation was noted. In addition, when distinguishing between the potential M-H adduct of one compound vs the M+Cl adduct of a different compound at the same m/z, the likelihood of ionisation in the mode was considered. However, it must be emphasised that especially for LA-REIMS metabolomics data, this poses a substantial challenge and necessitates further investigations/validations using among others MS/MS or LC-MS analysis.

In the next step, the generated list of m/z were inputted into Chemcalc (320), an online tool, to generate a list of potential chemical formulae for the monoisotopic mass of the specified m/z. The chemical formulae were imported into SciFinder (321) to generate metabolite matches. Furthermore, the potential features matched were checked in terms of their likelihood to be seen in negative or positive ionisation mode LA-REIMS.

For the list of tentatively identified features, literature references for compounds with an *m/z* of <5 ppm that were previously observed in faecal metabolomics studies (in the specific disease groups, i.e., CRC or IBD) were investigated. A biological hypothesis was generated, and features were cross-checked in HMDB faecal metabolomics database and LipidMaps, with the highest ranked matches, based on ppm error value (< 5ppm), taken as the tentative identification shown below. Mass accuracy or mass measurement error were calculated in ppm according to the following equation (322):

$$\Delta\, m^i \text{ in ppm} = \frac{(m^i - m^a)}{m^a} \times 10^6$$

$m^i$ = Measurement of a single accurate mass in Da (experimental mass)

$m^a$ = Theoretical exact mass

$\Delta\, m^i$ = mass deviation (in ppm)

The QToF instruments used in this work are modern high resolution accurate mass instruments so a mass accuracy of <5 ppm was utilised as a cut off for the cross-check for the

tentative identification of features. This is useful as with a <5 ppm mass accuracy, the elemental composition of relatively small molecules can usually be determined(323). According to manufacturer, for the present generation of QToF MS instruments and by using calibration and lock mass correction, mass accuracies within 3–5 ppm should be obtained in the 100–1,000 Da range (324). Furthermore, increasing the mass accuracy, that is decreasing the mass error for the identification of compounds, results in a substantial decrease in the number of potential formulae, see Figure 5 (325). Figure 5 also suggests that the number of potential formulae produced based on the cut off mass accuracy varies across the m/z range.

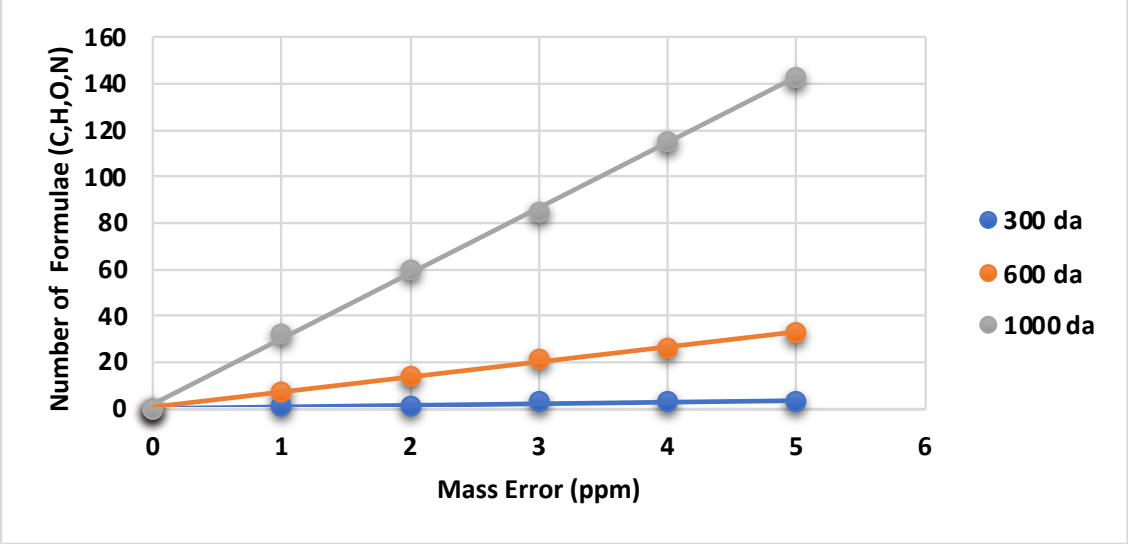



Figure 5: The effect of increasing mass accuracy for the number of identification of compounds with mass error (ppm) plotted against the number of potential formulae based on Carbon (C), Hydrogen (H), Oxygen (O), and Nitrogen (N) compositions at 300, 600, and 1000 Daltons respectively. The figure was adapted from Quenzer, et al 2002 (325).

### *3.3.3.6. Tentative identification – validation*

An attempt at the validation of several features of interest was made based on Ultra Performance Liquid Chromatography (UPLC)-MS based experiments.

#### 3.3.3.6.1. UPLC-MS based data dependent acquisition (DDA) experiments

LC-MS and MS/MS provide important information for metabolite identification. Typically, an untargeted approach is taken where a full MS scan is acquired. This is used to generate a list of m/z that are significantly different between case and control. MS/MS is typically carried out by selecting an *m/z* value and the corresponding ion being fragmented using collision induced dissociation (CID). The most accurate *m/z*, along with RT and MS/MS data and spectral reference libraries are then used for metabolite identification. In addition to MS/MS, data-dependent acquisition (DDA) analysis can be carried out to acquire fragmentation information (97). In DDA, an initial scan is followed by MS/MS acquisition; in the initial scan, ions above a pre-set abundance threshold are selected and fragmentation via MS/MS is automatically performed on features meeting the criteria (97). It is important to highlight that whilst DDA is an efficient technique that can carry out MS/MS on many precursor ions, it is limited to the most abundant ions. Furthermore, successful metabolite identification requires fragmentation of a specific ion of interest with the selection of specific collision energies, which might not be the case with DDA (97).

An attempt at validating tentative identifications (LA-REIMS analysis) in the CRC and IBD using UPLC-MS was carried out. For this analysis, six pooled faecal samples were prepared, by weighing out equal volumes (0.1 grams) of each faecal sample. The six groups included, 20 pooled CRC (CRC dataset), 20 pooled adenoma (CRC dataset), 20 pooled non-cancer control (CRC dataset), 20 pooled CD (IBD dataset), 20 pooled UC (IBD dataset), and 20 pooled non-IBD control (IBD dataset) samples. The CRC samples were analysed using MS and DDA mode, whereas the IBD samples were assessed using MS mode only

The pooled faecal samples underwent a modified monophasic IPA extraction according to Sarafian, et al (326). The organic solvent, isopropanol, was obtained from Sigma-Aldrich (Dorset, UK). Here, the sample preparation for lipid analysis was adapted for the analysis of faeces. Pre-cooled IPA 1:3 ratio was added to each Eppendorf of the pooled faecal samples. Faecal samples in IPA were vortexed for 1 minute and incubated for 2 hours at 4˚C. In the next step, beads were added for bead beating at 6500 G, 2x45 seconds. The samples were then vortexed for one minute and placed in 4˚C for protein precipitation overnight. The samples

were then centrifuged at 14,000 G for 10 minutes at 4˚C. The supernatant was transferred into a clean 2 ml Eppendorf vial, and the centrifugation step was repeated at 14,000 G for 10 minutes at 4˚C. The supernatant was transferred into a Waters Polypropylene Screw Neck Vial, 12 x 32 mm, 300 µL HPLC vial (Waters Corporation, UK) for UPLC-MS analysis. Furthermore, a quality control and blank samples were prepared to monitor the performance of the instrument throughout analysis.

The faecal extracts were analysed by the National Phenome Centre (NPC) (327) reverse phase lipid profiling method with a C8 column (Waters Acquity UPLC BEH C8, 1.7 µm, 2.1 x 100 mm (P/N: 186002878)).

#### 3.3.3.6.2. Liquid Chromatography

Chromatography analysis was performed using Acquity UPLC system (Waters Ltd., Elstree, U.K.) under UPLC parameters outlined in Table 5. The samples were injected onto a Waters Acquity UPLC BEH C8, 1.7 µm, 2.1 x 100 mm (P/N: 186002878) column at 55°C. Details of the composition of the mobile phases, injection volume, and flow rate used are provided in Table 5 and the gradient used is summarised in Table 6.

Table 5: UPLC instrument parameters

| | |
|---|---|
| Column | Waters Acquity UPLC BEH C8, 1.7 µm, 2.1 x 100 mm (P/N: 186002878) |
| Column temperature | 55 °C |
| PEEK tubing | Column to sample probe: ID = 0.004 inch, length = 38 cm |
| Sample Loop | Positive: 1 µL, negative: 2 µL |
| Injection volume and flow rate | 15 µL and 400 µL/min |
| Mobile Phase A | Water:isopropanol:acetonitrile 2:1:1, 5 mM ammonium acetate, 0.05% acetic acid, 20 µM phosphoric acid |

| | |
|---|---|
| Mobile Phase B | Isopropanol:acetonitrile 1:1, 5 mM ammonium acetate, 0.05% acetic acid |
| Seal Wash | Isopropanol:water 1:9, set to run every 0.5 minutes |
| Weak Wash | Water:isopropanol 1:4 |
| Strong Wash | Isopropanol |
| Lockspray | Leucine enkephalin 600 pg/µL (prepared in water:acetonitrile 1:1) |

Table 6: Chromatographic gradient used for lipid profiling

| Time (min) | %A | %B | Curve |
|---|---|---|---|
| Initial | 99.0 | 1.0 | Initial |
| 0.10 | 99.0 | 1.0 | 6 |
| 2.00 | 70.0 | 30.0 | 6 |
| 11.50 | 10.0 | 90.0 | 6 |
| 12.00 | 0.1 | 99.9 | 6 |
| 12.50 | 0.1 | 99.9 | 6 |
| 12.55 | 35.0 | 65.0 | 6 |
| 12.65 | 70.0 | 30.0 | 6 |
| 12.75 | 99.0 | 1.0 | 6 |
| 12.95 | 99.0 | 1.0 | 6 |
| 13.25 | 99.0 | 1.0 | 6 |

### 3.3.3.6.3. Lipid Profiling by UPLC–Quadrupole-Time-of-Flight (Q-TOF) Mass Spectrometry

Following chromatographic separation, MS analysis using a Xevo G2-S Q-TOF (Waters, Manchester, UK) was performed on the samples. Details of instrument and source parameters utilised are provided in Table 7. For both ion detection modes, leucine enkephalin was used at a rate of 30 µL/min for lock mass correction. Performance checks were run for MS lipid data acquisition of the faecal extract samples. These included a dual point lock mass check and an injection before and after the faecal samples of long-term reference (LTR)

plasma sample and blanks with internal standard (IS) to ensure instrument performance. Each pooled CRC cohort sample was analysed in MS and subsequently DDA mode, whereas each pooled IBD cohort sample was analysed in MS mode only. The run order in both negative and positive mode included, 1 blank injection, 1 LTR injection, 3 blank injections, faecal extract samples in provided order, 1 blank injection, 1 LTR injection. The NPC carried out the LC-MS analysis and provided the raw data files of all the acquired spectra.

Table 7: MS source parameters

| | |
|---|---|
| Sample Probe Angle | 7mm |
| Capillary Voltage | Positive: 2:0 / Negative: 1.5 |
| Sampling Cone | 25 |
| Source Offset | 80 |
| Source Temperature (°C) | 120 |
| Desolvation Temperature (°C) | 600 |
| Cone Gas Flow (L/hr) | 150 |
| Desolvation Flow (L/hr) | 1000 |
| Collision Energy (V) | Off |
| Gas | API On, Collision On |
| Detector Auto Gain | On |
| Auto Gain Controls | Custom auto gain check = ticked, fluidics = tune page, detector check time = 70 |
| Stepwave 2 Offset | 15 |
| Quad Profile | Auto |
| Lockspray flow (µL/min) | 15 |
| Lockspray Capillary Voltage (kV) | Positive: 3.0 / Negative: 2.0 |
| MS Function | TOF MS |
| Analyser mode | Sensitivity |
| Dynamic Range | Extended |
| Mass Range | 50-2000 Da |
| Scan Time | 0.1 sec |

| Data Format | Centroid |
|---|---|
| Lockspray function settings | Acquire lockspray & apply correction<br>Select appropriate Lock Spray file<br>Scan time = 0.15 sec, interval = 60 sec, scans to average = 4, Mass window = +/- 0.5 |

#### 3.3.3.6.3. MS Data analysis and feature identification

Data analysis was carried out by the candidate by using MassLynx v4.1 (Waters, Wilmslow, UK) software. The MS mode raw data was manually inspected in MassLynx v4.1 (Waters, Wilmslow, UK) for spectral signal from features of interest after pre-processing, see section *3.3.3. Data processing and statistical analysis*.The DDA tool function allowed for the investigation of features of interest and their fragmentation patterns. The *m/z* of the tentatively identified features (up to three decimal places) were checked for their presence in the list of MS mode spectra (IBD and CRC datasets).

Furthermore, only the CRC dataset was investigated for DDA hits in the LC-MS data. Further, the m/z of interest, were interrogated at each retention time eluting. A mass accuracy of <5 ppm was utilised as a cut off for the cross-check of LA-REIMS and LC-MS based features. Thereafter, for features that presented in both LA-REIMS and LC-MS in the CRC cohort, the accurate m/z, the retention time, and fragmentation pattern was subsequently used to interrogate spectral reference libraries for matches. Accurate m/z value of the molecular ion, isotope profile, retention time and fragmentation patterns observed in DDA mode were used for identification purposes.

## 3.4. Experiments

### 3.4.1. Optimisation studies

Patient recruitment and sample collection were carried out by the candidate at Charing Cross Hospital, Imperial College London NHS Trust. Ethical approval was granted by the Imperial College London ethics committee, with full local, ethical approval obtained (Research Ethics Committee reference 14/EE/0024).

#### *3.4.1.1. Laser optimisation*

Laser parameters of the LA-REIMS set-up were optimised to increase analytical throughput and sensitivity in this method optimisation study. Those under the age of 18, pregnant, or having a previous history of GI disease or surgery were not included. The study participants were provided with a FECOTAINER (AT Medical BV, The Netherlands), collection device along with instructions for collection at Charing Cross Hospital. Participants were asked to return the sample within one hour of bowel evacuation and faecal samples were immediately stored at -80˚C. All samples were collected within one week prior to analysis.

To prepare for LA-REIMS analysis, faecal samples were homogenised and prepared for two analytical workflows; the analysis of raw faecal samples and extracted faecal water (1:2 faeces: water) in both negative and positive ionisation modes. The faecal samples underwent an adapted and modified faecal water extraction according to Gratton, et al (114). Specifically, faecal water was extracted from each sample at a 2:1 water (HPLC-grade, Sigma Aldrich) to faeces ratio. Specifically, 15 g of faecal homogenate was transferred into a 50 mL centrifuge tube for faecal water extraction with 30 ml of water; experiments were performed in triplicate. The mixture was vortexed for 1 minute and centrifuged at 4°C at 18000 *x g* for 10 minutes. After centrifugation, the supernatant was aliquoted into a 15 ml centrifuge tube for further analysis.

To evaluate optimal laser parameters, 1.5 g of homogenised faecal sample was added to individual wells in a 24-well plate, Figure 6. For faecal water, 5µl of the extracts were added to 96-well Matrix Assisted Laser Desorption Ionisation (MALDI) spotting plates, Figure 7. The plates with loaded faecal sample and faecal water extracts were placed in a -80°C freezer overnight before analysis on the subsequent day. The following parameters were assessed: heating power 2-5 watts at 0.5 watts increments, laser beam Super Pulse on versus off, and

laser beam pulse (continuous, repeat 40 ms, repeat 60 ms, repeat 80 ms, or repeat 100ms). All possible combinations of laser parameters were tested, yielding a total of 70 conditions. The 70 conditions were tested for both faecal sample (negative and positive mode) and faecal water (negative and positive mode). Each well with faecal sample was tested under a different combination of laser parameters with the laser instructed to produce one burn per sample. Moreover, each condition was tested with four repeats (on 4 separate wells).

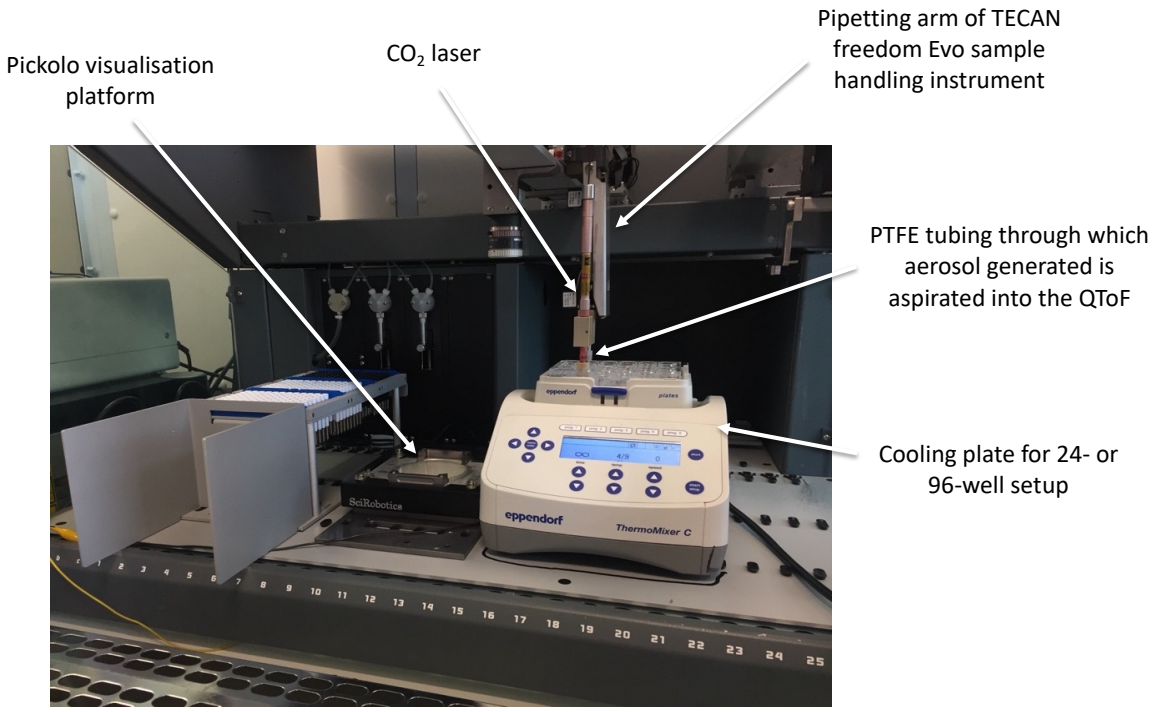



Figure 6: LA-REIMS instrument with cabinet cooling plate set to 4°C, $CO_2$ laser, and 24-well plate with loaded faecal sample (1.5 grams per well)

**A** **B** **C**

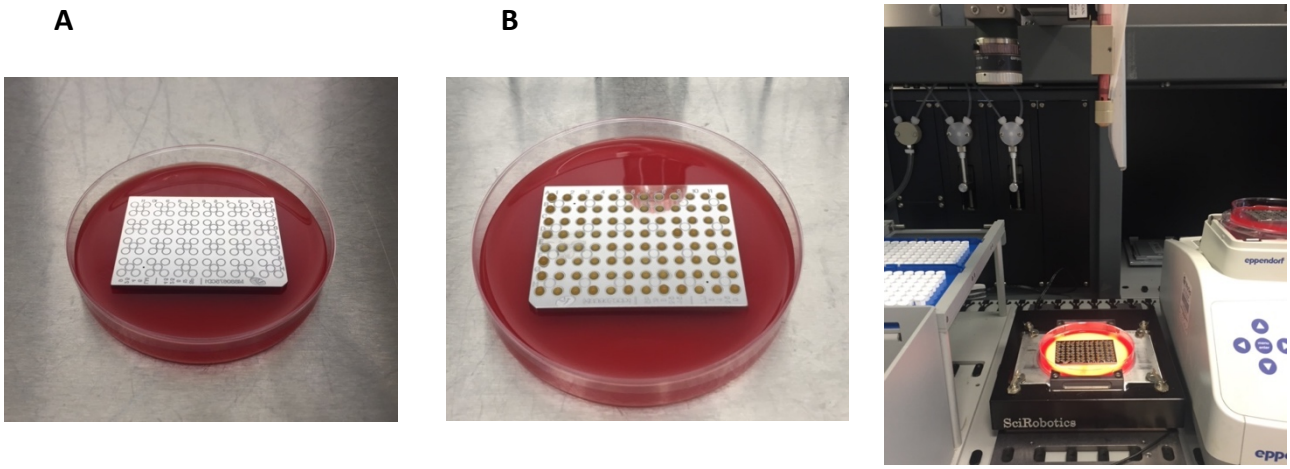



Figure 7: Faecal water extract LA-REIMS analysis with **A** 96-well MALDI spotting plate **B** 96-well MALDI plate with 5µl of faecal water extract added per well **C** 96-well MALDI plate with 5µl of faecal water extract per well with agar plate positioned on a Pickolo visualisation platform - an integrated camera captures an image of the setup and allows for the selection of samples to be analysed using the high-throughput LA-REIMS setup

Faecal samples were placed on a cooling plate set at 4°C (to keep the samples at 4°C throughout the run, to prevent extensive heating of the sample by the laser, as well as reduce metabolite degradation by the bacteria), see Figure 6. For faecal water analysis, 96-well MALDI spotting plate with sample were placed on the image visualisation platform, Figure 7 A-C, and the instrument then captured an image of the 96-well MALDI spotting plate with sample and areas of interest for subsequent analysis were selected for analysis. LA-REIMS analysis was carried out as described in section *3.3.2. LA-REIMS analysis*. Details of the negative and positive ion detection mode analysis and cleaning protocol are described in section *3.3.2.1. LA-REIMS negative and positive mode analysis switch* and *3.3.2.2. LA-REIMS cleaning protocol.* The flow rate of 0.2 mL/s was chosen based on previous optimisation studies (61) in microorganism analysis as the instrumentation method development was carried out after the laser optimisation. Further, before the instrumentation optimisation study, a 30 second time interval between subsequent sample analysis burns were utilised to avoid carry-over and contamination of the spectral signal between samples.

Acquired raw spectra underwent pre-processing utilising binning as described in section *3.3.3. Data processing and statistical analysis*. Thereafter, the four analytical groups, i.e., faecal sample negative ion detection mode, faecal sample positive ion detection mode, faecal water negative ion detection mode, and faecal water positive ion detection mode, were further investigated. For each group, the average signal intensity in the complex lipids (*m/z* 600-1000), low mass (*m/z* 200-500), and noise (*m/z* 50-51) ratios were calculated for each of the four replicates tested. For each group, the complex lipids (*m/z* 600-1000)-to-noise (m/z 50-51) ratios and low mass (*m/z* 200-500)-to-noise (*m/z* 50-51) ratios were calculated (for each of the four replicates of each combination of settings tested.

Specifically, the sum of the signal intensity of the complex lipids (*m/z* 600-1000) or the low mass (*m/z* 200-500) range were divided by the sum of the signal intensity of noise (m/z 50-51) region for each parameter combination tested. The average of the four replicates was utilised in the heatmap, after normalisation was applied. Heatmaps of the average signal intensities and the complex lipids (*m/z* 600-1000)-to-noise (m/z 50-51) ratios and low mass (*m/z* 200-500)-to-noise (*m/z* 50-51) ratios were created in R studio (V1.0.44) to allow for the selection of the combination of parameters yielding the highest overall signal-to-noise ratios.

#### *3.4.1.2. Instrumentation optimisation*

Instrumentation parameters were assessed next as a continuation of the laser optimisation study. Study participants were subject to the inclusion and exclusion criteria and were educated on sample collection techniques as described in section *3.4.1.1. Laser optimisation*. Faecal samples and faecal water (1:2 faeces: water) extracts were homogenised and prepared as described in section *3.4.1.1. Laser optimisation*. For faecal water, to allow for evaluation of % carry-over, 96-well MALDI spotting plates were arranged with each well with 5µl of homogenised faecal water followed by a well with 5µl of water (HPLC-grade, Sigma Aldrich) added. Similarly, for faecal samples, 24-well plates were organised by the repeated arrangement of wells with faecal sample followed by wells with equal volumes of agar added, see (Figure 8 A+B) for details of the sample arrangement.

The following parameters were assessed: intervals between subsequent sample burns (10s, 20s, 30s, 60s), and IPA (HPLC-grade, Sigma Aldrich) flow rates (0.25 mL/s, 0.3 mL/s, 0.35 mL/s). All possible combinations of instrumentation parameters were tested, yielding a total of 12 conditions. The 12 conditions were tested for both faecal sample (negative and positive mode) and faecal water (negative and positive mode). Each well with loaded faecal sample or water was tested under a different combination of instrumentation parameters with the laser instructed to produce one burn per sample. Moreover, each condition was tested with four repeats. LA-REIMS analysis as described in section *3.3.2. LA-REIMS analysis* was carried out.

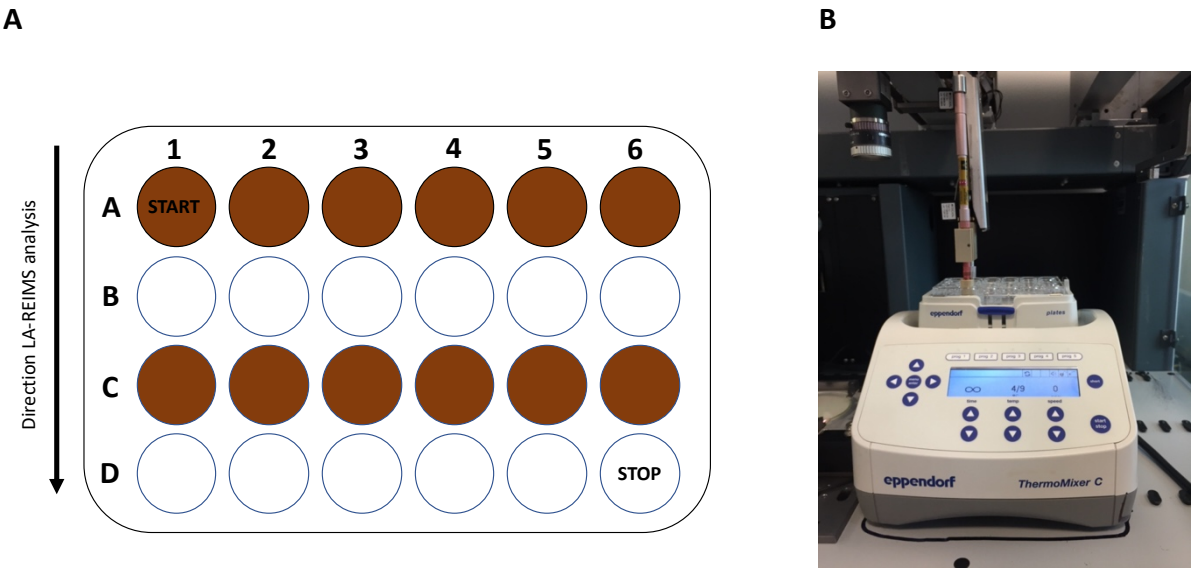



Figure 8: High-throughput LA-REIMS setup with A Schematic diagram of 24-well plate LA-REIMS setup with faecal sample organised by the repeated arrangement of wells with added homogenised faecal sample followed by wells with agar (brown circle = well loaded with faecal sample, white circle = well loaded with agar) B TECAN freedom Evo sample handling instrument with cooling plate set to 4°C, $CO_2$ laser, and 24-well plate with added sample

Acquired raw spectra underwent pre-processing utilising binning as described in section *3.3.3. Data processing and statistical analysis.* The resulting MS data was used to identify ten *m/z* features with the highest average intensities in both the low mass range (*m/z* 200-500) and the complex lipid range (*m/z* 600-1000) for the four chosen groups, i.e., faecal sample negative

mode, faecal sample positive mode, faecal water negative mode, and faecal water positive mode. In each group, for each replicate of a specific combination of instrumentation settings, the % carry-over was calculated for the ten specified *m/z* values in both the low mass and complex lipid region. This was achieved by dividing the sum of the signal intensity of the faecal sample or faecal water extract by the agar/water in the subsequent well. In each chosen group, for each setting combination tested, the average of four replicates for the % carry-over as well as standard deviations were calculated. Thereafter heat maps comparing the various settings were generated in R studio (V1.0.44) to allow for the selection of settings yielding the lowest % carry-over with the shortest time interval between burns.

#### *3.4.1.3. Water content study*

For this feasibility pilot study, a healthy volunteer, with no known history of GI disease, donated a faecal sample. Study participants were subject to the inclusion and exclusion criteria and were educated on sample collection techniques as described in section *3.4.1.1. Laser optimisation*. The study participant was provided with a FECOTAINER collection device along with instructions for collections at Charing Cross Hospital and asked to return the sample within one hour of bowel evacuation and faecal samples was immediately stored at -80˚C.

On the day of analysis, the faecal sample was thawed at RT in a Class II biological safety hood for 30 minutes and homogenised. Thereafter, the homogenised fresh faecal sample was prepared by adding HPLC grade water (VWR, Leicestershire, UK) to make up the following dilutions with four repeats each: raw faeces, water: faeces (1:1), water: faeces (1:2), water: faeces (1:4), water: faeces (1:6), water: faeces (1:8), and water: faeces (1:10). The homogenised faecal samples were vortexed, and 1.5 grams was added to individual wells into two 24-well plates.

The prepared 24-well plates with faecal samples were placed on the cooling plate set at 4°C for $CO_2$ LA-REIMS analysis, with optimised settings developed in terms of laser and instrumentation parameter implemented. Specifications for the analysis are outlined in section *3.3.2. LA-REIMS analysis*. Acquired raw spectra underwent pre-processing utilising

binning as described in section *3.3.3. Data processing and statistical analysis.* The processed data was utilised to build models that could determine whether separation based on faecal water content would be observed in unsupervised multivariate analysis, when analysing the same faecal sample with different ratios of HPLC grade water (VWR, Leicestershire, UK) added.

#### *3.4.1.4. Heterogeneity of metabolites*

A method development pilot study aimed at evaluating LA-REI-Mass Spectrometry Imaging (MSI) analysis of faeces, was setup at Charing Cross Hospital, Imperial College London. Participants were subject to the same inclusion and exclusion criteria and samples were collected as per section *3.4.1.1. Laser optimisation*.

Each participant was given a collection pack consisting of an insulated faecal collection kit - FECOTAINER ® (AT Medical BV, The Netherlands), two zip-lock freezer bags, gloves, and a freezer bag with ice packs inside. Participants were asked to provide the whole faecal samples on site (as part of their normal bowel evacuation routine) and return the samples within one hour from the time of evacuation. On receipt of the sample on each sampling day, full faecal samples were frozen at -150°C for 30 minutes and subsequently removed from the freezer, thawed for 30 minutes, and segmented into cross-sectional plates (5mm) at room temperature using a heavy-duty meat slicer (OSTBA, China), see Figure 9. The cross-sectional slides were stored at -80°C in individual plates until LA-REI-MSI analysis.

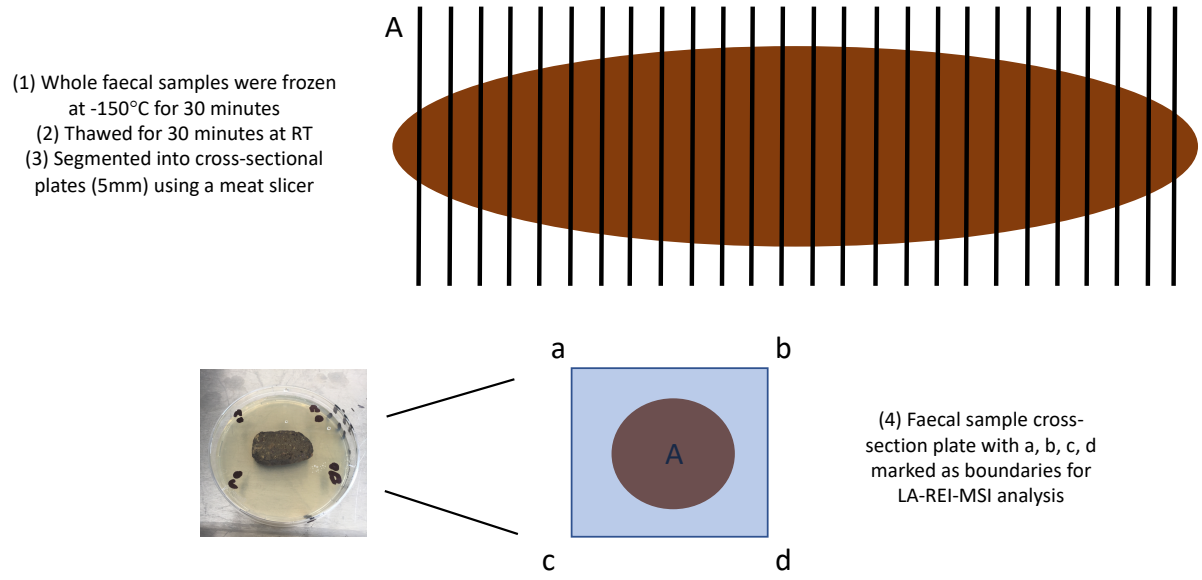



Figure 9: Whole faecal sample preparation for LA-REI-MSI: (1) Faecal samples are frozen for 30 minutes at -150°C, (2) subsequently thawed for 30 minutes at RT, (3) segmented into cross-sectional plates (A) (5mm) using a meat slicer, and (4) prepared for LA-REIMS analysis with a-d marked as area of imaging with 1 mm resolution

For LA-REIMS analysis, each cross-sectional plate was thawed at room temperature in a class 2 biological safety hood and analysed using the LA-REIMS Pickolo visualisation platform in both negative and positive ionisation modes, Figure 10. The optimised settings developed in terms of laser and instrumentation parameter were implemented in this study. Each sample slide was placed on an agar plate, Figure 11 A, and four corners (Figure 11 A a, b, c, and d) were marked as representing the edges of the square surface to be analysed via LA-REIMS at a 1 mm resolution.

The agar plate was positioned on the plate visualisation platform (Figure 11 B+C). An image was captured, and the operator selected the area of interest (here, square outlined by Figure 11 A a-d) for sample analysis. The robotic arm equipped with the $CO_2$ laser was set to fire a laser beam at the sample. The laser was set to fire in the pre-specified square, illustrated in (Figure 11A a-d), by producing a burn with a pixel size of 1 mm (3 seconds burn interval) at

the first location and then moving to the next position for the subsequent burn (time interval between burns five seconds) with a pointing depth of 2 mm. The laser continued to fire at 1 mm resolution until it reached the end of the pre-specified square (Figure 11 A a-d) as can be seen in (Figure 11D) and a detailed view in (Figure 11E). Specifications of the MS analysis are outlined in section *3.3.2. LA-REIMS analysis*.

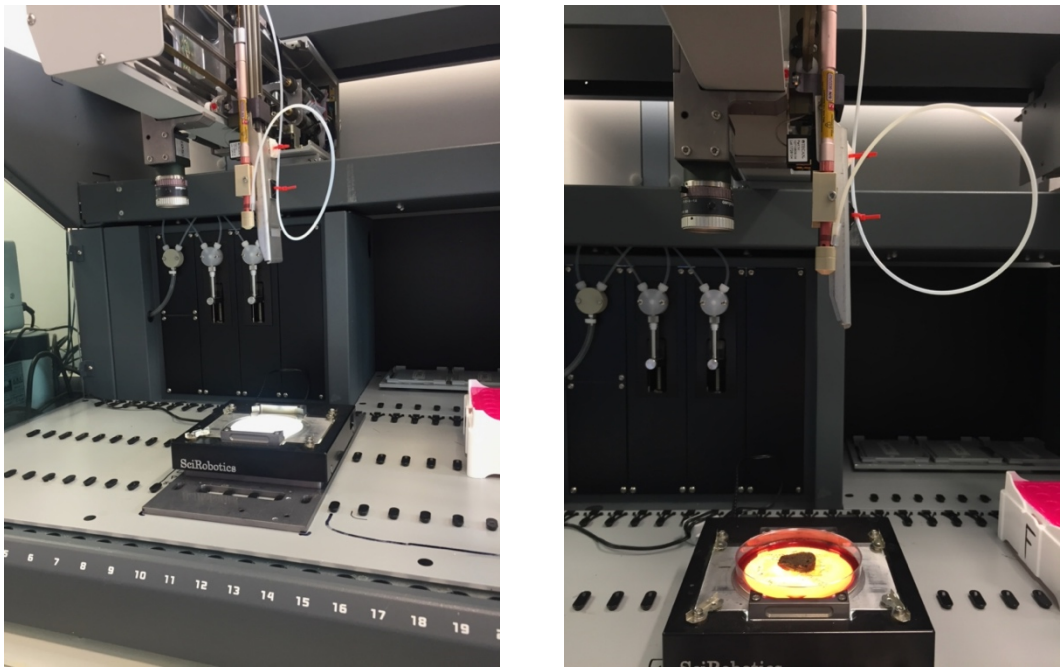



Figure 10: LA-REIMS Pickolo visualisation platform – an integrated camera captures image of sample setup and allows for the selection of samples to be analysed

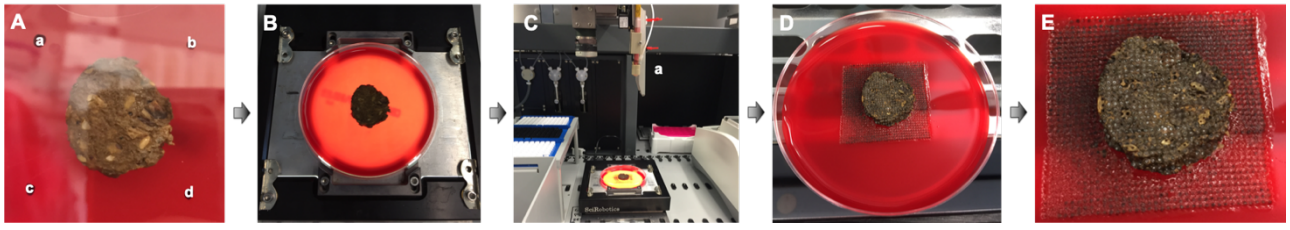


Figure 11: Main stages in LA-REI-MSI with A Faecal sample cross-sectional slide on agar plate with positions a,b,c, and d marked as the boundaries for analysis B Agar plate with sample positioned on the visualisation platform C Whole setup with (a) $CO_2$ laser in position for analysis D Agar plate with faecal sample after analysis E Detailed view of faecal sample with all individual burns at 1 mm resolution

Specifications for the LA-REIMS analysis is outlined in section *3.3.2. LA-REIMS analysis*. One spectrum was obtained for each burn and spectra were pre-processed utilising binning as described in section *3.3.3. Data processing and statistical analysis.* An in-house built statistical pipeline in R Studio (V 1.0.44) by Alvaro Perdones-Montero allowed for the allowed for the selection of each spectrum and mapping to specific locations within the cross-section, as observed in Figure 11 E. The top *m/z* features with highest relative intensity in the low mass and complex lipid region were tentatively identified as described in section *3.3.3.5 Metabolite Identification* and *3.3.3.6 Tentative identification – validation.* The putative features were examined along with their spatial distribution within the faecal slide cross-sections.

### 3.4.2. CRC study

Principal investigator (PI) Prof. Ramesh Arasaradnam and his team caried out the recruitment and collection of faecal samples at the University Hospitals Coventry and Warwickshire and Leceister General Hospital. Ethical approval was granted by Coventry & Warwickshire ethics committee as part of the FAMISHED study – 09/H1211/38. The recruitment of patients was carried out by Prof. Ramesh Arasaradnam and his team between January 2015 and January 2017. The inclusion criteria were patients referred via the national two-week wait cancer referral pathway for suspected CRC. The exclusion criteria included patients under the age of 18 and patients not having full colonic investigations. CRC and adenoma diagnosis was confirmed via histology. Study participants were given a 30 mL blue stool pot (Universal Sterilin, Thermo Scientific, UK), a faecal immunochemical test (FIT) device (Kyowa Medex Co Ltd, Japan) along with instructions for collections at the clinic visit. Patients were asked to return the faecal sample in the blue stool pot and the FIT Collection Picker after testing the

same faecal sample prior to colonic investigations directly to the laboratory at UHCW (Faecal samples were excluded if returned >4 days post-collection).

Faecal haemoglobin and faecal calprotectin were analysed for all UHCW patients who returned faecal samples by PI Prof. Ramesh Arasaradnam and his team at UHCW. A HM-JACKarc analyser (Kyowa Medex, Japan) was used for the analysis of the FIT kits, on a weekly basis. The detection limit at 7 micrograms of haemoglobin per gram of faeces was considered positive. Following faecal calprotectin analysis, faecal samples were stored at -80°C until they were sent to Charing Cross Hospital (Imperial College London NHS Trust) for LA-REIMS analysis. It must be noted that the patient recruitment, sample collection, as well as the FIT and Faecal calprotectin analysis was performed under PI Prof Ramesh Arasaradnam and his representative teams.

Frozen faecal samples were transported on dry ice on 17/01/2018 (transport time 3hr 35 min) from the University Hospitals Coventry and Warwickshire to Charing Cross Hospital (Imperial College London NHS Trust). Samples were immediately placed into a -80°C freezer. Between January 2019 and February 2019, LA-REIMS analysis of faecal sample and faecal water extracts (1:2 faeces:water) was carried out by the candidate in both negative and positive ion detection modes. Faecal samples were thawed (at 21°C in a Class II biological safety cabinet) and prepared for two analytical workflows.

The results of the optimisation work were implemented in the analysis of the CRC dataset. Faecal samples were randomised into 96-deep-well plates, with each well corresponding to the faecal sample of a single person, to be analysed using the optimised high-throughput LA-REIMS setup. Faecal samples were added to wells of the 96-deep-well plate using disposable 10 µL inoculating sterile loops (VWR International). Faecal water was extracted at a 1:2 faeces:water ratio as described in the protocol in section *3.4.1.1. Laser optimisation.* Faecal water samples were spotted at 5ul per well into 96-deep-well plates using a 0.5-10 µL volume pipette (Sigma-Aldrich, USA) with disposable tips.

The 96-deep well plate for faecal sample and water extract analysis were placed on a cooling

plate and a robotic arm equipped with the $CO_2$ laser fired a laser beam onto the sample. Specifications for the LA-REIMS analysis is outlined in section *3.3.2. LA-REIMS analysis*. For the CRC, in total, 6 x 96-deep well plates were analysed with LA-REIMS for both faecal sample and faecal water extracts. All acquired raw spectra and metadata (manually formatted according to specifications) were uploaded into an in-house built package developed in RStudio v1.1.419 for peak picking of LA-REIMS spectra, as described in section *3.3.3. Data processing and statistical analysis.*

After peak picking, the processed data was uploaded to Metaboanalyst 5.0 online platform was utilised for multivariate statistics to build diagnostic models of LA-REIMS (in comparison to FIT) for patients with cancerous lesions, adenomatous lesions, and no lesions based on faecal sample analysis. Furthermore, to assess the diagnostic performance and for sample classification purposes of faecal LA-REIMS, RFC algorithms were utilised. Receiver operating characteristic (ROC) curves were generated by Monte-Carlo cross validation (MCCV) using balanced subsampling, a method previously utilised in a LA-REIMS diagnostic accuracy study (318). Further, predictive models were built based on patient-related factors, including among others age, sex, BMI, ethnicity, comorbidities, medication use, smoking and drinking status. Specifically, the adenoma and cancer groups were interrogated for the extent of the disease, cancer/adenoma cell biology and specific histological phenotype. Statistically significant features between non-cancer control, adenoma and cancer patients were interrogated using univariate statistics.

Details of statistical analysis are described in *3.3.3.4. Further processing – Metaboanalyst*. The tentative identification of metabolites and the attempt at validation was carried out as described in section *3.3.3.5 Metabolite Identification* and *3.3.3.6 Tentative identification – validation.*

### 3.4.3. IBD study

At Imperial College London NHS Trust, PI Prof. Timothy Orchard and his team carried out at patient recruitment and sample collection with approval obtained from the London – Harrow Research Ethics Committee and the study was submitted for Site Specific Assessment (SSA) at each participating NHS Trust. Research Ethics Committee reference - 15/LO/0801.

A prospective, observational cohort study was performed at Imperial College London NHS Trust in the UK. LA-REIMS analysis of an IBD patient cohort was carried out as part of the "Refining the use of metabolic and metagenomic profiling in the diagnosis and characterisation of inflammatory bowel disease" study (PI Prof. Timothy Orchard), an observational cohort study collecting urine, faeces, blood and tissue of patients with UC and CD.

Patients were recruited from gastroenterology, diabetes, cardiology, respiratory clinics from Imperial College Healthcare Trust at the St Mary's Hospital and the Hammersmith Hospital sites, and London Northwest University Hospital's NHS Trust, at the Ealing Hospital site between August 2015 and November 2016. Healthy controls were recruited from subjects working at ICHT and Imperial College London (St Mary's Hospital). IBD diagnosis was based on clinical, radiological, endoscopic, and histological data taken from the patient's paper and electronic notes, and from the pathology, radiology, and endoscopy electronic reporting systems. Participants under the age of 18, pregnant, having stomas, IBD unclassified, or those having recently taken antibiotics (within 8 weeks), pre-or probiotics, or on liquid or treatment diets were excluded.

The recruitment process took place either via post/phone or via clinic visit. In the first scenario, participants were identified in clinic from medical notes according to study criteria and then a discussion with potential participants regarding the study was carried out and consent was obtained. Subsequently, a questionnaire was filled out and a time was arranged for sample collection at the point of visit or alternative visit. In the second case, participants were identified via clinic lists (according to inclusion criteria) 2 weeks prior to clinic date, a letter of invitation sent out to potential participants and a phone discussion was carried out

to invited participants 48 hours after letter of invitation sent. For participants who agreed to participate following phone call –parcels with sample pots and consent form were sent out.

The patients were then seen in clinic to go through consent form, check for questions and collect samples. Study participants were given a pack including a spoon cap sample container (Sigma-Aldrich, USA) along with instructions for collections at the clinic visit. Upon receipt, all samples were separated into microcentrifuge tubes (Eppendorf, Germany) prior to being frozen at -80°C on the same day (within 12 hours of being produced). It must be noted that the patient recruitment, sample collection, and patient metadata collection was performed under PI. Prof Timothy Orchard and his team. Faecal samples obtained from each patient were taken to Charing Cross Hospital (Imperial College London NHS Trust) for LA-REIMS analysis and stored at -80°C until analysis.

Between January 2019 and February 2019, faecal LA-REIMS analysis was carried out by the candidate. Faecal samples were thawed (at 21°C in a Class II biological cabinet) and prepared for two analytical workflows. The faecal sample and water preparation for the IBD faecal sample analysis was carried out using the 96-deep-well-plate setup, as previously outlined in section *3.4.2. CRC study*. The optimised settings from the laser and instrumentation experiments were implemented in this analysis. Specifications for the LA-REIMS analysis is outlined in section *3.3.2. LA-REIMS analysis*.

All acquired raw spectra and metadata (manually formatted according to specifications) were uploaded into an in-house built package developed in RStudio v1.1.419 for peak picking of LA-REIMS spectra, as described in section *3.3.3. Data processing and statistical analysis.* The diagnostic accuracy of LA-REIMS for patients with CD, UC and non-IBD control based on faecal samples was assessed utilising multivariate statistics. The supervised models build underwent cross validation using LOOCV and RFC and ROC curves were generated. Details are outlined in section *3.3.3.4. Further processing – Metaboanalyst.*

Further, patient related factors, including among others age, sex, BMI, ethnicity, comorbidities, medication use, smoking and drinking status were investigated using

unsupervised multivariate statistics to establish any major trends in the data based on these factors. Statistically different features between CD, UC and control were interrogated to identify discriminatory metabolites using univariate statistics, as described in 3.3.3.4. Further processing – Metaboanalyst. The tentative identification of metabolites and the attempt at validation was carried out as described in section *3.3.3.5 Metabolite Identification* and *3.3.3.6 Tentative identification – validation.*

# 4. Optimisation of LA-REIMS for faecal metabolomic analysis

## 4.1. Results and Discussion

The novel application of LA-REIMS to faecal sample analysis necessitates careful evaluation of sample collection, preparation as well as the optimisation of analytical parameters. The analysis of faeces introduces several challenges, including inter- and intra-individual variability based on diet, genetic, and environmental factors, variability in the water content of samples, as well as post-collection metabolite deterioration (microbial fermentation) (114). Faecal sample storage has been previously reported by our group (149) and is therefore out of the scope of this work.

Faecal samples were analysed as soon as possible after recruitment/ participant bowel emptying, based on the results reported by our group as well as existing recommendation in the faecal metabolomics field (149). If analysis was not feasible straight away, faecal samples were stored at -80°C until sample analysis and freeze/thawing cycles were kept at a minimum. The aim of this work was to optimise direct from crude sample analysis with little-to-no sample preparation as well as faecal water extract analysis, where mostly polar molecules, including low-molecular weight metabolites, could be targeted.

To investigate the polar, hydrophilic faecal metabolome in a more targeted fashion using LA-REIMS, faecal aqueous or faecal water extracts were investigated. This approach was utilised to target polar metabolites and to therefore assess a more inclusive coverage of the faecal metabolite profile. In the context of this work, faecal water extracts refer to the supernatant of a faecal aqueous extraction, details of the methodology can be found in section *3.3.1.1. Laser optimisation* (112,114). Faecal water extracts are polar and may contain food components, intestinal excretions and soluble host-microbial products (33). These metabolites play a key role in GI health and can therefore be targeted in LA-REIMS analysis. Furthermore, the consideration of faecal water extracts, as opposed to organic solvent

extractions was dictated by the need to develop a simple technique for sample preparation that could be carried out in hospitals or settings outside of the laboratory.

Metabolites are chemically diverse, and this often required to analyse the biological sample in both positive and negative ion detection modes. Metabolites containing only Carbon, Hydrogen and Oxygen may be detected in negative ion detection mode; on the other hand, metabolites, which contain Nitrogen, would be ionised in positive ion detection mode (152). In REIMS studies, negative ion detection mode has traditionally yielded superior spectral quality (149). However, for the purpose of this work it was deemed important to maximise metabolome coverage, and therefore carry out the optimisation work in both negative and positive ion detection modes.

### 4.1.1. Laser optimisation

Since the first development of REIMS, which was originally coupled with diathermy, a key modification has been the move to a laser-based setup. The REIMS interface has been coupled to several lasers, including ultraviolet (UV) and infrared (IR) lasers; most studies using the laser-coupled REIMS setup were based on the analysis tissue and cells, as well as biofluids with high water content, such as for instance milk (120,123,139,318,328). Details of the development and methodology of the LA-REIMS setup can be found in section *3.3.2. LA-REIMS analysis.*

A total of five healthy volunteers (n=5; 3F:2M), with no known history of GI disease, donated faecal samples for the LA-REIMS laser optimisation. Participant information is given in Table 8. Of the five faecal samples, two were graded as Type 3 and three as Type 4 according to the Bristol Stool Chart (BSC) (329).

Table 8: Participant Information for the five healthy volunteers is given

Note: Age is given in years. Smoking status, diet, and medication are self-reported

| Participant | Bristol Stool Scale | Bowel motions per day in the last 2 weeks | Age | Gender | Ethnicity | Alcohol (units/week) | Smoking Status (If smoker how many per day) | BMI | Diet | Medication | Antibiotics use | Probiotics/ prebiotics use | Vitamins, minerals, or dietary supplements | Known GI disorder | GI surgery in the past |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Volunteer 1 | 4 | 3/day | 22 | Male | White British | 10 | Smoker (5/day) | 20 | Vegetarian | None | Nil | None | Multivitamins | None | None |
| Volunteer 2 | 4 | 1/day | 22 | Female | White Other | 2 | Non-smoker | 25 | Meat eater | None | Nil | None | None | None | None |
| Volunteer 3 | 4 | 1/day | 23 | Male | White British | 20 | Smoker (1/day) | 24.6 | Meat eater | None | Nil | None | None | None | None |
| Volunteer 4 | 3 | 1/day | 22 | Female | British Asian | 5 | Non-smoker | 20.3 | Meat eater | None | Nil | Probiotics | None | None | None |
| Volunteer 5 | 3 | 1/day | 23 | Female | White Other | 5 | Non-smoker | 21.2 | Meat eater | None | Nil | None | Multivitamins | None | None |

Whilst REIMS has been utilised in the analysis of faecal samples, the use of high-throughput LA-REIMS in faecal sample and water extract analysis (negative and positive ion detection mode), presents a novel pipeline. Optimisation of laser operational parameters included power and pulse mode and length. Specifically, heating power 2-5 watts at 0.5 watts increments, laser beam super pulse on versus off, and laser beam pulse (continuous, repeat 40 ms, repeat 60 ms, repeat 80 ms, or repeat 100ms) were assessed, a simplified diagrammatic representation is provided in Figure 12. Heating power refers to Watts utilised for each laser burn, ranging from 2-5 watts at 0.5 watts increments. Super pulse off refers to a continuous wave operation (a continuous laser beam is emitted with a controlled output) whereas super pulse on refers to a pulsatile wave operation whereby short, high peak power pulses, with breaks in between pulses; both give the same energy per second, see Figure 12. Furthermore, laser beam pulse length (continuous, repeat 40 ms, repeat 60 ms, repeat 80 ms, or repeat 100ms) refers to the time interval of laser firing utilised, Figure 12.

To expand the range of applications of LA-REIMS analysis and to investigate the whole range of metabolites present in faeces, ranging from lipophilic to polar metabolites, LA-REIMS analysis of faecal water extracts was additionally considered in this thesis. Faecal water analysis offers a potential new target for studying water-soluble molecules key to the pathogenesis of GI diseases, which may not be visible using crude faecal sample analysis. Moreover, faecal water extraction is a simple, quick, and a low-cost preparation step in comparison with some organic solvent extractions.

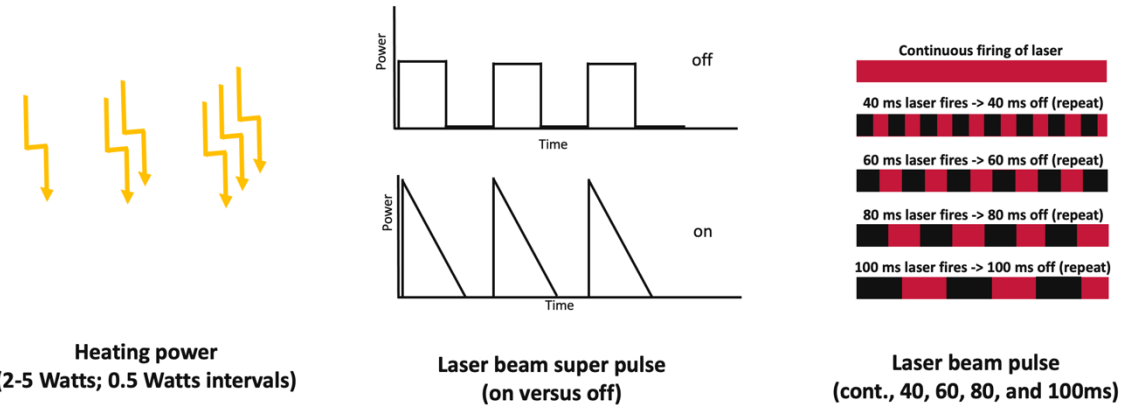



Figure 12: Parameters assessed for the laser optimisation of the high-throughput LA-REIMS instrument with heating power 2-5 watts at 0.5 watts increments, laser beam super pulse on versus off, and laser beam pulse (continuous, repeat 40 ms, repeat 60 ms, repeat 80 ms, or repeat 100ms) tested. All possible combinations of laser parameters were tested, yielding a total of 70 conditions.

The analysis was divided into (1) faecal sample and (2) faecal water, with both evaluated in (A) negative and (B) positive ion detection modes. Furthermore, signal intensity values were calculated for three spectral regions including an area defined as noise (m/z 50 to 51 m/z), fatty acid and low molecular weight metabolites (m/z 200-500), and complex lipid range (m/z 600-1000). Noise was defined as the m/z range 50-51 as was previously reported by our group in 2019 (328). The signal-to-noise ratios in the complex lipids (*m/z* 600-1000) and low mass (*m/z* 200-500) range were calculated, detailed methodology is described in section *3.4.1.1. Laser Optimisation.*

#### *4.1.1.1. Faecal sample negative ion detection mode – Signal-to-noise ratios*

For faecal sample analysed in negative ion detection mode, heating power 3.5 watts, laser beam pulse on/off at 100 ms, and super pulse off (3.5/100ms/off) was the optimum setting determined. Figure 13 shows the heatmaps of the complex lipids (*m/z* 600-1000)-to-noise (m/z 50-51) ratios and Figure 14 shows the heatmaps of the low mass (*m/z* 200-500)-to-noise

(m/z 50-51) ratios. 3.5/100ms/off was chosen as it has the highest signal-to-noise ratios in both the complex lipid and low mass/fatty acid range.

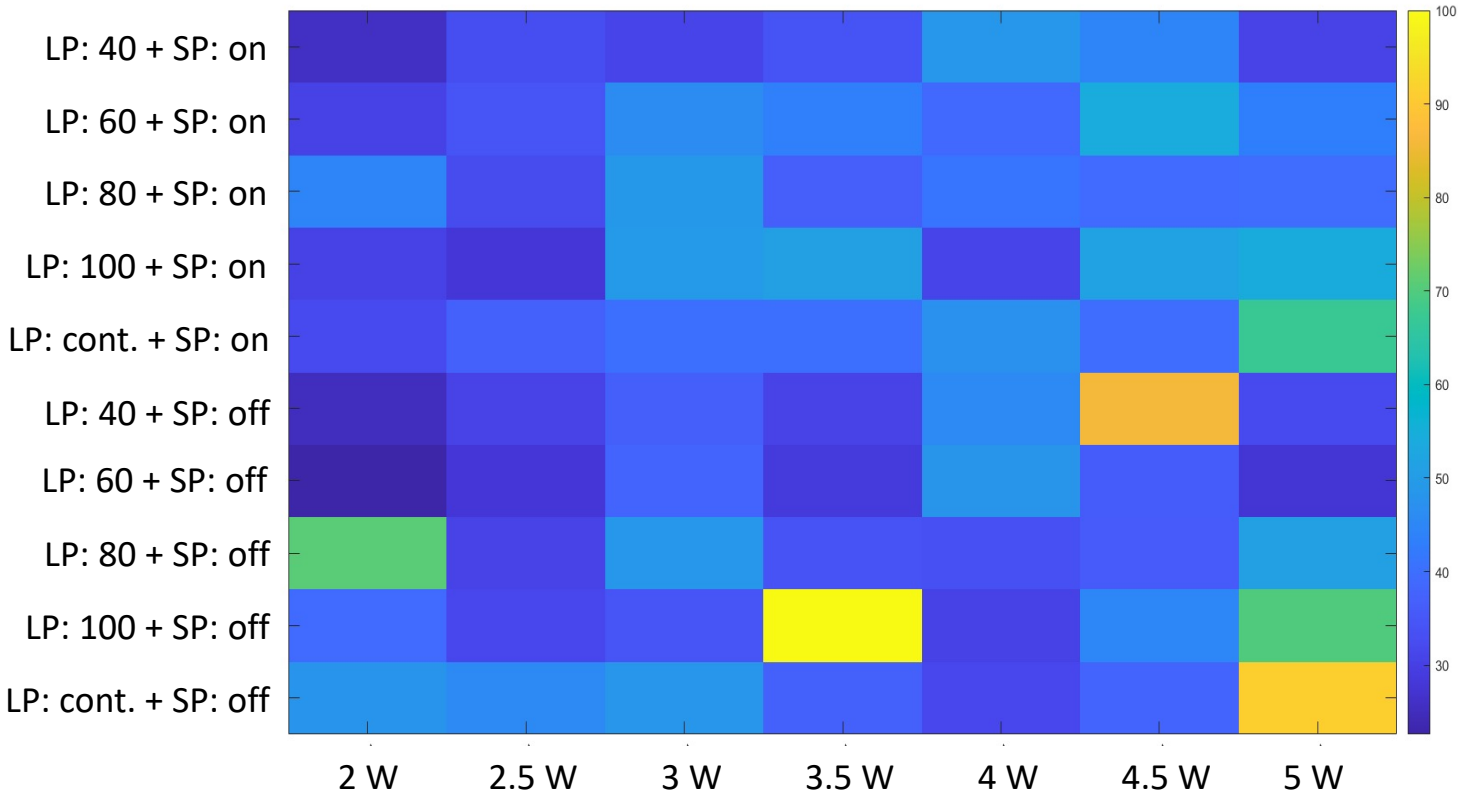



Figure 13: Heatmap of complex lipid (m/z 600-1000)-to-noise (m/z 50-51) ratio in faecal sample negative ion detection mode at parameters assessed: Heating power 2-5 watts at 0.5 watts increments, laser beam super pulse (sp) on versus off with laser beam pulse (bp) (continuous, repeat 40 ms, repeat 60 ms, repeat 80 ms, or repeat 100ms). The colour code represents the calculated signal-to-noise ratio at each parameter combination assessed.

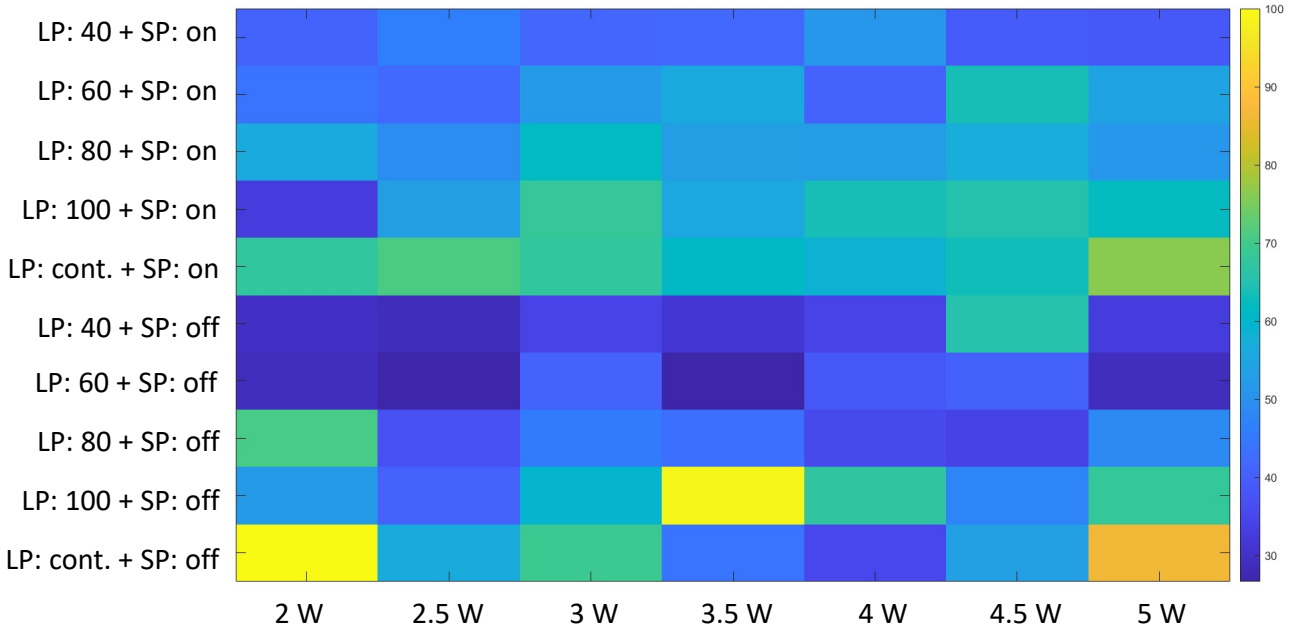



Figure 14: Heatmap of low-mass (m/z 200-500)-to-noise (m/z 50-51) ratio in faecal sample negative ion detection mode at parameters assessed: Heating power 2-5 watts at 0.5 watts increments, laser beam super pulse on versus off with laser beam pulse (continuous, repeat 40 ms, repeat 60 ms, repeat 80 ms, or repeat 100ms). The colour code represents the calculated signal-to-noise ratio at each parameter combination assessed.

#### *4.1.1.2. Faecal water negative ionisation mode – Signal-to-noise ratios*

For faecal water analysed in negative ion detection mode, heating power 3 watts, laser beam pulse continuous, and super pulse off (3/cont./off) was the optimum setting determined. Figure 15 shows the heatmaps of the complex lipids (*m/z* 600-1000)-to-noise (m/z 50-51) ratios and Figure 16 the low mass (*m/z* 200-500)-to-noise (m/z 50-51) ratios. In the consideration of the optimum parameter combination chosen for faecal water negative mode analysis, the setting providing the highest signal-to-noise ratio in the complex lipids (*m/z* 600-1000)-to-noise (m/z 50-51) was preferential, whilst also providing moderately high signal-to-

noise ratio in the low-mass (*m/z* 200-500)-to-noise (m/z 50-51) model. The settings combination at 3 watts, laser beam pulse continuous, and super pulse off (3.5/80ms/off), demonstrating the highest providing the highest signal-to-noise ratio in the low mass model was not chosen as this yielded a poor signal-to-noise ratio in the complex lipids model.

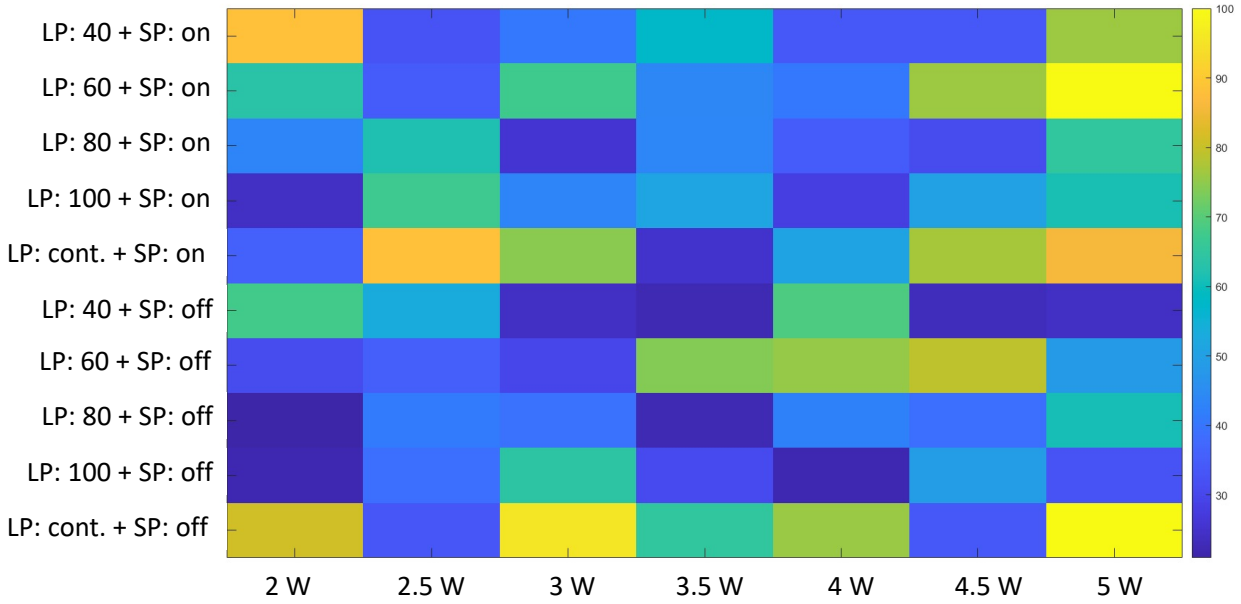



Figure 15: Heatmap of complex lipid (m/z 600-1000)-to-noise (m/z 50-51) ratio in faecal water negative ion detection mode at parameters assessed: Heating power 2-5 watts at 0.5 watts increments, laser beam super pulse (sp) on versus off with laser beam pulse (bp) (continuous, repeat 40 ms, repeat 60 ms, repeat 80 ms, or repeat 100ms). The colour code represents the calculated signal-to-noise ratio at each parameter combination assessed.

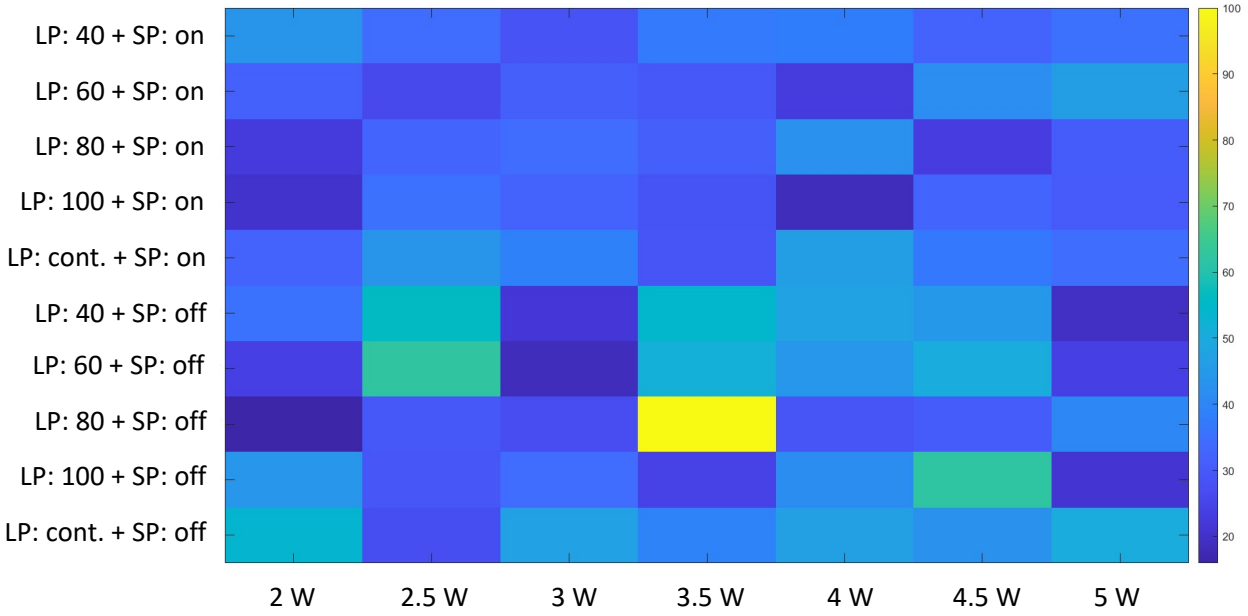



Figure 16: Heatmap of low mass (m/z 200-500)-to-noise (m/z 50-51) ratio in faecal water negative ion detection mode at parameters assessed: Heating power 2-5 watts at 0.5 watts increments, laser beam super pulse (sp) on versus off with laser beam pulse (bp) (continuous, repeat 40 ms, repeat 60 ms, repeat 80 ms, or repeat 100ms). The colour code represents the calculated signal-to-noise ratio at each parameter combination assessed.

#### *4.1.1.3. Faecal sample positive ion detection mode – Signal-to-noise ratios*

For faecal sample analysed in positive ion detection mode, heating power 2.5 watts, laser beam pulse continuous, and super pulse on (2.5/cont./on) was the optimum setting determined. Figure 17 shows the heatmap of the complex lipids (*m/z* 600-1000)-to-noise (m/z 50-51) ratio and Figure 18 shows the low mass (*m/z* 200-500)-to-noise (m/z 50-51) ratios. 2.5/cont./on was chosen as it has a high signal-to-noise ratios in the complex lipid region (Figure 17) whilst also providing a relatively high signal-to-noise ratio for the low mass model (Figure 18).

In the comparison of negative and positive ion detection mode for faecal sample analysis in the context of laser parameters, minimal differences across the heating powers were observed, with optimal heating powers ranging from 2.5-3.5 W. A slightly higher heating power, i.e., 3.5W vs 2.5W produced a higher signal-to-noise ratio for negative vs positive mode analysis, respectively. Although a small difference, possible explanations include a higher heating power needed in negative ion detection mode for the mobilisation of samples and a lower heating power needed in positive mode to reach a point of saturation for analysis.

The use of super pulse on vs off demonstrated variable signal-to-noise ratios based on ion detection mode utilised. Faecal sample positive ion detection mode demonstrated superior signal-to-noise ratios with Super Pulse on, whereas faecal sample negative ion detection mode demonstrated superior signal-to-noise ratios with super pulse off. In super pulse mode on, a short high-power beam is passed at the beginning of the analysis with breaks in between subsequent pulses whereas in super pulse off a continuous wave operation is used (a continuous laser beam is emitted with a controlled output). Furthermore, in terms of pulse mode, optimal laser beam pulse length was achieved with either continuous (faecal sample positive and faecal water negative ionisation mode) or repeat intervals 100ms (faecal sample negative ionisation mode) laser firing. Possible explanations for differences in super pulse mode and pulse length include differences in the threshold for fragmentations of complex lipid species, as previously hypothesised by in research previously published by our group (151) in the optimisation of LA-REIMS for bacterial sample analysis.

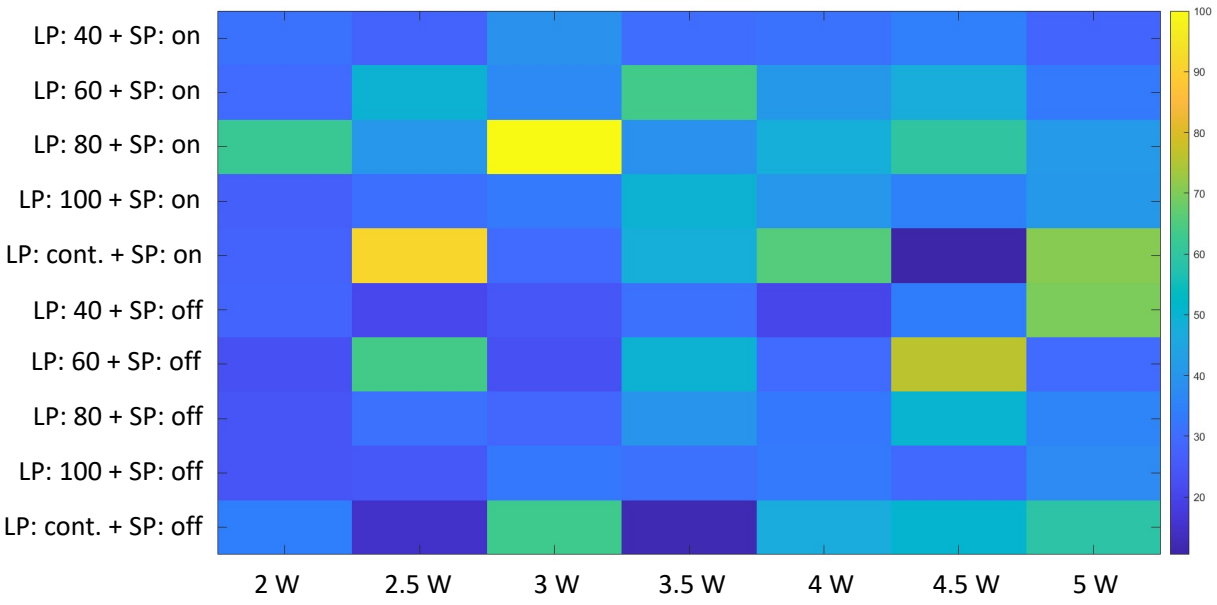



Figure 17: Heatmap of complex lipid (m/z 600-1000)-to-noise (m/z 50-51) ratio in faecal sample positive ion detection mode at parameters assessed: Heating power 2-5 watts at 0.5 watts increments, laser beam super pulse (sp) on versus off with laser beam pulse (bp) (continuous, repeat 40 ms, repeat 60 ms, repeat 80 ms, or repeat 100ms). The colour code represents the calculated signal-to-noise ratio at each parameter combination assessed.

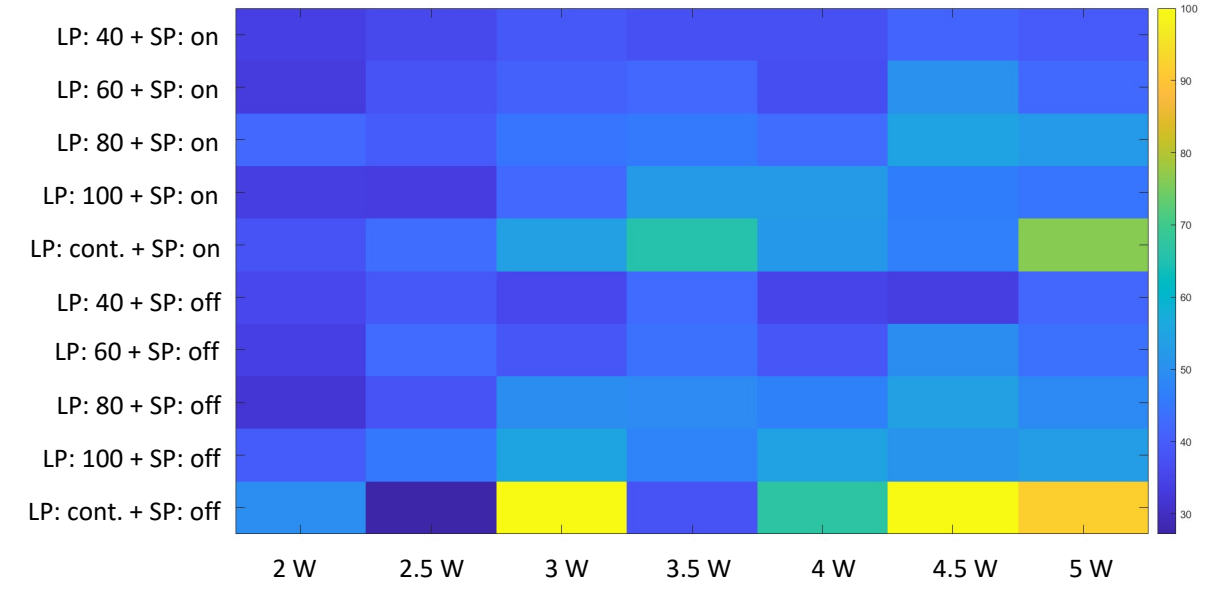



Figure 18: Heatmap of low-mass (m/z 200-500)-to-noise (m/z 50-51) ratio in faecal sample positive ion detection mode at parameters assessed: Heating power 2-5 watts at 0.5 watts increments, laser beam super pulse (sp) on versus off with laser beam pulse (bp) (continuous,

repeat 40 ms, repeat 60 ms, repeat 80 ms, or repeat 100ms). The colour code represents the calculated signal-to-noise ratio at each parameter combination assessed.

#### *4.1.1.4. Faecal water positive ion detection mode – Signal-to-noise ratios*

For faecal water analysed in positive ion detection mode, heating power 3 watts, laser beam pulse on/off at 80 ms, and super pulse on (3/80ms/on) was the optimum setting determined. Figure 19 shows the heatmap of the complex lipids (*m/z* 600-1000)-to-noise (m/z 50-51) ratio and Figure 20 shows the low mass (*m/z* 200-500)-to-noise (m/z 50-51) ratio. 3/80ms/on was chosen as it has the highest signal-to-noise ratios in the complex lipid region (Figure 19) whilst also providing a relatively high signal-to-noise ratio in the fatty acid range (Figure 20).

For faecal water extract analysis optimal heating powers observed were at 3 W for both negative and positive ion detection mode analysis. This may be due to the same heating power needed to reach the point of saturation in positive and negative ion detection mode for faecal water analysis. Overall, optimal heating powers tested ranged from 2.5 -3.5 W, demonstrating relatively small differences between the different conditions. Faecal water and sample positive ion detection mode demonstrated superior signal-to-noise ratios with super Pulse on, whereas faecal sample and water negative ionisation mode demonstrated superior signal-to-noise ratios with super pulse off. Furthermore, in terms of pulse mode, optimal laser beam pulse length was achieved with either continuous (faecal sample positive and faecal water negative ionisation mode) or with repeat intervals of 80 ms (faecal water positive ionisation mode) or repeat intervals 100ms (faecal sample negative ionisation mode) laser firing.

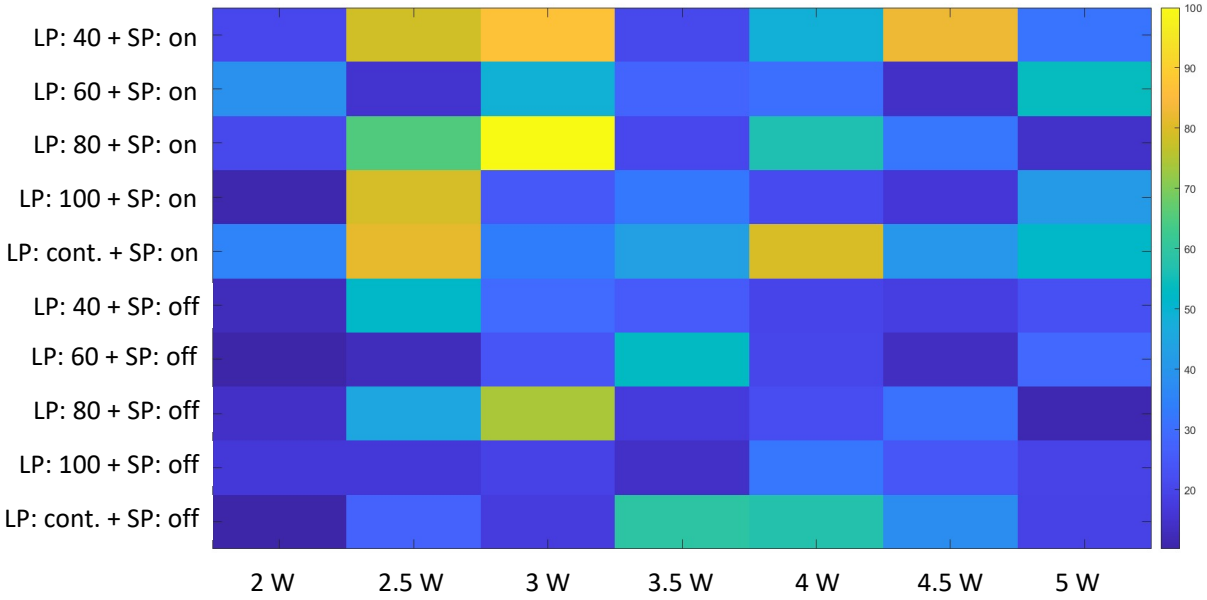



Figure 19: Heatmap of complex lipid (m/z 600-1000)-to-noise (m/z 50-51) ratio in faecal water positive ion detection mode at parameters assessed: Heating power 2-5 watts at 0.5 watts increments, laser beam super pulse (sp) on versus off with laser beam pulse (bp) (continuous, repeat 40 ms, repeat 60 ms, repeat 80 ms, or repeat 100ms). The colour code represents the calculated signal-to-noise ratio at each parameter combination assessed.

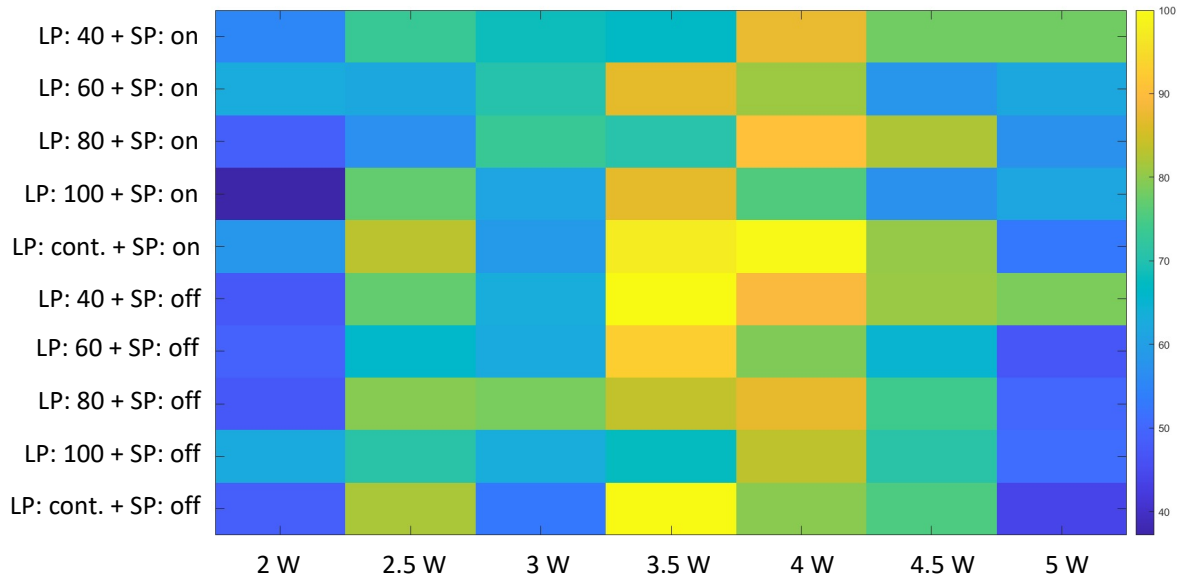


Figure 20: Heatmap of low mass (m/z 200-500)-to-noise (m/z 50-51) ratio in faecal water positive ion detection mode at parameters assessed: Heating power 2-5 watts at 0.5 watts increments, laser beam super pulse (sp) on versus off with laser beam pulse (bp) (continuous, repeat 40 ms, repeat 60 ms, repeat 80 ms, or repeat 100ms). The colour code represents the calculated signal-to-noise ratio at each parameter combination assessed.

Early development and optimisation work on REIMS bacterial (330) and faecal analysis (149) have been previously reported by our group (149). Parameters, including basic instrument parameters and voltages, lock-mass compound infusion, interface configuration, instrument ionisation mode, and the analysis mass range of m/z 50-1200 for acquisition were kept constant for the experiments of this thesis. Our group previously reported (149) the first application of REIMS monopolar diathermy to the metabolic profiling of faeces. Faecal samples (n=5 males) were used to optimise analytical parameters. The REIMS platform utilised radiofrequency electrical currents passed into the sample via the use of a stainless-steel monopolar probe. In terms of the optimisation carried out, this concentrated on identifying the optimum heating power in Watts (W), with levels between 10W and 50W, increasing in 5W increments, investigated. Both the low molecular weight region (m/z 50-500) and the lipid region (m/z 600-1000) demonstrated best signal intensity at 35 W.

Whilst this initial work in REIMS faecal metabolomics (149) optimised heating parameters, it was limited in both the number of participants in the study as well the variables assessed, i.e. limited to faecal sample, negative ion detection mode analysis, with heating power assessed only. It must also be noted, that since the initial work carried out in faecal REIMS analysis, the instrumentation saw a change from heating with a monopolar electrode to the use of a $CO_2$ laser using joule heating for sample analysis, necessitating a novel approach to the optimisation of instrumentation parameters.

In subsequent work, our group (151) presented the first application of the laser assisted REIMS setup for faecal metabolomics analysis. The lyophilised faecal samples from a type 2 diabetes mellitus cohort were analysed with LA-REIMS, with tentative identifications potentially associated with disease pathogenesis and demonstrated a disease classification

accuracy of 94%. Furthermore, our group (151) carried out a laser heating power optimisation, here different combinations of laser power (1 to 5 W in 0.5 W increments) and distance (1 mm to 5 mm in 1 mm increments) from the sample were tested. The optimal parameter combination result was a laser power setting was 4.0W at a 3 mm distance of laser to sample (151).

The optimal heating power of the laser of 4.0W for faecal sample negative ion detection mode analysis is line with the findings of the work presented in this chapter, i.e., faecal sample negative ionisation mode was found to have an optimal heating power of 3.5 W. This demonstrates similar findings for two separate independent sets of homogenised faecal samples. A possible explanation for the similarity in the results is that both the faecal samples previously published by our group (151) and the faecal samples assessed in the optimisation work were categorised as BSC 3-4. Therefore, the water content and consistency of the faecal samples analysed were similar, yielding a comparable matrix.

Furthermore, the faecal sample analysed by Cameron et al (151) were freeze-dried faeces reconstituted in water mixture, not crude faecal samples. As the benefits of LA-REIMS lies in sample preparation free high-throughput and automated analysis, an important factor to consider is the optimisation of laser and instrumentation parameters on crude faecal samples (no lyophilisation step). It has to be noted that the metabolic profile of crude faecal samples and lyophilised faecal samples have been shown to differ (33). Crude faeces and lyophilised faeces are two distinct matrices, that each require independent optimisation of parameters.

The optimisation by Cameron, et al (151) of parameters for faecal sample analysis was limited to laser power and optimal distance from laser to the faecal sample testing. Several parameters with regards to laser settings, such as for instance, laser beam Super Pulse and laser beam pulse (see section *3.4.1.1. Laser optimisation* for details) were not assessed. Furthermore, instrumentation parameters such as the testing of optimal solvent flow rate and time intervals between subsequent burns were not assessed for faecal samples. These parameters are important to consider as they may impact the ionisation potential of molecules present as well as the likelihood of fragmentation of complex molecular species.

### 4.1.2. REIMS instrumentation parameters optimisation

Reproducible analysis with high sample throughput and accuracy is of key importance to faecal metabolomics LA-REIMS fingerprinting. Instrumentation parameters were assessed to identify parameters combination providing the least % carry-over between samples and smallest time intervals between analyses. A total of four healthy volunteers (n=4; 3F:1M) were recruited for the LA-REIMS instrumentation optimisation of faecal samples. Participant information is given in Table 9. Of the four stool samples, two were graded as Type 3, one was graded as Type 5, and one as Type 4 according to the BSC (329).

Table 9: Participant Information for the five healthy volunteers

Note: Age is given in years. Smoking status, diet, and medication are self-reported.

| Participant | Bristol Stool Scale | Bowel motions per day in the last 2 weeks | Age | Gender | Ethnicity | Alcohol (units/ week) | Smoking Status (If smoker how many per day) | BMI | Diet | Medication | Antibiotics use | Probiotics/p rebiotics use | Vitamins, minerals, or dietary supplements | Known GI disorder | Major GI surgery in the past |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Volunteer 1 | 5 | 1/day | 23 | Male | White British | 20 | Smoker (1/day) | 24.6 | Meat eater | None | Nil | None | None | None | None |
| Volunteer 2 | 3 | 1/day | 22 | Female | British Asian | 5 | Non-smoker | 20.3 | Meat eater | None | Amoxicillin 500 mg 2/day on 6-day course | Probiotics | None | None | None |
| Volunteer 3 | 3 | 1/day | 23 | Female | White Other | 5 | Non-smoker | 21.2 | Meat eater | None | Nil | None | Multivitamins | None | None |
| Volunteer 4 | 4 | 1/day | 22 | Female | White Other | 2 | Non-smoker | 25 | Meat eater | None | Nil | None | None | None | None |

LA-REIMS instrumentation settings were optimised in terms of time intervals between subsequent sample burns (10s, 20s, 30s, 60s) and IPA flow rates (0.25 mL/s, 0.3 mL/s, 0.35 mL/s). More specifically, all possible combinations of instrumentation parameters were tested, yielding a total of 12 conditions. The 12 conditions were tested for both faecal sample and faecal water (both negative and positive ion detection modes).

In each group, for each replicate of a specific combination of instrumentation settings, the % carry-over was calculated for the ten specified *m/z* values in both the low mass (*m/z* 50-600) and complex lipid (*m/z* 600-1000) region. Detailed methodology is outlined in section *3.4.1.2. Instrumentation optimisation* for details.

In each chosen group, for each setting combination tested, the mean of four replicates for the % carry-over as well as standard deviations were calculated. Thereafter heat maps comparing the various settings were generated in R studio (V1.0.44) to allow for the selection of settings yielding the lowest % carry-over with the shortest time interval between burns.

#### *4.1.2.1. Faecal sample negative ion detection mode – % carry-over*

The % carry-over for the ten specified *m/z* values in the low mass (*m/z* 50-600), Figure 21, region, and the complex lipid (m/z 600-1200) region, Figure 22, was evaluated at a combination of parameters tested: intervals between subsequent sample burns (10s, 20s, 30s, 60s) and IPA flow rates (0.25 mL/s, 0.3 mL/s, 0.35 mL/s). IPA/0.25/30s was chosen as it is the combination of settings which provides the least % carry-over with the smallest time interval between burns to allow for a high-throughput setup. It is important to note, that in instances where parameter combinations provided similar % carry-over, the parameter combinations with smallest time interval between burns were chosen to allow for a high-throughput setup.

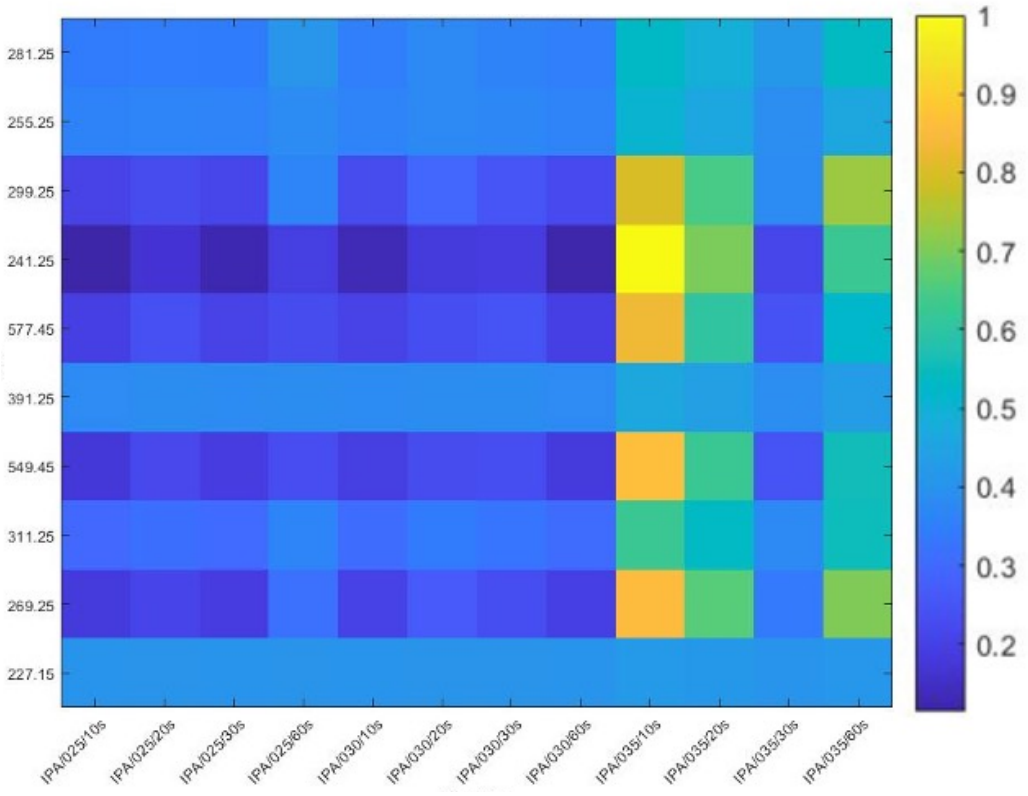



Figure 21: Faecal sample negative ion detection mode - Heatmap of the calculated % carry-over for the ten specified m/z values in the low mass (m/z 50-600) region evaluating a combination of parameters tested: intervals between subsequent sample burns (10s, 20s, 30s, 60s) and IPA flow rates (0.25 mL/s, 0.3 mL/s, 0.35 mL/s). The colour code represents the calculated % carry-over at each parameter combination assessed.

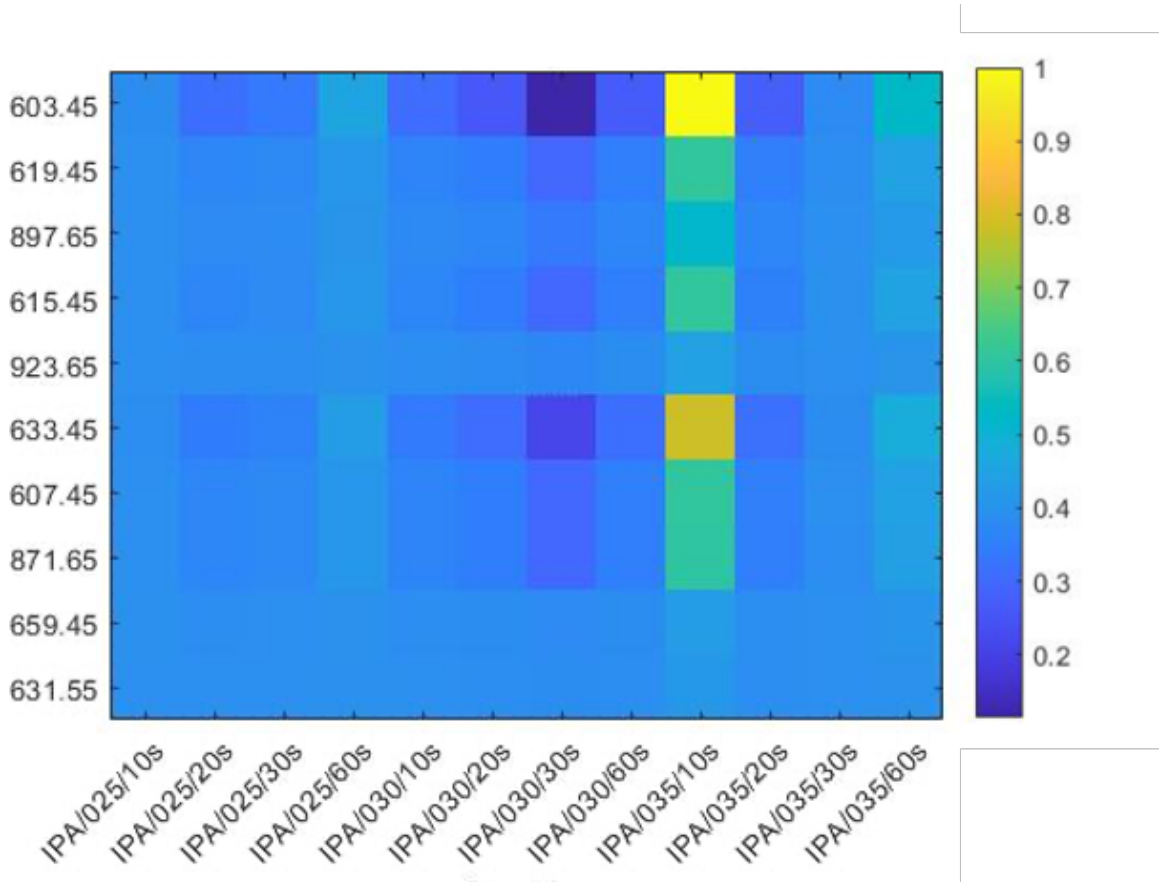



Figure 22: Faecal sample negative ion detection mode - Heatmap of the calculated % carry-over for the ten specified m/z values in the complex lipid mass (m/z 600-1200) region evaluating a combination of parameters tested: intervals between subsequent sample burns (10s, 20s, 30s, 60s) and IPA flow rates (0.25 mL/s, 0.3 mL/s, 0.35 mL/s). The colour code represents the calculated % carry-over at each parameter combination assessed.

#### *4.1.2.2. Faecal water negative ionisation mode – % carry-over*

For faecal water analysed in negative ion detection mode, IPA as the optimal solvent, at a flow rate of 0.35 mL/s, and a time interval between burns of 20 seconds (IPA/0.35/20s) was determined as optimum setting, see Figure 23-24. IPA/0.35/20s was chosen as it is the setting which provide the least % carry-over with the smallest time interval between burns to allow for a high-throughput setup in in both the low mass (Figure 23) and complex lipid (Figure 24) region.

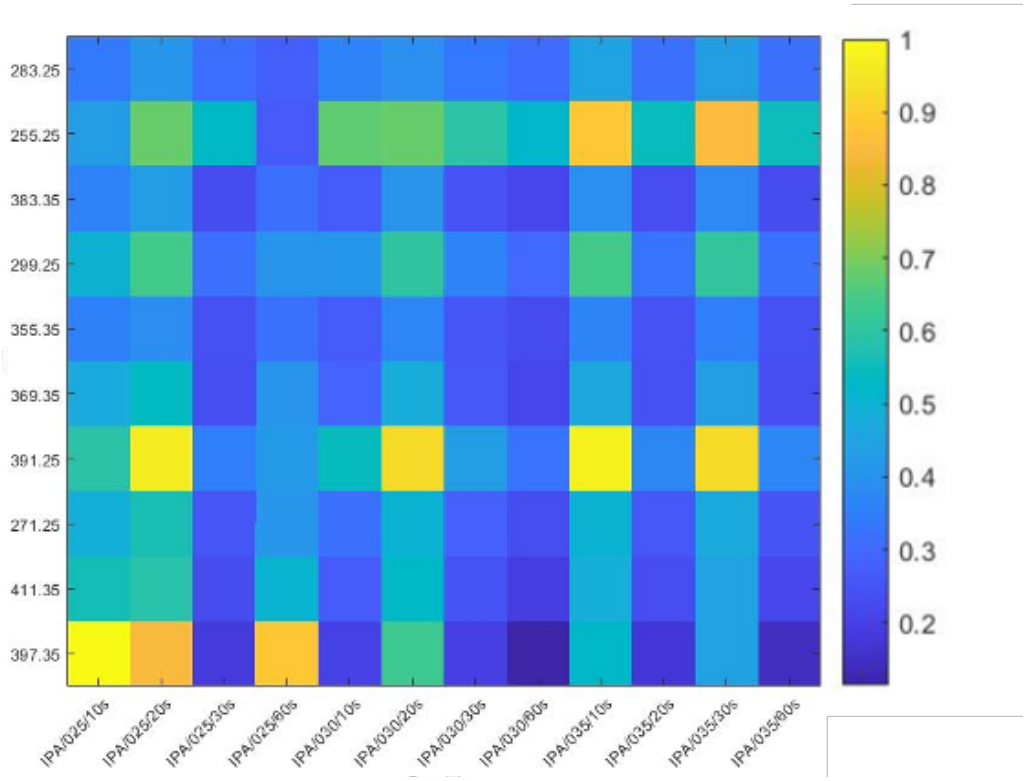



Figure 23: Faecal water negative ion detection mode - Heatmap of the calculated % carry-over for the ten specified m/z values in the low mass (m/z 50-600) region evaluating a combination of parameters tested: intervals between subsequent sample burns (10s, 20s, 30s, 60s) and IPA flow rates (0.25 mL/s, 0.3 mL/s, 0.35 mL/s). The colour code represents the calculated % carry-over at each parameter combination assessed.

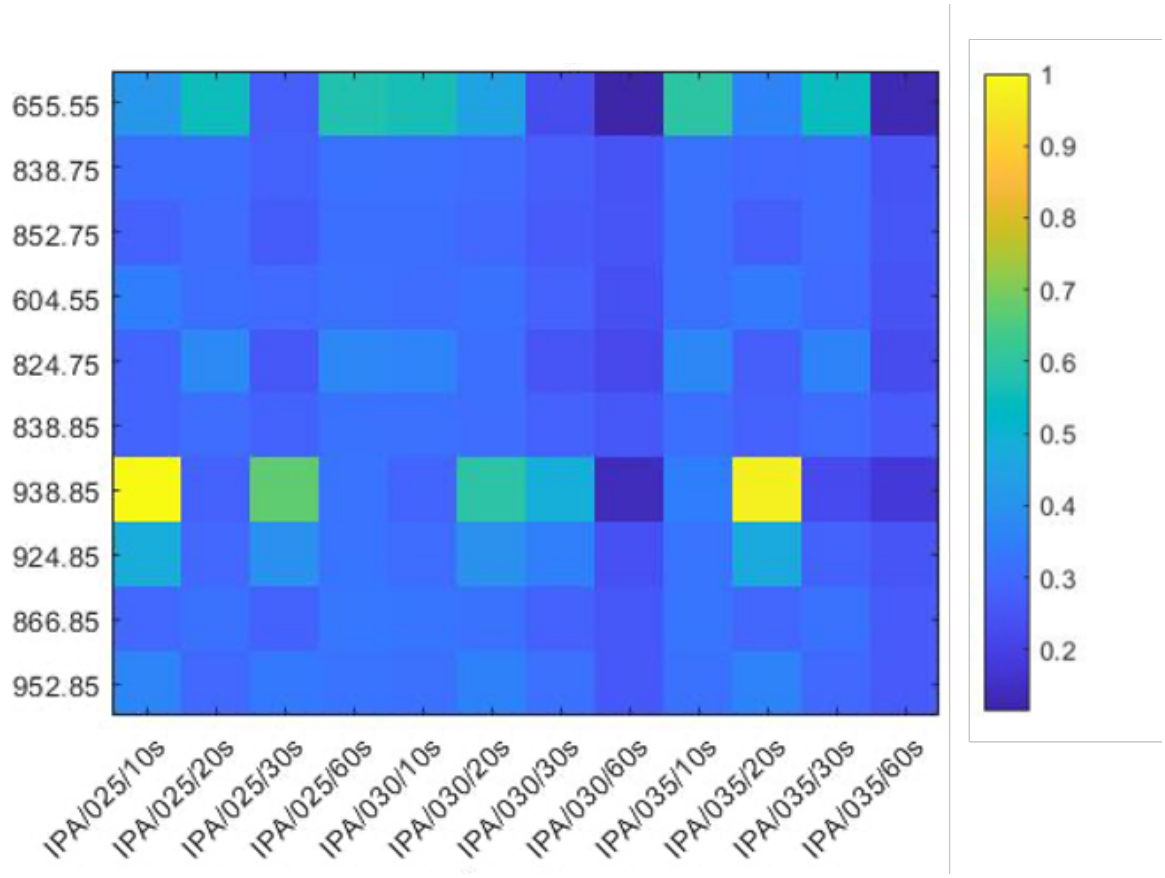



Figure 24: Faecal water negative ion detection mode - Heatmap of the calculated % carry-over for the ten specified m/z values in the complex lipid (m/z 600-1200) region evaluating a combination of parameters tested: intervals between subsequent sample burns (10s, 20s, 30s, 60s) and IPA flow rates (0.25 mL/s, 0.3 mL/s, 0.35 mL/s). The colour code represents the calculated % carry-over at each parameter combination assessed.

#### *4.1.2.3. Faecal sample positive ionisation mode – % carry-over*

The % carry-over for the ten specified *m/z* values in the low mass (*m/z* 50-600), Figure 25, and complex lipid (m/z 600-1200) region, Figure 26, was evaluated at a combination of parameters tested: intervals between subsequent sample burns (10s, 20s, 30s, 60s), IPA flow rates (0.25 mL/s, 0.3 mL/s, 0.35 mL/s). IPA/0.25/20s was chosen as it is the setting which

provide the least % carry-over with the smallest time interval between burns to allow for a high-throughput setup in both the low mass (Figure 25) and complex lipid (Figure 26) region.

In terms of instrumentation optimisation, faecal sample positive and negative ion detection mode analysis demonstrated the same optimal instrumentation parameters, with IPA flow rate of 0.25 mL/s and a time interval between burns of 20 seconds, whereas faecal water negative ion detection mode demonstrated an optimal rate at IPA/0.30/10s. Interestingly, ion detection mode does not seem to necessitate different instrumentation parameters when investigating faecal sample only, however, a change to a higher flow rate and smaller time interval between burns for faecal water vs faecal sample analysis is preferential.

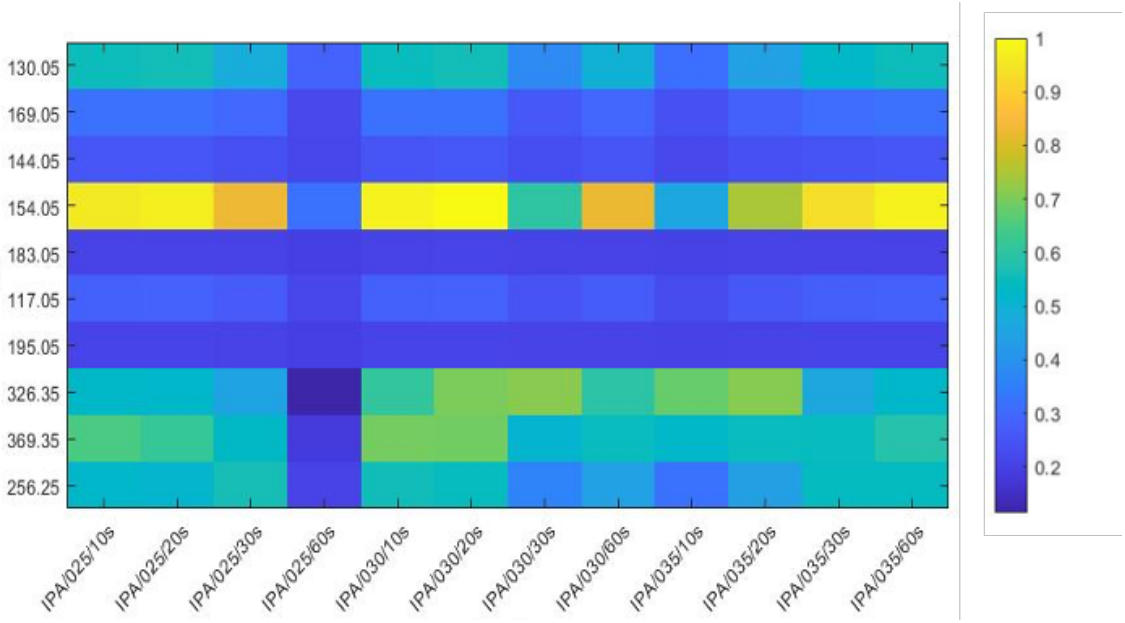



Figure 25: Faecal sample positive ion detection mode - Heatmap of the calculated % carry-over for the ten specified m/z values in the low mass (m/z 50-600) region evaluating a combination of parameters tested: intervals between subsequent sample burns (10s, 20s, 30s, 60s) and solvent IPA flow rates (0.25 mL/s, 0.3 mL/s, 0.35 mL/s). The colour code represents the calculated % carry-over at each parameter combination assessed.

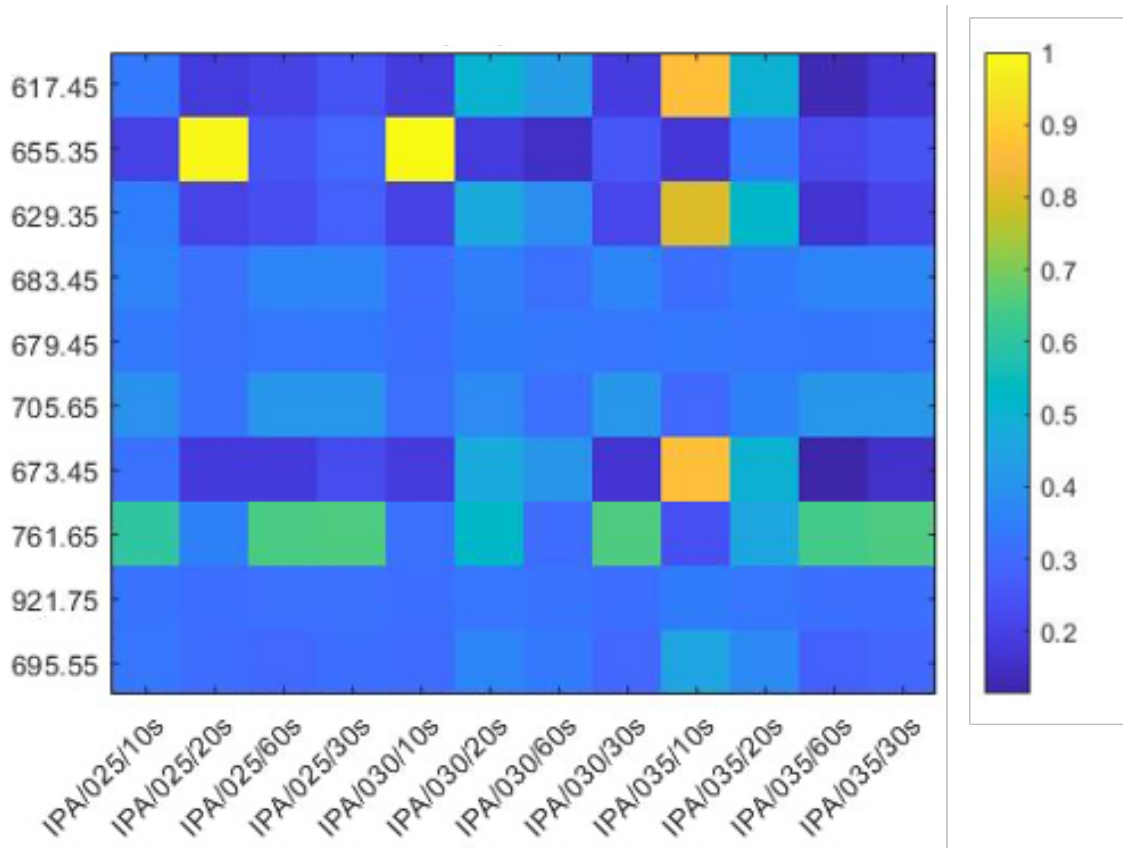



Figure 26: Faecal sample positive ion detection mode - Heatmap of the calculated % carry-over for the ten specified m/z values in the complex lipid mass (m/z 600-1200) region evaluating a combination of parameters tested: intervals between subsequent sample burns (10s, 20s, 30s, 60s) and IPA flow rates (0.25 mL/s, 0.3 mL/s, 0.35 mL/s). The colour code represents the calculated % carry-over at each parameter combination assessed.

#### *4.1.2.4. Faecal water positive ion detection mode – % carry-over*

For faecal water analysed in positive ion detection mode, IPA flow rate of 0.30 mL/s and a time interval between burns of 10 seconds (IPA/0.30/10s) was determined as optimum

setting, see Figure 27-28. IPA/0.30/10s was chosen as it is the setting which provide the least % carry-over with the smallest time interval between burns to allow for a high-throughput setup in both the low mass (Figure 27) and complex lipid (Figure 28) region.

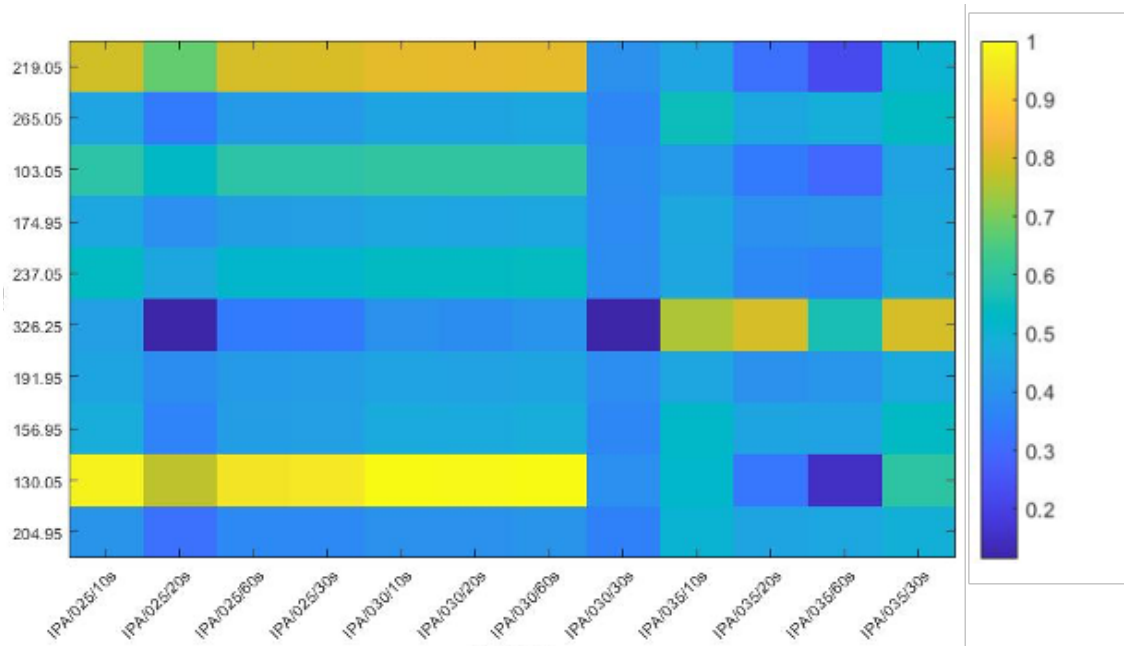



Figure 27: Faecal water positive ion detection mode - Heatmap of the calculated % carry-over for the ten specified m/z values in the low mass (m/z 50-600) region evaluating a combination of parameters tested: intervals between subsequent sample burns (10s, 20s, 30s, 60s) and solvent IPA flow rates (0.25 mL/s, 0.3 mL/s, 0.35 mL/s). The colour code represents the calculated % carry-over at each parameter combination assessed.

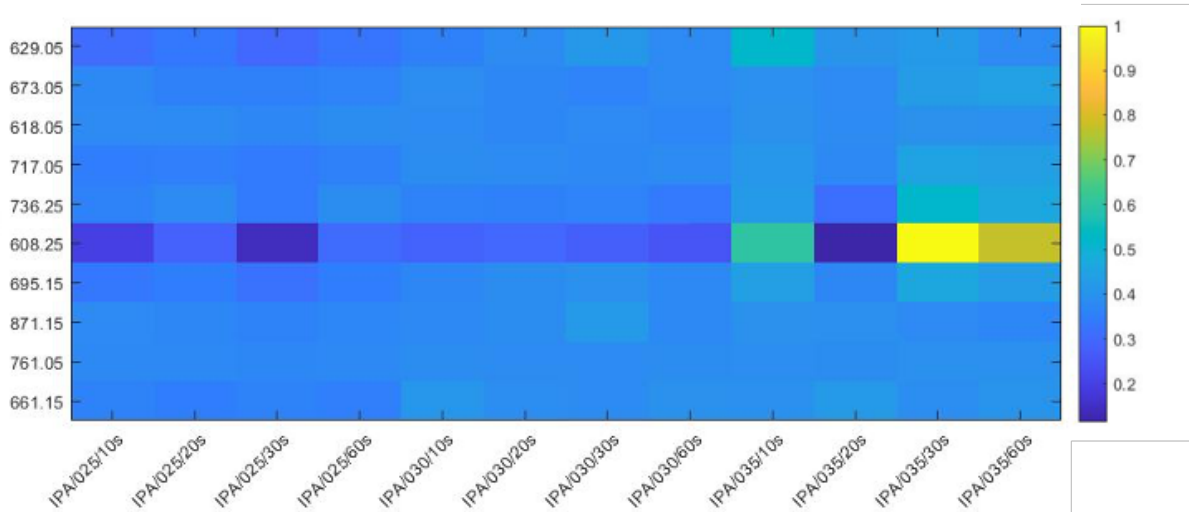



Figure 28: Faecal water positive ion detection mode - Heatmap of the calculated % carry-over for the ten specified m/z values in the complex lipid mass (m/z 600-1200) region evaluating a combination of parameters tested: intervals between subsequent sample burns (10s, 20s, 30s, 60s) and solvent IPA flow rates (0.25 mL/s, 0.3 mL/s, 0.35 mL/s). The colour code represents the calculated % carry-over at each parameter combination assessed.

In terms of instrumentation optimisation, faecal sample positive and negative mode analysis demonstrate the same optimal instrumentation parameters, that is IPA flow rate of 0.25 mL/s and a time interval between burns of 20 seconds. For faecal water analysis, a slightly higher flow rate of 0.30 mL/s and 0.35 3mL/s for negative and positive ion detection mode respectively demonstrated optimal settings.

Furthermore, similarly to faecal sample, faecal water positive ion detection mode, demonstrated optimal time interval between burns at 20 seconds, whereas faecal water negative ion detection mode was optimal at 10 seconds. A possible explanation for this is the more dilute faecal water allowing for less % carry-over between samples at a higher flow rate compared to faecal samples. Moreover, in terms of time interval between burns, all optimal intervals ranged between 10-20 seconds, with faecal water demonstrating the shortest time interval at 10 seconds. Interestingly, both sample and water extracts demonstrate least %

carry-over with similar time intervals between burns, indicating that the matrix of faecal sample versus faecal water does not play a large role in the development of carry-over between samples.

### 4.1.3. Summary of laser and instrumentation optimisation study results

The optimisation was carried out to increase analytical throughput and sensitivity and results were implemented into the analysis of all subsequent studies carried out in this work. A summary of the combinations of parameters for both laser and instrumentation optimisation is provided in Table 10.

Table 10: A summary of the combinations of parameters for laser and instrumentation faecal LA-REIMS optimisation for faecal sample and faecal water extracts (negative and positive ion detection mode)

| | Laser | Instrumentation |
|---|---|---|
| **Faecal sample negative ion detection mode** | Heating power: 3.5 W<br>Super Pulse: off<br>Pulse mode: 100 ms | Flow rate: 0.25 mL/s<br>Time interval: 20s |
| **Faecal sample positive ion detection mode** | Heating power: 2.5 W<br>Super Pulse: on<br>Pulse mode: continuous | Flow rate: 0.25 mL/s<br>Time interval: 20s |
| **Faecal water negative ion detection mode** | Heating power: 3 W<br>Super Pulse: off<br>Pulse mode: continuous | Flow rate: 0.3 mL/s<br>Time interval: 10 s |
| **Faecal water positive ion detection mode** | Heating power: 3 W<br>Super Pulse: on<br>Pulse mode: 80 ms | Flow rate: 0.35 mL/s<br>Time interval: 20s |

### 4.1.4. Limitations of the laser and instrumentation studies

A limitation of the current work is the small number of study participants, however previous studies in faecal metabolomics hold similar number (33). It must be noted that faecal samples from n=5 laser-, and n=4 instrumentation-experiment were homogenised to create a more representative faecal sample. Further, the optimisation work was limited to faecal samples BSC 3-4. Therefore as the water content of faeces has been shown to vary within the range of 63–86% across the mean value of studies (75% $H_2O$) (37), faecal samples categorised as BSC 1-2 (separate hard lumps) or 6-7 (watery, no solid material) may not be suitable for the parameters identified.

The study participants in both the instrumentation and laser optimisation work were all healthy individuals with no-known GI disease between the ages of 22-25. It must be noted that this study population differs systematically from the population of the interest of this thesis, i.e., patients who suffer from GI disease, e.g., CRC or IBD. However, it must be noted that samples were taken from multiple healthy individuals and homogenised to optimise a more representative sample; it may be argued that the analysis of a homogenate of multiple faecal samples may be a close enough approximation to the faecal matrix of people with GI disorders.

### 4.1.5. The effect of water content of samples on analysis

For this feasibility pilot study, a healthy volunteer, with no known history of GI disease, donated a faecal sample. Participant information is given in Table 11. The faecal sample was graded as Type 4 of the BSC (329).

Table 11: Participant Information for the healthy volunteer is given

Note: Age is given in years. Smoking status, diet, and medication are self-reported

| Participant | Bristol Stool Scale | Bowel motions per day in the last 2 weeks | Age | Gender | Ethnicity | Alcohol (units/week) | Smoking Status (If smoker how many per day) | BMI | Diet | Medication | Antibiotics use | Probiotics/ prebiotics use | Vitamins, minerals, or dietary supplements | Known GI disorder | GI surgery in the past |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Volunteer 1 | 4 | 1/day | 22 | Female | British Asian | 5 | Non-smoker | 20.3 | Meat eater | None | Nil | Probiotics | None | None | None |

The effect of water content on faecal LA-REIMS analysis was assessed by analysing faecal sample (n=1), at multiple ratios of HPLC grade water (VWR, Leicestershire, UK). The tested parameters included crude faeces, 1:1, 1:2, 1:6, 1:8, and 1:10 ratio of water: faeces with four repeats each. Unsupervised PCA analysis, Figure 29, indicates separation between samples with water added vs. crude raw faeces. Furthermore, intra-group variation seems to decrease as more water is added; a possible explanation is offered by water allowing for increased homogenisation of the sample. Details of the methodology are provided in section *3.4.1.3. Water content study*.

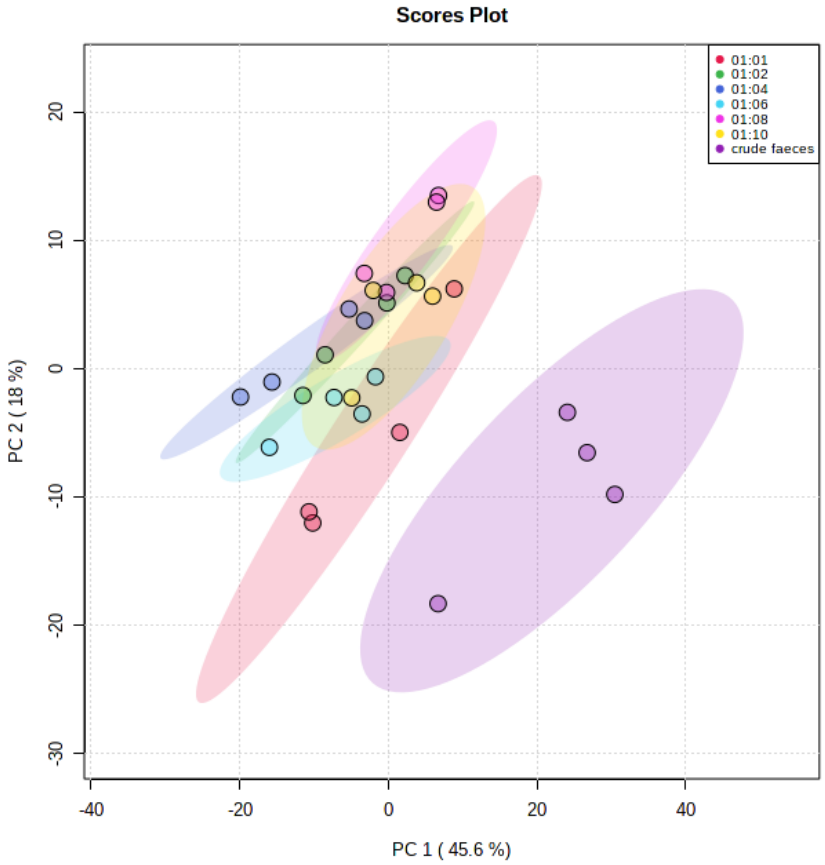



Figure 29: PCA scores plot of direct LA-REIMS faecal sample analysis at different concentrations (crude faeces, 1:1, 1:2, 1:6, 1:8, and 1:10 ratio of water: faeces) in the m/z range 50-1200 (negative ion detection mode). Scores plot of PCA analysis PC 1 (45.6%) and

PC2 (18%) tested crude and with different ratios of water added. Shaded region represents 95% confidence intervals.

Furthermore, the unsupervised PCA modelling was carried out for both the low mass (m/z 50-600) and the complex lipid region (m/z 600-1000) to determine whether the change in water content would have a different effect on either of these m/z ranges. The separation between samples with water added vs. raw faeces was more significant in the m/z range 50-600 (Figure 30A) than in the m/z range 600-1000 (Figure 30B). In the m/z range 600-1000 (Figure 30B), sample groups cluster together irrespective of the water content added to samples. The more polar lower molecular weight fatty acid region seems to be more strongly affected by the water content than the non-polar complex lipid region. It is important to note that the study was limited to a single sample and should be repeated with a higher sample size, the results should be interpreted accordingly. Furthermore, the faecal sample was not lyophilised at the beginning of the experiment. A priori lyophilisation likely serves to efficiently quench bacterial activity, limit oxidative reactions and immobilize enzymatic and non-enzymatic reactions (108), but does on the other hand change the faecal metabolomic profile.

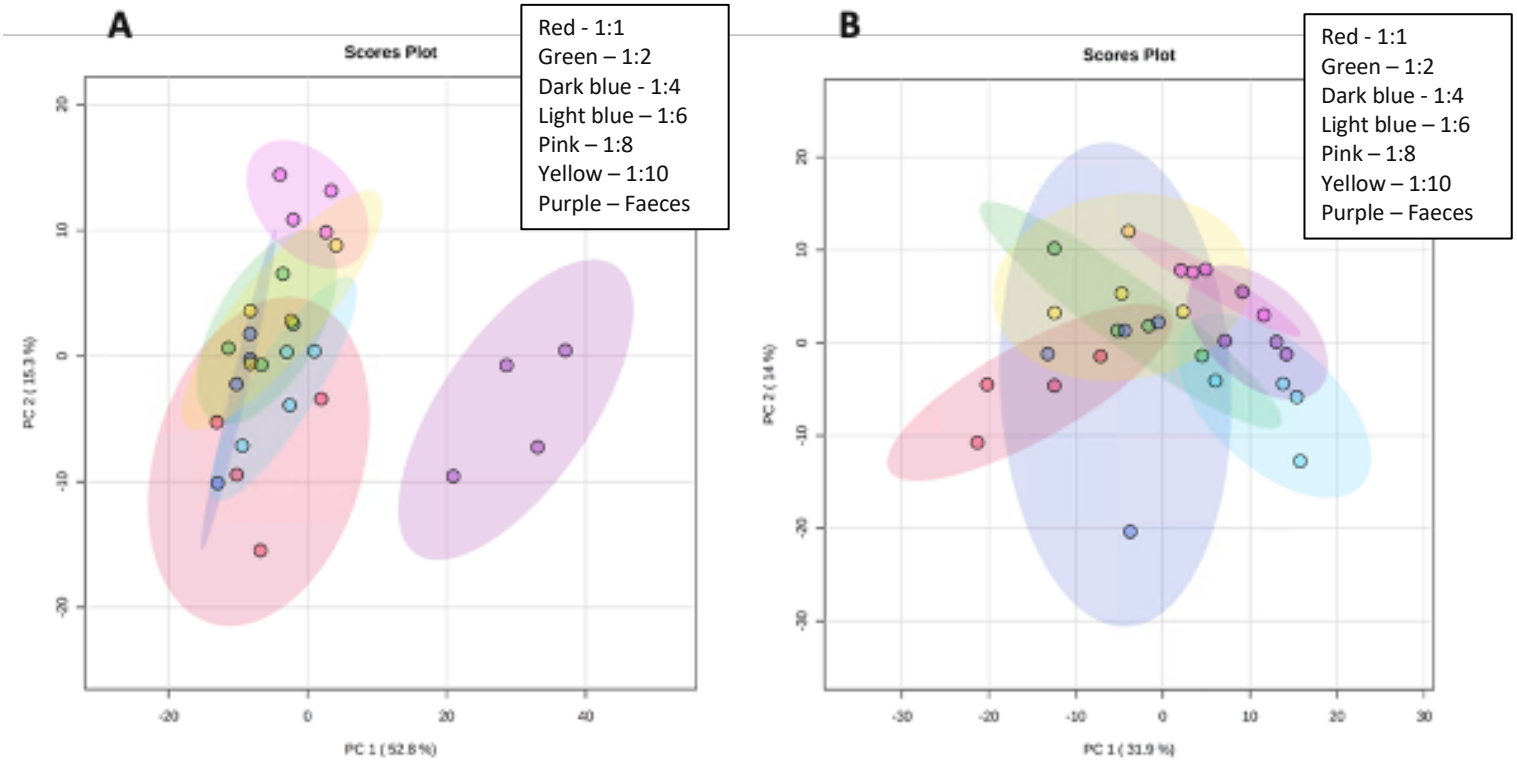


Figure 30: PCA scores plot of direct LA-REIMS faecal sample analysis at different concentrations (crude faeces, 1:1, 1:2, 1:6, 1:8, and 1:10 ratio of water: faeces) in (A) the *m/z* range 50-600 (negative ion detection mode) and (B) the *m/z* range 600-1000 (negative ion detection mode). Shaded region represents 95% confidence intervals.

The experiment evaluated the effect of water content of LA-REIMS faecal metabolomics analysis. However, this study missed to evaluate the importance of factors associated with amongst other, increased, or decreased water content, such as for instance colonic transit time, stool consistency, changes in the luminal microbial community and changes in terms of intestinal absorption mechanisms. Simply assessing water content through measurement, i.e., loss on drying experiments, or through the lyophilisation of samples is not sufficient, in controlling for all the associated effects of differences in water content. In addition to control for water content, stool consistency and transit time should be considered as a potential confounding factor and controlled for by assessment of BSC in faecal metabolomics experiments. Future work will investigate n>1 samples with a priori lyophilisation to allow for quantification of water content of samples and their impact on LA-REIMS analysis.

Furthermore, variations of faecal water content and present a challenge in the analysis of the metabolome. The median water content of faeces is 75% with a range of 60-85% (109,116). The necessity for control for variations in faecal water content has been established, with two main approaches, the direct and in-direct to correct for water content (33). For the in-direct approach, a correction for variation is carried out by normalisation (33). In faecal metabolomics studies, “constant sum normalisation” where each feature is normalised against the total ion count, is often used (33). This should reduce some of the issues of water content, however, has some issues such as bias brought on by matrix effects (331). Probabilistic quotient normalisation (PQN) is another technique for water content correction. PQN “calculates a reference spectrum and uses the median quotient estimated per variable to normalize each spectrum” (332). Although well suited for normalizing to water content, PQN demonstrates its limits faecal sample analysis, partly due to the strong effect of fibre content (332).

A more direct approach can be taken for correction of water content variability (33). For instance, faecal samples can be lyophilised by freeze-drying of homogenised wet faeces (33). The analysis of lyophilised samples leads to a loss of volatile metabolites, change in metabolic profile of the sample and it is time consuming (213). Another approach, although less accurate is to pellet the faecal sample in a cooled centrifuge and further analyse the weight of the pellet sample, which is depleted of most of the water (333). This technique is particularly good for the analysis of hydrophobic compounds but less suitable for the analysis of water-soluble metabolites.

It is important to consider variations in the faecal sample water content in the context of stool consistency and intestinal transit time. Changes in transit time and consistency of faeces, marked by diarrhoea or alternatively constipation, are frequently observed in the human population, especially with increasing age, sedentary lifestyles, and GI disturbances (329). GI transit time and consistency are of central importance to gut health and physiology, potentially affecting absorptive function (nutrients and water, small and large intestine respectively), changes to luminal microbial populations, and leading to diarrhoea or constipation with potential distressing symptoms for individuals (329).

Due to the invasive nature of direct measurement of intestinal transit time, this area has often been neglected in GI microbiome or metabolome studies. However, stool consistency measurement using the BSC, have been demonstrated to act as a proxy for colonic transit rate, allowing non-invasive assessment in a simple format (329,334). The BSC is made up of seven categories, with the highest number corresponding to loose stool and fast transit time, and the lowest score corresponding to hard stool with slower transit time (329). Each category reflects differences in water content of faecal samples – increased (fast transit time) or equally decreased water content (slow transit time) of faeces has potential effects on the luminal microbial communities. BSC can therefore be used to not only investigate stool consistency and GI transit time but also water activity and availability (334).

Recently, stool consistency and transit time measured via the BSC has been associated with gut microbiota composition. Vandeputte, et al(334) identified stool consistency to be strongly associated with faecal microbial species richness. Specifically, it was demonstrated that

species richness significantly declines with stool firmness (Spearmann's r=-0.45, p=0.0007), with the highest decrease in species richness observed in BSC 6-7. However, it must be noted that the study did not include a replication cohort and was limited to females aged 20 –55 years. Falony et al. (335) investigated faecal microbiome variation in a control population of two cohorts (total n=3948) and found (BSC) score to be the top feature covarying with faecal microbiome composition. On the other hand, Tigchgelaar et al (336) analysed the gut microbiota in relation to BSC in 1126 individuals and found no association between stool consistency and species richness. However, a positive correlation between the abundance of methanogens (*Methanobrevibacter), Clostridiaceae* and softer stool consistency was established. It is important to link and evaluate transit time, water availability, BSC, stool consistency, and the microbiota with the faecal metabolome. The metabolome should be evaluated as a complex system where changes in water content in faeces also represent fluctuations in GI transit time and stool consistency.

Overall, multiple techniques can be utilised for the correction of varying water content. As demonstrated in the experiment above, faecal water content impacts LA-REIMS metabolomic analysis. In this thesis, the studies discussed underwent indirect correction by constant sum normalisation or PQN. The benefit of LA-REIMS as a direct from sample analytical technique with minimal sample preparation needs to be emphasised. The use of an indirect technique for variation in faecal water content is the preferred modality for analysis in this thesis.

### 4.1.6. Spatial distribution of potential biomarkers in whole faecal samples

An important factor of consideration is the distribution of key metabolites in faeces, particularly whether the distribution of the features is homogenous or heterogenous throughout a faecal sample. This is of significance in terms of biomarker discovery and the subsequent translation from bench to bedside as a heterogenous distribution pattern poses challenges for sample collection and preparation.

The focus of this experiment was the spatial distribution of faecal metabolites and how this may affect the identification of biomarkers in faeces and the use of LA-REIMS as a high-

throughput diagnostic tool for CRC and IBD. To date, no consensus exists on the ideal collection method for faecal metabolomic analysis, with a number of studies only performing metabolomic analysis on parts of a faecal sample with limited sample homogenisation prior to analysis (33).The question driving this experiment was to what extent the distribution of metabolites of interest could impact biomarker discovery research in faecal metabolomics, specifically with the use of LA-REIMS as analytical tool.

Therefore, the spatial distribution of metabolites in faeces was investigated by developing a novel high throughput LA-REI-MSI pipeline for faeces. To date, MSI studies have concentrated on tissue and bacterial analysis rather than faeces (337).

A total of two healthy volunteers (n=2; 2F) were recruited and provided three faecal samples (sample 1 + 2 were provided by volunteer 1 on two consecutive days, whereas sample 3 was provided by volunteer 2) for the LA-REI-MSI analysis. Participant information is given in Table 12. Of the two stool samples, both were graded as Type 4 according to the BSC (329).

Table 12: Participant Information for the five healthy volunteers is given

Note: Age is given in years. Smoking status, diet, and medication are self-reported.

| Participant | Bristol Stool Scale | Bowel motions per day in the last 2 weeks | Age | Gender | Ethnicity | Alcohol (units/week) | Smoking Status (If smoker how many per day) | BMI | Diet | Medication | Antibiotics use | Probiotics/ prebiotics use | Vitamins, minerals, or dietary supplements | Known GI disorder | GI surgery in the past |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Volunteer 1 | 4 | 1/day | 22 | Female | British Asian | 5 | Non-smoker | 20.3 | Meat eater | None | Nil | Probiotics | None | None | None |
| Volunteer 2 | 4 | 1/day | 23 | Female | White British | 5 | Non-smoker | 21.2 | Meat eater | None | Nil | None | Multivitamins | None | None |

The faecal samples (n=3) were investigated using LA-REI-MSI. LA-REI-MSI was carried out <5hrs from bowel evacuation, cross-sections for analysis were taken from the faecal sample centre. In the low mass or fatty acid region (*m/z* 200-500), the most abundant peaks at *m/z* 255.24, 281.25, 279.25 demonstrated unique spatial distribution patterns in samples 1-3 and were putatively identified as palmitic acid, (18:1) FA, and (18:2) FA, respectively, see Figure 31. In the *m/z* 600-1000, the complex lipid region, the most abundant features at m/z 601.47, 619.48, 605.50, demonstrate distinct cluster patterns in the three samples, see Figure 32.

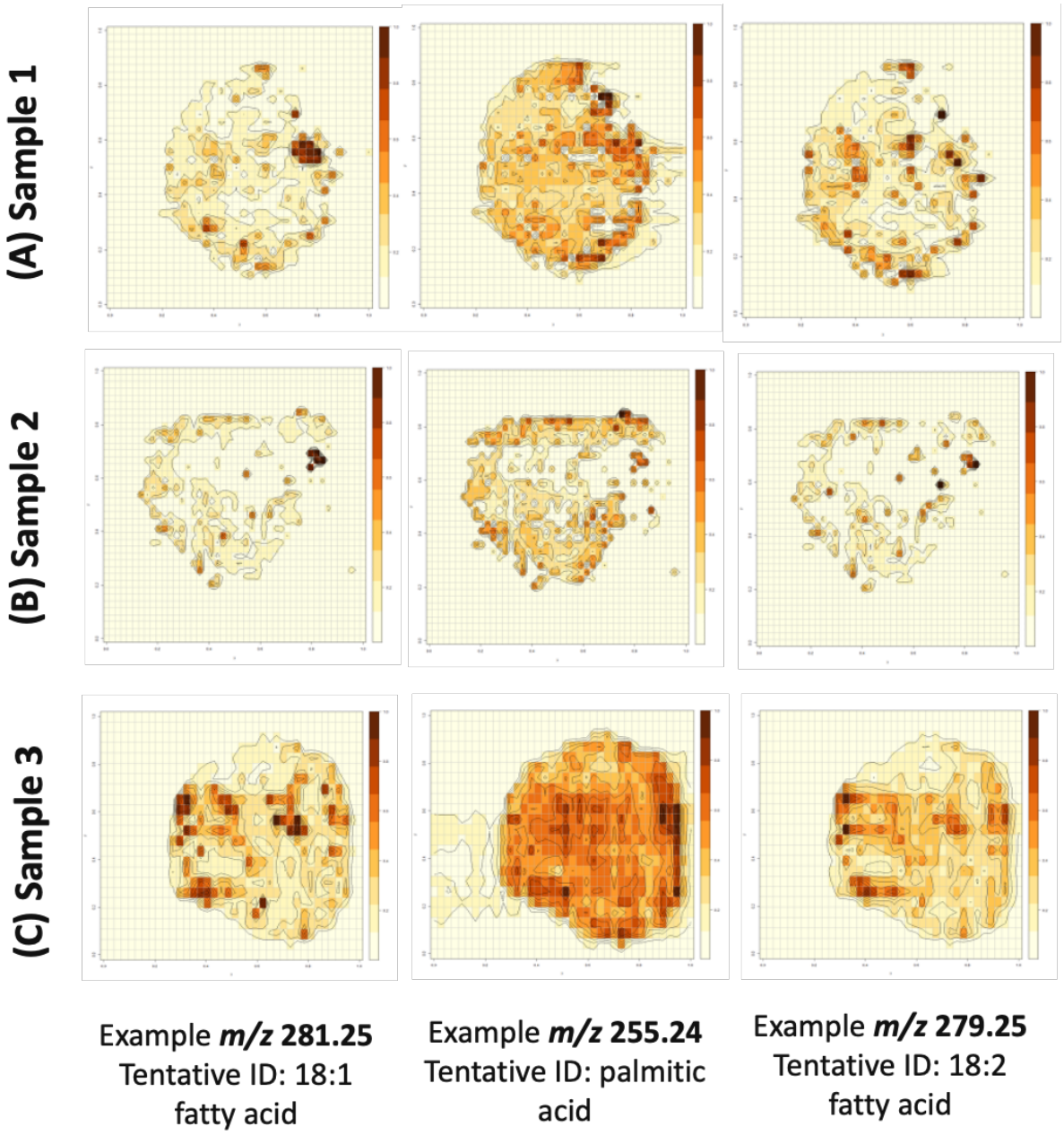



Figure 31: Heatmap depicts relative signal intensity across faecal sample cross-section of (A) Sample 1 (B) Sample 2 (C) Sample 3 at most abundant *m/z* 255.24, 281.25, 279.25, respectively. The colour code represents the relative signal intensity of *m/z* of interest with a resolution of 1mm (red=1, white=0).

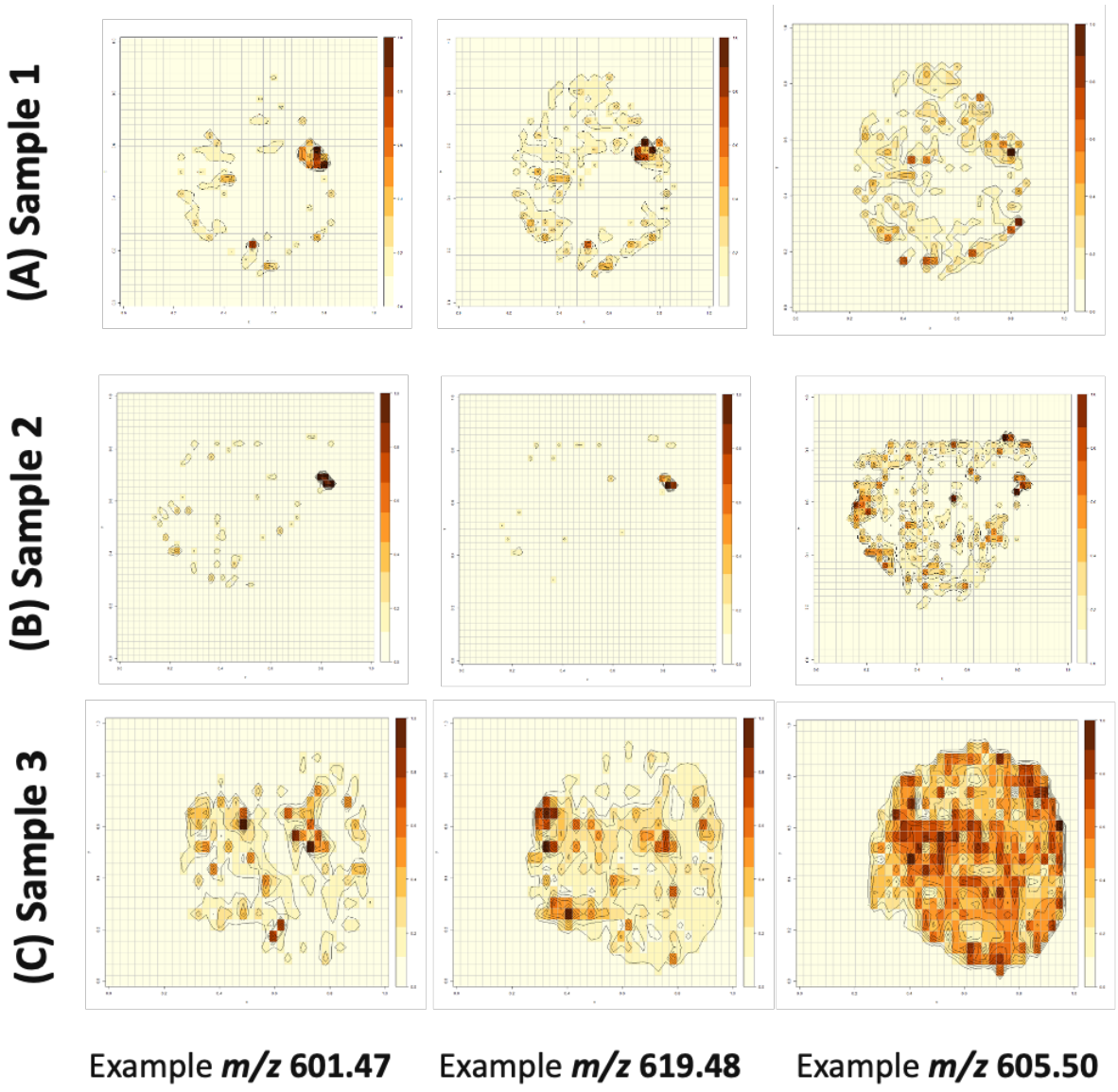



Figure 32: Heatmap depicts relative signal intensity across faecal sample cross-section of (A) Sample 1 (B) Sample 2 (C) Sample 3 at most abundant *m/z* 601.47, 619.48, 605.50, respectively. The colour code represents the relative signal intensity of m/z of interest with a resolution of 1mm (red=1, white=0).

The new pipeline of LA-REI-MSI allows for direct-from-sample analysis and visualisation of the relative abundance and spatial distribution of features of interest. Important limitations of this study include the small study population Future work will investigate the spatial distribution of features of interest, identified in the CRC and IBD studies, in faeces. LA-REI-

MSI is the first imaging technique to be used for the analysis of faeces and shows applicability for biomarker discovery research

## 4.2. Conclusion and Future work

The optimisation of laser and instrumentation parameters for the high-throughput LA-REIMS demonstrate improved signal-to-noise ratios with parameters yielding the least % carry-over between samples for high-throughput analysis. The results of the optimisation work were utilised for the analysis of all subsequent datasets described in this thesis. Overall, the results of the laser and instrumentation optimisation experiments were in line with previous work done by the group on faeces (in the case of heating power optimisation of faecal sample analysis) and different biofluids, tissue, and cells tested with the high-throughput LA-REIMS setup.

Future work will include the evaluation/optimisation of faecal samples categorised as BSC 1-2 (separate hard lumps) or 6-7 (watery, no solid material). Furthermore, the number of study participants will be increased, with a homogenate of a minimum of 5 samples for each BSC (1-7) tested for optimal optimisation parameters. Moreover, an independent validation of the optimised parameters identified in this analysis will be carried out, comparing the non-optimised vs optimised settings for faecal metabolomics analysis. This independent validation should be carried out using faecal samples from patients with the GI disease under investigation, e.g., CRC or IBD.

Pilot work on testing the potential impact of water content on faecal sample analysis LA-REIMS spectra demonstrates separation in multivariate analysis in the m/z range 50-1200 between crude faecal samples and faecal sample with different ratios water added. The impact appears higher in the *m/z* range 50-600 than in the *m/z* range 600-1000. However, the study was limited to n=1 sample and future work will be utilised of a homogenised sample (n>1) to assess the effect of water content after lyophilisation. An indirect approach towards correction for variation in water content for faecal samples is being advocated for LA-REIMS analysis. Furthermore, for the first time, LA-REI-MSI for mapping of metabolites in whole fresh or frozen human faecal samples was implemented. Future studies will utilise a targeted

metabolomics approach to investigate whole faecal samples from CRC and IBD patients utilising features of interest in each corresponding dataset.

## 5. Optimised LA-REIMS faecal metabolomics analysis for the detection of colorectal adenomas and cancer

CRC cells present unique metabolic fingerprints with dysregulated cellular metabolism (210). Metabolic changes from both cancer cells and the microbiome may present a specific phenotype that may be captured in the CRC faecal metabolome (33,338,339). Research in ambient MS as a technique for the investigation of the faecal metabolome in CRC has been either limited in the study participant numbers (340) or has lacked the optimisation of laser and instrumentation parameters for the analysis of the complex biospecimen (150).

The early detection of CRC with high diagnostic accuracy tests is key for effective treatment measures and ultimately patient outcome. The BCSP CRC screening has demonstrated efficacy, nonetheless, a low-to-moderate sensitivity for adenoma and early CRC detection exist (341). Here, LA-REIMS based faecal metabolomic analysis of adenomas and CRC was performed. This study presents the first application of optimised LA-REIMS to the investigation of the faecal metabolome in adenoma and CRC pathogenesis.

### 5.1. Results and Discussion

In the CRC study, a total of 244 frozen faecal samples (non-cancer control n=135, adenoma low grade dysplasia (LGD) n=79, adenoma high grade dysplasia (HGD) n=3, cancer n=27) were collected and analysed from 244 patients. Baseline demographic and clinical characteristics of CRC study patients are outlined in Table 13. Significant differences ($p$ <0.05) based on Fisher's exact test (categorical variable) and unpaired two-tailed T-test (continuous variable) were found in terms of mean age between non-cancer control vs adenoma ($p$ =0.0002), and non-cancer control vs cancer ($p$ =0.0001), see Table 13. Furthermore, significant differences ($p$ <0.05) were also found in terms of BMI between non-cancer control vs cancer (p=0.0314) and adenoma vs cancer ($p$ =0.0151), see Table 13. Seven patients were subsequently excluded: in two cases the peak peaking algorithm used for processing of the data indicated low signal-to-noise ratios with high levels of analytical noise; faecal sample volumes were limited in five cases. As a result, samples from 237 patients (non-cancer control n=131,

adenoma LGD n=77, adenoma HGD n= 3, CRC n=26) were used for further analysis. Subsequently the adenoma HGD group (n=3) was excluded from the analysis as it was limited to only three participants and did not allow for comparative analysis between the study groups.

Table 13: Baseline demographic and clinical characteristics of CRC study patients

| Characteristics | Non-cancer control | Adenoma patient (LGD) | Cancer and HGD patient | p value |
|---|---|---|---|---|
| Sample, n | 135 | 79 | 30 | |
| Age, mean (range) | 62.5 (21-88) | 68.7 (47-93) | 71.4 (38-88) | 0.0002*, 0.0001+, 0.0607• |
| Sex, n (%) | | | | 0.1564*, 1+, 0.2909• |
| Male | 64 (47.4) | 46 (58.2) | 14 (46.6) | |
| Female | 71 (52.6) | 33(41.8) | 16 (53.3) | |
| Ethnicity, n (%) | | | | 0.3072*, 0.3669+, 0.0609• |
| White | | | | |
| British | 122 (90.4) | 68 (86.1) | 28 (93.3) | |
| Irish | 2 (1.5) | 1 (1.27) | 1 (3.3) | |
| Other White | 2 (1.5) | 1 (1.27) | 1 (3.3) | |
| Not-white | | | | |
| Asian or Asian British | | | | |
| Indian | 6 (4.4) | 6 (7.6) | 0 (0) | |
| Pakistani | 1 (1.5) | 0 (0) | 0 (0) | |
| Black or Black British | | | | |
| Black African | 0 (0) | 0 (0) | 0 (0) | |
| Black Caribbean | 1 (1.5) | 1 (1.3) | 0 (0) | |
| Other Black | 1 (1.5) | 1 (1.3) | 0 (0) | |
| Other | 0 (0) | 1 (1.3) | 0 (0) | |
| Body Mass Index (BMI), mean (range) | 27. 8 (18.0-44.1) | 27.5 (17.9-37.5) | 30.0 (21.4-36.1) | 0.6747*, 0.0314+, 0.0151• |

* = normal vs adenoma

\+ = normal vs cancer

• = adenoma vs cancer

### 5.1.1. Negative vs positive ion detection and faecal sample vs faecal water extract analysis

The analysis of the CRC dataset was carried out for both faecal sample as well as faecal water (2:1 ratio water to faeces) extracts in positive and negative ion detection mode. Example raw spectra in negative and positive mode spectra for both faecal sample and faecal water extracts for a randomly chosen study participant in the non-cancer control group are shown in Figure 33.

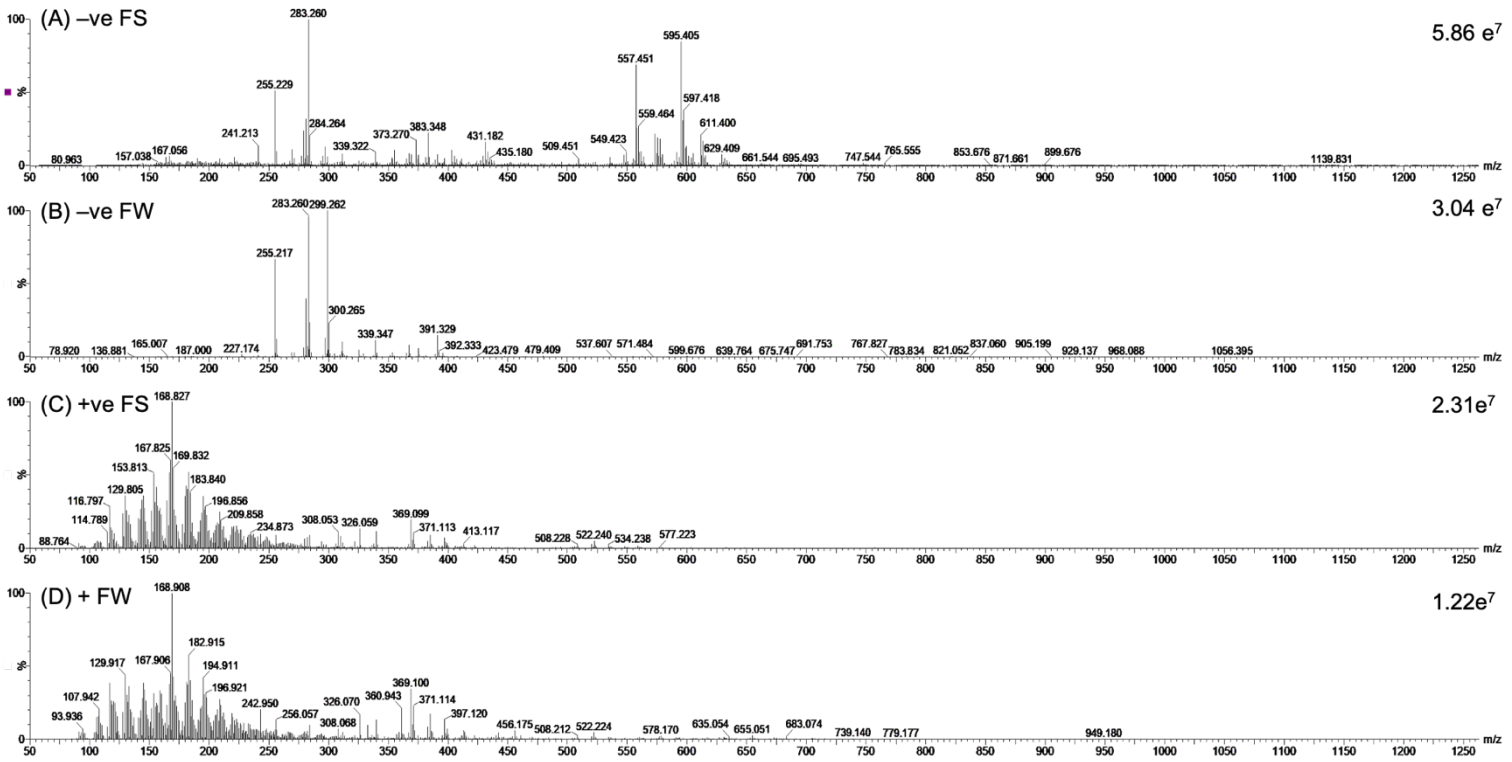



Figure 33: Example mass spectra of optimised faecal LA-REIMS analysis (A) negative ion detection mode faecal sample (-ve FS) (B) negative ion detection mode faecal water (-ve FW) (C) positive ion detection mode faecal sample (+ve FS) (D) positive ion detection mode faecal sample (+ve FW). Background subtraction and lock mass correction were carried out for spectra in the *m/z* range 50 to 1200.

From visual inspection of the mass spectra, negative ion detection mode contains higher levels of lipid species, characterised by peaks in the *m/z* range 550-1000. The optimisation in favour of detecting complex lipid species in faeces using LA-REIMS in CRC is key due to lipid metabolism playing a key role in cancer pathogenesis. Lipid dysregulation is a well-established

hallmark of carcinogenesis (167). Aberrations in lipid metabolic pathways in CRC tissue are commonly observed (342,343), and it has been hypothesised that these changes could be detected by analysing cells shed from colorectal adenoma or cancer into the GI lumen and subsequently faeces using ambient MS. Therefore, optimisation of data acquisition to favourably detect complex lipid species is important to increase sensitivity for detecting adenoma and cancer.

Furthermore, in depth visual investigation of peaks in positive ion detection mode vs negative ion detection mode show higher levels of background noise levels in positive mode (both faecal water and sample analysis), Figure 33. Furthermore, research published by our group indicates that negative ion detection mode provides richer mass spectra in terms of signal intensity and diversity of detectable ion species with LA-REIMS (149,344). Overall, after visual investigation of raw spectral data, negative ion detection mode was chosen for further statistical analysis using multivariate and univariate analysis; an in-depth analysis of the data in negative mode follows.

### 5.1.2. Faecal sample negative ion detection mode

#### *5.1.2.1. Unsupervised and supervised multivariate statistical analysis*

Faecal samples from 234 patients (non-cancer control n=131, adenoma LGD n=77, CRC n=26) were analysed using the optimised LA-REIMS setup in negative ion detection mode. The PCA model of the faecal LA-REIMS analysis of non-CRC control, adenoma LGD, and CRC patients is depicted in Figure 34. Details of the PCA analysis and why it was chosen as the preferred data analysis tool is summarised in section *3.3.3.4. Data analysis – Metaboanalyst.*

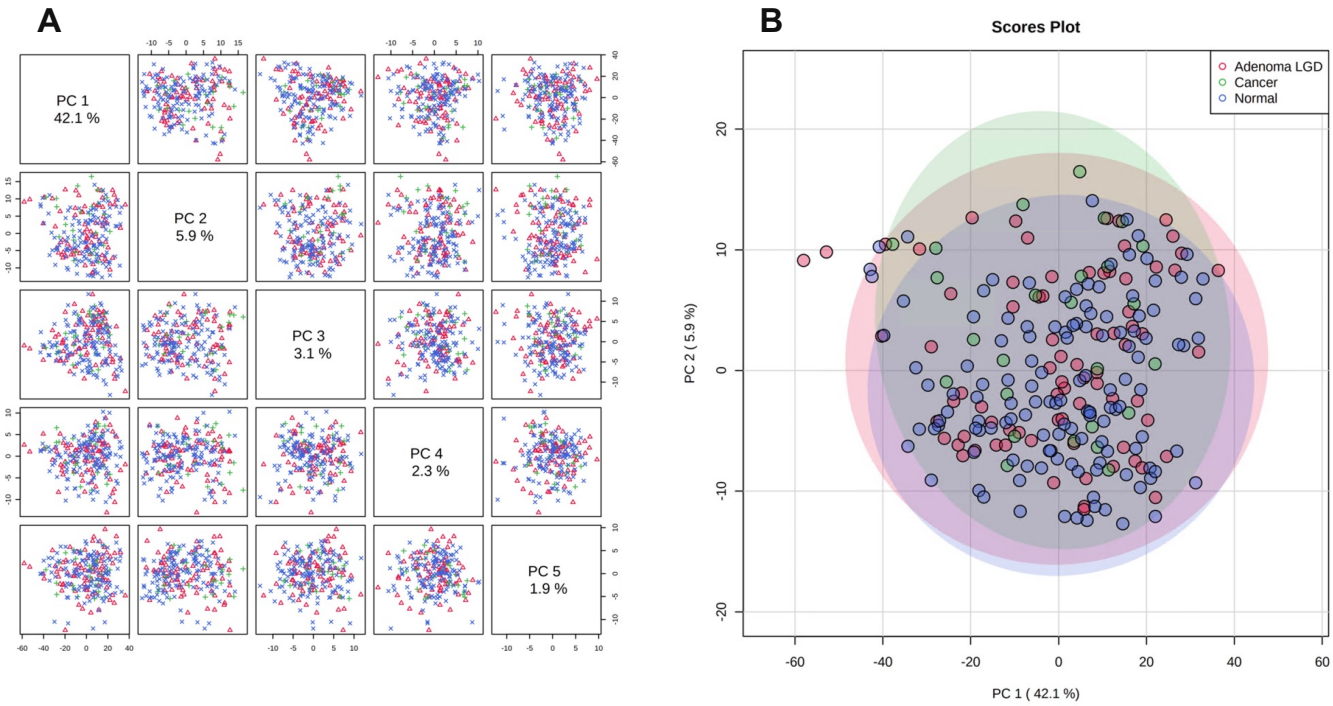



Figure 34: (A) Pairwise PCA scores plot for the top five PCs. (B) Scores plot of PCA analysis PC 1 (42.1%) and PC2 (5.9%) over the 50−1200 m/z range (negative ion detection mode) for faecal samples from non-cancer control, adenoma LGD, and CRC patients. Shaded region represents 95% confidence intervals.

The unsupervised multivariate analysis utilising PCA over the 50−1200 m/z range for faecal samples from non-cancer control, adenoma LGD, and cancer patient (negative ionisation mode) yielded no separation based on underlying disease pathology in both the pairwise PCA scores plot for the top five principal components (PC)s, Figure 34A, as well as the scores plot of PCA analysis PC 1 (42.1%) and PC2 (5.9%), Figure 34B.

Faeces is a highly complex and heterogenous matrix that contains in addition to the metabolome, several other end products from for instance food digestion. The heterogenous nature of the sample matrix as well as the direct from sample analysis (without extraction or preparation steps) offers a potential explanation for the lack of separation observed based on disease group when comparing the first five PC components of the PCA analysis. It should be noted that a further analysis of more PC components may yield a separation between the

groups of interest. However, as the main aim of this work was to establish group differences based on disease class, a supervised algorithm (PLS-DA) was utilised for further investigations.

To maximize the group separation and identify discriminating metabolites, supervised PLS-DA was used to build models of faecal LA-REIMS analysis of non-cancer control, adenoma LGD, and CRC patient patients in the *m/z* range 50-1200 (negative ion detection mode), Figure 34. Furthermore, the model was cross-validated using LOOCV with Q2 as performance measure (maximum components 5) and underwent permutation testing to check for potential overfitting of the model. Details of the PLS-DA analysis and why it was chosen as the preferred data analysis tool is summarised in section *3.3.3.4. Data analysis – Metaboanalyst*.

The supervised PLS-DA scores and loadings plot of non-cancer control, adenoma LGD, and CRC study groups are depicted in Figure 35 A+B. The model demonstrates separation between non-cancer control vs adenoma, whereas cancer tended to cluster with the adenoma group. LOOCV was utilised from which $Q^2$ and $R^2$ values were calculated, representing predictive capability, and explained variance, respectively. The model yielded $R^2$=0.885 and $Q^2$=0.698 when utilising a maximum of five components.

The model was validated using permutation analysis. In each permutation, a PLS-DA model is built between data (X) and permuted class labels (Y) (345). The graph in Figure 35C shows the distribution derived from the permuted samples. The highlighted bar represents the original sample. The further to the right of the distribution, the more significant is the separation between the two groups (345). Using 2000 permutation tests, the p value is reported as $p$<0.001, validating the predictive ability of the real model.

The supervised and unsupervised multivariate models indicate that faeces are highly heterogeneous. It is widely understood that the faecal metabolome shows large inter-individual and even intra-individual heterogeneity. Some studies investigating the faecal metabolome in CRC using unsupervised multivariate modelling have demonstrated a lack of separation or clustering based on disease classes alone (47). This is likely due to confounding factors such as diet, medications, gut microbiota variations, etc. However, supervised multivariate and univariate analysis, as a complementary method to unsupervised

techniques, have shown to be applicable in the analysis of datasets with increased heterogeneity (346) and often demonstrate separation based on disease type alone.

More in-depth confounding analysis to investigate potential modifiers of the faecal metabolome that could be responsible for the trends observed in both PCA and PLS-DA model were carried out investigating patient-related modifiers of the faecal metabolome, see section *5.1.3. Patient-related modifiers of the faecal metabolome assessment – analysis of confounding factors*. Furthermore, univariate statistics were utilised to predict specific *m/z* features significantly different between non-CRC control, CRC, and adenoma participants and tentative identification of the metabolites was carried out, see section *5.1.2.3. Univariate analysis.*

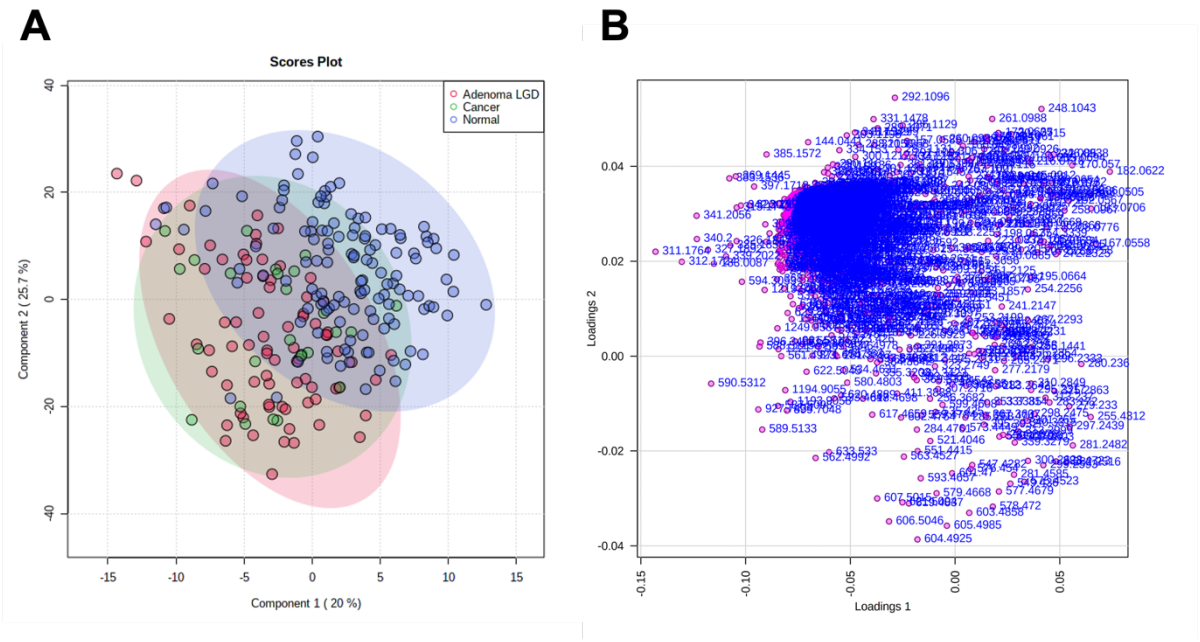


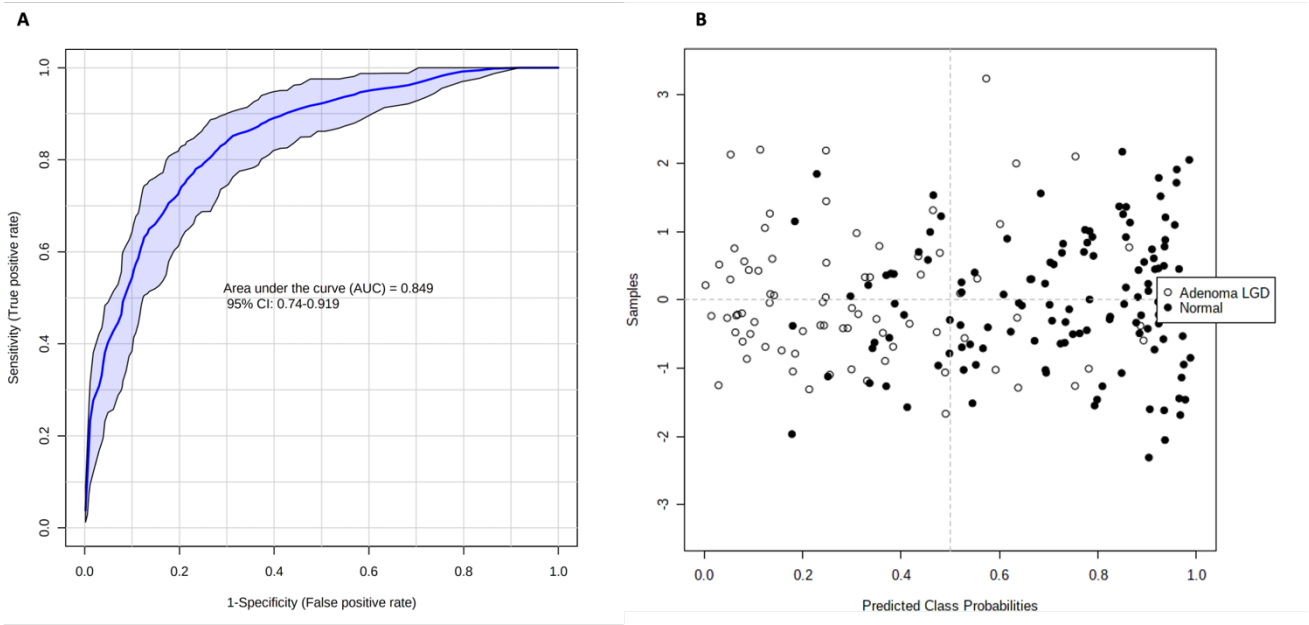
Figure 35: Supervised PLS-DA model for faecal samples from non-cancer control, adenoma LGD, and CRC patients in the *m/z* range 50-1200 (negative ion detection mode) (A) Scores plot (B) Loadings plot (C) Permutation test; select test statistic: Separation distance (B/W) with set permutation numbers: 2000 $p<0.01$. Red vertical error indicates observed data test statistic; y-axis indicated frequency of permuatations; x-axis indicates test statistic values.

#### *5.1.2.2. Diagnostic accuracy of LA-REIMS*

RFC algorithms were utilised for sample classification purposes and ROC curves were generated by MCCV to assess the diagnostic performance of faecal LA-REIMS for adenoma LGD vs non-cancer control, cancer vs non-cancer control, non-cancer control vs disease (adenoma LGD and CRC combined). Details of the RFC and ROC curve analysis and why they were chosen as the preferred data analysis tool are summarised in section *3.5.8. CRC.*

##### 5.1.2.2.1. LA-REIMS - Adenoma vs non-cancer control

RFC performance of the high-throughput faecal LA-REIMS setup for classification adenoma vs non-cancer control are shown in Figure 36, the ROC curve indicates an AUC of 0.849 (95% confidence interval CI: 0.74 to 0.919), Figure 36A. The predicted class probability plot in Figure 36B demonstrates correct classification of adenoma LGD and non-cancer control samples. LA-REIMS demonstrates a sensitivity of 80.5% and specificity of 80.2% in adenoma LGD vs non-cancer control patients (Figure 36C).




C

| | Actual class: Adenoma LGD | Actual class: Normal |
|---|---|---|
| Predicted class: Adenoma LGD | 62 | 26 |
| Predicted class: Normal | 15 | 105 |

Figure 36: RFC of the LA-REIMS analysis of adenoma LGD vs non-cancer control. (A) RFC method: ROC curve with AUC and 95% confidence interval (B) Predicted class probabilities plot utilising the best AUC classifier. Adenoma LGD represented by black dots, normal (non-cancer control) as white dots. (C) Confusion matrix of adenoma LGD vs non-cancer control

#### 5.1.2.2.2. LA-REIMS - Cancer vs non-cancer control

To investigate the non-cancer control vs cancer cohort using RFC, two cohorts were investigated, first the analysis of all non-cancer control (n=131) vs cancer (n=26), and second to perform a classification on an even group of participants, n=25 cancer patients and n=25 non-cancer control patients (randomly chosen) were studied. This two-factor approach was chosen to check any effects the large difference in group size may have on the RFC model. The RFC algorithm in unbalanced data can have a bias towards the dataset with the greater number of study participants, particularly if the difference is large (347). To investigate this

potential effect, the two analyses as described above were carried out and the model performances compared.

RFC classification performance evaluations are shown in Figure 37 (whole group analysis) and Figure 38 (selected even group analysis). For the whole group analysis, the ROC curve indicates an AUC of 0.942 (95% confidence interval CI: 0.787-0.993), Figure 37A. For the selected even group, the ROC curve indicates an AUC of 0.939 (95% confidence interval CI: 0.724 to 1), Figure 38A. The predicted class probability plot in Figure 37B demonstrates correct classification of cancer and non-cancer control samples for the whole cohort of cancer vs non-cancer control. In the whole group analysis, LA-REIMS demonstrates a sensitivity of 92.3% and specificity of 92.4% in cancer vs non-cancer control patients (Figure 37C). The predicted class probability plot in Figure 38 B demonstrates correct classification of cancer and non-cancer control samples for the even sample sub-set. Here, LA-REIMS demonstrates a sensitivity of 92.0% and specificity of 88.0% in cancer vs non-cancer control patients (Figure 38C).

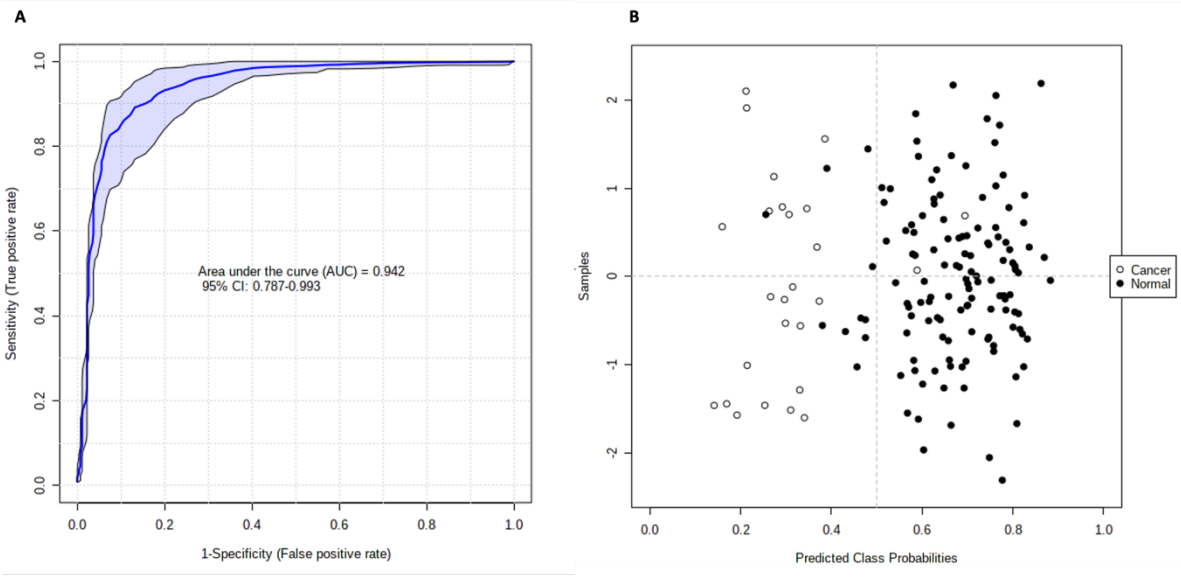



Figure 37: RFC of the LA-REIMS analysis of cancer (n=26) vs non-cancer control (n=131) for the whole study group population. (A) RFC method: ROC curve with AUC and 95% confidence

interval (B) Predicted class probabilities plot utilising the best AUC classifier. Cancer represented by black dots, Normal (non-cancer control) shown as white dots. (C) Confusion matrix of cancer (n=26) vs non-cancer control (n=131)

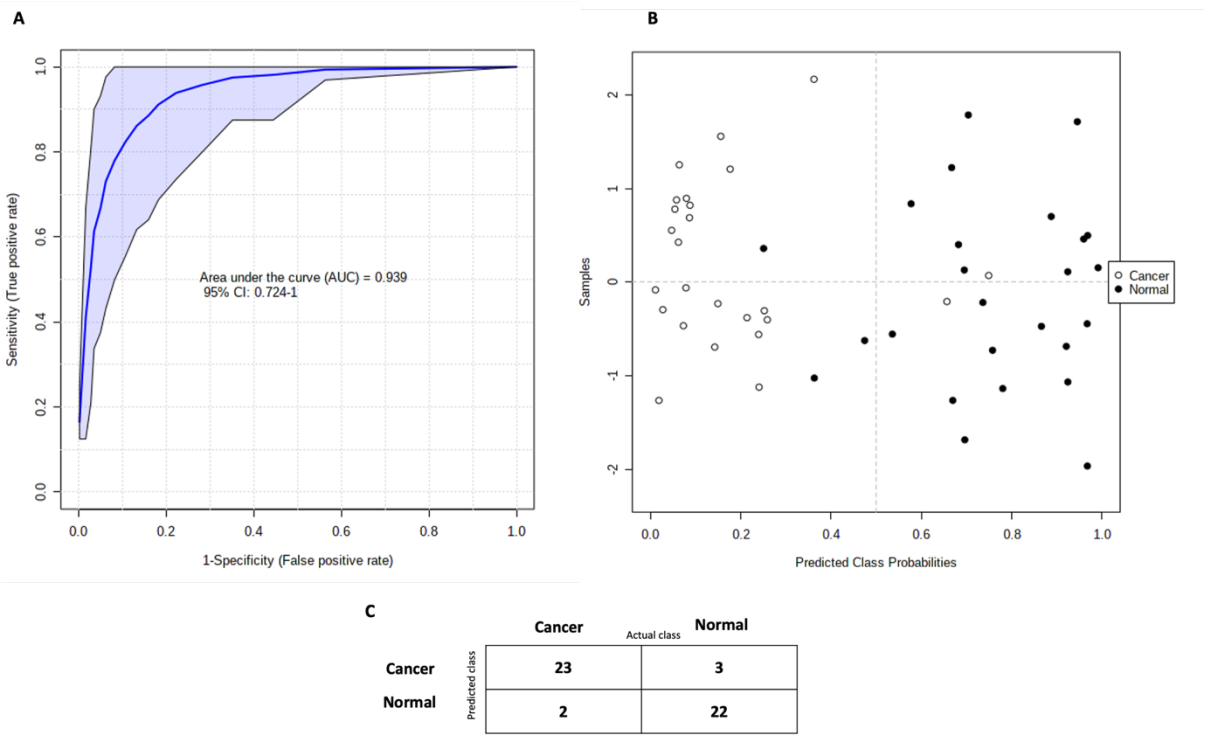



Figure 38: RFC of the LA-REIMS analysis of cancer (n=25) vs non-cancer control (n=25) for even number of study participants in each group (A) RFC method: ROC curve with AUC and 95% confidence interval (B) Predicted class probabilities plot utilising the best AUC classifier. Cancer represented by black dots, Normal (non-cancer control) shown as white dots. (C) Confusion matrix of cancer (n=25) vs non-cancer control (n=25)

#### 5.1.2.2.2.1. LA-REIMS - Non-cancer control vs disease (adenoma and cancer combined)

The RFC algorithm was utilised to assess the classification performance of the high-throughput faecal LA-REIMS setup for disease (adenoma and CRC) vs non-cancer control patients in the CRC cohort. RFC performance evaluation are shown in Figure 39, the ROC curve indicates an AUC of 0.81 (95% confidence interval CI: 0.72 to 0.88), Figure 39A. The predicted class probability plot in Figure 39B demonstrates correct classification of disease and non-cancer control samples. LA-REIMS demonstrates a sensitivity of 79.6% and specificity of 76.3% in disease vs non-cancer control patients, Figure 39C.

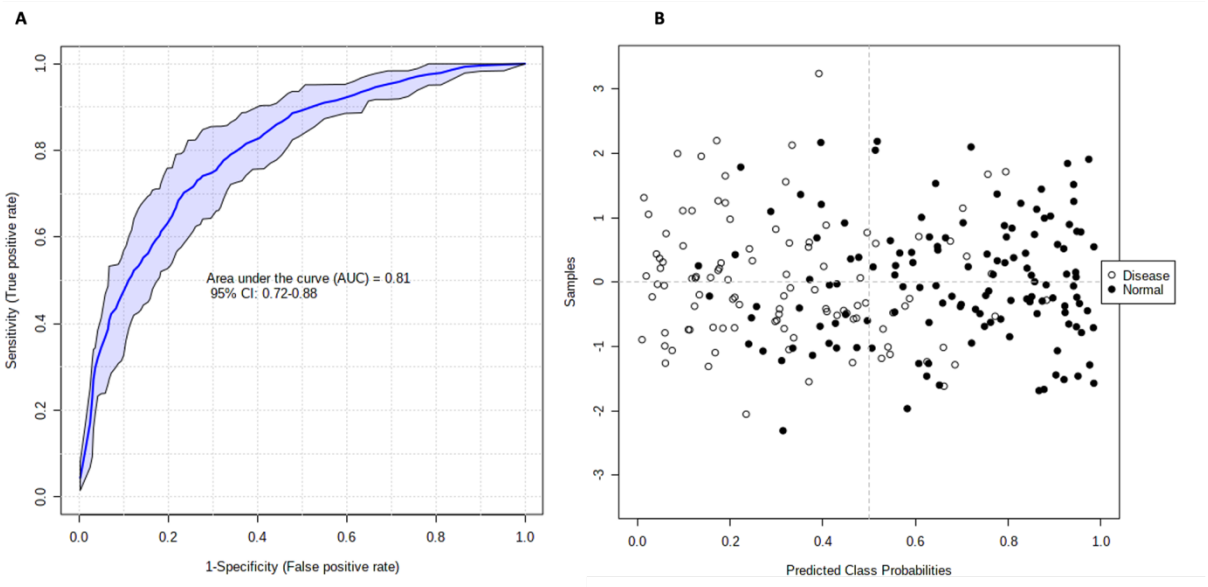
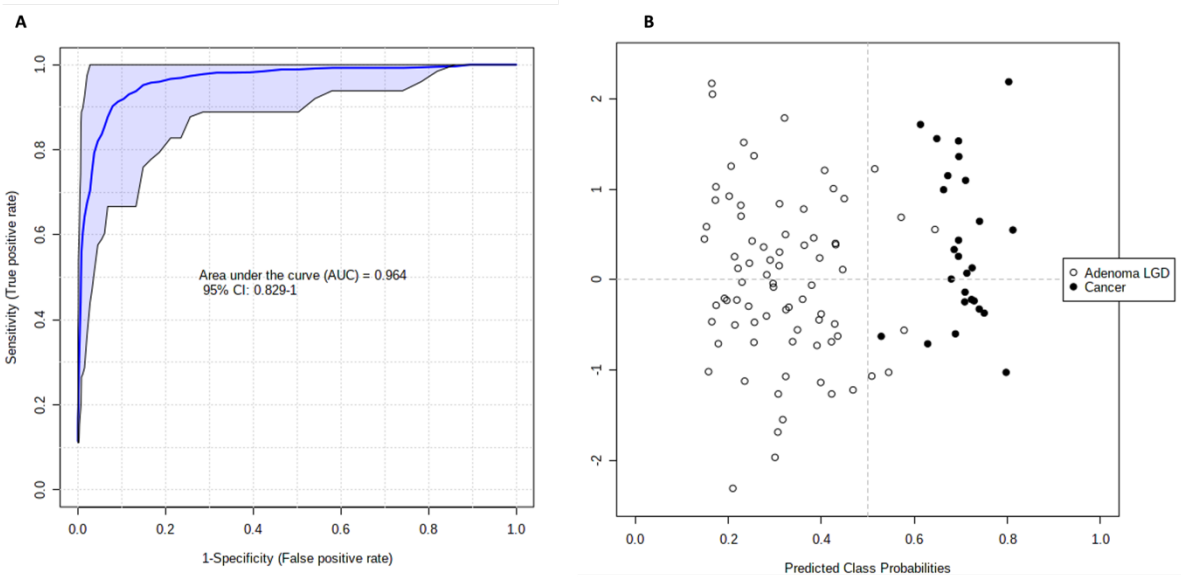



Figure 39: RFC of the LA-REIMS analysis of disease vs non-cancer control. (A) RFC method: ROC curve with AUC and 95% confidence interval (B) Predicted class probabilities plot utilising the best AUC classifier. Cancer represented by black dots, Normal (non-cancer control) shown as white dots. (C) Confusion matrix of disease vs non-cancer control

#### 5.1.2.2.2.2. LA-REIMS – Adenoma vs cancer

The RFC algorithm was utilised to assess the classification performance of the high-throughput faecal LA-REIMS setup for adenoma vs cancer patients in the CRC cohort. RFC performance evaluation are shown in Figure 40, the ROC curve indicates an AUC of 0.96 (95% confidence interval CI: 0.829 to 1), Figure 40A. The predicted class probability plot in Figure 39B demonstrates classification of disease and non-cancer control samples. LA-REIMS demonstrates a sensitivity of 93.4% and specificity of 96.2% in adenoma vs cancer patients, Figure 40C.




C

| Predicted class \ Actual class | Adenoma LGD | Cancer |
|---|---|---|
| Adenoma LGD | 71 | 1 |
| Cancer | 6 | 25 |

Figure 40: RFC of the LA-REIMS analysis of adenoma vs cancer. (A) RFC method: ROC curve with AUC and 95% confidence interval (B) Predicted class probabilities plot utilising the best

AUC classifier. Cancer represented by black dots, adenoma shown as white dots. (C) Confusion matrix of adenoma vs cancer

#### 5.1.2.2.2.3. Diagnostic accuracy of FIT and a comparison of LA-REIMS vs FIT

FIT performance was evaluated under PI Prof. Ramesh Arasaradnam and his team at the UHCW as described in detail in section *3.3.2. CRC study*. FIT testing has a high specificity for detecting advanced CRC, however for the detection of adenoma, i.e., precancerous lesions, FIT has a poor sensitivity; this results in precancerous lesions often remaining undetected (348). This is reflected in the results presented here, FIT demonstrates a high specificity of 87.2% in all models tested, whereas it shows variable sensitivities in cancer versus non-cancer control (81.8%), disease versus non-cancer control (44.8%), and adenomatous lesions vs non-cancer control (33.8%). In the CRC cohort FIT demonstrates a low sensitivity for adenoma LGD and disease detection; this highlights the need for improved screening modalities for early CRC detection.

Moreover, improved diagnostic accuracy of screening techniques is particularly important given the COVID-19 pandemic, which has resulted in a decrease in endoscopy numbers, producing a backlog of procedures, which may take up to one year to eliminate (349). Overall, low cost, high-throughput screening tests with high diagnostic accuracy for early cancer detection are needed. The diagnostic accuracy of LA-REIMS for adenomas and CRC was compared to the current screening technique used in the Bowel Cancer Screening Program in the UK, FIT.

For ease of comparison, a summary of LA-REIMS RFC model classification performance with sensitivities, specificities, and ROC curve with AUC and 95% confidence interval (CI) are provided in Table 14. FIT model sensitivities and specificities investigated are summarised in Table 15.

In the adenoma vs control model, LA-REIMS demonstrated a sensitivity of 80.5% and a specificity 80.2%. FIT yielded a sensitivity of 33.8% and specificity of 87.2% for adenomatous lesions vs non-cancer control, (Figure 41B). Additionally, LA-REIMS achieved a sensitivity of 92.0% and specificities of 88.0% in the cancer versus non-cancer control model, whereas FIT demonstrated a sensitivity of 81.8% and specificity of 87.2% in the cancer versus non-cancer control model (Figure 41A). FIT showed a sensitivity of 44.8% and specificity of 87.2% for disease versus non-cancer control patients, (Figure 41C), where REIMS demonstrated a sensitivity of 79.6% and specificity of 76.3%. Furthermore, LA-REIMS showed a sensitivity of 93.4% and specificity of 96.2% in adenoma vs cancer patients.

Overall LA-REIMS shows higher sensitivities for the detection of adenomas, cancer, and disease (adenoma and cancer combined) but a reduced specificity for CRC detection. FIT shows diagnostic accuracy for CRC, however not for adenomas. The results of the LA-REIMS modelling are promising but need to be validated in a prospectively recruited independent data set. The results of the CRC cohort indicate the utility of a potential combination of LA-REIMS and FIT.

**(A)**

| Predicted class \ Actual class | Cancer | Non-cancer control |
|---|---|---|
| Cancer | 18 | 17 |
| Non-cancer control | 4 | 116 |

**(B)**

| Predicted class \ Actual class | Adenoma | Non-cancer control |
|---|---|---|
| Adenoma | 25 | 17 |
| Non-cancer control | 49 | 116 |

**(C)**

| Predicted class \ Actual class | Disease | Non-cancer control |
|---|---|---|
| Disease | 43 | 17 |
| Non-cancer control | 53 | 116 |

Figure 41: FIT diagnostic performance confusion matrices with (A) cancer vs non-cancer control (B) adenoma LGD vs non-cancer control (C) disease vs non-cancer control

Note: FIT not available for (n=4 CRC and n=7 adenoma LGD)

Table 14: LA-REIMS RFC model classification performance for CRC cohort

| LA-REIMS | Sensitivity | Specificity | AUC + (95% CI) |
|---|---|---|---|
| Adenoma vs control | 80.5 | 80.2 | 0.849 (0.74-0.92) |
| Cancer vs control | 92 | 88 | 0.939 (0.72-1) |
| Disease vs control | 79.6 | 76.3 | 0.81 (0.72-0.88) |
| Adenoma vs cancer | 93.4 | 96.2 | 0.96 (0.83-1) |

Table 15: Summary of FIT diagnostic performance for CRC cohort

| FIT | Sensitivity | Specificity |
|---|---|---|
| **Adenoma vs control** | 33.8 | 87.2 |
| **Cancer vs control** | 81.8 | 87.2 |
| **Disease vs control** | 44.8 | 87.2 |

#### *5.1.2.3. Univariate analysis*

Faeces holds metabolites, lipids, and other molecules which are the downstream products of the processing occurring in the GI tract (33). Faeces is constant contact with intestinal wall and epithelial cells are shed into faeces (350–353). Adenomas and cancers tend to shed cells at an increased rate when compared to surrounding tissue due to the increased proliferation and are especially abundant in faeces. Furthermore, cancer pathogenesis is marked by changes to the glycolytic pathway and an upregulation of *de novo* fatty acid synthesis (342,354), the result is a dysregulation in metabolite and lipid abundances of both tissue and faeces.

Here, univariate statistics were utilised to predict specific metabolites significantly different between the control versus disease groups. Details of the Volcano and ANOVA analysis and why they were chosen as the preferred data analysis tool are summarised in section 3.3.3.4. Further processing – Metaboanalyst. Volcano plots were built to identify univariate features significantly different between adenoma LGD vs non-cancer control, Figure 42, and cancer vs non-cancer control, Figure 43. The volcano plot of adenoma LGD (red dots) vs non-cancer control (blue dots), Figure 42, identified 18 *m/z* features were more abundant in the adenoma LGD group, whereas 12 *m/z* features were associated with non-cancer control, 1448 *m/z* features were nonsignificant.

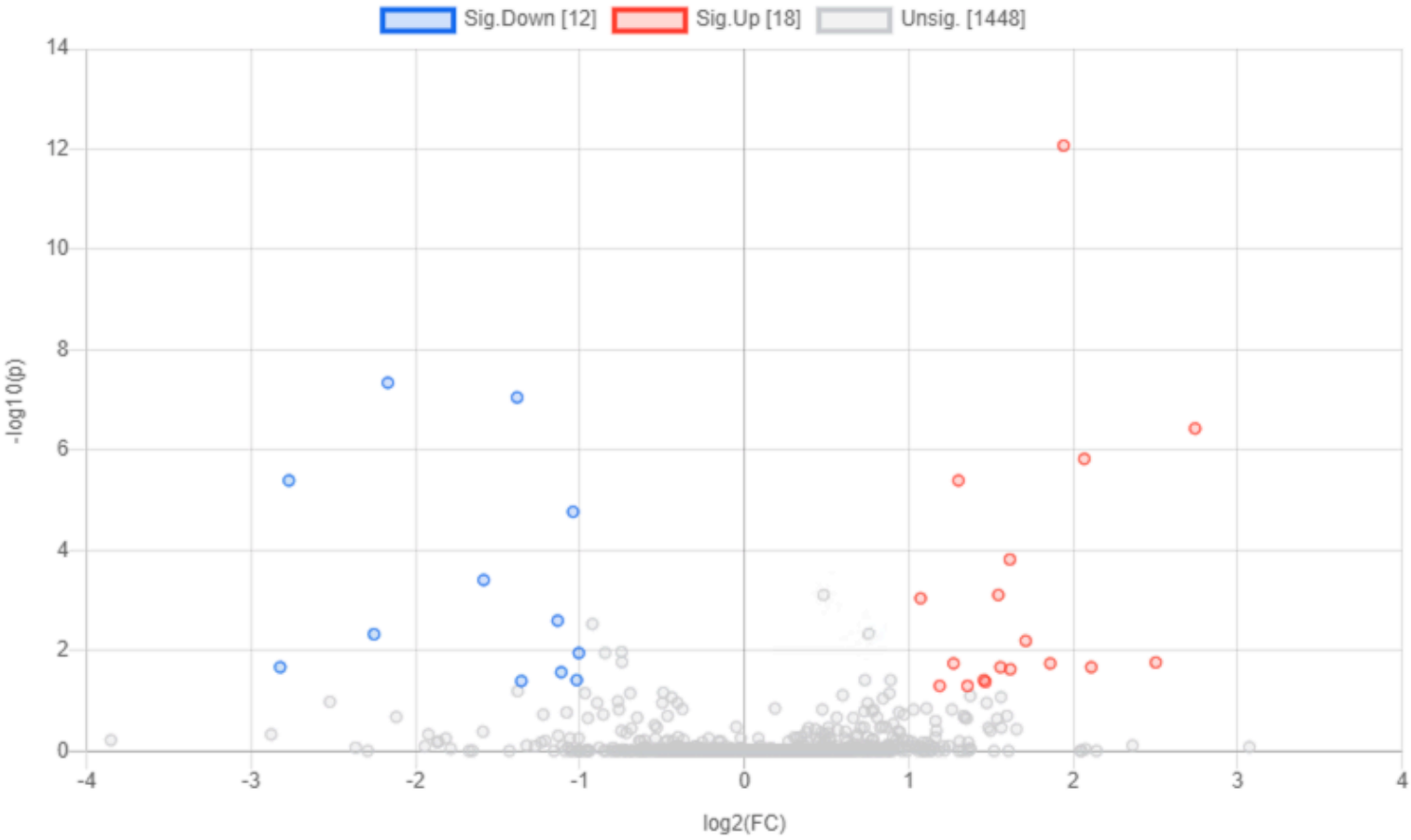



Figure 42: Volcano plot of adenoma LGD (red dots) vs non-cancer control (blue dots). Fold-change threshold set to 2; FDR corrected *p* value threshold set to <0.05; group variance set to unequal.

The volcano plot of cancer (red dots) vs non-cancer control (blue dots), Figure 43, identified 28 *m/z* features were more abundant in the cancer group, 17 *m/z* features were more abundant in the non-cancer control, 1433 m/z features were nonsignificant. Furthermore, ANOVA analysis, Figure 44, non-parametric version (Kruskal Wallis Test) of cancer vs adenoma LGD vs non-cancer control, along with Fisher's Least Significant Difference post-hoc test was carried out. FDR corrected *p* value of < 0.05 were accepted as significant.

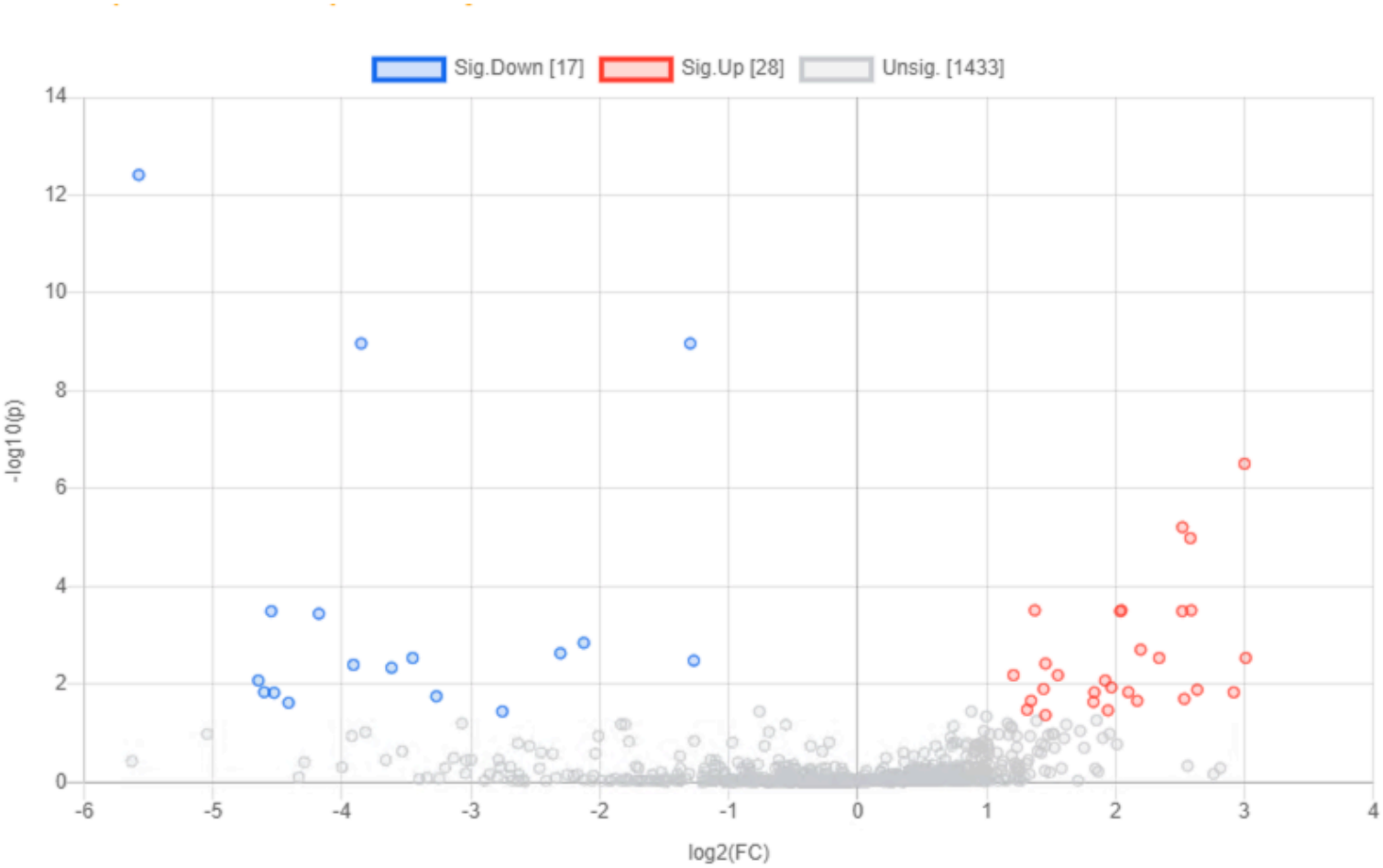



Figure 43: Volcano plot of cancer (red dots) vs non-cancer control (blue dots). Fold-change threshold set to 2; FDR corrected *p* value threshold set to <0.05; group variance set to unequal.

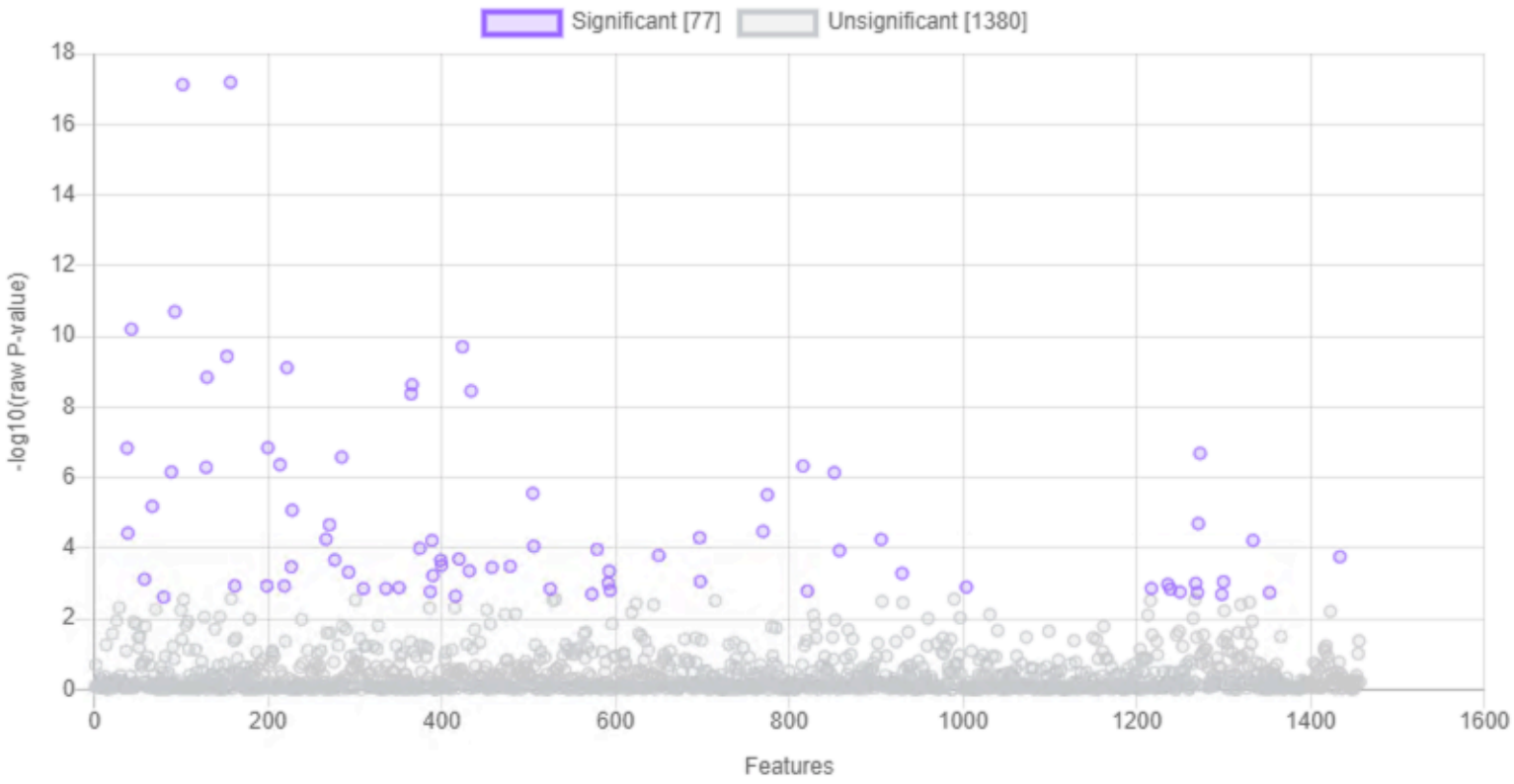



Figure 44: ANOVA non-parametric version (Kruskal Wallis Test) of cancer vs adenoma vs non-cancer control, along with Fisher's Least Significant Difference post-hoc test - FDR corrected *p* value of < 0.05 were accepted as significant.

The univariate methods (ANOVA, Volcano plots, etc.) were utilised to generate a list of potential biological importance. Annotation of the m/z of interest was carried out based on their accurate mass and the isotopic distribution, and tentative identification of metabolites was performed, see Table 17. A step-by-step description of the metabolite identification process utilised is given in section *3.3.3.5. Metabolite Identification* and *3.3.3.6. Tentative identification – validation.*

Table 16: Tentative identifications of m/z significantly different between either cancer, adenoma, non-cancer control based on ANOVA, along with Fisher's Least Significant Difference post-hoc test and volcano plots displaying fold change against FDR corrected P value from t tests (FDR corrected *p* value of < 0.05 were accepted as significant). Accurate masses, up to three decimal places after mass drift correction, were utilised for tentative identifications.

| Accurate m/z | Molecular formula of tentatively identified compound | Tentative ID | Adduct | Matched mass | ppm error |
|---|---|---|---|---|---|
| 167.060 | No Match | | | | |
| 181.074 | $C_6H_{12}O_6$ | Hexose | M-H | 181.0718 | 10 |
| 191.058 | $C_7H_{12}O_6$ | Quinic acid | M-H | 191.0556 | 12.56 |
| 193.070 | $C_7H_{14}O_6$ | 1-O-Methyl-myo-inositol or D-Glucoside | M-H | 193.0718 | 9 |
| 197.070 | No Match | | | | |
| 199.170 | $C_{12}H_{24}O_2$ | Dodecanoic acid | M-H | 199.1704 | 2 |
| 209.065 | $C_7H_{14}O_7$ | Sedoheptulose | M-H | 209.0667 | 8 |
| 217.067 | No Match | | | | |
| 381.338 | $C_{24}H_{46}O_3$ | FA(24:1(OH)) | M-H | 381.3374 | 1.57 |
| 601.461 | $C_{35}H_{66}O_5$ | Diacylglycerol (32:1) - chloride (CL) 35 | M+CL | 601.4604 | 1 |
| 603.477 | $C_{35}H_{66}O_5$ and $C_{35}H_{68}O_5$ | CL 37 Diacylglycerol (32:1) and CL 35 of Diacylglycerol (32:0) | M+CL | 603.4604 and 603.4761 | 1 and 10 |
| 605.493 | No Match | | | | |
| 619.472 | $C_{41}H_{66}O_5$ | Diacylglycerol (38:7) | M-H2O-H | 619.4726 | 1 |

Tentative identification of 13 *m/z* of interest are summarised in Table 16. Features at *m/z* 167.060, 197.070, 217.067, and 605.493 yielded no match on interrogation in literature databases. One feature, at *m/z* 193.070, yielded multiple hits on interrogation in literature databases. Features at m/z 181.072, 191.058, 199.170, 209.065, 381.338, 601.461, 603.477, and 619.472 were subsequently tentatively identified as hexose, quinic acid, dodecanoic acid, sedoheptulose, FA(24:1(OH)), DG (32:1) chloride (CL) isotope 35, and a combination of DG (32:1) CL isotope 37 and DG (32:0) CL isotope 35, and DG (38:7), respectively.

For more robust identification of *m/z* of interest identified using LA-REIMS, LC-MS analysis was carried out on pooled samples from this data set. MS/MS fragmentation was performed using DDA tool which enables fragmentation pattern analysis to be collected of the most abundant peaks in the samples. The *m/z* of the tentatively identified features (up to three decimal places) from the LA-REIMS analysis of the CRC were checked for their presence in the list of DDA hits from the LC-MS data, a list of the *m/z* identified in both is summarised in Table

17. A mass accuracy of <5 ppm was utilised as a cut off for the cross-check of LA-REIMS and LC-MS based features.

Table 17: List of tentatively identified *m/z* (up to three decimal places) from the faecal LA-REIMS analysis of the CRC cohort present in both the LA-REIMS and checked for their presence in the list of DDA hits from the LC-MS data.

| Accurate m/z | Molecular formula of tentatively identified compound | Tentative ID | Adduct | Matched mass | ppm error |
|---|---|---|---|---|---|
| 199.170 | $C_{12}H_{24}O_2$ | Dodecanoic acid | M-H | 199.1704 | 2 |
| 381.338 | $C_{24}H_{46}O_3$ | FA(24:1(OH)) | M-H | 381.3374 | 1.57 |

Thereafter, for features that presented in both LA-REIMS and LC-MS, the most accurate m/z, the retention time, and fragmentation pattern was subsequently used to interrogate spectral reference libraries for accurate matches for validation of tentative identifications. Figures 45 and 46 depict the comparative analysis of LA-REIMS vs DDA LC-MS for tentatively identified Dodecanoic acid and FA (24:1(OH))/ 2-hydroxy-(15Z)-tetracosanoate respectively.

Experimental [M-H]- at m/z 199.1709 was tentatively identified as dodecanoic/lauric acid (theoretical m/z at 199.1704, 2.5 ppm error), Figure 45. LA-REIMS and DDA LC-MS analysis were compared. In the LA-REIMS analysis, tentatively identified dodecanoic/lauric acid at m/z

199.170 was shown to be more abundant in control vs cancer ($p<0.001$; FDR corrected $p<0.001$) and adenoma LGD vs cancer ($p<0.001$; FDR corrected $p<0.001$), see Figure 45A. The chemical structure of Dodecanoic/Lauric acid is shown for reference in Figure 45B.

Furthermore, the tentatively identified dodecanoic/lauric acid at m/z 199.170 was also present in the LC-MS/MS analysis on the pooled non-cancer control sample, Figure 45C. The pooled LC-MS/MS analysis on CRC control with MS/MS (DDA collision energy ramp start (eV) 30.0 and ramp end (eV) 80.0) is depicted in Figure 45C. Figure 45D shows the extracted ion

chromatogram of the LC-MS/MS analysis at m/z 199.17 in the DDA analysis alongside retention time elution.

The tandem MS of the feature at m/z 199.17 shows the main parent ion at 199.17 and a product ion at m/z 181.159 $[M - H_2O]^-$ m/z, e.g. loss of $H_2O$, Figure 45C. Dodecanoic/Lauric acid is a medium chain fatty acid and a hydrophobic molecule, practically insoluble in water (355). In negative ion detection mode, Electrospray Ionisation (ESI) of free fatty acids effectively yield negative ions $[M - H]^-$, however in a way that leads to little or no fragmentation in tandem MS (356). The literature of lauric acid analysed using ESI negative ionisation tandem MS, yields a negative parent ions $[M - H]^-$ at *m/z* 199.17 and a product ions obtained following collisional activation at *m/z* 181.159 which corresponds to the expected loss of $H_2O$ (98). The literature on the MS and MS/MS of lauric acid based on same analytical technique utilised, are in line with the results obtained in this experiment, adding a level of proof in the metabolite identification validation. Furthermore, Figure 45D shows the extracted ion chromatograms of the LC-MS/MS analysis at m/z 199.17, with a retention time elution at 1.92.

Furthermore, a literature review was performed on the existence of Dodecanoic acid in faecal metabolomics studies in CRC to date. Brown et al 2016, carried out a pilot study investigating faecal samples of n=17 CRC patients using GC/MS and UPLC-MS/MS (357). The UPLC-MS/MS analysis demonstrated the presence of dodecanoic acid. Goedert et al 2014 (358), studied the faecal metabolome of 48 CRC and 102 matched controls using HPLC-GC/MS-MS. The analysis revealed the presence of dodecanoic acid in both the CRC and control cohorts, however with no significant difference between the two study groups. The differences in findings between the sample set presented in this thesis and the study carried out by Goedert et al 2014, may be attributed to variations in among others sample collection, extraction, and analytical techniques utilised.

Furthermore, for instance, Fauser et al 2013, (359) treated CRC intestinal cell lines with lauric acid, butyrate, or vehicle controls. In the comparative analysis with butyrate, lauric acid showed antineoplastic properties, including induction of apoptosis in the CRC cell line.

Moreover, Weng et al 2016, demonstrated that lauric acid can improve the sensitisation of Cetuximab in KRAS/BRAF mutated CRC by lauric acid induced miR-378 expression in mutated CRC cells (360).

Furthermore, Lappano et al 2017 (361), described the mechanisms through which lauric acid causes antiproliferative and pro-apoptotic effects in both breast and endometrial cancer cells. Lauric acid induces these effects by increasing reactive oxygen species levels, stimulating the phosphorylation of EGFR, ERK and c-Jun and inducing the c-fos (proto-oncogene) expression (361). Pro-Apoptotic cell death is induced by lauric acid promoting stress fibre formation and an increase of p21Cip1/WAF1 expression, upon lauric acid exposure in a p53-independent manner (361). It should be noted that this study was conducted in in both breast and endometrial cancer cells, not CRC. Overall, growing evidence exists of the antineoplastic properties of Lauric acid. The LA-REIMS analysis in the CRC presented in this chapter revealed statistically significant increase of the tentatively identified lauric acid in the adenoma vs cancer and control vs cancer. This is in line with the literature findings, which indicate an antineoplastic property of lauric acid. The presented work here is the first study, to the best of the candidate's knowledge, identifying a significant decrease in lauric acid for cancer patient's vs adenoma and cancer patient's vs non-control utilising ambient MS in faecal samples.

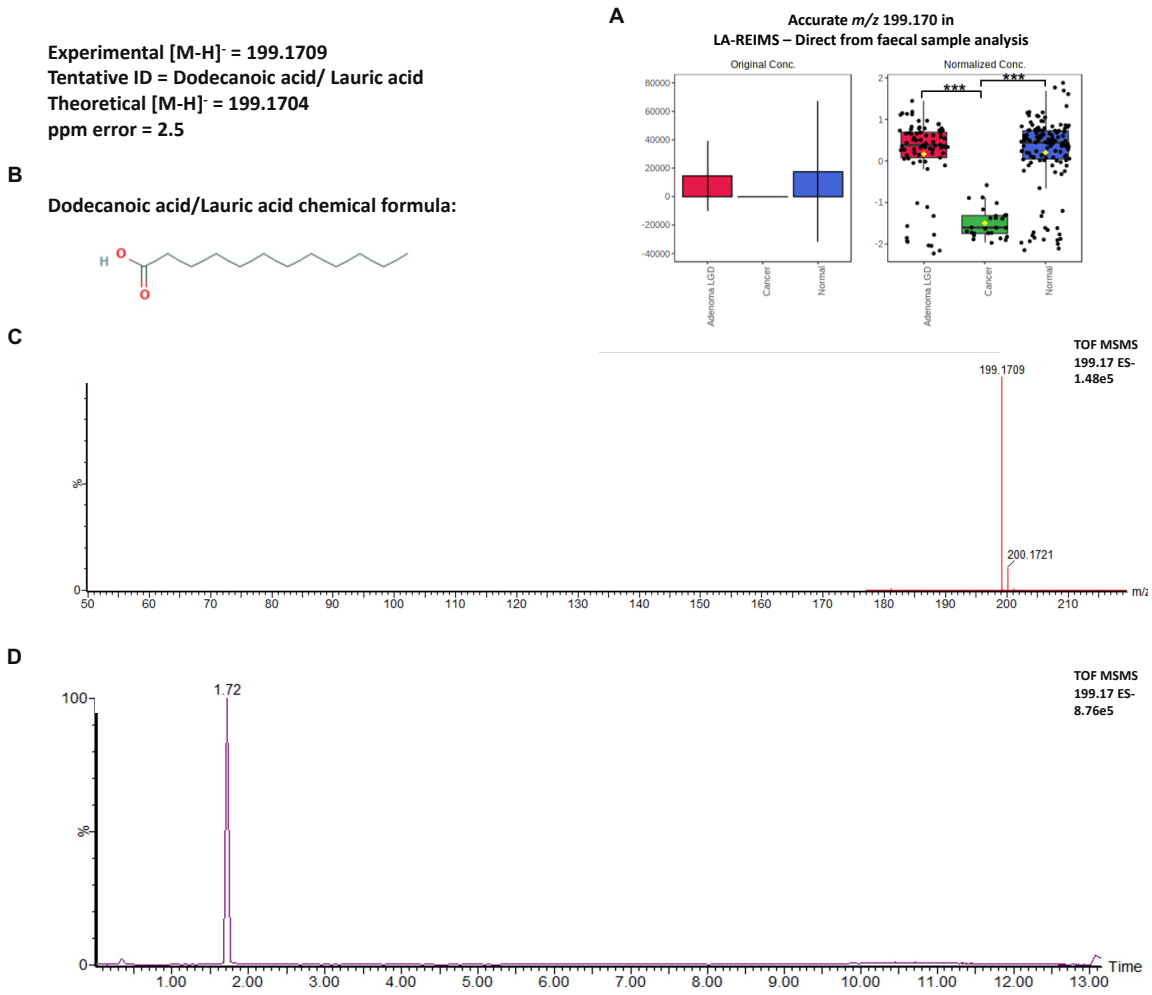



Figure 45: LA-REIMS and DDA LC-MS analysis of feature at *m/z* 199.17 tentatively identified as dodecanoic acid/ lauric acid. (A) Feature at *m/z* 199.170 in LA-REIMS direct from faecal sample analysis. Box plot of significant feature (FDR adjusted p value <0.05) determined using ANOVA (t test statistics) * <0.05, ** <0.01, *** <0.001 (B) Chemical structure of dodecanoic/lauric acid (C)Pooled LC-MS DDA analysis on CRC control with MS/MS (DDA collision energy ramp start (eV) 30.0 and ramp end (eV) 80.0) (D) Extracted ion chromatograms of the LC-MS/MS analysis at *m/z* 199.17.

LA-REIMS and DDA LC-MS analysis of feature at m/z 381.337 tentatively identified as FA (24:1(OH)) are shown in Figure 46.

Experimental [M-H]- at *m/z* 381.337 was tentatively identified FA (24:1(OH)) (theoretical *m/z* at 381.3374, ppm error of -0.26). LA-REIMS and DDA LC-MS analysis of feature at m/z 381.337 tentatively identified as FA (24:1(OH)) are shown in Figure 46. In the untargeted LA-REIMS analysis, tentatively identified FA (24:1(OH))/2-hydroxy-(15Z)-tetracosanoate at m/z 381.338 was demonstrated to be more abundant in adenoma vs cancer ($p<0.001$; false discovery rate (FDR) corrected p<0.001) and adenoma vs control ($p<0.001$; false discovery rate (FDR) corrected p<0.001), see Figure 45A.

Furthermore, the tentatively identified 2-hydroxy-(15Z)-tetracosanoate at m/z 381.338 was also present in the LC-MS/MS analysis of the pooled non-cancer control sample. MS/MS fragments observed in LC-MS DDA analysis on CRC control at m/z 381.337 are listed in Figure 46B. The corresponding pooled LC-MS/MS analysis on CRC control with MS/MS fragmentation at m/z 381.337 are depicted with areas magnified to aid visualisation of fragmentation patterns, Figure 46C (DDA collision energy ramp start (eV) 30.0 and ramp end (eV) 80.0).

In the LC-MS/MS, the parent ion at 381.337 and product ions at m/z 363.323, m/z 335.329, and m/z 307.301, which are likely expected losses of $H_20$, $H_2CO_2$, and $C_3H_6O_2$, respectively, were observed. The existing literature of hydroxylated fatty acids analysed using ESI negative ionisation tandem MS was considered. The fragmentation of hydroxy fatty acids was studied by Claeys et al, 2004 (362); a charge-driven loss of $CO_2$ then loss of $H_2$, yielding a stable alkoxide anion, was stipulated in hydroxylated fatty acids (362). The literature on the MS/MS fragmentation of hydroxylated free fatty acid are in line with the results obtained in this experiment, adding a level of proof in the metabolite validation.

In fatty acid biosynthesis and metabolism, as well as lipid peroxidation, intermediate products such as hydroxy fatty acids are formed (356). The 2-hydroxy-(15Z)-tetracosanoate described here is a hydroxylated monounsaturated fatty acid (MUFA). Furthermore, Figure 46D shows the extracted ion chromatogram of the LC-MS/MS analysis at m/z 381.337 in the DDA analysis alongside retention time elution at 3.50.

Furthermore, 2-hydroxy-(15Z)-tetracosanoate has been reported in numerous CRC metabolomics studies to date. Chen et al 2015, investigated endocannabinoids, ceramides, and free fatty acids in 47 pairs of human CRC tissues and adjacent non-CRC tissues using 3200 Q Trap LC-MS/MS system (Applied Biosystems, USA) coupled with a 1100-LC system (Agilent, China). Free fatty acids (FFA), saturated fatty acids (SFAs), MUFAs and polyunsaturated fatty acids were detected, with MUFAs from C14 - C24 consistently elevated in CRC tissues. Higher levels of expression of the C24:1 FFA was demonstrated in CRC tissue vs non-CRC tissue. Furthermore, levels of the C24:1 FFA were higher in tissue with lymph node metastasis vs tissue with no lymph node metastasis (363).

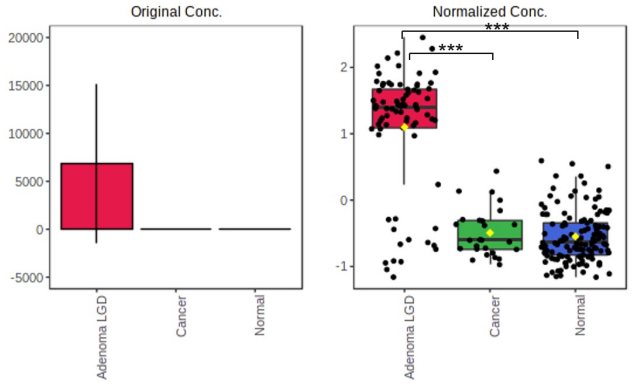
It should be noted that the LA-REIMS analysis indicated the presence of hydroxylated C24:1 fatty acid in faecal samples, with a significant increase of the tentatively identified 2-hydroxy-(15Z)-tetracosanoate in the adenoma vs cancer and adenoma vs control (for both $p < 0.001$). This indicates a potential involvement of the 2-hydroxy-(15Z)-tetracosanoate in early carcinogenesis with the hydroxylation potentially driving carcinogenesis. However, further investigation and validation of this potential mechanism are required.

To allow for further validation of the metabolite identities described above, the extracted faecal sample will be analysed alongside authentic standards using capillary GC or HPLC, combined with MS, to aid in the assignment of more definitive chemical structures. Of particular interest is the validation of the position of the -OH group on the carbon chain of the tentatively identified 2-hydroxy-(15Z)-tetracosanoate as this would point to the location of the oxidation reaction occurring. The oxidation may be either alpha, omega, peroxisomal beta oxidation, or mitochondrial beta oxidation. This is the first study, to the best of our knowledge, identifying a statistically significant increase of 2-hydroxy-(15Z)-tetracosanoate for adenoma vs cancer patients and adenoma vs non-cancer control utilising ambient MS on faecal samples.

**Experimental $[M-H]^-$ = 381.3373**
**Tentative ID = FA (24:1(OH))**
**Theoretical $[M-H]^-$ = 381.3374**
**ppm error = -0.26**

**A**

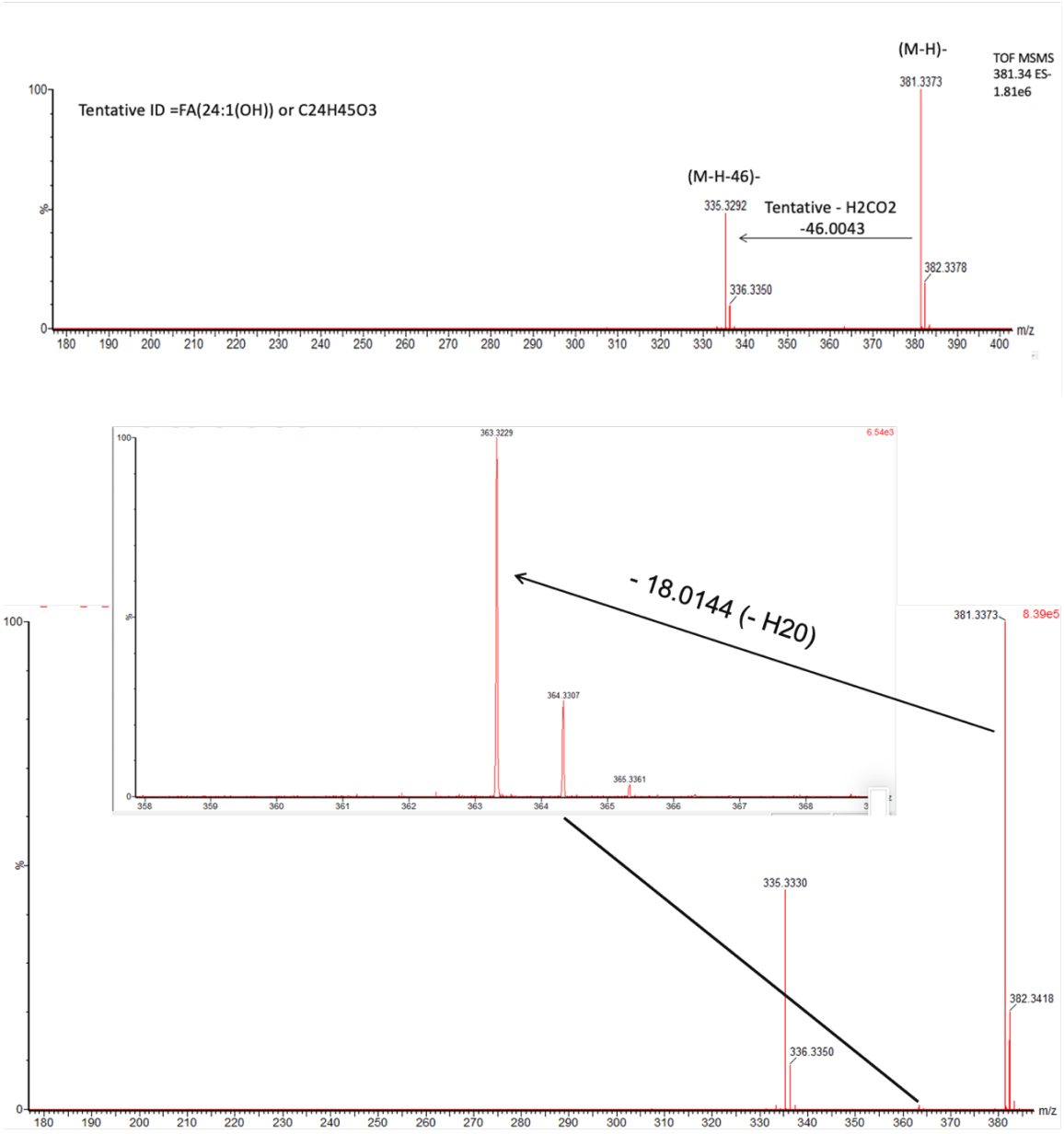
Accurate *m/z* 381.338 in
LA-REIMS – direct from faecal sample analysis




**B**

| Feature candidate | | | | Fragments | | | | |
|---|---|---|---|---|---|---|---|---|
| m/z for [M-H]- | Tentative ID | Theoretical mass | ppm error | Ion | ID | m/z fragment | Theoretical m/z fragment | ppm |
| 381.337 | FA(24:1(OH) | 381.337 | -0.26 | [M-H]- | (C24H43O2)- | 363.323 | 363.326 | -9.36 |
| | | | | [M-H]- | (C23H43O)- | 335.329 | 335.331 | -6.56 |
| | | | | [M-H]- | (C21H39O)- | 307.301 | 307.300 | 1.30 |

**C**


(M-H)-
TOF MSMS
381.34 ES-
1.81e6
Tentative ID =FA(24:1(OH)) or C24H45O3
381.3373
(M-H-46)-
335.3292
Tentative - H2CO2
-46.0043
382.3378
336.3350
m/z
363.3229
6.54e3
- 18.0144 (- H20)
364.3307
365.3361
381.3373
8.39e5
335.3330
382.3418
336.3350
307.3005
1.70e3
- 74.0368 (- C3H6O2)
308.3065
308.2958
305.2827
305.2039
307.5339
309.3106
311.1575
313.2889

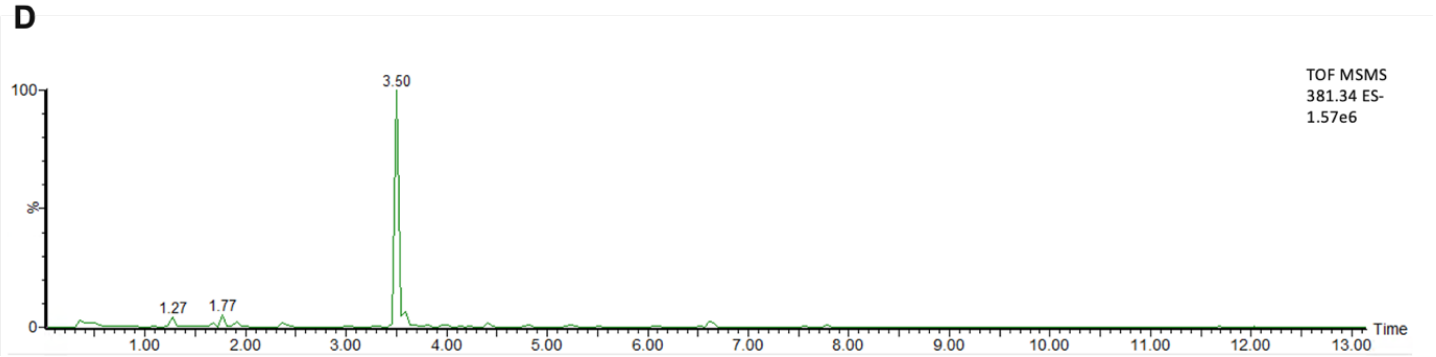



Figure 46: LA-REIMS and DDA LC-MS analyses of feature at m/z 381.337 tentatively identified as FA (24:1(OH)). (A) Accurate *m/z* 381.337 in LA-REIMS– Direct from faecal sample analysis. Box plot of significant feature (FDR adjusted *p* value <0.05) determined using ANOVA (t test statistics) * <0.05, ** <0.01, *** <0.001 (B) List of MS/MS fragments for LC-MS DDA analysis on CRC control (C) Pooled LC-MS DDA analysis on CRC control (DDA collision energy ramp start (eV) 30.0 and ramp end (eV) 80.0) – MS/MS fragmentation for m/z 381.337 depicted with areas magnified to aid visualisation of fragmentation patterns (D) Extracted ion chromatograms of the LC-MS/MS analysis at *m/z* 381.337

While the previous section discussed the LC-MS validated features, it is also important to consider the features which were shown to be significantly different between control, adenoma, and/or cancer based on the untargeted LA-REIMS analysis, Table 17, but which were only present in LA-REIMS and not the LC-MS analysis. To add a level of proof in validation measurements and increase identification power, LA-REIMS tandem MS experiments will be carried out, where possible, to obtain the fragmentation pattern of features of interest.

Although LA-REIMS faecal metabolomics presents a promising technique, the limitation of metabolite identification needs to be noted. Metabolite identifications are based on accurate mass measurement, which limits absolute quantitation and identification accuracy. For validation purposes and to increase identification power, tandem MS experiments will be carried out, to obtain the fragmentation pattern of features of interest. However, it has to be noted that the absence of chromatography, complicates tandem MS spectra, because isobaric and isomeric molecules cannot be separated (120). Furthermore, ion suppression in LA-REIMS

is a common problem which may pose additional difficulties for metabolite identification. LA-REIMS analysis was therefore complemented with more traditional LC-MS-based validation for the purpose of biomarker discovery.

### 5.1.3. Major patient-related modifiers of the LA-REIMS faecal metabolome in CRC

To assess study population related modifiers of the faecal metabolome, demographic information was assessed by building individual PCA models, Figure 47, including of age, gender, ethnicity, and BMI. As described previously in section *1.10. Colorectal cancer,* research indicates that CRC incidence and mortality rates increase with age, with the disease being most common in people older than 50 and an estimated 90% of cases and deaths occurring after this age (159,176). Furthermore, CRC is more common in men than women (364). Moreover, CRC incidence varies largely among ethnicities, for instance in the United States, African Americans and Native Americans have the highest incidence rates of CRC. Hispanic Americans show the same rates as do white Americans (157). It is important to consider that unequal exposure to risk factors and medical care might influence ethnic disparities (172). Obesity is an established risk factor for CRC, increased BMI (>25) has been associated with CRC (365).

The contributing factors of age, gender, ethnicity, and BMI-related risk of CRC were investigated using the PCA modelling in Figure 47. Age was investigated in the age ranges 21-50, 51-59, 60-74, 75-95 years, Figure 47A. Sex or gender (f=female, m=male) confounding analysis is shown in Figure 47B. Ethnicity was categorised as White, Asian, or Other, Figure 47C. BMI was categorised as <18=underweight, 18.5-24.9=normal, 25.0-29.9=overweight, >30.0=obese, Figure 47D.

In Figure 47A, PCA modelling indicates the greatest amount of variation lies within the age range (21-50), with age ranges, 51-59,60-74, 75-95 showing considerably less intra-group variation. Furthermore, the youngest age range group, 21-50, seems to cluster away from the oldest age range group 75-95, indicating a potential difference in the faecal metabolic profile of the two. The effect of age on the PCA modelling is important to consider when interpreting results of significant differences between disease cohorts as the effect of age may confound

the results observed. This is of special importance, as mean age was shown to be significantly different between the non-cancer control vs adenoma patient groups (*p* value = 0.0002) and between non-cancer control versus cancer patients (*p* value = 0.0001), see Table 13*: Baseline demographic characteristics of study cohort.*

Gender/sex show no clustering of the female and male participants, indicating a similar faecal metabolic profile between the two, Figure 47B. Furthermore, based on the baseline demographic characteristics table, no significant differences were established between the disease groups based on the number of male vs female participants involved. This indicates that overall, gender/sex does not lead to a visible change in the faecal metabolomic profile based on multivariate analysis, and the number of male vs female participants were evenly distributed between the study groups.

Ethnicity was not found to be significant, see Table 14. It is important to consider that most participants in the study cohort (92.5%) were White, with Blacks (2.1%) and Asians (5.4%) underrepresented, Table 13. The PCA plot of ethnicity, Figure 47C, shows a relatively homogenous distribution of the ethnicity categories, however, White clusters slightly to the right of Asian and Other.

In Figure 47D, PCA modelling indicates that the underweight cohort, shows the least amount of variation, whilst the remaining BMI groups show similar amounts of variation. However, it is important to consider that the underweight group contains fewer participants than the other BMI groups. Furthermore, the underweight group, seems to cluster away from the other BMI groups, indicating a difference in the faecal metabolic profile. Furthermore, mean BMI (+range) was shown to be significantly different between non-cancer control versus cancer patients (*p* value = 0.0314) and cancer versus adenoma patients (*p* value= 0.0151), see Table 14*: Baseline demographic characteristics of study cohort.* Overall, age and BMI seem to be the most important potential confounders in this CRC dataset.

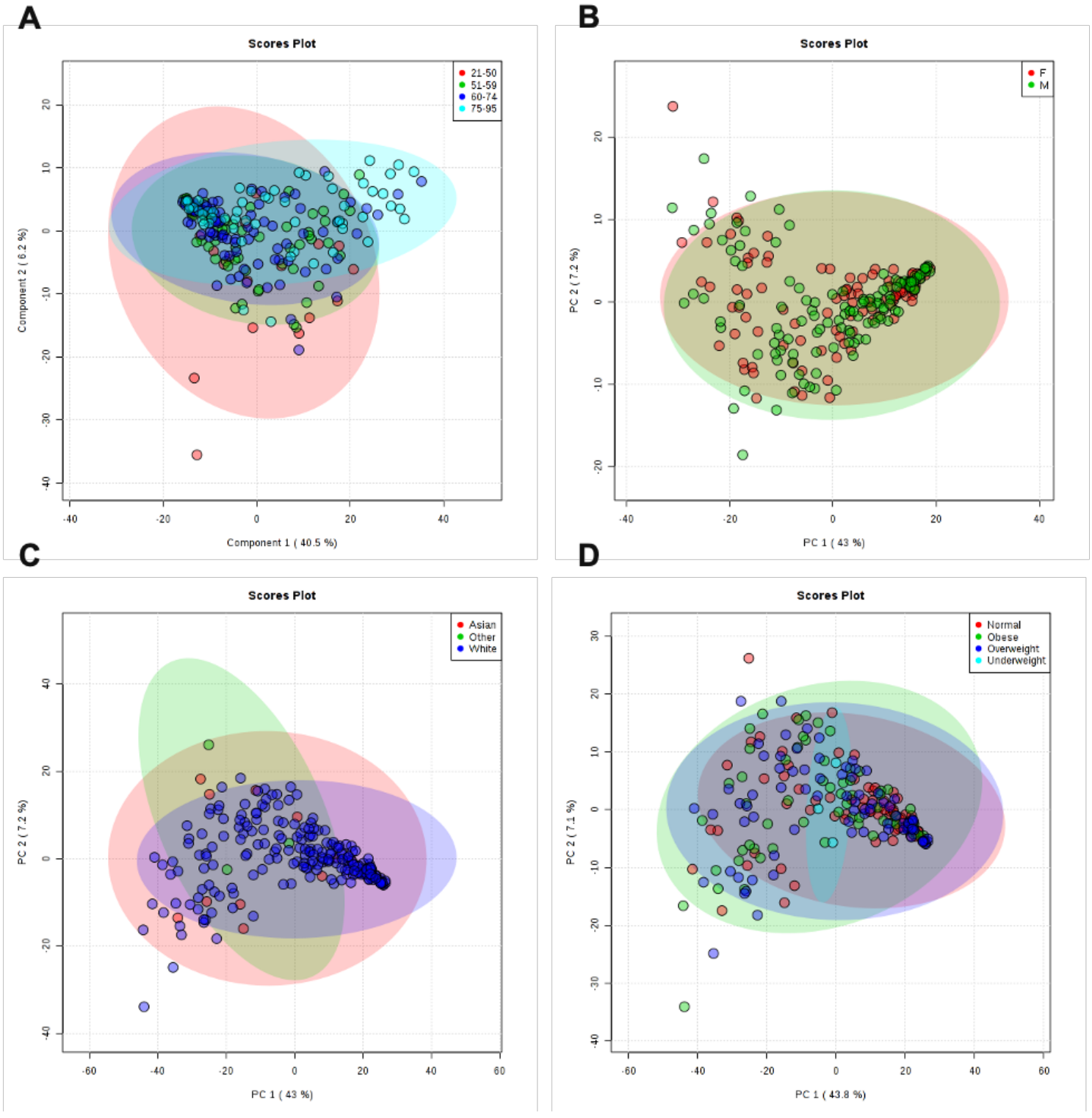



Figure 47: PCA scores plots in m/z range 50-1200 assessing A Age (ranges include: 21-50, 51-59, 60-74, 75-95), B Gender (f=female, m=male), C Ethnicity (categorised as White, Asian,

other), and D BMI (categorised as <18=underweight, 18.5-24.9=normal, 25.0-29.9=overweight, >30.0=obese). Shaded region represents 95% confidence intervals.

PCA models were built utilising existing comorbidity data to assess their potential contribution to the LA-REIMS faecal metabolomic analysis and the significant differences observed for the investigation of adenoma, non-cancer control, and cancer groups. To assess the potential confounding effect of haemorrhoids on the PCA on non-cancer control patients, a PCA model of non-cancer control vs (non-cancer control + haemorrhoids) was created, Figure 48A. To assess the potential effect of diverticular disease on adenoma clustering in unsupervised multivariate analysis, a PCA model of adenoma vs (adenoma + diverticular disease) was built, Figure 48B. Furthermore, a PCA model of adenoma + diverticular disease) vs (non-cancer control + haemorrhoids) was created, Figure 48C. For Figure 48A-C, no difference or distinct clustering was visible based on the comorbidities investigated or when comparing comorbidities as well as disease groups (Figure 48C).

Overall, this suggests that comorbidities do not impact the LA-REIMS faecal metabolome and the distribution of the cancer, adenoma, and non-cancer groups substantially. Age and BMI are factors that need to be considered when interpreting results of this work and should be considered when planning future studies in this area. The significant differences in mean age and BMI between the study groups is a limitation of this study. It is important to note that this work presents a superficial confounding analysis to check for main factors that may be driving the multivariate models. More detailed confounding analysis will be carried out in future work.

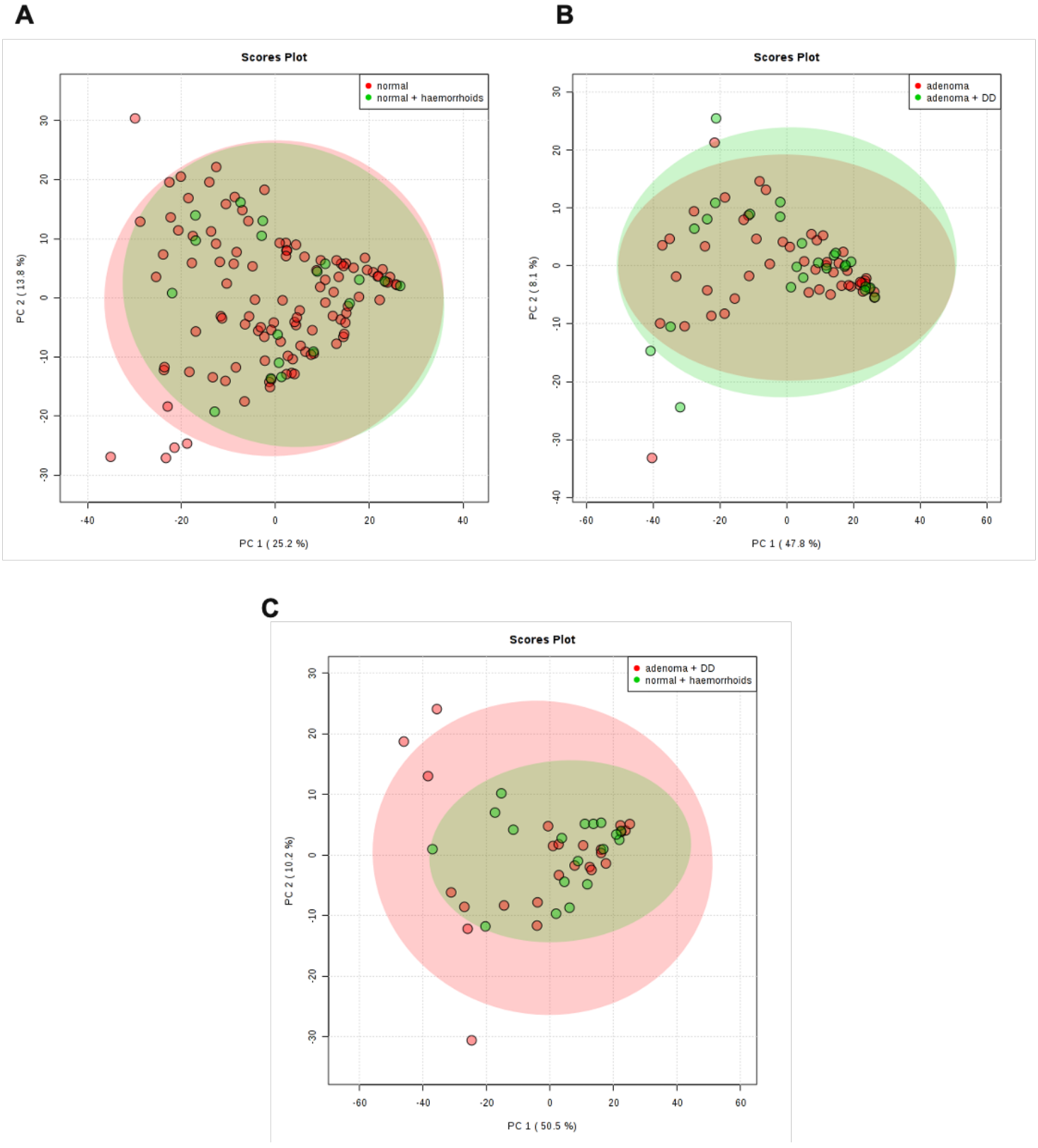



Figure 48: PCA scores plots in m/z range 50-1200 assessing A Non-cancer control vs (non-cancer control + haemorrhoids) B Adenoma vs (adenoma + diverticular disease) C (adenoma

+ diverticular disease) vs (non-cancer control + haemorrhoids). Note: DD indicates diverticular disease

### 5.1.4. Limitations

The results of the LA-REIMS modelling are promising but a major limitation is that the findings have not been validated in a prospectively recruited independent data set, this will be targeted in future work. The faecal samples investigated using LA-REIMS analysis underwent multiple freeze-thaw cycles and were stored at 4°C degrees for prolonged periods of time at the point of sample collection and shipment. It should be considered that freeze thaw cycles and prolonged storage at 4°C degrees can cause changes in the metabolomic profile of samples (33).

An important factor to note is that the detection of phenotypic dysregulation from bio-samples such as faeces is difficult using high-throughput techniques due to the complexity of the matrix. Bacterial species can modify metabolites present after evacuation of the faecal sample and the analysis can be affected significantly by confounding factors from for instance diet and water content. The faecal metabolome is heterogenous and highly influenced by dietary intake, with diet being one of the major causes of variation in faecal sample composition (37). The present dataset lacked detailed comprehensive dietary information, in the form of 24-hour or 72-hour dietary recall questionnaires; diet or time of food ingestion were not controlled for. Diet can act as a confounding factor in faecal metabolomics studies and needs to be controlled for or evaluated in the form of dietary questionnaires.

## 5.2. Conclusion and Future work

LA-REIMS demonstrates potential for accurate direct from faecal sample adenoma and CRC detection. This has potential as a novel screening tool for early CRC detection and could reduce the burden of unnecessary colonoscopy and optimise cancer referral pathways. The implementation of a testing pipeline with both FIT and LA-REIMS analysis performed on the same faecal sample of individuals to be triaged for further testing, that is colonoscopy, may

be considered. The study was limited in the number of cancer patients, validation in an independent prospectively recruited dataset is needed. Furthermore, validation of the tentatively identified biomarkers through comparative analysis with a standard will be performed using ambient MS and LC-MS based tandem MS techniques. A targeted prospectively recruited targeted study for detection of early colorectal cancer will enable more conclusive evidence as to their potential for CRC screening. Patient-related modifiers of the faecal metabolome, such as age and BMI need to be considered in more detail, as these were shown to be potential confounding factors in the CRC analysis. A further goal is to establish interactions between diet, microbiota, and metabolome, and linking this information to their role in the adenoma-carcinoma sequence to provide potential mechanistic insights into carcinogenesis. For this a detailed confounding analysis of specifically dietary factors and microbial interactions are needed.

## 6. Optimised LA-REIMS faecal metabolomics analysis for the detection of Inflammatory Bowel Disease

Various non-invasive faecal markers for IBD exist, these include among others lipocalin-2, lactoferrin and FC (282,366) . One of the most widely and promising tests to date is FC (282). In the UK, NICE guidelines recommend FC testing only for differentiating IBD and IBS in patients presenting to primary care with lower GI symptoms when referral for secondary care is considered and cancer is not suspected (367). Overall, FC is inflammation and not IBD specific, it demonstrates high diagnostic accuracy for differentiating organic GI disorders (IBD, CRC, etc.) from functional disorder (IBS) (286). However, since several diseases other than IBD can increase faecal disposal of calprotectin, its specificity to identify IBD is lower than desirable (289). Furthermore, a lack of evidence exists for the use FC for differentiating the CD from UC (282). To date, no single biomarker for the diagnosis, assessment of disease severity, and evaluation of treatment response for IBD exists (236).

An in-depth analysis and review on the current literature of markers in IBD diagnostics is provided in section *1.14.1. Diagnosis, assessing disease severity, and evaluating treatment response* and *1.15.1. Metabolomics and lipidomics in IBD.* The diagnostic pathways of IBD includes a combination of clinical, endoscopic, radiological, histological and/or biochemical examinations (368), with the current gold standard being endoscopic investigation and histology results, which are expensive, time-consuming and uncomfortable for the patient (268). A need for the development of non-invasive cheap biomarkers that provide disease specific IBD signatures as well as discriminate between CD and UC exist (291). This work presents a feasibility study of optimised faecal LA-REIMS as a diagnostic technique for IBD.

Optimised LA-REIMS would fit into the diagnostic pathway of IBD as a potential addition to FC for stratifying patients suspected of IBD in the primary care setting in a non-invasive manner (369). Furthermore, current guidelines stipulate that faecal samples from patients presenting to primary care with lower GI symptoms (when referral for secondary care is considered and cancer is not suspected) are collected for FC analysis (285). Faecal samples

collected could be divided for both FCB and direct optimised LA-REIMS analysis, with samples sent to a laboratory where LA-REIMS analysis would be carried out.

The main benefits of the LA-REIMS system include high-throughput, high-turnover, direct from sample analysis and real-time data generation (151). Furthermore, a single LA-REIMS instrument allows the analysis of >500 samples per day with an estimated cost of £20 per sample (estimated costs per sample included consumable, instrument depreciation, and personnel) (120). However, it must be noted that the candidate presents the first feasibility testing of this setup for IBD and further validation and method development studies are needed to assess the true translational potential of the optimised LA-REIMS instrument for the study of faecal markers in IBD.

## 6.1. Results

In total, 169 frozen faecal samples (CD n=60, UC n=60, and non IBD control n= 49) were analysed from 169 patients. Baseline demographics of the patients are outlined in Table 18. Statistical comparison between groups, *p* value of <0.05 considered significant, were analysed by a combination of Kruskal-Wallis test, Chi-squared test, and Mann-Whitney U test. Significant differences were found in terms of aminosalicylic acids (5-ASA), specifically mesalazine**,** between CD vs UC ($p$=0.017), see Table 18. Furthermore, significant differences were also found in terms of biologic ($p$=0.004) and non-IBD medication ($p$<0.001). Significant differences between CD, UC, and non-IBD controls were found in terms of the total co-morbidities ($p$<0.001), see Table 18.

Table 18: Baseline demographic characteristics of IBD study subjects.

| | CD, n = 60 | UC, n = 60 | Controls, n = 49 | P value[††] |
|---|---|---|---|---|
| Age (median) | 46 | 54 | 54 | 0.01[a] |
| Sex (M: F) | 25: 35 | 28: 32 | 21: 28 | 0.848[b] |
| Ethnicity (Cau:SA:Oth)[*] | 35:24: 3 | 30: 27: 1 | 22: 26: 1 | 0.534[b] |
| BMI (median) | 24.2 | 24.7 | 26.2 | 0.156[a] |
| Disease location[**] | L1: 10<br>L2: 31<br>L3: 18 | E1: 17<br>E2: 20<br>E3: 21 | -<br>-<br>- | -<br>-<br>- |
| Disease activity†: | Remission: 47<br>Active: 13 | Remission: 50<br>Active: 10 | -<br>- | 0.326 [b] |
| Operation | 9 | 4 | - | 0.180[b] |
| 5-ASA | 37 | 22 | - | **0.017** [b] |
| Oral immuno-modulator | 25 | 16 | - | 0.142[b] |
| Biologic | 13 | 2 | - | **0.004**[b] |
| Steroid | 2 | 3 | 1 | 0.673[b] |
| Non-IBD medication | 8 | 7 | 24 | **<0.001[b]** |
| Co-morbidity: | 7 | 13 | 25 | **<0.001[b]** |

[*]Ethnicity; Cau = Caucasian, SA = South Asian, Oth = Black/Other. **Montreal classification, CD L1 = ileal, L2 = colonic, L3 = ileo-colonic, UC E1 = proctitis, E2 = left sided, E3 = extensive. †CD HBI score, remission =<5 and active ≥5, UC SCCAI score, remission <5, active ≥5. ††Statistical comparison between groups, p value of <0.05 considered significant and are in bold. [a]Analysed by Kruskal-Wallis test, [b]analysed by Chi-squared test, [c]analysed by Mann-

Whitney U test. *Cohort contains subjects with no co-morbidities and subjects with non-IBD co-morbidities. Significant p values ($p < 0.05$) in bold.

### 6.1.1. Negative vs positive ion detection and faecal sample vs faecal water extract analysis

The analysis of the IBD dataset was carried out for both faecal sample as well as faecal water (2:1 ratio water to faeces) extracts in positive and negative ion detection mode. Example raw spectra of negative and positive mode spectra for both faecal sample and faecal water extracts for a randomly chosen study participant in the non-IBD control group are shown in Figure 49.

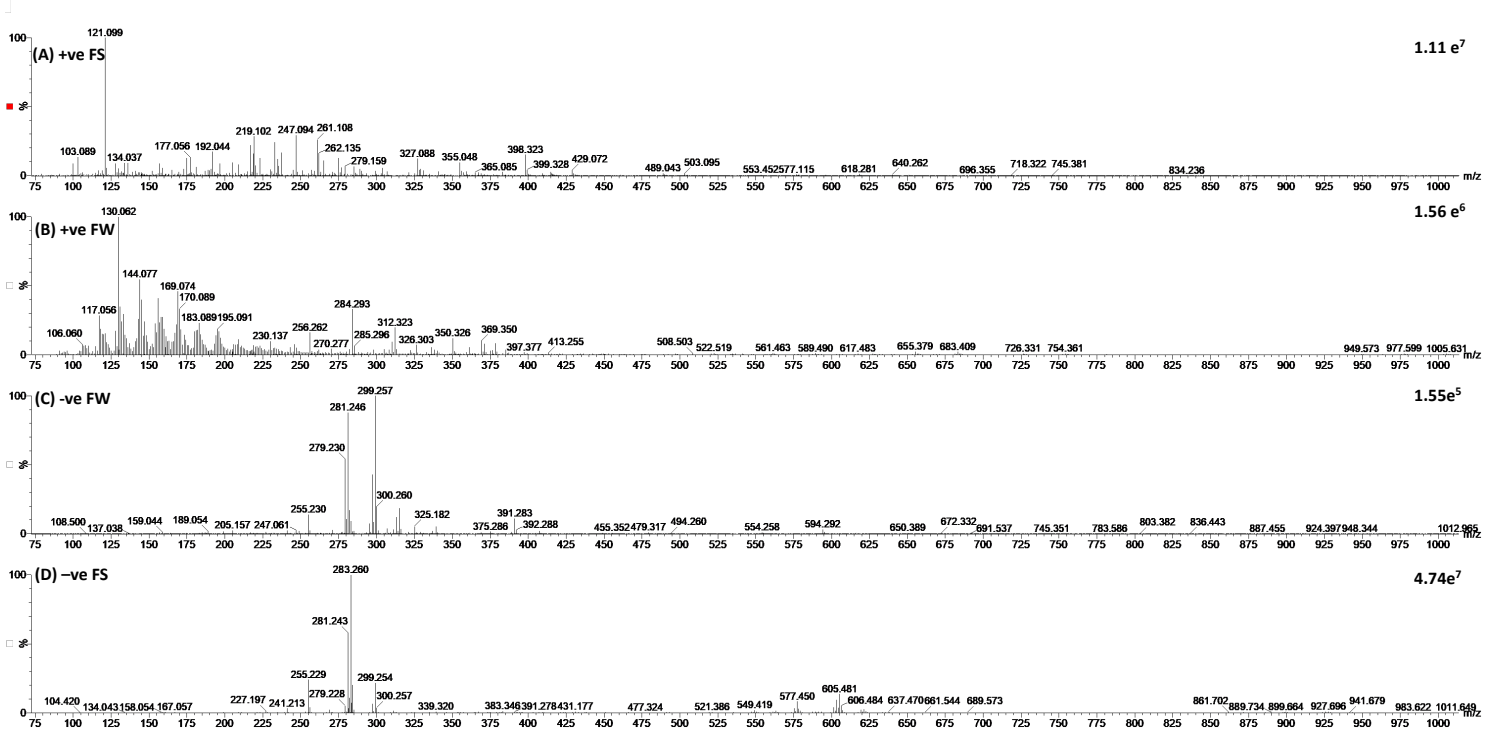



Figure 49: Example mass spectra of optimised faecal LA-REIMS analysis (A) Positive ion detection mode faecal sample (+ve FS) (B) positive ion detection mode faecal water (+ve FW) (C) negative ion detection mode faecal water (+ve FS) (D) negative ion detection mode faecal

sample (+ve FS). Background subtraction and lock mass correction were carried out for spectra in the *m/z* range 50 to 1200.

From visual inspection of the mass spectra, faecal sample negative ion detection mode spectra contain highest average intensity peaks in the *m/z* range 50-1200. Specifically, in the complex lipid region, faecal sample negative ion detection mode demonstrates high intensity of peaks. As previously described in section *5.2.1. Negative vs positive ion detection mode and faecal sample vs faecal water extraction,* complex lipids are key to GI disorder pathogenesis. The analytical methodology providing high signal intensity of *m/z* range where complex lipids are found, are therefore favourable.

Analytical noise levels appear higher in positive mode vs negative ion detection mode, for both faecal water and sample analysis, see Figure 49 A-D. Moreover, the optimisation work carried out in the analysis of the CRC cohort, *Chapter 5*, indicates a similar trend with positive ion detection mode. Furthermore, when comparing the spectra of faecal sample negative ion detection mode and faecal water negative ion detection mode, the average intensity of peaks of potential interest is significantly reduced in faecal water negative ion detection mode. The reduction in intensity may be due to the dilution of the faecal sample at a 2:1 ratio of water to faeces. Overall, after careful visual investigation of the whole range of the raw spectra (Figure 49 contains a typical spectra of mass range m/z 50-1200) faecal sample negative ion detection mode was chosen for further statistical analysis using multivariate and univariate analysis. An in-depth analysis of the data analysed in faecal sample negative mode follows.

### 6.1.2. Faecal sample negative ion detection mode

#### *6.1.2.1. Unsupervised and supervised multivariate statistical analysis*

Faecal samples from 169 patients (CD n=60, UC n=60, and non IBD control n= 49) were analysed using the optimised LA-REIMS setup in negative ion detection mode. The PCA in the 50−1200 *m/z* range for faecal samples from non-IBD control, CD, and UC patients (negative ionisation mode) yielded no separation based on underlying disease pathology in both the

pairwise PCA scores plot for the top five principal components (PC)s, Figure 50A, as well as the scores plot of PC 1 (48.7%) and PC2 (4.7%), Figure 50B. Details of the PCA analysis and why it was chosen as the preferred data analysis tool is summarised in section *3.3.3.4. Data analysis – Metaboanalyst*.

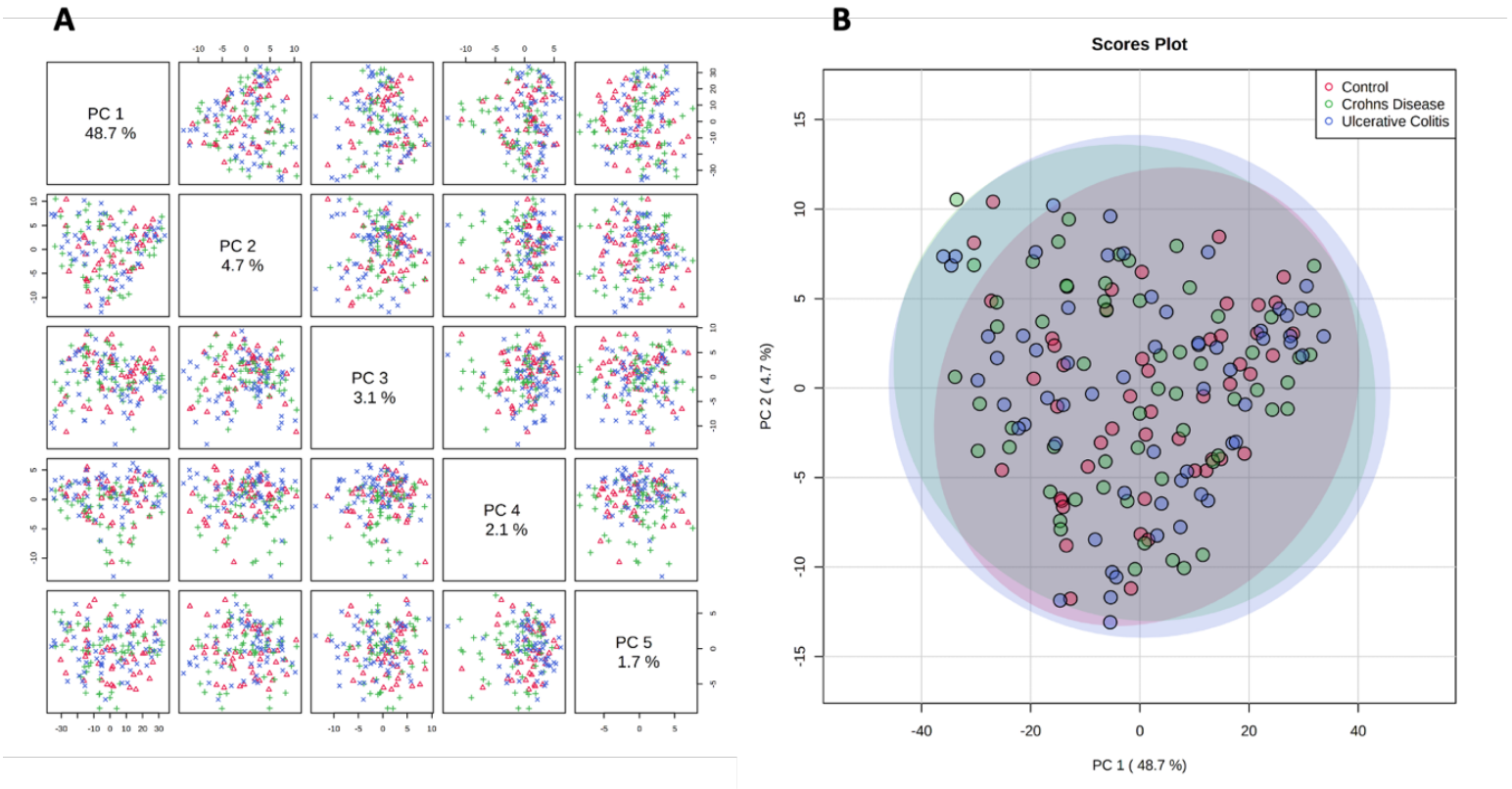



Figure 50: (A) Pairwise PCA scores plot for the top five PCs. (B) Scores plot of faecal LA-REIMS of PCA analysis PC 1 (48.7%) and PC2 (4.7%) over the 50−1200 *m/z* range for non-IBD control, CD, and UC patients (negative ion detection mode). Shaded region represents 95% confidence intervals.

To maximize the group separation, supervised PLS-DA model for faecal samples from non-IBD control, CD, and UC patients in the *m/z* range 50-1200 (negative ion detection mode) were built. Details of the PLSDA analysis and why it was chosen as the preferred data analysis tool is summarised in section *3.3.3.4. Data analysis – Metaboanalyst*. Cross-validation using LOOCV with Q2 as performance measure (maximum components 5) was performed and PLS-DA models underwent permutation testing. The PLS-DA model of non-IBD control, CD, and

UC patients show separation based on the three study groups, with UC vs non-IBD control demonstrating the largest separation and CD clustering in between UC and non-IBD control, Figure 51A+B. The PLS-DA yielded $R^2$=0.90 and $Q^2$=0.60. Permutation testing indicated model fitness, Figure 51C.

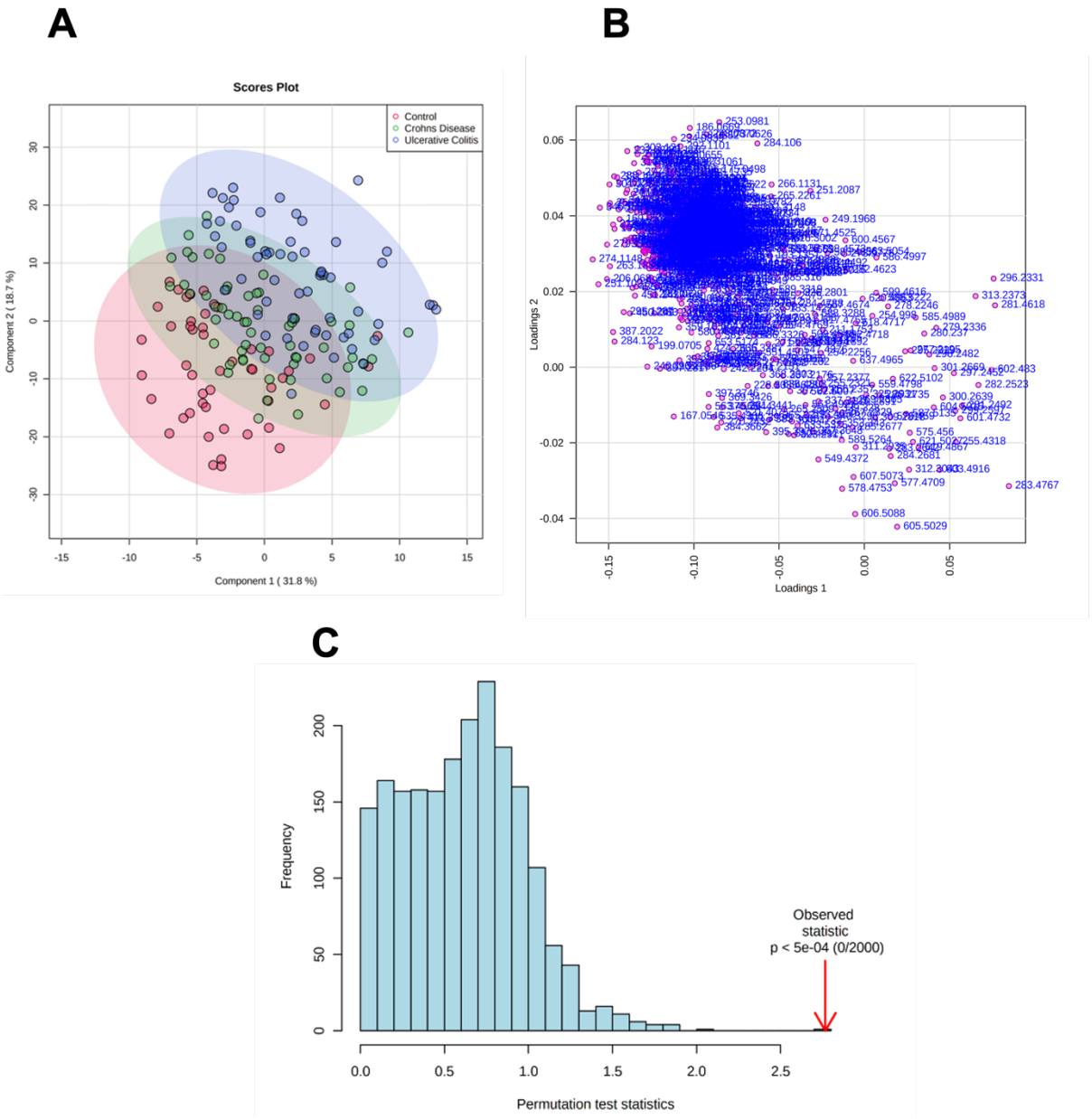


Figure 51: Supervised PLS-DA model for faecal samples from non-IBD control, CD, and UC in the *m/z* range 50-1200 (negative ion detection mode) (A) Scores plot (B) Loadings plot (C) Permutation test; select test statistic: Separation distance (B/W) with set permutation numbers: 2000 $p<0.001$. Red vertical error indicates observed data test statistic; y-axis indicated frequency of permuatations; x-axis indicates test statistic values.

#### *6.1.2.2. Diagnostic accuracy of LA-REIMS*

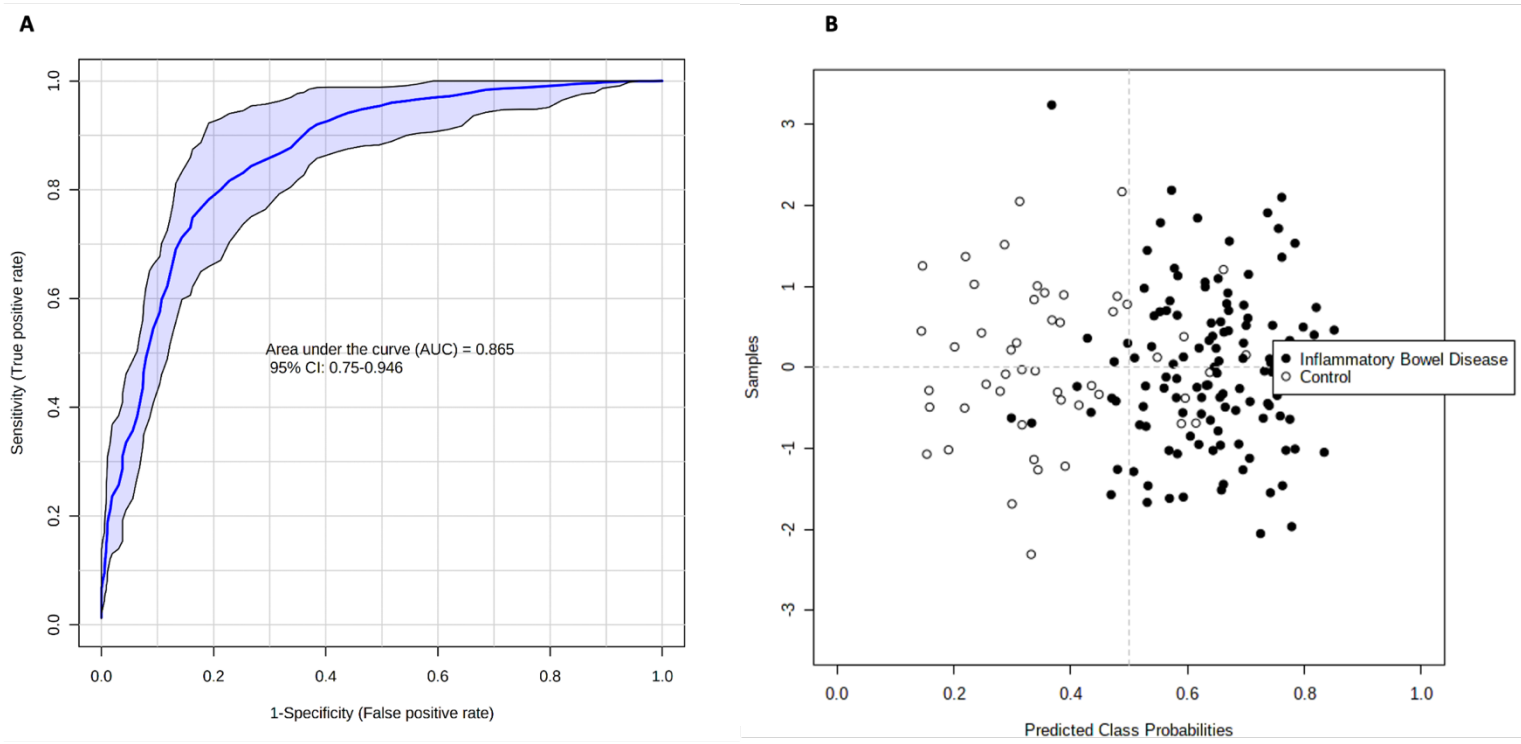
RFC algorithms were utilised for sample classification purposes and ROC curves were generated by MCCV to assess the diagnostic performance of faecal LA-REIMS for IBD vs non-IBD control, CD vs UC, and CD vs non-IBD control, and UC vs non-IBD control. Details of the RFC and ROC curve analysis and why they were chosen as the preferred data analysis modalities are summarised in section *3.5.9. IBD.*

##### 6.1.2.2.1. LA-REIMS – IBD vs non-IBD control

To investigate IBD vs non-IBD control using RFC, two cohorts were investigated, first the analysis of all IBD (n=120) vs non-IBD control (n=49), and second to perform a classification on an even group of participants, n=45 IBD randomly chosen participants and n=45 randomly chosen non-IBD control patients were studied. This two-factor approach was chosen to check any effects the large difference in group size may have on the RFC model.

RFC performance evaluation are shown in Figure 52 (whole group analysis) and Figure 53 (selected even group) analysis. For the whole group analysis, the ROC curve indicates an area under the curve (AUC) of 0.865 (95% confidence interval CI: 0.75-0.946), Figure 52A. Figure 53A shows a ROC curve indicating an AUC of 0.88 (95% confidence interval CI: 0.736 to 0.985), for selected even group analysis. The predicted class probability plot in Figure 52B demonstrates correct classification of IBD and non-IBD control samples for the whole cohort of IBD vs non-IBD control. In the whole group analysis, LA-REIMS demonstrates a sensitivity

of 90.0% and specificity of 83.7% in IBD vs non-IBD control patients (Figure 52C). The predicted class probability plot in Figure 53B demonstrates correct classification of IBD and non-IBD control samples for the sub-set cohort. Here, LA-REIMS demonstrates a sensitivity of 77.8% and specificity of 77.8% in IBD vs non-IBD control patients (Figure 53C).




C

| Predicted class \ Actual class | IBD | Control |
|---|---|---|
| IBD | 108 | 8 |
| Control | 12 | 41 |

Figure 52: RFC performance of the LA-REIMS analysis of IBD (n=120) vs non-IBD control (n=49) for the whole study group population. (A) RFC method: ROC curve with AUC and 95% confidence interval (B) Predicted class probabilities plot utilising the best AUC classifier. IBD represented by black dots, Normal (non-IBD control) as white dots. (C) Confusion matrix of IBD (n=120) vs non-IBD control (n=49).

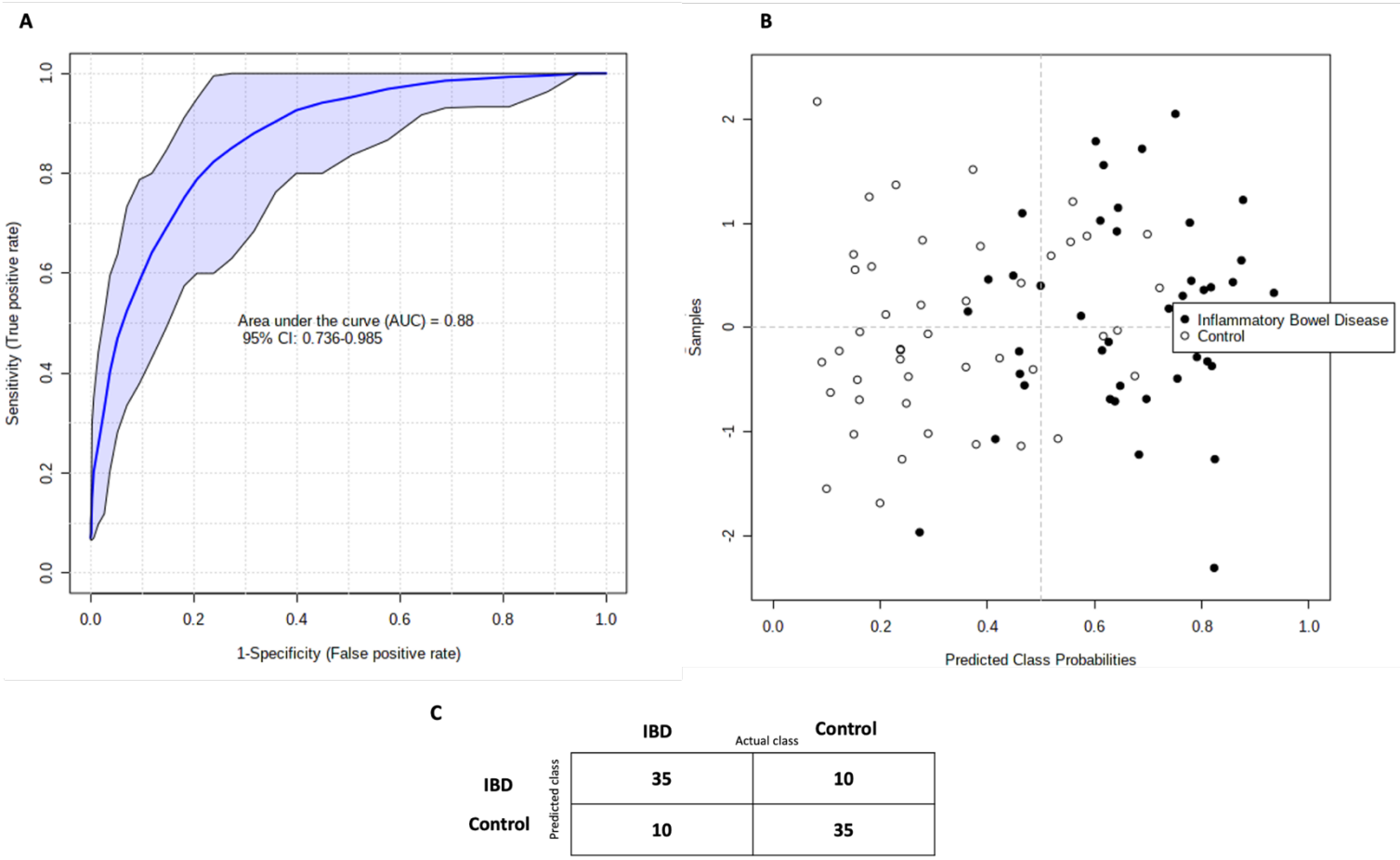
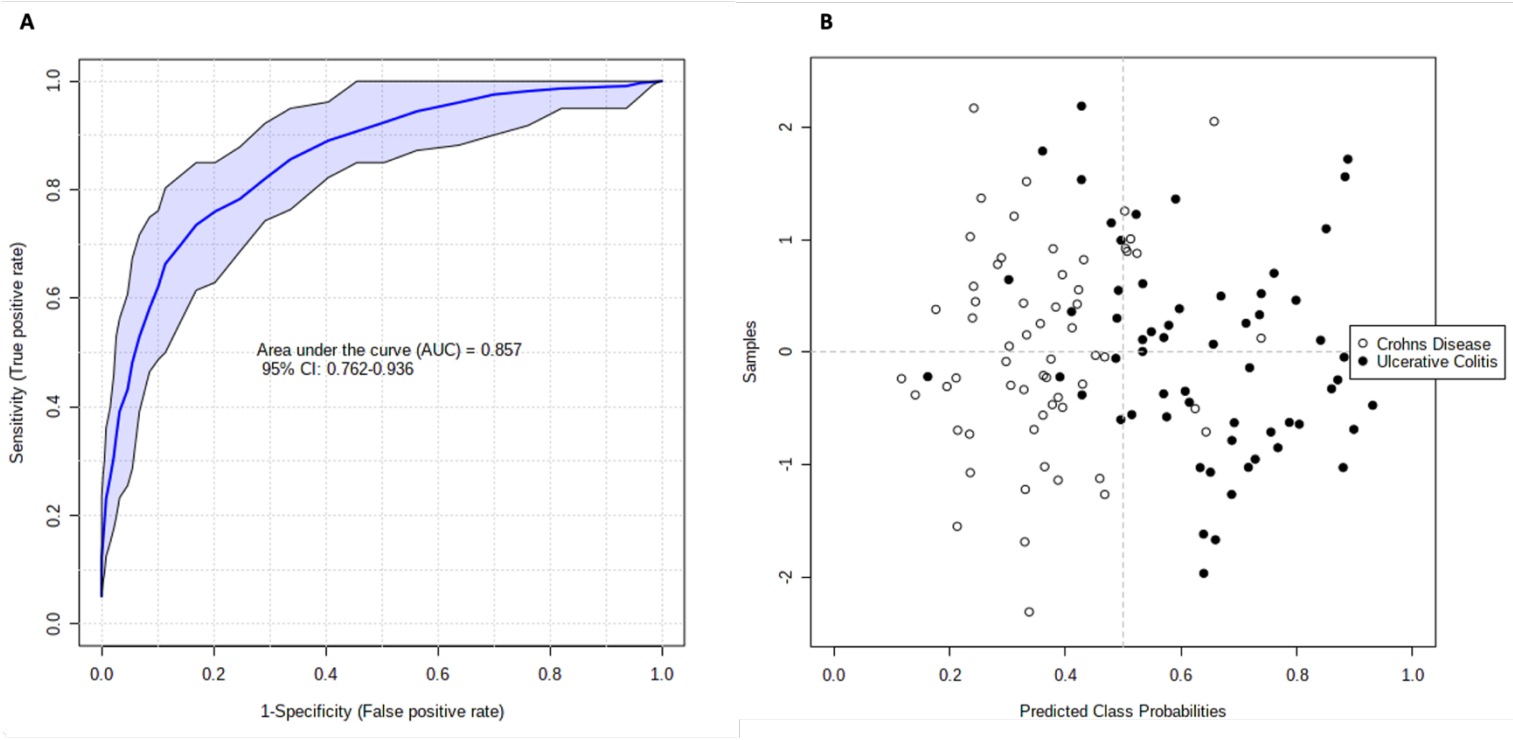



Figure 53: RFC performance of the LA-REIMS analysis of IBD (n=45) vs non-cancer control (n=45) for even number of study participants in each group. (A) RFC method: ROC curve with AUC and 95% confidence interval (B) Predicted class probabilities plot utilising the best AUC classifier. IBD represented by black dots, Normal (non-IBD control) as white dots. (C) Confusion matrix of IBD (n=45) vs non-cancer control (n=45).

#### 6.1.2.2.2. LA-REIMS – Crohn's Disease vs Ulcerative Colitis

The RFC algorithm was utilised to assess the classification performance of the high-throughput faecal LA-REIMS setup for CD vs UC study participants in the IBD cohort. RFC performance evaluation are shown in Figure 54, the ROC curve indicates an AUC of 0.857 (95% confidence interval CI: 0.762 to 0.936), Figure 54A. The predicted class probability plot

in Figure 54B demonstrates correct classification of CD and UC samples. LA-REIMS demonstrates a sensitivity of 85.0% and specificity of 76.7% in CD vs UC patients (Figure 54C).



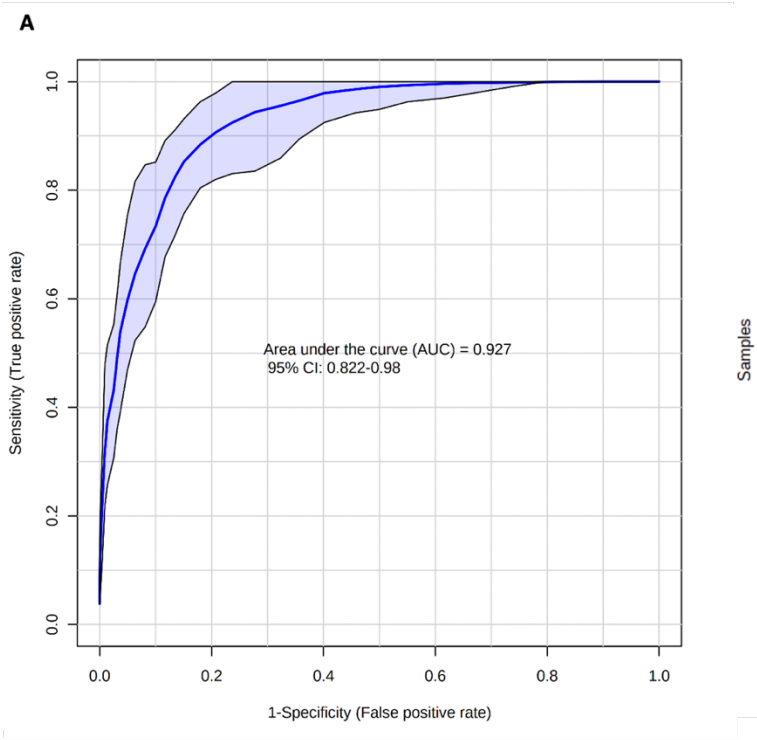

C

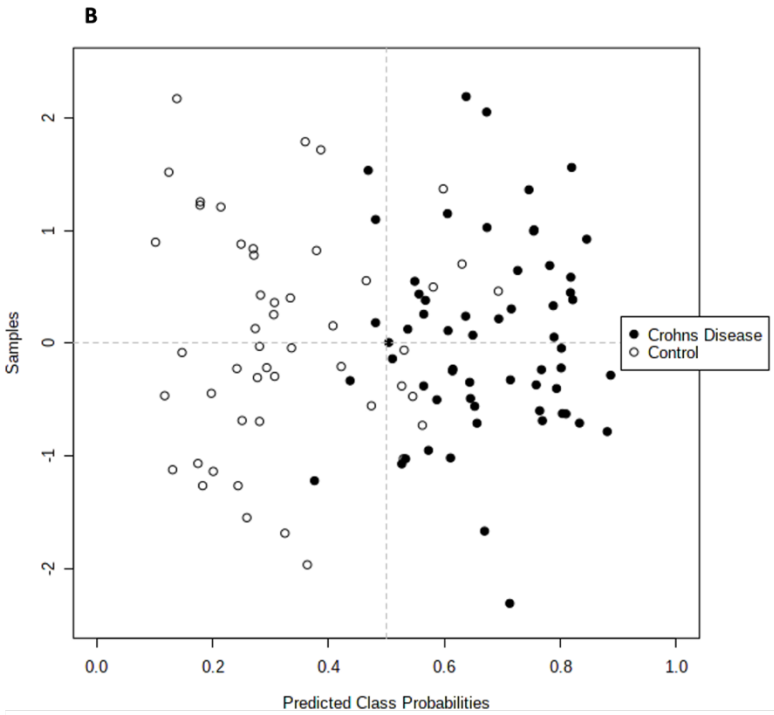
| | | Actual class: CD | Actual class: UC |
|---|---|---|---|
| Predicted class | CD | 51 | 14 |
| Predicted class | UC | 9 | 46 |

Figure 54: RFC performance of the LA-REIMS analysis of CD vs UC. (A) RFC method: ROC curve with AUC and 95% confidence interval (B) Predicted class probabilities plot utilising the best AUC classifier. UC represented by black dots, CD as white dots. (C) Confusion matrix of CD vs UC

#### 6.1.2.2.3. LA-REIMS – Crohn's Disease vs non-IBD control

The RFC machine learning algorithm was utilised to assess the performance of the high-throughput faecal LA-REIMS setup for CD vs non-IBD control study participants. RFC performance evaluation are shown in Figure 55, the ROC curve indicates an area under the curve (AUC) of 0.927 (95% confidence interval CI:  0.822 to 0.98), Figure 55A. The predicted class probability plot in Figure 55B demonstrates correct classification of CD and non-IBD control samples. LA-REIMS demonstrates a sensitivity of 76.7% and specificity of 81.6% in CD vs non-IBD control patients (Figure 55C).



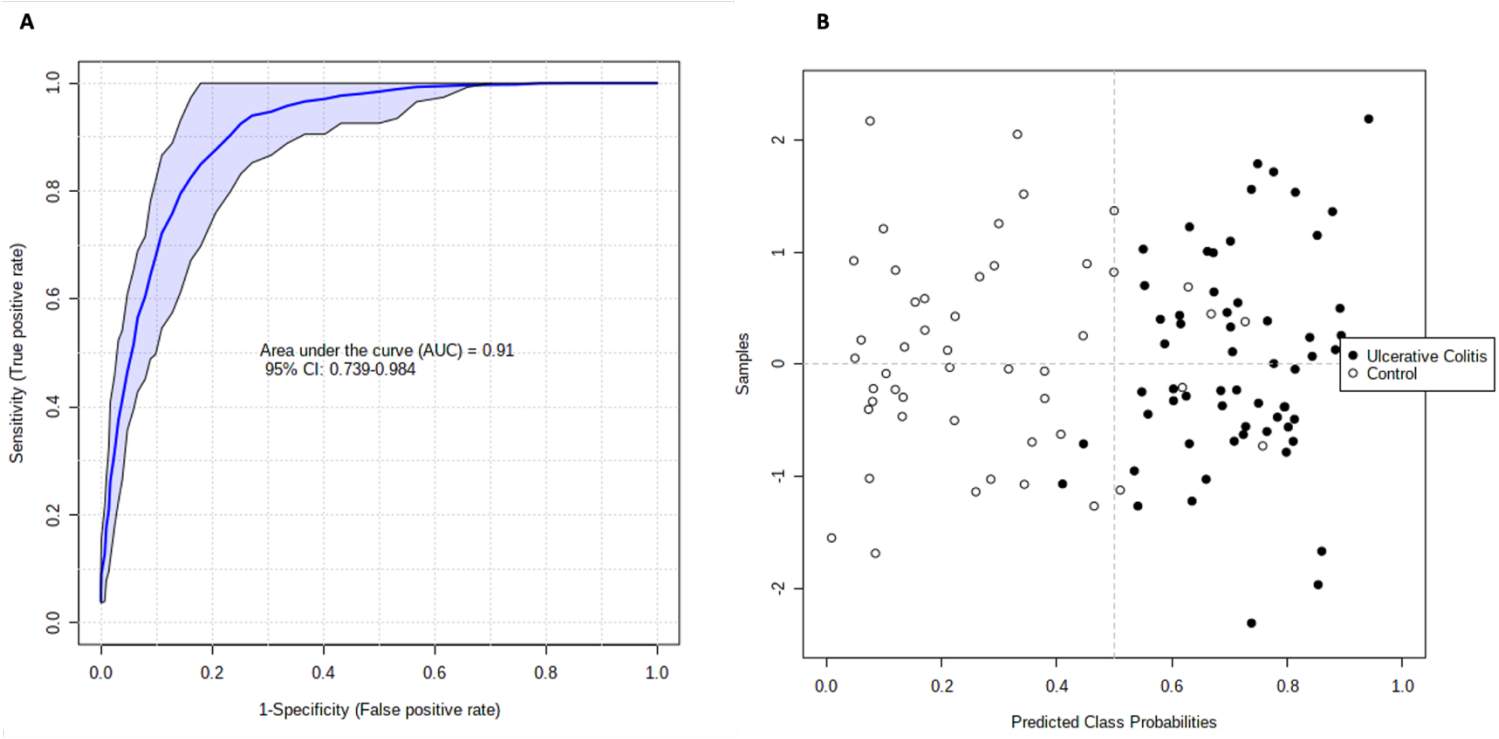

C

| Predicted class \ Actual class | CD | control |
|---|---|---|
| CD | 46 | 09 |
| control | 14 | 40 |

Figure 55: RFC performance of the LA-REIMS analysis of CD vs non-IBD control. (A) RFC method: ROC curve with AUC and 95% confidence interval (B) Predicted class probabilities plot utilising the best AUC classifier. CD represented by black dots, control as white dots. (C) Confusion matrix of CD vs non-IBD control

#### 6.1.2.2.4. LA-REIMS – Ulcerative Colitis vs non-IBD control

RFC classification performance of UC vs non-IBD control study participants are shown in Figure 56, the ROC curve indicates an AUC of 0.91 (95% confidence interval CI:0.739 to 0.984), Figure 56A. The predicted class probability plot in Figure 56B demonstrates correct classification of UC and non-IBD control samples. LA-REIMS demonstrates a sensitivity of 96.7% and specificity of 85.7% in UC vs non-IBD control patient (Figure 56C).




C

| Predicted class \ Actual class | UC | control |
|---|---|---|
| UC | 58 | 7 |
| control | 2 | 42 |

Figure 56: RFC performance of the LA-REIMS analysis of UC vs non-IBD control. (A) RFC method: ROC curve with AUC and 95% confidence interval (B) Predicted class probabilities plot utilising the best AUC classifier. UC represented by black dots, control as white dots. (C) Confusion matrix of UC vs non-IBD control

#### 6.1.2.2.5. Summary of LA-REIMS IBD diagnostic models

The diagnostic performance of optimised faecal LA-REIMS for IBD vs control, CD vs UC, CD vs control, and UC vs control RFC classification modelling is summarised in Table 19. The RFC model of UC vs control yielded the highest sensitivity and specificity, 96.7% and 85.7%, respectively. This high diagnostic accuracy was further validated by the separation of UC vs control study group in the PLS-DA model. Furthermore, in the second-best performing model, LA-REIMS achieved a sensitivity of 85.0% and a specificity of 76.7 % in the CD vs UC model. The worst performing model, CD vs control, yielded the lowest sensitivity of 76.7% and a specificity of 81.6%. LA-REIMS yielded a sensitivity of 77.8% and a specificity of 77.8% for the IBD versus non-IBD control model.

Overall, all models offer sensitivities and specificities between 75-97%. However, LA-REIMS RFC modelling demonstrates best performance in the classification of UC vs non-IBD control patients. The results of the LA-REIMS modelling are promising but need to be validated in a prospectively recruited independent data set. The results of the IBD cohort indicate the most promising application for LA-REIMS in the differentiation between UC and non-IBD control.

Table 19: Diagnostic accuracy of LA-REIMS for each model investigated

| LA-REIMS | Sensitivity | Specificity | AUC + (95% CI) |
|---|---|---|---|
| IBD vs control | 77.8 | 77.8 | 0.88 (0.736 - 0.985) |
| CD vs UC | 85.0 | 76.7 | 0.857 (0.762 - 0.936) |
| CD vs control | 76.7 | 81.6 | 0.927 (0.822 - 0.98) |
| UC vs control | 96.7 | 85.7 | 0.91 (0.739 - 0.984) |

### 6.1.2.3. Univariate Analysis

Univariate statistics were utilised to predict specific features significantly different between non-IBD control, CD, and UC. Details of the Volcano and ANOVA analysis and why they were chosen as the preferred data analysis tool are summarised in section *3.3.3.4. Data analysis – Metaboanalyst*.

Volcano plots were built to identify univariate features significantly different between CD vs non-IBD control (Figure 57), UC vs non-IBD control (Figure 58), and CD vs UC (Figure 59).
The volcano plot of CD (blue dots) vs non-IBD control (red dots), Figure 57, identified nine m/z features were more abundant in the CD group, whereas 14 *m/z* features were associated with non-IBD control, 66 *m/z* features were unsignificant.

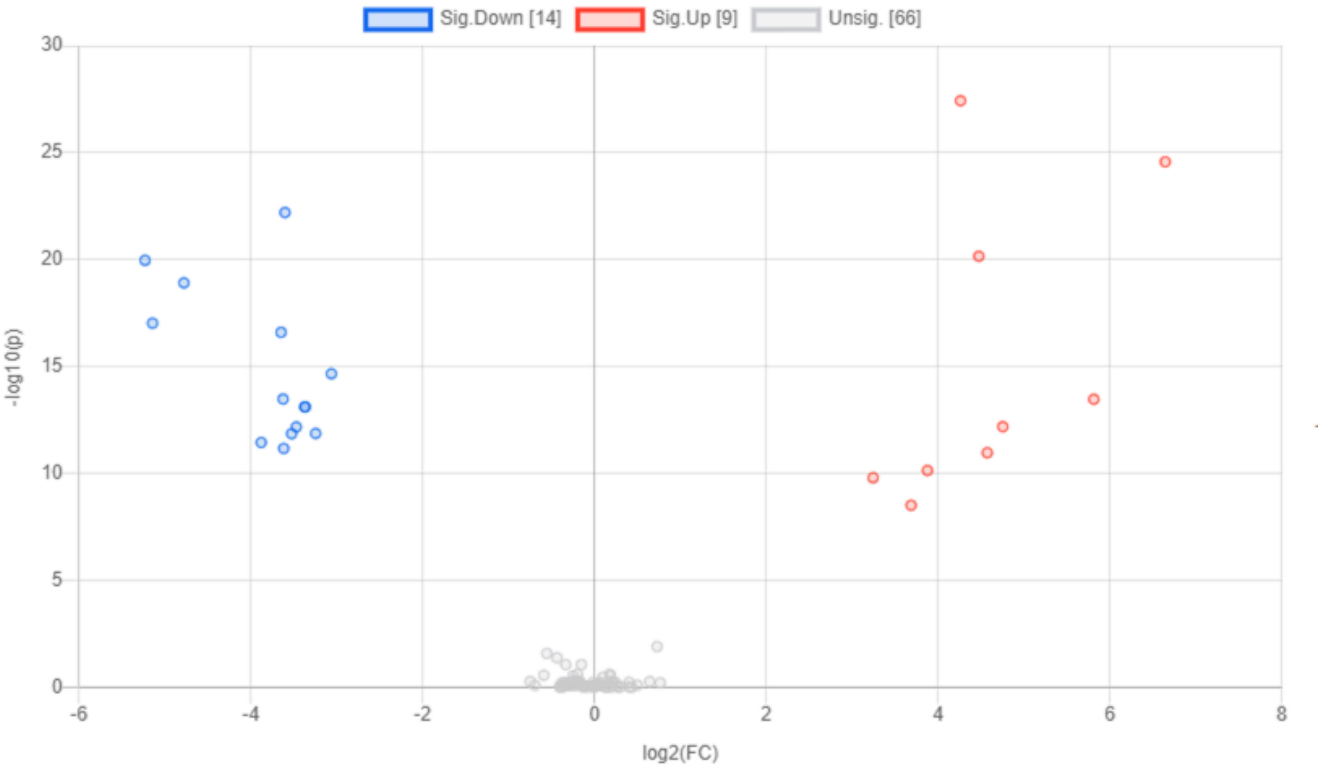



Figure 57: Volcano plot of CD (red dots) vs non-IBD control (blue dots). Fold-change threshold set to 2; FDR corrected *p* value threshold set to <0.05; group variance set to unequal.

The volcano plot of UC (red dots) vs non-cancer control (blue dots), Figure 58, identified five m/z features were more abundant in the UC group, 21 *m/z* features were more abundant in the non-IBD control, 59 *m/z* features were unsignificant.

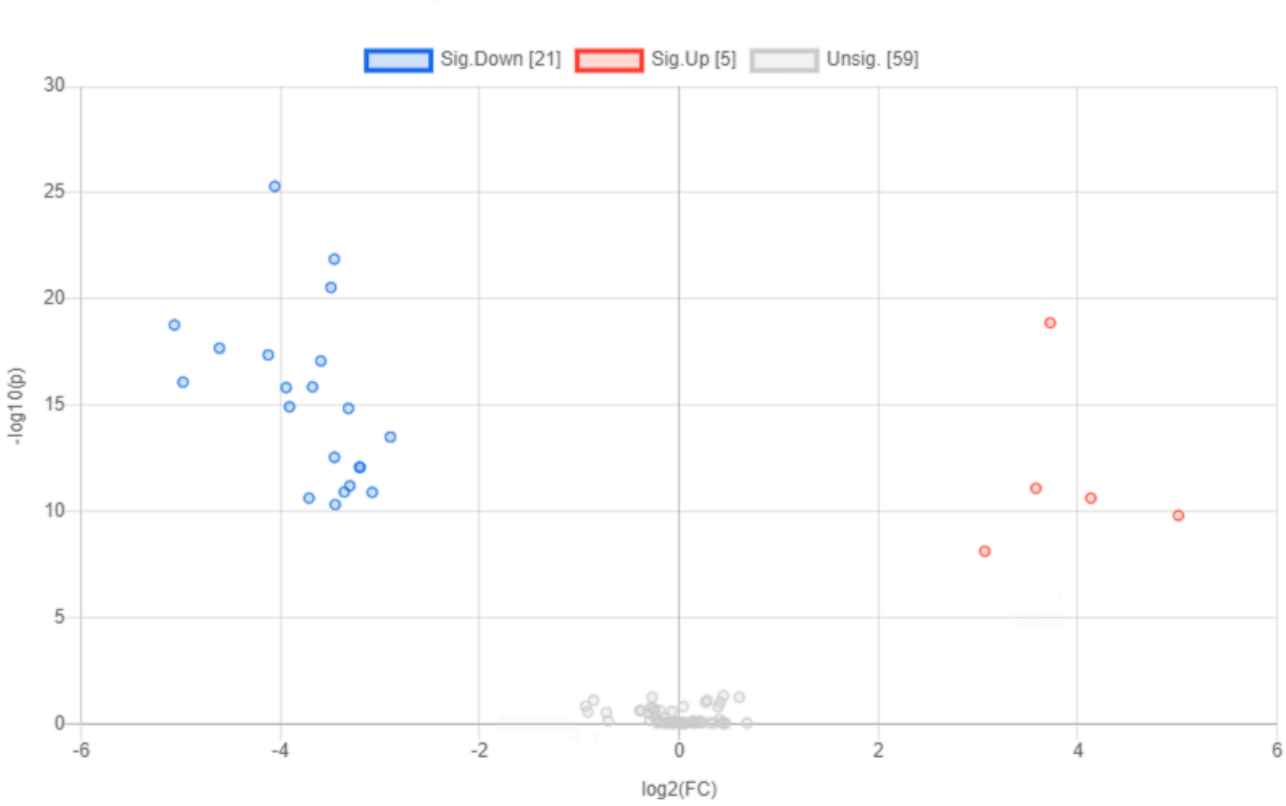



Figure 58: Volcano plot of UC vs non-IBD control. Volcano plot is a combination of fold change (FC) and t-test statistics. Fold-change threshold set to 2; FDR corrected *p* value threshold set to <0.05; group variance set to unequal.

The volcano plot of CD (blue dots) vs UC (red dots), Figure 59, identified 12 *m/z* features were more abundant in the CD group, four m/z features were more abundant in the UC group, 59 *m/z* features were unsignificant.

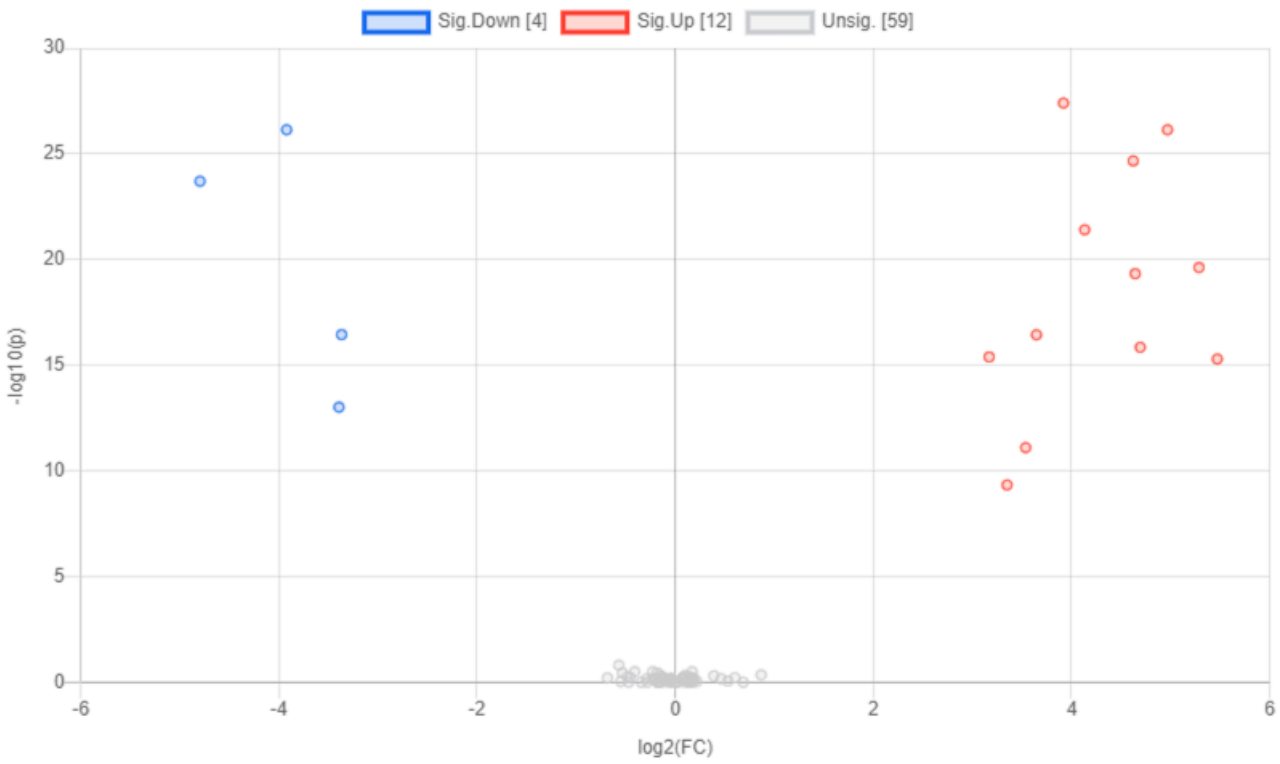



Figure 59: Volcano plot of CD vs UC. Volcano plot is a combination of fold change (FC) and t-test statistics. Fold-change threshold set to 2*;* FDR corrected p value threshold set to <0.05*;* *g*roup variance set to unequal.

Furthermore, ANOVA analysis, Figure 60, ANOVA non-parametric version (Kruskal Wallis Test) of CD vs UC vs con-IBD control, along with Fisher's Least Significant Difference post-hoc test was carried out. FDR corrected *p* values of less than 0.05 were accepted as significant.

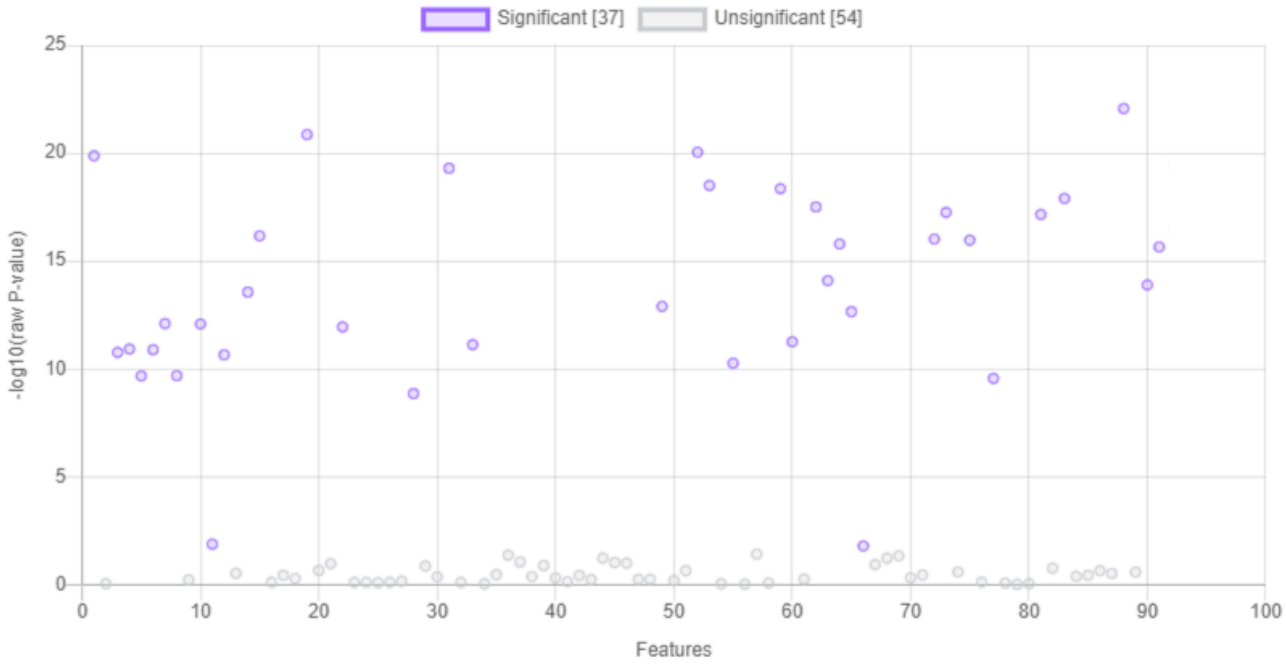



Figure 60: ANOVA non-parametric version (Kruskal Wallis Test) of CD vs UC vs non-IBD control, along with Fisher's Least Significant Difference post-hoc test - FDR corrected *p* value of < 0.05 were accepted as significant.

The univariate methods (ANOVA, Volcano plots, etc.) were utilised to generate a list of potentially biologically important features. Annotation of the *m/z* of interest was carried out based on their accurate mass and isotopic distribution, with tentative identification of metabolites performed, see Table 20. A step-by-step description of the metabolite identification process utilised is given in section *3.3.3.5 Metabolite Identification and 3.3.3.6 Tentative identification – validation.*

Tentative identification of the 17 *m/z* of interest is described in Table 20. Features at *m/z* 167.06, 325.184, 357.281, 549.428, 575.445, and 577.461 yielded no match on interrogation in literature databases. Four features, at *m/z* 195.049, 269.248, 279.233 and 339.327 yielded

multiple hits on interrogation in literature databases. Features at m/z 181.072, 313.237, 355.321, 533.455, 599.445, 601.46, and 603.476 were subsequently tentatively identified as hexose, octadecanedioic acid/*stearic acid,* 2-hydroxy-22:0 fatty acid, DG (31:1), DG (32:2), DG (32:1) chloride (CL) isotope 35, and a combination of DG (32:1) CL isotope 37 and DG (32:0) CL isotope 35, respectively. Interestingly, tentatively identified features hexose, DG (32:1), and DG (32:0) appeared as features of interest in both the IBD and the CRC datasets.

Table 20: Tentative identifications of *m/z* significantly different between either UC, CD, non-IBD control based on ANOVA and Fisher's Least Significant Difference post-hoc test and Volcano plots displaying fold change against FDR corrected *p* value from t tests (FDR corrected *p* value of < 0.05 were accepted as significant). Accurate masses of candidate m/z peaks, up to three decimal places after mass drift correction, were utilised for tentative identifications.

| Accurate m/z | Molecular formula of tentatively identified compound | Tentative ID | Adduct | Matched mass | ppm error |
|---|---|---|---|---|---|
| 167.06 | No Match | | | | |
| 181.072 | $C_6H_{12}O_6$ | Hexose | M-H | 181.0718 | 1.105 |
| 195.049 | $C_6H_{12}O_7$ | Galactonic acid, etc. (long list of possible matches) | M-H | 195.051 | -10.254 |
| 269.248 | $C_{17}H_{34}O_2$ | Heptadecanoic acid, etc. (long list of possible matches) | M-H | 269.2486 | -2.228 |
| 279.233 | $C_{18}H_{34}O_3$ | Ricinoleic acid, etc. (long list of possible matches) | M-H20-H | 279.2324 | 2.149 |
| 313.237 | $C_{18}H_{34}O_4$ | Octadecanedioic acid | M-H | 313.2384 | -4.469 |
| 325.184 | No Match | | | | |
| 339.327 | $C_{22}H_{44}O_2$ | Behenic acid, etc. (long list of possible matches) | M-H | 339.3269 | 0.295 |
| 355.321 | $C_{22}H_{44}O_3$ | 2-Hydroxy-22:0 fatty acid | M-H | 355.3218 | -2.252 |
| 357.281 | No Match | | | | |
| 533.455 | $C_{34}H_{64}O_5$ | Diacylglycerol (31:1) | M-H20-H | 533.457 | -3.749 |
| 549.428 | No Match | | | | |
| 575.445 | No Match | | | | |
| 577.461 | No Match | | | | |
| 599.445 | $C_{35}H_{64}O_5$ | Diacylglycerol (32:2) | M+Cl | 599.4448 | 0.334 |
| 601.46 | $C_{35}H_{66}O_5$ | Diacylglycerol (32:1) - chloride (CL) 35 | M+CL | 601.4604 | -0.665 |
| 603.476 | $C_{35}H_{66}O_5$ and $C_{35}H_{68}O_5$ | CL 37 Diacylglycerol (32:1) and CL 35 of Diacylglycerol (32:0) | M+CL | 603.4604 and 603.4761 | -0.166 |

Validation of tentative identifications of features of interest observed in LA-REIMS analysis was attempted using LC-MS analysis of pooled faecal sample extracts, described in detail in section *3.3.3.6 Tentative identification – validation.* The *m/z* of the tentatively identified features (up to three decimal places) from the LA-REIMS analysis of the CRC were checked for their presence in the LC-MS data. A list of the peaks identified in both is summarised in Table

21. Further, the m/z of interest, were interrogated at each retention time eluting. A mass accuracy of <5 ppm was utilised as a cut off for the cross-check of features present in LA-REIMS and LC-MS, Table 21.

Features that were present in both LA-REIMS and LC-MS mode at *m/z* 269.248, 279.233 and 339.327 yielded multiple hits on interrogation in literature databases. Feature at m/z 355.321 tentatively identified as 2-hydroxy-22:0 fatty acid was significantly increased in control vs CD and control vs UC (*p* value <0.001 for both; FDR corrected *p* value <0.001 for both), see Figure 61.

In future work, the analysis of tandem LA-REIMS and LC-MS/MS for tentatively identified features of interest will be carried out. Furthermore, for further validation of the metabolite identities, the extracted faecal sample will be analysed with the co-elution with standards using LC-MS and LA-REIMS to aid the assignment of more definitive structures to the observed *m/z.* While the previous section discussed features present in both the LA-REIMS and LC-MS analysis, it is also important to consider the features which were shown to be significantly different between UC, CD, and non-IBD control based on the untargeted LA-REIMS analysis, but which were not present in the LC-MS mode data. To add a level of proof in validation measurements and increase identification power, LA-REIMS tandem MS experiments will be carried out, where possible, to obtain the fragmentation pattern of features of interest.

Table 21: List of tentatively identified *m/z* (up to three decimal places) from the faecal LA-REIMS analysis of the IBD cohort present in both the LA-REIMS and LC-MS analysis

| Accurate m/z | Molecular formula of tentatively identified compound | Tentative ID | Adduct | Matched mass | ppm error |
|---|---|---|---|---|---|
| 269.248 | $C_{17}H_{34}O_2$ | Heptadecanoic acid, etc. (long list of possible matches) | M-H | 269.2486 | -2.228 |
| 279.233 | $C_{18}H_{34}O_3$ | Ricinoleic acid, etc. (long list of possible matches) | M-H20-H | 279.2324 | 2.149 |
| 339.327 | $C_{22}H_{44}O_2$ | Behenic acid, etc. (long list of possible matches) | M-H | 339.3269 | 0.295 |
| 355.321 | $C_{22}H_{44}O_3$ | 2-Hydroxy-22:0 fatty acid | M-H | 355.3218 | -2.252 |

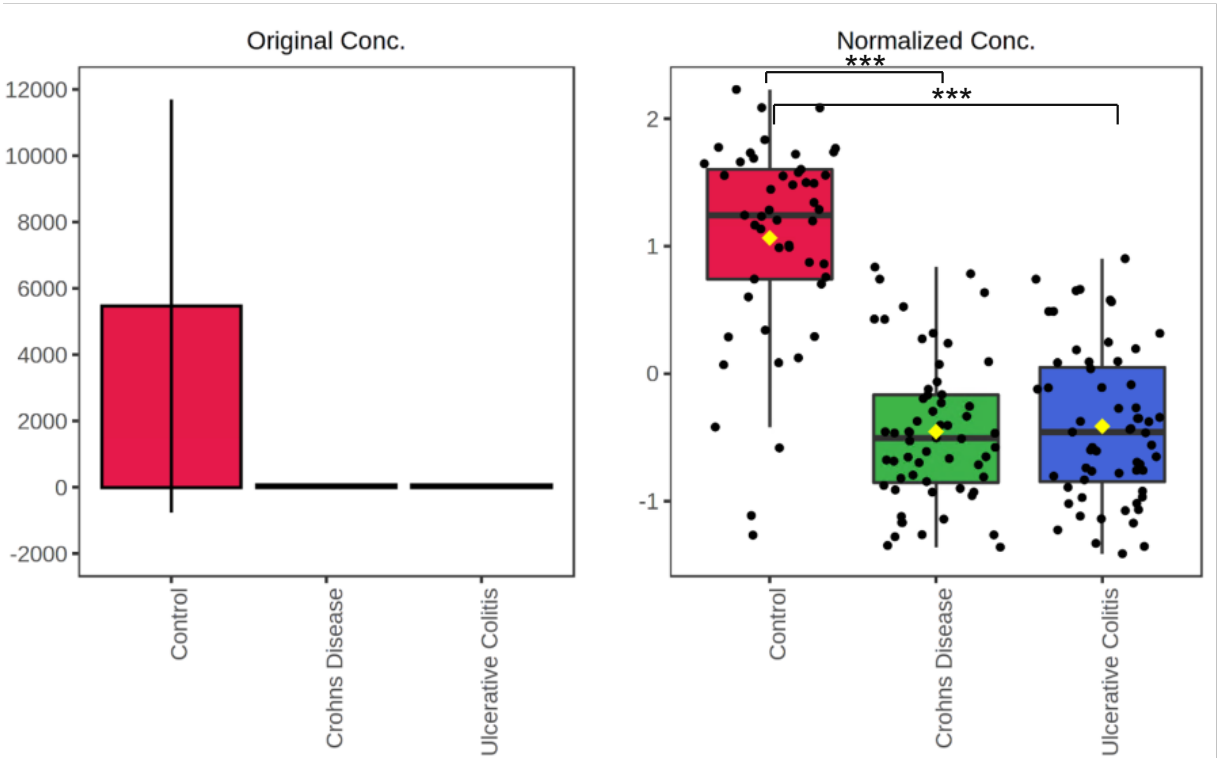



Figure 61: Box plot of m/z 355.321 tentatively identified as 2-hydroxy-22:0 fatty acid in LA-REIMS direct from faecal sample analysis determined using ANOVA (t test statistics) with FDR adjusted *p* value <0.05. * <0.05, ** <0.01, *** <0.001

### 6.1.3. Major patient-related modifiers of the LA-REIMS faecal metabolome in IBD

To establish variation in the study groups investigated, i.e., CD, UC, and non-IBD control, the baseline characteristics provided in Table 18 were assessed. Significant differences (<0.05) were found in terms of 5-aminosalicylic acid (5-ASA) medication intake between CD vs UC ($p$=0.017), see Table 18. Furthermore, significant differences (<0.05) between the study groups were also found in terms of biologic ($p$=0.004) and non-IBD medication ($p$<0.001). Significant differences between CD, UC, and non-IBD controls were found in terms of the total co-morbidities ($p$<0.001) see Table 18. Total comorbidities included diabetes mellitus, asthma, cardiovascular and inflammatory diseases.

The significant differences between the UC and CD study groups based on mesalazine (5-ASA), biologic and non-IBD medication intake is an important factor to consider when deducing conclusion from the results of this investigation, as the differences between the study groups may be influenced by the medications taken. The goals of drug treatment in IBD are the induction and maintenance of remission, mucosal healing, and avoiding surgical intervention and development of cancer (due to chronic inflammation) (370).

Mesalazine, also known as mesalamine or 5-ASA is an anti-inflammatory drug (371–373). There are very few studies looking at 5-ASA effects on the faecal microbiome and to the best of the candidate's knowledge none yet investigating the faecal metabolome. Xu et al (374) demonstrated a higher abundance of *Firmicutes* and lower levels of *Proteobacteria* in inflamed mucosae of untreated UC patients vs inflamed mucosae of 5-ASA treated UC patients (374). While this research demonstrates a potential direct impact of 5-ASA on the mucosal microbiome, further studies are needed to validate the results and investigate the faecal microbiome and the metabolome (374).

Significant differences (<0.05) between the study groups were also found in terms of biologic (p=0.004) medication. Biologics are drugs "that are derived partly or completely from living biological sources" (375), the most commonly used is TNF-α inhibitors such as adalimumab or infliximab, used both in UC and CD (376). Other biologics are integrin receptor antagonists (vedolizumab and natalizumab), and IL-12 and IL-23 antagonist, (ustekinumab) (376). There are very few studies investigating the direct effect of biologics on the faecal microbiome and to the best of the candidate's knowledge none yet investigating the faecal metabolome.

In a systematic review of 10 studies, Estevinho et al (377) demonstrated changes in faecal and colonic tissue microbiome in patients with IBD who responded to biologic agents. However, there were no differences in faecal and tissue microbiome among the various biologic agents assessed; all biologics produced decreases in relative abundances of *Escherichia* and *Enterococcus* and increases in genera that produce SCFAs (377). Overall, in the context of this thesis, it should be kept in mind that a significant difference in 5-ASA as well as biologics medication intake between the study groups exists, and that this may have affected the faecal metabolomic analysis.

To assess study population related modifiers of the faecal metabolome, the impact of variation in age, gender, ethnicity, and BMI on the LA-REIMS faecal metabolome in the studied patient cohort were assessed. Age was investigated in the age ranges 20-40, 41-50, 51-60, 61-70, and 71-85 years, Figure 62. Gender (f=female; m=male) confounding analysis was performed, Figure 63. Ethnicity was categorised as Caucasian, South Asian, and Black/Other, Figure 64. BMI was categorised as <18=underweight, 18.5-24.9=normal, 25.0-29.9=overweight, 30.0-34.9=obese, and >35.0 extremely obese, Figure 65.

In Figure 62, age was investigated in the PCA modelling, indicating no separation based on age groupings. Importantly, based on the baseline demographic characteristics table, Table 18, no significant differences were established between the study groups based on the age of participants involved.

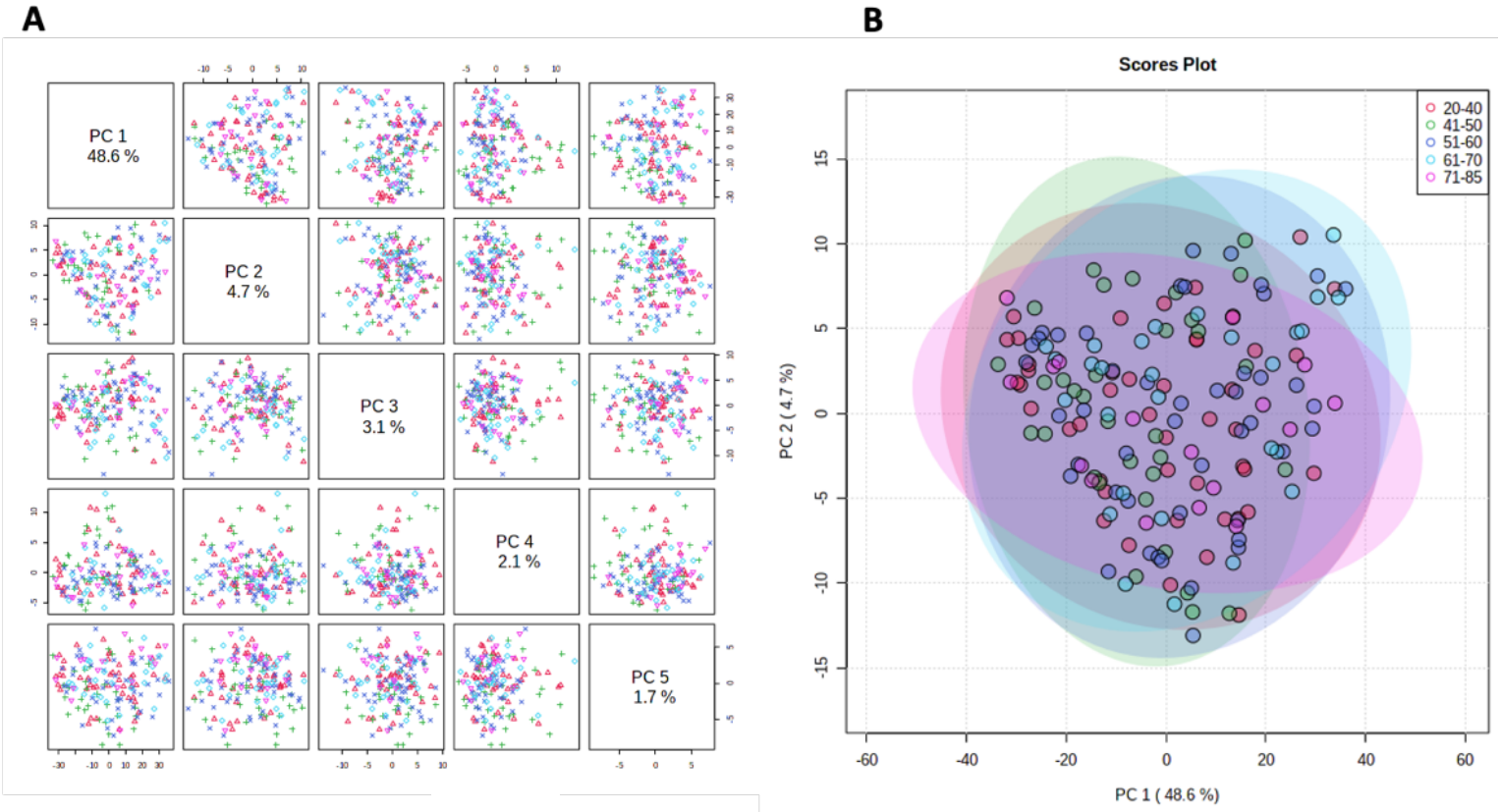


Figure 62: Age was performed in the age ranges 20-40, 41-50, 51-60, 61-70, and 71-85 years with A Pairwise PCA scores plot for the top five PCs in *m/z* range 50-1200 B Scores plot of PCA analysis PC 1 and PC2 in the *m/z* range 50−1200. Shaded region represents 95% confidence intervals.

In unsupervised analysis of the female and male participants in the IBD cohort, no separation based on gender was observed with a homogenous distribution of both groups, Figure 63 A+B. Based on the baseline demographic characteristics table, Table 18, no significant differences were established between the disease groups based on the number of male vs female participants involved.

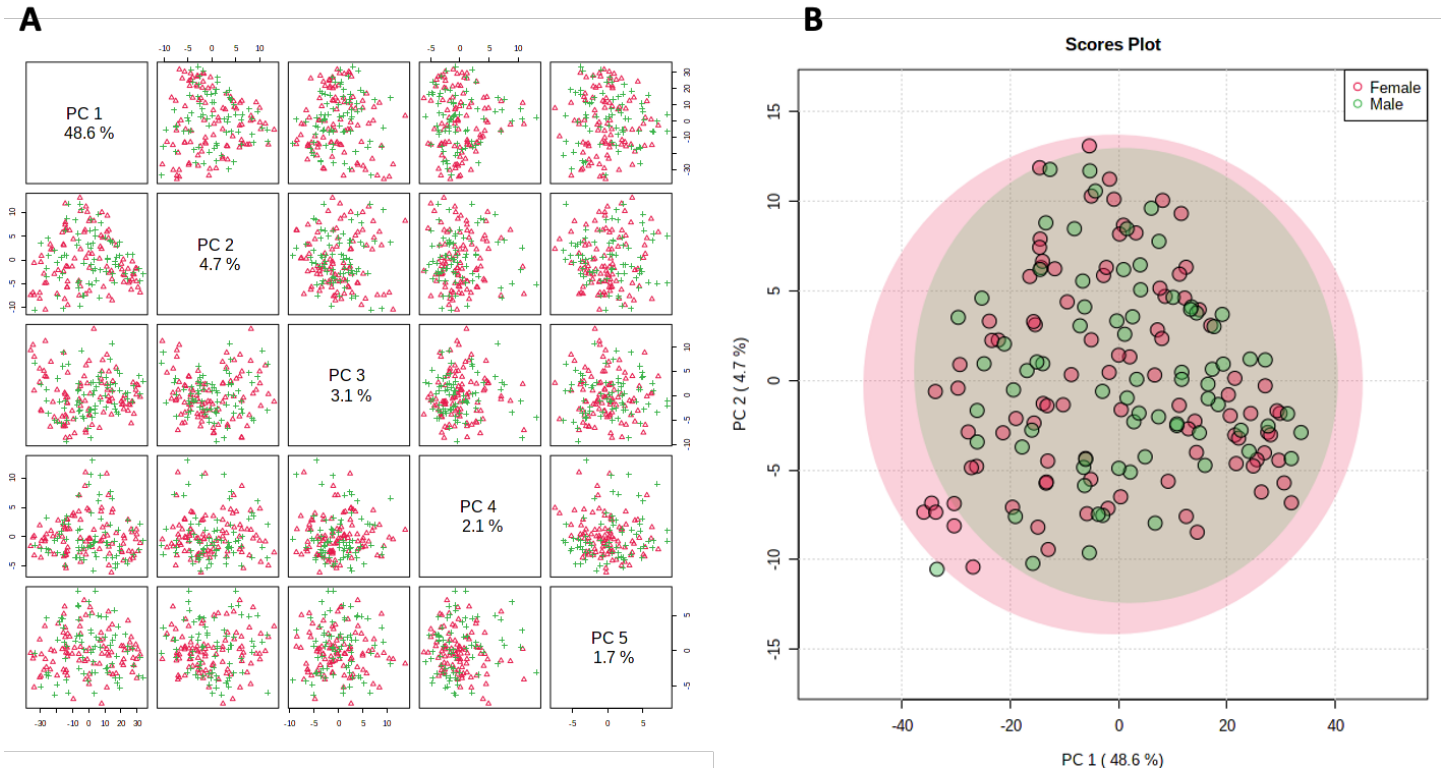



Figure 63: Gender (f=female; m=male) was performed using A Pairwise PCA scores plot for the top five PCs in *m/z* range 50-1200 B Scores plot of PCA analysis PC 1 and PC 2 in *m/z* range 50−1200. Shaded region represents 95% confidence intervals.

Ethnicity was not found to be significantly different between the participants present in each disease group investigated, baseline demographic characteristics Table 18. It is important to consider that most participants in the study cohort were Caucasians (n=87) and South Asians (n=77), with and Black/Other (n=5) underrepresented, Table 18. The PCA plot of ethnicity categories, Figure 64 A+B, shows Caucasian and South Asian clustering together: Black/Other clusters away slightly from Caucasian and South Asian.

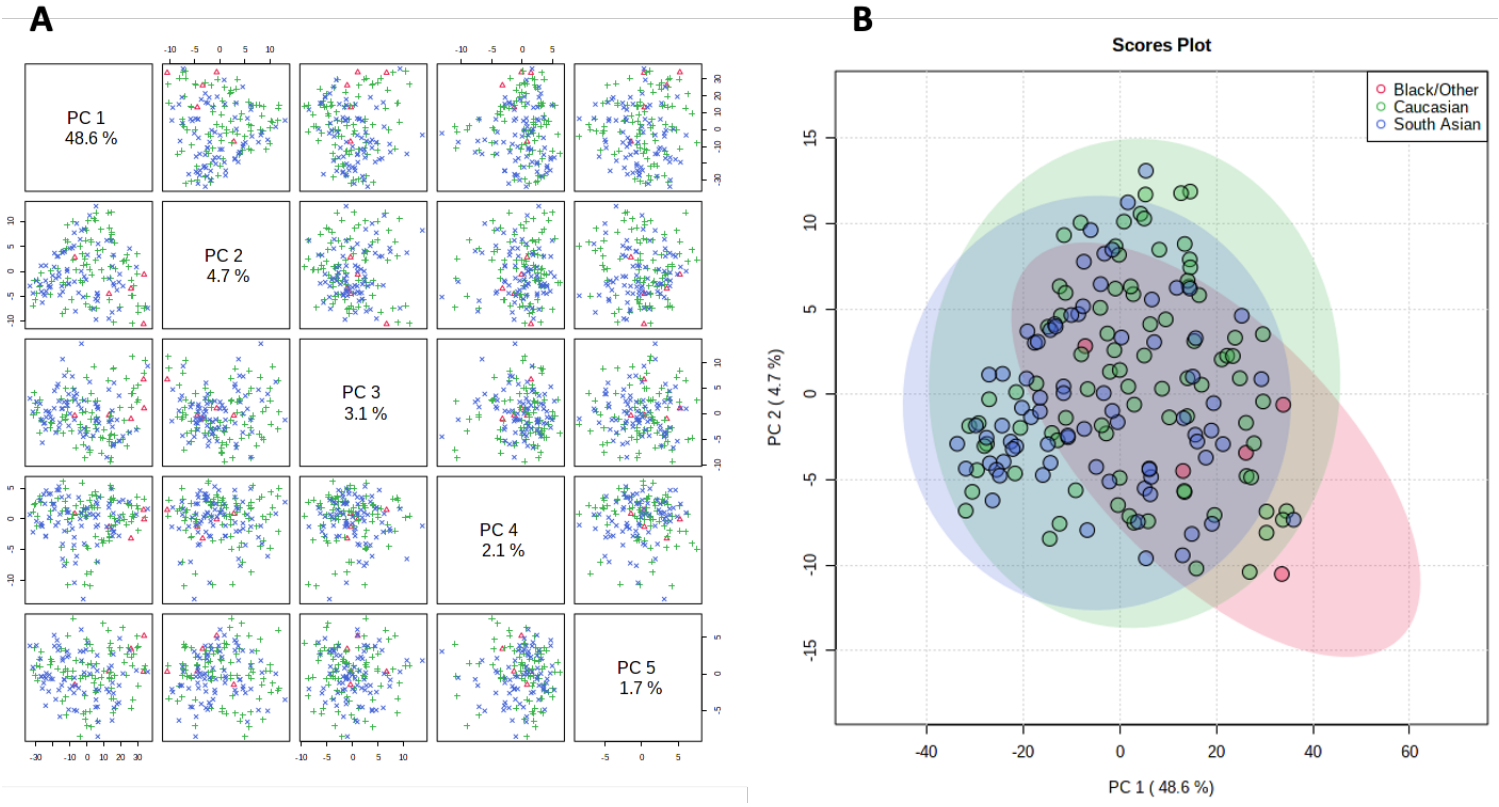



Figure 64: Ethnicity was categorised as Caucasian, South Asian, and Black/Other A Pairwise PCA scores plot for the top five PCs in *m/z* range 50-1200 B Scores plot of PCA analysis PC 1 and PC2 in *m/z* range 50−1200. Shaded region represents 95% confidence intervals.

In Figure 65 A+B, PCA modelling indicates that although homogenous in appearance, extremely obese patients cluster away from underweight patients. A possible explanation for

this finding may be differences in dietary patterns, with for instance extremely obese patients more likely to consume a high-fat diet compared to underweight patients (378). The baseline demographic characteristics data, Table 18, indicate no significant differences between the disease groups based on the BMI of participants involved.

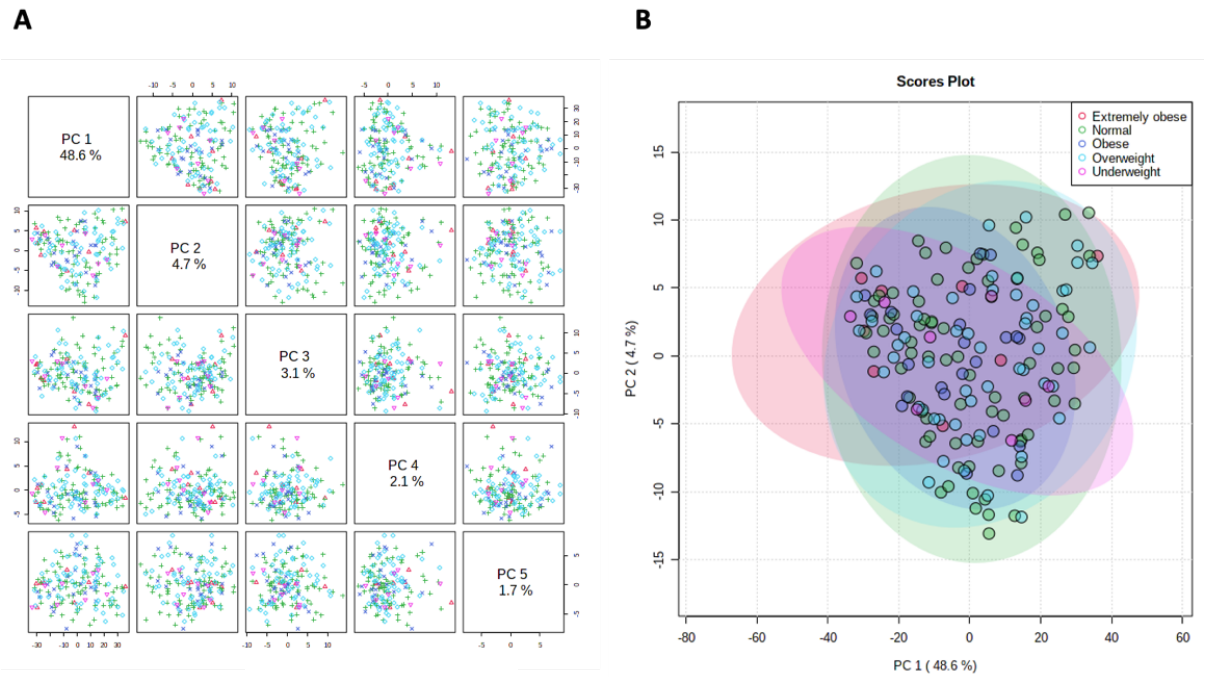



Figure 65: BMI was categorised as <18=underweight, 18.5-24.9=normal, 25.0-29.9=overweight, 30.0-34.9=obese, and >35.0 extremely obese A Pairwise PCA scores plot for the top five PCs in *m/z* range 50-1200 B Scores plot of PCA analysis PC 1 and PC2 in *m/z* range 50-1200. Shaded region represents 95% confidence intervals.

Overall, this analysis suggests that age and gender do not impact the LA-REIMS faecal metabolome substantially in this analysis. BMI and ethnicity indicate to play a role based on superficial multivariate analysis and needs to be considered when interpreting results of this work and more in-depth analysis of patient-related modifiers should be considered when planning future studies in this area.

### 6.1.4. Limitations

The IBD faecal samples investigated using LA-REIMS analysis underwent multiple freeze-thaw cycles and were stored at 4°C degrees for prolonged periods of time at the point of sample collection and transfer. It should be considered that changes in the metabolomic profile of samples may arise from many sources other than the intended study factor (33). Furthermore, the present dataset lacked detailed comprehensive dietary information, in the form of 24-hour or 72-hour dietary recall questionnaires, study participants were not controlled for diet or time of food ingestion.

FC analysis was not carried out for the IBD cohort. FC would have offered a useful comparison to LA-REIMS, as FC is sensitive marker for differentiating functional GI disease and IBD (286). Moreover, another limitation of this study is the potential confounding effect is stool consistency, that is BSC group variances between the study participants in each study group, which were not assessed at the point of recruitment. Furthermore, the study does not assess disease activity, that is active disease vs remission in CD and UC. In addition to FC, IBD scoring systems as well as endoscopic and histological activity should be assessed. However, due to the observational nature of the study design, this was limited in this work.

The faecal samples in this study were obtained from patients who had already been diagnosed with IBD and were treated with IBD medications. This limits the degree to which LA-REIMS can be assessed as a diagnostic tool for IBD. The faecal LA-REIMS based investigation of an IBD inception cohort would be favourable, as it would allow the assessment of the diagnostic test on a study cohort at the point of diagnosis (pre-treatment) (379). However, the recruitment of patients at this point is difficult as patients at the point of diagnosis often present with an active flare of IBD, necessitating immediate treatment with medications for symptomatic relief and to induce disease remission (274). This is further compounded by the established reluctance of patients' to provide faecal samples for research (380). A major limitation of this study is the lack of validation of results in a prospectively recruited independent data set. To further develop LA-REIMS into a diagnostic test, faecal samples from a patient-cohort at the point of diagnosis (inception cohort) and prior to

medication intake should be prospectively collected and analysed in a targeted manner to validate the findings of this work.

## 6.2. Conclusion and Future work

IBD, including CD and UC is a heterogenous disease group of the GI tract characterised by chronic inflammation often presenting with varying phenotype (237). Diagnosis of UC and CD is based on a complex investigation, including clinical findings, and endoscopic, radiological, histological and laboratory investigations (274).

LA-REIMS demonstrates potential high diagnostic accuracy for direct from faecal sample IBD detection. This has potential to optimise IBD diagnosis pathways and could reduce the burden of unnecessary endoscopic, radiological, histological and laboratory investigations. This is the first study, to the best of our knowledge, utilising optimised LA-REIMS faecal metabolomics analysis in IBD with model sensitivities and specificities between 75-97% demonstrating differences between UC vs CD, UC vs control, CD vs control, and IBD vs control study groups. Moreover, several tentatively identified features were shown to be significantly different between CD, UC, and control, including fatty acids, sugars, and diacylglycerol lipid species. Future work includes validation of the identified features of interest and a targeted study prospectively assessing the potential of faecal LA-REIMS in IBD diagnostics.

# 7. Discussion, Conclusion, and Future directions

## 7.1. Discussion

### 7.1.1. LA-REIMS for the analysis of faecal biospecimens in GI disease

This thesis lays the groundwork for the expansion of LA-REIMS applications to the study of the faecal metabolome in GI health and disease. The optimisation of laser and instrumentation parameters for high-throughput LA-REIMS demonstrate improved signal-to-noise ratios with parameters yielding the least % carry-over between samples for high-throughput faecal analysis.

Pilot work on testing the potential impact of water content on faecal sample analysis LA-REIMS spectra demonstrates separation in multivariate analysis in the m/z range 50-1200 between crude faecal samples and faecal sample with different ratios water added. The impact appears higher in the m/z range 50-600 than in the m/z range 600-1000. However, the study was limited to n=1 sample and future work will be utilised of a homogenised sample (n>1) to assess the effect of water content after lyophilisation. An indirect approach towards correction for variation in water content for faecal samples is being advocated for LA-REIMS analysis. Furthermore, for the first time, LA-REI-MSI for mapping of metabolites in whole fresh or frozen human faecal samples was implemented. Future studies will utilise a targeted metabolomics approach to investigate whole faecal samples from CRC and IBD patients investigating features of interest in each corresponding dataset.

LA-REIMS demonstrates potential for accurate direct from faecal sample adenoma and CRC detection. This has potential as a novel screening tool for early CRC detection and could reduce the burden of unnecessary colonoscopy and optimise cancer referral pathways. The implementation of a testing pipeline with both FIT and LA-REIMS analysis performed on the same faecal sample of individuals to be triaged for further testing, that is colonoscopy, may be considered. LA-REIMS demonstrates moderate-to-high diagnostic accuracy for direct from

faecal sample IBD detection. This has potential to optimise IBD diagnosis pathways and could reduce the burden of unnecessary endoscopic, radiological, histological and laboratory investigations. This is the first study, to the best of our knowledge, utilising LA-REIMS faecal metabolomics analysis in IBD with model sensitivities and specificities between 75-97% demonstrating differences between UC vs CD, UC vs control, CD vs control, and IBD vs control groups. Moreover, several tentatively identified features were identified to be significantly different between CD, UC, and control, including fatty acids, sugars, and diacylglycerol species. The CRC and IBD study carried out in this thesis are explorative in nature. A causal inference cannot be established between the features of interest and the time order of variables cannot be identified. Nevertheless, tentatively identified features of interest of the CRC and IBD study will be useful for directing future targeted studies.

### 7.1.2. Comparison with existing techniques

Current analytical techniques are limited in their applicability for studying the faecal metabolome. MS techniques such as LC-MS and GC-MS require time-intensive sample preparation (77). CE-MS and DI-MS demonstrate potential in their applicability in faecal metabolomics studies, but further research is needed to optimise and standardise the techniques before biological interpretation can be attempted. Additionally, NMR-based techniques lack sensitivity (52). Other ambient MS techniques, including DESI (121), DART (122) have not been utilised for the analysis of faecal metabolomics. In comparison, LA-REIMS has been previously applied to faecal metabolomics analysis (151) and presents a direct from sample high-throughput analysis.

A need for the development of improved early detection biomarker assays for CRC and IBD exists. IN CRC, FIT, the current screening modality in the BCSR in the UK, demonstrates high diagnostic accuracy for the detection of colorectal cancer but demonstrates a low to moderate sensitivity for adenomas (381). In IBD, diagnosis is based on a combination of clinical, laboratory, radiological, endoscopic, and histological criteria; there is no single test for diagnosing, monitoring, or evaluating treatment in IBD (264).

### 7.1.3. Strengths and limitations

LA-REIMS has not been developed as a biomarker discovery tool but has applications as a translational analytical technique. It has the potential to deliver the benefits of faecal metabolomics research to the clinical environment. Other analytical techniques, such as LC-MS, have greater potential as biomarker identifications tools and have been utilized here for validation purposes of tentative metabolite identifications using initial LA-REIMS analysis. The major advantage and benefit of REIMS is the near-real-time analysis of samples, at atmospheric pressure, with minimal sample preparation. To this end, this work aimed to design and optimise an analytical workflow that could be used in a high-throughput manner to standardise the technology for faecal sample analysis.

LA-REIMS untargeted metabolomics is limited in that only relative abundances/semi-quantification is achievable in general, unless isotopically labelled standard is utilised. Absolute concentrations of metabolites of interest could not be assessed. Due to a lack of chromatographic separation, LA-REIMS can demonstrate lower MS sensitivity, signal reproducibility, and higher matrix interference, compared to the more traditionally used LC-, or GC-based methods (382). LA-REIMS metabolite identification is typically based on accurate mass measurement, which limits identification accuracy (120). Tandem MS is typically utilised for metabolite identification. However, lack of chromatographic separation, will complicate tandem MS spectra interpretation (120).

LA-REIMS is affected by environmental conditions including temperature, humidity, and air pressure, leading to potential batch effects in data. Furthermore, ion suppression in LA-REIMS is a common problem. LA-REIMS may therefore be more suited as a fingerprinting technique for differentiation between healthy and disease groups or should be validated using more traditional LC-MS or GC-MS based techniques (120).

Further limitation should be considered with regards to the context of the analytical approach, rather than the instrument itself. A major consideration is the collection of faecal samples for LA-REIMS analysis, this includes both the complexity of obtaining samples for research purposes but also the post-evacuation metabolite changes associated with storage

at room temperature and an oxygen rich environment, as previously discussed in section *1.8. Challenges in faecal metabolomics.* For LA-REIMS to be used as a diagnostic tool for CRC and IBD analysis, a patient-friendly pathway for obtaining samples needs to be established.. Furthermore, the time from sample evacuation to the point of analysis needs to be kept minimal. This is not always feasible due to the difficulty associated with obtaining faecal samples, and thus presents a major limitation of this work. Here, where possible, time until analysis was limited and samples were stored at -80°C. A further limitation of faecal LA-REIMS analysis is the effect of dietary intake and other confounding factors such as patient/study participant differences in age, ethnicity, BMI, and medication use on the metabolome. Detailed comprehensive dietary information, in the form of 24-hour or 72-hour dietary recall questionnaires are needed. A further limitation is the lack of knowledge as to whether the tentatively identified metabolites of interest originate from exogenous factors, the host, the gut microbiome, or are the product of their co-metabolism. Further studies are needed to gain a better mechanistic insight into the role of features of interest in the pathogenesis of CRC and IBD.

However, the platform remains in developmental phase and more research is needed to establish the potential of the technology and develop LA-REIMS into a translational diagnostic technique. Furthermore, LA-REIMS allows for high-throughput direct from faecal sample analysis and results are obtained in near real time. Furthermore, the platform is semi-automated and allows samples to be assessed with minimal user intervention. LA-REIMS allows the analysis of 96 samples in approximately 3 hours for both negative and positive ion detection modes. This offers a viable alternative to existing metabolomics approaches and potential for use a screening platform for GI disorders.

## 7.2. Conclusions and Future directions

This works served as a first step towards the high-throughput LA-REIMS analysis of faecal metabolites in GI disease and has involved method development and optimisation of LA-REIMS for a novel sample type and disease. LA-REIMS has potential as a rapid, preparation-free method for the analysis of faecal samples and a screening and diagnostic technique for GI disorders. Validation of the tentatively identified biomarkers in the IBD and CRC cohort

through a prospectively recruited targeted study for detection of early colorectal cancer and alternatively the detection of IBD would enable more conclusive evidence as to their potential for CRC screening or IBD diagnosis. To gain a better mechanistic insight into the role of features of interest in the pathogenesis of CRC and IBD, *in vivo* and *in vitro* experiments will be carried out to investigate cause-effect relationships. The goal of future work is to study the pathways driving the interactions between diet, microbiota, metabolome, and the immune regulatory system and highlighting their roles in the adenoma-carcinoma sequence as well as the pathogenesis in IBD. Furthermore, a faecal LA-REIMS analysis of IBD and CRC datasets in congruence is being proposed. Future work will involve the testing of faecal LA-REIMS as a diagnostic tool for CRC vs IBD prediction. A prospective pilot study with a study population including both CRC and IBD patients should be considered. An important consideration of a non-invasive diagnostic tool for a GI disease is not only its ability to differentiate disease from non-disease but also the ability to pick up the GI pathology of interest from others with similar clinical presentations. For the translation of LA-REIMS from bench to bedside, its potential as a diagnostic technique for IBD and CRC needs to be studied whilst evaluating other GI conditions. IBD and CRC may present with similar clinical presentations and an accurate non-invasive diagnostic technique for differentiating between the two may allow for earlier diagnosis and therapeutic intervention, with potential to improve patient care.

# Bibliography


1. Zierer J, Jackson MA, Kastenmüller G, Mangino M, Long T, Telenti A, et al. The fecal metabolome as a functional readout of the gut microbiome. *Nature Genetics*. [Online] Springer US; 2018;50(6): 790–795. Available from: doi:10.1038/s41588-018-0135-7
2. Clish CB. Metabolomics: an emerging but powerful tool for precision medicine. *Molecular Case Studies*. [Online] 2015;1(1): a000588. Available from: doi:10.1101/mcs.a000588
3. McGuire AL, Gabriel S, Tishkoff SA, Wonkam A, Chakravarti A, Furlong EEM, et al. The road ahead in genetics and genomics. *Nature Reviews Genetics*. [Online] Springer US; 2020;21(10): 581–596. Available from: doi:10.1038/s41576-020-0272-6
4. Kaushika AK, DeBerardinis RJ. Applications of Metabolomics to Study Cancer Metabolism. *Physiology & behavior*. [Online] 2017;176(12): 139–148. Available from: doi:10.1016/j.bbcan.2018.04.009.Applications
5. Newgard CB. Metabolomics and Metabolic Diseases: Where Do We Stand? *Cell Metabolism*. [Online] Elsevier Inc.; 2017;25(1): 43–56. Available from: doi:10.1016/j.cmet.2016.09.018
6. Cantafio MEG, Grillone K, Caracciolo D, Scionti F, Arbitrio M, Barbieri V, et al. From single level analysis to multi-omics integrative approaches: A powerful strategy towards the precision oncology. *High-Throughput*. [Online] 2018;7(4): 1–20. Available from: doi:10.3390/ht7040033
7. Yu L, Li K, Zhang X. Next-generation metabolomics in lung cancer diagnosis, treatment and precision medicine: Mini review. *Oncotarget*. [Online] 2017;8(70): 115774–115786. Available from: doi:10.18632/oncotarget.22404
8. Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet*. [Online] 2014;15(1): 34–48. Available from: doi:10.1038/nrg3575
9. Vailati-Riboni M, Palombo V, Loor JJ. No Title. *Periparturient Diseases of Dairy Cows*. [Online] Springer; 2017. Available from: doi:https://doi.org/10.1007/978-3-319-43033-1_1
10. Zhang A, Wanying W. *Mass Spectrometry-Based Metabolomics in Clinical and Herbal Medicines*. Wiley; 2021.
11. Richard D. Beger, Warwick Dunn, Michael A. Schmidt, Steven S. Gross, Jennifer A. Kirwan, Marta Cascante, Lorraine Brennan, David S. Wishart, Matej Oresic, Thomas Hankemeier, David I. Broadhurst, Andrew N. Lane, Karsten Suhre, Gabi Kastenmüller, Susan J. S OF& RK-D. Metabolomics enables precision medicine: "A White Paper, Community Perspective". *Metabolomics*. 2016;12(146).
12. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biology*. [Online] Genome Biology; 2017;18(1): 1–15. Available from: doi:10.1186/s13059-017-1215-1
13. Kim Y, Maruvada P, Milner John. Metabolomics in biomarker discovery: future uses for cancer prevention. *Future Medicine*. [Online] 2008;4(1). Available from: doi:https://doi.org/10.2217/14796694.4.1.93
14. JOhnson CH, Gonzalez FJ. Challenges and Opportunities in Metabolomics. *Physiology & behavior*. [Online] 2016;176(1): 139–148. Available from:

doi:10.1016/j.physbeh.2017.03.040
15. Tennant DA, Durán R V., Gottlieb E. Targeting metabolic transformation for cancer therapy. *Nature Reviews Cancer*. [Online] Nature Publishing Group; 2010;10(4): 267–277. Available from: doi:10.1038/nrc2817
16. Wang JH, Byun J, Pennathur S. Analytical approaches to metabolomics and applications to systems biology. *Seminars in Nephrology*. [Online] 2010;30(5): 500–511. Available from: doi:10.1016/j.semnephrol.2010.07.007
17. Auffray C, Chen Z, Hood L. Systems medicine: The future of medical genomics and healthcare. *Genome Medicine*. [Online] 2009;1(1): 1–11. Available from: doi:10.1186/gm2
18. Furman D, Campisi J, Verdin E, Carrera-Bastos P, Targ S, Franceschi C, et al. Chronic inflammation in the etiology of disease across the life span. *Nature Medicine*. [Online] Springer US; 2019;25(12): 1822–1832. Available from: doi:10.1038/s41591-019-0675-0
19. Ahn AC, Tewari M, Poon CS, Phillips RS. The clinical applications of a systems approach. *PLoS Medicine*. [Online] 2006;3(7): 0956–0960. Available from: doi:10.1371/journal.pmed.0030209
20. Ahn AC, Tewari M, Poon CS, Phillips RS. The limits of reductionism in medicine: Could systems biology offer an alternative? *PLoS Medicine*. [Online] 2006;3(6): 0709–0713. Available from: doi:10.1371/journal.pmed.0030208
21. Karakitsou E, Foguet C, de Atauri P, Kultima K, Khoonsari PE, Martins dos Santos VAP, et al. Metabolomics in systems medicine: an overview of methods and applications. *Current Opinion in Systems Biology*. [Online] Elsevier Ltd; 2019;15(March): 91–99. Available from: doi:10.1016/j.coisb.2019.03.009
22. Karakitsou E, Foguet C, de Atauri P, Kultima K, Khoonsari PE, Martins dos Santos VAP, et al. Metabolomics in systems medicine: an overview of methods and applications. *Current Opinion in Systems Biology*. [Online] Elsevier Ltd; 2019;15: 91–99. Available from: doi:10.1016/j.coisb.2019.03.009
23. Azad R, Shulaev V. Metabolomics technology and bioinformatics for precision medicine. *Briefings in Bioinformatics*. [Online] 2018; Available from: doi:https://doi.org/10.1093/bib/bbx170
24. Alonso A, Marsal S, Julià A. Analytical methods in untargeted metabolomics: State of the art in 2015. *Frontiers in Bioengineering and Biotechnology*. [Online] 2015;3(MAR): 1–20. Available from: doi:10.3389/fbioe.2015.00023
25. Wiedmer SK, Hyötyläinen T. Selection of Analytical Methodology for Metabolomics. *RSC Chromatography Monographs*. [Online] 2013. p. 1–10. Available from: doi:10.1039/9781849737272-00001
26. Clarke CJ, Haselden JN. Metabolic Profiling as a Tool for Understanding Mechanisms of Toxicity. *Toxicologic Pathology*. [Online] 2008;36(1): 140–147. Available from: doi:10.1177/0192623307310947
27. Deo RC, Hunter L, Lewis GD, Pare G, Vasan RS, Chasman D, et al. Interpreting metabolomic profiles using unbiased pathway models. *PLoS Computational Biology*. [Online] 2010;6(2). Available from: doi:10.1371/journal.pcbi.1000692
28. Kell DB, Brown M, Davey HM, Dunn WB, Spasic I, Oliver SG. Metabolic footprinting and systems biology: The medium is the message. *Nature Reviews Microbiology*. [Online] 2005;3(7): 557–565. Available from: doi:10.1038/nrmicro1177
29. Vivanco F, Barderas MG, Laborde CM, Posada M, De La Cuesta F, Zubiri I, et al.

Metabolomic profiling for identification of novel potential biomarkers in cardiovascular diseases. *Journal of Biomedicine and Biotechnology*. [Online] 2011;2011(May 2014). Available from: doi:10.1155/2011/790132
30. Zhao Y-Y, Cheng X-L, Lin R-C. Chapter One - Lipidomics Applications for Discovering Biomarkers of DIseases in Clincal Chemistry. *International Review of Cell and Molecular Biology*. [Online] 2014;313: 1–26. Available from: doi:https://doi.org/10.1016/B978-0-12-800177-6.00001-3
31. Yang K, Han X. Lipidomics: Techniques, Applications, and Outcomes Related to Biomedical Sciences. *Trends in Biochemical Sciences*. [Online] 2016;41(11): 954–969. Available from: doi:10.1016/j.tibs.2016.08.010
32. Hyötyläinen T, Orešič M. Analytical Lipidomics in Metabolic and Clinical Research. *Trends in Endocrinology and Metabolism*. [Online] 2015;26(12): 671–673. Available from: doi:10.1016/j.tem.2015.08.006
33. Karu N, Deng L, Slae M, Guo AC, Sajed T, Huynh H, et al. A review on human fecal metabolomics: Methods, applications and the human fecal metabolome database. *Analytica Chimica Acta*. [Online] Elsevier Ltd; 2018;1030: 1–24. Available from: doi:10.1016/j.aca.2018.05.031
34. Wei Y. Efficacy of fecal sampling asa gut proxy in the study of chicken microbiota. *biorxiv*. [Online] 2018; Available from: doi:https://doi.org/10.1101/313577
35. Lavelle A, Lennon G, O'Sullivan O, Docherty N, Balfe A, Maguire A, et al. Spatial variation of the colonic microbiota in patients with ulcerative colitis and control volunteers. *Gut*. [Online] 2015;64(10): 1553–1561. Available from: doi:10.1136/gutjnl-2014-307873
36. Stearns JC, Lynch MDJ, Senadheera DB, Tenenbaum HC, Goldberg MB, Cvitkovitch DG, et al. Bacterial biogeography of the human digestive tract. *Scientific Reports*. [Online] 2011;1: 1–9. Available from: doi:10.1038/srep00170
37. Rose C, Parker A, Jefferson B, Cartmell E. The Characterization of Feces and Urine: A Review of the Literature to Inform Advanced Treatment Technology. *Critical reviews in environmental science and technology*. [Online] Taylor & Francis; 2015;45(17): 1827–1879. Available from: doi:10.1080/10643389.2014.1000761 [Accessed: 18th May 2018]
38. Loktionov A. Cell exfoliation in the human colon: Myth, reality and implications for colorectal cancer screening. *International Journal of Cancer*. [Online] 2007;120(11). Available from: doi:https://doi.org/10.1002/ijc.22647
39. Gregory KE, Bird SS, Gross VS, Marur VR, Lazarev A V, Walker WA, et al. Method development for fecal lipidomics profiling. *Analytical chemistry*. [Online] NIH Public Access; 2013;85(2): 1114–1123. Available from: doi:10.1021/ac303011k [Accessed: 18th May 2018]
40. O'Hara AM, Shanahan F. The gut flora as a forgotten organ. *EMBO reports*. [Online] 2006;7(7): 688–693. Available from: doi:10.1038/sj.embor.7400731
41. Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. *Nature Reviews Microbiology*. [Online] Springer US; 2021;19(1): 55–71. Available from: doi:10.1038/s41579-020-0433-9
42. Wu HJ, Wu E. The role of gut microbiota in immune homeostasis and autoimmunity. *Gut Microbes*. [Online] 2012;3(1): 4–14. Available from: doi:10.4161/gmic.19320
43. Lynch S V., Pedersen O. The Human Intestinal Microbiome in Health and Disease. *New England Journal of Medicine*. [Online] 2016;375(24): 2369–2379. Available from:

doi:10.1056/nejmra1600266
44. Putignani L, Del Chierico F, Vernocchi P, Cicala M, Cucchiara S, Dallapiccola B. Gut Microbiota Dysbiosis as Risk and Premorbid Factors of IBD and IBS Along the Childhood-Adulthood Transition. *Inflammatory Bowel Diseases*. [Online] 2016;22(2): 487–504. Available from: doi:10.1097/MIB.0000000000000602
45. Clarke G, Stilling RM, Kennedy PJ, Stanton C, Cryan JF, Dinan TG. Minireview: Gut Microbiota: The Neglected Endocrine Organ. *Molecular Endocrinology*. [Online] 2014;28(8): 1221–1238. Available from: doi:10.1210/me.2014-1108
46. Zierer J, Jackson MA, Kastenmüller G, Mangino M, Long T, Telenti A, et al. The fecal metabolome as a functional readout of the gut microbiome. *Nature Genetics*. [Online] 2018;50(6): 790–795. Available from: doi:10.1038/s41588-018-0135-7
47. Cubiella J, Clos-Garcia M, Alonso C, Martinez-Arranz I, Perez-Cormenzana M, Barrenetxea Z, et al. Targeted UPLC-MS Metabolic Analysis of Human Faeces Reveals Novel Low-Invasive Candidate Markers for Colorectal Cancer. *Cancers*. [Online] 2018;10(9): 300. Available from: doi:10.3390/cancers10090300
48. Marcobal A, Kashyap PC, Nelson TA, Aronov PA, Donia MS, Spormann A, et al. A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice. *ISME Journal*. [Online] Nature Publishing Group; 2013;7(10): 1933–1943. Available from: doi:10.1038/ismej.2013.89
49. Nannini G, Meoni G, Amedei A, Tenori L. Metabolomics profile in gastrointestinal cancers: Update and future perspectives. 2020;9327(20).
50. Probert C, Jones P, Ratcliffe N. A novel method for rapidly diagnosing the causes of diarrhoea. *Gut*. 2004;53: 58–61.
51. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*. [Online] 1999;29(11): 1181–1189. Available from: doi:10.1080/004982599238047
52. Emwas A-H. The Strengths and Weaknesses of NMR Spectroscopy and Mass Spectrometry with Particular Focus on Metabolomics Research. *Metabolomics : Methods and Protocols*. [Online] 2015. p. 1064–3745. Available from: doi:10.1007/978-1-4939-2377-9_13
53. Keun H, Athersuch T. Nuclear Magnetic Resonance (NMR)-Based Metabolomics. *Methods in Molecular Biology*. 2010. p. 321–334.
54. Hye KK, Sarantos K, Young CH. NMR analysis of faecal samples. *Clinical Metabolomics*. 2018;1730.
55. Keun HC, Ebbels TMD, Antti H, Bollard ME, Beckonert O, Schlotterbeck G, et al. Analytical reproducibility in 1H NMR-based metabonomic urinalysis. *Chemical Research in Toxicology*. [Online] 2002;15(11): 1380–1386. Available from: doi:10.1021/tx0255774
56. Dumas ME, Maibaum EC, Teague C, Ueshima H, Zhou B, Lindon JC, et al. Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: The INTERMAP study. *Analytical Chemistry*. [Online] 2006;78(7): 2199–2208. Available from: doi:10.1021/ac0517085
57. Kaddurah-Daouk R, Kristal BS, Weinshilboum R. Metabolomics: A Global Biochemical Approach to Drug Response and Disease. *Annual review of Pharmacology and Toxicology*. 2008;48: 653–683.
58. Lin Y, Stephenson MC, Xin L, Napolitano A, Morris PG. Investigating the metabolic

changes due to visual stimulation using functional proton magnetic resonance spectroscopy at 7 T. *Journal of Cerebral Blood Flow and Metabolism*. [Online] Nature Publishing Group; 2012;32(8): 1484–1495. Available from: doi:10.1038/jcbfm.2012.33
59. Veenstra TD. Metabomoics: the final frontier? *GenomeMedicine*. 2012;4(40).
60. Zhao L, Hartung T. *Metabonomics and Toxicology*. 2015. 209–231 p.
61. Cameron SJS, Takáts Z. Mass spectrometry approaches to metabolic profiling of microbial communities within the human gastrointestinal tract. *Methods*. [Online] Elsevier; 2018;149(February): 13–24. Available from: doi:10.1016/j.ymeth.2018.04.027
62. Greaves J, Roboz J. *Mass Spectormetry for the Novice*. CRC Press; 2014.
63. A. KW. Principles and application of LC·MS in new drug discovery. *Drug Discovery Today*. 2005;10(20): 1357–1367.
64. Jianga P, Lucy CA. Coupling normal phase liquid chromatography with electrospray ionization mass spectrometry: strategies and applications. *Analytical methods*. 2016;(35).
65. Maria G, Baeza-Baeza JJ, Guillermo R. Reversed Phase Liquid Chromatography. *Analytical Separation Science*. [Online] 2015; Available from: doi:https://doi.org/10.1002/9783527678129.assep008
66. Ren JL, Zhang AH, Kong L, Wang XJ. Advances in mass spectrometry-based metabolomics for investigation of metabolites. *RSC Advances*. [Online] Royal Society of Chemistry; 2018;8(40): 22335–22350. Available from: doi:10.1039/c8ra01574k
67. De Lacy Costello B, Amann A, Al-Kateb H, Flynn C, Filipiak W, Khalid T, et al. A review of the volatiles from the healthy human body. *Journal of Breath Research*. [Online] 2014;8(1). Available from: doi:10.1088/1752-7155/8/1/014001
68. Garner C, Smith S, De Lacy Costello B, White P, Spencer R, Probert C, et al. Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. *FASEB Journal*. [Online] 2007;21(8). Available from: doi:https://doi.org/10.1096/fj.06-6927com
69. Pasikanti KK, Ho PC, Chan ECY. Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*. [Online] 2008;871(2): 202–211. Available from: doi:10.1016/j.jchromb.2008.04.033
70. Garcia A, Barbas C. Gas Chromatography-Mass Spectrometry (GC-MS)-Based Metabolomics. *Methods in Molecular Biology*. 2010. p. 191–204.
71. Vernocchi P, Del Chierico F, Putignani L. Gut Microbiota Profiling: Metabolomics Based Approach to Unravel Compounds Affecting Human Health. *Frontiers of Microbiology*. [Online] 2016; Available from: doi:https://doi.org/10.3389/fmicb.2016.01144
72. Robledo VR, Smyth WF. Review of the CE-MS platform as a powerful alternative to conventional couplings in bio-omics and target-based applications. *Electrophoresis*. [Online] 2014;35(16): 2292–2308. Available from: doi:10.1002/elps.201300561
73. Mischak H, Schanstra JP. CE-MS in biomarker discovery, validation, and clinical application. *Proteomics - Clinical Applications*. [Online] 2011;5(1–2): 9–23. Available from: doi:10.1002/prca.201000058
74. Chekmeneva E, Dos Santos Correia G, Chan Q, Wijeyesekera A, Tin A, Young JH, et al. Optimization and Application of Direct Infusion Nanoelectrospray HRMS Method for Large-Scale Urinary Metabolic Phenotyping in Molecular Epidemiology. *Journal of*

*Proteome Research*. [Online] 2017;16(4): 1646–1658. Available from: doi:10.1021/acs.jproteome.6b01003
75. Hsu CC, Elnaggar MS, Peng Y, Fang J, Sanchez LM, Mascuch SJ, et al. Real-time metabolomics on living microorganisms using ambient electrospray ionization flow-probe. *Analytical Chemistry*. [Online] 2013;85(15): 7014–7018. Available from: doi:10.1021/ac401613x
76. Castrillo JI, Hayes A, Mohammed S, Gaskell SJ, Oliver SG. An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry*. [Online] 2003;62(6): 929–937. Available from: doi:10.1016/S0031-9422(02)00713-6
77. Scalbert A, Brennan L, Fiehn O, Hankemeier T, Kristal B, van Ommen B, et al. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics*. 2009;5(4).
78. Lee DY, Bowen BP, Northen TR. Mass spectrometry-based metabolomics, analysis of metabolite-protein interactions, and imaging. *BioTechniques*. [Online] 2010;49(2): 557–565. Available from: doi:10.2144/000113451
79. Rodrigues AM, Ribeiro-Barros AI, António C. Experimental design and sample preparation in forest tree metabolomics. *Metabolites*. [Online] 2019;9(12): 5–7. Available from: doi:10.3390/metabo9120285
80. Nalbantoglu S. Metabolomics: Basic Principles and Strategies. *Molecular Medicine*. [Online] 2019. Available from: doi:http://dx.doi.org/10.5772/intechopen.88563
81. Chen MX, Wang SY, Kuo CH, Tsai IL. Metabolome analysis for investigating host-gut microbiota interactions. *Journal of the Formosan Medical Association*. [Online] Elsevier Ltd; 2019;118: S10–S22. Available from: doi:10.1016/j.jfma.2018.09.007
82. Sugimoto M, Kawakami M, Robert M, Soga T. Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. 2012; 96–108.
83. De Livera AM, Dias DA, De Souza D, Rupasinghe T, Pyke J, Dedreia T, et al. Normalizing and Integrating Metabolomics Data. *Analytical Chemistry*. [Online] 2012;84: 10768–10776. Available from: doi:dx.doi.org/10.1021/ac302748b
84. Misra BB. Data normalization strategies in metabolomics: Current challenges, approaches, and tools. *European Journal of Mass Spectrometry*. [Online] 2020;26(3): 165–174. Available from: doi:10.1177/1469066720918446
85. Vettukattil R. Preprocessing of raw metabonomic data. *Methods in Molecular Biology*. [Online] 2015;1277: 123–136. Available from: doi:10.1007/978-1-4939-2377-9_10
86. Deininger SO, Cornett DS, Paape R, Becker M, Pineau C, Rauser S, et al. Normalization in MALDI-TOF imaging datasets of proteins: Practical considerations. *Analytical and Bioanalytical Chemistry*. [Online] 2011;401(1): 167–181. Available from: doi:10.1007/s00216-011-4929-z
87. van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics*. [Online] 2006;7: 1–15. Available from: doi:10.1186/1471-2164-7-142
88. Kvalheim OM, Brakstad F, Llang Y zeng. Preprocessing of Analytical Profiles in the Presence of Homoscedastic or Heteroscedastlc Noise. *Analytical Chemistry*. [Online] 1994;66(1): 43–51. Available from: doi:10.1021/ac00073a010
89. Yu T, Jones DP. Improving peak detection in high-resolution LC/MS metabolomics data using preexisting knowledge and machine learning approach. *Bioinformatics*

*(Oxford, England)*. [Online] 2014;30(20): 2941–2948. Available from: doi:10.1093/bioinformatics/btu430
90. Barberini L, Noto A, Saba L, Palmas F, Fanos V, Dessì A, et al. Multivariate data validation for investigating primary HCMV infection in pregnancy. *Data in Brief*. [Online] Elsevier; 2016;9: 220–230. Available from: doi:10.1016/j.dib.2016.08.050
91. Boccard J, Rudaz S. Extracting Knowledge from MS Clinical Metabolomic Data: Processing and Analysis Strategies. *Clinical Metabolomics*. 2018. p. 371–389.
92. Saccenti E, Hoefsloot HCJ, Smilde AK, Westerhuis JA, Hendriks M. Reflections on univariate and mutivariate analysis of metabolomics data. *Metabolomics*. 2014;10: 361–374.
93. Salek RM, Steinbeck C, Viant MR, Goodacre R, Dunn WB. The role of reporting standards for metabolite annotation and identification in metabolomic studies. *GigaScience*. [Online] 2013;2(1): 1. Available from: doi:10.1186/2047-217X-2-13
94. Sumner LW, Samuel T, Noble R, Gmbh SD, Barrett D, Beale MH, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. [Online] 2007;3(3): 211–221. Available from: doi:10.1007/s11306-007-0082-2.Proposed
95. Sumner LW, Amberg ÆA, Barrett ÆD, Beale ÆMH, Beger R, Daykin ÆCA, et al. Proposed minimum reporting standards for chemical analysis. 2007; 211–221. Available from: doi:10.1007/s11306-007-0082-2
96. Letter AWS. Common Reasons for HPLC Retention Time Drift , Variation or Change. 2015;(2): 2–5.
97. Jun Feng Xiao, Bin Zhou HWR. Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *Trends in Analytical Chemistry*. 2012;32: 1–14.
98. Murphy RC. *Tandem Mass Spectrometry of Lipids: Molecular Analysis of Complex Lipids*. 2014.
99. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean and JA. Untargeted metabolomics strategies – Challenges and Emerging Directions. *Physiology & behavior*. [Online] 2017;176(1): 139–148. Available from: doi:10.1016/j.physbeh.2017.03.040
100. Rosenberger L, Essen C Von, Khutia A, Kühn C, Urbahns K, Georgi K, et al. Crystalline Sponges as a Sensitive and Fast Method for Metabolite Identification : Application to Gemfibrozil and its Phase I and II Metabolites s. 2020; 0–6. Available from: doi:10.1124/dmd.120.091140
101. Viant MR, Kurland IJ, Jones MR, Dunn WB. ScienceDirect How close are we to complete annotation of metabolomes ? *Current Opinion in Chemical Biology*. [Online] The Author(s); 2017;36: 64–69. Available from: doi:10.1016/j.cbpa.2017.01.001
102. Poulin RX, Pohnert G. Simplifying the complex: metabolomics approaches in chemical ecology. *Analytical and Bioanalytical Chemistry*. [Online] Analytical and Bioanalytical Chemistry; 2019;411(1): 13–19. Available from: doi:10.1007/s00216-018-1470-3
103. Johnson CH, Gonzalez FJ. Challenges and opportunities of metabolomics. *Journal of Cellular Physiology*. [Online] 2012;227(8): 2975–2981. Available from: doi:10.1002/jcp.24002
104. Nicholson JK, Lindon JC. Metabolomics. *Nature*. 2008;455.
105. Klupczyńska A, Dereziński P, Kokot ZJ. Metabolomics in medical sciences - Trends, challenges and perspectives. *Acta Poloniae Pharmaceutica - Drug Research*. 2015;72(4): 629–641.

106. Castelli FA, Rosati G, Moguet C, Fuentes C, Marrugo-ramírez J, Lefebvre T, et al. *Metabolomics for personalized medicine : the input of analytical chemistry from biomarker discovery to point-of-care tests*. Analytical and Bioanalytical Chemistry; 2021.
107. Gika H, Theodoridis G, Extance J, Edge A, Wilson I. High temperature-ultra performance liquid chromatography-mass spectrometry for the metabonomic analysis of Zucker rat urine. *Journal of Chromatography*. 2008;871(2): 279–287.
108. De Spiegeleer M, De Graeve M, Huysman S, Vanderbeke A, Van Meulebroek L, Vanhaecke L. Impact of storage conditions on the human stool metabolome and lipidome: Preserving the most accurate fingerprint. *Analytica Chimica Acta*. [Online] Elsevier Ltd; 2020;1108: 79–88. Available from: doi:10.1016/j.aca.2020.02.046
109. Couch RD, Navarro K, Sikaroodi M, Gillevet P, Forsyth CB, Mutlu E, et al. The approach to sample acquisition and its impact on the derived human fecal microbiome and VOC metabolome. *PLoS ONE*. [Online] 2013;8(11): 1–13. Available from: doi:10.1371/journal.pone.0081163
110. Bezabeh T, Somorjai RL, Smith ICP. MR metabolomics of fecal extracts: Applications in the study of bowel diseases. *Magnetic Resonance in Chemistry*. [Online] 2009;47(SUPPL. 1). Available from: doi:10.1002/mrc.2530
111. Deda O, Gika HG, Wilson ID, Theodoridis GA. An overview of fecal sample preparation for global metabolic profiling. *Journal of Pharmaceutical and Biomedical Analysis*. [Online] Elsevier B.V.; 2015;113: 137–150. Available from: doi:10.1016/j.jpba.2015.02.006
112. Jansson J, Willing B, Lucio M, Fekete A, Dicksved J, Halfvarson J, et al. Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS ONE*. [Online] 2009;4(7). Available from: doi:10.1371/journal.pone.0006386
113. Loftfield E, Vogtmann E, Sampson JN, Moore SC, Nelson H, Knight R, et al. Comparison of collection methods for fecal samples for discovery metabolomics in epidemiologic studies. *Cancer Epidemiology Biomarkers and Prevention*. [Online] 2016;25(11): 1483–1490. Available from: doi:10.1158/1055-9965.EPI-16-0409
114. Gratton J, Phetcharaburanin J, Mullish BH, Williams HRT, Thursz M, Nicholson JK, et al. Optimized Sample Handling Strategy for Metabolic Profiling of Human Feces. *Analytical Chemistry*. [Online] 2016;88(9): 4661–4668. Available from: doi:10.1021/acs.analchem.5b04159
115. De Paepe E, Van Meulebroek L, Rombouts C, Huysman S, Verplanken K, Lapauw B, et al. A validated multi-matrix platform for metabolomic fingerprinting of human urine, feces and plasma using ultra-high performance liquid-chromatography coupled to hybrid orbitrap high-resolution mass spectrometry. *Analytica Chimica Acta*. [Online] Elsevier Ltd; 2018;1033: 108–118. Available from: doi:10.1016/j.aca.2018.06.065
116. BLISS D. Comparison of subjective classification of stool consistency and stool water content*1. *Journal of WOCN*. [Online] 1999;26(3): 137–141. Available from: doi:10.1016/s1071-5754(99)90031-1
117. Pinu FR, Beale DJ, Paten AM, Kouremenos K, Swarup S, Schirra HJ, et al. Systems biology and multi-omics integration: Viewpoints from the metabolomics research community. *Metabolites*. [Online] 2019;9(4): 1–31. Available from: doi:10.3390/metabo9040076
118. Clendinnen CS, Monge ME, Fernández FM. Ambient Mass Spectrometry in Metabolomics. *Physiology & behavior*. [Online] 2016;176(1): 139–148. Available from:

doi:10.1016/j.physbeh.2017.03.040
119. Cooks RG, Ouyang Z, Takats Z, Wiseman JM. Detection Technologies. Ambient mass spectrometry. *Science (New York, N.Y.)*. [Online] American Association for the Advancement of Science; 2006;311(5767): 1566–1570. Available from: doi:10.1126/science.1119426 [Accessed: 18th May 2018]
120. Van Meulebroek L, Cameron S, Plekhova V, De Spiegeleer M, Wijnant K, Michels N, et al. Rapid LA-REIMS and comprehensive UHPLC-HRMS for metabolic phenotyping of feces. *Talanta*. [Online] Elsevier B.V.; 2020;217(April): 121043. Available from: doi:10.1016/j.talanta.2020.121043
121. Takáts Z, Wiseman JM, Gologan B, Cooks RG. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science*. [Online] 2004;306(5695): 471–473. Available from: doi:10.1126/science.1104404
122. Wang C, Zhu H, Cai Z, Song F, Liu Z, Liu S. Newborn screening of phenylketonuria using direct analysis in real time (DART) mass spectrometry. *Analytical and Bioanalytical Chemistry*. [Online] 2013;405(10): 3159–3164. Available from: doi:10.1007/s00216-013-6713-8
123. Schäfer KC, Szaniszló T, Günther S, Balog J, Dénes J, Keserü M, et al. In situ, real-time identification of biological tissues by ultraviolet and infrared laser desorption ionization mass spectrometry. *Analytical Chemistry*. [Online] 2011;83(5): 1632–1640. Available from: doi:10.1021/ac102613m
124. Wójtowicz A, Wietecha-Posłuszny R. DESI-MS analysis of human fluids and tissues for forensic applications. *Applied Physics A: Materials Science and Processing*. [Online] Springer Berlin Heidelberg; 2019;125(5): 1–9. Available from: doi:10.1007/s00339-019-2564-2
125. Cooks RG, Ouyang Z, Takats Z, Wiseman JM. Ambient mass spectrometry. *Science*. 2006;311.
126. Ifa DR, Wiseman JM, Song Q, Cooks RG. Development of capabilities for imaging mass spectrometry under ambient conditions with desorption electrospray ionization (DESI). *International Journal of Mass Spectrometry*. [Online] 2007;259(1–3): 8–15. Available from: doi:10.1016/j.ijms.2006.08.003
127. Kertesz V, Van Berkel GJ. Improved imaging resolution in desorption electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry*. [Online] 2008;22(17): 2639–2644. Available from: doi:10.1002/rcm.3662
128. Parrot D, Papazian S, Foil D, Tasdemir D. Imaging the Unimaginable: Desorption Electrospray Ionization - Imaging Mass Spectrometry (DESI-IMS) in Natural Product Research. *Planta Medica*. [Online] 2018;84(9–10): 584–593. Available from: doi:10.1055/s-0044-100188
129. Wiseman JM, Evans CA, Bowen CL, Kennedya JH. Direct analysis of dried blood spots utilizing desorption electrospray ionization (DESI) mass spectrometry. *Analyst*. [Online] 2010;135: 720–725. Available from: doi:https://doi.org/10.1039/B922329K
130. Calligaris D, Caragacianu D, Liu X, Norton I, Thompson CJ, Richardson AL, et al. Application of desorption electrospray ionization mass spectrometry imaging in breast cancer margin analysis. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 2014;111(42): 15184–15189. Available from: doi:10.1073/pnas.1408129111
131. Keriana KS, Jarmuscha AK, V. Pirroa MOK, Mastersonb TA, Chengc L, Cooks G. Differentiation of Prostate Cancer from Normal Tissue in Radical Prostatectomy

Specimens by Desorption Electrospray Ionization and Touch Spray Ionization Mass Spectrometry. *Physiology & behavior*. [Online] 2016;176(1): 139–148. Available from: doi:10.1016/j.physbeh.2017.03.040
132. Wiseman JM, Puolitaival SM, Takáts Z, Cooks RG, Caprioli RM. Mass spectrometric profiling of intact biological tissue by using desorption electrospray ionization. *Angewandte Chemie - International Edition*. [Online] 2005;44(43): 7094–7097. Available from: doi:10.1002/anie.200502362
133. Kevin Range and DMYAM. Multivariate Statistical Identification of Human Bladder Carcinomas Using Ambient Ionization Imaging Mass Spectrometry. *Bone*. [Online] 2012;23(1): 1–7. Available from: doi:10.1038/jid.2014.371
134. Abbassi-Ghadi N, Veselkov K, Kumar S, Huang J, Jones E, Strittmatter N, et al. Discrimination of lymph node metastases using desorption electrospray ionisation-mass spectrometry imaging. *Chem Communication*. [Online] 2014;50(28): 3661–3664. Available from: doi:10.1039/c3cc48927b
135. Gerbig S, Golf O, Balog J, Denes J, Baranyai Z, Zarand A, et al. Analysis of colorectal adenocarcinoma tissue by desorption electrospray ionization mass spectrometric imaging. *Analytical and Bioanalytical Chemistry*. [Online] 2012;403(8): 2315–2325. Available from: doi:10.1007/s00216-012-5841-x
136. Dória ML, McKenzie JS, Mroz A, Phelps DL, Speller A, Rosini F, et al. Epithelial ovarian carcinoma diagnosis by desorption electrospray ionization mass spectrometry imaging. *Scientific Reports*. [Online] Nature Publishing Group; 2016;6(November): 1–11. Available from: doi:10.1038/srep39219
137. Eberlin1 LS, Norton I, Dill AL, Golby AJ, Ligon KL, Santagata S, et al. Classifying Human Brain Tumors by Lipid Imaging with Mass Spectrometry. *Bone*. [Online] 2011;23(1): 1–7. Available from: doi:10.1161/CIRCULATIONAHA.110.956839
138. Crawford E, Gordon J, Wu J-T, Musselman B, Liu R, Yu S. Direct analysis in real time coupled with dried spot sampling for bioanalysis in a drug-discovery setting. *Bioanalysis*. [Online] 2011;3(11). Available from: doi:https://doi.org/10.4155/bio.11.99
139. Schäfer KC, Dénes J, Albrecht K, Szaniszló T, Balogh J, Skoumal R, et al. In vivo, in situ tissue analysis using rapid evaporative ionization mass spectrometry. *Angewandte Chemie - International Edition*. [Online] 2009;48(44): 8240–8242. Available from: doi:10.1002/anie.200902546
140. Alexander J, Gildea L, Balog J, Speller A, McKenzie J, Muirhead L, et al. A novel methodology for in vivo endoscopic phenotyping of colorectal cancer based on real-time analysis of the mucosal lipidome: a prospective observational study of the iKnife. *Surgical endoscopy*. [Online] Springer; 2017;31(3): 1361–1370. Available from: doi:10.1007/s00464-016-5121-5 [Accessed: 18th May 2018]
141. St John ER, Balog J, McKenzie JS, Rossi M, Covington A, Muirhead L, et al. Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: towards an intelligent knife for breast cancer surgery. *Breast Cancer Research*. [Online] BioMed Central; 2017;19(1): 59. Available from: doi:10.1186/s13058-017-0845-2 [Accessed: 18th May 2018]
142. Tzafetas M, Mitra A, Paraskevaidi M, Bodai Z, Kalliala I, Bowden S, et al. The intelligent knife (iKnife) and its intraoperative diagnostic advantage for the treatment of cervical disease (Proceedings of the National Academy of Sciences of the United States of America (2020) 117 13 (7338-7346)). *Proceedings of the National Academy*

*of Sciences of the United States of America*. [Online] 2020;117(31): 18892. Available from: doi:10.1073/pnas.2013655117
143. Balog J, Sasi-Szabó L, Kinross J, Lewis MR, Muirhead LJ, Veselkov K, et al. Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. *Science Translational Medicine*. [Online] 2013;5(194). Available from: doi:10.1126/scitranslmed.3005623
144. Golf O, Strittmatter N, Karancsi T, Pringle SD, Speller AVM, Mroz A, et al. Rapid evaporative ionization mass spectrometry imaging platform for direct mapping from bulk tissue and bacterial growth media. *Analytical Chemistry*. [Online] 2015;87(5): 2527–2534. Available from: doi:10.1021/ac5046752
145. Strittmatter N, Jones EA, Veselkov KA, Rebec M, Bundya JG, Takats Z. Analysis of intact bacteria using rapid evaporative ionisation mass spectrometry. *Chemical Communications*. 2013;(55).
146. Bolt F, Cameron SJS, Karancsi T, Simon D, Schaffer R, Rickards T, et al. Automated High-Throughput Identification and Characterization of Clinically Important Bacteria and Fungi using Rapid Evaporative Ionization Mass Spectrometry. *Analytical Chemistry*. [Online] 2016;88(19): 9419–9426. Available from: doi:10.1021/acs.analchem.6b01016 [Accessed: 18th May 2018]
147. Bardin EE, Cameron SJS, Perdones-Montero A, Hardiman K, Bolt F, Alton EWFW, et al. Metabolic Phenotyping and Strain Characterisation of Pseudomonas aeruginosa Isolates from Cystic Fibrosis Patients Using Rapid Evaporative Ionisation Mass Spectrometry. *Scientific Reports*. [Online] 2018;8(1): 1–10. Available from: doi:10.1038/s41598-018-28665-7
148. Cameron SJS, Bolt F, Perdones-Montero A, Rickards T, Hardiman K, Abdolrasouli A, et al. Rapid Evaporative Ionisation Mass Spectrometry (REIMS) Provides Accurate Direct from Culture Species Identification within the Genus Candida. *Scientific Reports*. [Online] Nature Publishing Group; 2016;6(1): 36788. Available from: doi:10.1038/srep36788 [Accessed: 18th May 2018]
149. Cameron SJS, Alexander JL, Bolt F, Burke A, Ashrafian H, Teare J, et al. Evaluation of Direct from Sample Metabolomics of Human Feces Using Rapid Evaporative Ionization Mass Spectrometry. *Analytical Chemistry*. [Online] 2019;91(21): 13448–13457. Available from: doi:10.1021/acs.analchem.9b02358
150. Paizs P. *BSc thesis: Development of novel biomarkers for colon cancer screening using Rapid Evaporative Ionisation Mass Spectrometry (REIMS) – A prospective validation study*. Imperial College London; 2018.
151. Cameron SJS, Perdones-Montero A, Van Meulebroek L, Burke A, Alexander-Hardiman K, Simon D, et al. Sample Preparation Free Mass Spectrometry Using Laser-Assisted Rapid Evaporative Ionization Mass Spectrometry: Applications to Microbiology, Metabolic Biofluid Phenotyping, and Food Authenticity. *Journal of the American Society for Mass Spectrometry*. [Online] 2021; Available from: doi:10.1021/jasms.0c00452
152. Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, et al. Global metabolic profiling procedures for urine using UPLC-MS. *Nature Protocols*. [Online] Nature Publishing Group; 2010;5(6): 1005–1018. Available from: doi:10.1038/nprot.2010.50
153. Genangeli M, Heeren RMA, Porta Siegel T. Tissue classification by rapid evaporative ionization mass spectrometry (REIMS): comparison between a diathermic knife and

CO2 laser sampling on classification performance. *Analytical and Bioanalytical Chemistry*. [Online] Analytical and Bioanalytical Chemistry; 2019;411(30): 7943–7955. Available from: doi:10.1007/s00216-019-02148-8
154. *Bowel cancer incidence statistics | Cancer Research UK*. [Online] Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/incidence#ref- [Accessed: 18th May 2018]
155. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. [Online] 2018;68(6): 394–424. Available from: doi:10.3322/caac.21492
156. J F, M E, F L. *Global Cancer Observatory: Cancer Today*. [Online] Available from: http://gco.iarc.fr/today/data/factsheets/cancers/10_8_9-Colorectum-fact-sheet.pdf
157. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors. *Przeglad Gastroenterologiczny*. [Online] 2019;14(2): 89–103. Available from: doi:10.5114/pg.2018.81072
158. Montalban-Arques A, Scharl M. Intestinal microbiota and colorectal carcinoma: Implications for pathogenesis, diagnosis, and therapy. *EBioMedicine*. [Online] Elsevier B.V.; 2019;48: 648–655. Available from: doi:10.1016/j.ebiom.2019.09.050
159. Keum NN, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nature Reviews Gastroenterology and Hepatology*. [Online] Springer US; 2019;16(12): 713–732. Available from: doi:10.1038/s41575-019-0189-8
160. Hagland HR, Berg M, Jolma IW, Carlsen A, Søreide K. Molecular pathways and cellular metabolism in colorectal cancer. *Digestive Surgery*. [Online] 2013;30(1): 12–25. Available from: doi:10.1159/000347166
161. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990;61(5): 759–767.
162. Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and Familial Colon Cancer. *Gastroenterology*. [Online] Elsevier Inc.; 2010;138(6): 2044–2058. Available from: doi:10.1053/j.gastro.2010.01.054
163. Taylor DP, Burt RW, Williams MS, Haug PJ, Cannon-Albright LA. Population-Based Family History-Specific Risks for Colorectal Cancer: A Constellation Approach. *Gastroenterology*. [Online] Elsevier Inc.; 2010;138(3): 877–885. Available from: doi:10.1053/j.gastro.2009.11.044
164. Baglietto L, Jenkins MA, Severi G, Giles GG, Bishop DT, Boyle P, et al. Measures of familial aggregation depend on definition of family history: Meta-analysis for colorectal cancer. *Journal of Clinical Epidemiology*. [Online] 2006;59(2): 114–124. Available from: doi:10.1016/j.jclinepi.2005.07.018
165. Johns LE, Houlston RS. A Systematic Review and Meta-Analysis of Familial Colorectal Cancer Risk. *American Journal of Gastroenterology*. [Online] 2001;96(10): 2992–3003. Available from: doi:10.1111/j.1572-0241.2001.04677.x
166. Carethers JM, Jung BH. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. *Gastroenterology*. [Online] Elsevier, Inc; 2015;149(5): 1177-1190.e3. Available from: doi:10.1053/j.gastro.2015.06.047
167. Pakiet A, Kobiela J, Stepnowski P, Sledzinski T, Mika A. Changes in lipids composition and metabolism in colorectal cancer: A review. *Lipids in Health and Disease*. [Online] 2019;18(1): 1–21. Available from: doi:10.1186/s12944-019-0977-8

168. Brown R, Short S, Williams C. Colorectal Cancer and Metabolism. *Physiology & behavior*. [Online] 2017;176(3): 139–148. Available from: doi:10.1007/s11888-018-0420-y.Colorectal
169. Warburg O, Wind F, Negelein E. I . Killing-Off of Tumor Cells in Vitro . *The journal of General Physiology*. 1927;8(6): 519–530.
170. Xu XD, Shao SX, Jiang HP, Cao YW, Wang YH, Yang XC, et al. Warburg effect or reverse warburg effect? a review of cancer metabolism. *Oncology Research and Treatment*. [Online] 2015;38(3): 117–122. Available from: doi:10.1159/000375435
171. Jiao S, Peters U, Berndt S, Brenner H, Butterbach K, Caan BJ, et al. Estimating the heritability of colorectal cancer. *Human Molecular Genetics*. [Online] 2014;23(14): 3898–3905. Available from: doi:10.1093/hmg/ddu087
172. Lansdorp-Vogelaar I, Kuntz KM, Knudsen AB, Van Ballegooijen M, Zauber AG, Jemal A. Contribution of screening and survival differences to racial disparities in colorectal cancer rates. *Cancer Epidemiology Biomarkers and Prevention*. [Online] 2012;21(5): 728–736. Available from: doi:10.1158/1055-9965.EPI-12-0023
173. Yamada A, Komaki Y, Komaki F, Micic D, Zullow S, Sakuraba A. Risk of gastrointestinal cancers in patients with cystic fibrosis: a systematic review and meta-analysis. *The Lancet Oncology*. [Online] Elsevier Ltd; 2018;19(6): 758–767. Available from: doi:10.1016/S1470-2045(18)30188-8
174. Zhang Y, Liu H, Li L, Ai M, Gong Z, He Y, et al. Cholecystectomy can increase the risk of colorectal cancer: A meta-analysis of 10 cohort studies. *PLoS ONE*. [Online] 2017;12(8): 1–17. Available from: doi:10.1371/journal.pone.0181852
175. Gillessen S, Templeton A, Marra G, Kuo YF, Valtorta E, Shahinian VB. Risk of colorectal cancer in men on long-term androgen deprivation therapy for prostate cancer. *Journal of the National Cancer Institute*. [Online] 2010;102(23): 1760–1770. Available from: doi:10.1093/jnci/djq419
176. Observatory TGC. *Cancer Fact Sheets: Colorectal Cancer*. [Online] Available from: http://gco.iarc.fr/today/fact-sheets-cancers
177. Calle EE, Kaaks R. Overweight, obesity and cancer: Epidemiological evidence and proposed mechanisms. *Nature Reviews Cancer*. [Online] 2004;4(8): 579–591. Available from: doi:10.1038/nrc1408
178. Farahani H, Mahmoudi T, Asadi A, Nobakht H, Dabiri R, Hamta A. Insulin Resistance and Colorectal Cancer Risk: the Role of Elevated Plasma Resistin Levels. *Journal of Gastrointestinal Cancer*. [Online] 2020;51(2): 478–483. Available from: doi:10.1007/s12029-019-00260-7
179. World Cancer Research Fund International /, American Institute for Cancer Research. Continuous Update Project Report: Diet , Nutrition , Physical activity and Stomach cancer. *Stomach Cancer Report 2016*. 2016;
180. *International Agency for Research on Cancer: Agents classified by the IARC Monograohs*. [Online] Available from: https://monographs.iarc.fr/list-of-classifications
181. Rossi M, Anwar MJ, Usman A, Keshavarzian A, Bishehsari F. Colorectal cancer and alcohol consumption—populations to molecules. *Cancers*. [Online] 2018;10(2). Available from: doi:10.3390/cancers10020038
182. Botteri E, Iodice S, Raimondi S, Maisonneuve P, Lowenfels AB. Cigarette Smoking and Adenomatous Polyps: A Meta-analysis. *Gastroenterology*. [Online] 2008;134(2): 388–395. Available from: doi:10.1053/j.gastro.2007.11.007

183. Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and Colorectal Cancer. *Jama*. [Online] 2008;300(23): 2765. Available from: doi:10.1001/jama.2008.839
184. Hou L, Jiang J, Liu B, Nasca PC, Wu Y, Zou X, et al. Association between smoking and deaths due to colorectal malignant carcinoma: A national population-based case-control study in China. *British Journal of Cancer*. [Online] 2014;110(5): 1351–1358. Available from: doi:10.1038/bjc.2014.9
185. Liang PS, Chen TY, Giovannucci E. Cigarette smoking and colorectal cancer incidence and mortality: Systematic review and meta-analysis. *International Journal of Cancer*. [Online] 2009;124(10): 2406–2415. Available from: doi:10.1002/ijc.24191
186. Giovannucci E, Martinez ME. Tobacco, Colorectal Cancer, and Adenomas: a Review of the Evidence. *Journal of the National Cancer Institute*. 1996;88(23).
187. Jahani-Sherafat S, Alebouyeh M, Moghim S, Amoli HA, Ghasemian-Safaei H. Role of gut microbiota in the pathogenesis of colorectal cancer; a review article. *Hepatol Bed Bench*. 2018;11(2): 101–109.
188. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. [Online] 2010;464(7285): 59–65. Available from: doi:10.1038/nature08821
189. Wong SH, Yu J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nature Reviews Gastroenterology and Hepatology*. [Online] Springer US; 2019;16(11): 690–704. Available from: doi:10.1038/s41575-019-0209-8
190. Saus E, Iraola-Guzmán S, Willis JR, Brunet-Vega A, Gabaldón T. Microbiome and colorectal cancer: Roles in carcinogenesis and clinical potential. *Molecular Aspects of Medicine*. [Online] Elsevier; 2019;69(March): 93–106. Available from: doi:10.1016/j.mam.2019.05.001
191. *Cancer survival by stage at diagnosis for England (experimental statistics) - Office for National Statistics*. [Online] Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancersurvivalbystageatdiagnosisforenglandexperimentalstatistics/adultsdiagnosed20122013and2014andfollowedupto2015#survival-by-stage [Accessed: 18th May 2018]
192. *Bowel cancer screening in England · Bowel Cancer UK*. [Online] Available from: https://www.bowelcanceruk.org.uk/media-centre/blog/bowel-cancer-screening-in-england/ [Accessed: 8th May 2018]
193. Lee JK, Liles EG, Bent S, Levin TR, Corley DA. Accuracy of Fecal Immunochemical Tests for Colorectal Cancer. *Annals of Internal Medicine*. [Online] American College of Physicians; 2014;160(3): 171–181. Available from: doi:10.7326/M13-1484 [Accessed: 8th May 2018]
194. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget Stool DNA Testing for Colorectal-Cancer Screening. *New England Journal of Medicine*. [Online] Massachusetts Medical Society; 2014;370(14): 1287–1297. Available from: doi:10.1056/NEJMoa1311194 [Accessed: 18th May 2018]
195. Brenner H, Tao S. Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. *European Journal of Cancer*. [Online] 2013;49(14): 3049–3054. Available from: doi:10.1016/j.ejca.2013.04.023 [Accessed: 8th May 2018]

196. Chiu H, Lee Y, Tu C, Chen C, Tseng P, Liang J, et al. Association Between Early Stage Colon Neoplasms and False-negative Results From the Fecal Immunochemical Test. *Clinical Gastroenterology and Hepatology*. [Online] 2013;11(7): 832-838.e2. Available from: doi:10.1016/j.cgh.2013.01.013 [Accessed: 8th May 2018]
197. Shapiro JA, Bobo JK, Church TR, Rex DK, Chovnick G, Thompson TD, et al. A Comparison of Fecal Immunochemical and High-Sensitivity Guaiac Tests for Colorectal Cancer Screening. *Nature Publishing Group*. [Online] 2017; Available from: doi:10.1038/ajg.2017.285 [Accessed: 8th May 2018]
198. MORIKAWA T, KATO J, YAMAJI Y, WADA R, MITSUSHIMA T, SHIRATORI Y. A Comparison of the Immunochemical Fecal Occult Blood Test and Total Colonoscopy in the Asymptomatic Population. *Gastroenterology*. [Online] 2005;129(2): 422–428. Available from: doi:10.1016/j.gastro.2005.05.056 [Accessed: 18th May 2018]
199. Park D Il, Ryu S, Kim Y-H, Lee S-H, Lee CK, Eun CS, et al. Comparison of Guaiac-Based and Quantitative Immunochemical Fecal Occult Blood Testing in a Population at Average Risk Undergoing Colorectal Cancer Screening. *The American Journal of Gastroenterology*. [Online] 2010;105(9): 2017–2025. Available from: doi:10.1038/ajg.2010.179 [Accessed: 22nd May 2018]
200. Hernandez V, Cubiella J, Gonzalez-Mao MC, Iglesias F, Rivera C, Iglesias MB, et al. Fecal immunochemical test accuracy in average-risk colorectal cancer screening. *World journal of gastroenterology*. [Online] Baishideng Publishing Group Inc; 2014;20(4): 1038–1047. Available from: doi:10.3748/wjg.v20.i4.1038 [Accessed: 18th May 2018]
201. Erben V, Bhardwaj M, Schrotz-King P, Brenner H. Metabolomics biomarkers for detection of colorectal neoplasms: A systematic review. *Cancers*. [Online] 2018;10(8): 1–24. Available from: doi:10.3390/cancers10080246
202. Metallo CM. Expanding the reach of cancer metabolomics. *Cancer Prevention Research*. [Online] 2012;5(12): 1337–1340. Available from: doi:10.1158/1940-6207.CAPR-12-0433
203. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nature Medicine*. [Online] Springer US; 2019;25(6): 968–976. Available from: doi:10.1038/s41591-019-0458-7
204. Zhgun ES, Ilyina EN. Fecal Metabolites As Non-Invasive Biomarkers of Gut Diseases. *Acta Naturae*. [Online] 2020;12(2): 4–14. Available from: doi:10.32607/actanaturae.10954
205. WU D, YANG JJ, YANG F, ZHANG BY, DU J, WANG YF, et al. Analysis of Alkaline and Neutral Volatile Metabolites in Feces by Gas Chromatography-Tandem Mass Spectrometry. *Chinese Journal of Analytical Chemistry*. [Online] Changchun Institute of Applied Chemistry, Chinese Academy of Sciences; 2017;45(6): 837–843. Available from: doi:10.1016/S1872-2040(17)61019-3
206. Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW, et al. Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 2008;105(10): 3805–3810. Available from: doi:10.1073/pnas.0708897105
207. Cangelosi GA, Meschke JS. Dead or alive: Molecular assessment of microbial viability. *Applied and Environmental Microbiology*. [Online] 2014;80(19): 5884–5891. Available from: doi:10.1128/AEM.01763-14

208. Marcobal A, Kashyap PC, Nelson TA, Aronov PA, Donia MS, Spormann A, et al. A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice. *ISME Journal*. [Online] Nature Publishing Group; 2013;7(10): 1933–1943. Available from: doi:10.1038/ismej.2013.89
209. Trošt K, Ahonen L, Suvitaival T, Christiansen N, Nielsen T, Thiele M, et al. Describing the fecal metabolome in cryogenically collected samples from healthy participants. *Scientific Reports*. [Online] 2020;10(1): 1–8. Available from: doi:10.1038/s41598-020-57888-w
210. Zhang F, Zhang Y, Zhao W, Deng K, Wang Z, Yang C, et al. Metabolomics for biomarker discovery in the diagnosis, prognosis, survival and recurrence of colorectal cancer: A systematic review. *Oncotarget*. [Online] 2017;8(21): 35460–35472. Available from: doi:10.18632/oncotarget.16727
211. Kim M, Vogtmann E, Ahlquist DA, Devens ME, Kisiel JB, Taylor WR, et al. Fecal metabolomic signatures in colorectal adenoma patients are associated with gut microbiota and early events of colorectal cancer pathogenesis. *mBio*. [Online] 2020;11(1): 1–16. Available from: doi:10.1128/mBio.03186-19
212. Yang Y, Misra BB, Liang L, Bi D, Weng W, Wu W, et al. Integrated microbiome and metabolome analysis reveals a novel interplay between commensal bacteria and metabolites in colorectal cancer. *Theranostics*. [Online] 2019;9(14): 4101–4114. Available from: doi:10.7150/thno.35186
213. Sinha R, Ahn J, Sampson JN, Shi J, Yu G, Xiong X, et al. Fecal microbiota, fecal metabolome, and colorectal cancer interrelations. *PLoS ONE*. [Online] 2016;11(3): 1–13. Available from: doi:10.1371/journal.pone.0152126
214. Monedeiro F, Monedeiro-Milanowski M, Ligor T, Buszewski B. A Review of GC-Based Analysis of Non-Invasive Biomarkers of Colorectal Cancer and Related Pathways. *Journal of Clinical Medicine*. [Online] 2020;9(10): 3191. Available from: doi:10.3390/jcm9103191
215. *The LIPID MAPS Lipidomics Gateway*. [Online] Available from: https://www.lipidmaps.org/
216. Fahy E, Cotter D, Sud M, Subramaniam S. Lipid classification, structures and tools. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*. [Online] 2011;1811(11): 637–647. Available from: doi:10.1016/j.bbalip.2011.06.009
217. Van Meulebroek L, De Paepe E, Vercruysse V, Pomian B, Bos S, Lapauw B, et al. Holistic Lipidomics of the Human Gut Phenotype Using Validated Ultra-High-Performance Liquid Chromatography Coupled to Hybrid Orbitrap Mass Spectrometry. *Analytical Chemistry*. [Online] 2017;89(22): 12502–12510. Available from: doi:10.1021/acs.analchem.7b03606
218. Pasternak G, Aebisher D, Filip R, Bartusik-Aebisher D. Inflammatory bowel disease: clinical aspects. *European Journal of Clinical and Experimental Medicine*. [Online] 2019;16(4): 341–345. Available from: doi:10.15584/ejcem.2018.4.12
219. Segal JP, Mullish BH, Quraishi MN, Acharjee A, Williams HRT, Iqbal T, et al. The application of omics techniques to understand the role of the gut microbiota in inflammatory bowel disease. *Therapeutic Advances in Gastroenterology*. 2019;12(1–13).
220. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*. [Online] Elsevier Inc.; 2012;142(1): 46-54.e42.

Available from: doi:10.1053/j.gastro.2011.10.001
221. Alatab S, Sepanlou SG, Ikuta K, Vahedi H, Bisignano C, Safiri S, et al. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Gastroenterology and Hepatology*. [Online] 2020;5(1): 17–30. Available from: doi:10.1016/S2468-1253(19)30333-4
222. Jones GR, Lyons M, Plevris N, Jenkinson PW, Bisset C, Burgess C, et al. IBD prevalence in Lothian, Scotland, derived by capture-recapture methodology. *Gut*. [Online] 2019;68(11): 1953–1960. Available from: doi:10.1136/gutjnl-2019-318936
223. Hendrickson BA, Gokhale R, Cho JH. Clinical aspects and pathophysiology of inflammatory bowel disease. *Clinical Microbiology Reviews*. [Online] 2002;15(1): 79–94. Available from: doi:10.1128/CMR.15.1.79-94.2002
224. Feakins RM. Inflammatory bowel disease biopsies: Updated British Society of Gastroenterology reporting guidelines. *Journal of Clinical Pathology*. [Online] 2013;66(12): 1005–1026. Available from: doi:10.1136/jclinpath-2013-201885
225. Khan I, Ullah N, Zha L, Bai Y, Khan A, Zhao T, et al. Alteration of Gut Microbiota in Inflammatory BowelDisease (IBD): Cause or Consequence? IBD TreatmentTargeting the Gut Microbiome. *Pathogens*. 2019;8(3): 1–28.
226. Ek WE, D'Amato M, Halfvarson J. The history of genetics in inflammatory bowel disease. *Annals of Gastroenterology*. 2014;27(4): 294–303.
227. Thompson NP, Driscoll R, Pounder RE, Wakefield AJ. Concordance for inflammatory bowel disease in twin pairs. 312: 95–96.
228. Duerr RH. Genome-Wide Association Studies Herald a New Era of Rapid Discoveries in Inflammatory Bowel Disease Research. *Gastroenterology*. [Online] 2007;132(5): 2045–2049. Available from: doi:10.1053/j.gastro.2007.03.082
229. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. 2001;411(May): 603–606.
230. Ferguson LR, Shelling AN, Browning BL, Huebner C, Petermann I. Genes, diet and inflammatory bowel disease. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*. [Online] 2007;622(1–2): 70–83. Available from: doi:10.1016/j.mrfmmm.2007.05.011
231. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature Genetics*. [Online] 2007;39(2): 207–211. Available from: doi:10.1038/ng1954
232. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Huett A, Green T, et al. Genome-wide association study identifies five novel susceptibility loci for Crohn's disease and implications a role for autophagy in disease pathogenesis. *Nature Genetics*. [Online] 2007;39(5): 596–604. Available from: doi:10.1038/ng2032.Genome-wide
233. Murthy A, Li Y, Peng I, Reichelt M, Katakam AK, Noubade R, et al. A Crohn's disease variant in Atg16l1 enhances its degradation by caspase 3. *Nature*. [Online] Nature Publishing Group; 2014;506(7489): 456–462. Available from: doi:10.1038/nature13044
234. Duerr RH, Taylor KD, Brant SR, Rioux JD, Mark S, Daly MJ, et al. A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. 2015;314(5804): 1461–1463. Available from: doi:10.1126/science.1135245.A

235. Kordjazy N, Haj-Mirzaian A, Haj-Mirzaian A, Rohani MM, Gelfand EW, Rezaei N, et al. Role of toll-like receptors in inflammatory bowel disease. *Pharmacological Research*. [Online] Elsevier Ltd; 2018;129: 204–215. Available from: doi:10.1016/j.phrs.2017.11.017
236. Younis N, Zarif R, Mahfouz R. Inflammatory bowel disease: between genetics and microbiota. *Molecular Biology Reports*. [Online] Springer Netherlands; 2020;47(4): 3053–3063. Available from: doi:10.1007/s11033-020-05318-5
237. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of IBD. *Nature*. [Online] 2011;474(7351): 307–317. Available from: doi:10.1038/nature10209.Genetics
238. Loftus E V. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology*. [Online] 2004;126(6): 1504–1517. Available from: doi:10.1053/j.gastro.2004.01.063
239. Li X, Sundquist J, Hemminki K, Sundquist K. Risk of inflammatory bowel disease in first- and second-generation immigrants in Sweden: A nationwide follow-up study. *Inflammatory Bowel Diseases*. [Online] 2011;17(8): 1784–1791. Available from: doi:10.1002/ibd.21535
240. Benchimol EI, Mack DR, Guttmann A, Nguyen GC, To T, Mojaverian N, et al. Inflammatory bowel disease in immigrants to Canada and their children: A population-based cohort study. *American Journal of Gastroenterology*. [Online] Nature Publishing Group; 2015;110(4): 553–563. Available from: doi:10.1038/ajg.2015.52
241. Lee SH, Yun Y, Kim SJ, Lee E-J, Chang Y, Ryu S, et al. Association between Cigarette Smoking Status and Composition of Gut Microbiota: Population-Based Cross-Sectional Study. *Journal of Clinical Medicine*. [Online] 2018;7(9): 282. Available from: doi:10.3390/jcm7090282
242. Mahid SS, Minor KS, Soto RE, Hornung CA, Galandiuk S. Smoking and inflammatory bowel disease: a meta-analysis. *Mayo Clin Proc*. 2006;81(11): 1462–1471.
243. Higuchi LM, Khalili H, Chan AT, Richter JM, Bousvaros A, Fuchs CS. A prospective study of cigarette smoking and the risk of inflammatory bowel disease in women. *American Journal of Gastroenterology*. [Online] 2012;107(9): 1399–1406. Available from: doi:10.1038/ajg.2012.196
244. Ananthakrishnan AN, Khalili H, Higuchi LM, Bao Y, Korzenik JR, Giovannucci EL, et al. Higher predicted vitamin D status is associated with reduced risk of crohn's disease. *Gastroenterology*. [Online] Elsevier Inc.; 2012;142(3): 482–489. Available from: doi:10.1053/j.gastro.2011.11.040
245. Kaplan GG, Jackson T, Sands BE, Frisch M, Andersson RE, Korzenik J. The risk of developing Crohn's disease after an appendectomy: a meta-analysis. *American Journal of Gastroenterology*. 2008;103(11).
246. Koutroubakis IE, Vlachonikolis IG, Kouroumalis EA. Role of appendicitis and appendectomy in the pathogenesis of ulcerative colitis: A critical review. *Inflammatory Bowel Diseases*. [Online] 2002;8(4): 277–286. Available from: doi:10.1097/00054725-200207000-00007
247. Xu L, Lochhead P, Ko Y, Claggett B, Leong RW, Ananthakrishnan AN. Systematic review with meta-analysis: Breastfeeding and the risk of Crohn's disease and ulcerative colitis. *Physiology & behavior*. [Online] 2018;176(1): 139–148. Available from: doi:10.1111/apt.14291.Systematic
248. Kaplan GG, Hubbard J, Korzenik J, Sands BE, Panaccione R, Ghosh S, et al. The

inflammatory bowel diseases and ambient air pollution: A novel association. *American Journal of Gastroenterology*. [Online] Nature Publishing Group; 2010;105(11): 2412–2419. Available from: doi:10.1038/ajg.2010.252
249. Koloski NA, Bret L, Radford-Smith G. Hygiene hypothesis in inflammatory bowel disease: A critical review of the literature. *World Journal of Gastroenterology*. [Online] 2008;14(2): 165–173. Available from: doi:10.3748/wjg.14.165
250. Klein A, Eliakim R. Non steroidal anti-inflammatory drugs and inflammatory bowel disease. *Pharmaceuticals*. [Online] 2010;3(4): 1084–1092. Available from: doi:10.3390/ph3041084
251. Mawdsley JE, Rampton DS. Psychological stress in IBD: New insights into pathogenic and therapeutic implications. *Gut*. [Online] 2005;54(10): 1481–1491. Available from: doi:10.1136/gut.2005.064261
252. Ledder O. Antibiotics in inflammatory bowel diseases: Do we know what we're doing? *Translational Pediatrics*. [Online] 2019;8(1): 42–55. Available from: doi:10.21037/tp.2018.11.02
253. Lichtenstein GR. Diet and inflammatory bowel disease. *Gastroenterology and Hepatology*. 2019;15(3): 127.
254. Bilski J, Brzozowski B, Mazur-Bialy A, Sliwowski Z, Brzozowski T. The role of physical exercise in inflammatory bowel disease. *BioMed Research International*. [Online] Hindawi Publishing Corporation; 2014;2014. Available from: doi:10.1155/2014/429031
255. C M, M A, J S, A T, F C, N B, et al. Unstable composition of the fecal microbiota in ulcerative colitis during clinical remission. *The American Journal of Gastroenterology*. 2008;103(3): 643–648.
256. Sartor RB. Microbial Influences in Inflammatory Bowel Diseases. *Gastroenterology*. [Online] 2008;134(2): 577–594. Available from: doi:10.1053/j.gastro.2007.11.059
257. Sepehri S, Kotlowski R, Bernstein CN, Krause DO. Phylogenetic analysis of inflammatory bowel disease associated Escherichia coli and the FimH virulence determinant. *Inflammatory Bowel Diseases*. [Online] 2009;15(11): 1737–1745. Available from: doi:10.1002/ibd.20966
258. Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut*. [Online] 2006;55(2): 205–211. Available from: doi:10.1136/gut.2005.073817
259. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, Gratadoux JJ, et al. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 2008;105(43): 16731–16736. Available from: doi:10.1073/pnas.0804812105
260. Zhang YZ, Li YY. Inflammatory bowel disease: Pathogenesis. *World Journal of Gastroenterology*. [Online] 2014;20(1): 91–99. Available from: doi:10.3748/wjg.v20.i1.91
261. Vavricka SR, Schoepfer A, Scharl M, Lakatos PL, Navarini A, Rogler G. Extraintestinal manifestations of inflammatory bowel disease. *Inflammatory Bowel Diseases*. [Online] 2015;21(8): 1982–1992. Available from: doi:10.1097/MIB.0000000000000392
262. Lamb CA, Kennedy NA, Raine T, Hendy PA, Smith PJ, Limdi JK, et al. British Society of

Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. *Gut*. [Online] 2019; gutjnl-2019-318484. Available from: doi:10.1136/gutjnl-2019-318484
263. Tontini GE, Vecchi M, Pastorelli L, Neurath MF, Neumann H. Differential diagnosis in inflammatory bowel disease colitis: State of the art and future perspectives. *World Journal of Gastroenterology*. [Online] 2015;21(1): 21–46. Available from: doi:10.3748/wjg.v21.i1.21
264. Nikolaus S, Schreiber S. Diagnostics of Inflammatory Bowel Disease. *Gastroenterology*. [Online] 2007;133(5): 1670–1689. Available from: doi:10.1053/j.gastro.2007.09.001
265. Stange EF, Travis SPL, Vermeire S, Beglinger C, Kupcinskas L, Geboes K, et al. European evidence based consensus on the diagnosis and management of Crohn's disease: Definitions and diagnosis. *Gut*. [Online] 2006;55(SUPPL. 1): 1–15. Available from: doi:10.1136/gut.2005.081950a
266. Wehkamp J, Götz M, Herrlinger K, Steurer W, Stange EF. Inflammatory Bowel Disease. *Deutsches Arzteblatt International*. [Online] 2016;113(5): 72–81. Available from: doi:10.3238/arztebl.2016.0072
267. NICE. *NICE guidelines: Faecal calprotectin diagnostic tests for inflammatory diseases of the bowel*.
268. Moum B, Ekbom A, Vatn MH, Aadland E, Sauar J, Lygren I, et al. Inflammatory bowel disease: Re-evaluation of the diagnosis in a prospective population based study in south eastern Norway. *Gut*. [Online] 1997;40(3): 328–332. Available from: doi:10.1136/gut.40.3.328
269. Henderson P, Anderson N, Wilson DC. The Diagnostic Accuracy of Fecal Calprotectin During the Investigation of Suspected Pediatric Inflammatory Bowel Disease: A Systematic Review and Meta-Analysis. *AJG*. 2013;109(5): 637–645.
270. Crispin A, Birkner B, Munte A, Nusko G, Mansmann U. Process quality and incidence of acute complications in a series of more than 230000 outpatient colonoscopies. *Endoscopy*. [Online] 2009;41(12): 1018–1025. Available from: doi:10.1055/s-0029-1215214
271. Frolkis AD, Dykeman J, Negrón ME, Debruyn J, Jette N, Fiest KM, et al. Risk of surgery for inflammatory bowel diseases has decreased over time: A systematic review and meta-analysis of population-based studies. *Gastroenterology*. [Online] Elsevier, Inc; 2013;145(5): 996–1006. Available from: doi:10.1053/j.gastro.2013.07.041
272. Satsangi J, Silverberg MS, Vermeire S, Colombel JF. The Montreal classification of inflammatory bowel disease: Controversies, consensus, and implications. *Gut*. [Online] 2006;55(6): 749–753. Available from: doi:10.1136/gut.2005.082909
273. Levine A, Griffiths A, Markowitz J, Wilson DC, Turner D, Russell RK, et al. Pediatric modification of the Montreal classification for inflammatory bowel disease: The Paris classification. *Inflammatory Bowel Diseases*. [Online] 2011;17(6): 1314–1321. Available from: doi:10.1002/ibd.21493
274. Lamb CA, Kennedy NA, Raine T, Hendy PA, Smith PJ, Limdi JK, et al. British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. *Gut*. [Online] 2019;68: s1–s106. Available from: doi:10.1136/gutjnl-2019-318484
275. Walsh AJ, Bryant R V., Travis SPL. Current best practice for disease activity assessment in IBD. *Nature Reviews Gastroenterology and Hepatology*. [Online] Nature Publishing

Group; 2016;13(10): 567–579. Available from: doi:10.1038/nrgastro.2016.128
276. Regueiro M, Rodemann J, Kip KE, Saul M, Swoger J, Baidoo L, et al. Physician assessment of ulcerative colitis activity correlates poorly with endoscopic disease activity. *Inflammatory Bowel Diseases*. [Online] 2011;17(4): 1008–1014. Available from: doi:10.1002/ibd.21445
277. Gisbert JP, González-Lama Y, Maté J. Role of biological markers in inflammatory bowel disease. *Gastroenterology and Hepatology*. 2007;30(3): 117–129.
278. Vermeire S, Van Assche G, Rutgeerts P. Laboratory markers in IBD: Useful, magic, or unnecessary toys? *Gut*. [Online] 2006;55(3): 426–431. Available from: doi:10.1136/gut.2005.069476
279. Barnes EL, Burakoff R. New Biomarkers for Diagnosing Inflammatory Bowel Disease and Assessing Treatment Outcomes. *Physiology & behavior*. [Online] 2016;176(1): 139–148. Available from: doi:10.1016/j.physbeh.2017.03.040
280. Iijima H, Shinzaki S, Tsujii M, Takehara T. Biomarkers in the diagnosis and therapy of inflammatory bowel diseases. *Nihon rinsho. Japanese journal of clinical medicine*. [Online] 2012;70(5): 869–873. Available from: doi:10.1053/j.gastro.2010.11.058.The
281. Kolho KL, Pessia A, Jaakkola T, de Vos WM, Velagapudi V. Faecal and Serum Metabolomics in Paediatric Inflammatory Bowel Disease. *Journal of Crohn's & colitis*. [Online] 2017;11(3): 321–334. Available from: doi:10.1093/ecco-jcc/jjw158
282. Freeman K, Willis BH, Fraser H, Taylor-Phillips S, Clarke A. Faecal calprotectin to detect inflammatory bowel disease: A systematic review and exploratory meta-analysis of test accuracy. *BMJ Open*. [Online] 2019;9(3): 1–11. Available from: doi:10.1136/bmjopen-2018-027428
283. Konikoff MR, Denson LA. Role of fecal calprotectin as a biomarker of intestinal inflammation in inflammatory bowel disease. *Inflammatory Bowel Diseases*. [Online] 2006;12(6): 524–534. Available from: doi:10.1097/00054725-200606000-00013
284. Khaki-Khatibi F, Qujeq D, Kashifard M, Moein S, Maniati M, Vaghari-Tabari M. Calprotectin in inflammatory bowel disease. *Clinica Chimica Acta*. [Online] 2020;510: 556–565. Available from: doi:10.1016/j.cca.2020.08.025
285. NICE - National Institute for Health and Care Excellence. Faecal calprotectin diagnostic tests for inflammatory diseases of the bowel. 2013;(October): 1–55. Available from: http://nice.org.uk/guidance/dg11
286. Waugh N, Cummins E, Royle P, Kandala NB, Shyangdan D, Arasaradnam R, et al. Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel diseases: Systematic review and economic evaluation. *Health Technology Assessment*. [Online] 2013;17(55). Available from: doi:10.3310/hta17550
287. Tibble JA, Sigthorsson G, Foster R, Forgacs I, Bjarnason I. Use of surrogate markers of inflammation and Rome criteria to distinguish organic from nonorganic intestinal disease. *Gastroenterology*. [Online] 2002;123(2): 450–460. Available from: doi:10.1053/gast.2002.34755
288. Gisbert JP, McNicholl AG. Questions and answers on the role of faecal calprotectin as a biological marker in inflammatory bowel disease. *Digestive and Liver Disease*. [Online] 2009;41(1): 56–66. Available from: doi:10.1016/j.dld.2008.05.008
289. Bjarnason I. The use of fecal calprotectin in inflammatory bowel disease. *Gastroenterology and Hepatology*. 2017;13(1): 53–56.
290. Lehmann FS, Burri E, Beglinger C. The role and utility of faecal markers in inflammatory bowel disease. *Therapeutic Advances in Gastroenterology*. [Online]

2015;8(1): 23–36. Available from: doi:10.1177/1756283X14553384
291. Storr M, Vogel HJ, Schicho R. Metabolomics : is it useful for IBD ? *Curr Opin Gastroenterol*. [Online] 2014;29(4): 378–383. Available from: doi:10.1097/MOG.0b013e328361f488.Metabolomics
292. Schicho R, Shaykhutdinov R, Ngo J, Nazyrova A, Schneider C, Panaccione R, et al. Quantitative metabolomic profiling of serum, plasma, and urine by 1H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals. *Journal of Proteome Research*. [Online] 2012;11(6): 3344–3357. Available from: doi:10.1021/pr300139q
293. Dawiskiba T, Deja S, Mulak A, Zabek A, Jawień E, Pawełka D, et al. Serum and urine metabolomic fngerprinting in diagnostics of inflammatory bowel diseases. *World Journal of Gastroenterology*. [Online] 2014;20(1): 163–174. Available from: doi:10.3748/wjg.v20.i1.163
294. Williams HRT, Willsmore JD, Cox IJ, Walker DG, Cobbold JFL, Taylor-Robinson SD, et al. Serum metabolic profiling in inflammatory bowel disease. *Digestive Diseases and Sciences*. [Online] 2012;57(8): 2157–2165. Available from: doi:10.1007/s10620-012-2127-2
295. Ooi M, Nishiumi S, Yoshie T, Shiomi Y, Kohashi M, Fukunaga K, et al. GC/MS-based profiling of amino acids and TCA cycle-related molecules in ulcerative colitis. *Inflammation Research*. [Online] 2011;60(9): 831–840. Available from: doi:10.1007/s00011-011-0340-7
296. HR W, IJ C, DG W, BV N, VM P, SE M, et al. Characterization of inflammatory bowel disease with urinary metabolic profiling. *The American Journal of Gastroenterology*. 2009;104(6): 1435–1444.
297. Stephens NS, Siffledeen J, Su X, Murdoch TB, Fedorak RN, Slupsky CM. Urinary NMR metabolomic profiles discriminate inflammatory bowel disease from healthy. *Journal of Crohn's and Colitis*. [Online] European Crohn's and Colitis Organisation; 2013;7(2): e42–e48. Available from: doi:10.1016/j.crohns.2012.04.019
298. Ahmed I, Greenwood R, Costello B, Ratcliffe N, Probert CS. Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease. *Alimentary Pharmacology and Therapeutics*. [Online] 2016;43(5): 596–611. Available from: doi:10.1111/apt.13522
299. Van Malderen K, De Winter BY, De Man JG, De Schepper HU, Lamote K. Volatomics in inflammatory bowel disease and irritable bowel syndrome. *EBioMedicine*. [Online] Elsevier B.V.; 2020;54: 102725. Available from: doi:10.1016/j.ebiom.2020.102725
300. Duboc H, Rajca S, Rainteau D, Benarous D, Maubert MA, Quervain E, et al. Connecting dysbiosis, bile-acid dysmetabolism and Gut inflammation in inflammatory bowel diseases. *Gut*. [Online] 2013;62(4): 531–539. Available from: doi:10.1136/gutjnl-2012-302578
301. Huttenhower C, Kostic AD, Xavier RJ. Inflammatory bowel disease as a model for translating the microbiome. *Immunity*. [Online] Elsevier Inc.; 2014;40(6): 843–854. Available from: doi:10.1016/j.immuni.2014.05.013
302. Jacobs JP, Goudarzi M, Singh N, Tong M, McHardy IH, Ruegger P, et al. A Disease-Associated Microbial and Metabolomics State in Relatives of Pediatric Inflammatory Bowel Disease Patients. *Cmgh*. [Online] Elsevier Inc; 2016;2(6): 750–766. Available from: doi:10.1016/j.jcmgh.2016.06.004
303. Melnik A V., Da Silva RR, Hyde ER, Aksenov AA, Vargas F, Bouslimani A, et al. Coupling

Targeted and Untargeted Mass Spectrometry for Metabolome-Microbiome-Wide Association Studies of Human Fecal Samples. *Analytical Chemistry*. [Online] 2017;89(14): 7549–7559. Available from: doi:10.1021/acs.analchem.7b01381
304. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature Microbiology*. [Online] Springer US; 2019;4(2): 293–305. Available from: doi:10.1038/s41564-018-0306-4
305. Lavelle A, Sokol H. Gut microbiota-derived metabolites as key actors in inflammatory bowel disease. *Nature Reviews Gastroenterology and Hepatology*. [Online] Springer US; 2020;17(4): 223–237. Available from: doi:10.1038/s41575-019-0258-z
306. Marchesi JR, Holmes E, Khan F, Kochhar S, Scanlan P, Shanahan F, et al. Rapid and noninvasive metabonomic characterization of inflammatory bowel disease. *Journal of Proteome Research*. [Online] 2007;6(2): 546–551. Available from: doi:10.1021/pr060470d
307. Bjerrum JT, Wang Y, Hao F, Coskun M, Ludwig C, Günther U, et al. Metabonomics of human fecal extracts characterize ulcerative colitis, Crohn's disease and healthy individuals. *Metabolomics*. [Online] 2015;11(1): 122–133. Available from: doi:10.1007/s11306-014-0677-3
308. Le Gall G, Guttula K, Kellingray L, Tett AJ, Hoopen R ten, Kemsley EK, et al. Correction: Metabolite quantification of faecal extracts from colorectal cancer patients and healthy controls. *Oncotarget*. [Online] 2019;10(17): 33278–33289. Available from: doi:10.18632/oncotarget.26766
309. Braun A, Treede I, Gotthardt D, Tietje A, Zahn A, Ruhwald R, et al. Alterations of phospholipid concentration and species composition of the intestinal mucus barrier in ulcerative colitis: A clue to pathogenesis. *Inflammatory Bowel Diseases*. [Online] 2009;15(11): 1705–1720. Available from: doi:10.1002/ibd.20993
310. Fischbeck A, Leucht K, Frey-Wagner I, Bentz S, Pesch T, Kellermeier S, et al. Sphingomyelin induces cathepsin D-mediated apoptosis in intestinal epithelial cells and increases inflammation in DSS colitis. *Gut*. [Online] 2011;60(1): 55–65. Available from: doi:10.1136/gut.2009.201988
311. Abdel Hadi L, Di Vito C, Riboni L. Fostering Inflammatory Bowel Disease: Sphingolipid Strategies to Join Forces. *Mediators of Inflammation*. [Online] Hindawi Publishing Corporation; 2016;2016(Figure 1). Available from: doi:10.1155/2016/3827684
312. Cameron S. *Direct from Sample Metabolomics Using Rapid Evaporative Ionisation Mass Spectrometry*. [Online] NIH Metabolomics Interest Group Webinar Series; 2018. Available from: https://www.youtube.com/watch?v=M5tV7P-pptk
313. Perdones-Montero A. *peak_picking_CO2_LA_REIMS_automatic_pipeline*. 2018.
314. Perdones-Montero A. *REIMS Peak picking pipeline (NO IMAGING)*. 2019.
315. Sztáray J, Memboeuf A, Drahos L, Vékey K. Leucine enkephalin - A mass spectrometry standard. *Mass Spectrometry Reviews*. [Online] 2011;30(2): 298–320. Available from: doi:10.1002/mas.20279
316. Lever J, Krzywinski M, Altman N. Points of Significance: Principal component analysis. *Nature Methods*. [Online] Nature Publishing Group; 2017;14(7): 641–642. Available from: doi:10.1038/nmeth.4346
317. Worley B, Powers R. Multivariate Analysis in Metabolomics. *Current Metabolomics*. [Online] NIH Public Access; 2013;1(1): 92–107. Available from: doi:10.2174/2213235X11301010092 [Accessed: 18th May 2018]

318. Paraskevaidi M, Cameron SJS, Whelan E, Bowden S, Tzafetas M, Mitra A, et al. Laser-assisted rapid evaporative ionisation mass spectrometry (LA-REIMS) as a metabolomics platform in cervical cancer screening. *EBioMedicine*. [Online] Elsevier B.V.; 2020;60: 103017. Available from: doi:10.1016/j.ebiom.2020.103017
319. Junhua Zhu RBC. Formation and decompositions of chloride adduct ions, [M + Cl]−, in negative ion electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry*. 11(11): 932–941.
320. Patiny L, Borel A. *ChemCalc: a building block for tomorrow's chemical infrastructure.* [Online] Available from: doi:10.1021/ci300563h
321. *SciFInder - Chemical Abstracts Service*.
322. Brenton AG, Godfrey AR. Accurate mass measurement: Terminology and treatment of data. *Journal of the American Society for Mass Spectrometry*. [Online] Elsevier Inc.; 2010;21(11): 1821–1835. Available from: doi:10.1016/j.jasms.2010.06.006
323. Waters. *Mass Accuracy and Resolution*. [Online] Available from: https://www.waters.com/waters/en_US/Mass-Accuracy-and-Resolution/nav.htm?cid=10091028&locale=en_US
324. Chernushevich I V., Loboda A V., Thomson BA. An introduction to quadrupole-time-of-flight mass spectrometry. *Journal of Mass Spectrometry*. [Online] 2001;36(8): 849–865. Available from: doi:10.1002/jms.207
325. Quenzer TL, Robinson JM, Bolanios B, Milgram E, Greig MJ. The effect of increasing mass accuracy for unambiguous identification of compounds. *Automated accurate mass analysis using FTICR mass spectrometry*. 2002.
326. Sarafian MH, Gaudin M, Lewis MR, Martin FP, Holmes E, Nicholson JK, et al. Objective set of criteria for optimization of sample preparation procedures for ultra-high throughput untargeted blood plasma lipid profiling by ultra performance liquid chromatography-mass spectrometry. *Analytical Chemistry*. [Online] 2014;86(12): 5766–5774. Available from: doi:10.1021/ac500317c
327. Centre NP. *National Phenome Centre*. [Online] Available from: https://phenomecentre.org/ [Accessed: 14th May 2021]
328. Cameron SJS, Bodai Z, Temelkuran B, Perdones-Montero A, Bolt F, Burke A, et al. Utilisation of Ambient Laser Desorption Ionisation Mass Spectrometry (ALDI-MS) Improves Lipid-Based Microbial Species Level Identification. *Scientific Reports*. [Online] 2019;9(1): 1–8. Available from: doi:10.1038/s41598-019-39815-w
329. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scandinavian Journal of Gastroenterology*. [Online] 1997;32(9): 920–924. Available from: doi:10.3109/00365529709011203
330. Strittmatter N. *Development of novel mass spctrometric methods for the characterisation and identification of microorganisms*. Imperial College London; 2016.
331. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in1H NMR metabonomics. *Analytical Chemistry*. [Online] 2006;78(13): 4281–4290. Available from: doi:10.1021/ac051632c
332. Kohl SM, Klein MS, Hochrein J, Oefner PJ, Spang R, Gronwald W. State-of-the art data normalization methods improve NMR-based metabolomic analysis. *Metabolomics*. [Online] 2012;8: 146–160. Available from: doi:10.1007/s11306-011-0350-z
333. Bazanella M, Maier T V., Clavel T, Lagkouvardos I, Lucio M, Maldonado-Gòmez MX, et al. Randomized controlled trial on the impact of early-life intervention with

bifidobacteria on the healthy infant fecal microbiota and metabolome. *American Journal of Clinical Nutrition*. [Online] 2017;106(5): 1274–1286. Available from: doi:10.3945/ajcn.117.157529
334. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut*. [Online] 2016;65(1): 57–62. Available from: doi:10.1136/gutjnl-2015-309618
335. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. *Science*. [Online] 2016;352(6285): 560–564. Available from: doi:10.1126/science.aad3503
336. Tigchelaar EF, Bonder MJ, Jankipersadsing SA, Fu J, Wijmenga C, Zhernakova A. Gut microbiota composition associated with stool consistency. *Gut*. [Online] 2016;65(3): 540–542. Available from: doi:10.1136/gutjnl-2015-310328
337. Chughtai K, Heeren RMA. Mass Spectrometric Imaging for Biomedical Tissue Analysis - Chemical Reviews (ACS Publications). *Chemical reviews*. [Online] 2011;110(5): 3237–3277. Available from: doi:10.1021/cr100012c.Mass
338. Lee JK, Liles EG, Bent S, R. TL, Corley DA. Accuracy of Fecal Immunochemical Tests for Colorectal Cancer: Systematic Review and Meta-analysis. *Ann Intern Med*. [Online] 2014;160(3). Available from: doi:10.7326/M13-1484.Accuracy
339. Imperiale TF, Gruber RN, Stump TE, Emmett TW. Performance Characteristics of Fecal Immunochemical Tests for Colorectal Cancer and Advanced Adenomatous Polyps. *Annals of Internal Medicine*. [Online]. Available from: doi:https://doi.org/10.7326/M18-2390
340. Huq T. *BSc thesis: Direct from sample Rapid Evaporative Ionization Mass Spectrometry (REIMS) as a screening tool for the early detection of colorectal cancer- a pilot prospective observational study*. Imperial College London; 2017.
341. Niedermaier T, Tikk K, Gies A, Bieck S, Brenner H. Sensitivity of Fecal Immunochemical Test for Colorectal Cancer Detection Differs According to Stage and Location. *Clinical Gastroenterology and Hepatology*. [Online] 2020; Available from: doi:https://doi.org/10.1016/j.cgh.2020.01.025
342. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. [Online] Nature Publishing Group; 2009;458(7239): 719–724. Available from: doi:10.1038/nature07943
343. Ecker J, Benedetti E, Kindt ASD, Höring M, Perl M, Machmüller AC, et al. The Colorectal Cancer Lipidome: Identification of a Robust Tumor-Specific Lipid Species Signature. *Gastroenterology*. [Online] 2021;161(3): 910-923.e19. Available from: doi:10.1053/j.gastro.2021.05.009
344. Strittmatter N, Rebec M, Jones EA, Golf O, Abdolrasouli A, Balog J, et al. Characterization and identification of clinically relevant microorganisms using rapid evaporative ionization mass spectrometry. *Analytical Chemistry*. [Online] 2014;86(13): 6555–6562. Available from: doi:10.1021/ac501075f
345. Dutta M, Joshi M, Srivastava S, Lodh I, Chakravarty B, Chaudhury K. A metabonomics approach as a means for identification of potential biomarkers for early diagnosis of endometriosis. *Molecular BioSystems*. [Online] 2012;8(12): 3281–3287. Available from: doi:10.1039/c2mb25353d
346. Ikeda A, Nishiumi S, Shinohara M, Yoshie T, Hatano N, Okuno T, et al. Serum metabolomics as a novel diagnostic approach for gastrointestinal cancer. *Biomedical*

*Chromatography*. 2011;26: 548–558.
347. O'Brien R, Ishwaran H. A random forests quantile classifier for class imbalanced data. *Pattern Recognition*. [Online] Elsevier Ltd; 2019;90: 232–249. Available from: doi:10.1016/j.patcog.2019.01.036
348. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget Stool DNA Testing for Colorectal-Cancer Screening. *New England Journal of Medicine*. [Online] Massachusetts Medical Society; 2014;370(14): 1287–1297. Available from: doi:10.1056/NEJMoa1311194 [Accessed: 8th May 2018]
349. Ho KMA, Banerjee A, Lawler M, Rutter MD, Lovat LB. Predicting endoscopic activity recovery in England after COVID-19: a national analysis. *The Lancet Gastroenterology and Hepatology*. [Online] Elsevier Ltd; 2021;6(5): 381–390. Available from: doi:10.1016/S2468-1253(21)00058-3
350. Turroni S, Brigidi P, Cavalli A, Candela M. Microbiota–Host Transgenomic Metabolism, Bioactive Molecules from the Inside. *Journal of Medicinal Chemistry*. [Online] American Chemical Society; 2018;61(1): 47–61. Available from: doi:10.1021/acs.jmedchem.7b00244 [Accessed: 24th May 2018]
351. Bojanova DP, Bordenstein SR. Fecal Transplants: What Is Being Transferred? *PLoS biology*. [Online] Public Library of Science; 2016;14(7): e1002503. Available from: doi:10.1371/journal.pbio.1002503 [Accessed: 24th May 2018]
352. Okumura R, Takeda K. Roles of intestinal epithelial cells in the maintenance of gut homeostasis. *Experimental & Molecular Medicine*. [Online] Nature Publishing Group; 2017;49(5): e338. Available from: doi:10.1038/emm.2017.20 [Accessed: 24th May 2018]
353. Bullen TF, Forrest S, Campbell F, Dodson AR, Hershman MJ, Pritchard DM, et al. Characterization of epithelial cell shedding from human small intestine. *Laboratory Investigation*. [Online] Nature Publishing Group; 2006;86(10): 1052–1063. Available from: doi:10.1038/labinvest.3700464 [Accessed: 18th May 2018]
354. Currie E, Schulze A, Zechner R, Walther TC, Farese R V. Cellular fatty acid metabolism and cancer. *Cell Metabolism*. [Online] Elsevier Inc.; 2013;18(2): 153–161. Available from: doi:10.1016/j.cmet.2013.05.017
355. DS W, YD F, A M, AC G, K L. *HMDB 4.0 — The Human Metabolome Database*. https://hmdb.ca/metabolites/HMDB0000638.
356. Robert C Murphy. Chapter 1: Fatty Acids. *Tandem Mass Spectrometry of Lipids: Molecular Analysis of Complex Lipids*. [Online] 2014. p. 1–39. Available from: doi:10.1039/9781782626350-00001
357. Brown DG, Rao S, Weir TL, O'Malia J, Bazan M, Brown RJ, et al. Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. *Cancer and Metabolism*. [Online] Cancer & Metabolism; 2016;4(1): 1–12. Available from: doi:10.1186/s40170-016-0151-y
358. Goedert JJ, Sampson JN, Moore SC, Xiao Q, Xiong X, Hayes RB, et al. Fecal metabolomics: Assay performance and association with colorectal cancer. *Carcinogenesis*. [Online] 2014;35(9): 2089–2096. Available from: doi:10.1093/carcin/bgu131
359. Fauser JK, Matthews GM, Cummins AG, Howarth GS. Induction of apoptosis by the medium-chain length fatty acid lauric acid in colon cancer cells due to induction of oxidative stress. *Chemotherapy*. 2013;59(3): 214–224.
360. Weng WH, Leung WH, Pang YJ, Hsu HH. Lauric acid can improve the sensitization of

Cetuximab in KRAS/BRAF mutated colorectal cancer cells by retrievable microRNA-378 expression. *Oncology Reports*. [Online] 2016;35(1): 107–116. Available from: doi:10.3892/or.2015.4336
361. Lappano R, Sebastiani A, Cirillo F, Rigiracciolo DC, Galli GR, Curcio R, et al. The lauric acid-activated signaling prompts apoptosis in cancer cells. *Cell Death Discovery*. [Online] 2017;3(1): 1–9. Available from: doi:10.1038/cddiscovery.2017.63
362. Moe MK, Strøm MB, Jensen E, Claeys M. Negative electrospray ionization low-energy tandem mass spectrometry of hydroxylated fatty acids: A mechanistic study. *Rapid Communications in Mass Spectrometry*. [Online] 2004;18(15): 1731–1740. Available from: doi:10.1002/rcm.1545
363. Chen L, Chen H, Li Y, Li L, Qiu Y, Ren J. Endocannabinoid and ceramide levels are altered in patients with colorectal cancer. *Oncology Reports*. [Online] 2015;34(1): 447–454. Available from: doi:10.3892/or.2015.3973
364. *Bowel cancer statistics | Cancer Research UK*. [Online] Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/i [Accessed: 18th May 2018]
365. Parkin DM, Boyd L. 8. Cancers attributable to overweight and obesity in the UK in 2010. *British Journal of Cancer*. [Online] 2011;105(S2): S34–S37. Available from: doi:10.1038/bjc.2011.481 [Accessed: 18th May 2018]
366. Vrabie R, Kane S. Noninvasive markers of disease activity in inflammatory bowel disease. *Gastroenterology and Hepatology*. 2014;10(9): 576–584.
367. Turvill J, Aghahoseini A, Sivarajasingham N, Abbas K, Choudhry M, Polyzois K, et al. Faecal calprotectin in patients with suspected colorectal cancer: A diagnostic accuracy study. *British Journal of General Practice*. [Online] 2016;66(648): e499–e506. Available from: doi:10.3399/bjgp16X685645
368. Maaser C, Sturm A, Vavricka SR, Kucharzik T, Fiorino G, Annese V, et al. ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: Initial diagnosis, monitoring of known IBD, detection of complications. *Journal of Crohn's and Colitis*. [Online] 2019;13(2): 144–164. Available from: doi:10.1093/ecco-jcc/jjy113
369. Hanauer SB. Advances in IBD Fecal Biomarkers for the. 2011;7(6): 396–398.
370. Iacucci M, De Silva S, Ghosh S. Mesalazine in inflammatory bowel disease: A trendy topic once again? *Canadian Journal of Gastroenterology*. [Online] 2010;24(2): 127–133. Available from: doi:10.1155/2010/586092
371. Ak A, D Z, Gordon M. Oral 5-aminosalicylic acid for maintenance of medically-induced remission in Crohn's disease. *Cochrane Database of Systematic Reviews*. 2016;9.
372. Moja L, Danese S, Fiorino G, Del Giovane C, Bonovas S. Systematic review with network meta-analysis: Comparative efficacy and safety of budesonide and mesalazine (mesalamine) for Crohn's disease. *Alimentary Pharmacology and Therapeutics*. [Online] 2015;41(11): 1055–1065. Available from: doi:10.1111/apt.13190
373. WC L, Y W, JK M. Aminosalicylates for induction of remission or response in Crohn's disease. *Cochrane Database of Systematic Reviews*. 2016;7.
374. Xu J, Chen N, Wu Z, Song Y, Zhang Y, Wu N, et al. 5-Aminosalicylic Acid Alters the Gut Bacterial Microbiota in Patients With Ulcerative Colitis. *Frontiers in Microbiology*. [Online] 2018;9(JUN): 1–13. Available from: doi:10.3389/fmicb.2018.01274
375. Morrow T, Felcone LH. Defining the difference: What Makes Biologics Unique. *Biotechnology healthcare*. [Online] 2004;1(4): 24–29. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/23393437%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3564302
376. Chan HC hey, Ng SC. Emerging biologics in inflammatory bowel disease. *Journal of Gastroenterology*. [Online] Springer Japan; 2017;52(2): 141–150. Available from: doi:10.1007/s00535-016-1283-0
377. Estevinho MM, Rocha C, Correia L, Lago P, Ministro P, Portela F, et al. Features of Fecal and Colon Microbiomes Associate With Responses to Biologic Therapies for Inflammatory Bowel Diseases: A Systematic Review. *Clinical Gastroenterology and Hepatology*. 2020;
378. A Golay, E Bobbioni. The role of dietary fat in obesity. *Int J Obes Relat Metab DIsord*. 1997;21.
379. Porta Mi. *A dictionary of Epidemiology (6 ed.)*. Oxford University Press; 2016.
380. Bolte LA, Klaassen MAY, Collij V, Vila AV, Fu J, van der Meulen TA, et al. Patient attitudes towards faecal sampling for gut microbiome studies and clinical care reveal positive engagement and room for improvement. *PLoS ONE*. [Online] 2021;16(4 April): 1–13. Available from: doi:10.1371/journal.pone.0249405
381. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget Stool DNA Testing for Colorectal-Cancer Screening. *New England Journal of Medicine*. [Online] 2014;370(14): 1287–1297. Available from: doi:10.1056/nejmoa1311194
382. Xiao JF, Zhou B, Ressom HW. Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *TrAC - Trends in Analytical Chemistry*. [Online] 2012;32: 1–14. Available from: doi:10.1016/j.trac.2011.08.009

# Appendix

## Appendix 1: Permissions to reprint

Permissions to reprint were obtained from the following publications and can be found attached to this document:

1. Cameron SJS, et al. Sample Preparation Free Mass Spectrometry Using Laser-Assisted Rapid Evaporative Ionization Mass Spectrometry: Applications to Microbiology, Metabolic Biofluid Phenotyping, and Food Authenticity. Journal of the American Society for Mass Spectrometry. 2021; (151)

2. Zhang A, Sun H, Wang P, Han Y, Xijung W. Modern analytical techniques in metabolomics analysis. Analyst. 2012;(2)

CCC Marketplace™



| | | | |
|---|---|---|---|
| **Order Date** | 10-Aug-2021 | **Type of Use** | Republish in a thesis/dissertation |
| **Order License ID** | 1139569-1 | | |
| **ISSN** | 0003-2654 | **Publisher** | ROYAL SOCIETY OF CHEMISTRY |
| | | **Portion** | Image/photo/illustration |

## LICENSED CONTENT

| | | | |
|---|---|---|---|
| **Publication Title** | The Analyst | **Rightsholder** | Royal Society of Chemistry |
| **Article Title** | Modern analytical techniques in metabolomics analysis. | **Publication Type** | Journal |
| | | **Start Page** | 293 |
| **Author/Editor** | ROYAL SOCIETY OF CHEMISTRY (GREAT BRITAIN), SOCIETY OF PUBLIC ANALYSTS (GREAT BRITAIN), SOCIETY OF PUBLIC ANALYSTS AND OTHER ANALYTICAL CH, SOCIETY FOR ANALYTICAL CHEMISTRY., CHEMICAL SOCIETY (GREAT BRITAIN), SOCIETY OF PUBLIC ANALYSTS (GREAT BRITAIN). | **End Page** | 300 |
| | | **Issue** | 2 |
| | | **Volume** | 137 |
| **Date** | 01/01/1876 | | |
| **Language** | English | | |
| **Country** | United Kingdom of Great Britain and Northern Ireland | | |

## REQUEST DETAILS

| | | | |
|---|---|---|---|
| **Portion Type** | Image/photo/illustration | **Distribution** | Worldwide |
| **Number of images / photos / illustrations** | 1 | **Translation** | Original language of publication |
| **Format (select all that apply)** | Print, Electronic | **Copies for the disabled?** | No |
| | | **Minor editing privileges?** | No |
| **Who will republish the content?** | Author of requested content | **Incidental promotional use?** | No |
| **Duration of Use** | Life of current edition | **Currency** | GBP |
| **Lifetime Unit Quantity** | Up to 499 | | |
| **Rights Requested** | Main product | | |

## NEW WORK DETAILS

| | | | |
|---|---|---|---|
| **Title** | PhD Thesis | **Institution name** | Imperial College London |
| **Instructor name** | Petra Paizs | **Expected presentation date** | 2021-09-28 |

## ADDITIONAL DETAILS

| | | | |
|---|---|---|---|
| **Order reference number** | Thesis | **The requesting person / organization to appear on the license** | Petra Paizs |

## REUSE CONTENT DETAILS

| | | | |
|---|---|---|---|
| **Title, description or numeric reference of the portion(s)** | Analytical Workflow | **Title of the article/chapter the portion is from** | Modern analytical techniques in metabolomics analysis. |
| **Editor of portion(s)** | Zhang, Aihua; Sun, Hui; Wang, Ping; Han, Ying; Wang, Xijun | **Author of portion(s)** | Zhang, Aihua; Sun, Hui; Wang, Ping; Han, Ying; Wang, Xijun |
| **Volume of serial or monograph** | 137 | **Issue, if republishing an article from a serial** | 2 |
| **Page or page range of portion** | 293-300 | **Publication date of portion** | 2012-01-01 |

## CCC Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

   3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

   3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

   3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of

use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4. In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is

not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

8.1. User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

8.3. The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

8.5. The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to support@copyright.com.

v 1.1

# Publication list

1. Peer-reviewed publications

Sam E Mason, Eftychios Manoli, Liam Poynter, James Alexander, Petra Paizs, Afeez Adebesin, Robert Goldin, Ara Darzi, Zoltan Takats, James Kinross. Mass spectrometry transanal minimally invasive surgery (MS-TAMIS) to promote organ preservation in rectal cancer.
Surgical Endoscopy 2020.

Sam E Mason, Eftychios Manoli, James Alexander, Liam Poynter, Lauren Ford, Petra Paizs, Afeez Adebesin, James McKenzie, Rob Goldin, Ara Darzi, Zoltan Takats, James M Kinross. Lipidomic Profiling of Colorectal Lesions for Real-Time Tissue Recognition and Risk-Stratification using Rapid Evaporative Ionisation Mass Spectrometry Annals of Surgery, 2021.

2. Conference presentations

2.1.Poster

Lauren L Ford[1]; Daniel Simon[1]; Alvaro Perdones Montero[1]; Natasha Jiwa[1]; Yuchen Xiang[1]; Eftychios Manoli[1]; Vincen Wu[1]; James Higginson[1]; Petra Paizs[1]; Sam Mason[1]; Stefania Maneta-Stavrakaki[1]; Daniel Mcgill[1]; James Kinross[1]; Thomas Hankemeier[2]; Zoltan Takats. Direct from swab high throughput mass spectrometry for detection of SARS-COV-2. ASMS, USA, 2021.

Petra Paizs, Monika Widlak, Alvaro Perdones-Montero, Maria Sani, Lauren Ford, James Alexander, Simon Cameron, Ramesh Arasadnam, James Kinross, Zoltan Takats. High-throughput fecal metabolic profiling for the early detection of colorectal cancer using a direct mass spectrometry assay. AACR, USA, 2021 (online)

Petra Paizs, Alvaro Perdones-Montero, James Kinross, Simon Cameron, Zoltan Takats.
Laser Assisted - Rapid Evaporative Ionisation – Mass Spectrometry Imaging Direct from sample mapping of faecal metabolites without sample preparation. Metabolomics Conference 2019, The Hague, Netherlands.

James Lewis, Sam Mason, Petra Paizs, David J.Brinkman, Simon Cameron, Alvaro Perdones-Montero, Ara Darzi, Zoltan Takats, Wouterde Jonge, and James M.Kinross
The Preoperative Faecal Lipodome But Not Microbial Diversity Predicts Post-Operative Ileus in Elective Colorectal Surgery. Digestive Diseases Week 2019, San Diego, USA.

Sam Mason, Liam Poynter, Efty Manoli, James Alexander, Petra Paizs, Afeez Adebesin, Robert Goldin, Ara Darzi, Zoltan Takats, James Kinross. Mass Spectrometry Transanal Minimally Invasive Surgery (MS-TAMIS) For Organ Preservation in Rectal Cancer, SAGES Annual Meeting 2019, Baltimore, USA

Petra Paizs, Eftychios Manoli, Sam E Mason, James Alexander, Zsolt Bodai, Emma White, Afeez Adebesin, Jonathan Hoare, Robert Goldin, Ara Darzi, James M Kinross, Zoltan Takats. Rapid Evaporative Ionization Mass Spectrometry (REIMS) Analysis of the mucosal lipidome has a high diagnostic accuracy for Adenomas and early Colorectal Cancer. American Association of Cancer Research (AACR) 2018, Chicago, USA.

Petra Paizs, Eftychios Manoli, Sam E Mason, James Alexander, Zsolt Bodai, Emma White, Afeez Adebesin, Jonathan Hoare, Robert Goldin, Julia Balog, Steven D Prongles, Ara Darzi, James M Kinross, Zoltan Takats.Rapid Evaporative Ionisation Mass Spectrometry (REIMS) Analysis of the mucosal lipidome to distinguish Adenomas and early Colorectal Cancer. American Society of Mass Spectrometry (ASMS) 2019, San Diego, 2018.

Eftychios Manoli; Zsolt Bodai; Petra Paizs; Julia Balog; Steven D Pringle; Ara Darzi; Zoltan Takats; Philip Pratt. Automated chemically driven robotic surgery using Rapid Evaporative Ionisation Mass Spectrometry (REIMS). American Society of Mass Spectrometry (ASMS) 2019, San Diego, 2018.

Sam Mason, James Alexander, Emma White, Zsolt Bodai, Eftychios Manoli, Petra Paizs, Alaisdair Scott, Afeez Adebesin, Jonathan Hoare, Robert Goldin, Ara Darzi, Zoltan Takats, James Kinross. Prospective Observational Cohort Study of Rapid Ionization Mass Spectrometry (Reims) for Near-Real Time Diagnosis and Stratification of Colorectal Cancer, Digestive Diseases Week 2018, Washington, DC, USA.

## 2.2.Oral Presentations

P. Paizs, A. Perdones-Montero, J. Kinross, S. Cameron, and Z. Takats.
Optimising Rapid Evaporative Ionization Mass Spectrometry (REIMS)-Mass Spectrometry Imaging (MSI) for the direct from sample mapping of faecal metabolites. Informal Meeting on Mass Spectrometry 2019, Fiera di Primiero, Italy.

Petra Paizs, Alvaro Perdones-Montero, James Kinross, Simon Cameron, Zoltan Takats.
Optimising Laser Assisted - Rapid Evaporative Ionisation - Mass Spectrometry Imaging (LA-REI-MSI) for the Spatially Resolved Analysis of Faecal Metabolites, Mass Spectrometry: Applications to the Clinical Lab (MSACL) 2019, Salzburg, Austria.