Imperial College London

Exploring the utility of metabolic profiling in stratifying patient groups in Inflammatory Bowel Disease

Thesis submitted by

Lucy Charlotte Hicks

For the Degree of Doctor of Philosophy

Department of Medicine

Imperial College London

2018

Supervisors

Professor Timothy R. Orchard and Dr Horace R.T. Williams

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.

Copyright Declaration

This thesis is dedicated to my wonderful family, Wesley, Lara and Xanthe

with all my love

Abstract

The pathogenesis of IBD, involving dynamic interactions between the microbiome, innate and adaptive immune systems, genetics and environmental factors, is a major focus of academic interest, in order to reveal more about the heterogeneous clinical course of the disease and in pursuit of improved therapeutic targets.

Metabonomics has been previously used with a variety of biofluids to successfully distinguish IBD from controls, but the complex metabolic data also have potential to unlock insights into pathogenesis and better understand how to better stratify patients for personalised clinical care.

In the largest urinary metabonomics IBD study to date, changes in the white European cohort confirmed previous published findings, highlighting discriminatory metabolites of gut microbial and inflammatory pathway sources. Significant metabolic differences were seen when comparing IBD patients and controls from South Asia to white North Europeans, demonstrating the influence of ethnicity on the metabolic profile and showing metabolite changes related to host-nutrition-microbiome interactions.

Results from longitudinal measurements of the IBD metabolome in the same individuals over several years indicate relative stability despite the relapsing-remitting course of the disease and different treatments. This early finding suggests clinical outcomes may only have subtly discernible changes on metabolic profiles, potentially limiting its application as a disease-monitoring tool.

16S rRNA profiling, employed to characterise the microbiome, showed reduced microbial diversity in IBD and 4 key bacterial genera - *Veillonella*, *Acidaminococcus*, *Lactobacillus* and *Streptococcus* - associated with disease. Significant urinary and faecal metabolites in the same patients were correlated with these bacteria to demonstrate the feasibility of multi-omic integration in IBD.

Furthermore, the breath VOC profiles of IBD patients obtained by SIFT-MS were distinct from those of heathy controls, with the significant compounds originating from microbial sources, and inflammatory pathways, demonstrating the potential of this technology and another facet to metabolic profiling in IBD.

Acknowledgements

First and foremost I would like to thank my supervisors Professor Tim Orchard and Dr Horace Williams, without whose support, guidance and patience, it goes without saying this thesis would not have been possible. Sincere thanks to them both for keeping faith and interest in my work over all these (long) years. Specifically thanks to Professor Orchard for taking me on in the first place, always seeing the big picture and making time in his incredibly busy schedule for me. And to Dr Williams for providing endless help and advice, down to the small and practical details.

I would like to thank members of the South Kensington Biomolecular department without whom I wouldn't know what the term metabonomics even meant. I am grateful to Beatriz Jimenez for all her help at the beginning, and subsequently Jon Swann who was generous enough to step in when I had no scientific support and gave plenty of invaluable advice in person and by Skype. His colleague Marine Letertre was also extremely kind in assisting with my micriobiome analysis work.

I am also grateful to the other members of the luminal research group at Marys – Sam Powles, for always making me laugh, endless coffees and distractions; Leo Chong for several days pipetting in the lab with me; the historical samples kindly given by David Walker.

The microbial work would not have been possible without the generous support of Professor Julian Marchesi and his laboratory. I'm grateful to Julie McDonald for patiently explaining Mothur to me, and to Ben Mullish for trying to optimise my data on SIMCA.

For their collaboration and help with the breath project I am thankful to Frank Huang, Sacheen Kumar and Professor George Hanna, and also to Mohammed Najim for help collecting samples.

Ultimately, I wouldn't have finished this thesis without the support and encouragement of friends, firstly Nowlan Selvapatt, who always knows how to talk sense into me and has always been there with the cherry vodka when required. Also to all my friends at Mary's who made my days in academic research so much fun and kept me going even when my enthusiasm wavered – 'St Mary's fellows' (SMF) Nikhil Vergis and Shevanthi Nayagam, as well as Ameet Dhar for all his generosity and advice, and all the others on the 10th floor.

Thanks to the unwavering support of Wesley, who must have been completely fed up of hearing about this thesis long ago, but never complained. And thanks to the joyful distraction of my two beautiful girls, who have undoubtedly significantly delayed the completion of this work. Thanks as well to my parents who always believed in me, and provided us with childcare to enable me to write up.

Last, but definitely not least, a special mention for Harry Antoniades who was an invaluable support, mentor, academic enthusiast and wonderful friend; but who sadly never got the chance to see this thesis. I blame him for convincing me to do a PhD in the first place, but I hope he would be proud of me completing this. This thesis is written in his memory.

Table of Contents

Chapter 2: Materials and Methods

2.1 Study participants and clinical data

2.1.1 Clinical phenotyping of IBD patients

2.1.2 Metabonomic and dietary information

2.2 Sample collection and preparation

2.2.1 Urine samples

2.2.2 Faecal sample collection and water extraction for NMR

2.2.3 Extraction of DNA from faeces for microbial analysis

2.2.3.1 Preparation of the reagents used

2.2.3.2 DNA extraction

2.2.3.3 Quantification of DNA

2.3 Microbial diversity analysis of faecal samples

2.3.1 16S rRNA gene analysis

2.3.1.1 Illumina 2 step technique

2.3.1.2 16S data processing

2.3.1.3 Microbial composition analysis

2.3.1.4. Diversity index analysis

2.3.1.5. Beta diversity analysis

2.3.1.6. Differential Abundance analysis

2.4 Urinary and faecal water NMR spectroscopy

2.4.1 NMR spectroscopy

2.4.1.1 Basics of NMR Spectroscopy

2.4.2 Preparation of urine samples for NMR

2.4.3 Preparation of faecal water samples for NMR

2.4.4 NMR Data acquisition and pre-processing

2.4.4.1 Data acquisition

2.4.4.2 Pre-processing

2.5 Breath testing with SIFT-MS

2.5.1 Exhaled breath capture technique

3.5.2.2.2 O-PLS-DA

3.5.2.2.2.1 Comparison of IBD patients by diagnosis

3.5.2.2.2.2 Validation of results

3.5.2.2.2.3 Comparison of IBD patients in ethnic cohorts

3.5.2.2.2.4 Comparing ethnicities

3.5.2.2.2.5 Comparison of first and second-generation Asians

3.5.2.2.2.6 Comparison by disease location

3.5.2.2.2.7 Comparison by surgical status

3.5.2.2.2.8 Comparison by gender

3.5.2.2.2.9 Validation of storage methodology

3.6 Discussion

3.6.1. Hypothesis driven targeted metabolite analysis

3.6.2 Multivariate pattern-recognition analysis

3.6.3 Significant metabolites altered in IBD

3.6.3.1 Hippurate

3.6.3.2 Other bacterially derived metabolites

3.6.2.3 Energy cycle metabolites

3.6.2.4 Short chain fatty acids

3.6.2.5 Amino Acids

3.6.2.6 Organic acids

3.6.2.7 Novel metabolites

3.6.4 Examining the effect of ethnicity on the metabolic profile

3.6.4.1 Metabolic profiling comparing different ethnicities

3.6.4.2 Examining IBD in different ethnicities

3.6.4.3 Ethnicity and the effect of diet on the metabolome and microbiome

3.6.5 Other analyses of data

3.6.5.1 Comparison of 1^{st} vs 2^{nd} generation South Asians

3.6.5.2 Disease location

3.6.5.3 Effect of surgery

3.6.5.4 Validation analysis: Gender

Table of Figures

Table of Tables

Supplement 1: Ethical approval from Research Ethics Committee

Supplement 2: Control and patient information sheets

Supplement 3: Median values of 19 metabolites for each group

Supplement 4: p values and corrected p values for all comparisons

List of Abbreviations

^{1}H	Proton
16S rRNA	16S ribosomal Ribonucleic acid
5-ASA	5-aminosalicylate
6MP	6-mercaptopurine
7FCV	7-fold cross validation
Anti-TNF	Anti-tumour necrosis factor
CD	Crohn's disease
DNA	Deoxyribonucleic acid
EB	Exhaled breath
FID	Free-induction decay
HBI	Harvey-Bradshaw index
IBD	Inflammatory bowel disease
IL-10	Interleukin-10
IL-$10^{-/-}$	Interleukin-10 gene-deficient
LOOCV	Leave-one-out-cross validation
NMR	Nuclear magnetic resonance
NOD	Nucleotide oligomerisation domain
NPV	Negative predictive value
O-PLS-DA	Partial least squares discriminant analysis orthogonal signal correction
OTU	Operational taxonomic unit
PAG	Phenylacetylglycine
PCs	Principal components
PCA	Principal components analysis
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PLS-DA	Partial least squares discriminant analysis
PPV	Positive predictive value
SCCAI	Simplified clinical colitis activity index
SCFA	Short chain fatty acid
SIFT-MS	Selected Ion Flow Tube Mass Spectrometry
STOCSYE	Statistical Total Correlation Spectroscopy Editing
TOCSY	Total correlation spectroscopy
TMA	Trimethylamine
TMAO	Trimethylamine-N-oxide
TSP	3-trimethylsilyl-(2,2,3,3-$^{2}H_4$)-1-propionate
RSPA	Recursive segment-wise peak alignment
UC	Ulcerative colitis
VOC	Volatile organic compound

Chapter 1: Introduction

1.1 Inflammatory bowel disease

1.1.1 IBD pathology and diagnosis

The inflammatory bowel diseases (IBD), of which Crohn's disease and ulcerative colitis (UC) are the principle types, are lifelong inflammatory conditions of the gastrointestinal tract (1, 2). The main IBD sub-types share commonalities – both are chronic, idiopathic conditions characterised by episodes of relapse and remission (3). Clinically there are overlapping symptoms, which are usually reflective of disease location, and often reflective of disease severity. Diarrhoea and urgency, with or without abdominal pain, are the most common symptoms. In UC and Crohn's colitis blood is often present in stool. Constitutional symptoms of fatigue and weight loss are variable. In Crohn's disease symptoms can be more heterogeneous and non-specific. Some pathological mechanisms are common to both and many of the medical treatments are shared.

However there are important features specific to each disease. Ulcerative colitis is, by definition, restricted only to the colon (although backwash ileitis can sometimes occur, this is considered a secondary phenomenon from the colonic inflammation) whereas Crohn's disease can affect any part of the gastrointestinal tract, although typically is found in the terminal ileum and/or colon. Features specific to Crohn's include perianal involvement and fistulation to other organs.

Histo-pathologically the diseases are distinct, with different 'typical' features, although in clinical practice it can sometimes be impossible to separate on the grounds of the histological findings alone (3). Crohn's disease is usually patchy, with deep trans-mural involvement and

non-caseating granulomata are more commonly (but not ubiquitously) seen. UC usually has continuous mucosal inflammation and crypt abscesses are a more frequent finding.

The international standard of phenotype subtyping of Crohn's and UC is the Montreal classification (4) as shown in table 1.1.

Crohn's disease		Ulcerative colitis	
Age at diagnosis	A1: Below 16 years A2: Between 17 and 40 A3: Above 40 years old		
Location	L1: Ileal L2: Colonic L3: Ileocolonic L4: Isolated upper GI disease[a]	Extent	E1: ulcerative proctitis E2: Left sided (disease distal to splenic flexure) E3: Extensive / pancolitis (disease proximal to splenic flexure)
Behaviour	B1: non-stricturing, non-penetrating B2: stricturing B3: penetrating p: perianal disease[b]		

Table 1.1. Montreal classification of Crohn's and ulcerative colitis

[a] modifier which can be added to L1-3 when upper GI disease present; [b]modifier which can be added to B1-3 when perianal disease present

Hence it is recommended that the diagnosis of IBD and its subtype is a clinical one, based on a combination of clinical findings, endoscopy, histology, radiology and biochemical investigations (1, 2). An estimated 10% of IBD is termed 'IBD unclassified' (IBD-U) - this is a colitis which fulfils neither the diagnostic criteria of Crohn's or UC (5).

1.1.2 Treatment and identifying response and progression

Medical therapies for IBD are targeted and escalated according severity of disease or non-response, and include steroids, immunosuppression and anti-tumour necrosis factor (anti-TNF)

therapies. Surgical management is indicated for treatment of complications (including fistulae, strictures, perforations) and in those patients with fulminant disease where medical therapy has failed or those intolerant of medical treatments. Surgery is required more often in Crohn's disease, with a lifetime risk of 80% compared to 20% in UC (6), and a high recurrence of Crohn's (60% at 10 years) with requirements for further surgery (7).

Aminosalicylates (5-Aminosalicylic acid / 5ASA) treatments are administered topically or orally, and can be used to treat mild / moderate flares of UC. However their main role is for maintenance of clinical remission (8) (with an additional benefit being possible reduction of colorectal cancer risk (9)) and thus a large proportion of UC patients will be prescribed these long-term.

Steroids are indicated in acute flares of IBD, and whilst they have potent anti-inflammatory effects, their systemic side effect profile (including osteoporosis, adrenal suppression, insulin resistance, glaucoma and psychosis) limit long term use. To avoid repeated or prolonged steroid courses immunomodulators, including azathioprine and 6-mercaptopurine (thiopurines) or methotrexate (folic acid antagonist), are given, with evidence of efficacy in maintenance of clinical remission in both types of IBD (10, 11), and reduction of post-surgical recurrence in Crohn's (12).

Over recent years the armoury of IBD therapeutic options has grown, and continues to develop. Monoclonal antibodies (mAbs) to anti-tumour necrosis factor α (anti-TNFs), known as biological agents (biologics), have been established therapy for moderate to severe disease for the last 15 years (13). Originally only Infliximab (chimeric mouse/human whole monoclonal antibody) and Adaulimumab (humanised whole monoclonal antibody) were available for use in IBD, but newer licenced drugs include Golimumab (another humanised whole monoclonal

antibody), Vedolizumab (monoclonal antibody selective to the α4β7 integrin-receptor) and Ustekinumab (humanised monoclonal antibody to interleukins 12 and 23) as well as lower-cost biosimilar TNF-inhibitors which have recently become available (14). Late phase trials of small molecules also show promising results (15).

Whilst there is good evidence that biologic treatments are effective in both reducing clinical symptoms and improving quality of life measures, as well as mucosal healing and disease modification (16, 17), up to 40% of patients have primary or secondary treatment failure or drug intolerance (18). Additionally, these treatments are expensive, and administration carries a risk of adverse events, most commonly opportunistic infections, but also rarely demyelination, heart failure or malignancy.

The previously conventional algorithm of 'step-up' treatment, escalates the potency of treatment as the disease continues to relapse, off-setting the higher risks of drug side effects and cost against the need to treat more progressive and aggressive disease. Current guidelines indicate biologics are indicated for the use of moderate or severely active IBD which in non-responsive to conventional therapy or in those who are intolerant or who have contraindications to conventional treatment (8). However this reactive management strategy, treating clinical symptomology and response, can fail to change the natural history of disease progression and complications (19).

The alternative 'top-down' paradigm – initiating more intensive combined therapies early after diagnosis with the expectation of modifying the long term disease course – has both health economic implications and risks 'over-treating' up to 30% of patients (20). Long term outcomes from 'step-up' and 'top-down' approaches, including remission, hospitalisation and surgery, appear to be equivocal (21).

Therefore one of the current challenges facing IBD specialists is personalising treatment to identify the right people in which to initiate the right treatment at the right time, in order to maximise effectiveness and minimise adverse events. The new concept of 'accelerated step-up' treatment, aims for more individualised proactive management of patients according to severity and prognosis with the goal of improving outcomes (22).

At present there are no validated biomarkers available in clinical practice which can predict response to therapy (23); the search in translational medicine is to identify biomarkers of treatment response, or methods to stratify patients according to disease phenotype (15).

Decision making is currently a clinical one, based on multiple factors including patient choice, age and disease features. Patients of all ethnicities are currently treated the same, despite evidence that ethnic groups have different clinical IBD phenotypes (24, 25) and show different treatment failure rates (26). It would be very valuable to better understand the contribution of ethnicity on disease phenotype and disease outcomes.

1.1.3. The multifactorial concept of IBD Pathogenesis

The 4 key components of IBD pathogenesis, as it is currently understood, are likely to be genetics, environment, immunological response and the gut microbiome (27). Whilst research in each of these areas has moved our understanding forward, the complex interconnection between these elements has not yet been fully explained. The constantly evolving interactions between these components means that cause versus consequence is difficult to determine, and the role of each is dynamic and changeable (28).

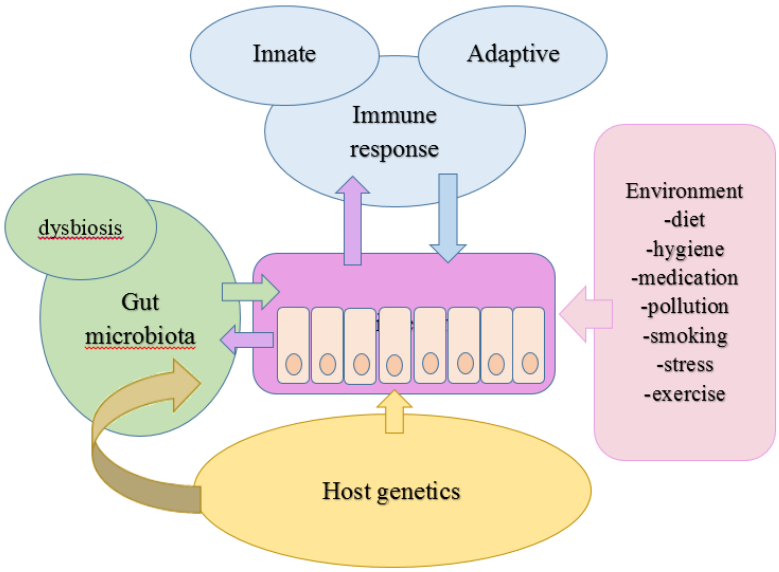


Figure 1.1: Conceptualised pathogenesis of IBD

The current concept of pathogenesis (figure 1.1) explains an aberrant and persistent inflammatory response to environmental stimuli, of which the most significant is likely to be altered gut bacteria, in individual with predisposing genetic factors relating to both innate and adaptive immunity, bacterial defence and gut barrier function (29).

1.1.3.1. Genetics and immunology

Familial clustering of cases first suggested a genetic component to IBD, with a positive family history conferring a 8-10 fold relative risk of IBD (30) and twin concordance of 30-50% in monozygotic twins with Crohn's (compared to 15% in UC) (31). A number of genetic associations were described, particularly with HLA polymorphisms (eg HLA DRB1*0103 in UC (32)) , but it was not until identification of the nucleotide-binding oligomerization domain

containing protein 2 (NOD2) / caspase recruitment domain-containing protein 15 (CARD15) gene variant on chromosome 16 (33) and its relationship to ileal Crohn's was established that a specific gene was linked to disease susceptibility. Progression in genetic technologies enabled the identification of single nucleotide polymorphisms (SNPs); and the largest genome wide association study (GWAS) to date identified more than 163 gene loci associated with IBD in Caucasians with 110 common to both UC and Crohn's (34). More recently 28 further loci were discovered in a cross-ethnicity study, demonstrating population specific genetics (35).

The NOD2/CARD15 gene codes for an intracellular receptor for bacterially produced muramyl dipeptide (MDP), which stimulates autophagy via activation of NFκB pathways. It is also directly involved in T-cell response regulation and is involved with both innate and adaptive pathways. Other identified genes relate to autophagy defects, including ATG16L1 and IRGM, affecting intracellular homestasis as well as antibacterial resistance and phagocytosis (36).

Advances in genetic research have highlighted the combined effects of gene mutations leading to significant defects in innate and adaptive immunological pathways, toll-like receptors (allowing recognition of pathogen-associated microbial patterns) and a reduction in inducible regulatory T cells (37).

Although the progression of our understanding of these genetic and immunological mechanisms reveals key insights into disease pathogenesis, the identified genes to date only explain less than 15% of the heritability (genetic risk) of IBD (28) – hence so called 'missing heritability'- suggesting other significant factors implicated in pathogenesis, most likely environmental and immunological. Whilst this does not dispute the importance of genetic susceptibility in IBD it demonstrates that the pathogenic model is much more complex. Epigenetics – the phenomenon of heritable modifications in gene expression without alterations

in DNA – may explain some of this effect (38), and may be described as the interface between DNA and our external cellular environment (39). It is recognised that in IBD the environment and gut microbiota influence the immune response via epigenetics to determine if inflammation is instigated and/or precipitated (28)

1.1.3.2. Environment and diet

Epidemiological studies have linked IBD to, amongst other factors, smoking (40), stress (41), appendicetomy (42), non-steroidal anti-inflammatories (43) and vitamin D deficiency (44). Smoking has the strongest and most replicated causal association; with meta-analyses demonstrating an adverse outcome of smoking in Crohn's – both in disease risk, disease progression and post-operative recurrence (45), but an inversely protective effect in UC risk and outcomes (46).

Epidemiological data show a rapid rise in the incidence and prevalence of IBD worldwide over the last 70 years. Whilst conventionally considered a disease of the 'Western' world (highest global rates in North America and Northern Europe (47)) the fastest rise in incidence is Asia and Eastern Europe in line with industrialisation of these countries (48). Whilst likely to be the result of multifactorial environmental causes, diet and improved sanitation ('hygiene hypothesis') are the two obvious factors implicated. The effect of diet on disease or the microbiome is notoriously difficult to fully establish, due to recall bias, misreporting and the heterogeneity of most diets. No specific diet or dietary constituent has been shown to cause or ameliorate IBD, but the significant impact of nutrients on the microbiota is established (49). A high-fat, high-dairy, high-calorie diet was linked to increased rates of colitis in a Japanese study (50), with dietary constituents acting as luminal 'antigens' in animal models (51). The relationship between diet and the intestinal microbiome, has been demonstrated in ecological

studies, with dominant Bacteroides and low Firmicute species found in the gut microbiome of rural African children on a high fibre diet, compared to the inverse ratio seen in European children consuming animal proteins, high sugar and low fibre intake (52). Again, animal models reinforce this link between diet and the microbiota, with bacterial species enriched by a high-fat, high-sugar diet causing obesity in genetically predisposed mice (53). There is evidence that diet not only modifies the composition of the gut microbiota, but also affects the abundance or absence of available substrates for host and bacterial metabolic processes, including maintenance of mucosal immunity (54).

1.1.3.3. Gut microbiome

1.1.3.3.1. Gut microbial function and composition

The gut microbiome plays an essential role in maintenance of host health or disease (55) with a balanced symbiosis of the human-microbial axis in the healthy physiological state (56). The GI tract bacterial community forms the largest and most complex ecosystem in the human body comprising of 10^{13} to 10^{14} microorganisms (57). Human genes are outnumbered by collective resident microbial genes by a factor of 100 with the huge majority of these resident in the gastrointestinal tract. The gut microbiota can be regarded as a metabolic super-organ (58) which plays the following essential roles:

- synthesis of essential short-chain fatty acids and vitamins
- development of the immune system and protection from pathogenic species (59)
- influence of energy homeostasis and body mass (60)
- drug metabolism (61)

There is overwhelming evidence for the microbiome impacting on disease status through human–microorganism functional interactions, with variations in the microbiome associated with metabolic diseases including diabetes (62), amongst many others (63, 64). Microbial alterations have been connected to inflammation, co-morbidities and markers of frailty in the elderly (65) and other systemic diseases (coronary heart disease (66) and rheumatoid arthritis (67)) by linkage of specific metabolic phenotypes shown to be driven by microbial metabolism. IBD has emerged as one of the most strongly established diseases linked to the gut microbiota (68) and is one of the most researched diseases in this area.

It has been established that the composition and diversity of the microbiome is determined in childhood, reaches stability in adulthood and declines in stability again towards old age (69). IBD has a bi-modal incidence, with the largest peak aged 15-30 and a second smaller peak in the 50-70 years group (70) and a smaller, but rapidly increasing paediatric (<15 years old) cohort (71). These early and late peaks coincide with the most variable composition of the microbiome and may point to lack of resilience of the gut microbes in resisting fluctuations initiating IBD, or may suggest different contributions of the microbiome in disease instigation and/or propagation (72).

Multiple influences affect bacterial configuration, especially during the initial establishment phase (64), including host genetics, environmental exposures (including antibiotic use (73) and method of delivery (74)), contact with microbes / hygiene and diet (75). Whilst the gut microbiota composition shows some fluctuation with time within an individual (76) the functional productivity appears to be more stable (77), suggesting that overall function rather than the content is most important.

The microbial concentration increases along the length of the GI tract from mouth to anus, with the maximal diversity in the distal gut (caecum, colon, rectum). Gene-sequencing studies have identified over over 1800 genera and 15,000 phylotypes in healthy subjects (78) although the colonic microbiome is strongly predominated (over 80%) by *Bacteroides* and *Firmicutes* taxa (79). Importantly studies have shown that low abundance species can play unique key molecular functions in the gut, so that even small perturbations in the balance can cause and perpetuate microbial dysfunction (80).

1.1.3.3.2. Dysbiosis in IBD; cause or consequence?

Technological advances have refined identification and cataloguing of the microbiome, including fungi, viruses, eukaryotes as well as bacterial species. The microbiota in IBD has been well studied in both animal models and human subjects with a variety of techniques.

Two main pieces of evidence underline the key role of dysbiosis in IBD. Firstly, there is substantial and consistent evidence that an alteration in the composition of the microbiome – termed dysbiosis – is seen in IBD (81-83), however specific bacterial findings have not been entirely similar between studies, this in part due to the multiple other influences on gut bacteria including diet, age and geography (84). One result repeatedly seen however is the loss of species richness (α-diversity) seen in IBD (85, 86). Reduced species richness is also seen in monzygous twins discordant for Crohn's (87). There is also evidence that the microbial diversity in inflamed tissue is reduced compared to non-inflamed colonic tissue (88).

Studies identifying specific changes in the microbial contents have been somewhat contradictory (37) with most groups documenting an increase in *Bacteroidetes* phyla in IBD

(89, 90) but others a decrease (91). *Gammaproteobacteria* is noted to be increased in both types of IBD (92, 93).

No single causative organism has been recognised in IBD but one noteworthy and consistent finding is the reduction in *Faecalibacterium prausnitzii* in Crohn's disease (94-96) and to a slightly lesser extent UC (97). *F. prausnitzii* is part of the *C. leptum* group (of the *Firmicutes* phylum*)* which are one of the important bacterial sources of butyrate – the major energy source for colonocytes, which also exerts significant immunomodulatory and anti-inflammatory effects (98). Separate to the function of butyrate *F. prausnitzii* has been shown to display direct anti-inflammatory properties in murine *in vitro* and *in vivo* models (94).

In addition to butyrate, other short chain fatty acid (with immunoregulatory function) producing bacteria have been found to be altered in IBD, specifically *Ruminococcaceae* family (85) and *Leuconostocaceae* (72) suggesting the change in these bacteria is linked to the functional alterations seen in IBD. Other microbial functions associated with IBD include decreases in amino acid biosynthesis, increase in auxotrophy (inability to synthesize certain substances required for growth and metabolism) and increased oxidative stress (72).

Secondly, the aberrant immune response to bacteria can be demonstrated in animal models with the development of colitis upon the introduction of bacteria into genetically susceptible animals (99). In humans defunctioning stomas are effective in improving perianal and colonic Crohn's disease (100), while restoration of the faecal stream restarts inflammatory symptoms (101). Additionally antibiotics (anti-microbials) have been demonstrated to ameliorate intestinal inflammation (102) and are an important therapy in pouchitis (inflammation of surgically created neo-rectum) (103); evidence which suggests the manipulation of the microbiome as a potential target for therapeutic strategies.

There is also evidence of the interplay between host genetics, the immune system and the microbiota. Alterations in the gut microbial make-up are seen in patients carrying NOD2 and ATG16L1 risk alleles, specifically a reduction in *Faecalibacterium* and increase in *Escherichia* (92), tying together genotype with dysbiosis. Murine models show that NOD2-deficient animals have a dysbiosis in the terminal ileum with more *Bacteroides, Firmicutes*, and *Bacillus* present. Cells expressing NOD2 (mostly Paneth cells of the terminal ileum) are unable to effectively kill bacteria and therefore cannot regulate the ileal microbiota (104).

It is not clear from simply cataloguing the intestinal microbes how much the dysbiosis is the ignition or the fuel for chronic inflammation, or what role specific bacteria are related to in the inflammatory process (105). It is likely to be more important to explore the functional output of the microbiome and the subsequent disrupted metabolic processes than simply identifying and quantifying the bacteria [55].

An interdisciplinary approach may offer an approach to answering these questions, and metagenomic studies have started to explore the functional role of gut bacteria in IBD.

1.1.3.3.3. Exploring the gut microbiota

Studying the gut microbiota was previously limited by culture based techniques which failed to identify a large proportion of the microbial community. Next generation sequencing (NGS) has revolutionised the approach to understanding the commensal microbes living in the human body. High throughput technology and falling cost has enabled the application of NGS to the exploration of the gut microbiome and its function in health and disease.

In NGS 16S rRNA is extracted and amplified from the total DNA by PCR techniques, then sequenced, and bioinformatic analysis is applied to provide the relative phylogenetic composition of a sample. Reference genomes can be used as a comparison to infer the functional capacity of the gut microbiome [85].

Combined metagenomic (study of the variation of species in a microbial sample) and metaproteomic (study of the proteins in a biological sample) techniques in IBD patients have demonstrated the reduced ability of bacteria to self-produce amino acids and an up-regulated ability to transport these amino acids from areas of inflammation as well as increased oxidative stress pathways (106) .

Despite the valuable data thus far, there remain significant gaps in the knowledge of how microbes act as pathogens, and the relevance of perturbed host-microbial pathways and subsequent functions. A more multifaceted approach to investigating the altered functional and dynamic relationships of each component in pathogenesis is required to start to unravel more about the relationship between the gut microbiota and IBD pathology.

1.2 Characterising IBD using Nuclear Magnetic Resonance Spectroscopy and Metabonomics

1.2.1. Metabonomics

Metabonomics is the technique of identifying and quantifying low-molecular mass metabolites contained in a biological specimen using, most commonly, 1H nuclear magnetic resonance (NMR) spectroscopy or mass spectroscopy platforms (107). The information set generated is large and complex, representing the entire biological and metabolic functions of an individual, as well as external influences, at the time point of sampling.

NMR spectroscopy is an untargeted technique, with a short analysis time, providing comprehensive structural information of the most abundant metabolites contained in a sample; typically 20-70 identifiable compounds including amino and organic acids, sugars, amines, nucleosides, phenolic compounds, osmolytes, and lipids (108). This is a powerful and sensitive investigative tool, as even minor external or internal perturbations cause detectable variations in the metabolic profile. Thus this can reveal novel insights into disease mechanisms or the effects of environmental factors that may not be identified by other techniques. However there are multiple potential confounders that must be considered when using metabonomics; standardised methods (including sample collection and processing) must be employed with rigour to ensure reproducibility and large cohorts are required for validation. Clinical phenotypic data is essential to allow meaningful recovery of information and translation of results into the clinical arena.

Thus far metabonomics has been used in drug development (109), toxicology studies (110) and animal models (111) as well as basic science clinical research in many disease states. The strategy of combining metabonomics data with other ‘omic’ techniques, known as ‘systems

biology approach', involves modelling of complex biological systems through complementary pathways. This thesis involves work incorporating microbiological data (discussed further in chapter 1.3.1) with metabonomics and understanding the correlations between the two. Metabonomics and its integration with other technologies, has an evolving role in clinical research exploring mechanisms of disease pathogenesis, with the potential to be developed and applied in the clinical field for diagnostics or monitoring purposes.

Metabonomics can be performed on a wide variety of biofluids including urine, plasma, serum or faecal water, as well as biological vapours including breath, the headspace of biofluids, surgical gas vapour, or using intact tissue specimens. Each of these types of sampling has comparative advantages and disadvantages and may require invasive sampling, as well as different processing techniques. They may also be combined to provide complementary metabolic information from the same individual, reflecting the integrated systems biology approach above.

1.2.2 Metabonomics in IBD

Metabonomics (also referred to as metabolomics in the literature and this thesis) may potentially address key issues in IBD at several levels from bench to bedside - helping to characterise and dissect the pathological mechanisms described above by exploring the small molecules that play a functional role in the interaction between microbial communities and the immune system. Clinical applications may potentially include a novel, non-invasive diagnostic method, but rather than identifying a biomarker for the disease, this knowledge is more likely promote the understanding of an individual's specific disease phenotype allowing optimisation of individualised treatment.

Metabonomics has been applied in IBD research using a variety of biofluids, but most often using faecal water, urine, plasma/serum or colonic biopsies (112). These samples are easily handled and mostly non-invasively acquired (although colonic biopsies require colonoscopy). Urine and faecal water do not even require venepuncture for collection.

All the above biofluids have demonstrated differences in metabolic profiles between IBD patients and healthy controls (113-116) and several have been able to discriminate sub-types of UC and Crohn's from one another (117-119).

Serum and plasma studies and colonic biopsy samples (not studied in this thesis) show more information about host metabolism, and less inter-individual variation (120) but do not contain much information about the gut microbiota or its function. Studies in serum and plasma are summarised in table 1.2.

Changes in the serum/plasma and colonic biopsies mostly relate to altered amino acid metabolism (115, 121) possibly related to underlying malabsorption or increased catabolism due to inflammation. One study went further to show that the amino acid profile ('AminoIndex') could distinguish patients with Crohn's from UC from healthy controls, as well as identify active disease from quiescent within the same diagnosis (122).

Sample size	Author	Analytical platform	Differences between IBD and healthy controls	Differences between Crohn's and UC
CD (n=24) UC (n=20) HC (n=23)	Williams (119)	^{1}H-NMR	↓ low-density lipoproteins and HDLs in IBD	↑ N-acetylglycoprotein in Crohn's ↓ Lipids and choline in Crohn's
CD (n=19) UC (n=24) HC (n=17)	Dawiskiba (116)	^{1}H-NMR	↑ N-acetylated compounds and phenylalanine in IBD ↓ low-density lipoproteins and very low-density lipoproteins in IBD	Unable to differentiate
CD (n=20) UC (n=20) HC (n=40)	Schicho (115)	^{1}H-NMR	↑ methanol, mannose, and amino acids in IBD ↓ urea, citrate and acetate in IBD	Weak statistical models
CD (n=21) UC (n=13) HC (n=17)	Ooi (121)	GC/MS	↑ Fumaric acid in UC ↑ Fumaric acid, malic acid, and succinic acid in Crohn's	↓ Aconitic acid in Crohn's ↑ Succinic acid in Crohn's
CD (n=165) UC (n=222) HC (n=210)	Hisamatsu (122)	Plasma aminograms	↓ histidine and tryptophan in IBD	Differences in amino acids between active UC and Crohn's

Table 1.2: Table of metabonomic studies using serum/plasma in IBD

Metabonomics have been examined in animal IBD models (also not studied in this thesis), with the advantage that they present less challenges than human subjects in accounting for potential confounders such as diet, genetics and environment. Mouse models show metabonomic alterations in urine (123, 124) and serum (124) associated with development of colitis. Urinary changes were mostly relating to methylamines and other metabolites produced by gut bacteria; serum metabolic changes reflect alteration in energy pathways including ketone bodies.

1.2.2.1. Faecal water metabolic profiling in IBD

Faecal water profiling was the first biofluid to demonstrate changes relating to IBD in humans. The strengthening evidence for gut dysbiosis as a central pathological mechanism attracted interest in the use of faecal water or extracts to explore the metabolic relationship between host and bacteria. Faecal profiles consist of metabolites from host cellular pathways, bacterial microbiota and xenobiotics (molecules produced externally to the host organism – mostly breakdown products of nutrition or drugs). They also contain human commensal co-metabolites which are produced by a combination of microbial and human metabolism (125).

Sample size	Author	Analytical platform	Differences between IBD and healthy controls	Differences between Crohn's and UC
CD (n=10) UC (n=10) HC (n=13)	Marchesi (117)	^{1}H-NMR	↑ amino acids in IBD ↓ SCFAs, trimethylamine and methylamine in IBD	↑ amino acids and glycerol in Crohn's ↓ butyrate and acetate in Crohn's
CD (n=44) UC (n=48) HC (n=21)	Bjerrum (113)	^{1}H-NMR	↑ amino acids in IBD ↓ SCFAs in IBD *unable to differentiate CD from healthy	↓ Aspartic acid and glutamate in inactive Crohn's vs inactive UC
UC (n=13) HC (n=22)	Le Gall (126)	^{1}H-NMR	↑ levels of taurine and cadavarine in UC	N/A
CD(n=10 twins) HC(n=7 twins)	Jansson (127)	ICR-FT/MS	↑ amino acids and bile acid metabolites in IBD	N/A
CD (n=50) UC (n=82) HC (n=51)	Santoru (128)	GC-MS ^{1}H-NMR and LC-QTOF-MS	↑ biogenic amines, amino acids, lipids in IBD ↓ B group vitamins in IBD	Poor statistical models

Table 1.3: Table of metabonomic studies using faecal extracts in IBD – adapted from (125) and (112)

Marchesi et al (117) demonstrated depletion of gut bacterial metabolites in IBD stool including short chain fatty acids (SCFAs) (predominantly butyrate and acetate) and bacterially-produced food breakdown products (methylamine and trimethylamine) consistent with dysbiosis in these individuals. Higher amino acid concentrations were seen in the faecal IBD samples, consistent with increased gut loss of protein through an inflammatory malabsorption process. Jansson et al (127) discovered multiple differentiating masses discriminating concordant and discordant twins with Crohn's and healthy twin pairs using Fourier Transform Ion Cyclotron Resonance Mass Spectrometer, and related them to bile acid and fatty acid biosynthesis, and tyrosine pathways amongst others. A 1H-NMR study (126) found elevated levels of taurine in faecal water of UC patients compared to controls – this is a metabolite produced by bacterial breakdown of bile products in the colon, and/or an intracellular anti-oxidant and anti-inflammatory. Cadavarine was also detected in higher quantities in the UC cohort than controls, with the authors suggesting this was due to an alteration in microbes. As in the Marchesi study, SCFAs and branched chain fatty acids (BCFAs) were also reduced.

The study by Bjerrum et al (113) also studied faecal extracts with ^{1}H-NMR to investigate metabolic differences in active and inactive IBD cohorts and healthy controls. Some differences were consistent with previous data. Firstly, a decline in SCFAs (butyrate and propionate), which are products of bacterial fermentation of polysaccharides, and are an important energy substrate for colonic tissue (129). They also display anti-inflammatory properties via inhibition of NFκB pathways (130). Another replicated finding was the greater abundance of amino acids (in this study lysine, alanine, tyrosine, phenylalanine, and glycine) in active IBD compared to controls, potentially related to gut malabsorption. However in this study recognised significant confounders were not excluded – namely intestinal surgery, treatment with anti-TNF-α antibodies and potentially other medication. On exclusion of these

subgroups the results changed dramatically, and the significance of comparative models switched (some models became invalid whilst others became valid), highlighting the necessity to closely control for clinical variables and the careful interpretation of results.

Santoru et al (128) published a more recent study of faecal water metabolic profiling using three analytical techniques – NMR, gas chromatography–mass spectrometry and liquid chromatography quadrupole time-of-flight mass spectrometry. In their IBD cohort they also found more polyamines and amino acids which was postulated to be due to malabsorption, oxidative stress mechanisms or dysbiosis.

In addition to solid or liquid faecal samples, more recently the volatile contents of faecal headspace have been investigated in IBD – this will be discussed further later in this chapter (1.2.4.1.).

All these above studies demonstrate metabolic profiling of faeces can reveal important information relating to functional changes from the dysbiosis of normal bacterial ecology as well as malabsorption and inflammation pathways in IBD.

1.2.2.2. Urinary profiling in IBD

Urine is also a useful biofluid to study gut pathology and there have been several studies examining the differentiating metabolites in urine between IBD disease types, disease activity and location (see table). Urine is arguably the most easily obtained biofluid, with a high acceptability to patients, is stable over time and requires minimal pre-analysis preparation. Urine contains both gut bacterial metabolites and bacterial host commensal co-metabolites which are absorbed from the colon into the systemic circulation and excreted in urine. Studies

have shown bacterial perturbations can be detected in urine metabolic profiles of both animal models (131) and in human disease (55).

A potential limitation of this approach is that there may be more inter-individual variation in urine profiles in the general population than that seen in serum/plasma, including age and gender related changes, as well as environmental influences such as diet, which all must be accounted for.

Sample size	Author	Analytical platform	Differences between IBD and healthy controls	Differences between Crohn's and UC
CD (n =86) UC (n =60) HC (n=60)	Williams (118)	^{1}H-NMR	↑ glycine, methylhistidine, guanidinoacetate in IBD ↓ hippurate in IBD	↑ formate in CD ↓ hippurate and 4-cresol sulphate in CD
CD (n =30) UC (n =30) HC (n=60)	Stephens (114)	^{1}H-NMR	↓ succinate, citrate and hippurate in IBD	No differentiation found between groups after controlling for surgery.
CD (n =20) UC (n =20) HC (n=40)	Schicho (115)	^{1}H-NMR	↑ mannitol, allantoin, xylose, and carnitine in IBD ↓ hippurate and betaine in IBD	No differences
CD (n =19) UC (n =24) HC (n=17)	Dawiskiba (116)	^{1}H-NMR	↓ taurine, citrate, succinate and hippurate in IBD	No differences
UC (n =68) HC (n=25)	Bjerrum (132)	^{1}H-NMR	No differences	N/A

Table 1.4: Table of metabonomic studies using urine in IBD - adapted from (125) and (112)

Williams et al (118) first demonstrated changes in metabolic profiles in urine with 1H NMR spectroscopy in a cohort of IBD patients and healthy controls. In addition to distinguishing disease from healthy, the 2 subtypes of IBD could be separated by multivariate techniques. The most significantly reduced metabolite (in Crohns greater than UC) was hippurate, a human bacterial co-metabolite produced by bacterial fermentation of polyphenols, purines or aromatic amino acids, which undergoes hepatic conjugation and urinary excretion (133). This finding was replicated in work by Stephens (114) and Schicho (115) also found hippurate levels were less in IBD. Clostridia species, which are lowered in IBD, are associated with hippurate levels (134) suggesting dysbiosis as the cause.

In the Crohn's cohort of the Williams study, 4-cresol sulphate was significantly reduced and formate increased relative to UC and controls. 4-cresol sulphate is a metabolite of tyrosine (135), produced primarily by Clostridia and Bacteroidetes species, both of which have been shown to be lower in the bacterial bowel communities in IBD, specifically Crohn's (86) . Formate is the signature metabolite produced primarily by Escherichia coli (136) and other Enterobacteriaceae, which have been shown to increase in active Crohn's disease in microbial molecular studies (137).

Stephens et al (114) were also able to use multivariate techniques to distinguish IBD from healthy controls although UC could not be separated from Crohn's. Metabolites differentiating IBD from controls included gut bacterial co-metabolites as seen above (hippurate, formate, methanol, acetate and methylamine) as well as succinate and citrate – metabolic intermediates in the tricarboxylic acid (TCA) energy cycle. The amino acids asparagine, lysine, and histidine were also reduced relative to controls, with authors suggesting intestinal malabsorption as the cause, similar to findings(117) in faecal studies.

Schicho and colleagues studied serum, plasma and urinary metabonomics in an IBD cohort and compared to healthy controls. Robust models could discriminate between diseased and non-diseased individuals, but not identify the sub-types of IBD from each other. Citrate and succinate were lowered in IBD as in the Stephens study (114), as well as hippurate, as seen in the 2 studies above (114, 118). Other discriminating metabolites included allantoin and tryptophan (raised levels in UC) and lactate and carnitine (elevated levels in both Crohn's and UC relative to controls). These findings correlated with urinary metabolic changes in a mouse model of colitis by the same authors (124).

An integrated metabonomic study in UC performed by Bjerrum et al (132) was the only study not to be able to differentiate either active or inactive UC from healthy controls with urinary metabonomics. However other samples from the same cohort demonstrated differences in mucosal biopies and colonocytes.

The consistent findings of urinary metabonomic studies this far highlight alterations in metabolites secondary to disturbances in the gut microbiome, as well as alterations in energy pathways and malabsorption, demonstrating the validity of urinary metabolic profiling to investigate the pathogenesis of IBD.

1.2.3. Integration of complementary metabolic and microbiological approaches

The above approaches have been successful in identifying potential features specific to IBD that distinguish it at a metabolic and microbiological level from healthy people. The next challenge is to apply these techniques together in the same patient group to try and start to unravel the complex genetic-environmental-immune interaction in IBD at a functional descriptive level (138).

Ultimately the improved characterisation of IBD phenotypes with distinctive pathogenic mechanisms will allow better tailored treatments and decision making in the clinical environment (139).

1.2.4 Novel techniques of metabolic profiling in IBD

In addition to more traditional biofluids as discussed above, the application of metabolic profiling has expanded to examining gaseous biological samples to investigate airborne gas phase biomarkers in IBD.

1.2.4.1. Volatile Organic Compounds (VOCs) and the volatilome

Biological volatile organic compounds (VOCs), defined as carbon-based chemicals with a high vapour pressure in room atmospheric conditions, are produced from a multitude of sources within the human body, including, but not exclusively, exhaled breath, sweat, urine, saliva, blood and faeces, and comprise the so-called 'volatilome' (volatile metabolic signature of an individual) (140). As with other metabonomic data these VOCs are recognised to be a measure of an entire system's biology; produced by both host and bacterial host co-metabolic processes and representative of the combined pathological and physiological processes of an individual at any particular time. In addition, VOCs can detect metabolites difficult or impossible to measure by blood tests or alternative means, representing otherwise unmeasurable processes (58) and have been shown to relate to gut disorders, including dysbiosis, measurable distant to the GI tract (141).

Developments of more sophisticated micro and nanotechnology have allowed the detection and identification of minute concentrations of molecules in breath (discussed below) and the

headspace (air above a sample in a sealed container) of urine and faeces and have related these to gastrointestinal diseases, more recently IBD.

The testing of faeces is a rational starting point for VOC testing in gastroenterology, as faeces consists largely of the direct products of digestion and bacterial metabolism, and has characteristic odiferous contents, which have been noted to change distinctly with disease (142).

Sample size	Author	Analytical platform	Differences between IBD and healthy controls	Differences between Crohn's and UC
CD (n =22) UC (n =20) HC (n=19)	Walton (143)	GC-MS	↑ ester and alcohol derivates of SCFAs and indole in Crohn's (not UC)	↑ ester and alcohol derivates of SCFAs and indole in Crohn's compared to UC
CD (n =62) UC (n =48) HC (n=109)	Ahmed (144)	GC-MS	↑ aldehydes (heptanal, propanal, pentanal) in active CD; ↑ Isobutanol, undecane and Methoxy-phenyl-oxime in UC	Not compared
CD (n =83) UC (n =68) HC (n=40)	De Preter (145)	GC-MS	↓ MCFAs and protein metabolites in IBD	↑ Propanal, methylene chloride, benzaldehyde, acetone and 2-Methyl-propanal in Crohn's colitis
CD (n =117) UC (n =100) HC (n=109)	Ahmed (146)	GC-MS	↑ Heptanal, 1-octenol, 2-piperidinone in CD; Bicyclohexane, 4-methylene-1-methylethyl in UC ↓ SCFAs in active Crohn's.	↑ propyl-acetate, 1-octen-3-ol, 2-piperidinone in active CD compared to active UC ↓ 2-butanal in active CD compared to UC

Table 1.5: Table of metabonomic studies using faecal VOCs in IBD – adapted from (125) and (112)

Early GC-MS data discovered a potential relationship between Clostridia species (147), and later other anaerobic bacteria (148) with altered levels of SCFA VOCs in clinical specimens. Garner et al analysed faecal samples with GC-MS in cohorts of healthy donors, patients with bacterial gastrointestinal infections (*C. difficile*, and *C. jejuni*) and in UC, demonstrating significant differences between the groups which enabled excellent predictive diagnostic accuracy (149). GC-MS has subsequently been applied in the analysis of faecal VOCs including a cohort of UC and Crohn's patients by Walton et al. (143). This showed a distinct difference in 8 of the 13 analysed VOCs between IBD (and irritable bowel syndrome) and healthy volunteers, with statistically significantly elevated levels of SCFA derivatives in Crohn's, which normalised after treatment was initiated.

De Preter et al. found a reduction in medium chain fatty acids (MCFAs) (pentanoate, hexanoate, heptanoate, octanoate and nonanoate) and protein metabolites (3-methyl-1H-indole, p-cresol, dimethyl sulfide, 2-methyl butanoate, methyl propyl sulfide and methyl-2-propenyl disulphide) in the IBD cohort compared to controls (145). MCFAs have not been widely reported in association with IBD, although in piglets the addition of MCFAs to the diet has been shown to alter microbially produced metabolites in the small bowel (150). And as seen in other studies (113, 117) De Preter showed SCFAs were reduced in IBD although not to a statistically significant level. This study also linked some metabolites to disease activity in both IBDs, but no rigorous relationship to disease location.

Furthermore in a study by Ahmed et al (144) irritable bowel disease could be separated from both types of active IBD as well as healthy controls, using GC-MS of faecal headspace VOCs, suggesting the potential for developing this for diagnostic purposes in this condition. Using the same method, this group went on to examine a larger IBD population with specific

subgroups with active and inactive disease to compare to controls (146). Distinct VOC profiles separated active from inactive Crohn's and from healthy volunteers, as well as small bowel Crohn's from controls, and colonic Crohn's from UC. In contrast however profiling was not able to separate active from inactive UC or from controls. Authors propose the altered VOCs are a phenomenon of gut microbial dysbiosis in line with other findings.

VOCs relating to the gut microbial composition to are also detectable in the headspace of urine using electronic-nose and Field Asymmetric Ion Mobility Spectroscopy (FAIMS), enabling separation of UC from Crohn's patients, as well as identification of active from quiescent disease (although numbers in the subgroups of this study very small) (151).

1.2.4.2. Exhaled breath analysis

In the search for novel, non-invasive approaches, exhaled breath (EB) analysis has become of increased interest in recent years and is a rapidly evolving frontier in medical research for diseases of the lung and beyond. Breath has been recognised historically as a clinical sign of hepatic encephalopathy (foetor hepaticus) or diabetic ketoacidosis, and ammonia and acetone were among the first compounds to be objectively measured. Pauling et al's landmark (152) paper in 1971 using gas-liquid partition chromatography was the first to identify over 250 substances which represented blood-borne low molecular weight (typically 300 g mol^{-1}) (153) compounds exchanged at the blood/air interface (later called VOCs) of alveoli, that were detectable in exhaled breath (EB). This was the first recognition that breath constituents in minute amounts, representing systemic physiology and pathology, could be identified and quantified.

Current roles for breath testing include monitoring of asthma and detection of airways inflammation, regulation of gases during anaesthetic, alcohol measurement, diagnosis of

neonatal jaundice, and early diagnosis of heart transplant rejection (154). In gastroenterology breath analysis has mostly been used for diagnosis of bacterial conditions such as small intestinal bacterial overgrowth or helicobacter pylori by detecting a chosen end-product of bacterial activity. Up until now the application of breath testing in a clinical context has been limited due to the insensitivity of the available methods requiring pre-concentration manoeuvres (155) and the problems of analysing multiple compounds (156). But there is great potential to exploit the sensitivity of newer technology in analysing the minute fractions of volatile products in EB that reflect systemic conditions. Significant advances in technology such as gas chromatography mass spectrometry (GC-MS) (157) and selective ion flow mass spectrometry (158), enables accurate and reproducible measurement of breath compounds, as well as an opportunity to increase our understanding of the underlying metabolic and biological pathways linked to breath in IBD. (159)

Usually the diagnosis of endoluminal gastroenterological disease requires invasive tests including endoscopy and biopsy. These investigations tend to be expensive, time consuming and carry rare, but serious, possible complications. They can also be relatively uncomfortable and/or unpleasant for patients. Whilst breath testing is unlikely to replace the gold standard diagnostic algorithm in IBD, it may potentially offer a complementary approach to disease monitoring or treatment response; and integration of this data with other methods may yield as-yet undiscovered metabolic evidence relating to pathogenesis. Advantages of breath analysis are that it is non-invasive and highly acceptable to patients. Compared to other metabolic profiling techniques it is relatively inexpensive (160) but still information-rich. Technically its advantages are that no sample preparation is required, and processing and analysis are fast.

Exhaled air is composed mostly of nitrogen, oxygen, carbon dioxide, water vapour and inert gases, as well as, in a relatively tiny fraction, VOCs at a part-per-billion-by-volume range (161). These are a mixture of exogenous and endogenous gaseous components in an approximately 50:50 proportion. Exogenous compounds are present at a negative alveolar gradient, ie the concentration of VOC is greater in the atmosphere than in the breath. The source of these are mostly environmental, including pollutants, and to a lesser degree, diet. Endogenous VOCs are present at a positive alveolar gradient, signifying the source is from within the person sampled. These are generated from pathological and physiological host metabolic pathways, as well as bacterial metabolism (162).

Endogenous VOCs include hydrocarbons (ethane, pentane and isoprene), oxygen-containing compounds (acetone, acetaldehyde, methanol, ethanol, and propanol), sulphur-containing compounds (dimethylsulfide, carbon disulphide) and nitrogen-containing substances (ammonia and dimethyl/trimethylamine) (161).

1.2.4.2.1. Exhaled breath VOCs in clinical research

Alterations in EB profiles have been recognised in *inter alia* gastro-oesophageal cancer(163), fatty liver disease in children (164), Alzheimers disease (165), idiopathic Parkinson's disease (166) and thyroid cancer (167).

There are several developed methodologies of measuring breath VOCs, each with its own pros and cons; and equally these approaches may be combined and complementary to one another. One potential drawback of developing technology is that there is no established consensus of standard operating procedure and not all potential confounders are fully understood (159). A brief comparison of breath analyser techniques is shown below:

Technique	Advantages	Disadvantages
Gas chromatography mass spectrometry (GC-MS)	Complete profile recognition	Not done in real-time Time consuming Needs sample pre-concentration
Ion mobility spectrometry	Faster than GC-MS	No complete profile recognition Not done in real-time Time consuming
Proton transfer reaction mass spectrometry (PTR-MS)	Real time High sensitivity (<ppb) Fast	No complete profile recognition No single VOC identification Limited number of VOCs detectable
Selected ion flow tube mass spectrometry (SIFT-MS)	Real time High sensitivity (<ppb) Fast No maximum of VOCs measured	No complete profile recognition

Table 1.6: Comparison of various analytical techniques in breath research – adapted from (159)

Despite this, the ease of use of exhaled breath analysis makes it an attractive option, as although investigating GI disease and the gut microbiome has been most commonly been performed with faecal analysis, the drawbacks of using stool samples may have limited translation into wide scale clinical research (168). This may be in part due to cost and/or difficulty collecting samples, as well as the recognition that bacterial function rather than composition is equally important (as discussed earlier) to measure and understand. To this end exhaled breath analysis may be supplementary to faecal analysis, and other biofluid metabonomics, and may help answer some vital pathophysiological questions (159).

1.2.4.2.2. Exhaled breath VOCs in IBD

Single EB compounds have thus far been the most studied in IBD, with most published work focusing on pentane and ethane as markers of inflammation and oxidative stress. Kokoszka first demonstrated that breath pentane was associated with intestinal inflammation in rodent colitis models, and subsequently revealed elevated breath pentane in patients with intestinal inflammation, as defined by a positive indium-111-labeled leukocyte scan (169). Sedghi et al. measured breath ethane and pentane in patients with diagnosed IBD and compared this with longitudinal clinical scores, endoscopic data and chemiluminescence of rectal tissue, but found only ethane was positively correlated with disease activity, pentane was not (170). Another study showed that three breath alkanes - ethane, propane, and pentane - were all significantly different in patients with active IBD compared to controls (171). Dryahina (172) et al. examined pentane in a cohort of healthy volunteers and IBD patients. Breath pentane was significantly higher in both subtypes of IBD (greater in CD than UC) than controls, with a very high AUC (0.927) for diagnosing IBD. Although there were statistically significantly different mean levels of pentane between UC and CD, there was a relatively low AUC (0.68) when applying pentane as a discriminatory test and no difference in quantified pentane between the groups with active compared to quiescent disease. However the same group went on to show that in a larger cohort studied, pentane levels were significantly different in patients with disease activity compared to those in complete remission suggesting that pentane concentrations correlated with activity (173). The above studies propose oxidative stress causing excessive lipid cellular fatty acid peroxidation during as the responsible pathway for production of alkanes in blood, subsequently detectable in breath.

Other single EB compounds measured in IBD include: exhaled nitric oxide which was shown to be higher in IBD than controls, and relatively more elevated in UC than CD and in those with active disease (174); methane which is lower in breath in IBD than healthy controls (175); and exhaled carbon monoxide which was not shown to differ in IBD (176).

A single compound is unlikely to be sufficiently sensitive to be able to diagnose or monitor chronic diseases, or reveal information about complex pathological or metabolic processes, especially with possible external influences. Originally breath analysis in IBD had concentrated on single VOCs, however it is apparent though that simultaneous measurement of multiple VOCs to form a breath profile will provide more significant information. (58) Measuring multiple VOCs to form a breath 'profile' or 'fingerprint' is more likely to allow discrimination between conditions, as it reflects all of the concurrent physiological processes occurring at one time (159).

A multi-compound breath analysis in IBD in a paediatric cohort (177) using SIFT-MS breath analysed 21 compounds, of which three alkenes (1-octene, 1-decene, (E)-2-nonene) were identified that could collectively distinguish IBD patients from controls. They were unable to distinguish CD from UC and there was no correlation with disease activity.

FAIMs has also been applied in an adult IBD population, with a good ability to distinguish IBD breath profiles from controls, and a moderate ability to separate UC from Crohn's (178). One of the limitiations of FAIMS is that it cannot accurately identify metabolic compounds, rather relying on a 'breath print' or breath 'bio-signature' to differentiate between conditions. SIFT-MS, whilst more expensive, has a much higher diagnostic power and superior ability to quantify chemicals (178), which is essential in understanding the origin of the altered breath profile.

1.3. The study of IBD and ethnicity

Epidemiological studies show that IBD, which was once a disease almost exclusively confined to Western developed countries, is becoming a global phenomenon, associated with significant mortality and morbidity. IBD prevalence has increased substantially in the last 60 years, at time doubling every decade, with the highest rates currently seen in Canada, Northern Europe and Australia (47). The most recent systematic review of world-wide population-based studies estimated a prevalence of IBD of 0.3% in Western countries, with most of these countries showing a stable or decreasing incidence of the disease (179).

There is considerable variation in the burden of IBD in different geographic regions, mostly seen in a West versus East 'divide' as well as between different ethnic groups in the same location (47). In contrast to the West, the incidence of both Crohn's and UC is accelerating in newly-industrialised countries (179) although limited data makes exact percent calculations impossible.

Most research to date has been based on Caucasian populations (180), with relatively few population based studies of IBD in South Asia (only 6 on the most recent meta-analysis) (179), and even less research in non-Caucasian IBD patient cohorts in Western countries (181). IBD appears to manifest itself as a different clinical phenotype in different ethnic groups (24, 25, 180) and studies suggest that response to therapy and disease progression may be affected by ethnicity (26).

One challenge of studying ethnicity is that race is defined geographically, most often by the continent from which a person originated, whereas the concept of ethnicity is usually self-identified, and encompasses cultural, religious and social aspects as well (182). In this way ethnicity can be considered a social definition rather than a scientific definition (183).

A potential confounder when studying ethnicity, is that through integration of ethnic communities over history, ethnicity terminology is often inadequate to describe the heterogeneity of populations (183), and many individuals self-identify as a mixture, or as more than one ethnicity (184). As with any self-identified characteristic there is potential for mis-reporting. However an epidemiological perspective would strongly favour the use of self-identified ethnicity as the best descriptor in medical research (182). In the studies in this thesis, as with most other medical research into ethnicity, self-reporting is the most practical way of collecting and categorising individuals.

Studying IBD in different ethnic groups offers the chance to reveal significant components of the complex paradigm of pathophysiology IBD. Ethnic subgroups are genetically different from one another, but the phenomenon of migration provides an interesting model to test the effect of environment on the development of IBD, helping to untangle the relationship between genetics and environment.

Additionally, there is interest in this area as the epidemiology of the UK population is changing and through studying IBD in different ethnic groups there is the opportunity better understand IBD in specific patient populations. According to the 2011 National census, England and Wales is becoming more ethnically diverse with rising numbers of all minority ethnic groups (185) including the percentage of self-reported ethnic groups of Asian origin (including Indian, Pakistani, Bangladeshi and other Asian origin) all increasing in size between 2001 and 2011. In this PhD the definition of South Asian includes Indian, Pakistani and Bangladeshi – the three largest groups of Asians in the UK (185) as well as Sri Lankan, as these countries are geographically close. This group make up the second highest percentage (7.5%) of the English and Welsh resident population after White (White British and any other White) (86%). London

is the most ethnically diverse area, with the greatest proportional change over recent years (White British group falling between 2001 and 2011 by 14.9% and Asian increasing by 3%).

In the UK the rate of UC is higher in the South Asian population than non-South Asian with rates of 10.8 per 100,000 compared to 5.3 (186) with the inverse true in Crohn's with an incidence of 3.1 per 100,000 versus 5.3 (187).

In Scotland the relative risk (adjusted for age and socioeconomic deprivation) of UC is higher in Indian men and Pakistani men and women (188). For Crohn's the relative risk is higher in Pakistani men, but only slightly higher in women, with data missing for the Indian sub-population (188).

1.3.1. IBD manifests differently clinically according to ethnicity

It has been shown that IBD phenotype and disease outcomes differs between Asians and Caucasians (189) when studied in native populations of Asia (see table). In combined data from Asian countries the main differences in Asian cohorts compared to Western patients is the male predominance and higher rates of perianal disease in Crohn's, less familial aggregation, less requirement for surgery and fewer extra-intestinal manifestations (EIMs) (190).

However, data amalgamated in this particular paper were compiled from other parts of Asia including Eastern Asia (China, Korea and Taiwan) and South Eastern Asia (Malaysia, Singapore), and in this thesis only patients from South Asia were studied.

Clinical characteristics of IBD	West	Asia
Gender	CD: Female predominance	CD: Male predominance
	UC: Equal gender distribution	UC: Equal gender distribution
Peak age of diagnosis	20–30 years old for CD 30–40 years old for UC	Similar to the West but smaller second peak for CD and UC
CD phenotype	Equal disease distribution with isolated colonic disease predominance in some studies	Ileo-colonic disease predominant Perianal disease more common (33–40%)
UC phenotype	Approximately 30% for proctitis, distal colitis and extensive colitis	Disease distribution similar to the West Milder disease course
Extra-intestinal manifestation	21%–41%	Overall lower frequency (6%–14%)
Primary sclerosing cholangitis	2-7%	0-1%
Colorectal cancer	3%–5%	Lower rates (0%–1.8%)
Colectomy rates	–	Variable in Asia but lower than the West especially for UC
Family history of IBD	10%–25%	Lower rates of familial aggregation (0%–3%)

Table 1.7: Characteristics of IBD in Western and Asian populations - Adapted from (190)

There are a substantial paucity of data lacking from individual South Asian countries including Bangladesh and Pakistan (179). The Indian Society of Gastroenterology (ISG) Task Force reported the first national audit of IBD in 2012, with the main findings being a considerably higher rate of EIMs (over 50% for both disease types) and a higher proportion of extensive UC and relatively rare proportion of proctitis, in contrast to the pooled data from Asia. (191)

The clinical phenotype of South Asian populations in the UK has also been studied. Walker et al showed that South Asians living in London demonstrated a distinct disease phenotype compared to Caucasians in the same area (181). There were some consistencies with the data from Asians studied in native countries, but also some differences, possibly suggesting that disease phenotype manifests differently dependent on environment. Pan-colitis was more common in Asian patients compared to Northern Europeans in the UK (63 vs 42.5%) and proctitis was rare in comparison (9.9 vs. 26.1%). In this cohort there was a higher rate of paediatric Crohn's amongst South Asians and less smokers. The Asian Crohn's cohort had statistically less penetrating disease and less requirement for surgery.

In a pure Bangladeshi cohort studied in London the findings were different (192). Crohn's was more prevalent than UC in the Bangladeshi group compared to Caucasians which is contradictory to other studies of South Asians (186, 192), possibly because of environmental factors related to religion (alcohol, diet) (192). In this analysis there was no difference between the ethnic groups with Crohn's in terms of phenotype, however Bangladeshi's were more likely to require steroids and treatment escalation with thiopurines and anti-TNFs. In the UC cohort, Bangladeshis had more extensive UC than Caucasians but did not require more surgery or escalated medical treatment.

When comparing first generation (born outside of the UK) and second generation (born in the UK) South Asian IBD patients, time spent in the UK did not appear to influence the rate of UC - consistent with other epidemiological data (186) – as seen but did the risk of Crohn's (risk increasing with time) (181).

Although data is limited to small cohorts, there is some evidence that ethnicity affects treatment response in UC to 5-ASAs / sulfasalazine therapy with 49.6% relapsing in an Asian study

population (193) compared to 72% relapsing in a Western cohort (194). Although it is worth noting that the two groups were not directly compared. Steroids in UC appear to be comparably effective (189) in both ethnicities. There may be a higher response rate to biologics (unpublished data) in Asian countries although this has not been fully investigated and use of these therapies is limited (189). There is no data specifically in South Asian patients in Asia or the UK.

1.3.2. The changing global picture of IBD

The areas with the most significant and rapid increases are developing countries, including Asia (195) whilst the incidence in developed countries appears to be stable (196). Although some of this is due to improved access to healthcare and diagnosis reporting in developing areas, it is certainly a true epidemiological occurrence. In fact incidence rates and prevalence values from developing and under-developed countries are likely underestimated (47). Whilst the genetics of a population can be considered to be essentially unchanged over time, environment must account for at least a substantial influence on the growing incidence in these regions. Influences such as improved hygiene and socioeconomic development (196) are proposed to explain the epidemiological pattern of IBD following that in the West but lagging by 50 years (197). These factors interplay with other pathological mechanisms discussed above, including diet and microbial dysbiosis.

1.3.3. The effect of migration on IBD in a population

Genetic susceptibility in Asian populations is different compared to Caucasians (198). Less familial clustering has been seen in non-Caucasian populations than is observed in Caucasians (199) and there is a global difference in the susceptibility genes for IBD seen in different

geographical and racial groups. The major NOD2/CARD15 polymorphisms associated with susceptibility to Crohn's in Caucasians are rare or absent in Indian populations and a polymorphism in the interleukin 23 receptor (ILR23R) gene was not protective against IBD in the Indian cohort as it is in Caucasians(200).

Migration provides an ideal model in which to study the effect of environment on ethnic groups, and there is a clear increased risk of developing IBD in those who migrate from a low prevalence area to a high prevalence region (201). Data have shown that second generation Asians in the UK are at even higher risk than the indigenous population of developing IBD (202, 203). It is also seen that IBD is more common in urban than rural regions within the same country (204) and there may be a variety of different environments risks to account for both of these findings including diet, lifestyle behaviour, medication, pollution and sanitation (205). The effects of all of these factors are notoriously difficult to measure as they cannot be adequately controlled for in prospective studies and therefore we rely on observational data.

In order to understand the temporal and geographical trends of IBD more studies are needed from developing countries and in non-Caucasian populations (47). Through examining the data in IBD from ethnically and genetically different populations fundamental aspects of IBD pathogenesis can be examined and applied in the context of a diverse multicultural clinical population.

1.4. Scope of this thesis

This thesis explores the further use of metabolic profiling in IBD, in a variety of stratified patient groups, and examines different aspects of how metabonomics can be better understood and refined, to ultimately improve its translational impact to the clinical arena. Each of the 4 chapters includes the results, as well as a discussion and conclusion within the chapter.

Chapter 2: Materials and Methods

To prevent repetition this is an overall shared methodology chapter as several studies use the same techniques. Specifics relating to each individual dataset are further defined in each individual chapter.

Chapter 3: Exploring the effect of ethnicity on urinary metabolic profiles in IBD

This part of the thesis studies urinary metabonomic data from ethnically different IBD populations in order to better understand the important metabolic differences between these groups.

Hypothesis: Ethnicity has a significant effect on the urine metabolic profile in inflammatory bowel disease which may relate to the microbiome, and reflects that South Asians are an under-recognised cohort in IBD research.

Chapter 4: Effects of time on urinary metabonomic signatures in IBD

This chapter examines the stability of the urinary metabolic profile in IBD over time, in order to improve interpretation of this methodology, to assess changes associated with disease and to examine the potential of urinary metabonomics as a predictor of disease outcome.

Hypothesis: Urinary metabonomics may separate patients who subsequently develop disease progression from those with a less aggressive phenotype, and metabolic profiles are altered by disease complications and medical interventions over time.

Chapter 5: Pilot Study integrating microbial and metabonomic data in IBD

In this cross-sectional study metabonomic data are integrated with bacterial compositional data to explore correlations between key urinary and faecal metabolites and the microbiome in IBD.

Hypothesis: Marker metabolites identified in urine and stool differentiating IBD from controls can be directly correlated to dysbiosis in the disease through identified alterations in gut bacteria.

Chapter 6: Exhaled breath VOC analysis as a novel metabolic profiling technique in IBD

Here, a novel metabolic approach to non-invasive testing in IBD is performed, using breath analysis for the first time in adults to explore the exhaled compounds associated with the disease.

Hypothesis: Exhaled breath VOC profiles differ in IBD and can be used to discriminate patients from controls.

Chapter 2: Materials and Methods

2.1 Study participants and clinical data

Ethical approvals for the studies were obtained from St. Mary's Research Ethics Committee (Ref 05/Q0403/106) with additional later amendments from the Fulham NRES Committee.

All patients gave written, informed consent and received a patient information sheet. Healthy volunteers also gave written, informed consent and received a volunteer information sheet. Demographic data including age and self-reported ethnicity was recorded.

2.1.1 Clinical phenotyping of IBD patients

IBD patients were diagnosed with IBD based on rigorous review of clinical data, endoscopy, histology and imaging (4, 146). Diseases were sub-classified according to the Montreal classifications (206). Disease activity was assessed using the Harvey-Bradshaw index (HBI) for CD patients (207) or the simplified clinical colitis activity index (SCCAI) for UC patients (208). Disease was considered active if the score was ≥5.

Disease information was collected from a thorough review of medical notes, clinical letter database, radiology and pathology results for each patient.

IBD medications were recorded as 5-aminosalicylates, immunosuppressants (defined as azathioprine or 6-mercaptopurine (6MP)), anti-TNFs (infliximab or adalimumab) and steroids(209).

Patients with significant co-morbidities were also excluded. These were considered to be systemic conditions which may influence the metabolome, and included cardiovascular disease, respiratory disease, neurological conditions, infections, cancer and atopy.

For historical samples, recorded demographic and phenotypic data was used and a review of any available historical records was performed. Where corresponding information was missing for patients their samples were excluded from the analysis or sub-analysis.

2.1.2 Metabonomic and dietary information

In addition to this information, patients and volunteers recorded dietary data (24-hour recall and usual weekly), alcohol use, smoking history, exercise and other medications. No dietary restrictions were imposed to simulate real life circumstances; however it was recorded if a participant was vegetarian or not.

2.2 Sample collection and preparation

2.2.1 Urine samples

Urine was collected in 30ml universal sample containers (Sigma-Aldrich, USA). Random urine was taken rather than first void (early morning) samples as these have been shown to exhibit more intra-individual variation due to lifestyle and diet than those collected later in the day (210). Samples were collected within 6 hours of production.

Specimens were centrifuged at 14,000 rpm for 20 minutes to removed solid particulate matter, and then aliquoted and stored at -80^{0}C in microcentrifuge tubes (Ependorff, Germany) until NMR analysis.

2.2.2 Faecal sample collection and water extraction for NMR

Stool was collected in a 30ml universal sample container with spoon cap (Sigma-Aldrich, USA) and frozen at -80^{o}C on the same day of sample donation (within 12 hours) until analysis.

To extract faecal water 500mg of faeces was mixed with sterile phosphate buffered saline in a 2:1 ratio and vortexed for 5 minutes. The resultant mixture was centrifuged for 10 minutes at 18000*g* at 4ºC. The supernatant was stored in aliquots of 600µL in microcentrifuge tubes for NMR analysis.

2.2.3 Extraction of DNA from faeces for microbial analysis

Extraction of DNA from faecal samples was performed using the QIAamp Fast DNA Stool Mini Kit. Laboratory materials were supplied by Qiagen, unless otherwise stated. Centrifugation was performed at 14,000rpm, at room temperature, unless stated otherwise.

2.2.3.1 Preparation of the reagents used

Buffers AW1 and AW2 were diluted to correct concentration with 100% ethanol. Buffers AL and InhibitEx ® were incubated at 21ºC to dissolve any precipitates and all buffers were mixed thoroughly before use.

2.2.3.2 DNA extraction

For each specimen 180-220mg of stool was weighed into a 2ml microcentrifuge tube (Ependorff, Germany) and kept frozen on ice. 1ml of InhibitEx® buffer was added to each sample, and then vortexed for up to 5 minutes to ensure sample was homogenised.

The suspension was then incubated to 95ºC in a thermomixer for 5 minutes to lyse cells. Samples were then mixed again on the vortex for 15 seconds. Samples were centrifuged for 1 minute to pellet stool particles or until there was no solid matter remaining in the suspension.

15ul of Proteinase K was added to new 1.5ml microcentrifuge tubes and 200ul of supernatant was added. These were mixed before addition of 200ul of Buffer AL and a further 15 seconds

on the vortex to mix samples. Tubes were heated to 70ºC for 10 minutes before 200ul of 96-100% ethanol was added and mixed again on the vortex.

600ul of lysate was pipetted into a QIAamp spin column in a 2ml collection tube and centrifuged for 1 minute, or until all filtrate had passed through the spin column. The collection tube containing filtrate was discarded and replaced.

500ul of Buffer AW1 was inoculated on the spin column and centrifuged for 1 minute and the collection tube containing filtrate was discarded and replaced.

500ul of Buffer AW2 was added and centrifuged for 3 minutes. In order to prevent any residue of Buffer AW2 contaminating the QIAamp spin column, the spin tube was discarded and replaced again, and spin columns centrifuged for an additional 3 minutes.

QIAamp spin columns were placed in 1.5ml microcentrifuge tubes and 200ul Buffer ATE was pipetted onto the membrane of the column. After 1 minute at room temperature samples were centrifuged to elute DNA. Microcentrifuge tubes were then stored at -20ºC until quantification of DNA was performed.

2.2.3.3 Quantification of DNA

Quantitation of DNA was performed using a Qubit™ fluorometer and quantitation assays. Standards #1 and #2 were used to calibrate the fluorometer prior to reading samples. 200µl of each sample was transferred to 0.5ml Qubit® assay tubes and incubated at 20ºC for 2 minutes before each reading.

Fluorescence was measured and expressed as a concentration of DNA in the assay tube in ng/mL.

2.3 Microbial diversity analysis of faecal samples

Culture-independent analysis of the gut microbiome has improved the exploration of the gut microbiome and its role as a central driver in host health and disease. Not only is a significant percentage of the gut microbiome as yet uncultured (79) but newer techniques including metagenomic and 16S rRNA gene-based methods are also able to describe the functions of the microbes as well as identify the species present. Whilst it is known that the luminal contents of the gut differ from the mucosal-associated microbiota (211) it is established that faecal microbial analysis is more practical, safer and delivers a valuable mechanism into investigating the gut microbiota (212).

2.3.1 16S rRNA gene analysis

16S ribosomal RNA analysis is becoming a prevalent method of rapid, cheaply and accurately quantifying the bacterial populations within clinical samples. It enables the description of bacterial diversity within a sample by creating inventories of 16S rRNA genes. This is not to be confused with metagenomics analysis which is the measurement of the entire genetic data within a sample.

Firstly DNA is extracted from the biological samples, then polymerase chain reaction (PCR) is performed on specified regions of the 16S gene using ‘universal’ primers aimed to include as wide a range of microorganisms as possible. The generated data is then compared to validated microbial sequencing databases for identification purposes.

Data demonstrates the clustering of related sequences at a particular level of identity, called operational taxonomic units (OTUs). OTU counts for each organism in a sample can be shown with a level of 97% usually being representative of a species and 95% being representative of a genus. Some bacteria can only be identified at the genus or family level rather than species.

2.3.1.1 Illumina 2 step technique

Samples were analysed by Research and Testing Laboratories, Lubbock, Texas, with the Illumina 2-step technique. Methodology below supplied by Research and Testing:

V1-V2 regions were amplified for sequencing in a two-step process. The forward primer was constructed with (5'-3') the Illumina i5 sequencing primer (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG) and the forward gene specific primer. The reverse primer was constructed with (5'-3') the Illumina i7 sequencing primer (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG) and the reverse gene specific primer. Amplifications were performed in 25 ul reactions with Qiagen HotStar Taq master mix (Qiagen Inc, Valencia, California), 1ul of each 5uM primer, and 1ul of template. Reactions were performed on ABI Veriti thermocyclers (Applied Biosytems, Carlsbad, California) under the following thermal profile: 95°C for 5 min, then 25 cycles of 94°C for 30 sec, 54°C for 40 sec, 72°C for 1 min, followed by one cycle of 72°C for 10 min and 4°C hold.

Products from the first stage amplification were added to a second PCR based on qualitatively determine concentrations. Primers for the second PCR were designed based on the Illumina Nextera PCR primers as follows:

Forward - AATGATACGGCGACCACCGAGATCTACAC[i5index]TCGTCGGCAGCGTC
Reverse - CAAGCAGAAGACGGCATACGAGAT[i7index]GTCTCGTGGGCTCGG.

The second stage amplification was run the same as the first stage except for 10 cycles.

Amplification products were visualized with eGels (Life Technologies, Grand Island, New York). Products were then pooled equimolar and each pool was size selected in two rounds

using Agencourt AMPure XP (BeckmanCoulter, Indianapolis, Indiana) in a 0.7 ratio for both rounds. Size selected pools were then quantified using the Quibit 2.0 fluorometer (Life Technologies) and loaded on an Illumina MiSeq (Illumina, Inc. San Diego, California) 2x300 flow cell at 10pM.

2.3.1.2 16S data processing

Mothur (213) was used to process the 16S rRNA gene sequences generated by the Illumina MiSeq platform. Forward (R1) and reverse (R2) reads were paired, ensuring any low-count reads were excluded. Sequences containing ambiguous bases were removed and duplicates removed, leaving only unique sequences which were then aligned to customised references. Poorly aligned or redundant sequences were culled, leaving groups of sequences to be sorted into groups with abundances. Taxonomy was assigned against the SILVA rRNA database (214).

Data were transferred from Mothur format to be used in Microbiomeanalyst software (215) online.

2.3.1.3 Microbial composition analysis

Relative bacterial abundance (percentage) was measured at phylum, class, order, family and genus levels and shown in stacked bar charts.

Heatmap visualisation with clustering dendrogram was produced to visualise the relationship between bacteria and samples as per their direct (Euclidean) distance. Heatmaps use a colour super-imposed on a data matrix to visualise the larger and smaller values. Hierarchical clustering demonstrates similarities (close distance) by linking data to one another in a ranked

(most similar) order of height. This is used as an overview to highlight associations with particular bacteria.

2.3.1.4. Diversity index analysis

There are a variety of metrics which can be used to describe the diversity of biological samples. Estimators of 'within sample' (alpha) diversity can measure the richness (number of species per sample) and evenness (includes the relative abundance of difference species in a sample). The Observed OTU count and Chao1 index describe richness alone, Shannon index and Fisher's index describe both richness and evenness of the samples. These indices can be measured and compared between groups and a p value ascribed to determined significance.

As each of the indices are calculated slightly differently (ie are slightly difference representations of the same biodiversity) it is helpful to measure and quote several indices when comparing groups.

2.3.1.5. Beta diversity analysis

To compare the composition of different bacterial communities between groups beta diversity is calculated. This is a multivariate technique which measures the distance (or dissimilarity) between every pair of samples, to form a dissimilarity matrix.

Principal coordinate analysis (PCoA) (also known as multidimensional scaling, MDS) enables visualisation of the matrix in low dimensional (2D or 3D) space to show similar clusters, or highlight outliers, based on their Euclidean distance. In this there is similarity to primary component analysis (PCA, described later) however the difference between the two is that PCA is based on similarities and PCoA on dissimilarities (216).

The interpretation of a PCoA is whether samples cluster together based on the degree of dissimilarity, and further interpretation beyond this is limited. Each axis has an eigenvalue which describes the variance accounted for in that axis. The higher the cumulative eigenvalues the better the PCoA displays differences. R^2 values are sometimes displayed to quantify the variance demonstrated, with models being tested for significance with a p value.

PCoA can be constructed using a variety of metrics to measure distance (dissimilarity) including Bray-Curtis (Non-phylogeny based, taking abundance into account), Unweighted UniFrac (based on phylogenetic branches, and OTU count) and Weighted UniFrac (phylogenetic and OUT count and abundance) (217).

2.3.1.6. Differential Abundance analysis

The significant bacteria (at each taxonomic level) between groups were identified with EdgeR (218), a bioinformatics package which calculates statistical significance of the bacteria accounting for biological variability. P values (<0.05 considered significant) were corrected for false discovery rate.

2.4 Urinary and faecal water NMR spectroscopy

2.4.1 NMR spectroscopy

Nuclear magnetic resonance spectroscopy is used throughout this thesis to allow biological samples (urine and stool) to be interrogated for the molecular structures within it and thereby the metabolic composition. NMR relies on a radio frequency pulse to cause the atomic nuclei of a sample to resonate at a specific frequency whilst the sample is held in a strong external magnetic field (219).

2.4.1.1 Basics of NMR Spectroscopy

Some atomic nuclei, including protons (^{1}H), possess the quantum mechanical property called spin, whereby when placed in a magnetic field (B_0), they align in one of two orientations - either a higher (parallel) or lower (anti-parallel) energy spin state. Applying a radiofrequency excitation pulse (B_1) causes a transition in the energy state of the nuclei, and once this is removed the nuclei recover to equilibrium state, releasing absorbed energy. This signal, known as free-induction decay (FID) is detected and measured by receiver coils, and characterised by the time it takes to return to original spin (spin-lattice relaxation - T1) and the amplitude of the relaxation in the transverse plan (spin-spin relaxation - T2) (220). The FID is resolved into a frequency domain spectrum by Fourier transformation, displaying the energy difference between the protons two spin states. This is displayed as spectrum, where the chemical shift δ (expressed in parts per million) describes the difference between the resonant frequency of the nucleus and a reference signal, in this case 3-trimethylsilyl-1-1-(2,2,3,3,-$^{2}H_4$) propionate (TSP), a chemically non-reactive standard.

The specific chemical shift of each nuclei is determined by several factors, including chemical shielding, electronegativity and electron density and in addition to this the interaction of neighbouring nuclei in a covalent bond, known as 'spin-spin coupling', which gives rise to multiple peaks in a spectrum determined by the number of protons bonded to adjacent atoms. Prior knowledge of the multiple peaks and chemical shifts allows identification of the molecules present in a complex mixture. The signal intensity is proportional to the number of nuclei contributing to that signal, and so by integrating the area under a peak a relative quantification of a compound can be made (219).

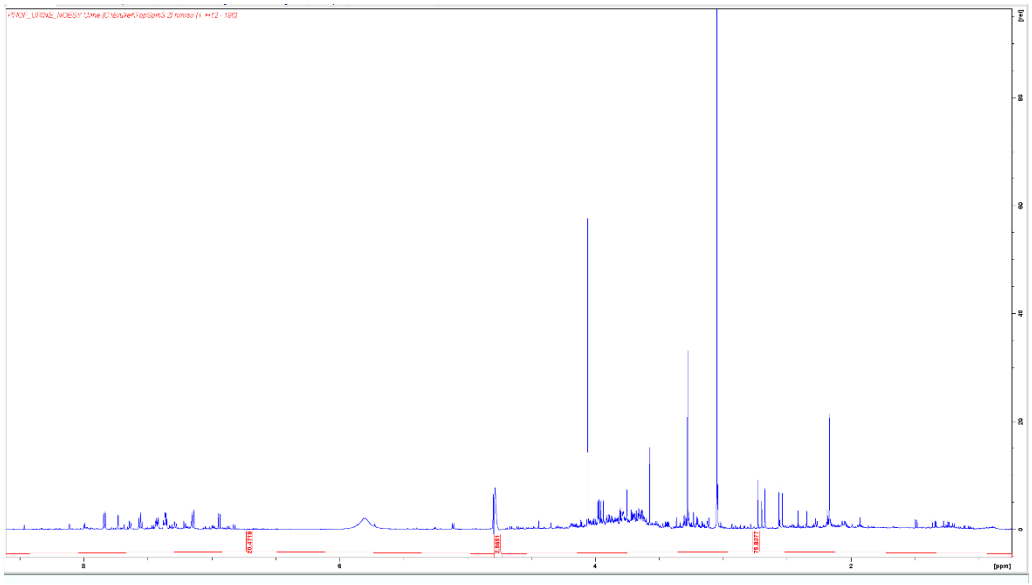


Figure 2.1: Example of NMR spectrum

2.4.2 Preparation of urine samples for NMR

Urine samples were prepared using a standardised method (221). Laboratory agents were supplied by Sigma-Aldrich, USA, unless otherwise stated.

400µL of urine was mixed with 200µL of phosphate buffer (pH7.4) to stablise pH. The phosphate buffer contained TSP to act as a reference standard, and D_20 to act as a field lock. Samples were centrifuged at 12,000*g* for 5 minutes then 550µL was transferred to sterilised 5mm NMR tubes.

2.4.3 Preparation of faecal water samples for NMR

Faecal water samples were prepared by the same standardised method in 2.4.2.

2.4.4 NMR Data acquisition and pre-processing

2.4.4.1 Data acquisition

NMR spectra was acquired on an Avance 600MHz NMR spectrometer (Bruker Biospin) using standardised protocol with water presaturation. Samples were randomised to reduce the effect of temperature shift and field drift during the acquisition. Standard 1-dimenssional NOESY spectra were acquired over 32 scans (4 dummy scans) with the receiver gain at 90.5 with a 90^{o} high power pulse angle for 13.35μseconds (Bruker pulse programme: noesypr1d).

^{1}H-^{1}H correlation spectroscopy (COSY) and ^{1}H-^{1}H total correlation spectroscopy (TOCSY) NMR spectra were also acquired to aid assignment of peaks to metabolites.

2.4.4.2 Pre-processing

Acquired free induction decays for one-dimensional data were multiplied by an exponential window function with a line broadening of 0.3Hz, followed by a Fourier transformation. Phasing and baseline correction were performed using in-house software. Referencing was set to TSP (δ 0.0). Spectra were visualised using Topspin (3.5, Bruker BioSpin 2006) then imported into Matlab R2014a (Version 8.5, Mathworks) for processing and analysis.

Regions containing the water resonance, urea (in urine) and TSP were removed (see individual chapters for specific chemical shift details). Full resolution spectra were used. The advantages of this are that assignment of peaks is more accurate, and the full data information is retained. Alternatively 'binning' can be used, where the spectral intensity from small pre-defined regions are combined into a variable. This makes the reduced spectral dataset more manageable, and can overcome some effect of chemical shift, however it reduces the accuracy of data and can confuse statistical modelling. In order to overcome the effects of peak shift associated with

variations in pH and ionic strength of samples, the spectra were aligned using an in-house algorithm: recursive segment-wise peak alignment (RSPA) (222). This allows better comparison of spectra and removes artefactual variation to improve robustness of analysis (222).

Clinical samples invariably contain a variety of peaks representing xenometabolites produced from medications. There is debate in the literature as to the optimal method to account for these when analysing data (223). Other published works have excluded large regions (δ 2.12-2.22 and δ 4.70-9.50) of the NMR spectra to control for this (117), but this excludes a substantial proportion of the biologically relevant metabolite data, including the aromatic region in which many important peaks previously reported to be altered in IBD are located. Three approaches were used in this thesis. Firstly, the most rigorous way to prevent xenometabolites influencing the statistical models was to exclude all of the samples which included drug peaks. However, as many IBD patients take 5-ASA medications and or paracetamol (which both have large dominant drug peaks in the aromatic region), this approach was only possible in the largest study (ethnicity: Chapter 3) otherwise the subgroups became too small for accurate analysis. This is the most robust method, as it not only removes the xenometabolite signals themselves, but also removes the effect of medication on the other metabolic constituents. Alternatively, in the smaller studies, the regions in which the drug peaks were present were either individually removed from the spectra prior to processing the data, or were selectively removed using an in-house script for Statistical Total Correlation Spectroscopy (STOCSY) Editing (224). This edits out the resonances highly statistically correlated with a chosen peak, at a pre-defined covariance set by the user (usually 0.9). It allows removal of the drug peak whilst leaving the unrelated resonances intact and without the need to exclude the entire region. This was performed on 5-ASA (225) and paracetamol (226) metabolites in some studies in this thesis.

Median-fold change normalisation (also known as probabilistic quotient) (227) was performed to account for metabolome-wide effects such as urine dilution / stool concentration. This calculates a ratio (quotient) of each variable in relation to a reference spectrum calculated from all the samples. This identifies changes in a sample which are due to biological effect, rather than overall changes to affecting the whole sample such as dilution.

Scaling and mean centring was then performed. Scaling is performed to prevent the models being dominated by the large peaks of high concentration metabolites (228). There are several alternative methods of scaling and the optimal method is best determined by the specific dataset itself (229). Univariate scaling divides each variable by its standard variation thereby giving all peaks equal weighting (of one). Pareto scaling uses the square root of the standard deviation as a scaling factor instead of the standard deviation alone, and allows enhancement of the contribution of small concentration metabolites to be visualised. Logarithmic scaling can also be used as an alternative, which removes heteroscedasticity (values with different variabilities in a dataset) and reduces the differences between very high and low value metabolites by pseudo-scaling (230).

2.5 Breath testing with SIFT-MS

SIFT-MS uses selected precursor ions (H_3O^+, NO^+ and $O2^+$) injected into flowing helium gas carrier, to ionise the trace compounds in a gaseous sample, in this case breath. These precursor ions do not react significantly with the major components of air (nitrogen, oxygen etc) but form characteristic product (analyte) ions with trace volatiles; thus with knowledge of the ion-molecule chemistry, accurate identification of trace compounds can be made to a parts-per-billion by-volume (ppbv) or parts-per-million by-volume (ppmv) level. As the sample is

introduced at a known flow rate, hence determining the reaction time, the count rates can be converted to quantification allowing sensitive and accurate real-time trace-gas analysis.

SIFT-MS can run in 2 modes: Full scan mode (FS) in which the downstream analytical mass spectrometer/detection system is scanned over a pre-determined time period over a range of mass to charge (m/z) values, or multi-ion mode (MIM), during which the spectrometer is rapidly switched between selected m/z values to target selected trace gas species. MIM is more sensitive and reproducible than FS mode (231) and therefore was employed for this thesis.

2.5.1 Exhaled breath capture technique

Breath capture technique and SIFT-MS details are described in chapter 6.4.2 – 6.4.3.

2.6 Data analysis

Data treatment and statistical analysis was performed in either Excel, GraphPad Prism 6 (GraphPad Software Inc. CA, USA) or IBM SPSS statistics 21 (SPSS Inc., Chicago, IL). For analysis of phenotypic data the Mann-Whitney U test or Kruskal-Wallis test (as appropriate) were applied for continuous variables and Fisher's exact test for categorical variables.

2.6.1 Targeted analysis of hypothesised metabolites in NMR spectra

Targeted analysis of metabolites was performed to identify and elucidate important metabolites in the models which had been hypothesised as contributing to separation in the datasets. The area under the peak for each specified metabolite was integrated giving a relative index, from which the median was used to compare groups. Kruskall-Wallis and / or Mann-Whitney U test were calculated and to correct for multiple comparisons the Benjamini-Hochberg or Bonferroni correction was applied.

2.6.2 Multivariate statistical analysis

The most common approach to interpreting the highly complex nature of metabonomic datasets (NMR or SIFT-MS) is with multivariate statistical analysis. These techniques try and reduce the dimensionality of the data and improve the ease of handling and understanding of pattern recognition. For the NMR data this was undertaken in Matlab, for the SIFT-MS data Pirouette v4.0 (Informetrix, Inc. Bothwell WA, USA) was used.

2.6.2.1 Principle component analysis

Principle component analysis (PCA) is an 'unsupervised' approach, by which it means a model is constructed without knowing the class to which each sample belongs. This provides an overall structure of the samples relative to each other, and can show clustering relationships as well as outliers (biological or pathological). The scores are displayed on a scores plot which is a 2D graphical representation of the co-ordinates of each sample in 3-D space.

'Principal components' (PCs) represent the combinations of metabolites accounting for the variation of the dataset as a whole, with the first principal component being the strongest contributor, the second PC being the second greatest etc.

2.6.2.2 Partial Least Squares Discriminant Analysis

Partial Least Squares discriminant analysis (PLS-DA) is used to is to identify class differences from multivariate dataset. By assigning class membership an algorithm is generated to expose the separation between classes, expressed as a score in space (232). However not all variation is related directly to the class, and therefore an orthogonal component can be added (to become O-PLS-DA), so that the described variation is attributable to the ascribed class, allowing enhanced biological interpretability. Of note, adding an orthogonal component does not offer

increased statistical advantage; nor necessarily change the model in PLS-DA but can aid interpretation of models (233) In a similar fashion OPLS or PLS can show the association between continuous measurements and the data rather than discrete classes.

2.6.2.2.1 Loadings

Loadings describe the relative influence of the variables driving the separation in the model. Loadings can be visualised in a pseudo-spectrum in which positive or negative correlations related to class are demonstrated, enabling the identification of metabolites contributing to the model (234, 235).

2.6.2.2.2 Validation

'Over-fitting' by which samples are forced into models to allow class separation is a risk of PLS-DA statistics; validation is therefore essential to ensure reliability of the model. A type of internal cross-validation must be employed – either leave-one-out-cross validation (LOOCV) or 7-fold cross validation (7FCV). These techniques create multiple permutations of the model to test its statistical strength, by creating a training set (either by excluding each sample in turn (LOOCV) or every seventh sample (7FCV)), and attempting to re-predict the class of the excluded sample back into the model. In this way each sample also acts as a control for the remainder of the dataset. So whilst randomly generated class assignment can show good separation on a model, it would not stand up to internal cross validation.

The quality assessment (Q^2) statistic is generated as a measure of the predicted and original data to provide a qualitative measure of the predictability of the model (where $Q^2 = 1$ indicates perfect predictability).

The R^2 value is also be given, allowing an estimation of how well the model explains the variation in the data. Highly disparate R^2 and Q^2 values suggest forced 'over-fitting' of the data.

Permutation calculation is further evidence of the reliability of a model, which unlike internal cross validation does not require leaving out portions of data. By generating random models (100 up to 1,000 as specified) it tests the null hypothesis that the Q^2 of the model is derived by chance, and thus can give a p value in which $p<0.05$ would be considered significant.

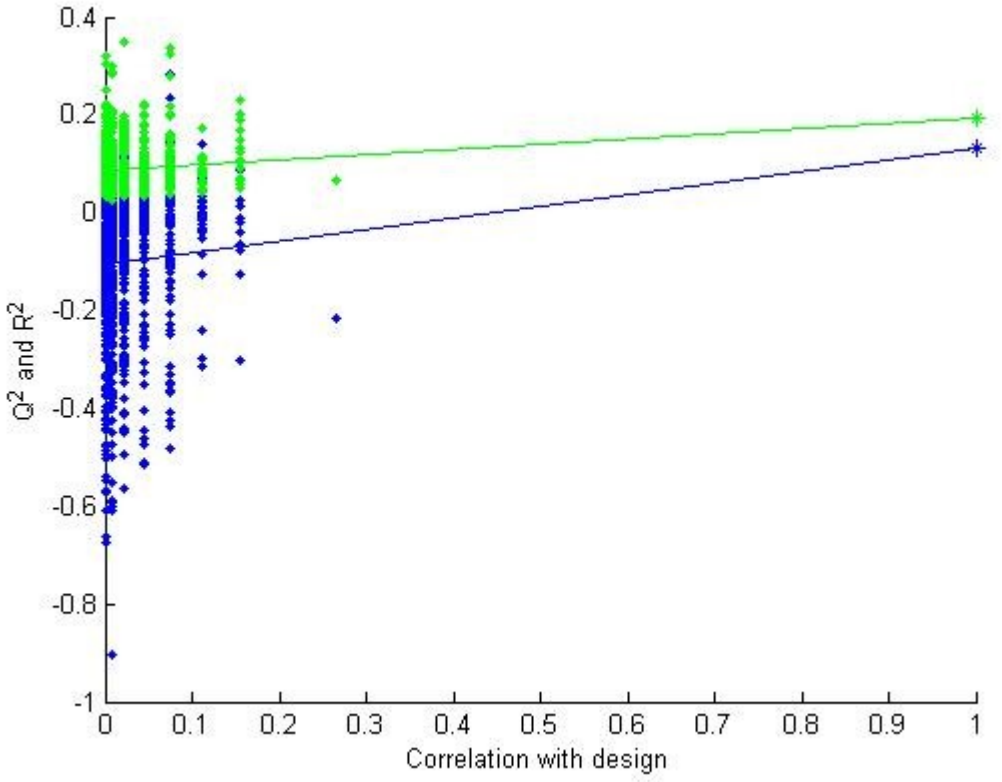


Figure 2.2: Permutation testing plot (p=0.001 with 1000 permutations)

2.6.2.3 Assignment strategy for NMR spectra

Human urinary and faecal metabolites were identified in the spectra according to published data or the Human Metabolome Database – an electronic database which describes over 40,000 metabolites with chemical and spectral information, including 1600 drugs / drug metabolites. If a peak could not be assigned this way a STOCSY was performed to show associated peaks, and Chenomx NMR Suite (8.2) used. This is a software package which allows importation of spectra for comparison against known metabolites to improve identification (236).

Once identified the area under the peak for each metabolite was specified in ppm and integrated with a software script (integrate_JMP.m) to give the correlation coefficient (*r)*. To ensure that the metabolites were statistically significantly correlated, the non-directional probability was calculated for each *r* value, where a probability of <0.05 was considered significant.

2.7 Integration of metabonomic and bacterial data

Matlab R2014a (Version 8.5, Mathworks) was employed to analyse the correlations between the metabonomic and 16S rRNA bacterial data using Spearman's correlations. The statistically significant metabolites identified from the metabonomic data, and the candidate bacteria differentiating the groups were combined to generate heatmap correlograms. Each correlation was tested for significance with a p value ≤ 0.05 considered to be significant. Only significant correlation coefficients were plotted, with blank spaces for no correlation. Each result was represented as a coloured pie charts, in red (positive) or blue (negative) with the proportion of pie representing strength of correlation.

Chapter 3: Exploring the effect of ethnicity on urinary metabolic profiles in IBD

3.1 Summary

In this large study, the urinary metabolic profiles of South Asian and white Northern European IBD patients were compared to healthy controls and to each other. Two complementary techniques were used to analyse the data to improve the validity of results and further investigate the effect of IBD and ethnicity on the metabolome.

In the largest IBD metabonomics study to date, univariate and multivariate analysis of the white Northern European results were consistent with previous studies differentiating IBD from controls and subtypes from one another. Confirmatory metabolites were seen as previously shown, relating to the gut microbiome and inflammatory related energy cycles. Several new differentiating metabolites were also identified.

In South Asians Crohn's could be robustly separated from controls using OPLSDA, but UC could not; despite univariate analysis showing differences between ethnicities in specific measured metabolites. Reasons for this are explored and discussed.

In comparing both healthy controls and groups of IBD of both ethnicities, interesting and novel differences were seen in both univariate and multivariate models. The metabolites holding differential power between Northern Europeans and South Asians in both healthy volunteers and IBD were hippurate and p-cresol, both microbial co-metabolites. When examining the metabolome alterations, it is likely that diet, genetics and gut bacteria combine in so-called a host- nutrition-microbiome interaction. These findings demonstrate that in order to realise the potential of metabolomics in an IBD population it is essential to account for influences of ethnicity and diet when stratifying patient groups.

3.2 Aims and hypothesis

i. The aim was to examine and compare the urinary metabolic profiles of IBD patients and controls from white Northern European and South Asian backgrounds.

ii. It was hypothesised that IBD patients of both ethnic backgrounds would show alterations in urinary metabolites demonstrable by metabolomics, and these metabolites would be related to the gut microbiota, tricarboxylic acid (TCA) energy cycles and amino acids, as demonstrated in previous literature.

iii. It was hypothesised that South Asian IBD patients would share some similarities to their white counterparts, enabling them to be distinguished from healthy controls using urinary metabonomics. But it was also hypothesised that South Asian patients and controls would have differences in the metabolic profile compared to white Northern European equivalents reflecting their different backgrounds.

3.3 Introduction

The heterogeneity of the inflammatory bowel diseases, including genetic predisposition (198), clinical phenotype (181) and treatment response (26) has been demonstrated across different ethnicities. IBD has been sparsely studied in populations other than Caucasians (181), but with the fastest increase in incidence of IBD in the developing world (195), and an incidence in South Asian migrants superseding that of the native UK population (202), it is essential to understand more about IBD among different populations. In the UK the phenotype of the population is also changing; South Asians are the second largest ethnic group after whites in the UK, with a rising percentage of South Asians and falling white British percentage (237). Recent consensus statistics also record the number of IBD patients of non-Caucasian origin

across Western Europe is increasing (238). Hospitalisation and diagnosis of IBD in minority ethnic groups are increasing significantly, with the most marked increase in Asians (239).

Current management algorithms used in the UK are based on guidelines derived from largely Caucasian data (240), but with the advent of personalised medicine, it is essential to be able to tailor the approach to managing IBD in individual patients, accounting for variables including ethnicity.

It has been well established in the white UK population that urinary NMR profiling can separate patients with Crohn's and UC from healthy controls (118), with gut bacterial metabolites having a dominant role in differentiating these groups (118). It is also established that the gut microbiota differs according to both geography and ethnicity (241) but the interrelationship of ethnicity, environment and the gut microbiome is not clear. Research in different ethnic populations are needed to address these questions and better understand IBD in these cohorts.

The design of this study was to examine urinary metabonomics in large groups of IBD patients and healthy controls from white and South Asian ethnicities to help understand the similarities and differences between disease groups from these backgrounds. It was also designed to gain insight into this research technique in a new cohort and how ethnicity may affect the way these data are interpreted.

3.4. Methods

3.4.1 Subjects

The study group totalled 542, of which 405 were patients and 137 were self-reported healthy controls. Diagnosis of IBD was based on clinical, endoscopic and histological data collected from notes and electronic radiology/pathology databases.

IBD patients were recruited from St Mary's and Hammersmith Hospitals (Imperial College NHS Healthcare Trust) in gastroenterology clinic. And also from West Middlesex Hospital, Ealing Hospital, Northwick Park Hospital and John Radcliffe Hospital, Oxford.

Patients with significant co-morbidities were excluded. These were considered to be systemic conditions which may influence the metabolome and included cardiovascular or respiratory diseases, neurological conditions, infections, cancer and atopy. Other exclusion criteria were: pregnant females, anyone on antibiotics or pro/pre-biotics, patients with stomas or those on a therapeutic diet (semi-elemental, elemental or parenteral nutrition) for IBD.

Patients and controls of the following self-reported ethnicities were included:

- White: British / Irish / any other white European background – this group is referred to as white, white European or Caucasian in this study.
- Asian: Indian / Pakistani / Bangladeshi / Sri Lankan

The following ethnicities were excluded:

- Black: African / Caribbean / any other
- Chinese: Chinese / any other
- Mixed: Asian and white / Black African or Caribbean and white / any other mixed

Metadata was collected for all patients and controls as previously reported in chapter 2 including vegetarian status.

3.4.2 NMR Spectroscopic Analysis

Midstream urine samples were taken, aliquoted and stored at −80 °C in siliconised microvials (Sigma-Aldrich, USA) until analysis. Before analysis, samples were thawed to room temperature and centrifuged at 13 000 rpm for 20 min at 4 °C.

Samples were randomised and analysed by 1D ^{1}HNMR spectroscopy by standardised technique as previously described in chapter 2. 6 quality controls were used which were composite of all samples, one sample run in each box. This was used for validation to detect drift during data acquisition and for statistical analysis as a representative spectrum.

3.4.3 Demographic and phenotypic data analysis

Phenotypic data was analysed using the Mann-Whitney U test and Fisher's exact test or the Chi-squared test (X2) as appropriate. For the phenotypic data a Holm-Bonferroni adjustment was applied to correct for multiple comparisons.

3.4.4. Data Acquisition and Pre-Processing

Spectral analysis was performed using Matlab in-house software (spec_preproc_v5). 4 samples were excluded due to inadequate water suppression. Resonances for TSP (δ -0.04 to 0.2), water (δ 4.63-4.923) and urea (δ 5.518 – 6.197) were removed. A further sample was excluded due to an abnormal base line reading. Spectral alignment (RSPA (222) and manual) and probabilistic quotient normalization was applied (227).

3.4.5 Hypothesis driven analysis of individual metabolites

Targeted profiling was performed using integral regions of 19 metabolites hypothesised to be different between groups. Metabolites were chosen on the basis that they had previously been

identified in other studies as differentiating between IBD subtypes and/or controls (114-116, 118, 132), as well as bacterial host-co-metabolites (table 3.1).

The median relative levels of metabolites were compared across groups and the Kruskal-Wallis test (GraphPad Prism v5.0 (GraphPad Software Inc., USA)) applied to test for differences. Benjamini-Hochberg correction with a false discovery rate of 0.1 was applied to account for multiple comparisons. The Mann-Whitney U test was used in to clarify differences between groups and p values corrected for multiple comparisons.

Analysis was performed to compare groups by diagnosis, and also according to ethnicity.

Metabolite	***Formula***	***Multiplicity***	***Chemical shift (ppm)***
Alanine	βCH_3; CH	**d**	**1.49**
		q	3.79
Acetate	$C_2H_3O_2$	**s**	**1.92**
Citrate	CH_2; CH_2	**d**	**2.55**
		d	2.67
Creatine	CH3; CH_2	**s**	**3.05**
		s	3.92
Creatinine	N-CH_3; N-CH_2	s	3.06
		s	**4.08**
4-Cresol sulphate	CH_3; CH; CH	**s**	**2.35**
		d	7.21
		d	7.28
Dimethylamine	N-$(CH_3)_2$	**s**	**2.73**
Formate	CH	**s**	**8.46**
Glycine	CH_2	**s**	**3.57**

Glycolate	CH_2	**s**	**3.95**
Guanidoacetate	CH_2	**s**	**3.80**
Hippurate	CH_2; CH; CH; CH	d	3.97
		t	7.55
		t	7.64
		d	**7.83**
Lactate	CH_3; CH	**d**	**1.33**
		q	4.14
Methanol	CH_3; -OH	**s**	**3.34**
3-Methylhistidine	CH; N- CH_3; CH; CH	m	3.24
		s	**3.74**
		s	7.02
		s	7.63
Succinate	$C_4H_6O_4$	**s**	**2.39**
Trans-aconitate	$C_6H_6O_6$	s	3.74
		s	**6.93**
Trigonelline	$C_7H_7NO_2$	s	4.43
		m	8.07
		m	8.82
		s	**9.11**
Trimethylamine-N-oxide	N-$(CH_3)_3$	**s**	**3.27**

Table 3.1: Table listing metabolites in urine chosen for targeted analysis, and chemical formula. For each metabolite the chemical shift of their ^{1}H NMR peaks, along with multiplicity, are shown. **Bold and greyed box** indicates the shift selected for peak integration in this experiment.

s=singlet, d=doublet, t=triplet, m=multiplet, dd=doublet of doublets

3.4.6 Multivariate analysis of groups

A PCA was constructed to visualise outliers, and distance from the model (DModX) was applied to exclude those above the critical value (Dcrit).

To assess the differences in metabolic data OPLS-DA were performed using Matlab (R21014a, The MathsWorks inc) and Simca (version 14.1, Umetrics, Umeå, Sweden). Pareto scaling (mean-centred data divided by the square root of the standard deviation – used to visualise small or medium variations in data (228)) or log transformation was applied and 7-fold cross validation was undertaken in each model. R^2 and Q^2 values are described, and for positive Q^2 values permutation testing or CV-ANOVA testing was performed to calculate a p value (considered significant if value <0.05).

Metabolites identified as contributing to the model were identified from the loadings plot using a combination of resources including Human Metabolome Database and Chenomx Profiler (Chenomx NMR Suite 8.1). Statistical evidence from STOCSY Matlab (R2014b, The MathsWorks, MA, USA) was used to confirm relationships between peaks correlating to metabolites.

The maximum r value (correlation coefficient) for each metabolite was measured and the significance calculated using http://vassarstats.net/rsig.html where a non-directional p value of <0.05 was considered significant.

3.5 Results

3.5.1 Phenotypic dataset

Table 3.2 shows the phenotypic data of the patients and controls included in this study.

	White European			South Asian		
	Healthy controls	CD	UC	Healthy controls	CD	UC
n	98	160	123	39	42	80
Age (years) Median (Range)	31 (18-67)	36 (19-79)	42 (17-80)	40 (21-70)	43 (18-76)	42 (18-77)
Male	41 (42%)	76 (48%)	59 (48%)	19 (49%)	14 (33%)	59 (73%)
Smokers	3 (3%)	5 (3%)	5 (4%)	0	4 (10%)	6 (8%)
Vegetarians*	3 (3%)	2 (1%)	5 (4%)	19 (50%)	10 (25%)	19 (23%)
Active disease[a*]		28 (17%)	17 (14%)		4 (10%)	13 (16%)
HBI (median)		2			2	
SCCAI (median)			1			2
IBD surgery		56 (35%)	2 (2%)		20 (48%)	0

Age at diagnosis: Median (range)		24.5 (12-77)	29 (11-73)		31.5 (12-57)	27 (10-77)
Medication use in preceding 4 weeks[b]		M1 : 20	M1 : 54		M1 :12	M1 :58
		M2 : 78	M2 : 40		M2 :18	M2 : 33
		M3 : 23	M3 : 5		M3 : 5	M3 : 2
Disease location [c]		L1 : 24	E1 : 29		L1 : 7	E1: 6
		L2 : 50	E2 : 32		L2 : 16	E2 : 26
		L3 : 68	E3 : 45		L3 : 19	E3: 48
Disease behaviour [d]		B1 : 114			B1 : 26	
		B2 : 12			B2 : 7	
		B3 : 12			B3 : 4	

Table 3.2: Characteristics of the study population

[a]As defined by disease activity index ≥5 [b] Medications: M1: 5-ASA; M2: immunosuppressants (Azathioprine/6-MP); M3:anti-TNF [c] Location (206): L1:ileal disease; L2: colonic disease; L3:ileocolonic disease; E1:proctitis; E2:disease limit distal to splenic flexure; E3:disease proximal to the splenic flexure [d] Behaviour (206): B1:inflammatory; B2:structuring; B3:fistulating * small amounts patient data missing.

The cohorts were compared for demographic differences using the Chi-squared test for categorical variables and Mann-Whitney *U* test for continuous variables (Table 3.3). The phenotypic data of the IBD patients were compared in Table 3.4. P values of <0.05 were considered statistically significant.

	White European	South Asian	P value
Male:female [a]			
Controls	41:57	19:20	0.94
CD	76:84	14:28	0.10
UC	59:64	59:21	**<0.01**
Age (median) [b]			
Controls	31	40	**<0.01**
CD	36	43	0.09
UC	42	42	0.5
Smokers [a]			
Controls	3 (3%)	0	0.27
CD	5 (3%)	4 (10%)	0.08
UC	5 (4%)	6 (8%)	0.29

Table 3.3: Comparison of the study participants

[a]Analysed by Chi-squared test [b]Analysed by Mann-Whitney *U* test

Bold values are P values that showed significance (P<0.05).

	White	South Asian	P value
Age at diagnosis (median) [b]			
CD	24.5 (12-77)	31.5 (12-57)	**0.00338**
UC	29 (11-73)	27 (10-77)	0.65
Disease location [a]			
CD	L1 : 24	L1 : 7	0.97
	L2 : 50	L2 : 16	0.73

	L3 : 68	L3 : 19	0.76
	p: 26	p:11	0.14
UC	E1 : 29	E1: 6	**0.0006**
	E2 : 32	E2 : 26	0.74
	E3 : 45	E3: 48	**0.017**
Disease activity [a]			
Active CD	28	4	0.17
Active UC	17	13	0.68
Disease behaviour [a]			
CD	B1 : 114	B1 : 26	0.10
	B2 : 12	B2 : 7	0.08
	B3 : 12	B3 : 4	0.69
Surgery [a]			
CD	56 (35%)	20 (48%)	0.13
UC	2 (2%)	0	0.25
Medications [a]			
CD	M1 : 20	M1 :12	**0.011**
	M2 : 78	M2 :18	0.49
	M3 : 23	M3 : 5	0.68
UC	M1 : 54	M1 :58	**0.000062**
	M2 : 40	M2 : 33	0.21
	M3 : 5	M3 : 2	0.55

Table 3.4: Comparison of IBD patient cohorts [a]Analysed by Chi-squared test [b]Analysed by Mann-Whitney *U* test. Significant P values (P<0.05) in bold

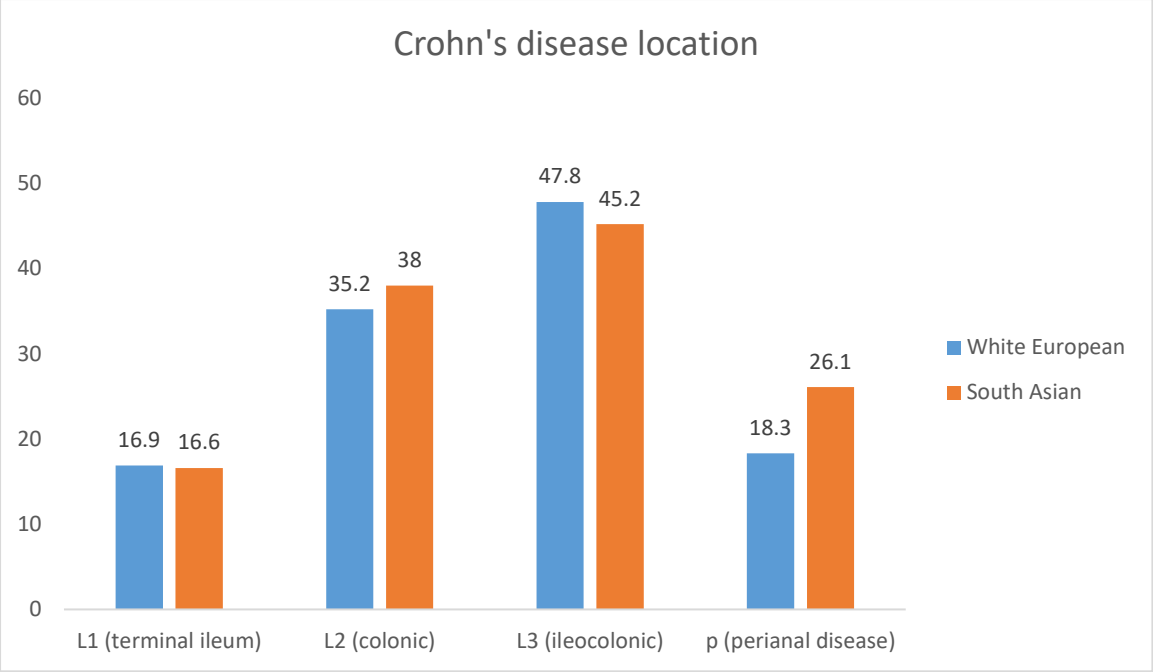


Figure 3.1: Graph comparing Crohn's disease location for white and South Asian patients.

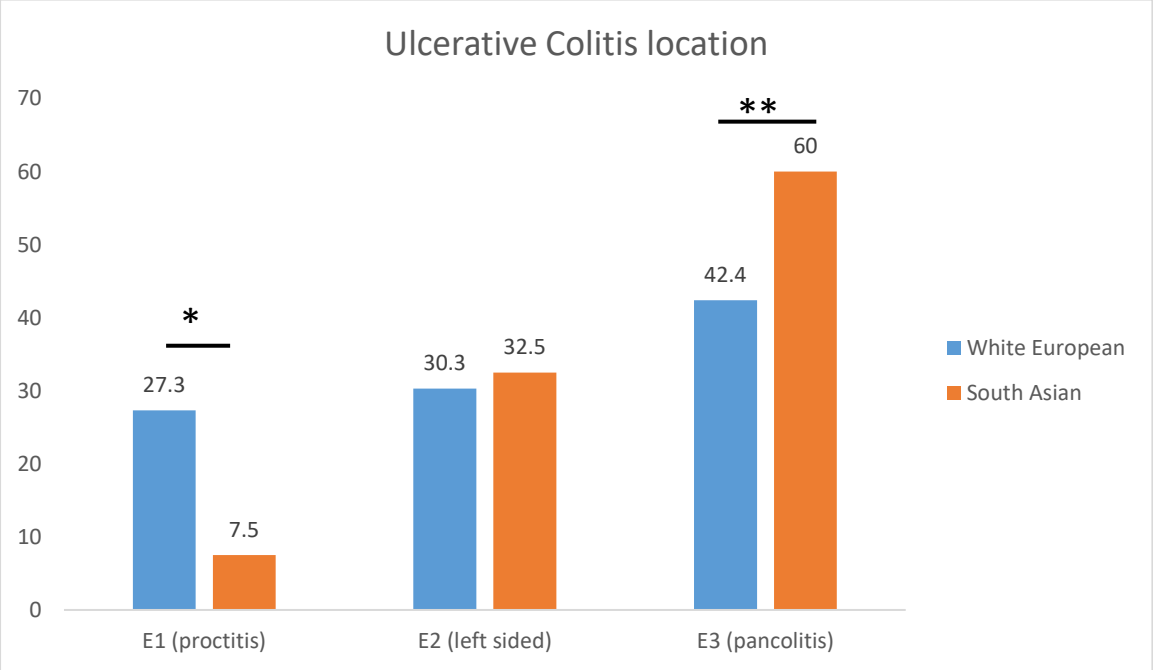


Figure 3.2: Graph comparing UC disease location for white European and South Asian patients. Significant p values are shown. *p = 0.0006 **p = 0.17

The South Asian UC cohort had significantly more male subjects and the South Asian control cohort was older (median age 40 years compared to 31 in Caucasian controls).

In terms of phenotype, the South Asian Crohn's cohort was significantly older at diagnosis (31.5 vs 24.5 years). A much smaller percentage of South Asians had proctitis compared to Caucasians in the UC group (figure 3.2). There were more South Asians on 5-ASAs than those in the Caucasian group. No other significant differences were seen.

3.5.2 NMR urinary analysis

Typical 1H NMR urinary spectra for each group are shown in figure 3.3.

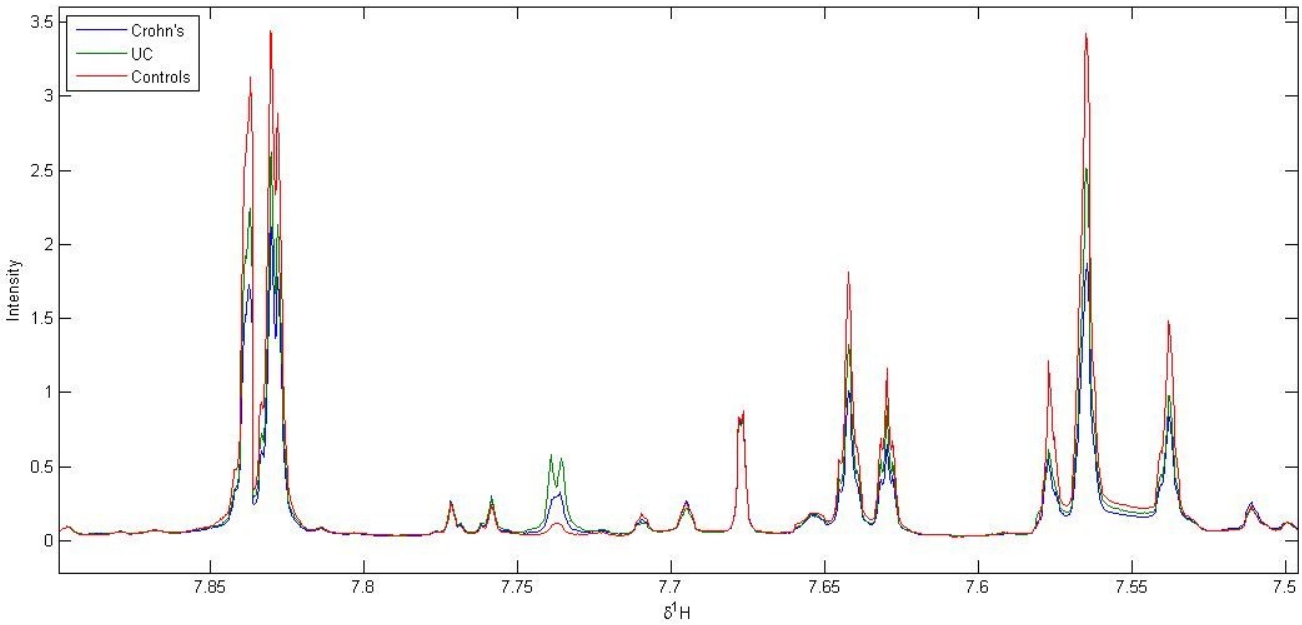


Figure 3.3: Median spectra from Crohn's (blue), UC (green) and healthy controls (red) showing average differences in metabolites in part of the aromatic region.

3.5.2.1 Hypothesis driven targeted metabolite analysis

Median relative values for the 19 compounds as measured in each group are shown in supplement 1.

3.5.2.1.1 Analysis of metabolites by disease

Table 3.5 shows p values for statistically significant metabolites differing between healthy controls, Crohn's and UC, in the three comparison groups; samples of all ethnicities, Europeans and South Asians.

Samples of all ethnicities				
Statistically significant metabolites with p values between Controls, UC and Crohn's				
	p value across all three groups[a]	p value Controls vs CD [b]	p value Controls vs UC [b]	p value UC vs CD [b]
Hippurate	**<0.0001***	**<0.0001***	**0.0003***	0.1640
Glycine	**0.0105***	**0.0156***	0.4864	**0.0069***
Formate	**0.0141***	0.5671	**0.0056***	0.0243*
Methanol	0.0334	**0.0074***	0.1430	0.4557

Table 3.5: p values for statistically significant metabolites differing between healthy controls, Crohn's and UC

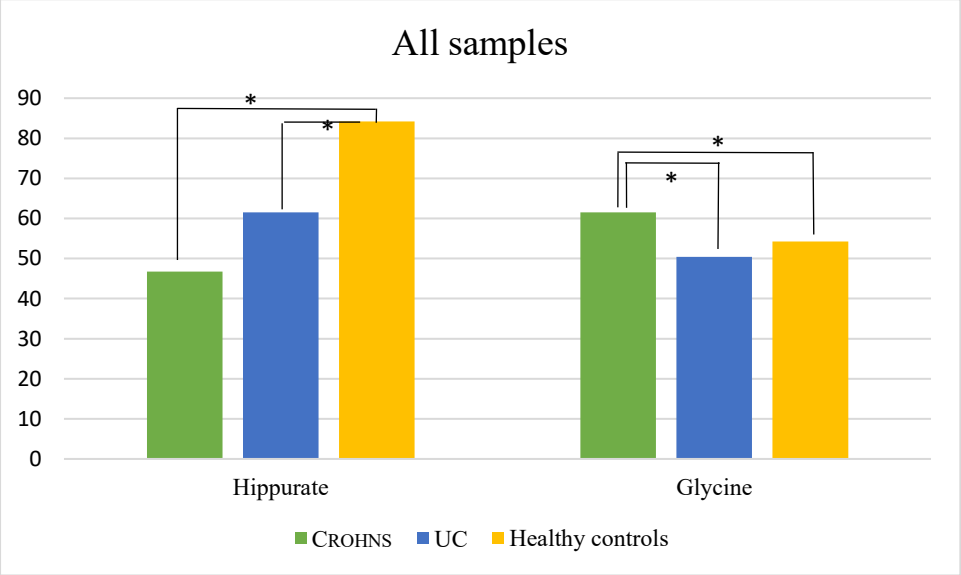
[a] as measured by Kruskal-Wallis test. [b] as measured by Mann-Whitney *U* test

*bold values remain significant post multiple comparison testing with Benjamini-Hochberg correction

When samples of all ethnicities were analysed together (table 3.5) hippurate (measured as a relative integrated area) was the most significantly different metabolite differentiating across

all groups ($p<0.0001$). After applying correction for multiple testing, glycine and formate remained statistically significant between the groups, but methanol did not.

Further analysis compared metabolites from two groups to each other with Mann-Whitney *U* test. All p values and corrected p values (q values) are shown in supplement 2.



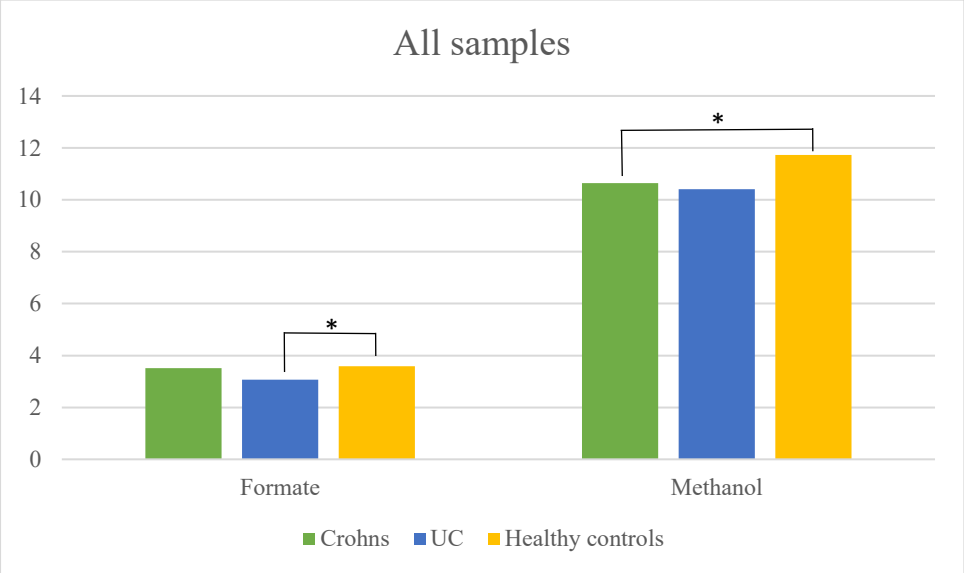


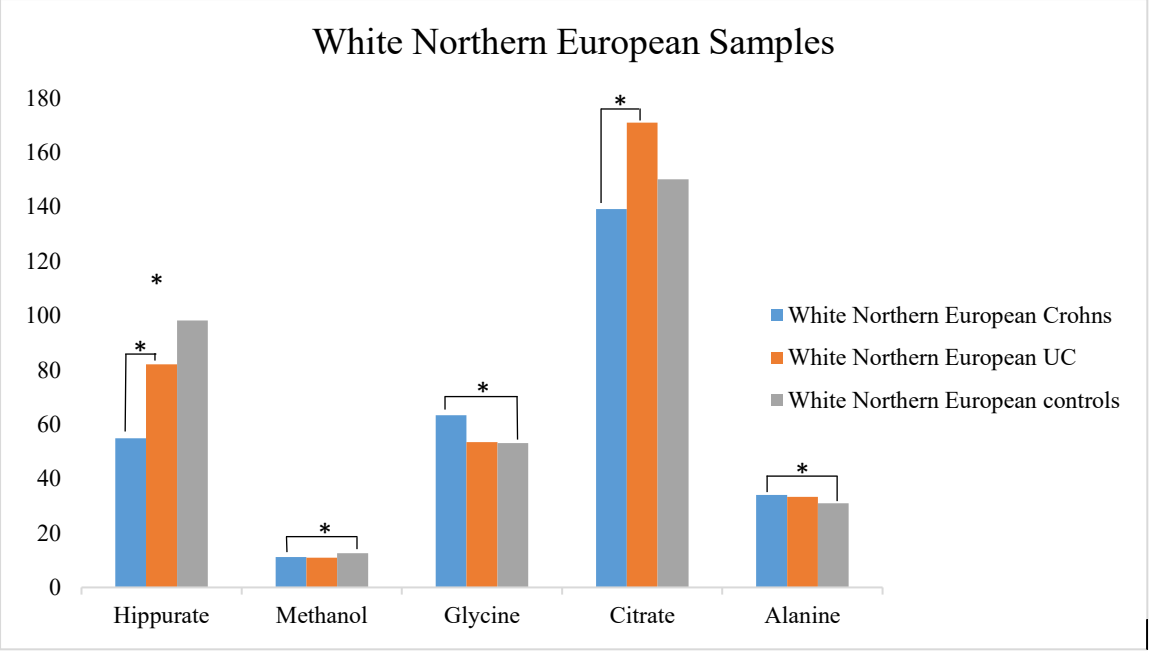
Figure 3.4: Relative integrated area of significant metabolites in each group with statistically significant p values between groups shown by *

3.5.2.1.2 Analysis of ethnic specific cohorts

In Europeans (table 3.6) hippurate was again strongly significant ($p<0.0001$), with p values for methanol and glycine remaining significant after multiple correction testing. In South Asians hippurate was also the most dominant discriminatory metabolite, but the other significantly different metabolites – TMNO, succinate and formate differed to findings in Caucasians.

Statistically significant metabolites with p values between Controls, UC and Crohn's				
White Europeans				
	p value across all three groups[a]	p value Controls vs CD [b]	p value Controls vs UC [b]	p value UC vs CD [b]
Hippurate	**<0.0001***	**<0.001***	0.1317	**0.0024***
Methanol	**0.0134***	**0.0031***	0.1027	0.3578
Glycine	**0.0154***	**0.0056***	0.5198	0.0536
Citrate	0. 0353*	0.5140	0.0572	**0.0114***
Alanine	0.0265*	**0.0100***	0.0517	0.7575
South Asian				
Hippurate	**0.0022***	**0.0004***	0.0508	**0.0459***
Trimethylamine-N-oxide	**0.0046***	**0.0007***	0.1127	**0.0368***
Succinate	**0.0055***	**0.0013***	0.0701	0.0644
Formate	**0.0068***	0.0722	**0.0019***	0.6604

Table 3.6: p values for statistically significant metabolites differing between healthy controls, Crohn's and UC in each separate ethnic group. [a] as measured by Kruskal-Wallis test. [b] as measured by Mann-Whitney *U* test. *bold values remain significant post multiple comparison testing with Benjamini-Hochberg correction



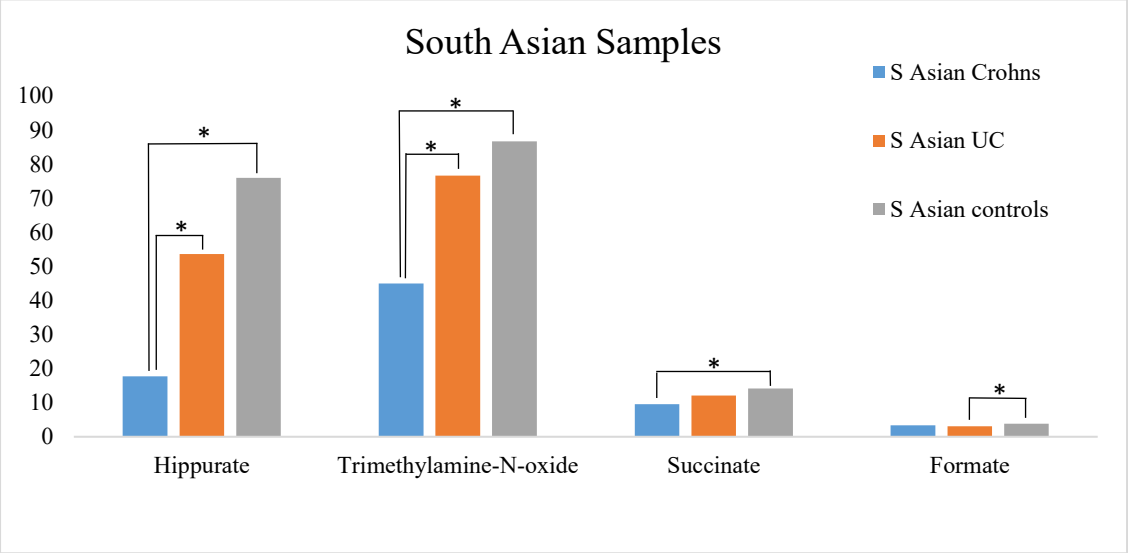


Figure 3.5: Relative integrated area of significant metabolites in each group, divided into ethnicity, with statistically significant p values between groups shown by *

3.5.2.1.3 Marker metabolites between ethnicities

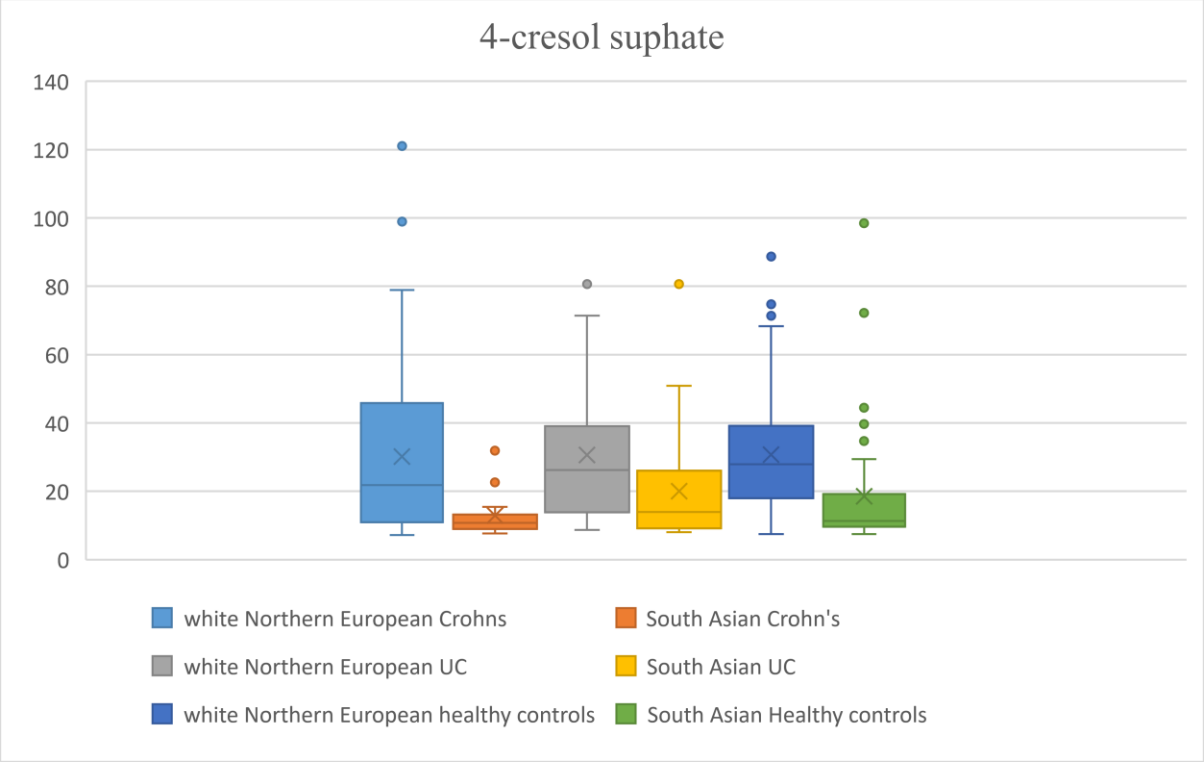
Table 3.7 shows p values for comparisons of South Asians and Europeans compared in groups of controls, Crohn's and UC.

Statistically significant metabolites with p values between white Europeans and South Asians			
	Healthy controls European : South Asian	Crohn's European : South Asian	UC European : South Asian
4-cresol sulphate	**<0.0001***	**0.0019***	**0.0003***
Hippurate	**0.0032***	**0.0016***	**0.0011***
Trimethylamine-N-oxide	0.1167	**0.0001***	**0.0336***
Succinate	0.6939	**0.0099***	**0.0114***
Alanine	**<0.0001***	0.9965	**0.0058***
Methanol	**0.0016***	0.4401	0.6910
Citrate	0.0397*	0.9133	0.1693
Lactate	0.0216*	0.5274	0.1162
Formate	0.0229*	0.4046	0.6253

Table 3.7: p values for statistically significant metabolites differing between cohorts.

[a] as measured by Kruskal-Wallis test. [b] as measured by Mann-Whitney *U* test ***bold** values remain significant post multiple comparison testing with Benjamini-Hochberg correction

When comparing Caucasians to South Asians, groups of Crohn's, UC and the healthy controls all had significantly higher 4-cresol sulphate and hippurate in the urine.



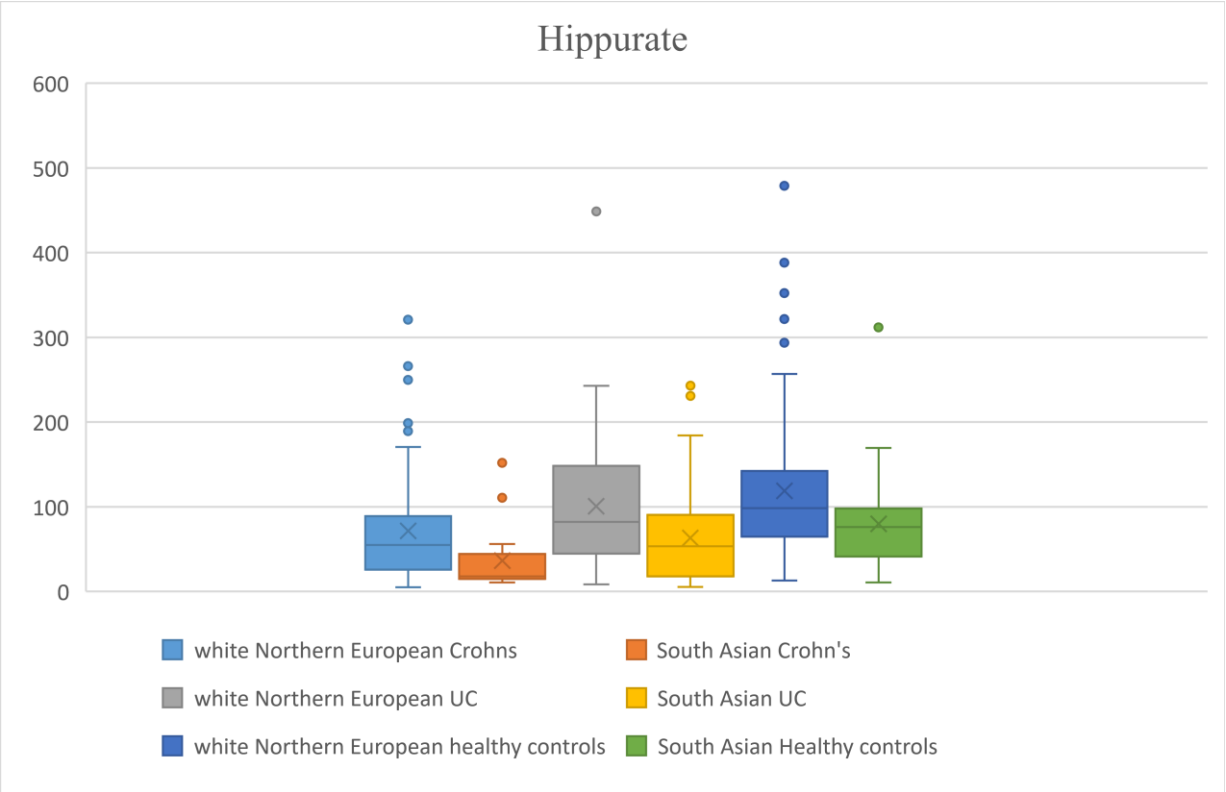


Figure 3.6: Box and whisker plots for 4-cresol sulphate and hippurate, measured for each group, with statistically significant p values shown between comparative groups of ethnicity.

Metabolites which differed between ethnic groups of IBD patients were TMAO and succinate (in both Crohn's and UC) and alanine (in UC only). In healthy participants alanine (higher in South Asians) and methanol (lower in South Asians) were significantly different. Citrate, lactate and formate were altered but p values were not significant after multiple correction testing.

3.5.2.2 Multivariate analysis

3.5.2.2.1 PCA

From preliminary analysis of the PCA, drug peaks (xenometabolites) were seen driving separation between IBD patients and healthy controls – these were identified as 5-ASA metabolites and paracetamol from the loadings plots using Statistical Total Correlation Spectroscopy (STOCSY) (figure 3.7).

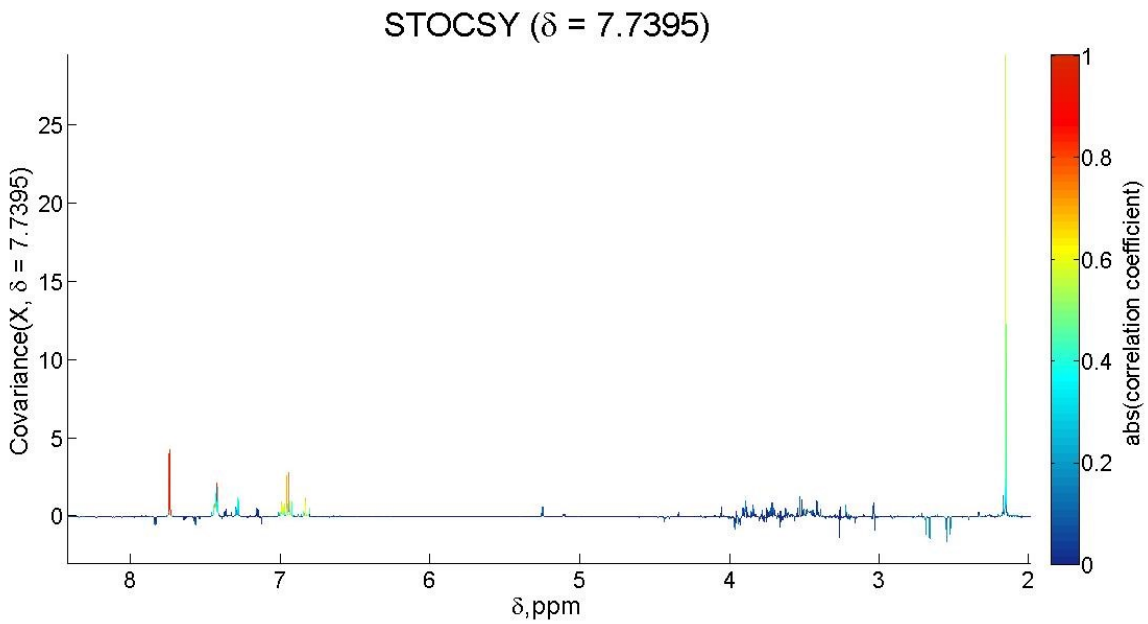


Figure 3.7: STOCSY plot with driver peak at δ 7.73. Peaks highlighted in increasing shades of red display a high degree of correlation (>0.7) with the driver peak, whereas those in decreasing shades of blue are weakly correlated and unrelated.

As discussed in chapter 2 there are a variety of ways to handle this potential confounder. Removing regions with drug peaks or using an statistical application to remove the drugs peaks (STOCSY-E) are reasonable methods when the size of the groups are smaller, as no samples need to be excluded. But as xenometabolites not only create specific drug peaks, they may also exert an influence on the metabolic constituents themselves, the most rigorous technique is to remove all samples identified with such xenometabolite peaks. As the groups in this analysis were large enough to remove participants, still leaving large numbers in each cohort, this was the chosen methodology here.

205 samples containing peaks at resonances identified as 5-ASA drugs and/or paracetamol (117, 226) (table 3.8) (144 patients had recorded 5-ASAs in the drug history) were removed, leaving 358 samples to include for analysis.

Metabolite	***Multiplicity***	***Chemical shift (ppm) δ***
N-acetyl-5-aminosalicylate	**s**	2.17
	d	6.95
	d	7.40
	d	7.70
Paracetamol glucuronide	**s**	2.17
	d	7.13
	d	7.36
Paracetamol sulphate	**s**	2.18
	d	7.31
	d	7.70

Table 3.8: Table listing chemical shifts for the peaks for 5-ASA and paracetamol. s=singlet, d=doublet

Other IBD drugs including steroids, anti-TNFs and thiopurines have not been reliably detected and quantified in human urine according to the Human Metabolome Database and therefore patients on these medications were not excluded. As one study (114) suggested that anti-TNFs altered the metabolic profile, analysis was later performed to look at whether patients on anti-TNFs could be distinguished with OPLS-DA. 7 outliers were excluded from the PCA as gross outliers and the PCA shown in figure 3.8.

The phenotypic data for the new dataset is shown below (table 3.9).

	White European			South Asian		
	Healthy controls	CD	UC	Healthy controls	CD	UC
n	83	96	69	39	25	46
Age (years) Median (Range)	31 (18-67)	35 (19-72)	44 (21-78)	40 (21-70)	41 (18-71)	44 (18-69)
Male	39 (47%)	40 (42%)	36 (52%)	19 (49%)	14 (56%)	21 (46%)
Smokers	3 (4%)	2 (2%)	3 (4%)	0	3 (12%)	5 (10%)
Vegetarians*	3	2	3	16	3	8

	(3%)	(2%)	(5%)	(41%)	(17%)	(21%)
1st generation* (%)				23 (59%)	8 (40%)	18 (69%)
Active disease [a] * (%)		26%	20%		8%	6 %
HBI (median)		2			1	
SCCAI (median)			1			1
IBD surgery		40 (42%)	2 (3%)		13 (52%)	0
Medications use in preceding 4 weeks[b]		M1 : 6	M1 : 12		M1 :2	M1 : 19
		M2 : 50	M2 : 19		M2 :13	M2 : 16
		M3 : 7	M3 : 0		M3 : 5	M3 : 1
Disease location [c]		L1 : 22	E1 : 19		L1 : 2	E1: 7
		L2 : 40	E2 : 20		L2 : 9	E2 : 10
		L3 : 26	E3 : 24		L3 : 11	E3: 24
Disease behaviour [d]		B1 : 76			B1 : 13	
		B2 : 6			B2 : 4	
		B3 : 4			B3 : 6	

Table 3.9: Characteristics of the study population with 5-ASAs removed

[a]As defined by disease activity index ≥5 [b] Medications: M1: 5-ASA; M2: immunosuppressants (Azathioprine/6-MP); M3:anti-TNF [c] Location (206): L1:ileal disease; L2: colonic disease; L3:ileocolonic disease; E1:proctitis; E2:disease limit distal to splenic flexure; E3:disease proximal to the splenic flexure [d] Behaviour (206): B1:inflammatory; B2:structuring; B3:fistulating * small amounts patient data missing.

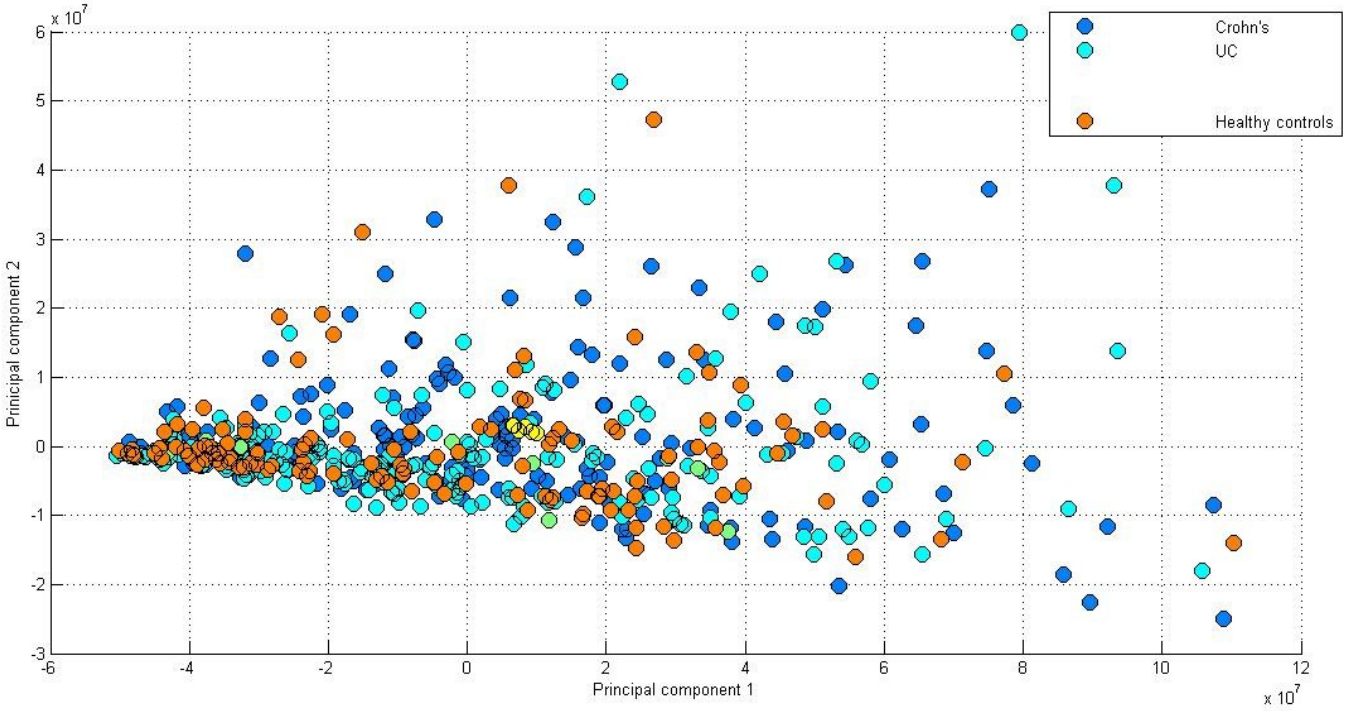


Figure 3.8: PCA scores plot of CD patients, UC and controls.

3.5.2.2.2 O-PLS-DA

O-PSL-DA models with one predictive component (to limit over-fitting) were created to compare cohorts according to disease types and comparisons made to healthy controls. In addition to 7-fold cross-validation, permutation testing was performed for each model with a positive Q^2 for statistical robustness. Firstly all ethnicities were included, then sub-groups of each ethnicity were compared. Major discriminatory metabolites were identified from the loadings and the significantly correlated metabolites from each model described.

3.5.2.2.2.1 Comparison of IBD patients by diagnosis

An example of a cross-validated O-PLS-DA model is shown in figure 3.9 showing distinct separation of samples based on diagnosis. Figure 3.10 shows the corresponding loadings plot for the model with significant metabolites documented in the table.

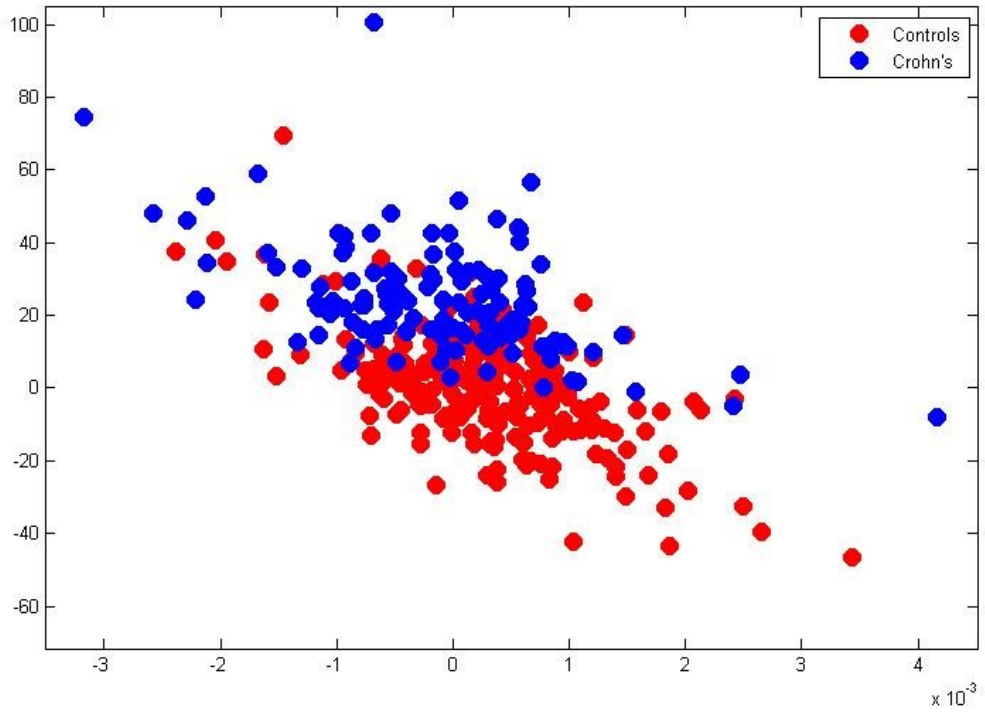


Figure 3.9: Cross-validated O-PLS-DA scores plot of healthy controls and Crohn's patients.

The R^2 and Q^2 of the O-PLS-DA models to compare groups by diagnosis are shown in table 3.10 with the list of differentiating metabolites identified from the loadings plot listed in order of importance.

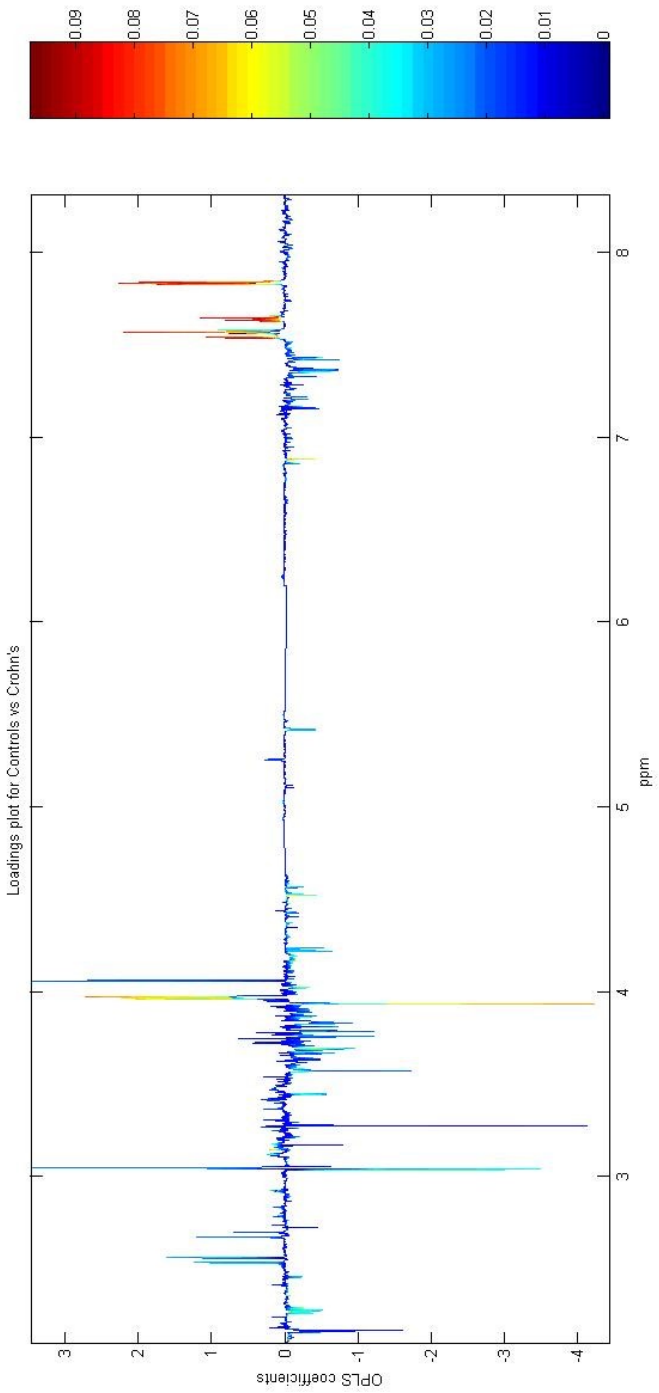


Figure 3.10: Loadings plot for OPLS-DA model of healthy controls and Crohn's patients.

All samples					
	n	R2	Q2	P value (1000 permutation testing)	Significantly correlated* metabolites *in order of importance*
H vs IBD	351	0.596	0.627	0.001	**Hippurate ↑:↓** Ascorbate ↓:↑ 1 methyl-histidine ↓:↑ **NMND ↑:↓** N-acetyl glycoproteins ↓:↑ **Dimethylglyceine ↑:↓** 4-HPA ↓:↑ Creatine ↓:↑ Guanidoacetate ↓:↑
H : CD	243	0.6593	0.6225	0.001	**Hippurate ↑:↓** NAG fraction ↓:↑ 4-HPA ↓:↑ **Cis-trans-aconitate (TCA) ↑:↓** Creatine ↓:↑ **4-DTA ↑:↓** **Lactic acid ↑:↓** Gamma-amino-N-butyrate ↓:↑ **Alanine ↑:↓** **Ascorbate ↑:↓** Glycine ↓:↑ Pantothenate ↓:↑ Phenyacetylglutamine (PAGn) ↓:↑ Guanidoacetate ↓:↑ Creatine ↓:↑ **Methanol ↑:↓**
H : UC	226	0.7983	0.7567	0.007	**Hippurate ↑:↓** Fatty acyl chain protons ↓:↑ **Furoylglycine ↑:↓** HMB ↓:↑ **NMND ↑:↓** Ascorbate ↓:↑ **Formate ↑:↓** **Lactate ↑:↓** Citrate ↓:↑
UC : CD	221	0.88	0.2818	0.001	N-Acetylglutamic acid ↓:↑ **Alpha-hydroxyisobutyrate (2-hydroxyisobutyrate) ↑:↓** Phenyacetylglutamine (PAGn) ↓:↑ Creatine ↓:↑ **4-DTA ↑:↓** Unidentified **BHMB ↑:↓** Trigonelline ↓:↑ Formate ↓:↑ **Lactate ↑:↓** **Citrate ↑:↓** Acetate ↓:↑ Glycine ↓:↑ **Hippurate ↑:↓**

Table 3.10: Summary of O-PLS-DA models constructed from all samples. R^2 (% of variance accounted for) and Q^2 (predictability of the model) are given, as well as the p value for the model after testing. Discriminatory metabolites are given in order of importance. *r value with $p<0.05$ **Bold ↑:↓** indicates positive direction of relative change in first group relative to second, e.g. in Healthy vs IBD hippurate is higher in healthy controls compared to IBD patients.

Models of all mixed ethnicity samples showed robust separation between healthy controls compared to all IBD patients, Crohn's patients and UC patients, as well as between UC and Crohn's disease. The models stool up to robust testing of 1000 permutations with significant p values in all cases.

By measuring the maximum r value for each metabolite identified from the loadings plot, hippurate was identified as dominating the models separating healthy controls from CD and UC. Other metabolites with statistically significant correlation values discriminating healthy controls from Crohn's were ascorbate, methyl-histidine, NMND, N-acetyl glycoproteins, and dimethylglyceine. For healthy controls versus UC patients the strongest metabolites were (after hippurate) fatty acyl chain protons, furoylglycine, beta-hydroxy-beta-methylbutyrate (HMB), 1-Methylnicotinamide (NMND) and ascorbate.

In the comparison of IBD subtypes N-Acetylglutamic acid, alpha-hydroxyisobutyrate (2-hydroxyisobutyrate), phenyacetylglutamine (PAGn), creatine and 4-DTA (4-deoxythreonic acid) were important.

3.5.2.2.2.2 Validation of results

Other similar studies have shown that after further sub-analysis of groups their results were no longer significant or models were less strongly predictive. Therefore validation was performed.

i. For validation, to ensure that the metabolic alterations were due to disease and not to the effect of medication (118) the comparisons were remodelled using only IBD patients and healthy controls taking no medications at all (both IBD or non-IBD medications).

Participants on no medications at all				
	n	R^2	Q^2	P value (100 permutation testing)
H vs IBD	186	0.4433	0.2322	0.01
H vs CD	136	0.7942	0.7411	0.01
H vs UC	124	0.8797	0.8475	0.01
UC vs CD	92	0.8819	0.5512	0.01

Table 3.11: Summary of O-PLS-DA models from participants on no medications. R^2, Q^2 and p value for each model are described.

These models showed stronger Q^2 values, (with the exception of the healthy vs Crohn's) with significant p values for participants on no medications, corroborating the above results (table 3.11).

ii. In another paper exclusion of post-surgical patients removed the ability to distinguish between UC and CD (114). Therefore as a further validation, the analysis was repeated with all patients (54) who had undergone surgery removed.

Surgical patients excluded				
	n	R^2	Q^2	P value (100 permutation testing)
H vs IBD	297	0.5829	0.2074	0.01
H vs CD	193	0.7443	0.6972	0.01
H vs UC	224	0.8012	0.7582	0.02
UC vs CD	167	0.9019	0.2481	0.04

Table 3.12: Summary of OPLS-DA models from surgical patients excluded. R^2, Q^2 and p value for each model are described.

These comparisons confirmed that the models all remained valid and robust (table 3.12).

iii. Furthermore it has been suggested that patients on anti-TNF treatments showed significant differences in their metabolite profiles and in one paper exclusion of such patients rendered discriminant analysis non-significant (114). Therefore the 22 patients on anti-TNFs were removed from the models for validation.

Participants on anti-TNF treatment excluded				
	n	R^2	Q^2	P value (100 permutation testing)
H vs IBD	329	0.6290	0.2033	0.01
H vs CD	224	0.6798	0.6365	0.01
H vs UC	221	0.7996	0.7548	0.01
UC vs CD	201	0.8888	0.2822	0.01

Table 3.13: Summary of OPLS-DA models repeated with patients on anti-TNF treatments excluded. R^2, Q^2 and p value for each model are described.

All models remained valid by permutation testing suggesting anti-TNFs do not significantly alter the models (table 3.13).

3.5.2.2.2.3 Comparison of IBD patients in ethnic cohorts

Groups were further analysed as separate cohorts according to ethnicity to examine if the same metabolic changes were seen in South Asians and white Europeans.

White Europeans					
	n	R^2	Q^2	P value (1000 permutation testing)	Significantly correlated* metabolites *in order of importance*
H : CD	177	0.6335	0.627	0.003	**Beta-aminoisobutyrate (3-aminoisobutyrate) ↑:↓** **Hippurate ↑:↓** Unassigned ↓:↑ Creatine ↓:↑ N-Acetylglutamate ↓:↑ 4-hydroxybutyrate ↓:↑ Glycine ↓:↑ Alanine ↓:↑ **Cholate ↑:↓** Guanidoacetate ↓:↑ **Creatinine ↑:↓** **Methanol ↑:↓** 4-cresol sulphate ↓:↑
H : UC	145	0.8149	0.7691	0.012	Fatty acyl chain protons ↓:↑ **Hippurate ↑:↓** Citric acid ↓:↑ Citrate ↓:↑ Cholate ↓:↑ Alanine ↓:↑ Ascorbate ↓:↑ Succinate ↓:↑ Glycolate ↓:↑
UC : CD	156	0.8817	0.4578	0.008	N-Acetylglutamate ↓:↑ **Lipids in LDL ↑:↓** **Hippurate ↑:↓** **Alpha-hydroxyisobutyrate (2-hydroxyisobutyrate) ↑:↓** **Citrate ↑:↓** Phenyacetylglutamine (PAGn) ↓:↑ Creatine ↓:↑ Acetate ↓:↑ **Methyl-histidine ↑:↓** **Isovalerylglycine ↑:↓**

Table 3.14: Summary of OPLS-DA models constructed from white European patients and controls. For each model the R^2 ,Q^2 and p value for the model are given. Major discriminatory metabolites for each model are described in order of importance.

*r value with $p<0.05$ **Bold ↑:↓** indicates positive direction change in first group relative to second

The Caucasian group showed similarly robust models as the mixed ethnicity group (table 3.14).

South Asians					
	n	R^2	Q^2	P value	Significantly correlated* metabolites *in order of importance*
H : CD	64	0.94	0.54	0.005	Butyrate ↓:↑ **1-Methylnicotinamide (NMND) ↑:↓** **Nicotinamide ribotide ↑:↓** **Hippurate ↑:↓** **Glucose ↑:↓** 4-cresol sulphate ↓:↑ Trigonelline ↓:↑ **Nicotinamide ↑:↓** Isoleucine ↓:↑
H : UC	85	0.18	0.69	0.16	
UC : CD	48	0.40	0.21	>0.50	;,

Table 3.15: Summary of O-PLS-DA models in South Asians. R^2 ,Q^2 and p value for the models are given. Major discriminatory metabolites for each model are described in order of importance. *r value with $p<0.05$ **Bold ↑:↓** indicates positive direction change in first group relative to second

In the South Asian patients, there was a strong model separating Crohn's from healthy controls, with butyrate, nicotinamide metabolites and hippurate most strongly influencing (table 3.15).

However, in the South Asian cohort O-PLS-DA models were unable to reproducibly identify South Asian UC from controls or the two subtypes of IBD. There were no significant models to interrogate for discriminatory metabolites.

Subgroups were analysed to examine whether differences were due to dietary status, with vegetarians being compared to omnivores in each ethnicity and combined (table 3.16). No valid models could separate these groups.

Comparison by diet				
	n	R^2	Q^2	p value (100 permutation testing)
Vegetarians vs omnivores (all ethnicities)	351	0.04	0.707	>0.5
Vegetarians vs omnivores (South Asian)	104	0.0001	0.1169	>0.5
Vegetarians vs omnivores (white European)	244	0.0211	0.0183	>0.5

Table 3.16: Summary of O-PLS-DA models comparing vegetarian and omnivore status for all participants, and in groups of specific ethnicity. R^2, Q^2 and p value for the models are given.

The number of patients in the subgroups were too unequal to allow comparison of IBD subtype according to vegetarian status (for example only 2 vegetarian Caucasian CD patients compared to 96 omnivores).

3.5.2.2.2.4 Comparing ethnicities

Multivariate analysis was also applied to investigate the differences between samples of difference ethnicities according to diagnosis (table 3.17).

	n	R^2	Q^2	p value*	Significantly correlated** metabolites *in order of importance*
Healthy controls Caucasian : South Asian	122	0.321	0.1555	**0.001**	**Alanine ↑:↓** **Triglycerides ↑:↓** Phenyacetylglutamine (PAGn) ↓:↑ **4 cresol sulphate ↑:↓** **Acetic acid ↑:↓** Hippurate ↓:↑ Methanol ↓:↑ **Lysine ↑:↓** Beta-aminoisobutyrate (3-aminoisobutyrate) ↓:↑ Lactate ↓:↑ **Citrate ↑:↓** Trans-aconitate ↓:↑ Tartrate ↓:↑ **Formate ↑:↓** **Glycolate ↑:↓**
Crohn's Caucasian : South Asian	111	0.1431	0.0284	**0.011**	**Glutamate ↑:↓** **Glucose ↑:↓** **4-amino Hippurate ↑:↓** 4-cresol sulphate (p-CS) ↓:↑ Succinate ↓:↑ **Methyl-histidine ↑:↓** **Acetate ↑:↓** Hippurate ↓:↑ Glycine ↓:↑
UC Caucasian : South Asian	94	0.4771	0.1762	0.33	

Table 3.17: Summary of OPLS-DA models comparing ethnicity by each diagnosis and in healthy controls. R^2, Q^2 and p value for the models are given. Major discriminatory metabolites for each model are described in order of importance.

*p value calculated with 1000 permutation testing **r value with $p<0.05$ **Bold ↑:↓** indicates positive direction change in first group relative to second

There were significant differences between healthy controls of white Europeans and South Asian ethnicity. Although the Q^2 value was not particularly high, the model was robust and identified 15 metabolites with significant correlation values within the model.

In healthy adults the strongest metabolites were alanine, triglycerides, phenyacetylglutamine (PAGn), 4 cresol sulphate, acetic acid and hippurate. All of these were higher in Caucasians except for PAG and hippurate.

In the Crohn's group there was again a strong model; glutamate, glucose, 4-amino hippurate and 4-cresol sulphate were the strongly correlated metabolites.

In the UC model although there was a positive Q^2 value it did not pass permutation testing and therefore cannot be analysed further.

3.5.2.2.2.5 Comparison of first and second generation South Asians

It was not possible to separate by O-PLS-DA according to 1st vs 2nd generation in either controls or IBD patients. However these groups were small as data was missing from a proportion (group sizes = controls: 39, CD: 20, UC: 26).

3.5.2.2.2.6 Comparison by disease location

To investigate if metabolic differences could separate isolated colonic Crohn's disease (L2) from UC, comparisons were made between the two, in the mixed cohort and individual ethnicity groups (table 3.18).

Comparison by disease location				
		n	R^2	Q^2
All patients	Colonic CD: UC	154	0.3289	-0.1523
White European	Colonic CD: UC	97	0.4872	-0.2418
South Asian	Colonic CD: UC	56	0.5007	-0.1362

Table 3.18: Summary of OPLS-DA models comparing colonic Crohn's disease to UC, in groups of mixed ethnicity and for white Europeans and South Asians separately. R^2 and Q^2 values for the models are given.

There were no valid models in any ethnicity group suggesting colonic Crohn's cannot be separated from UC by urinary metabonomics.

3.5.2.2.2.7 Comparison by surgical status

To examine the effect of surgery, groups were compared of those who had undergone intestinal resection versus those who had never had surgery (table 3.19).

Surgical analysis				
	n (patients with previous surgery/total)	R2	**Q2**	p value *
All IBD patients	55/222	0.2351	-0.0088	
Crohn's patients	51/119	0.4277	**0.211**	0.02
UC patients	2/102	0.0382	-0.135	

Table 3.19: Summary of OPLS-DA models separating patients with previous intestinal resections from all other IBD patients, and post-surgical patients with Crohn's and UC separately. R^2, Q^2 and p value for the models are given.

*p value with 100 permutation testing

A difference was shown in the Crohn's patient group, suggesting that surgery has a demonstrable effect on the metabolic profile in Crohn's.

3.5.2.2.2.8 Comparison by gender

To compare to the published literature, which has consistently demonstrated differences between males and females with NMR, samples were compared by gender (table 3.20).

Gender analysis			
	R^2	Q^2	p value* (100 permutations)
All	0.526	0.5225	**0.01**
White Europeans	0.501	0.4949	**0.01**
South Asians	0.595	0.5846	0.18
Healthy controls	0.545	0.5316	**0.01**
IBD patients	0.517	0.5132	**0.01**

Table 3.20: Summary of O-PLS-DA models testing separation of males and females in the whole cohort, each ethnicity, healthy adults and IBD patients. R^2, Q^2 and p value for the models are given. *p value with 100 permutation testing

Data were examined for gender differences and all groups showed similar Q^2 values with a significant p value except the South Asian cohort. Loadings plots showed the discriminant metabolites were a higher creatinine in male subjects (this dominated the Caucasian, healthy and IBD models) and higher creatine (most significant in the South Asian group) and citrate in females of all groups.

3.5.2.2.2.9 Validation of storage methodology

To verify that there was no effect from freezer storage of samples over time, comparisons were made of samples collected and stored for less than 1 year compared to more than one year (table 3.21).

Storage time analysis			
	n	R^2	Q^2
All samples acquired	542	0.2910	-0.0103
Samples used for OPLSDA analysis	351	0.2592	-0.0375

Table 3.21: Summary of O-PLS-DA models testing samples stored for <1 year compared to >1 year. R^2 and Q^2 values for the models are given.

Q^2 values for samples stored longer compared to recently were negative, showing no metabolic differences between those sample sets.

3.6 Discussion

Although several studies have used urinary metabolic profiling to investigate IBD, this is a significantly larger sized study than publications in this area to date, and the first time it has been done specifically in groups of different ethnicities. Both hypothesis driven targeted analysis and multivariate statistical modelling were used as methods to demonstrate metabolic differences between diseased and non-diseased, and in two different ethnic cohorts. Using two statistical approaches to the data analysis provides complementary results, and also enables this dataset to be compared to previously published findings.

Summary and discussion of results by each method

3.6.1. Hypothesis driven targeted metabolite analysis

Specific metabolite analysis is a targeted approach to investigate hypothesis derived differences between groups (242). Whilst metabonomics is an enormously powerful and information rich technique, one of the main analytical challenges is extracting accurate quantitative information from spectra whilst accounting for effects of the chemical environment such as small pH or temperature changes during acquisition (242). By pre-identifying specific metabolites thought to relate to underlying pathological mechanisms, and integrating their peak resonances, a quantifiable value compares groups directly. With this approach the choice of post-processing technique (such as scaling) has less effect on results, there are less potential effects of xenometabolites and it avoids the complex interpretation required in multivariate mathematical modelling. It also addresses the potential criticism of 'top down' metabolic profiling and allows comparison with previous peer-reviewed data.

The chosen metabolites were based on previously published results from other studies profiling IBD (114-116, 118, 132) hypothesising that the metabolic profile in IBD reflects the microbiome, energy cycle disruption and colonocyte metabolic pathways.

In this study the pattern of alteration in chosen metabolites between groups were largely consistent with previously reported results (hippurate, formate, methanol, glycine, alanine and succinate) (114-116, 118, 132) although reduced TMAO in South Asian IBD appears to be a new finding.

3.6.2 Multivariate pattern-recognition analysis

Multivariate statistical methods are proven analytical techniques developed to handle the large number of metabolites and identify discriminating features between cohorts (243). This approach enables low concentration metabolites to be identified and can highlight subtle patterns between biological cohorts relating to the disease processes.

When applied to the whole dataset of over 350 individuals of mixed ethnicities, O-PLS-DA demonstrated findings in keeping with other studies (114-116, 118), clearly separating IBD from healthy controls using larger cohorts than any previously published data in this area. Each model was rigorously tested with permutation testing – a step not taken in many published metabonomic analyses. Furthermore, subgroups were removed in turn that had in previous papers been suggested as distorting or swaying the models (114). Removal of all patients on medications, post-surgical patients and those on anti-TNF treatments did not affect the ability to discriminate between IBD and controls or between the subtypes of IBD. The Q^2 values of each model were high (HC:CD 0.623 and HC:UC 0.757) and comparable, or better, than published papers (114, 116, 118).

Multivariate models were also able to discriminate UC from Crohn's, which has only been shown in the Williams paper (118) but not replicated in other studies (114-116). This may be because the Williams cohort was larger in size than the other studies. In this thesis the Q^2 value of the CD:UC comparison was the lowest of all the models (0.282), but stood up to vigorous permutation testing, and the Q^2 improved with removal of all patients on medications (IBD and non-IBD related). This may suggest that it is certainly possible to discriminate IBD subtypes with metabonomics, but the discrimination is more subtle and potentially affected by other confounders.

This study, using participants of mixed ethnicity tested the 'all-comers' approach, that IBD has a significant effect on the metabolome, detectable in a group of mixed ethnicity.

However, when separating out the cohorts, it was seen that in the South Asian cohort only Crohn's disease was able to be identified from controls by multivariate analysis. OPLS-DA models separating South Asian UC patients from controls and the IBD subtypes from one another failed permutation testing; suggesting that Crohn's has a stronger effect on the urine metabolome than UC, which correlates with other studies (114, 118). However, in the earlier univariate analysis, several key metabolites were significantly different between the Asian UC group and controls.

These new data in a different ethnic group highlight that the effect of ethnicity on the metabolic profile, which was not considered in other previous papers, is important in IBD and must be accounted for.

3.6.3 Significant metabolites altered in IBD

Loadings plots identified discriminating metabolites associated with IBD, many metabolites had been described previously by other authors, some were described for the first time in IBD.

3.6.3.1 Hippurate: the most significant metabolite?

Hippurate dominated the separation between groups regardless of ethnicity, underlining its significance in the IBD metabolome. Reduced urinary hippurate has been the most consistently demonstrated finding in IBD metabolic profiling, with consensus of results across all papers (114-116, 118). Levels of hippurate are lowest in Crohn's compared to UC compared to controls.

Hippurate is a mammalian-microbial co-metabolite formed by the glycine conjugation of benzoic acid in mitochondria (244). The ubiquitous nature of hippurate, being associated with a wide range of disease states and external influences, means that its relationship with IBD must be carefully interpreted.

It has been shown to vary in multiple pathological and physiological states, including diabetes (245), hypertension (246) and obesity (247), amongst others making determining the origin of its perturbation complex. Additionally, all of the three systemic conditions above are also associated with gut dysbiosis (244). Although diabetes and hypertension were exclusion criteria in this study, BMI was not measured (incomplete data collection), and obesity cannot be accounted for.

The current paradigm in IBD suggests that low hippurate is linked to gut dysbiosis (115). One postulated mechanism for this is specific bacterial species change; urinary hippurate levels have been correlated with the abundance of *Clostridia* in the gut microbiome (134). Decreased

Clostridia species in Crohn's disease has been shown in several studies (68), and also in UC (86). These bacteria have essential roles modulating the gut immune response through maintenance of epithelial barrier integrity (91).

Food is also a significant alternate source of hippurate, including black tea (248), fruit and vegetables (249). In the dietary histories taken from patients in this study, food rich in polyphenolic compounds (benzoate producers) (133) were specifically documented to account for this potential source of hippurate. No group had a significantly higher intake of black, green or herbal teas, carbonated drinks, berries or yoghurt, excluding this as a cause of altered hippurate in the urine.

To further endorse the link between the disturbance of hippurate and gut dysbiosis in IBD Williams et al (133) demonstrated normal benzoate conjugation in Crohn's patients, and by accounting for accurate dietary intake, were able to exclude other reasons for lowered hippurate levels in IBD patients. They concluded from this that disturbances in hippurate seen in IBD were likely to be the result of alterations in microbial composition and its interaction with human metabolic pathways.

3.6.3.2 Other bacterially derived metabolites

There is strong evidence that the microbiome is one of the main influences on the metabolome in IBD. In this study, as well as hippurate, several other bacterial metabolites or bacterial-host co-metabolites were found to be important candidates in distinguishing IBD from controls including methanol, formate and 4-cresol sulphate.

Methanol was reduced in both UC and Crohn's on univariate analysis (but only statistically significantly in Crohn's) and also was a significant discriminator in the multivariate separation of Crohn's from controls. This confirms the same finding in three other studies (114-116).

Interestingly, formate was reduced in South Asian IBD patients – this result was also seen in studies by Dawiskiba (116), Stephens (114) and Schicho(115) (ethnicity of patients unknown). However in our cohort of white North Europeans formate was higher in IBD, and this was the same pattern shown by Williams et al (118) in their Caucasian study where formate was higher in Crohn's compared to UC and controls. Formate is also a gut microbial-host produced co-metabolite, produced by bacteria including enterobacteria (136) which has been shown to be altered in IBD (68). This may suggest that IBD related metabolic changes are related to alterations in microbial communities which are specific to ethnic group.

4-cresol sulphate was identified as a driver in the models separating Crohn's from controls in both ethnic groups. 4-cresol sulphate (also known as p-cresol sulphate) is produced by bacterial metabolism of the amino acid tyrosine, specifically by *Clostridia* and *Bacteroides*. Interestingly these are the same bacteria (*Clostridia* and *Bacteroides*) shown to be reduced in IBD (68, 86).

Phenyacetylglutamine (PAG), a gut-microbial co-metabolite (62)was also significant in the models differentiating IBD from controls further reinforcing the evidence for dysbiosis in the disease.

Trimethylamine-N-oxide (TMAO), another bacterial-host co-metabolite, was reduced in South Asian Crohn's patients relative to UC and healthy people. This appears to be a new finding, as no difference in TMAO was shown in either the Williams (118) or Stephens (114) papers (and similarly in this dataset no difference in TMAO was seen in Caucasians in this study). However murine models have previously demonstrated alterations in urinary TMAO with the initiation of colitis in IL-10 animals (123).

3.6.2.3 Energy cycle metabolites

Citrate, succinate and cis-trans-aconitate, all TCA cycle intermediaries, helped discriminate IBD from controls. Consistent findings were shown in the published studies from all other groups (114-116, 118) pointing to increased energy requirements and more rapid turnover of metabolic energy pathways. Creatine, a cellular fuel metabolite, was also elevated in the Schicho study (115) as well as mouse colitis models (124).

In South Asians, succinate, an energy metabolite formed as part of the Kreb's cycle (tricarboxylic acid (TCA) cycle) was significantly reduced in Crohn's compared to controls (with a trend towards significance in UC, p=0.070) in accordance with other published studies (114-116). This reflects disruptions in energy metabolic pathways and/or increased energy demand in the inflammatory state (116). Succinate and other TCA intermediates were also significantly altered in a mouse colitis model (250), suggesting that the development of colitis increases energy requirements or reduces absorption of nutrients for synthetic energy pathways due to gut damage.

N-Acetylglutamic acid, another part of the TCA energy cycle was also lowered in UC compared to Crohn's.

In the South Asian group there were several nicotinamide metabolites identified separating Crohn's from controls. These form part of the tryptophan / nicotinamide adenine dinucleotide (NAD) energy pathway and are linked to high energy turnover states (251). Urinary nicotinamide metabolite disruptions have been reported in stressed animals as a mimic for IBD (251), as well as autism (252), Parkinson's (253) and major depression (254)

3.6.2.4 Short chain fatty acids

Short chain fatty acids are produced by the saccharolytic fermentation of oligosaccharides, but are more strongly correlated with faecal metabolic profiling, as discussed in chapter 5.

In this urine study, butyrate, the main energy substrate for colonocytes, was a key marker in differentiating Crohn's patients from controls in South Asians. It has been used in published studies as a surrogate marker for gut permeability (255).

Higher levels of Low density lipoproteins (LDLs) and isovalerylglycine (a metabolite of fatty acids) were identified in UC patients compared to Crohn's. This suggest a disturbance in lipid metabolism, which was also shown in a DSS-induced acute colitis in animal models (256).

Acetate also differentiated UC from Crohn's in the white European population and mixed group.

3.6.2.5 Amino Acids

Alanine contributed significantly to the models differentiating controls from Crohn's and UC in multivariate analysis in Northern Europeans and was also significant in the univariate results in the same cohort. The strongest effect seen was in Crohn's.

The amino acid glycine was significantly higher in Caucasians with Crohn's compared to their controls, as seen in the Williams' study (118).

Dawiskiba (116) also found higher levels of glycine and alanine associated with active IBD, thought to be related to their anti-inflammatory properties or impairment of macrophage and neutrophil activation (257). Amino acids and their derived metabolites have been also shown to be increased in serum and plasma(115), faecal samples (117) and colonocytes (258) in IBD.

In a novel finding, UC patients had higher hydroxy-beta-methylbutyrate, a leucine metabolite, is associated with muscle loss in critically ill patients (259) and suggesting a catabolic state. And methylhistidine (not previously described) was altered in IBD, most specifically Crohn's.

3.6.2.6 Organic acids

Organic acids, produced by nucleic acid metabolism, including beta-aminoisobutyrate and alpha-hydroxyisobutyrate were reduced in IBD. This finding was also seen in the study by Dawiskiba where changes were linked to altered energy pathways related to IBD inflammation (116). Beta-aminoisobutyrate has also been shown to be altered in the urine of cancer patients (260) related to higher cell turnover.

3.6.2.7 Novel metabolites

Most of the above metabolites are altered in a consistent pattern to published literature; however some metabolites identified in this study are described here for the first time in IBD.

2-hydroxyisobutyrate, raised in UC, has been linked to chronic oxidative stress, secondary to alterations to the rate of glutathione synthesis which then changes the rate of urinary excretion of 2-hydroxyisobutyrate (261). 4-Deoxythreonic acid (DTA), also higher in UC patients, is a carboxylic acid which so far has only been described as being associated with type 1 diabetes (262) and advanced pregnancy (263).

Other novel metabolites were guanidoacetate and pantothenate, the origin of these perturbations is not clear.

3.6.4 Examining the effect of ethnicity on the metabolic profile

There is minimal published data on urinary metabonomics in healthy populations of different ethnicities, and none in inflammatory bowel disease. One paper suggests urinary metabonomics changes related to race (and independent of diet and other variables) are linked to the gut microbiome being different across ethnicities (264). A further paper showed metabonomic variation across geographical regions of Europe in healthy adults – postulated to be linked to diet (265) but difficult to directly link. To the best of my knowledge this thesis has examined the first data comparing urinary metabolic profiles in healthy patients and IBD patients to understand the contribution of ethnicity on our metabolome.

Previous urinary metabonomics studies in IBD discussed in this thesis were carried out by Danish (132), Polish (116) and Canadian (114, 115) authors, none of which stated the ethnicity of participants in their papers. The study by Williams (118) was performed in an exclusively Caucasian UK cohort.

3.6.4.1 Metabolic profiling comparing different ethnicities

4-cresol sulphate and hippurate were found at significantly lower levels in South Asians compared to Caucasians regardless of disease or health. These metabolites that are consistently changed across both types of IBD and healthy controls are therefore likely to reflect the difference in ethnicity.

4-cresol sulphate is mostly found in animal proteins and thus urinary excretion of 4-cresol sulphate has been shown to be lower in vegetarians than omnivores (135). As there are more vegetarians in the South Asian group this may explain some of the difference between

ethnicities. It is not clear if this reduced 4-cresol sulphate is related to lower protein intake in vegetarians (thereby reducing available substrate), or increased dietary fibre (increased carbohydrate source for microbial growth and therefore less waste product) as these dietary patterns often co-exist (135). Furthermore, it may be that the long term adaptive microbial changes in response to diet (266) determine our gut bacterial enterotype, such that species known to synthesise 4-cresol sulphate may be reduced in these individuals(267) in preference to other species who show higher energy harvest from vegetarian sources. The effect of diet on the metabolome is further discussed below.

Hippurate was also significantly lower in South Asian Crohn's patients, UC patients and in healthy volunteers compared to their Caucasian counterparts. As discussed above, hippurate is mammalian-microbial co-metabolite, and is associated with dietary intake of certain foods (specifically fruit, vegetables, green and black tea (249)) and is also linked to the gut microbiome (244).

Both hippurate and 4-cresol sulphate are metabolic co-metabolites and therefore altered by dysbiosis, diet or host; or most likely an interaction of all three.

Whilst the literature on comparative metabonomics across different ethnic groups is limited, it is well established that the microbiome is altered by both geography and ethnicity (84). Walsh et al demonstrated in a multi-centre European study that geographical region had a large influence on the urine metabolic profile (265). Ethnicity itself encompasses the effect of genetics, environment and diet, all of which also impact on the composition of the microbiome (268).

When comparing healthy controls of different ethnicity, in addition to hippurate and 4-cresol sulphate, other bacterial metabolites including methanol formate and PAG were different, also

suggesting a microbial origin. Energy cycle metabolites (citrate and trans-aconitate), organic acids (3-aminoisobutyrate) and triglycerides were altered, which may be dietary or genetic influences.

By demonstrating that the urine metabonomic profile varies between ethnicities, it highlights another confounder which must be accounted for when studying metabonomics across populations and in large multi-country research. Whilst age, sex and diet are often reported, ethnicity or cultural background is often lacking. It is relatively easy to standardise aspects of methodology in order to compare data, it is more challenging to describe differences between populations geographically and ethnically (265). However these effects, even if small compared to other influences on the metabolome such as disease, must be better understood in order to accurately interpret metabonomic data.

3.6.4.2 Examining IBD in different ethnicities

This study shows that the metabolic profiles of IBD patients are different in South Asians and white Northern Europeans.

In Crohn's the microbial co-metabolites 4-cresol sulphate, hippurate and TMNO were identified as statistically significant in the univariate analysis, as well as succinate. In UC the identical metabolites were affected, as well as alanine. OPLS-DA demonstrated that Crohn's appears to be metabolically different between the two ethnicities, with models showing glutamate, glucose, 4-amino-hippurate, 4-cresol sulphate, succinate, methyl-histidine, acetate, hippurate, and glycine identified in the loadings plot. Although the Q^2 value was small (0.0284) the p value was significant ($p = 0.011$).

However, in UC the OPLS-DA model failed validation testing. This is unexpected given that the univariate analysis showed differences, and that the phenotype of South Asian UC in this set was different (more pancolitis and less proctitis) to the Caucasian population. The reasons for this finding are not clear, especially as the univariate analysis showed clear differences here.

The lack of valid OPLSDA models may be a limitation of using multivariate analysis in smaller groups which lack the power of larger cohorts (number in each group of South Asians was 25-46, compared to Caucasian 69-96). Another reason reliable models cannot sometimes be built with PLS is if the groups are non-homogenous or have a high intra-class variability [52]. Although it is difficult to quantify this, it can be seen that the South Asian samples have a wider distribution on the PCA plot, suggesting a wider variation within the dataset.

One previous study has catalogued and described gut microbial changes that differ between South Asian and Caucasian UC patients, demonstrating distinct changes associated with ethnicity and disease severity (269).

Although there are no comparable data in IBD, other metabonomics studies in different ethnic groups show distinct differences in Hispanic and non-Hispanic pregnant women (270), in oesophageal cancer in groups of Han, Kazak and Uygur patients [60] and colorectal cancer in different ethnicities (271).

My study and other published studies suggest that both microbiota composition and metabolic function differ according to ethnicity, and this is likely to drive the differences seen metabolic urinary profiles in IBD.

3.6.4.3 Ethnicity and the effect of diet on the metabolome and microbiome

Nutritional research identifies diet as a major driver of the microbiota composition and function (272). It can be seen that both long standing food patterns, as well as short-dietary manipulation can alter the composition and diversity of the gut microbiota (273). And both the host metabolism and metabolic profile is altered by diet in the healthy host and the patient (272). Cardiovascular and cancer risks, as well as overall mortality has been linked to diet (274). It is therefore clear that diet plays an important role in determining health and disease through nutrition-host-microbiome interactions, although the exact role of microbes and their nutritional metabolic metabolites varies from disease to disease.

Previous metabonomic studies have tried to explore the effect of diet on the metabolic profile, and link microbial composition to diet-dependent gut microbial metabolites (275). Several population based studies have examined metabolic diversity in stratified groups to find associations with metabolic profiles and diet (246, 276). Others have examined short term nutritional interventions (249, 277).

One published study specifically comparing the urine metabolome in vegetarians and omnivores showed vegetarians have higher urinary excretion of citrate and less TMAO, methylhistidine and phenylalanine than omnivores (278). In this thesis the same pattern is seen in the healthy cohort analysis (higher citrate in the urine of South Asians, lower levels of TMAO and methylhistidine) consistent with proportionally more vegetarians in the South Asian controls group.

Alanine (higher levels in South Asians), a non-essential amino acid, has been associated with a high animal protein diet, but can also be found in vegetarian diets (legumes, pulses, beans, nuts). As mentioned above, several bacterial metabolites differed between the groups.

Although diet must be rigorously considered as a significant potential explanation for some the differences seen in metabolic profiles between ethnicities it is unlikely to be the single cause of the differences seen. It is likely that diet has an effect on metabonomics both directly, and also in influencing the gut microbial composition itself, which would contribute to separating Caucasian from South Asian metabolic profiles in both disease and healthy adults.

It also must be considered that in patients with gastrointestinal diseases, including IBD it is very common for patients to alter their diet to try and alleviate symptomatology (279). Literature suggests that more than 76% of IBD patients do this at some point during their disease (280). When examining the effect of diet on different ethnicities in IBD it may be that this manipulation in nutritional intake may have a variable, and unaccounted for, effect on the metabolic profile.

One of the challenges when assessing diet is its complex and varied composition, and how to compare cohorts to each other, or understand the contribution of each of the foodstuffs. Often definitions such as Western versus rural diet are used (52) or omnivorous versus vegetarian (281). In this thesis diet was defined as omnivorous or vegetarian (no red or white meat or fish, consumes dairy / eggs) and a 24 hour dietary recall taken. This was so that groups could be compared, and so that any metabolic differences could be interrogated to see if diet could explain them. However it is important to acknowledge that there are differences in the composition of a vegetarian diet compared to a carnivorous diet, beyond the exclusion of meat consumption. Vegetarians tend to consume more fibre and omega-6 polyunsaturated fatty acids, and less unsaturated fats, but deficiency in dietary iron and vitamin B12 is more common (272). To properly investigate the effect of diet a study would have to be appropriately controlled with comparable sized vegetarian and omnivorous groups and full food diaries used.

However, in this study it did help begin to unravel possible reasons behind differences in ethnic metabolic profiles.

As seen in phenotypic data, South Asians not only manifest IBD phenotypically differently, but also display a different metabolic profile to Caucasians with IBD. Differences are also shown between groups of healthy controls that highlight the need to consider ethnicity when using metabonomics, and that a combination of genetics, diet (beyond a simple vegetarian:omnivore comparison) and microbiome account may for these differences. Significantly these findings underline the importance of further research in this area in different ethnic groups, as well as the need to control for ethnicity as a potential confounder when using metabonomics in a clinical population.

3.6.5 Other analyses of data

3.6.5.1 Comparison of 1st vs 2nd generation South Asians

Although it was not possible to separate these groups, this analysis was limited by missing data and unequal group sizes.

3.6.5.2 Disease location

There were no valid models constructed to distinguish colonic Crohn's from UC in either ethnic (or mixed ethnicity) group. Stephens (114) and De Preter (145) also could not separate these groups with multi-variate techniques, suggesting that colonic Crohn's is metabolically similar to UC in the urinary profile. It should also be considered that there may also be an element of misdiagnosis between the 2 diseases, confusing the data.

3.6.5.3 Effect of surgery

Models showed an effect of surgery in the Crohn's patients. The effect of surgery has previously been shown to affect predictive models in IBD (114). This finding was also seen in Chapter 4.

3.6.5.4 Validation analysis: Gender

As samples are normalised within the dataset (mean-fold normalisation) to account for dilutional change, this type of data cannot be quantitatively compared to other papers. For validation the data were examined for gender differences which have been well established. The findings replicated previous studies in healthy controls (282-284) showing differences between male and female cohorts and the distinguishing metabolites were the same as seen in other studies. The groups were further sub-analysed as a healthy cohort, all IBD patients, Caucasians and South Asians. All models showed very similar Q^2 value and all showed higher creatinine in males and higher citrate and creatine in females, as shown in other studies (282). Only the South Asian group had a non-significant p value – this is likely because it was the smallest of the sub-analysis groups.

3.6.5.5 Validation analysis: Storage methodology

In one study freezing has been shown to have a very subtle effect on NMR data (285), although other papers report no effect (286, 287). Overall the reproducibility of NMR pre- and post-freezing is 'remarkable' (285). Freezing procedures may have minute effects on pH dependent metabolite peaks, but not the concentration changes of metabolites, however targeted univariate analysis overcomes this as the exact peak position is not quite as relevant (285).

As some older samples were used (to maximise numbers and improve the power of the study), comparison was made using O-PLS-DA modelling to check if there was any effect of freezing on the samples. There was no separation seen on scores plots and no positive Q^2 values between the historical and recent datasets. This excludes a significant effect of storage on the results.

3.6.5.6 Disease activity

One study suggests urine metabonomics can differentiate active IBD from quiescent disease (116), whilst other studies suggest there are no metabolic changes detectable in urine related to clinical or biochemical markers of active inflammation (118). Several urine studies did not analyse disease activity in their cohort (114, 115).

Findings in serum and plasma seem more consistent however, and have linked TCA cycle-related molecules and amino acids with disease activity indices (288), favouring the use of these biofluids, for the study of inflammation in IBD.

Urine is the preferred approach to reveal microbial effects on the metabolic signature as more urinary metabolites are bacterially-derived or produced by host-bacteria co-metabolism (62). Results show that gut bacteria do not appear to differ depending on disease activity, or when comparing samples from actively inflamed areas or not (289, 290), and this may be why it is more difficult to relate urinary metabonomic profiles to activity indices.

In this study, numbers in the active and quiescent disease groups were unequal and therefore O-PLSDA analysis would be less valid. Although it was accounted for in the clinical parameters, this study was not designed to investigate the effects of disease activity on metabolites, but in designing a study to further examine this, it would appear that a combined serum/plasma and urine study would be ideal.

3.6.6 Strengths of this study

This is, to date, the largest cohort of IBD patients ever analysed for metabolic profiling, and whilst a power calculation is not possible in metabonomics, the larger the group, the stronger the validity of the results.

The unique aspect of this study is the analysis of ethnicity in a disease area in which lower prevalent ethnicities are at higher risk of IBD in the UK and in which research is minimal and lacking. Previous studies have examined metabonomics in IBD, this work highlights the significant effect of ethnicity on the metabolic signature, in both health and disease states, and something that must be controlled for when using metabolic profiling.

Several analytical methods applied to the data ensure the robustness of the discriminatory power of the technique in investigating IBD and the effect of ethnicity. A targeted approach rather than 'all' identifiable metabolites tactic tests the hypothesis of other published data that gut microbial differences, as well as energy cycle metabolites and amino acids are responsible for separation seen in IBD metabolic profiles. In addition, for the targeted profiling analysis multiple comparison corrections were applied – other papers have not performed this on their data (114). In addition to this, re-analysis was performed on participants recorded as not taking any medications at all to remove the effect of xenometabolites on results.

3.6.7 Limitations of the study

Groups in the OPLS-DA analysis were smaller because those on 5-ASAs were excluded. There is no consensus approach to dealing with xenometabolites in the spectrum and some controversy exists around this (132). Other groups have chosen to exclude large regions of the spectra (117, 132) but include all samples. The regions containing 5-ASA and paracetamol metabolites also contain other biologically important metabolites – including glutamate,

glutamine, glutathione, cholate, homocysteine, cholate, glycocholate, methionine, methyl-succinate, proline, cystathionine, asparagine, o-cresol, phenol, methylguanidine, tyramine, tyrosine, π-methylhistidine, amongst others. Therefore excluding these portions from analysis would lose a potentially enormous amount of physiological data and skew or alter the results (223). When this approach has been used in previous studies of urinary metabolic profiling, no differences were found between clinical cohorts (132) suggesting that these areas contain crucial biologically relevant metabolites.

In this chapter excluding the samples containing 5-ASAs was the most rigorous approach, as not only does it exclude the effect of the peaks from xenometabolites, but also the effect of drugs on other metabolites and associated systemic metabolic differences. Although this does not replicate 'real life' in one sense, it could be considered a similar approach to investigating newly diagnosed patients before initiation of treatment commences.

A further potential confounder is the unequal sizes of ethnic cohorts; nearly double the white participants than South Asian. Although this is a 'real life' ratio, it may have affected the ability of the multivariate analysis to demonstrate differences in the South Asian cohort. Despite this it should be stated that group sizes were significantly larger than most other total cohort sizes in similar papers.

Dietary differences must be acknowledged as a potential confounder - one which is very difficult to control for when analysing metabolic data. Whilst dietary data was meticulously collected from each participant (including specific dietary constituents known to affect the metabolome such as green/black tea intake, as well as 'usual' diet) to aid analysis and explain results, without imposing dietary restrictions it is always difficult to truly compare diet between individuals. Partly this is because dietary recall is notoriously inaccurate (291) which makes

the determination and quantification of nutritional effect difficult. In addition the 'Western' diet, rich in saturated fat, refined carbohydrates, and food additives, is acknowledged as an independent risk factor for development of IBD (292) through the effects on the gut microbiota, effects on intestinal permeability, and inflammatory promotion. For ethnic migrants in the UK it is also likely that there is an overlap in diet between more 'traditional' South Asian diet and 'Western' diets and that a more hybrid intake of foods are consumed.

Future studies could consider excluding vegetarians, or directly comparing vegetarian-only groups, although the cohort sizes for sub-analysis in this study were too small.

In addition to diet, it is well recognised that obesity is related to dysbiosis and alterations in the metabolic profile (293). As some of the body mass index (BMI) data was missing for this study, it was impossible to account for the possible effect obesity on the results.

Lack of data to further characterise ethnicity (first / second generation status) was missing from some subjects, as well as length of time of residency in the UK was not accounted for. It may be over simplifying to categorise patients based on ethnicity alone, and may have been improved by stratifying according to time in residence of UK to account for environment as well.

Additionally, disease activity was measured by a relatively weak indicator, as HBI and SCCAI are subjective. Better biochemical parameters would be CRP, or ideally a faecal calprotectin, which are more sensitive and objective biomarkers. At the time of the study calprotectin was not available and the ethics for blood sampling not in place.

3.6.8 Future work

As with most diseases, IBD is a complex multifactorial perturbation of intra and extracelluar pathways in addition to the influences of the gut microbiome, genetics and external environment. Metabolomics is a powerful platform with which to explore these relationships in difference clinical cohorts and phenotypes, but understanding them can be complex and challenging. Whilst this data demonstrates the relationship between ethnicity, IBD and certain metabolites, further investigation is required to explore the underlying mechanisms of this, and extrapolate how much is cause or effect. Furthermore it is necessary to explore specific pathophysiological processes, including dysbiosis, to interpret metabolomic data in the context of clinical disease. The challenge is to be able to further refine the understanding of metabolic data in order to be able to specifically apply this methodology to a clinical cohort.

Importantly this study demonstrates that IBD research should be performed in different ethnic groups to represent the heterogeneous, and evolving, ethnic mix of our UK patient population. As with all results, validation studies are beneficial, and in this cohort testing the metabolite models in IBD and healthy controls, along with other gastrointestinal diseases such as infection or irritable bowel syndrome could be useful.

Characterisation of active and inactive disease with calprotectin or histology, would help link metabolic changes directly to inflammation, as HBAI and SCAI, although well validated, have been shown to lack correlation to ileocolonoscopy (294). Ideally this would be done in a longitudinal design to enable the inflammatory and healing processes to be examined and explore the temporal relationship of metabonomics in IBD.

It would also be valuable to look at metabolic profiling in more homogenous, specific groups combining the use of genotyping and phenotyping (for example ileal disease and NOD-

2/CARD-15 status). Also it would be interesting to stratify genotypes in different ethnic groups, in whom some IBD genotype associations have been shown to be different (295) to try and disentangle the tightly associated factors of ethnicity and environment.

Furthermore the future of 'omics' technology is likely to be the integration of metabolic, genetic and proteomic data allowing the investigation of the perturbation seen in IBD at the level of genes, proteins and metabolite in both the microbes and host (296).

3.7 Advances in Knowledge

- This large study confirms that IBD has a strong metabolic signature in urine, with consistent alterations in characteristic microbial metabolites and energy cycle molecules relating to inflammation.
- It has shown that urinary metabolic signatures are affected by ethnicity, in both disease and healthy states, in one of the few studies to directly compare metabonomics in different ethnic cohorts. Whilst metabolic profiling reveals valuable information about 'marker' metabolites and related pathophysiological processes, these results highlight the importance of recognising the multiple endogenous and exogenous influences on interpretation of results.
- In translating this research towards understanding and treating IBD better, studies must be performed in different ethnicities to be able to apply meaningful results to a heterogeneous patient cohort, and improve a personalised approach to IBD.

Chapter 4: Effects of time on urinary metabonomic signatures in IBD

4.1 Summary

Urine samples from IBD patients collected at two time points separated by between 7 and 9 years were analysed by H^1 NMR spectroscopy and the metabonomic profiles were compared. PCA showed clustering of individual pairs of samples indicating intra-individual stability. When analysed with multivariate statistical techniques there were no differences between the baseline samples and the repeat samples, despite a variety of clinical outcomes amongst the cohort, nor could a model be created to predict a worse disease outcome.

This may suggest that urinary metabolic signatures are stable over time, despite different disease behaviours and that stability outweighs more subtle longitudinal metabolic changes in the urine in IBD. This may limit the application of metabolic profiling in disease monitoring and prediction of treatment response, but further larger studies are required.

It has been shown that the urinary metabolome is altered by intestinal surgery, a finding consistent with other papers and results in other chapters of this thesis. Furthermore, although only a small preliminary study, signals were seen which may indicate that metabolic profile may be able to predict future surgical requirements and help stratify high risk patients in the clinical setting.

4.2 Aims and hypothesis

1. The aim was to compare baseline urinary metabolic profiles with repeated samples 7-9 years later in a cohort of IBD patients, and to analyse if clinical outcome influenced a separation in metabolic profiles which may be able to predict outcome.
2. It was hypothesised that urinary metabonomics may separate patients who subsequently develop disease progression from those with a less aggressive phenotype, and that metabolic profiles are altered by disease complications and medical interventions over time.

4.3 Introduction

The theory of the unique individual metabolic phenotype, or fingerprint, has been proposed, suggesting that our own metabolic phenotype is highly specific to an individual and largely invariant over time (297). Whilst the contributions of various environmental influences such as diet, lifestyle and age are well recognised and must be accounted for when interpreting metabolic profiles, the underlying stable part of a metabolic phenotype is thought to remain unique to a person even over a period of years (298).

Assfalg et al demonstrated that in a cohort of 22 individuals in whom multiple urinary samples were analysed (approximately 40 samples across 3 months), distinguishing features were present enabling correct assignment to the donor using supervised pattern recognition analysis with an accuracy of almost 100% (297). Furthermore, in a longitudinal study of healthy volunteers, Bernini et al collected 1849 urine samples from 31 adults over 3 years and showed that over the examined time period the individual metabolic signal was largely invariable, with all individuals being able to be reassigned to their own metabolic phenotype with a very high

correlation value (298). With metabonomic data, small variations are detectable across multiple sampling time-points in a person. Bernini described these as either 'spikes' – day-to-day variations which were attributable to diet or exercise, or 'waves' – gradual alterations persisting for several days which were strongly linked to gut microbial activity, or more significant and persistent (sometimes across years) changes called 'jumps'. These 'jumps' were attributable mainly to changes in the gut microbiota and were not influenced by external factors such as diet or lifestyle as described by the metadata.

This finding was confirmed in a shorter study assessing daily and diurnal changes in urinary metabolic profiles, which again showed relatively small deviations in an individual's own profile from external and / or temporary influences (120). Overall in healthy people, inter-individual variation hugely outweighs the effect of intra-individual variation, suggesting there is a unique metabolic core signature, which is preserved with time.

However it is not clear whether or not in patients, the effects of medications, surgery or disease itself, may more significantly alter the metabolome over time and whether or not these changes are predictable of future outcomes and/or reflect disease progression.

Longitudinal metabonomics have been used to study other diseases, including viral infections (299), bacterial infection (300) and chronic kidney disease (301). Temporal changes in metabolite expression, related to up or down regulation of biological pathways, can be linked to disease initiation, progression and/or resolution, suggesting that this approach may show potential for monitoring or predicting disease. A previous study showed that urinary metabolic profiles were different in active and quiescent IBD (116), suggesting that there are detectable changes during a disease flare, and thus longitudinal metabonomics have been proposed to have potential for tracking disease activity and response to treatment.

Pharmacometabonomics has demonstrated how metabonomics can be used to predict drug metabolism, efficacy and safety (302), and other studies have identified metabolic markers of future cardiovascular disease risk (303). In IBD, whilst there are some clinical, serological and endoscopic predictors of disease course (304), no single validated tool is available to accurately determine disease susceptibility, natural history, and therapeutic response. As metabonomic profiling is a reflection of genotype, gut microbiome and system metabolism, it may be able to identify specific phenotypes, predictive of aggressive subtypes of IBD, or non-responders to therapy for example. If this proves true, it may in future have a role in personalising the treatment of these patients.

The above studies conducted in healthy people (120, 297, 298) showed stability of the metabonomic profile over time. Only one previous IBD study has analysed any longitudinal urinary metabonomics to date. Williams et al (118) used repeat samples from IBD patients approximately 4 months apart, as part of a larger study, to validate measured levels of discriminatory metabolites. They found mean concentrations of key marker metabolites were unchanged over this time period in the group. It was not shown whether the cohorts from different time points could be separated, but instead suggested that the perturbations in the IBD metabolome were longitudinally reproducible.

It has not been fully investigated as to whether more subtle outcomes of the disease can be measured with metabolic profiling. This study aimed to assess if IBD, a chronic relapsing and remitting condition, changes the metabolic profile dynamically over time, or whether the disease diagnosis itself has a prolonged and unchanged effect on the metabolic signature. The design of this study was to compare the urinary metabonomics of the same cohort of patients re-sampled after a 7-9 year gap. And also to examine if there were changes identifiable in the

baseline cohort that could predict future disease outcome over the coming years, which may be translatable into clinical application.

4.4. Methods

4.4.1 Subjects

39 patients with IBD (based on clinical, endoscopic and histological evidence) at St Mary's Hospital (Imperial NHS Healthcare Trust) who had been sampled as part of a previous study between the years 2006 and 2008 were re-recruited to provide a further random urine sample.

Identical metadata was collected at both time sampling points (as previously reported in chapter 2.) In addition CRP levels were collected for the second set of samples (not collected at baseline).

Clinic letters, radiology and endoscopy reports were reviewed to record medication, surgical interventions, endoscopic findings and complications. Clinical outcome was defined as 'progression' if there was either

i. an escalation in medication required (initiation of immunosuppression or biologics)
ii. progression of disease location or phenotype (Montreal Classification)
iii. requirement of surgery (intestinal resection)

4.4.2 NMR Spectroscopic Analysis

Urine was analysed by 1D ^{1}HNMR spectroscopy. All samples were prepared as described previously (chapter 2) and randomised during preparation and spectral acquisition. An

exponential window function with a line broadening of 0.3 Hz was applied prior to Fourier transformation to all 1D NMR spectra.

4.4.3 Data Acquisition and Pre-Processing

Spectra were imported into Matlab using spec_preproc_v5 (in-house software) and phasing, baseline correction and referencing to TSP was performed. Resonances for TSP (δ -0.2 to 0.2), water (δ 4.553 -4.923) and urea (δ 5.645 – 6.167) were removed. Spectra were aligned using first an automated recursive segment-wise peak alignment (in-house script) (222) followed by manual alignment to minimise shifts in the peaks due to pH variation during acquisition. Data was normalised by probabilistic quotient normalization method to reduce inter-sample concentration variability and allow samples to be directly compared (227).

A PCA was constructed to identify significant outliers, and distance from the model (DModX) calculated to identify those above the critical value (Dcrit); these were excluded from the model. PCA was used to visualise the association of pairs of samples to one another.

For the supervised analysis O-PLS and O-PLS-DA models were constructed in Matlab using scripts as described in chapter 2. O-PLS models were used to show associations between the metabolic variables and continuous measurements, and O-PLS-DA models used for discrete variables. The discriminatory power of each model was tested with 7-fold cross validation (7FCV). For each multivariate model the Q^2 (predictive ability) of the model was calculated. A small or negative value indicated there was no robust discrimination possible between classes. A positive Q^2 was tested using the permutation method – where 100 permutations are generated, and a robust value was considered as within the 95th percentile ($p<0.05$).

4.5 Results

4.5.1 Phenotypic dataset

Table 4.1 shows the phenotypic data of the patients included in this study.

One patient had the diagnosis changed from Crohn's to UC in the time period. For comparative models this individual was excluded.

	CD Baseline (Time = 0)	CD (Time 1)	UC Baseline (Time = 0)	UC (Time 1)
n	22	21	17	18
Age (years) Median (Range)	42 (23-66)	48 (30-73)	39 (17-61)	48 (25-69)
Median time between samples (years)	7		7	
Male	8 (36%)		12 (70%)	
Smokers	5	2	1	2
Active disease [a]	8 (36%)	6 (28%)	3 (17%)	2 (11%)
CRP	[b]	0.9	[b]	1.9
IBD surgery	8	11	2	2
Progression of disease at time 1 : combined outcomes (%)		12 (57%)		3 (17%)
Required (further) surgery		4		0
Required addition of immunosuppressants		4		1
Required initiation of biologics		5		0
Disease location / behaviour progression		2		2
Developed other significant disease diagnosis		2		2
Age at diagnosis: Median (range)	24 (12-53)		30 (10-63)	
Medications use in preceding 4 weeks [c]	M1 : 9	M1 : 5	M1 : 8	M1 : 13
	M2 : 7	M2 : 11	M2 : 8	M2 : 8
	M3 : 1	M3 : 5	M3 : 0	M3 : 1
	M4 : 0	M4 : 0	M4 : 0	M4 : 1
Disease location [d]	L1 : 3	L1 : 2	E1 : 7	E1: 5
	L2 : 8	L2 : 8	E2 : 1	E2 : 2
	L3 : 10	L3 : 12	E3 : 9	E3: 11
Disease behaviour [e]	B1 : 10	B1 : 9		
	B2 : 5	B2 : 6		
	B3 : 7	B3 : 8		

Table 4.1: Characteristics of the study population

[a]As defined by disease activity index ≥5 [b]Data not available [c] Medications: M1: 5-ASA; M2: immunosuppressants (Azathioprine/6-MP); M3:anti-TNF; M4:steroids [d] Location (206): L1:ileal disease; L2: colonic disease; L3:ileocolonic disease; E1:proctitis; E2:disease limit distal to splenic flexure; E3:disease proximal to the splenic flexure [e] Behaviour (206): B1:inflammatory; B2:structuring; B3:fistulating

4.5.2 Statistical analysis

4.5.2.1 PCA

When visualising the spectral data, two samples from the same patient were dominated by huge glucose peaks and therefore removed from the dataset. A PCA was constructed with the remaining 76 samples. Variables were mean-centred and scaled to unit variance, to prevent the model being dominated by large values and a Hotellings T2 eclipse was generated to visualise outliers. From the loadings of the PCA model spectral peaks correlating to paracetamol and 5-ASA (117, 226) were identified as driving the model and therefore the regions at δ2.134-2.187, δ7.113-7.182 and 7.337-7.402 were removed and the PCA remade.

From this 2 samples were shown as strong outliers (1 Crohn's, 1 UC) which would incorrectly sway the model and therefore should be excluded. To identify any moderate outliers (which would not be good fit to the model) the distance from the model (DModX) was calculated for each sample, and those above the critical value (Dcrit) were excluded from the model. 1 further moderate outlier (Crohn's) was thus excluded. Of the excluded outliers 2 samples contained drug peaks (unidentified) and one contained very high levels of dimethylamine in the spectra (no known reason identified).

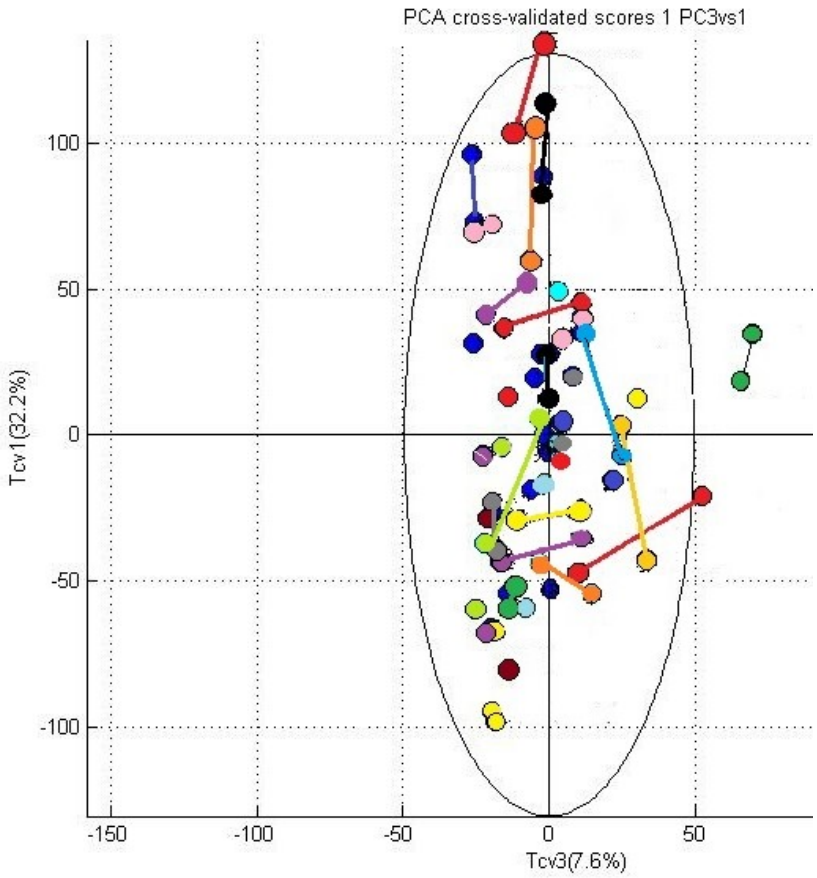


Figure 4.1: PCA scores plot showing coloured pairs of samples taken from same patients

The PCA scores plot shows clustering of pair of samples from the same individuals, as shown with coloured links, with most samples from the same person being closely clustered.

4.5.2.2 O-PLS-DA and O-PLS regression

4.5.2.2.1 Comparison of longitudinal samples

For the supervised analysis O-PLS-DA models were constructed to see if there was a difference between baseline samples and repeat samples – this was done for all combined samples and each IBD subtype (table 4.2). All Q2 values were negative (no robust predictive model could be made) suggesting there were no distinct differences between the cohorts when sampled 7 years apart.

	Comparison	Q^2 value
All samples	Baseline vs time 1	-0.123
UC patients		-0.005
CD patients		-0.087

Table 4.2: Summary of OPLS-DA models to separate baseline samples from time point 1 across all samples, and by disease subtype. Q^2 for each model are described.

4.5.2.2.2 Separating samples by disease subtype and disease activity

Models were made to look for significant differences in the UC versus CD cohorts from all samples, at baseline and at time 1. For positive Q^2 values a permutation test was run to calculate a p value to test significance. See table 4.3. The models were unable to accurately separate UC from CD when analysed at separate time-points (baseline or time 1) or with all samples together.

	Comparison	Q^2 value	P value
All samples	UC : CD	0.0286	0.11
Time 0		0.061	0.22
Time 1		0.0046	0.37

Table 4.3: O-PLS-DA models to distinguish UC from Crohn's samples

No models could be created to show differentiation of active disease from inactive disease. (table 4.4).

	Comparison	Q^2 value	P value
All samples	Active : inactive disease	-0.093	
CD		0.0574	0.13
UC		-0.06	

Table 4.4: O-PLS-DA models to distinguish patients with active disease (HBI or SCAI ≥5) from quiescent disease. Q^2 for each model is described, p value calculated for positive Q^2 only.

4.5.2.2.3 Analysis of gender cohorts

When analysing according to gender, models for each separate time point showed differences between sexes. All Q^2 values were positive with significant p values (table 4.5). The loadings plots for these three models identified the same discriminatory metabolites – creatinine (most significant), hippurate, and citrate.

	Comparison	Q^2 value	P value
All samples	gender	0.2051	0.01*
Time 0		0.141	0.02*
Time 1		0.282	0.01*

Table 4.5: O-PLS-DA models to distinguish IBD patients by gender. Q^2 and p value for each model are described. *significant if ≤0.05

4.5.2.2.4 Modelling according to clinical outcome and medications

Multivariate models were calculated to see if groups of patients with different clinical outcomes could be separated. Specifically baseline samples were compared according to future outcome, to see if retrospectively those patients who went on to require (further) surgery or escalation of treatment could be 'predicted'.

These models were unable to discriminate between samples from patients that had stable disease over the time period between samples, or patients that had disease progression. However the subgroup of patients who had undergone surgery showed robust separation with significant p value on testing. Baseline samples also showed a strong model 'predicting' those who would go on to require (further) surgery (all CD), although it is worth noting that this group were small (n=4).

Analysis showed no difference between subgroups of patients according to taking different medications (table 4.6).

Timepoint	Comparison	Q^2 value	P value
All samples	Progressive: stable disease[a]	0.0746	0.12
Baseline		0.0008	0.26
Timepoint 1		0.0008	0.13
All samples	Post-surgical patients vs never had surgery	0.326	0.01*
Baseline	Patients who went on to require future surgery	0.015	0.03*
Baseline	Patients who went on to require future biologics	-0.36	
Baseline	Patients who went on to require future immunosuppression	-0.148	
All samples	Taking immunosuppression	-0.0776	
Baseline		0.013	0.21
Timepoint 1		-0.193	
All samples	Taking 5-ASA	-0.078	
Baseline		-0.413	
Timepoint 1		0.0865	0.11
All samples[b]	Taking Anti-TNFs	0.065	0.47

Table 4.6: O-PLS-DA models to compare groups based on disease outcome and medication use. [a]as defined above as either an escalation in medication required (initiation of immunosuppression or biologics), progression of disease location or phenotype (Montreal Classification) or requirement of surgery (intestinal resection). [b]Separate time point analysis not performed because of small numbers. *significant if ≤0.05

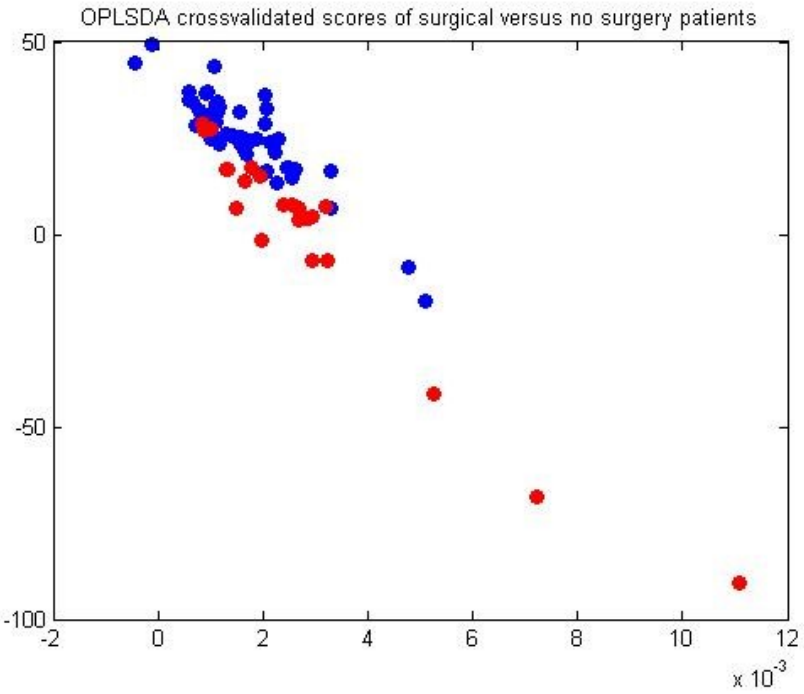


Figure 4.2: O-PLS-DA cross validated scores comparing post-surgical patients and patients who have never had surgery. Red spots are those who have had surgery, blue spots are patients who have never had surgery.

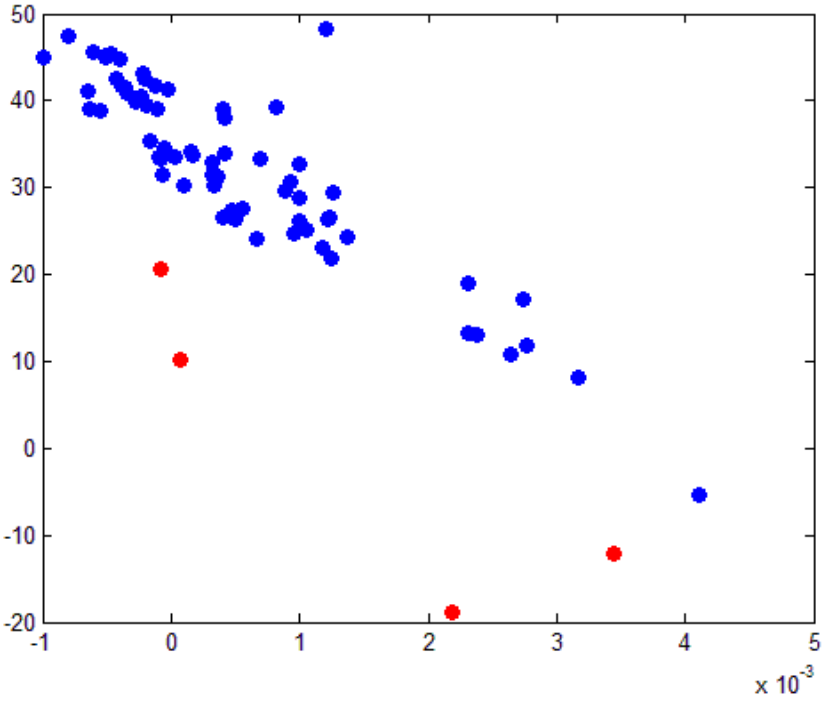


Figure 4.3: O-PLS-DA cross validated scores of baseline samples showing patients who went on to require future surgery (red).

4.5.2.2.5 Modelling according to CRP and age

OPLS regression was used to investigate if there was a relationship between the metabolic data and continuous data; CRP at time point 1 (data missing for baseline samples) and age (analysed as combined data and for each subgroup at each time point) (table 4.7). There were no statistically significant associations seen.

Variable	Subgroup	Time point	P value
CRP		Time 1	0.97
Age	All samples	All samples	0.06
		Time 0	0.37
		Time 1	0.23
	CD	All samples	0.24
		Time 0	0.42
		Time 1	0.79
	UC	All samples	0.26
		Time 0	0.88
		Time 1	0.37

Table 4.7: OPLS regression models for CRP and age

4.6 Discussion

This preliminary study was undertaken to examine whether urinary metabolic profiles in IBD alter over a prolonged time, and whether there may be metabolic markers which could separate those patients who subsequently had disease progression from those in whom disease remained stable. To date there are no longitudinal data in IBD metabonomics to compare this to in the literature; this is the first work (as far as known) examining the effects of time on metabonomic signatures in this disease area.

Results from this cohort do not appear to confirm the hypothesis that subtle IBD-related changes may be demonstrated in longitudinal metabolic studies, and in fact, the overall stability of individual profiles may make revealing and interpreting subtle changes challenging.

4.6.1 PCA analysis

One of the most complex aspects of handling metabonomic data is analysis and interpretation of longitudinal data (305), especially when there are only a small number of samples from each individual. However longitudinal data potentially hold key information describing the evolution of biological processes in a population and its relationship to dynamic disease. Such studies are required to enable metabonomic data to be applied for monitoring or prognostication purposes (299). Alternative analytical approaches such as Markov and Bayesian models require multiple samples from an individual to construct a trajectory curve (305), for which this study was not adequately powered. In this study PCA was used, as it is a well suited technique to handle a longitudinal dataset (306), and variation in the data can be displayed and related to a priori information.

PCA visualisation of these data showed clustering of pairs of samples from the same patient in many cases. Whilst a correlation coefficient cannot be calculated from only 2 samples from the same individual, there are clear intra-individual similarities displayed on the PCA, despite a wide array of clinical outcomes within the cohort, both disease and non-disease related, as well as multiple external factors which have changed between the sampling time points. As discussed it has been shown that in healthy adults inter-individual variation is more significant than intra-individual variation over time suggesting that the unique metabolic signature remains relatively stable, and this may hold true in patients, even with the effects of disease over time and medications.

4.6.2 Multivariate analysis

4.6.2.1 Longitudinal analysis

When analysed as a whole group, and also separately as UC and Crohn's, there were no significant differences between the collective samples at the two time points. This may suggest that the metabonomic signatures are not measurably changed by time; or it may reflect the fact that the cohorts in this study are too small to detect differences.

There may be too many confounders, including a wide range of clinical outcomes in the cohort, to extract subtle alterations in metabolic detail. Interpreting the multiple influences on complex NMR data can be challenging, and to analyse in detail the effect of specific confounders would require much larger and more stratified cohorts. This study highlights that whilst the diagnosis of IBD has strong, detectable metabolic consequences, the more subtle variations from disease progression or medications may be difficult to accurately extract and would need careful interpretation in other studies.

4.6.2.2 Disease subtypes

In this small cohort, multivariate analyses were unable to separate the disease subtypes from one another, which is the same finding as other urinary metabonomic studies of similar sizes (114-116). Only Williams et al (118) have been able to differentiate UC from CD using urinary metabonomics and this may be due to the significantly larger patient numbers in their study.

4.6.2.3 Gender differences

When analysing the effect of gender, there was a clear difference between males and females which has been well described in literature (282-284). Creatinine (higher in males) and citrate (higher in females) are the most consistently described metabolites. The loadings plot from the models showed creatinine, hippurate and citrate as most influential in distinguishing between gender groups, as seen in other studies, and also as discussed in Chapter 3.

4.6.2.4 Progression and the effect of surgery

Models were unable to separate the group of patients in whom disease had progressed from those in whom disease remained stable, at either time point or when all sampled were analysed together. No predictive models could be created from baseline samples to indicate those whom would require escalation of medical treatment to biologics or immunosuppression.

Although this is a small snapshot study, it may indicate that subtle changes may prove difficult to use to predict outcomes or monitor treatments.

In the sub-analysis of baseline samples, the 4 patients who had never had previous IBD surgery, but subsequently went on to require resection, could be identified from a model with a significant p value. Interpretation of this result should be made cautiously in view of the small number of patients and low Q^2. But is an interesting finding that could be examined in more

detail in future investigations, as it may have significant potential as a clinically relevant biomarker.

The model was also able to separate the subgroup of patients who had undergone bowel resection from patients who had never had surgery. The effect of surgery has previously been shown to affect predictive models in IBD (114) and shows that surgery exerts a strong metabolic effect, possibly through alterations in the microbiome.

4.6.2.5 Disease activity

In this study there was no clear separation in O-PLS-DA models between IBD patients with active disease and those in remission. Models were also tested in UC and Crohn's separately and no separation seen. Further analysis of each disease at single time points was not possible because of small numbers. This is consistent with a study in UC (132) which was also unable to show differences in active or inactive colitis. A later study however, was able to differentiate between active and quiescent IBD (116), suggesting that this may be possible, although there are no subsequent studies to confirm this.

4.6.2.6 Medications

There were no differences seen on multivariate analysis between any subgroups on medications. The 5-ASA drug metabolites were removed in pre-processing, and immunosuppressants and anti-TNFs have not been shown to have detectable xenometabolites in urine. Stephens et al (114) did show a difference in the urine metabonomics of Crohn's patients on anti-TNF treatments, although it is not stated what metabolic differences were shown (whether this was detectable drug or systemic metabolic effect). In our study one of the outliers removed from the model was a Crohn's patient on anti-TNF treatment, but as there were only 6 patients on anti-TNFs in total, interpretation of this subgroup is limited.

4.6.3 Limitations of the study

The major limitation to this work is the size of study and only having a pair of samples from each patient rather than several repeated samples. Additionally the study would be improved by a control group of healthy volunteers monitored over the same time period for comparison.

A further limitation is that within a relatively small group of IBD patients there was a wide diversity of clinical outcomes, including medications, surgery, other diseases. Although this reflects real life, it does affect the ability to build a predictive model if subgroups are very small and/or heterogenous.

4.6.4 Future work

In order to develop metabonomics as a biomarker tool for IBD follow-up, a clear association with time and disease progression must be established. This pilot work suggests there may not be a large enough or significant alteration detectable using this technique and this may therefore limit its use in the clinical arena as a monitoring tool.

To adequately assess these questions a prospective study is required, with more frequent samples, in a larger cohort, to assess effect of time, natural history of disease, response to treatments and requirement for surgery.

4.7 Advances in Knowledge

- Changes in the urinary metabolic profiles of IBD patients appear stable over a period of several years. This may make the longitudinal use of this technique for monitoring

disease difficult, without further understanding of the more subtle effects of activity, medication and disease progression.

- Metabonomics may be able to predict the requirement for future surgery, but this is a preliminary finding.
- Surgery appears to have a significant effect on the urinary metabolic phenotype of an IBD patient which is likely to be linked to microbiome changes.

Chapter 5: Pilot Study integrating microbial and metabonomic data in IBD

5.1 Summary

Dysbiosis is widely recognised as being integral to the pathogenesis and perpetration of IBD, although there is no accepted consensus description of the microbial changes or its exact relationship with gut inflammation. Metabonomics have been successfully applied to demonstrate metabolic changes in non-invasive biofluids in IBD.

Whilst metabonomics provides useful information about cellular perturbations, and microbiomics described important gut microbial changes, the integration of these 'omic' techniques may potentially reveal more about the functional interaction of the host-microbiome-metabolome in IBD.

This pilot study was undertaken to explore the integration of tri-omics in IBD for the first time. In this chapter 16S rRNA profiling of faeces, as well as metabolic profiling of urine and faecal water are examined in 46 adults (34 with IBD), and the discriminatory metabolites correlated with marker bacteria. Consistent with other chapters, urinary metabolic profiling strongly discriminated between UC, Crohn's and healthy controls, with significant metabolites being microbial or inflammatory in origin. Faecal water metabonomics clearly separated UC from controls, although Crohn's could not be robustly identified. As seen in other published data the metabolites were mostly faecal short chain fatty acids as well as microbial co-metabolites.

16S rRNA profiling confirmed reduction in diversity in IBD compared to controls, and in differential abundance analysis, *Veillonella*, *Acidaminococcus*, *Lactobacillus* and *Streptococcus* were demonstrated to differ between groups. Correlation coefficient testing revealed the relationships between specific bacterial genera and significant discriminatory metabolites.

5.2 Aims and hypothesis

1. The aim was to describe gut microbial composition and diversity of species in IBD patients and compare them with healthy adults using 16S rRNA profiling of stool.
2. The second aim was to examine with metabonomics the urine and stool of the same cohort, and correlate the most abundant NMR visible metabolites present in these biofluids with the bacterial species identified from the gut microbiome.
3. The hypothesis was that marker metabolites identified in urine and stool differentiating IBD from controls can be directly correlated to dysbiosis in the disease through identified alterations in gut bacteria.
4.

5.3 Introduction

The gut microbiome has become a high priority for IBD research due to the overwhelming evidence of ecological and functional perturbations seen in the disease (72). Whilst there is a large and growing volume of studies examining the composition and metabolic potential of gut microbes, the precise role of the microbiome in IBD remains unclear. Dysbiosis in this disease has been well established by studies, with an almost unanimous finding of a reduction in biodiversity compared to healthy controls (307). Not all microbial studies have shown consistent results however, and it is likely that the exact 'IBD microbiome' is heterogeneous and individual-specific. Not only do intrinsic host factors affect the microbial milieu, including ethnicity and genetics, there are also multiple external factors including diet and environment.

Twin studies (concordance rates in identical twins of 40-50% in CD and only 10% in UC) demonstrate that beyond genetics, environment plays a significant role in disease risk (308). The increasing incidence in Westernised countries is occurring at a rate too rapid to be genetic,

and may well be at least partly explained by the ‘hygiene hypothesis’, where less early exposure to microbes is critical to shaping development of the immune response, and determining the future risk of chronic inflammatory disorders (308).

Further evidence from genetic analyses link IBD with loci associated with immune responses to bacteria (309), and these specific gene alleles have also been tied directly to changes in the microbial community (92).

Despite most healthy adults having a microbiome highly variable from one another, and variable over time (77), the metabolic ability of the bacterial genes actually remains very stable (77). So in addition to taxonomic profiling, one of the important challenges remains better understanding the function of the gut microbiome and how this ties into the complex paradigm of disease initiation, propagation and healing. Hence it is not sufficient to merely catalogue the microbiome, but correlation with functional data is essential.

16S studies are a powerful, rapid and relatively cheap technique to catalogue and quantify bacterial species, however, unlike metagenomics, their main limitation is that they do not generate function or metabolic data. Metabonomics identifies the metabolites present in a biofluid in order to describe the metabolic and biological processes of an individual. By combining this approach with microbial profiling in the same cohort, gut-microbial-host interactions can be studied, and correlations between bacteria and metabolic products can be examined.

A recently published study in an Italian IBD cohort has shown correlations between 16S microbial data and faecal water (128). In a similar approach, in this chapter both stool and urine metabonomics are correlated with gut microbes in the same patients, to evaluate if robust associations can be determined with this data integration approach.

5.4. Methods

5.4.1 Subjects

46 adult subjects, all Caucasian, consented to donate stool and urine samples (given on the same day). 34 were IBD patients (14 UC and 20 CD) at St Mary's Hospital and 12 were self-reported healthy volunteers. Exclusion criteria for IBD patients included intestinal surgery, therapeutic diet, other significant co-morbidities (as defined in chapter 3) or pregnancy. Metadata and dietary data were recorded, as previously described.

5.4.2. Urine and Stool Sample Preparation

Samples were prepared as described in sections 2.2.1 and 2.2.2. and preparation for NMR was conducted as in sections 2.4.2 and 2.4.3.

5.4.3. Urine and Faecal water NMR Spectroscopic Analysis

^{1}H NMR spectroscopy was performed on the samples in a randomised order on the Avance 600MHz NMR spectrometer (Bruker Biospin) using the protocol described in 2.4.1.

5.4.4. Urine data Pre-Processing and Analysis

Phasing, baseline correction and removal of urea, water and TSP peaks was performed in Matlab as written in section 2.4.2. For this dataset the regions were defined as TSP δ -0.2 – 0.2, water δ 4.51-5.00 and urea δ 5.52 – 6.19. RSPA, hand alignment and PQ normalisation were performed (see section 2.4.2.).

STOCSY (235) was performed on 5-ASA resonances and STOCSYE (224) used to edit out peaks with a correlation value of greater than 0.7, using peaks at 2.17, 7.73 as references. This

was checked against metadata and confirmed that all samples who had peaks removed were patients taking 5-ASAs.

PCA and OPLS-DA multivariate analyses were performed, with metabolites identified from valid models using assignment strategy described in chapter 2.

5.4.5. Faecal water data Pre-Processing and Analysis

The faecal water NMR data was pre-processed as per the urine. In this dataset TSP was removed in the regions δ -0.2 – 0.2 and water δ 4.863 to 4.703. Normalisation by probabilistic quotient method was applied.

5.4.6. DNA extraction and 16S rRNA gene sequencing

Faecal water was extracted from stool samples and DNA extracted as per section 2.2.3 and 2.3.1. 16S rRNA gene sequencing was performed on V1-V2 regions using Illumina MiSeq 2 step technique. Bacterial community composition was assigned using SILVA taxonomy (310) which ranks to the level of genus.

5.4.7. Microbiome analysis

Microbial composition, diversity profiling, as well as analysis of differential abundance was performed using microbiomeanalyst.ca (215) Statistical analysis was with the Mann–Whitney *U* test (comparison of groups) with a p value ≤ 0.05 considered to be significant.

5.4.8. Correlation of gut microbiota with urinary and stool metabonomics

Statistical correlation was analysed using Matlab R2014a (Version 8.5, Mathworks). Correlations were determined using Spearman's correlation tests. Heatmap correlograms were generated to display the relationship between the significant metabolites identified from each

biofluid and the bacteria identified as differentiating between IBD and controls. Correlations were tested with a p value ≤0.05 considered to be significant. Significant correlation coefficients were plotted as coloured pie charts (non-significant correlation coefficients represented as blank spaces) representing positive (red) or negative (blue) association and pie chart to demonstrate strength of correlation.

5.5 Results

5.5.1 Phenotypic dataset

Table 5.1 shows the characteristics of the subjects in this chapter.

The UC population was significantly older than the Crohn's group or controls. There were no statistically significant differences between the groups in terms of gender, smokers, or between proportions of patients with active disease. The UC group used more 5-ASAs.

	Healthy controls	CD	UC	p value[e]
n	12	20	14	
Age (years) Median (Range)	34 (23 - 67)	45.5 (23 - 79)	53 (25 – 76)	0.001*
Male Number (%)	6 (50%)	6 (30%)	6 (42%)	0.50
Smokers Number (%)	2 (17%)	0	2 (14%)	0.52
Active disease[a]		6 (30%)	2 (14%)	0.29
Medication use in preceding 4 weeks[b]		M1 : 5	M1 : 9	0.02*
		M2 : 5	M2 : 4	0.81
		M3 : 5	M3 : 1	0.17
Disease location [c]		L1 : 8	E1 : 6	
		L2 : 8	E2 : 4	
		L3 : 4	E3 : 4	
Disease behaviour [d]		B1 : 10		
		B2 : 8		
		B3 : 2		

Table 5.1 Characteristics of the study population

[a]As defined by disease activity index ≥5 [b] Medications: M1: 5-ASA; M2: immunosuppressants (Azathioprine/6-MP); M3:anti-TNF [c] Location (206): L1:ileal disease; L2: colonic disease; L3:ileocolonic disease; E1:proctitis; E2:disease limit distal to splenic flexure; E3:disease proximal to the splenic flexure [d] Behaviour (206): B1:inflammatory; B2:structuring; B3:fistulating [e]p value as calculated by Chi squared test or Kruskal-Wallis *significant if ≤0.05

5.5.2 NMR urinary analysis

5.5.2.1 Principal Components Analysis

PCA plot showed 2 outliers (identified as one Crohn's patient and one healthy control) and these were excluded from the model and any subsequent analysis. These 2 individuals had the highest BMIs of all subjects (32.8 and 38.6). There were no moderate outliers (as defined by above the critical distance from the model Dcrit) to exclude.

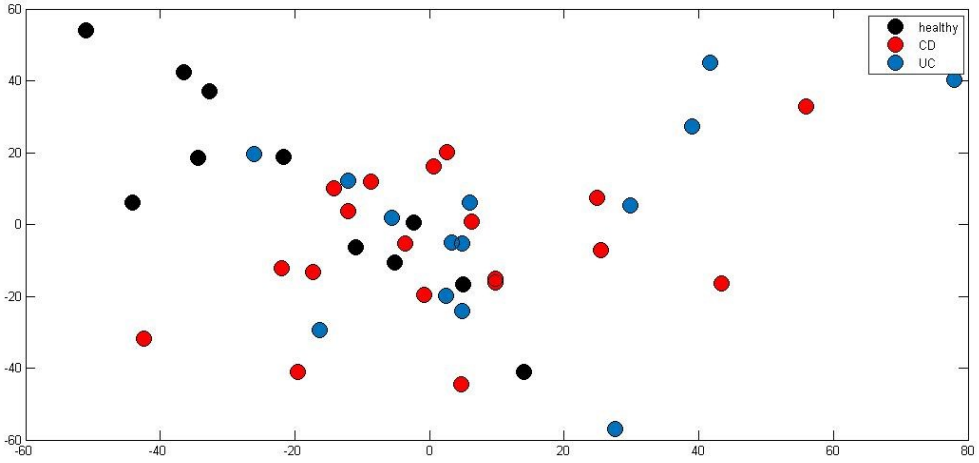


Figure 5.1: PCA scores plot of all included urine samples, coloured according to class with 2 components displayed. Outliers have been removed.

5.5.2.2 O-PLS-DA modelling

OPSLDA models were made to compare groups according to diagnosis. The summary of models is shown in table 5.2.

Urine samples				
	n	R^2	Q^2	P value (1000 permutations)
Healthy vs IBD	44	0.74	0.33	**0.001***
Healthy vs CD	30	0.33	0.28	**0.028***
Healthy vs UC	25	0.34	0.43	**0.020***
UC vs CD	33	0.03	0.10	0.220

Table 5.2: Summary of OPLS-DA models of urine comparing 3 groups. R^2, Q^2 and p value for each model are described. *significant if ≤0.05

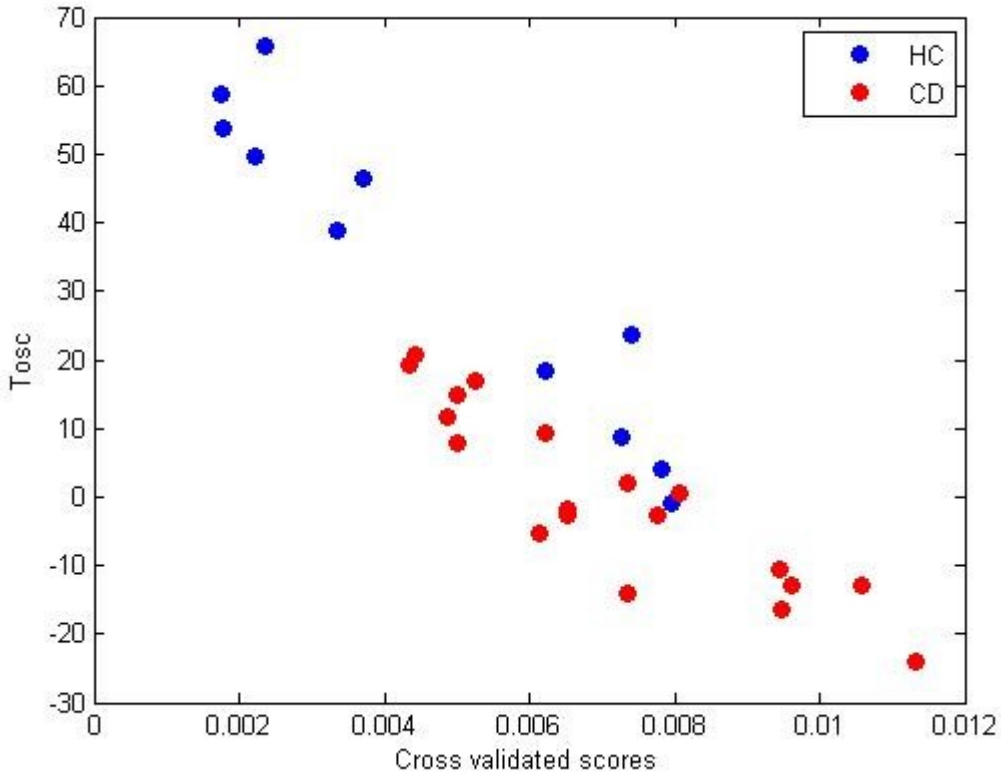


Figure 5.2: OPLS-DA cross-validated scores for healthy controls and Crohn's patients.

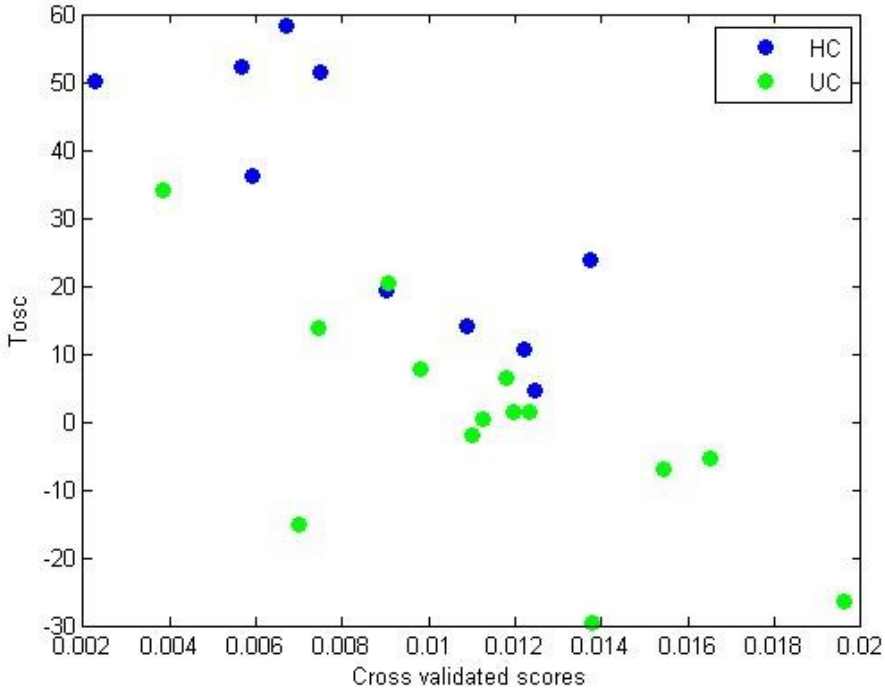


Figure 5.3: O-PLS-DA cross-validated scores for healthy controls and UC patients.

The O-PLS-DA models were able to show robust (7-fold-cross-validation) and significant (post permutation testing) separation between healthy controls and both types of IBD. In this small group it was not possible to distinguish between CD and UC.

The corresponding coefficient plots (loadings) for the valid models were examined to identify metabolites associated with the differences between the groups.

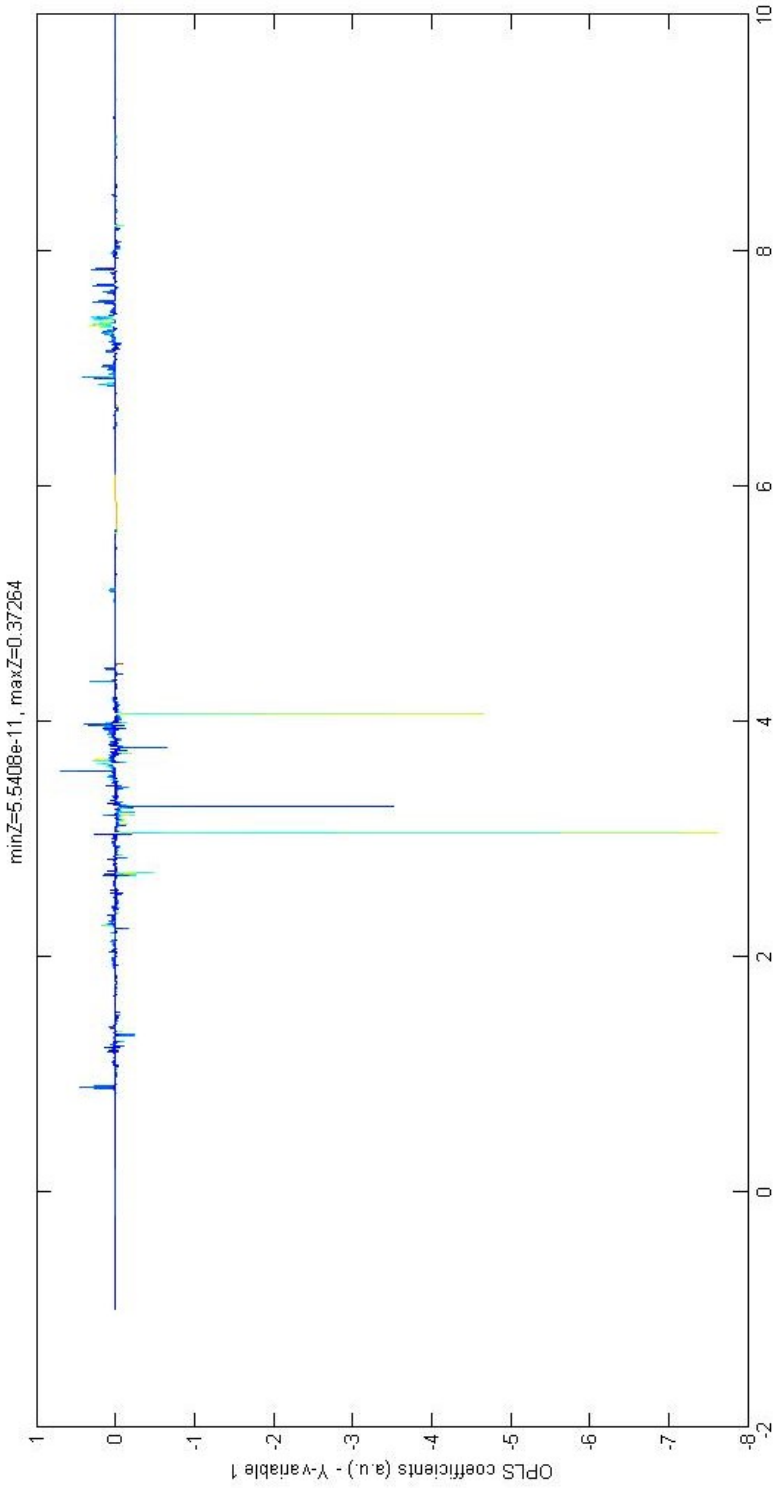


Figure 5.4: Coefficient plot for urine OPLS-DA model of healthy controls and Crohn's patients colour-coded by the correlation coefficient (r^2).

	Significantly correlated* metabolites in urine *in order of importance*
Healthy vs Crohn's	**N-methyl-2-pyridone-5-carboxamide ↑:↓** **Methylnicatinamide ↑:↓** **1-Methylnicotinamide (NMND) ↑:↓** **Dimethylamine ↑:↓** **2-Hydroxyisobutyrate ↑:↓** **Creatinine ↑:↓** Gamma-hydroxybutyrate (4-hydroxybutyrate) ↓:↑ Malonic acid ↓:↑ **Valine ↑:↓** **Dimethylsulfone ↑:↓** Phenylacetylglycine (PAG) ↓:↑ **1,9-dimethylurate ↑:↓** **Succinate ↑:↓** **Hippurate ↑:↓**
Healthy vs UC	3-Hydroxyphenylacetate (3-HPA) ↓:↑ **Trimethylamine ↑:↓** **N-acetyl-L-aspartic acid↑:↓** **Unidentified (d) δ 1.468 ↑:↓** **Valine ↑:↓** **2-hydroxyisobutyrate ↑:↓** 3-Methyl-L-histidine ↓:↑ **N-butyrlglycine ↑:↓** **5-hydroxylysine ↑:↓** **Methylnicatinamide ↑:↓** **Butyrate↑:↓** Unidentified (dd) δ 1.981 ↓:↑ 3,4-Dihydroxyphenylacetate ↓:↑ Gamma-hydroxybutyrate (4-Hydroxybutyrate) ↓:↑ **Creatinine↑:↓**

Table 5.3: Significantly correlated urine metabolites identified from loadings plots. *r value with $p<0.05$ **Bold ↑:↓** indicates positive direction of relative change of metabolite in first group relative to second.

The corresponding loadings plots show the differences in the metabolic profiles of healthy people and those with CD were related to N-methyl-2-pyridone-5-carboxamide, methylnicatinamide, NMND, dimethylamine and 2-Hydroxyisobutyrate. Hippurate featured lower on the list. When comparing healthy people and UC patients 3-HPA, trimethylamine, n-acetyl-L-aspartic acid and an unidentified compound (with doublet at δ 1.468).

The area under the peak of selected metabolites identified by multivariate modelling as being significantly different between groups were measured. These selected urinary metabolite values were used to correlated with microbial data in a correlogram to identify significant correlations.

5.5.3 NMR faecal water analysis

5.5.3.1 Principal Components Analysis

As per analysis of urine samples, initial PCA was used to identify significant outliers. One sample (UC patient) was subsequently removed.

When STOCSYE was applied to remove 5-aminosalicylic acid resonances, some xenometabolites still appeared in the spectral data. Therefore in line with published data in faecal water NMR, the regions of 2.156 – 2.172 and 6.78 – 7.775 containing these peaks were removed. Although this did cut out a proportion of spectral data, it allowed interpretation of multivariate analysis without the effect of 5-ASA and its metabolite, N-5-ASA dominating.

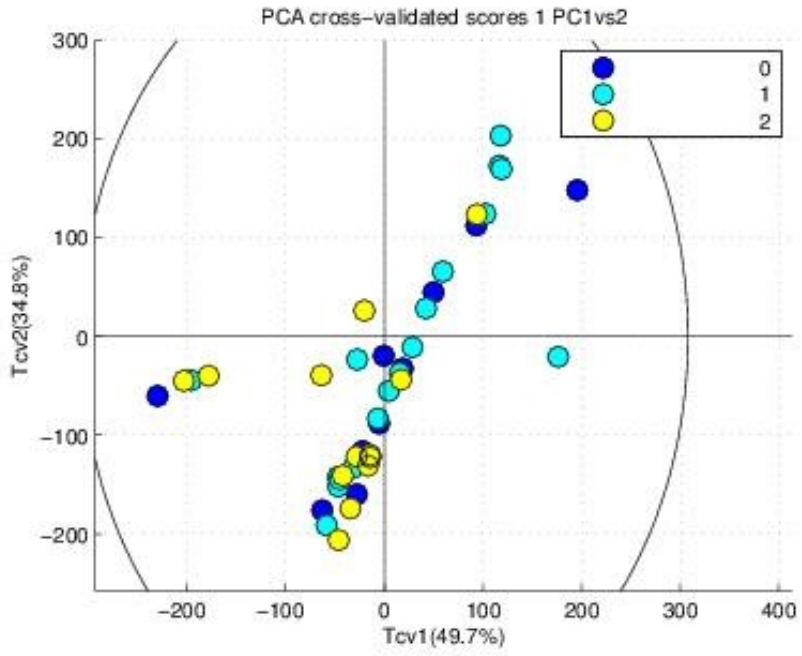


Figure 5.5: Cross-validated PCA score plot of all included faecal water samples, coloured according to class with 2 components chosen. (0 = healthy controls, 1 = Crohn's patients and 2 = UC).

5.5.3.2. O-PLS-DA modelling

For the supervised analysis OPLS-DA models were constructed to compare each pairwise comparison.

Faecal water samples				
	n	R^2	Q^2	P value (1000 permutations)
H vs IBD	45	0.6205	0.332	0.67
H vs CD	31	0.437	0.447	0.58
H vs UC	25	0.168	0.173	**0.010***
UC vs CD	33	0.017	-0.04	0.78

Table 5.4: Summary of PLS-DA models of faecal water comparing 3 groups: Healthy controls, Crohn's and UC. *significant if ≤0.05

Despite reasonable Q^2 values, only the model comparing healthy controls with UC showed robust separation after permutation testing. This model was interrogated for significant metabolites (r value with calculated p value <0.05) responsible for separating the groups.

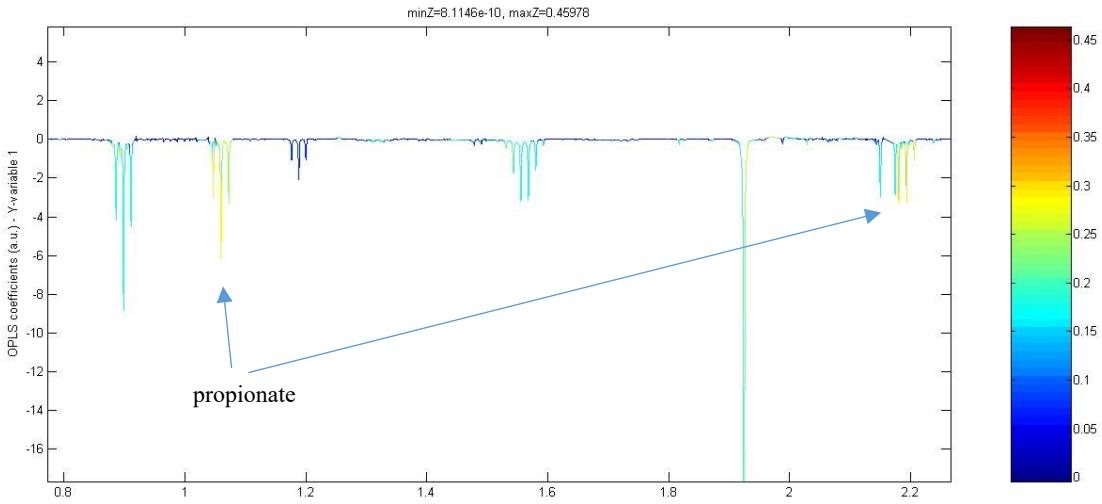


Figure 5.6: Coefficient plot for faecal water OPLS-DA model of healthy controls and UC colour-coded by the correlation coefficient (r^2).

	Significantly correlated* metabolites in urine *in order of importance*
Healthy vs UC	**Propionate ↑:↓** **Proline ↑:↓** **Isoleucine ↑:↓** **Acetate ↑:↓** **Aspartate ↑:↓** **Butyrate ↑:↓** Trimethylamine ↓:↑ **Alanine ↑:↓** **Threonine ↑:↓** **Leucine ↑:↓** **Methylamine ↑:↓**

Table 5.5: Significantly correlated faecal water metabolites identified from loadings plots of healthy patients versus UC. *r value with p<0.05 **Bold ↑:↓** indicates positive direction of relative change of metabolite in first group relative to second.

Propionate, proline, isoleucine, acetate, aspartate and butyrate were the most correlated metabolites found. All of these were higher in healthy adults than in UC.

5.5.4. Microbial Analysis

5.5.4.1. Microbial composition analysis

The gut microbiome in the majority of subjects was predominantly *Bacteroidetes* and *Firmicutes* with a considerable variation in the ratio of these in all groups. Two subjects (both with Crohn's) had a strong predominance of *Proteobacteria*.

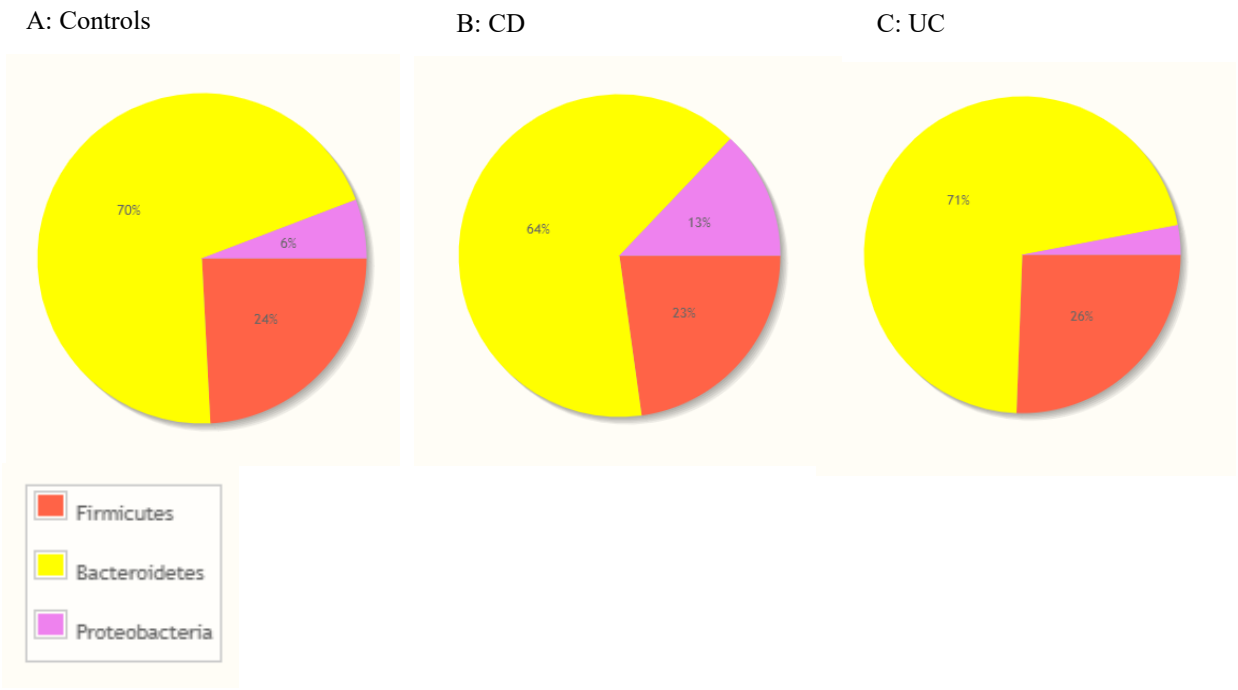


Figure 5.7: Mean phylum-level abundance for Controls, Crohn's and UC

Mean phylum-level relative abundances (percentage) are shown in figure 5.7. UC patients had a similar ratio of *Bacteroidetes* and *Firmicutes* to healthy controls, whereas both these species were reduced in the Crohn's group.

Microbial composition of samples at order, family and genus level are shown in figures 5.8, 5.9, 5.10 and 5.11.

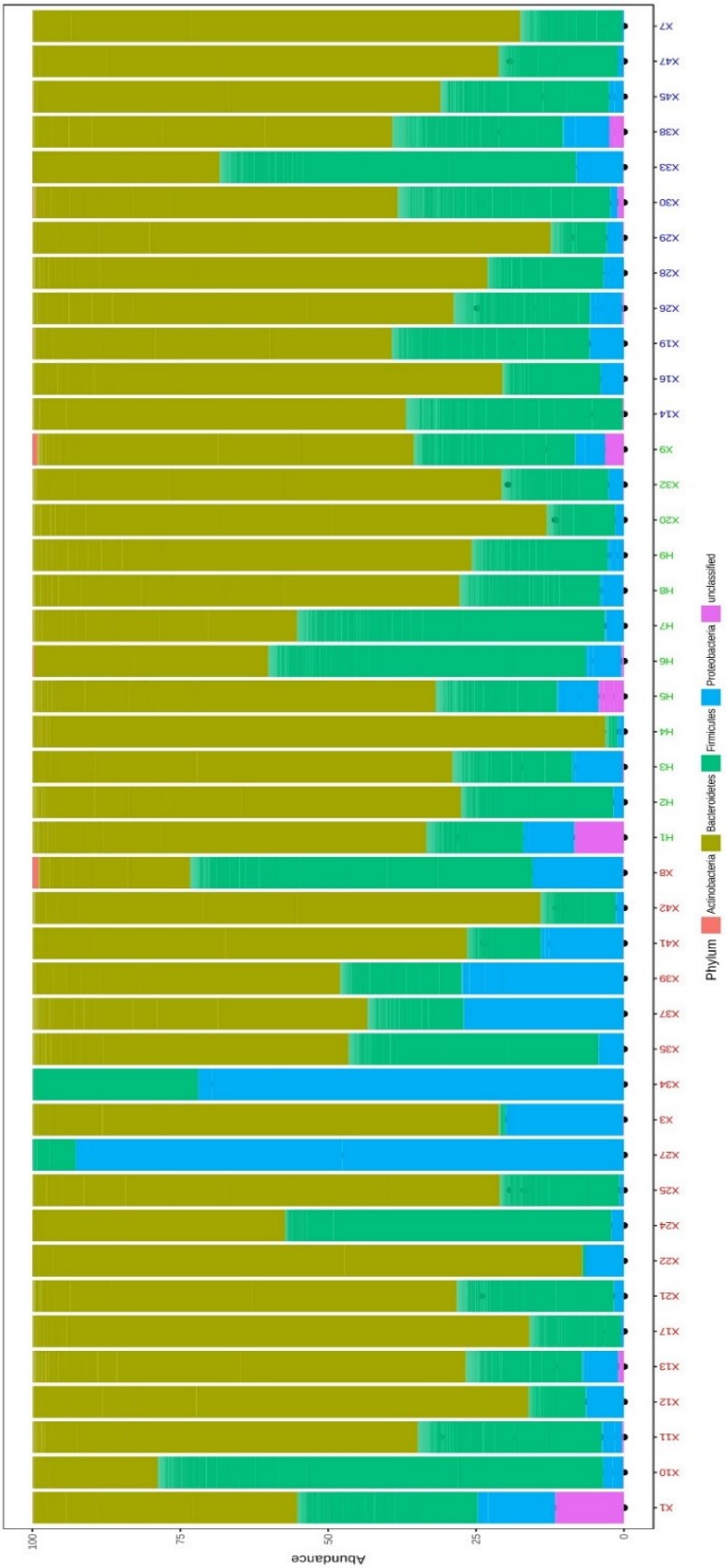


Figure 5.8: Stacked bar plot showing variation in bacterial relative abundances at the phylum level for Crohn's (red labels), Controls (green labels) and UC (blue labels).

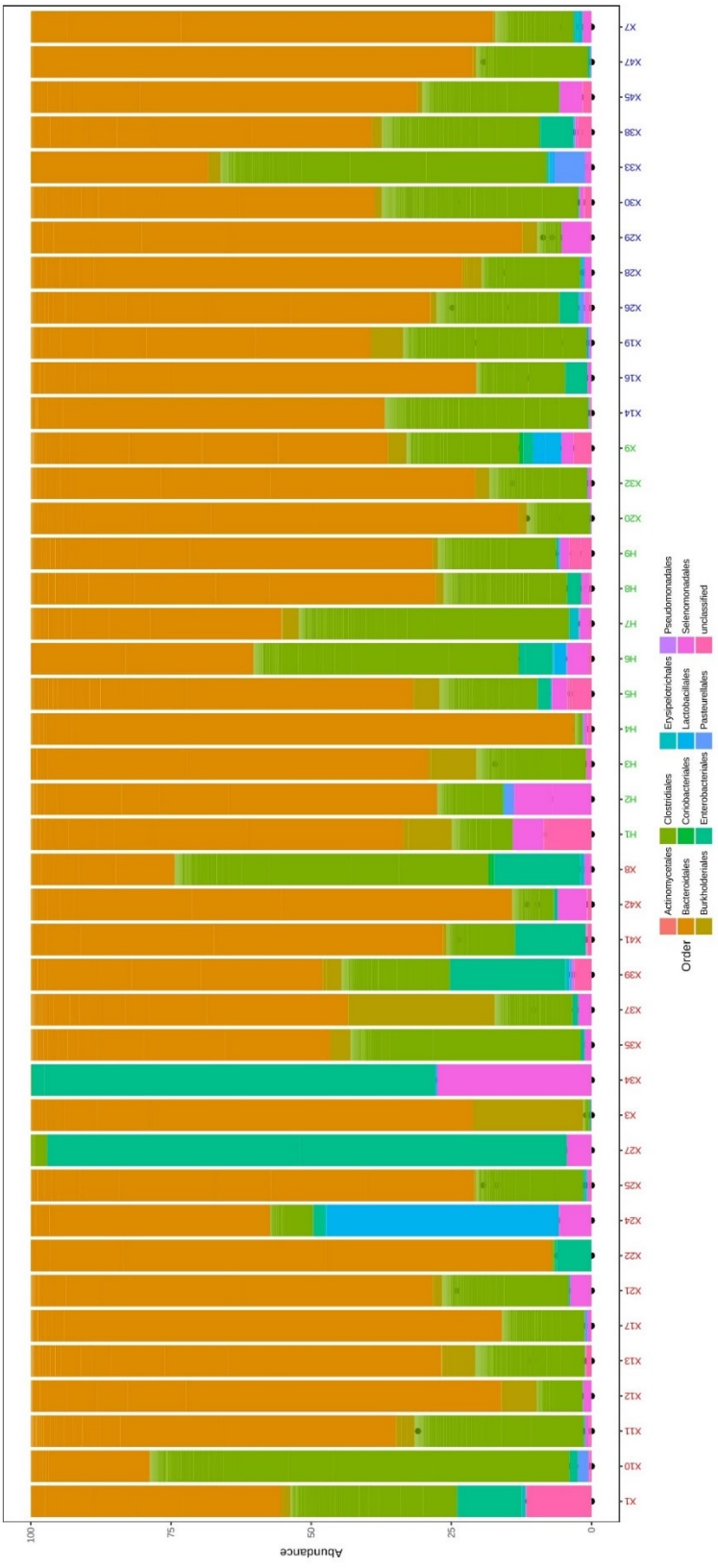


Figure 5.9: Stacked bar chart (percentage abundance) of abundance profile (order). CD (red labels), Controls (green labels) and UC (blue labels).

Figure 5.10: Stacked bar chart (percentage abundance) of abundance profile at family level for Crohn's (red labels), Controls (green labels) and UC (blue labels).

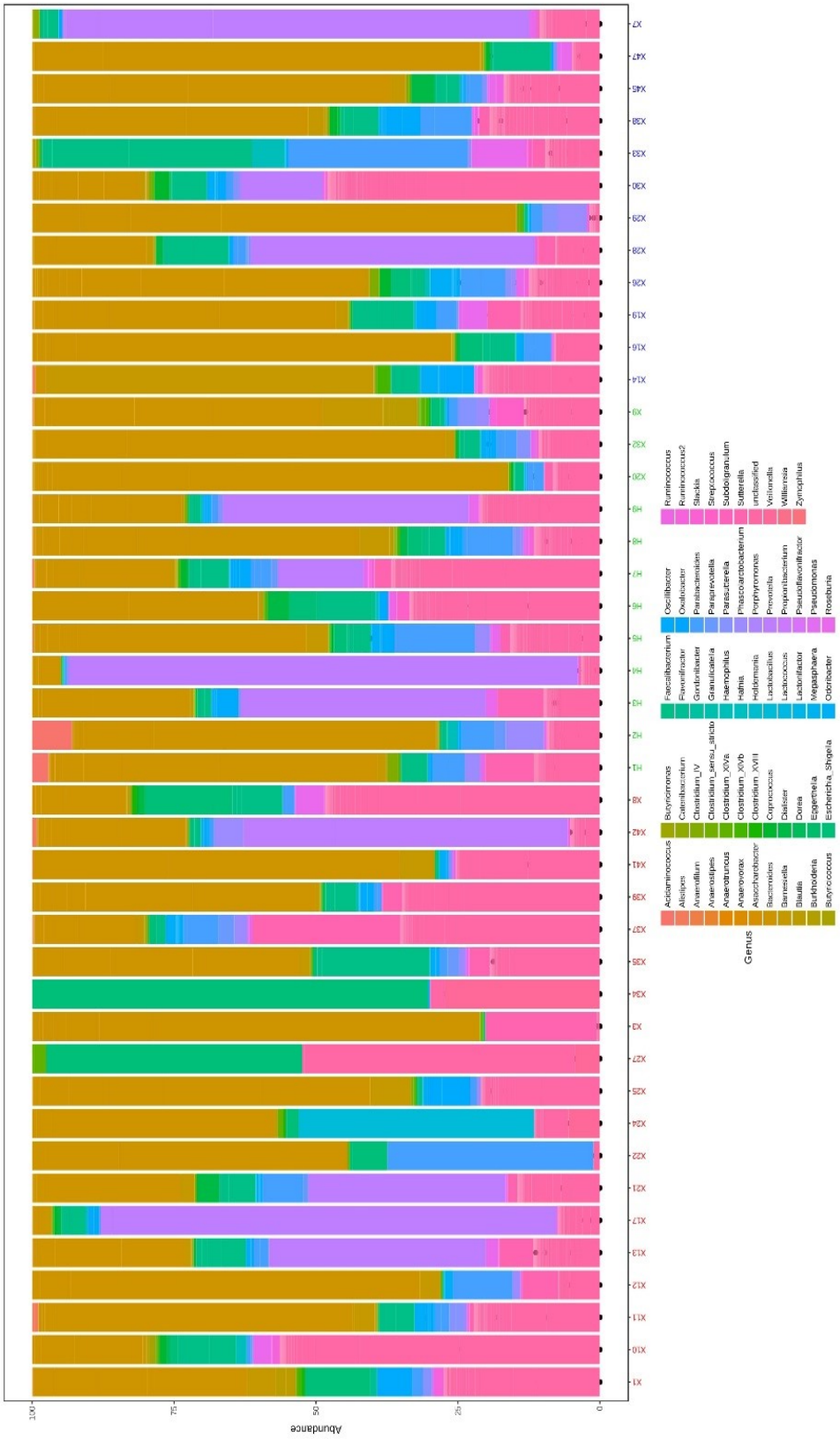


Figure 5.11: Stacked bar chart (percentage abundance) of abundance profile at genus level for Crohn's (red labels), Controls (green labels) and UC (blue labels).

Heatmap visualisation was used to show the relative percentage of 16S rRNA gene sequences assigned to each bacterial genus across the samples analysed. Heatmap colours range between dark red (higher abundance) and dark blue (lower abundance). Hierarchical clustering helps display the relationships between bacteria and samples according to the direct distance between them. There was significant intra-individual variation and inter-individual variation seen on the heatmap.

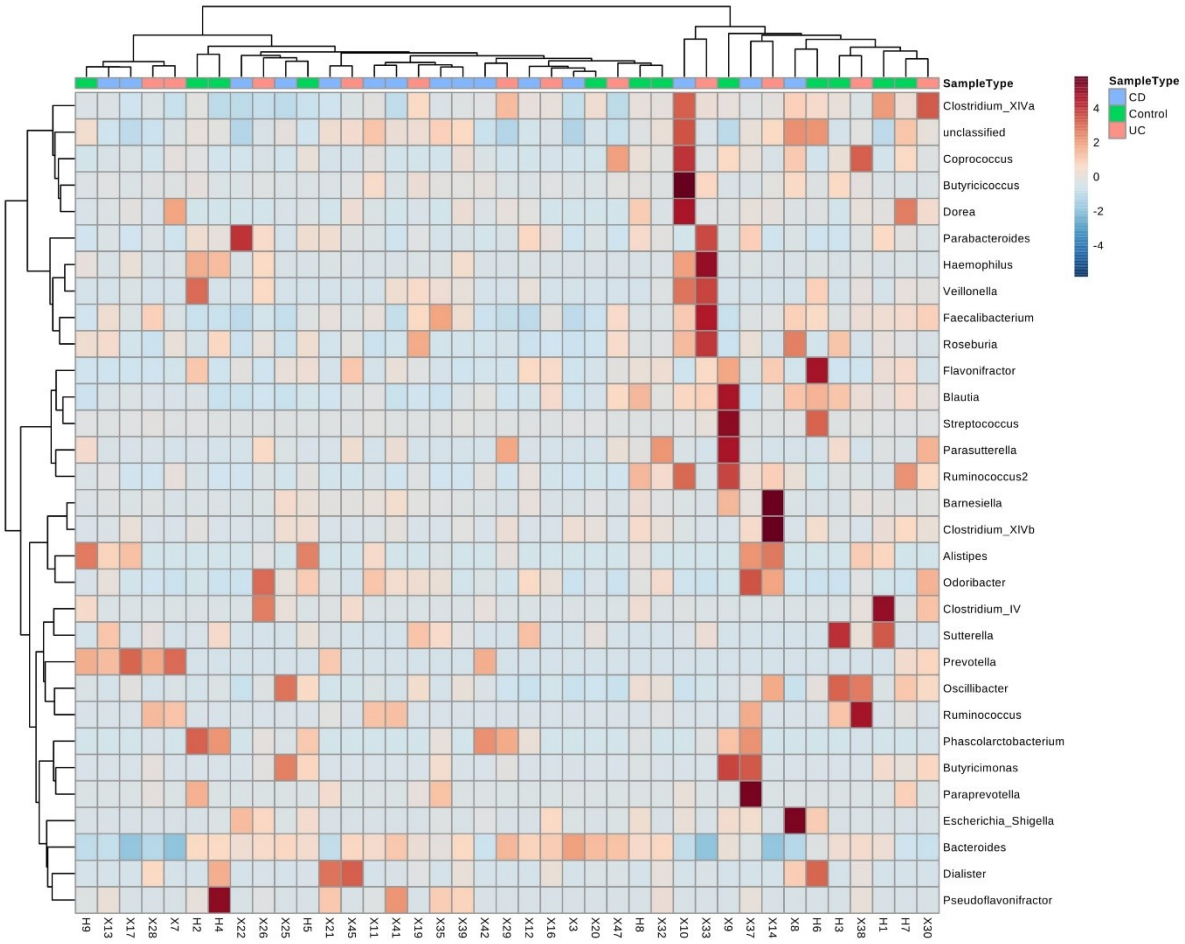
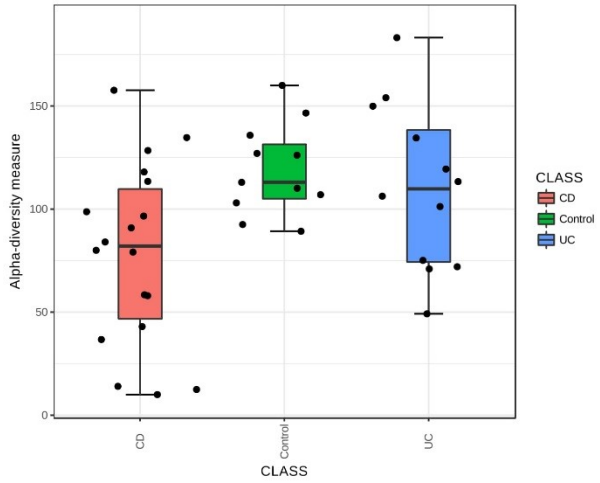


Figure 5.12: Hierarchical clustering and heatmap visualisation across all samples according to their Euclidean (direct) distance.

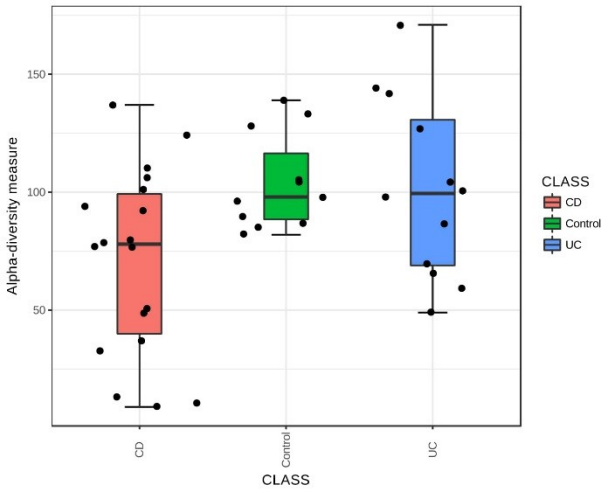
5.5.4.2. Comparison of Alpha diversity

Microbiota diversity within each stool sample (alpha diversity) at OTU level was tested with 4 measures: Observed, Chao-1, Shannon index and Fisher indices. All indices showed lower diversity in the Crohn's compared to UC, compared to controls group, with significant p values in all except the Shannon index.

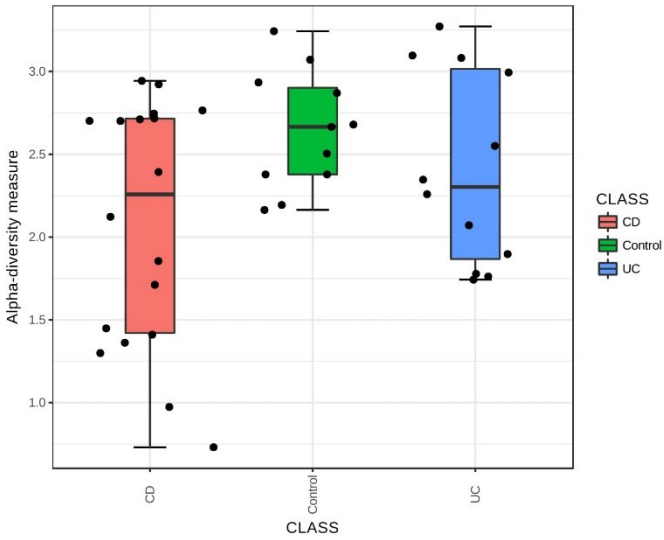
A:



B:



C:



D:

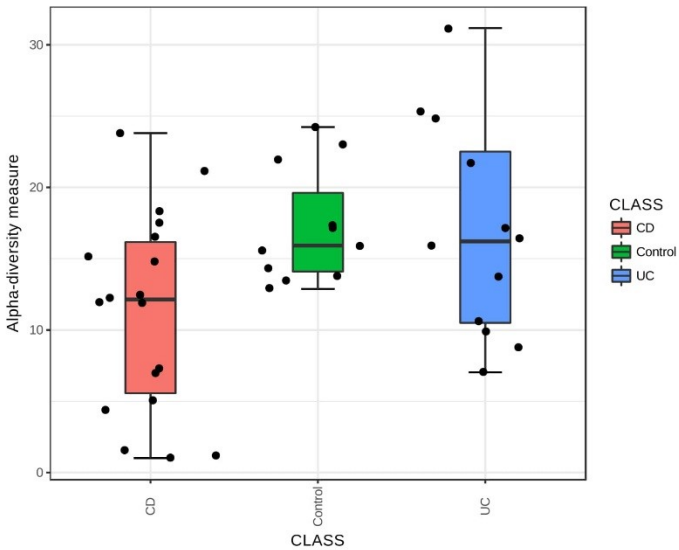


Figure 5.13: Alpha diversity box plots comparing each class for A: Chao-1 richness, B: observed diversity, C: Shannon and D: Fisher Index

Alpha diversity profiling	Significance testing (p value)
Chao 1 index	0.0325*
Observed diversity of OTUs	0.0496*
Shannon Index	0.173
Fisher Index	0.0496*

Table 5.6: Significance testing of 4 indices of alpha diversity between groups. *indicates significant p value ≤ 0.05

5.5.4.3. Beta diversity analysis

Principal coordinate analysis (PCoA) was performed for comparative analysis (beta diversity) between groups. For this low abundance features (which may be due to sequencing errors or low-level contaminations) and low variance features (which are unlikely to be associated with the conditions under study) were filtered out.

PCoA based on unweighted UniFrac distances (R^2 0.068038; p-value < 0.041) showed a moderate dispersal of CD patients away from the cluster of controls and UC patients (figure 5.14). This dissimilarity analysis shows that principal coordinate analysis (PCoA) can discriminate the IBD subtypes from healthy controls. However, there is significant overlap between the three groups as seen in figure 5.15

Weighted UniFrac and Bray-Curtis models had similar R^2 values (0.045 and 0.053 respectively) but non-significant p values.

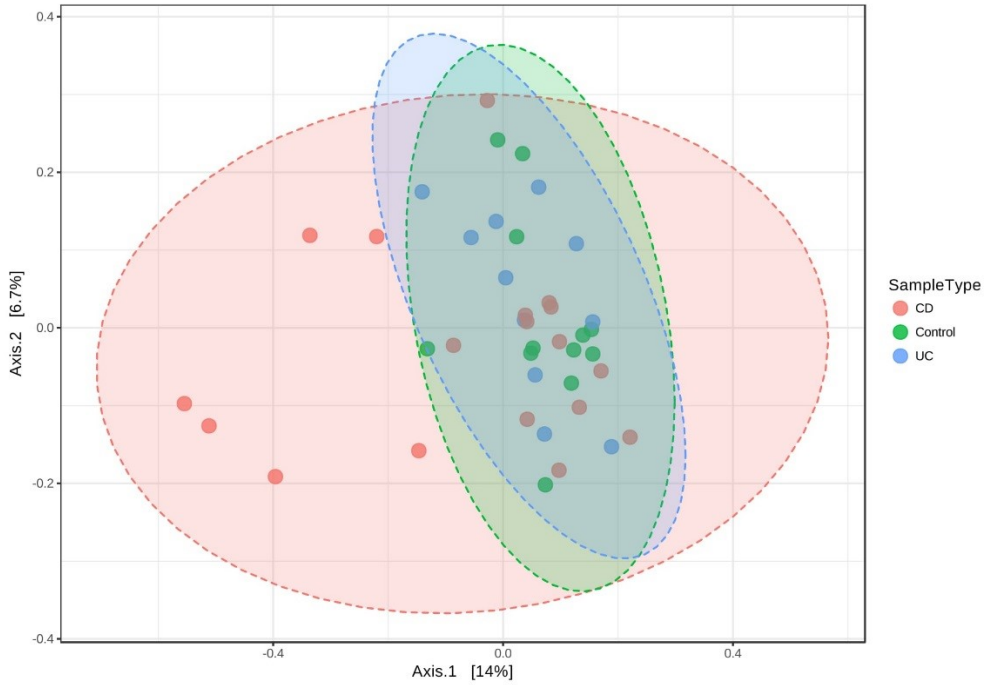


Figure 5.14: Principal Co-ordinate Analysis based of unweighted UniFrac distances, with 2 main axes explaining just over 20% of variance in the model. (R^2 0.068; p-value < 0.04)

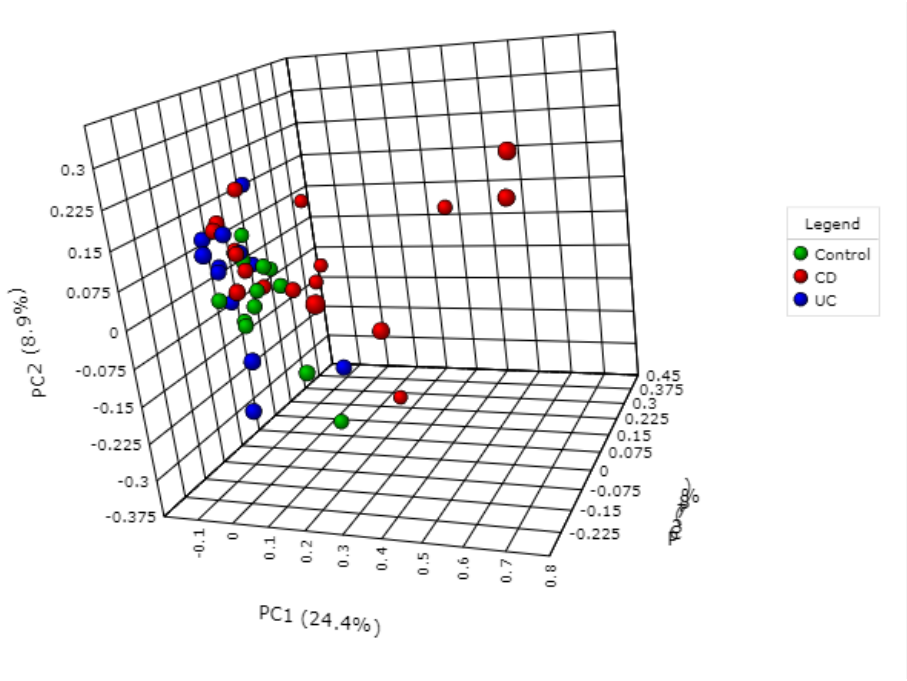


Figure 5.15: 3-D model of PCoA based on unweighted Unifrac distance at OTU level (PERMANOVA 0.064; p-value < 0.058)

5.5.4.3. Differential Abundance analysis

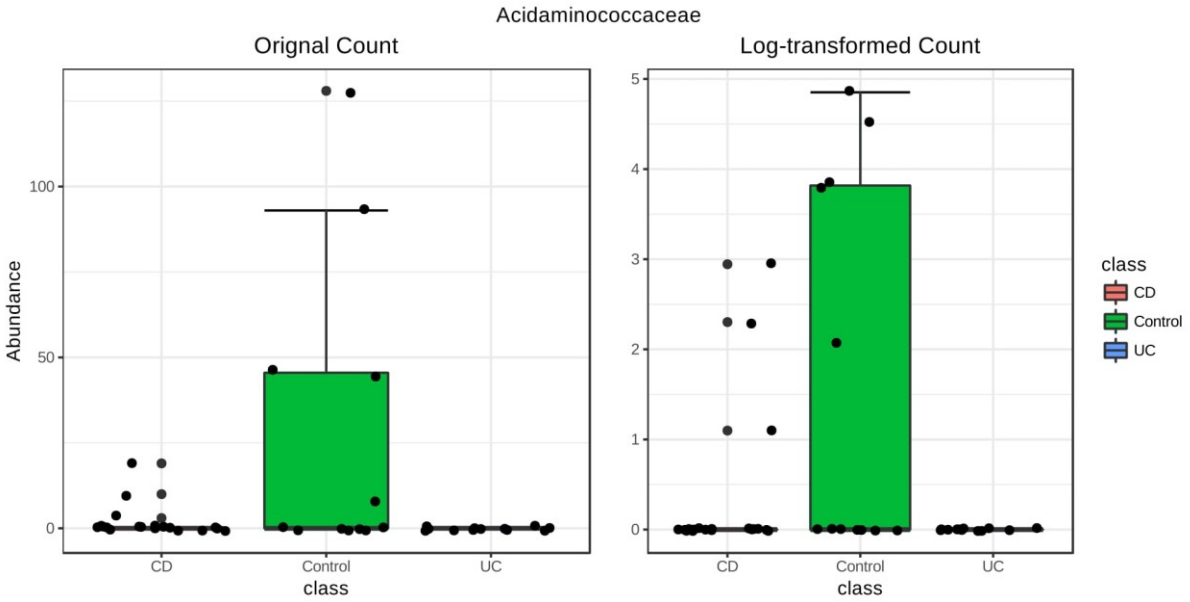
The bioinformatics package EdgeR (218) was used to examine differential expression of counts between groups (by diagnosis) to identify statistically significant bacteria, ranked by p value and adjusted for false discovery rate (FDR).

Significant results between cohorts showed: 28 OTUs identified (not shown), 4 bacteria at genus level taxonomy, 5 at family, 1 by order, 1 by class. The results are shown in table 5.7 and figures 5.16 and 5.17.

Taxonomy	Name	p value	FDR corrected
Class	*Gammaproteobacteria*	2.049×10^{-5}	1.844×10^{-4}
Order	*Enterobacteriales*	1.314×10^{-4}	0.0015
Family	*Streptococcaceae*	1.2169×10^{-5}	2.5578×10^{-4}
Family	*Lactobacillaceae*	2.1315×10^{-5}	2.5578×10^{-4}
Family	*Enterobacteriaceae*	1.4701×10^{-3}	0.011761
Family	*Rikenellaceae*	5.1829×10^{-3}	0.031097
Family	*Sutterellaceae*	9.6839×10^{-3}	0.046483
Genus	*Acidaminococcus*	9.57×10^{-9}	5.17×10^{-7}
Genus	*Lactobacillus*	2.69×10^{-5}	4.0175×10^{-4}
Genus	*Streptococcus*	2.78×10^{-5}	4.0175×10^{-4}
Genus	*Veillonella*	2.98×10^{-5}	4.0175×10^{-4}

Table 5.7: Significant features from each taxonomic group as identified by EdgeR ranked by p value and FDR.

A:



B:

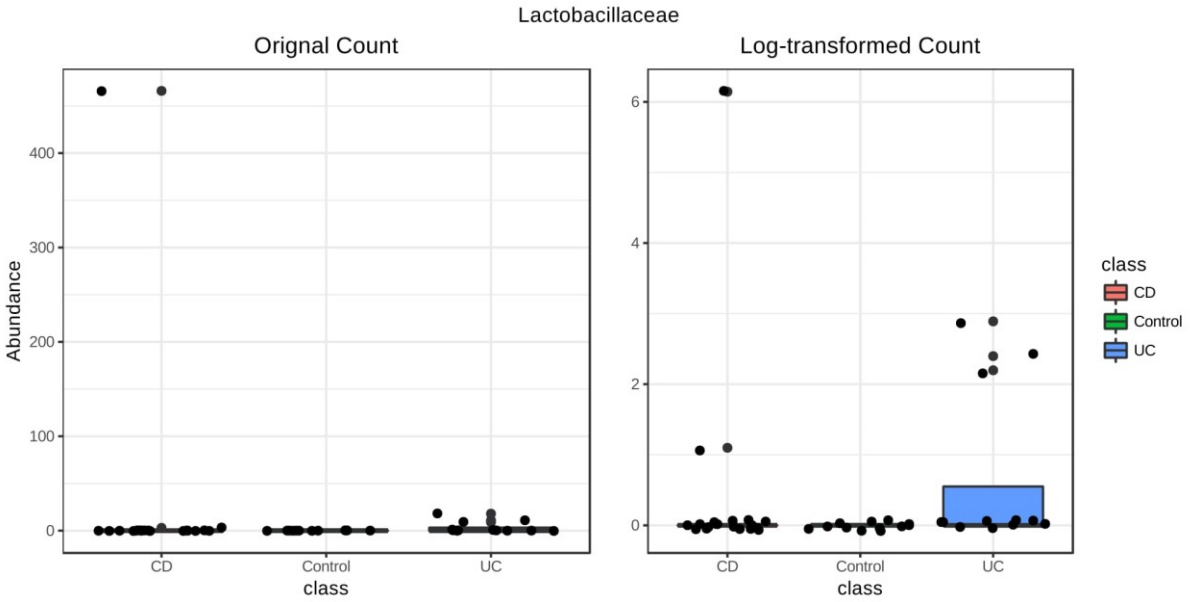
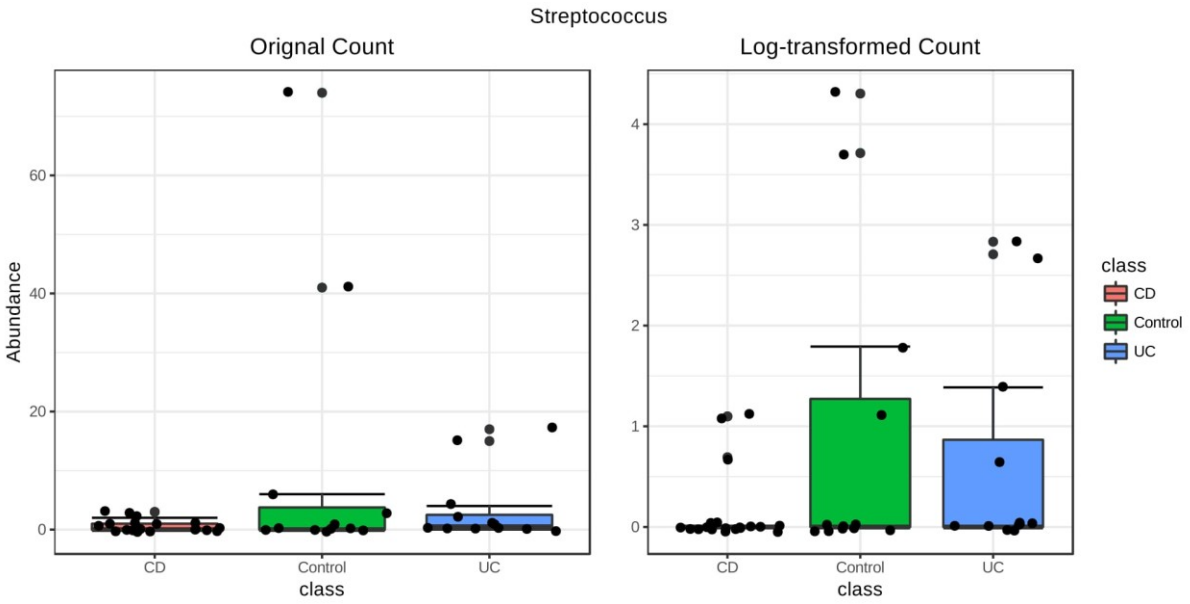


Figure 5.16: Original and log transformed counts of bacteria identified (genus) with statistically significant differential abundance (A: *Acidaminococcus*, B: *Lactobacillus*)

C:



D:

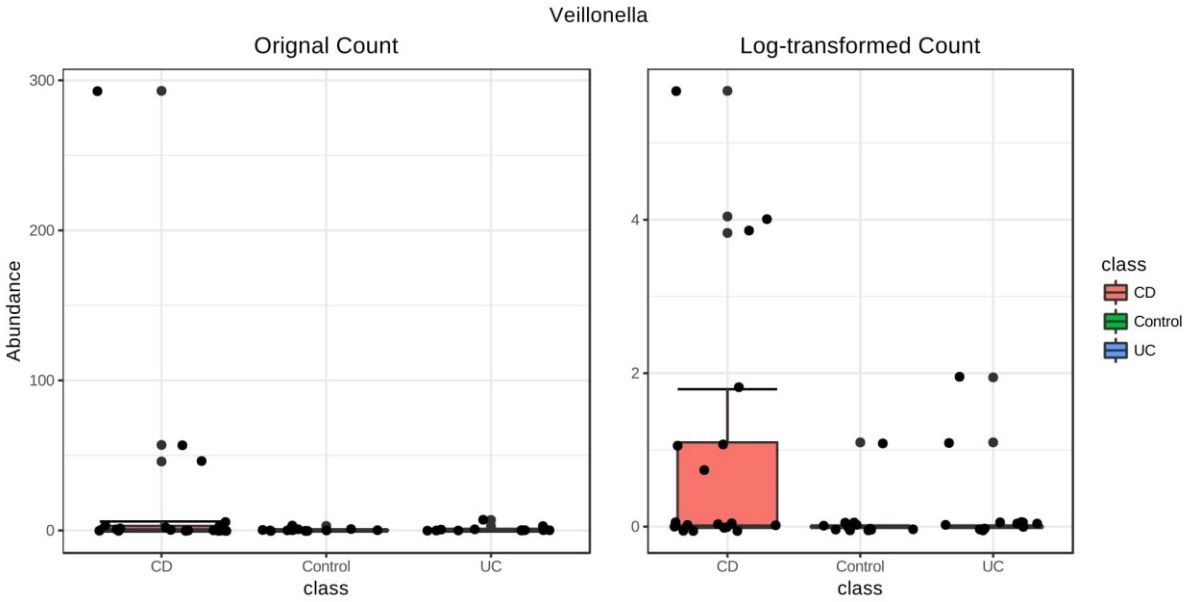


Figure 5.17: Original and log transformed counts of bacteria identified (genus) with statistically significant differential abundance (C: *Streptococcus* and D: *Veillonella*).

5.5.5 Integration of metabonomic and microbial data

5.5.5.1. Urine metabolic profiles and microbial data

The metabolites identified as separating UC, Crohn's patients and controls in the metabolic profiles were correlated with the candidate microbes in a correlogram to identify significant relationships.

Squares within the red box represent relationships between bacteria and metabolites, with statistically significant correlations shown as a pie chart. Colour is used to demonstrate a positive (red) or negative (blue) correlation with the pie chart area being proportional to the value of the correlation coefficient.

Overall in the urine there is an inverse and significant correlation between hippurate and trigonelline and *veillonella* (figure 5.18).

In the Crohn's patient's correlation matrices for urine metabolites show trigonelline is negatively correlated with *veillonella*, and there are positive associations between p-cresol sulphate and *acidaminococcus*, and hydroxyisobutyrate and *veillonella* (figure 5.19).

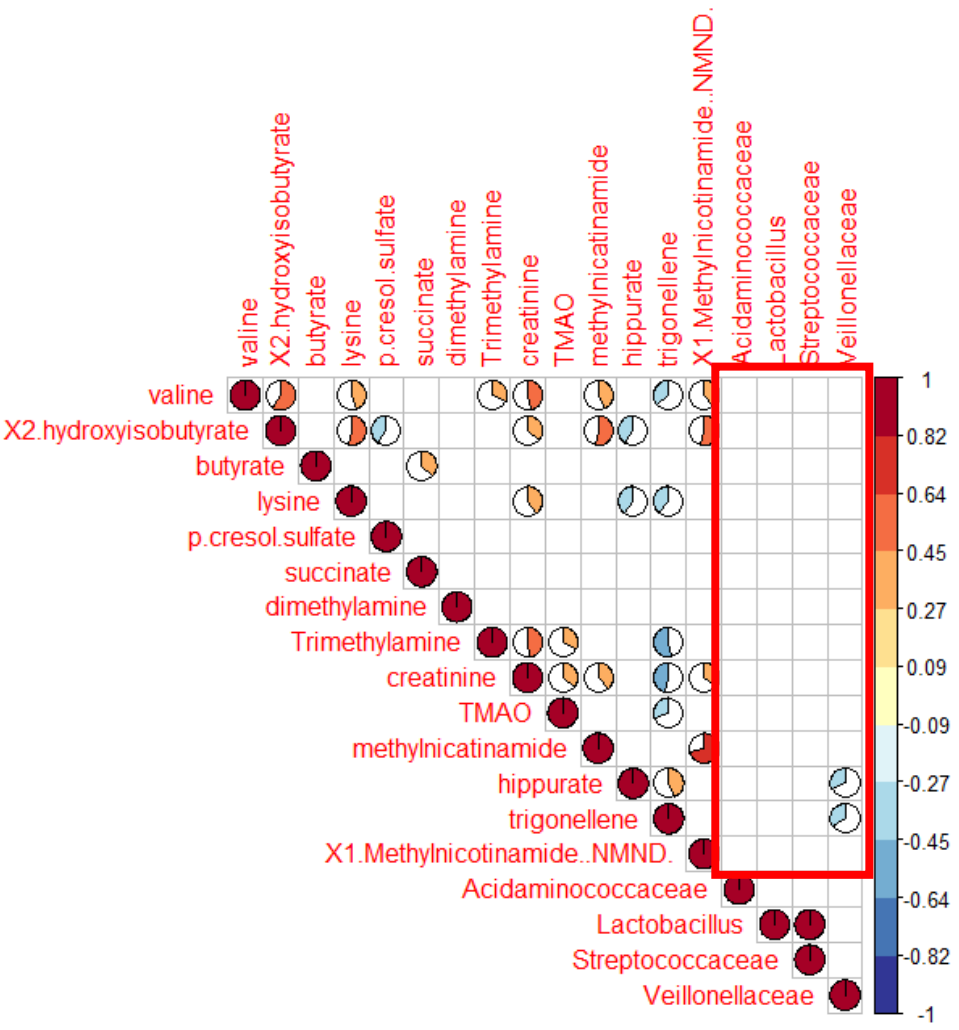


Figure 5.18: Correlogram showing Spearman's correlation between key bacteria (genus) and urine metabolites in all samples. Blank squares mean no significant correlation. Red circles demonstrate a positive correlation, while blue ones designate a negative correlation. In this plot the pie chart area is proportional to the value of the correlation coefficient.

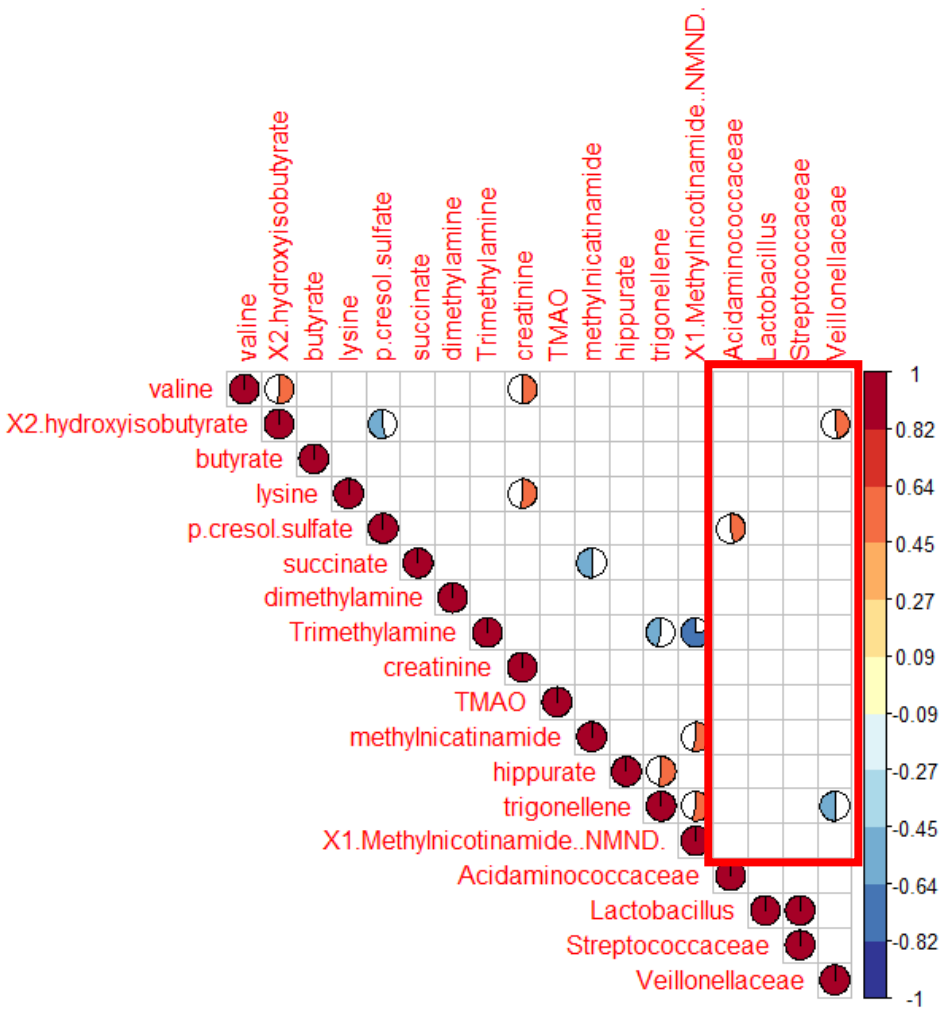


Figure 5.19: Correlogram showing Spearman's correlation between key bacteria (genus) and urine metabolites in Crohn's patients. Blank squares mean no significant correlation. Red circles demonstrate a positive correlation, while blue ones designate a negative correlation. In this plot the pie chart area is proportional to the value of the correlation coefficient.

A: UC

B: Healthy adults

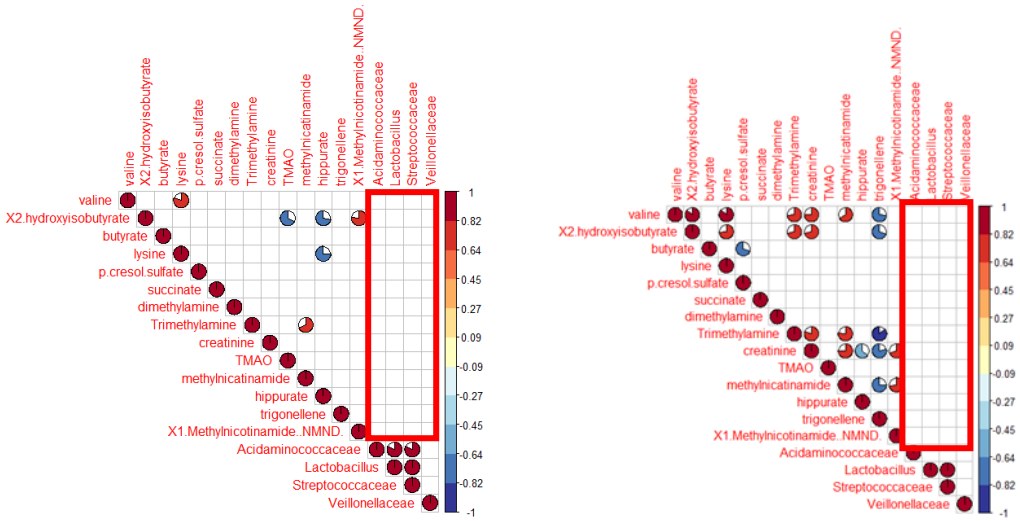


Figure 5.20: Correlogram showing Spearman's correlation between key bacteria (genus) and urine metabolites in A: UC and B: healthy controls.

In UC and healthy adults there are no direct correlations between metabolites and specific bacteria (figure 5.20).

5.5.5.1. Faecal water metabolic profiles and microbial data

As with the urine data, the faecal water marker metabolites were integrated with the significant microbes distinguishing the groups in a correlogram according to the Spearman's coefficients.

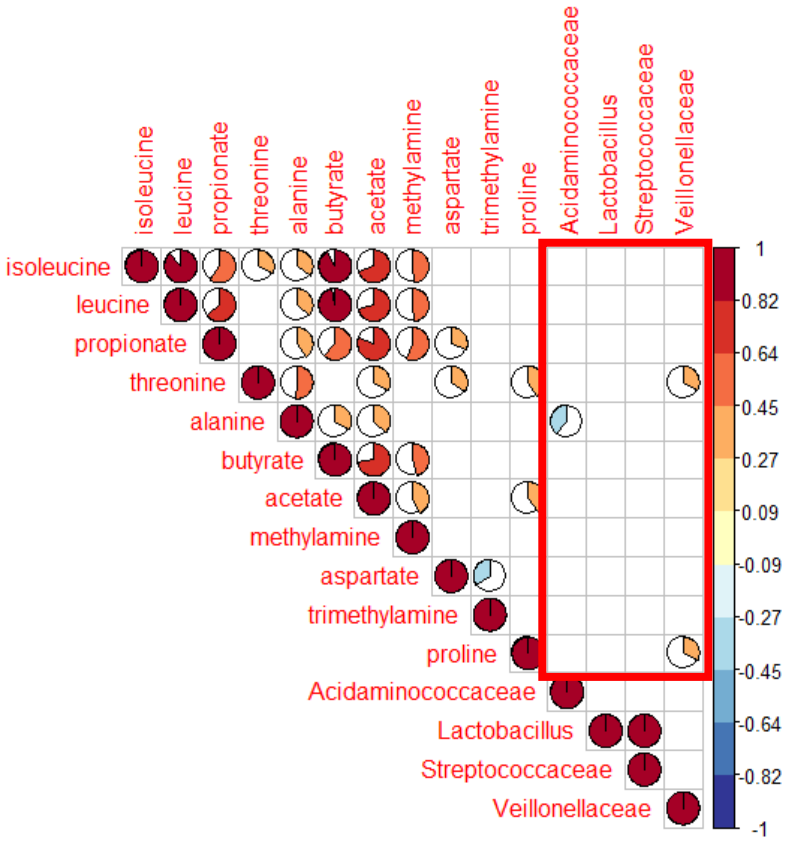


Figure 5.21: Correlogram showing Spearman's correlation between key bacteria (genus) and faecal water metabolites in all samples. Blank squares mean no significant correlation. Red circles demonstrate a positive correlation, while blue ones designate a negative correlation. In this plot the pie chart area is proportional to the value of the correlation coefficient.

In the correlation plot (figure 5.21) there are positive correlations between *veillonella* and both threonine and proline. There is also a negative correlation of alanine with *acidaminococcus*.

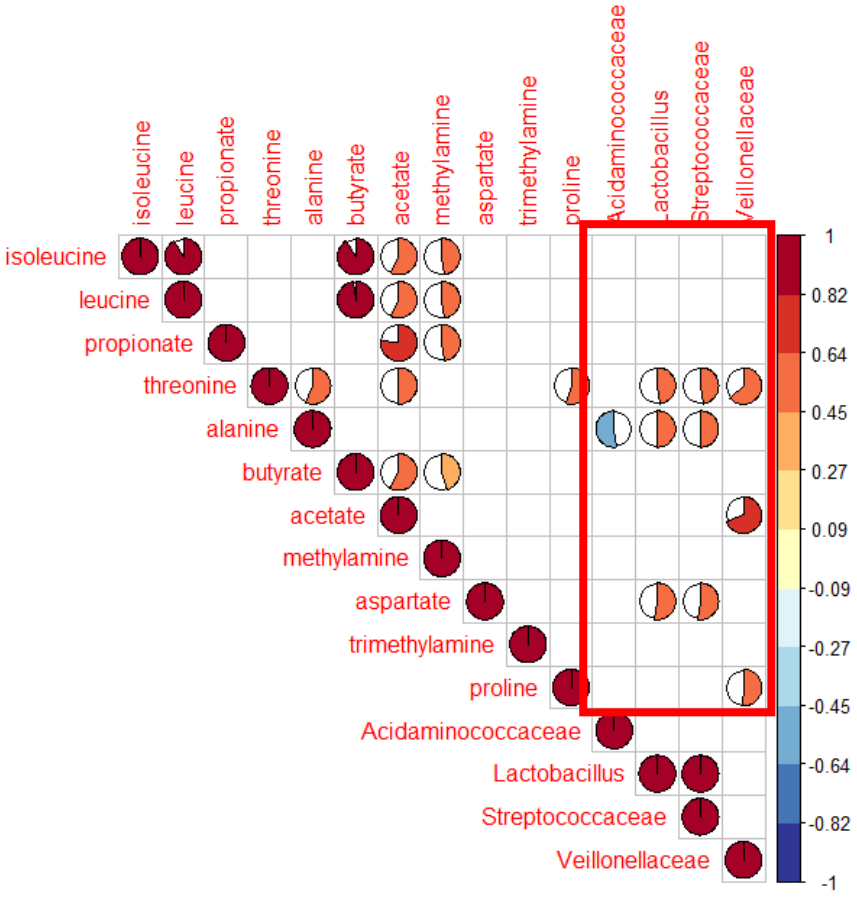


Figure 5.22: Correlogram showing Spearman's correlation between key bacteria (genus) and faecal water metabolites in Crohn's patients. Blank squares mean no significant correlation. Red circles demonstrate a positive correlation, while blue ones designate a negative correlation. In this plot the pie chart area is proportional to the value of the correlation coefficient.

In the stool of the Crohn's patients there are several bacteria positively correlated with threonine, alanine and aspartate (figure 5.22). Acetate is strongly associated with *veillonella*, and proline also correlated with this. There is an inverse correlation with alanine and *acidaminococcus*.

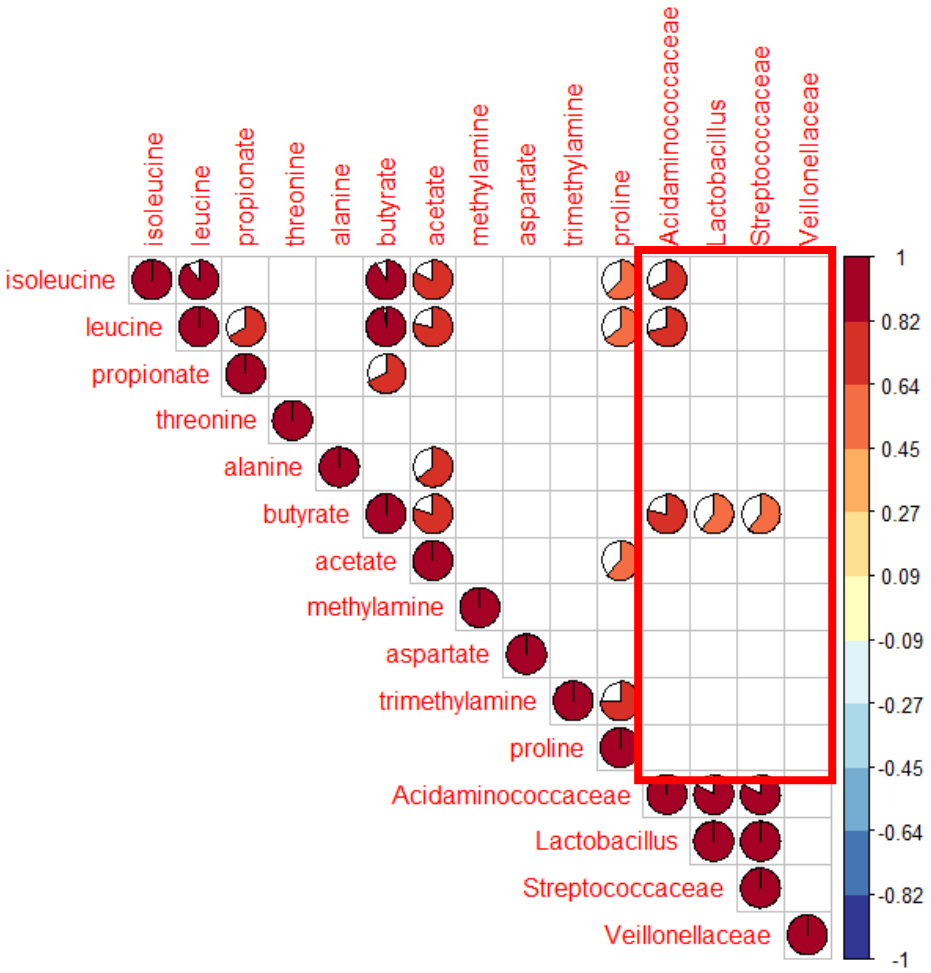


Figure 5.23: Correlogram showing Spearman's correlation between key bacteria (genus) and faecal water metabolites in UC. Blank squares mean no significant correlation. Red circles demonstrate a positive correlation, while blue ones designate a negative correlation. In this plot the pie chart area is proportional to the value of the correlation coefficient.

In UC (figure 5.23), there are positive associations with butyrate and several bacteria (*acidaminococcus*, *lactobacillus* and *streptococcus*) and also isoleucine and leucine with *acidaminococcus*.

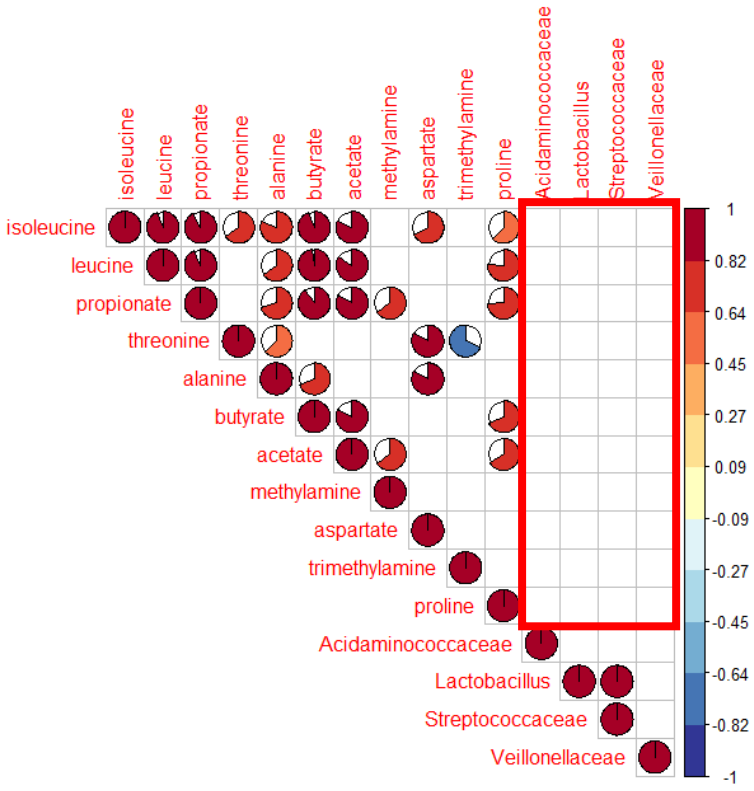


Figure 5.24: Correlogram for faecal metabolites and bacteria in in healthy controls

In healthy adults (figure 5.24) there are no direct correlations between faecal metabolites and specific bacteria, which is confirmatory that the bacteria identified with IBD are normal in the healthy state.

5.6 Discussion

This study used urine and faecal water metabonomics, as well as 16S rRNA bacterial gene analysis to extract and correlate microbial data with metabolic data in IBD patients with a healthy control cohort for comparison. Whilst dysbiosis in IBD is a well-established phenomenon (105), whether the disordered host-microbiome relationship is a pre-requisite for inflammation, or is a consequence of disease, is not clear.

Only one recently published study has also looked at similar correlations in faecal water and bacterial profiling (128) in IBD. In our study the addition of urine, which reflects more host systemic metabolic alterations, was also combined in a three-way omics integration.

Our study is a pilot, demonstrating that important functional correlations can be established by measuring microbial and metabolic disruptions within the same individuals, as stated in the hypothesis.

Although only a small study, findings were consistent with other papers using one analytical technique. 16S microbial data confirmed reduced alpha diversity in IBD, specifically in Crohn's, with enrichment in *Proteobacteria*, as a results of reduced *Bacteroidetes* and *Firmicutes* – the dominant phyla in the healthy stable microbiome. Specific bacteria including *Streptococcus* and *Veillonella* were identified as differentiating between IBD and controls; changes that have been described in other IBD studies (128).

Metabonomic changes in urine mirrored findings in the rest of this thesis, with Crohn's and UC being robustly distinguished from healthy controls. Faecal water metabolic changes were distinct in UC but not valid in Crohn's, likely reflecting the direct effect of colitis on faecal constituents. Published data in faecal water have been less reproducible than urine, although

in this dataset it is possible that separation of the Crohn's cohort from controls was not possible due to methodology (as discussed later) and / or the small group.

The correlations between candidate microbial genera in the different groups and metabolic markers showed directional changes specific to each disease. There were many strong and significant correlations between the microbiome and faecal metabolites in Crohn's and some in UC. Urinary profiles showed less correlations.

5.6.1. Urine metabonomics

As shown in previous chapters, urine metabonomics can demonstrate metabolic differences between healthy adults and patients with CD and UC separately, as well as combined together. Supervised analysis showed robustly tested models could demonstrate separation between groups with reasonable Q^2 values (0.28 for healthy vs CD and 0.43 for healthy vs UC) despite a smaller number of samples/people than previous chapters and other published data.

Obese outliers were removed from analysis as it has been shown that obesity alters both the microbiome and the metabonomic profile (293). All other participants had a BMI of less than 30 (not obese).

UC and CD did not separate on PLS-DA models, similar to previously published data and other smaller studies.

Metabolites identified as separating Crohn's patients from healthy controls corroborate those identified in chapter 3 and other studies, including hippurate, 4-hydroxybutyrate, creatinine and Phenylacetylglycine (PAG).

Other metabolites were also identified as being associated with IBD including methylnicotinamide - a nicotinamide metabolite which has been found to have anti-inflammatory properties (311) and that has been linked to inflammatory processes in the cardiovascular system (312) - as well as dimethylamine and dimethylsulfone, both of which have microbial sources (261).

Some of the same metabolic signals were seen differentiating UC from healthy adults as were identified in CD, including 2-hydroxyisobutyrate, methylynicatinamide, 4-hydroxybutyrate and creatinine.

The two most significant metabolites in the model were 3-Hydroxyphenylacetate (3-HPA) and trimethylamine (TMA). 3-HPA, which is a microbial aromatic compound involved in the tyrosine metabolism pathway, has antioxidant effects (261), and this was higher in the UC group relative to controls. It has been linked to liver inflammation (313) but not to IBD previously.

TMA (lower in UC than controls) is produced in the gut from dietary sources of choline by microbial enzymes (314) (see figure 5.25). Alterations in TMA have been shown to be directly associated with gut dysbiosis, as well as metabolic disorders (diabetes, obesity), cancer and cardiovascular disease (315). Animal models of colonic inflammation (IL-10 deficient mice) also show alterations of urinary TMA (123).

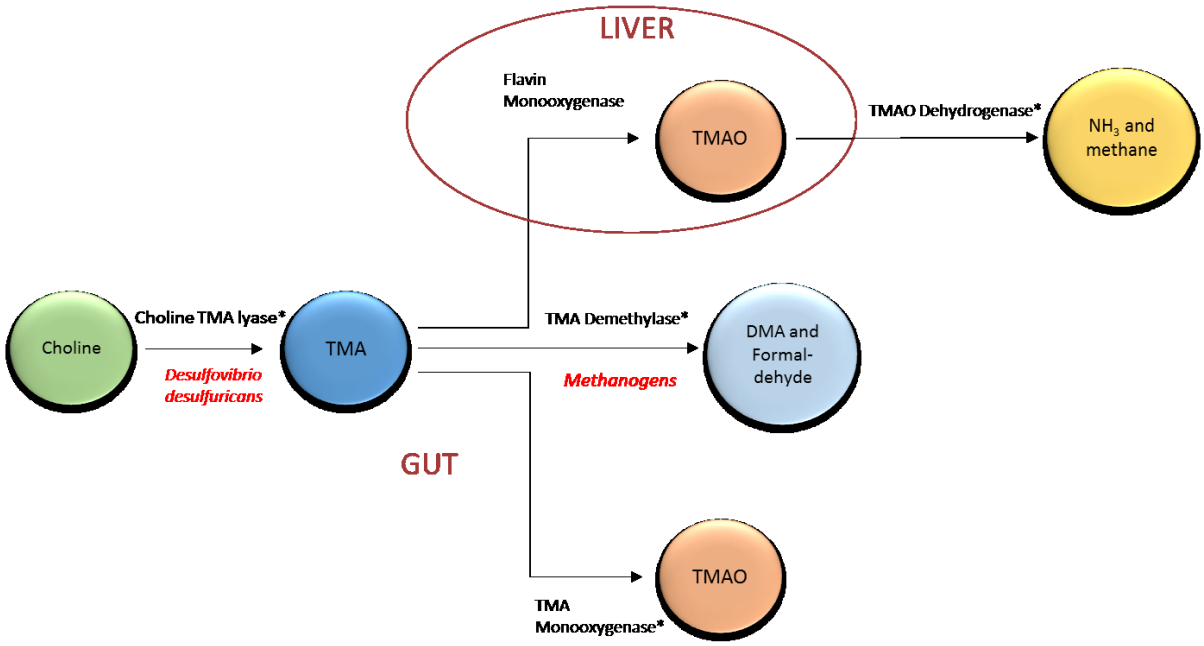


Figure 5.25: TMA pathway in the human GI tract (adapt from (315))

Trimethylamine oxide (TMAO), which is produced via a microbiome–host metabolic pathway from TMA (either by hepatic oxidation or by microbial enzyme TMO monooxygenase) has previously been proposed as a novel IBD-biomarker (316), as plasma levels were found to be lower in IBD. TMAO was also lower in urine of the Caucasian UC group (compared to controls, and compared to CD) in Chapter 3 of this thesis. TMA levels (see below) in faecal water are also shown to be altered in IBD in this study and others (117), suggesting that the TMA metabolic pathway can be affected at several different levels by microbial perturbations.

5.6.2. Faecal metabonomics

Only UC could be differentiated from controls with faecal water metabolic profiling, with multivariate analysis failing to be able to robustly separate Crohn's from healthy samples or from UC. This may suggest that the colonic inflammation of UC has a greater effect on the faecal metabolic profile, whilst the reverse is seen in urinary profiles, where Crohn's which is

a more systemic inflammatory condition, exerts a stronger effect on urine metabolic profiles than UC.

In this population the Crohn's cohort contained a mix of small bowel and colonic disease (disease location L1:8 patients, L2:8, L3:4). Larger studies allowing analysis by disease location may determine whether the changes seen in UC are disease specific or relate to colonic inflammation more generally.

Another reason for the inability to separate the Crohn's group may be the removal of parts of the spectrum in the aromatic region to avoid the xenometabolites in this region; an area where potentially valuable metabolic information exists. This method (117) had to be performed on the faecal water data as the STOCSYE (224) software could not remove resonances from 5-ASA adequately to prevent them influencing the models.

In the Bjerrum paper (113) the PLS-DA models of both active and inactive CD vs controls were initially valid, but once patients post-intestinal surgery were removed from the models, the Crohn's models became invalid, and only active UC could be differentiated from healthy. As surgery was an exclusion criteria in this study, the results in our Crohn's cohort are not dissimilar to these published data.

Whilst Marchesi found a strong faecal signal with Crohn's (117), the phenotypic data of these patients (disease activity, location, behaviour or surgery) is not known. More recently, in a larger Italian study, investigators were also able to distinguish Crohn's from controls with faecal water metabonomics, suggesting that cohort size was likely to be the most limiting factor of our findings.

Metabolites associated with UC included propionate (most significant metabolite), butyrate and acetate, which are the main short chain fatty acids (SCFAs) produced by gut microbes

during fermentation of complex carbohydrates (126). All of these were lower in the UC group than controls, which is entirely consistent with other published data (113, 117, 126). *Clostridium coccoides* and *Clostridium leptum* groups have a principal role in the production of SCFAs (132) and these species have been found to be reduced in UC (317, 318). The SCFAs, particularly butyrate, are an essential energy source for colonocytes (319), and SCFAs have anti-inflammatory properties via inhibition of aTNF activation in the NFkappaB pathway (98, 130). It has also been shown that active inflammation in UC is associated with impaired butyrate oxidation (320), suggesting that SFCA depletion is both a consequence of inflammatory dysbiosis and a promoter of on-going inflammation (113).

This study did not examine the metabolome at species level, but given these findings future studies should attempt to associate the reduction in SCFAs with changes in *Clostridium coccoides* and *Clostridium leptum* directly.

Methylamine and TMA also strongly contributed to separation in UC; this was also seen in the Marchesi study (117) and the Santoru study (128). These mammalian-microbial co-metabolites are likely to have been altered by dysbiosis.

5.6.3. Bacterial community identification

Phylum level analysis of the microbial community in this cohort revealed a lower percentage of *Bacteroidetes* and *Firmicutes* in Crohn's patients than in UC or controls. The same findings have been seen in several studies (72, 321).

Enrichment of *Proteobacteria* was seen in the Crohn's group (13%) versus healthy controls (6%) and UC patients (3%). Whilst *Bacteroidetes* and *Firmicutes* dominate the gut microbiome in most healthy adults, in dysbiosis the loss of homeostasis gives *Proteobacteria* the opportunity to increase, and is therefore seen as a marker of microbiome instability (322).

Increased relative percentages of *Proteobacteria* (from approx. 4.5% in healthy adults) (322) have been shown in metabolic disorders (13.2%) (323) and in IBD (14.9%) (93). Other studies have proposed that *Proteobacteria* may have an initiator role in the inflammatory response and propagation of IBD (324).

Diversity analysis revealed the lowest alpha diversity indices in Crohn's patients, compared to UC and healthy adults. All 4 indices, accounting for richness and evenness were lower (only Shannon index had a non-significant p-value) across the IBD groups, as seen consistently in studies (91, 325, 326).

Dysbiosis is the most substantial marker of IBD in microbial studies thus far, although it is unclear whether it is a reflection of an abnormal immune response and inflammation, or the trigger that starts these (308). Most likely there is overlap between these two events occurring in a genetically susceptible individual, with multiple microbial drivers and bacteria-host-immune interactions simultaneously.

Principal coordinate analysis (PCoA) was used to demonstrate dissimilarity (separation) between the groups. And whilst unweighted UniFrac distances could statistically model separation, the R value was low (R^2 0.068; p-value < 0.041) and there was significant overlap between the groups. This is to be expected in human data and suggests small but measurable differences between bacterial communities. Although the groups in this study are small, it does demonstrate a relative difference in microbial composition when comparing the groups.

5.6.4. Specific bacteria differing in IBD

Although a multitude of studies have profiled the IBD microbiome, there are inconsistencies in the ages, phenotypes, ethnicities and treatments of the cohorts examined. Whilst dysbiosis is almost universally seen, specific microbial changes are less demonstrable, and no particular

recipe of microbial constituents has been linked to either diagnosis or shown to predict outcomes (72). Although it is unlikely that one, or a few microbial candidates are responsible, some particular bacterial perturbations have been repeatedly linked with IBD; specifically a reduction in *Faecalibacterum prausnitzii* in ileal CD, and an increase in Adherent–invasive *Escherichia coli* (AIEC).

Faecalibacterum prausnitzii (from the Clostridia class of Firmicutes) is a butyrate producing species with known anti-inflammatory properties (327). In this study there was no statistical difference in *Faecalibacterum* (genera level) measured between groups, although only 60% of the Crohn's cohort had ileal involvement, and species level classification was not available. Although overall *Proteobacter* was higher in the Crohn's patients, no difference was seen in E-coli when comparing across groups.

Specific bacteria identified at the genus level as differentiating between cohorts in this study were *Veillonella, Acidaminococcus, Lactobacillus* and *Streptococcus*.

5.6.4.1. Veillonella

Veillonella (Firmicutes) was raised in the stool of Crohn's patients compared to UC and controls. The same finding was seen in the Santoru study (128). This bacteria has been shown to be related to recurrent inflammation in CD post-operatively (328). *Veillonellaceae* are lactate-fermenting bacteria, and have also been found to be enriched in CD (regardless of location) in several paediatric studies (329, 330).

5.6.4.2. Acidaminococcus

Metagenomic studies have been mostly consistent in demonstrating a reduction in *Firmicutes* (except *Lactobacillus*) in stool samples in IBD, particularly Crohn's (321) In this study

Acidaminococcaceae (gram negative, *Firmicute* phylum) were profoundly reduced in both the CD and UC cohorts compared to controls. This genus has previously been shown to be associated with a dysbiosis seen in type 1 diabetes in a paediatric cohort (331), although not been shown to be specifically identified in IBD previously.

5.6.4.3. Lactobacillus

In this cohort *Lactobacilli* (Firmicutes phylum) were enriched in the UC cohort significantly compared to CD and controls. In one published study, over 11 different types of *lactobacillus* species were shown to be altered in UC compared to controls (332) (some relatively increased, others decreased). *Lactobacillus* was also found be shown to be higher Crohn's samples in a different study (82) using mucosal samples.

Probiotics, including *Lactobacillus* species have been postulated as a treatment for IBD (103). In a mouse model administration of one species (*L. crispatus*) was found to aggravate chemically induced colitis, whilst another species ameliorated it (*L. fermentum*). From 16S rRNA sequencing it is not possible to classify to a species level and so it is not clear which *Lactobacillus* was higher in the UC group in this study.

5.6.4.4. Streptococcus

Similarly, *Streptococcus* (also Firmicutes phylum) was significantly lower in CD in this cohort. The recent Italian paper also highlighted this bacteria as a significant IBD microbiota biomarker (128). Contradictory to this finding, a paediatric study showed a gain in *Streptococcus* taxa in treatment naive paediatric Crohn's patients, (330) but this has not been reported in adults.

5.6.5 Integration of data

The statistically significant microbial genera, as shown by differential abundance, and the highly discriminate metabolites from each biofluid (r value with $p<0.05$) were correlated and ranked according to Spearman's analysis.

Overall there are seen to be more significant, and stronger correlations between bacteria and faecal metabolites than urinary ones, as may be expected. In the stool samples in Crohn's there were 9 positive associations, and one negative association between bacteria and metabolites. The strongest correlation was with acetate and *Veillonellaceae*, the only negative was with *Acidaminococcaceae* and alanine. The Spearman correlation coefficient was less strong for UC than for Crohn's, which was also the finding seen by Santoru et al (128). In faecal water samples of UC patients *Acidaminococcus* was the strongest influence with 3 positive metabolic associations, and *Lactobacillus* and *Streptococcus* had one each. This demonstrates clearly that gut bacteria are directly altering the metabolic constituents of the stool.

In urine there were less metabolic relationships identified, although *Veillonellaceae* was positively linked to hydroxyisobutyrate and inversely related to trigonelline in Crohn's. *Acidaminococcus* was negatively associated with alanine in the same group.

In healthy adults there are no direct correlations of either urinary or faecal metabolites with the bacteria, this is clearly because the specified bacteria are the ones relating to disease in this group.

This study tested the feasibility of integrating cross-sectional omics data in a cohort of IBD patients and controls, to demonstrate patterns and relationships between the gut microbiota and metabolic markers in urine and stool. These results suggest that a larger study with a similar integrated design will reveal more about microbial-host-metabolic mechanisms and drive

forward the understanding of the role of the gut microbiome in pathogenesis, progression and treatment of IBD.

Without further correlating the function of the microbiome with metabolic products, the significance of dysbiosis in IBD cannot be properly understood. Multi-omic approaches can offer more than metabolic profiling alone, to help draw out meaningful and clinically relevant information. The natural development of these studies is to further incorporate proteomics and genomics to demonstrate functional relationships within an individual, with a high degree of confidence.

This study successfully demonstrates that integrating these omic technologies is an important step forward in robustly associating specific bacteria with IBD related metabolic perturbations.

5.6.6. Study design and Limitations

This was a pilot study to test the feasibility and value of combining multi-omic datasets in IBD. Although this study has demonstrated methodology to combine metabolic and microbial signals in a well phenotyped group of patients, the major limiting factor is the small group sizes. It is a novel study design, integrating 3 different datasets to establish direct links between bacteria and the metabolic state of the patients, but ideally more samples would improve the power of the study.

Although 16S rRNA gene sequencing is a powerful method of measuring the bacterial composition of the microbiome to a genus level, it lacks the specificity to accurately describe species, because at this level the read sequences are nearly identical to other bacteria in the reference database (333). Further to this, there may be multiple strains of each bacterial species,

each with separate metabolic functions (333) which cannot be described by 16S. Whilst this method can be used to infer the metabolic capacity of a microbial community by correlating phylogenetic trees and clusters of genes shared between taxa, it remains inference not measurement. Other authors have done this type of pathways analysis (128) but more work is needed with these bioinformatic tools to improve the accuracy of enrichment analyses (334).

In contrast, metagenomics (high through-put or shot-gun sequencing) examines the entire genetic material in a sample, and if sufficient sequencing depth is applied, also holds the potential to directly assess metabolic microbial potential and function (335). Additionally high through-put or shot-gun sequencing also identifies non-bacterial organisms (archaea, viruses, virophages, and eukaryotes) which have also been proposed to contribute to host immune response regulation (336). The pay-off for this superior taxonomical and functional resolution of metagenomics however, is time and cost.

The challenge of discovering correlations between metabonomics and gut bacteria remains difficult when co-handling two (or three) large sets of multivariate biological data (337). As such there are no benchmarked and validated computational approaches, especially for metabonomics (334). In this study the data has been analysed for co-occurrence detection, and biological associations have been shown between significant bacteria and metabolites. Alternative approaches used with metagenomics data and mathematical modelling are network inference tools such as SparCC or CoNet (338), however these have not be verified for metabonomic data (337). It may be possible to develop these network construction tools for integrating metabolic and bacterial data, which may capture more biological correlations.

As described earlier the microbiome composition is also affected by multiple external factors, to a lesser or greater extent according the individual. These include (amongst others)

medication, age, smoking, environment, antibiotics, surgery (93) and the effects on the microbiome of each of these is variable and immeasurable (321). To account as fully as possible for this, accurate phenotypic information was taken from patients, although as a real life cohort, the groups were heterogeneous in some aspects, for example an older UC cohort.

Additionally in this study the stool microbiome was measured. This was chosen because it is non-invasively obtained, easily handled and did not require bowel cleansers for sample collection at colonoscopy (which may affect the microbiome). However there are data recognising that the mucosal-associated microbiome (MAM) is not the same as the faecal-associated microbiome (FAM) and that MAM may be altered more by active disease (79). It is still worth studying the FAM because this is an easier biofluid to obtain from patients and therefore one that is more amenable to a larger, future study, or for development as a potential biomarker.

The above factors may affect the similarity between results in this study and previously published work.

When studying metabonomics the usual caveats must also be considered, including external influences and confounders on the metabolic signature such as diet, medication and co-morbidities (as discussed in chapter 3). However, for the first time the metabolic data extracted from 2 different biofluids has been integrated with microbial data, and relationships between the two established. This is a helpful first step in trying to elucidate more about the gut-microbial-host interactions, but it remains a causal link, and further mechanistic investigations are required.

Future work should include a much larger study with sub-populations of well-characterised IBD including active and inactive disease (with corresponding faecal calprotectin). It would be

valuable to study faecal metagenomics to further explore the microbiome beyond just bacteria, and to extract functional data in order to better establish metabolic-microbiome relationships in IBD.

5.7 Advances in Knowledge

- Relationships between gut bacteria and candidate metabolic molecules in biofluids can be robustly established with omic integration, demonstrating a direct link between dysbiosis and metabolic profiles in IBD.
- Bacteria appear to have more and stronger metabolic correlations in stool than in urine; and there are more seen in Crohn's than UC, as other studies have also shown.

Chapter 6: Exhaled breath VOC analysis as a novel metabolic profiling technique in IBD

6.1 Summary

Volatile organic compounds from the exhaled breath of IBD patients and controls were analysed using SIFT-MS. Here multivariate profiling was employed to demonstrate distinct separation of groups according to diagnosis, and concentrations of individual VOCs were compared. The VOCs which were statistically significantly different between the groups were used to construct a ROC with integrated areas under the curve ranging between 0.74 and 0.86.

The discriminatory compounds identified are associated with bacterial dysbiosis and oxidative stress, both mechanisms implicated in the pathogenesis of IBD.

SIFT-MS breath profiling is a novel and developing technology which will enrich our understanding of the volatilome as part of the metabolic signature of an individual, and may become a useful adjunct in IBD clinical research.

6.2 Aims and hypothesis

1. The aims were to identify and analyse the exhaled breath VOCs in IBD patients and controls with SIFT-MS; and to compare VOC profiles between IBD subtypes, and between patients and controls using multivariate and univariate analytical techniques.
2. The hypothesis was that exhaled breath VOC profiles differ in IBD and can be used to discriminate patients from controls.

6.3 Introduction

The exploration of novel technologies may help to address some of the unanswered questions in IBD and shed light on pathophysiological pathways in this disease. Breath analysis has related VOCs to gastrointestinal pathologies including oesophago-gastric and colorectal cancers (163, 339) as well as the pathological processes of dysbiosis and inflammation (162).

SIFT-MS has previously demonstrated measurable breath patterns in coeliac disease (340) and upper gastrointestinal cancers (163) and this study was to test the hypothesis that breath profiles would be different in IBD patients when compared to controls using this technology. It was hypothesised that the different gut bacteria in IBD, as well as inflammatory change, would contribute to alterations in breath VOCs.

As discussed earlier in this thesis, breath analysis in IBD has previously been limited to examination of single or a few compounds (170, 172), showing differences in individual VOCs but not in overall breath profiles. At the time this work was published, the only multi-compound study to date had been in a paediatric population (177), and whilst this showed discriminatory differences between IBD and controls, the cohort was unfasted, and the potentially huge impact of diet on the results was not addressed.

This study was a pilot of exhaled breath analysis in adults with well characterised IBD, and a matched control group. The objective was to demonstrate that alterations in breath VOC concentrations can be used to measure pathological processes in IBD, and also to investigate the origin of these specific compounds and their relationship to pathogenesis.

6.4. Methods

6.4.1 Subjects

38 patients with IBD were recruited at St Mary's Hospital (Imperial NHS Healthcare Trust) who had an established diagnosis of Crohn's disease or ulcerative colitis based on clinical, endoscopic and histological evidence . Disease activity was scored according to the Harvey-Bradshaw index (207) or simple clinical colitis index (208) with active disease considered as a score ≥5 in each category.

20 healthy controls were recruited with no self-reported major comorbidities. All participants were Caucasian.

Exclusion criteria were: significant co-morbidities, pregnant females, those who had taken antibiotics or pro/pre-biotics in the last 4 months, patients with stomas or those on a therapeutic diet (semi-elemental, elemental or parenteral nutrition) for IBD.

6.4.2 Exhaled breath capture

Patients were fasted for a minimum of 4 hours prior to breath collection. Breath samples were collected using a standardised procedure reported in published studies (341) (163) . Patients and controls were rested for 20 minutes prior to collection to minimise the possible effect of exercise (342) and were seated during sampling.

Mixed alveolar breath samples were captured in double layer (2 × 25 μm) Nalophan (Kalle UK Ltd., Witham, U.K.) bags via a 1 mL Luer lok syringe (Terumo Europe, Leuven, Belgium) secured with a pull tight security seal. The capacity of the bag was 2 litres and bags were washed with dry synthetic air (BOC Ltd.,Guildford, U.K.) prior to use. Each bag was single use only.

Participants were asked to perform a deep inhalation (as close to maximum as possible) followed by a complete expiration, followed by a deep inhalation again. At this point the plunger was removed from the syringe and participants blew through the barrel of the syringe to fill the bag. The plunger was immediately reinserted into the syringe and samples taken to the SIFT-MS laboratory within the hospital, sealed in a Tupperware container to reduce environmental contamination. Breath samples were analysed within 2 hours of sampling.

6.4.3 SIFT-MS

Breath profiles were characterised using a Profile-3 SIFT-MS instrument (Instrument Science, Crewe, UK). Breath samples were held in an incubator at 37°C for 5 minutes prior to sampling. Multi-ion mode was used with an analysis time of 60 seconds for selected VOCs. Gas carrier rate was set at 20mL/min with a temperature of 80°C. Table 6.1 shows the selected VOCs and individual precursor ions used.

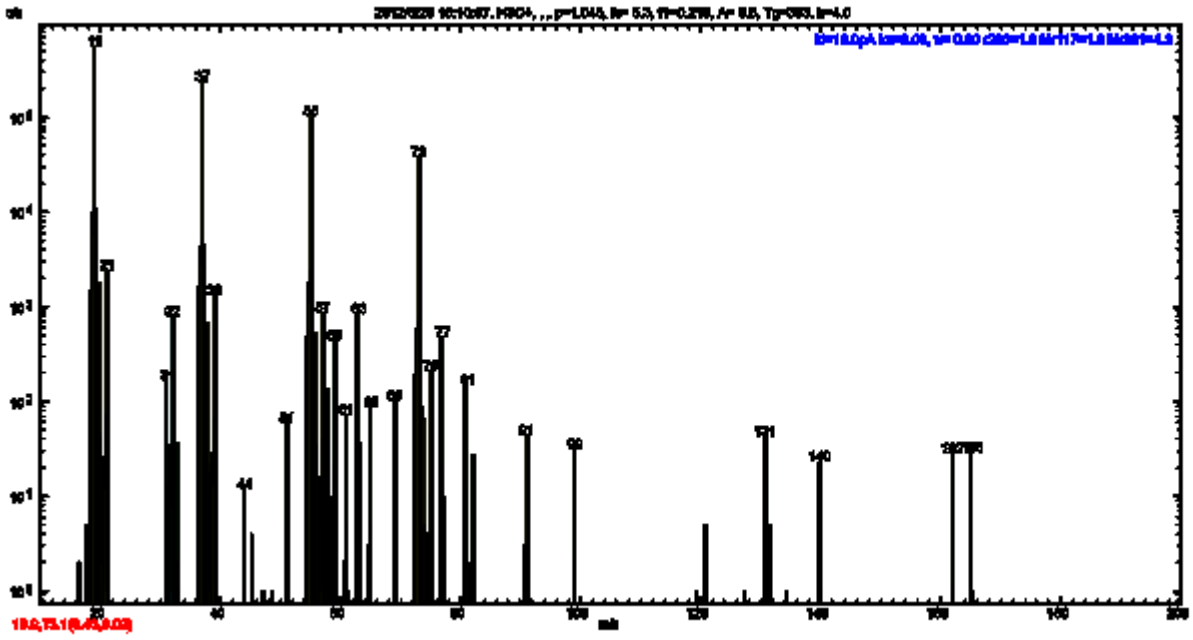


Figure 6.1 Example of SIFT-MS VOC profile obtained

	Molecular formula	Precursor ion	Characteristic product ions
Acetic acid	$C_2H_4O_2$	NO^+	$NO^+ \cdot CH_3COOH$, $NO^+ \cdot (H_2O)CH_3COOH$
Pentanoic acid	$C_5H_{10}O_2$	H_3O^+	$C_5H_{10}O_2H^+$
Hexanoic acid	$C_6H_{12}O_2$	H_3O^+	$C_6H_{12}O_2H^+$, $C_6H_{12}O_2H^+(H_2O)$
Propanal	C_3H_6O	NO^+	$C_3H_5O^+$
Butanal	C_4H_8O	NO^+	$C_4H_7O^+$
Pentanal	$C_5H_{10}O$	NO^+	$C_5H_9O^+$
Hexanal	$C_6H_{12}O$	NO^+	$C_6H_{11}O^+$
Heptanal	$C_7H_{14}O$	NO^+	$C_7H_{13}O^+$
Octanal	$C_8H_{16}O$	NO^+	$C_8H_{15}O^+$
Nonanal	$C_9H_{18}O$	NO^+	$C_9H_{17}O^+$
Decanal	$C_{10}H_{20}O$	NO^+	$C_{10}H_{19}O^+$
Methanol	CH_4O	H_3O^+	CH_5O^+, $CH_5O^+(H_2O)$
Propanol	C_3H_8O	H_3O^+	$C_3H_7^+$,$C_3H_7^+(H_2O)$
Butanol	$C_4H_{10}O$	H_3O^+	$C_4H_9^+$,$C_4H_9^+(H_2O)$
Pentanol	$C_5H_{12}O$	H_3O^+	$C_5H_{11}^+$, $C_5H_{11}^+(H_2O)$
Phenol	C_6H_6O	NO^+	$C_6H_6O^+$, $C_6H_6O^+(H_2O)$
Methyl phenol	C_7H_8O	O_2^+	$C_7H_8O^+$, $C_7H_8O^+(H_2O)$
Ethyl phenol	$C_8H_{10}O$	NO^+	$C_8H_{10}O^+$, $C_8H_{10}O^+(H_2O)$
Acetone	C_3H_6O	NO^+	$NO^+ \cdot C_3H_6O$
Dimethyl sulphide	C_2H_6S	H_3O^+	$C_2H_6SH^+$
Dimethyl disulphide	$C_2H_6S_2$	H_3O^+	$C_2H_6S_2H^+$
Hydrogen sulphide	H_2S	H_3O^+	H_3S^+
Carbon disulphide	CS_2	O_2^+	CS_2^+
Ammonia	NH_3	O_2^+	NH_3^+, $NH_3^+(H_2O)$
Hydrogen cyanide	HCN	H_3O^+	H_2CN^+
Isoprene	C_5H_8	NO^+	$C_5H_8^+$

Table 6.1: Molecular formula of VOCs measured, precursor ions used and characteristic product ions detected.

The VOCs chosen included those previously demonstrated to be altered in gastrointestinal disease (163), those derived from bacterial sources (141, 162) and those abundant in the healthy population for comparison (table 6.1).

6.4.4 Data analysis

GraphPad Prism 6 (GraphPad Software Inc. CA, USA) and IBM SPSS statistics 21 (SPSS Inc., Chicago, IL) were used to analyse phenotypic data using the Mann-Whitney U test or Kruskal-Wallis test for continuous variables and Fisher's exact test for categorical variables.

PCA and O-PLS-DA (as described previously in Chapter 2) were performed on the data using Pirouette v4.0 (Infometrix, Inc. Bothell WA, USA)(343). Datasets underwent pre-processing with a $\log^{10}$ transformation. PCA allowed visualisation of clustering and outliers. Loadings plots for each principal component were derived to explain the variables responsible for the patterns shown in the scores plots, and determine the most contributory VOCs in the modelling. In order to validate the models both leave-one-out cross validation (LOOCV) and 7-fold cross validation (7FCV) were applied. The Q^2 was calculated as a measure of the predictability of the model.

For each model the misclassification (confusion) matrix was derived, giving the number of correctly and incorrectly predicted samples in each group. From the confusion matrices the sensitivity and specificity of each model was calculated using the following:

Sensitivity (%) = true positive / [true positive + false negative]
Specificity (%) = true negative / [true negative + false positive]
Positive predictive value (%) = true positive / [true positive + false positive]
Negative predictive power (%) = true negative / [true negative + false negative]

	CD (actual)	Control (actual)
CD (predicted)	[True positive]	[False positive]
Control (predicted)	[False negative]	[True negative]

Table 6.2: Example of misclassification matrix for Crohn's disease vs controls.

Univariate comparisons of VOC concentrations between cohorts were also performed using the Mann−Whitney U test. A Receiver Operator Characteristic (ROC) analysis was performed for each comparative analysis using statistically significant VOCs (as defined by Mann−Whitney U test $p \leq 0.05$) for each paired cohorts. Area under the curve was calculated with 95% confidence intervals.

6.5 Results

6.5.1 Phenotypic dataset

Phenotypic and disease data of the patients and controls is shown in table 6.3. Phenotypic data of the groups were compared with Chi-squared test for categorical variables and Mann-Whitney *U* test for continuous variables. There were no statistically significant differences (defined as p value ≤ 0.05) between the groups.

	Healthy controls	CD	UC
n	18	18	20
Age (years) Median (Range)	44.5 (29-82)	45.5 (21-78)	49.5 (23-66)
Male Number (%)	8 (44%)	7 (39%)	10 (50%)
Smokers Number (%)	2 (11%)	3 (17%)	4 (20%)
Active disease[a]		6 (33%)	6 (30%)
Medication use in preceding 4 weeks[b]		M1 : 6	M1 : 18
		M2 : 9	M2 : 8
		M3 : 4	M3 : 0
		M4: 2	M4: 2
Disease location [c]		L1 : 2	E1 : 7
		L2 : 8	E2 : 8
		L3 : 8	E3 : 5
Disease behaviour [d]		B1 : 11	
		B2 : 4	
		B3 : 3	

Table 6.3: Demographic data and disease characteristics of participants

[a]As defined by disease activity index ≥5 [b] Medications: M1: 5-ASA; M2: immunosuppressants (Azathioprine/6-MP); M3:anti-TNF; M4: steroids [c] Location (206): L1:ileal disease; L2: colonic disease; L3:ileocolonic disease; E1:proctitis; E2:disease limit distal to splenic flexure; E3:disease proximal to the splenic flexure [d] Behaviour (206): B1:inflammatory; B2:structuring; B3:fistulating

6.5.2 Multivariate analysis

The dataset was first visualised on a PCA. There were no outliers and therefore no VOC data were excluded.

O-PLS-DA models were then built for the three comparative groups: Crohn's cohort vs controls (figure 6.2), UC vs controls (figure 6.3) and Crohn's vs UC (figure 6.4). Each comparative group showed clear separation between classes.

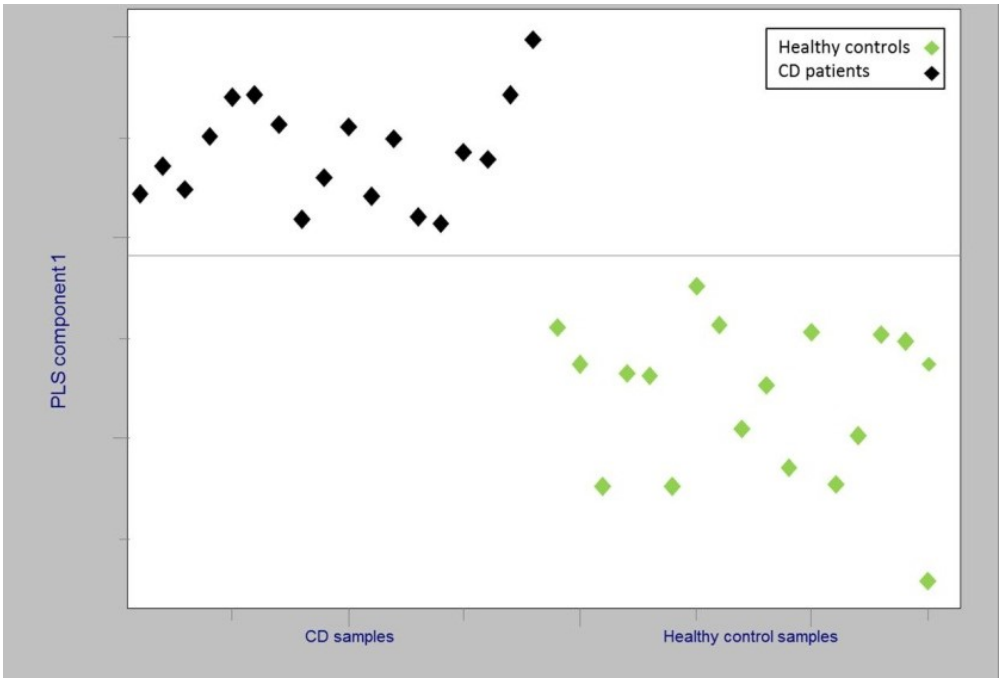


Figure 6.2: O-PLS-DA cross-validated scores for healthy controls and Crohn's patients.

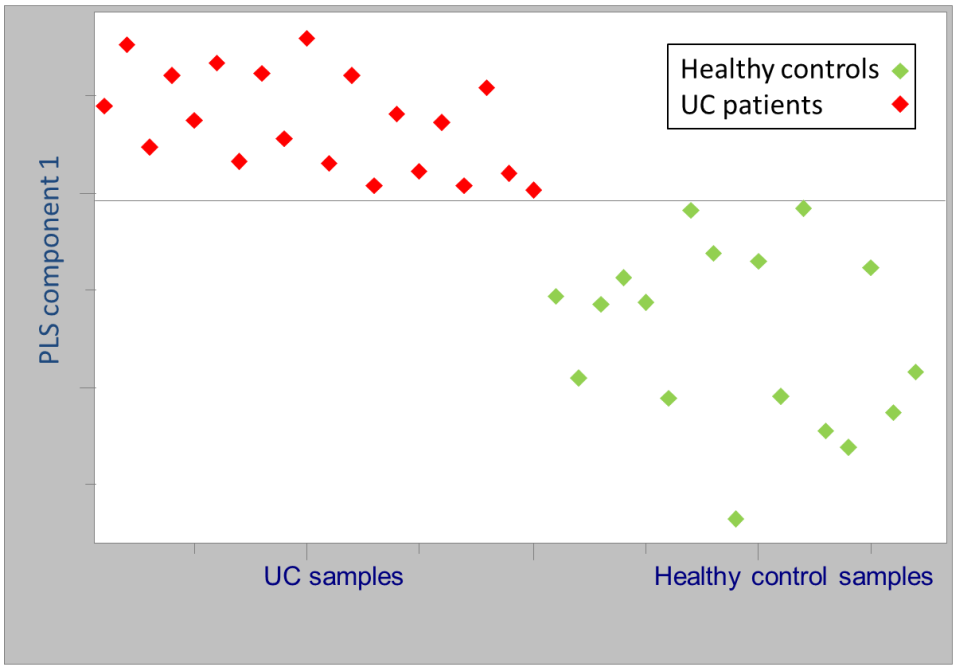


Figure 6.3: O-PLS-DA cross-validated scores for healthy controls and UC patients.

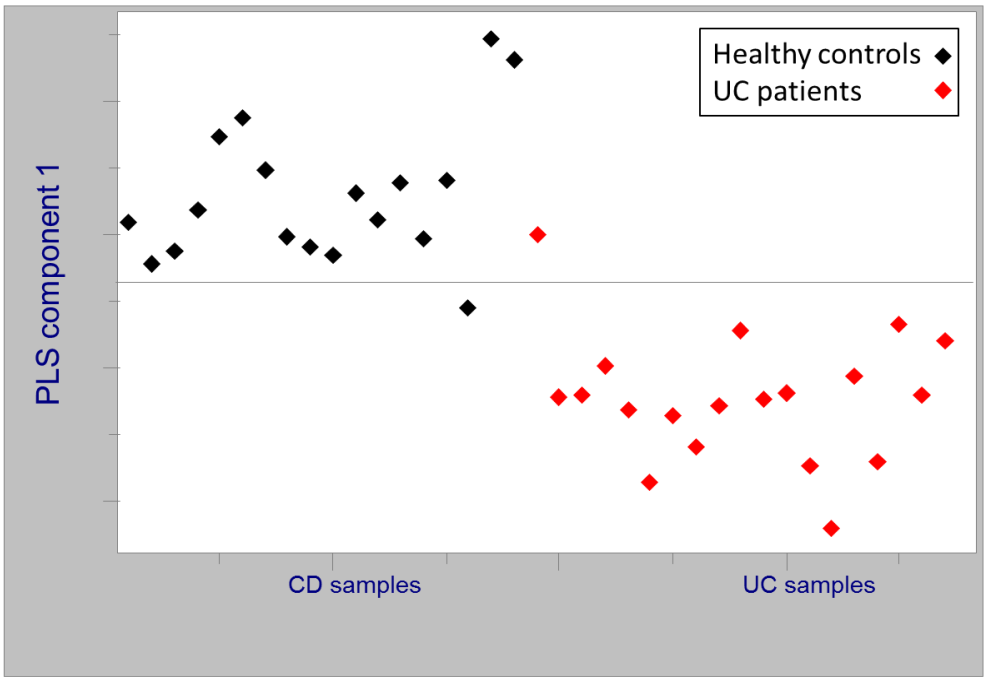


Figure 6.4: O-PLS-DA cross-validated scores for Crohn's and UC patients.

For each O-PLS-DA model the sensitivity and specificity was calculated from the confusion matrices. For each comparison the values were between 88 – 95%. The strongest model was Crohn's disease vs healthy controls with a sensitivity and specificity both of 94.4% and the highest Q^2 and R^2 values (0.78 and 0.79 respectively). There were no differences in the models when either LOOCV or 7FCV was used. For each model the regression vectors and loadings were analysed to show the VOCs contributing most to the separation.

	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	Q^2 value	R^2 value	Three most contributory VOCs in O-PLS-DA model
CD vs HC	94.4	94.4	94.4	94.4	0.78	0.79	Dimethyl sulphide Ammonia acetone
UC vs HC	90.5	94.4	95.0	88.9	0.66	0.72	Acetone Methanol ammonia
CD vs UC	88.9	90.0	88.9	90.0	0.69	0.71	Dimethyl sulphide Ammonia acetone

Table 6.4: Sensitivity and specificity values of the O-PLS-DA models created; Q^2 and R^2 values and VOCs most contributory to model construction.

6.5.3 Univariate analysis

Individual measured concentrations on VOCs were compared from each group. Those compounds which different significantly (as defined by a Mann Whitney U-test p value of <0.05) between the cohorts are shown in table 6.5.

				p values		
VOCs (ppbv)	CD	UC	Healthy control	CD vs Healthy control	UC vs Healthy controls	CD vs UC
Dimethyl sulphide	556.2 [303.9-1143.2]	357.2 [228.7-698.7]	302.1 [182.4-503.9]	**0.0223**	0.438	0.0815
Hydrogen cyanide	19.2 [12.4-27.4]	12.8 [10.0-17.9]	19.9 [9.9-25.1]	0.6224	0.1782	**0.0276**
Hydrogen sulphide	1.1 [0.5-2.3]	3.1 [2.0-8.1]	4.1 [1.8-8.8]	**0.0062**	0.539	**0.0022**
Butanal	3.1 [2.3-4.1]	1.5 [1.0-3.0]	1.9 [1.1-2.7]	**0.0270**	0.9366	**0.0388**
Nonanal	1.9 [1.0-2.7]	1.3 [0.7-2.5]	0.9 [0.3-1.5]	**0.0098**	0.3322	0.1924
Ammonia	342.1 [163.0-470.4]	207.4 [133.5-376.2]	415.2 [246.0-789.4]	0.2904	**0.0101**	0.0578

Table 6.5. Measured concentrations in ppbv (parts per billion by volume) of significant VOCs from CD, UC and Healthy control groups. Median values are given with interquartile range. Significant p values shown in bold.

Dimethyl sulphide (p=0.0223), hydrogen sulphide (p=0.0062), butanal (p=0.027) and nonanal (p=0.0098) differed significantly between the CD group and healthy controls. In the UC cohort there was a significantly lower levels of ammonia (p=0.0101) compared to healthy controls. Between UC and CD patients the measured hydrogen cyanide (p=0.0276), butanal (p=0.0388) and hydrogen sulphide (p=0.0022) were different. And although they did not reach statistical significance, the two compounds contributing most strongly to the multivariate model comparing these groups – dimethyl sulphide and ammonia - showed trends towards significance (p= 0.0815; p=0.0578 respectively).

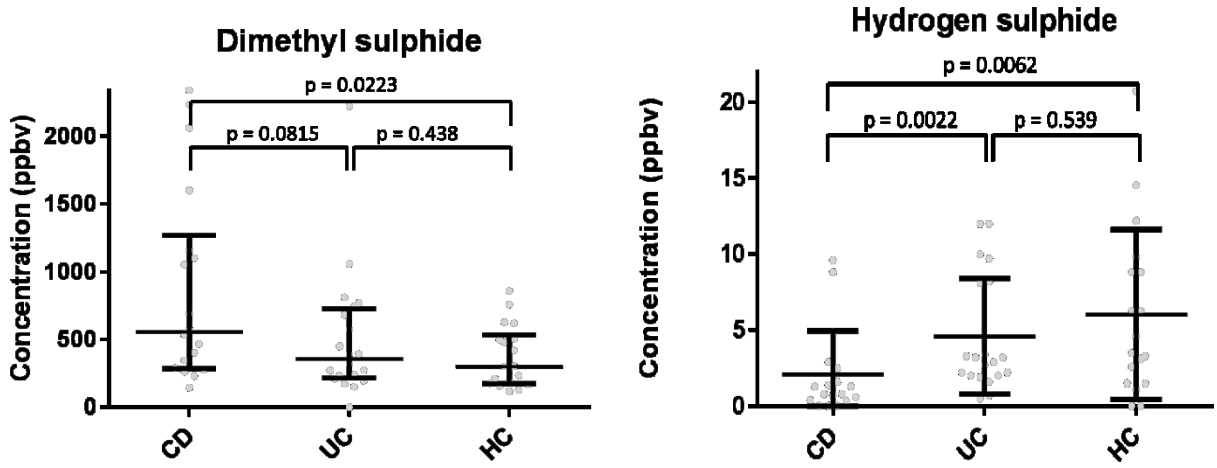


Figure 6.5: Scatter plots of obtained concentrations of dimethyl sulphide and hydrogen sulphide, showing median and 95% confidence bars, with p values between groups shown.

Using the discriminatory compounds (see table 6.4) for each comparison, ROC analysis gave an integrated AUC of 0.864 for CD vs healthy controls, 0.742 for UC vs controls and 0.828 for UC vs CD (table 6.6).

	AUC	95% confidence intervals
CD vs healthy controls	0.864	0.749-0.980
UC vs healthy controls	0.742	0.581-0.902
CD vs UC	0.828	0.699-0.956

Table 6.6: Summary of ROC analysis using significant VOCs to distinguish cohorts.

Further subgroup analysis of disease activity (as defined by HBI and SCAI scores >5) did not show any significant differences, although the groups were small. The cohorts were too small and / or uneven to allow any meaningful analysis comparing disease location or medication use.

6.5.4 Validation

For validation of the technique, the concentrations of abundant VOCs (defined as compounds uniformly present in the exhaled breath of the normal population at concentrations of several hundred ppbv (163)) were also measured. Isoprene, acetone and methanol concentrations did not differ between the three groups (table 6.7) and the values fell within the ranges reported using this technology from other groups (163, 341, 344)

							Comparative studies		
VOCs	CD	UC	Healthy control	CD vs HC	UC vs HC	CD vs UC	HC (163)	HC (341)	HC (344)
Mean concentration (ppbv)				p values			Mean concentration (ppbv)		
Isoprene	69.3	47.8	59.4	0.9196	0.0597	0.0639	64	52	n/a
Acetone	268.5	351.7	247.9	0.9045	0.2041	0.2737	338	372	363
Methanol	217.4	235.0	227.9	0.7123	0.6269	0.7107	248	159	238

Table 6.7: Concentration of abundant VOCs in this study and comparison with healthy controls of other studies

6.6 Discussion

This pilot study analysed exhaled breath profiles in adult IBD for the first time. Results confirmed the hypothesis that SIFT-MS would be able to detect a significant alteration in breath VOCs induced by IBD. Multivariate and univariate approaches were employed to highlight compounds relating to the disease which may reveal insights in to pathological mechanisms.

Previous to this work a SIFT-MS study in a paediatric cohort was able to distinguish between IBD and controls, but unable to separate the sub-types (177). However this study was able to distinguish CD from UC with good sensitivity and specificity.

Subsequent to this study being published, a FAIMS (field asymmetric ion mobility spectroscopy) breath study showed specific 'breathprints' associated with IBD, which also were distinct between Crohn's and UC, but could not identify specific compounds for interpretation (178).

Following this study, two other groups published SIFT-MS breath studies in IBD, and both were also able to distinguish IBD patients from non-IBD with high accuracy (173, 345).

This work expands on earlier VOC studies (177) and demonstrates the potential utility of breath profiling in IBD.

6.6.1 Understanding the VOC profile in IBD

To interpret these results it is necessary to investigate the origin of the VOCs altered in IBD breath profiles to understand the relationship with the disease.

6.6.1.1. Bacterial VOCs: Hydrogen sulphide, dimethyl sulphide, ammonia and hydrogen cyanide

As discussed in this thesis, dysbiosis plays an essential part in the development and continuity of IBD, and it has been described how other metabonomic studies have highlighted bacterial by-products as discriminatory markers of the disease (118, 143, 146). In this study the compounds which different most significantly between Crohn's patients and controls were dimethyl sulphide (DMS) (increased) and hydrogen sulphide (H_2S) (decreased). These VOCs are mostly produced by gut bacteria (346) but also, in smaller amounts, by human cells (347, 348). They have also been linked to the initiation of mucosal inflammation (349). Patel et al also showed lower concentrations in hydrogen sulphide in IBD, although no difference was seen in DMS (177). This may be partly explained as their absolute concentrations of VOCs were only shown for the mixed IBD cohort (not sub-types), and in this study DMS was only statistically higher in Crohn's, not in UC, so this finding may have been lost in the mixed IBD group.

More recently Reider et al (345) also showed significantly elevated levels of DMS in the breath of IBD patients compared to controls; hydrogen sulphide levels were not statistically altered.

Hydrogen sulphide and DMS have also been linked to several systemic inflammatory conditions including infective endocarditis and hepatic cirrhosis; and DMS has also been associated with lung cancer, cystic fibrosis and chronic hepatitis (350). Increased exhaled breath levels of hydrogen sulphide have been shown in patients with small intestinal bacterial overgrowth correlating to activity of intestinal sulphate-reducing bacteria (351). In the gut hydrogen sulphide has been shown to have discrepant cytoprotective and cytotoxic effects on colonic mucosa, and likely has a physiologic regulatory function in healthy people (348).

Although it has been hypothesized that gut inflammation is related to hydrogen sulphide and sulphate-reducing bacteria in ulcerative colitis, evidence is thus far inconclusive and the true role of hydrogen sulphide in colonic disease is not clear (352).

In the UC cohort the ammonia concentration was significantly lower than that of healthy controls. Ammonia is formed by amino acid metabolism in the gastrointestinal tract, as well as directly from gastrointestinal bacteria (353). It has been studied as a volatile biomarker in a variety of conditions including *Helicobacter Pylori* infection, cirrhosis and renal failure (354). The colon has the highest concentration of ammonia in the body (354) but the link to IBD is unclear.

Hydrogen cyanide (HCN) is produced by leucocytes and has been associated with neutrophil activation (355) in infection. The source of most exhaled HCN is most likely primarily from the oral cavity and respiratory tract. Studies have shown it to be produced by bacteria in vivo and in vitro studies (356). HCN has been proposed as a biomarker of specific infections including *Pseudomonas aeruginosa* infection in cystic fibrosis (357, 358), and *Burkholderia cepacia* (356) however findings of other studies have been contradictory (355). Other data shows HCN breath levels are actually lower in pneumonia patients (359) and therefore this remains controversial and the link to IBD is not clear.

6.6.1.2 Aldehyde VOCs: butanal and nonanal

Other significant discriminating compounds identified in CD were elevated levels of butanal and nonanal; VOCs belonging to the aldehyde group. Breath aldehydes are elevated in lung (360), breast (361), gastric and oesophageal cancers (163) and likely represent oxidative stress and tumour cell metabolic pathways.

Specifically, nonanal has been shown to be higher in the breath those with upper GI cancer (163), and was higher in a cohort of lung cancer patients (both smokers and non-smokers) (360), and ovarian cancer patients compared to healthy volunteers (362). Butanal was measured at a significantly higher level in the breath of both gastric and oesophageal cancer patients (163) as well as in lung cancer (363).

One of the phenomena associated with inflammation in IBD is oxidative stress (364) caused by reactive oxygen molecules (365). Studies have linked other breath markers of oxidative stress (not measured in this study) including pentane (172) and ethane (170, 171) to IBD, suggesting that possible biomarkers of GI inflammation are detectable with this methodology.

6.6.2 Applications and potential limitations of the study

6.6.2.1 Strengths of the study

The novel aspect of this work is that it was the first multi-compound breath study in adult IBD, and 19 VOCs in this study have been examined in this disease for the first time. The IBD patients were robustly characterised, groups were well matched and accurate metadata was collected. Importantly the participants were all fasted, unlike other published studies, as the effects of diet on the breath profile is not fully recognised. The sampling technique described has previously been shown to minimise exogenous contributions on VOC concentrations. For validation the measured concentrations of the abundant VOCs were similar to those in 4 published studies of healthy adults (163, 341, 344) underlining the reproducibility and legitimacy of this technology.

The advantages of this modus operandi are that is immediate, non-invasive and acceptable to patients. Results have shown robust models separating cohorts of Crohn's, UC and controls with good sensitivity and specificity. With these interesting and novel findings, further investigation is warranted to assess the potential application of SIFT-MS in IBD.

6.6.2.2 Addressing potential confounders

Although the methodology of breath sampling and measurement used this study has been validated in several published papers (163, 341, 366), with over 2000 published breath studies there is no overall consensus in design or accepted 'normal' values of breath metabolites. Until recently breath research was limited by the sensitivity and accuracy of technology but the new challenge is to develop reliable and reproducible standard operating procedures, as well as further understand the data generated. Further studies are required, on bigger clinical cohorts, to verify these findings, as well as allow sub-group analysis to investigate the effect of disease activity and location on the IBD breath profile.

There are multiple possible confounders to breath metabolite quantification data which must be addressed and minimised (367). Patient demographics, including age and gender (368), as well as ethnicity (369) can affect VOCs, although in our group all participants were Caucasian and age and sex were matched between cohorts. It is known that diet has a complex contributory effect on breath profiles (370), and whilst patients in this study were fasted for 4 hours, the full effect of starvation on breath profiles is not known and no accepted consensus for this (178).

It is established that tobacco use can influence VOCs in a healthy population (371) (specifically nonanal) through inflammatory and oxidative stress pathways. Smokers were included in the

study although the numbers were low and there was no statistical difference in the proportion of smokers in the three cohorts.

Other possible contaminants include environmental pollutants and atmospheric chemicals, efforts were made to reduce the potential effect of these by sealing the samples in boxes for transportation and performing analysis within 2 hours of sampling. Treatments that alter bowel flora, such as antibiotics or laxatives (372) can also affect breath analysis, and therefore patients on these were excluded.

As discussed earlier, breath pentane and ethane levels have been linked to IBD. These two compounds were not measured in this pilot study as the VOCs chosen were hypothesised to relate to dysbiosis, or have been previously demonstrated to be abnormal in upper GI inflammation with this technique (373). Although pentane has been correlated with active IBD, it has also been shown to alter in other systemic inflammatory conditions including rheumatoid arthritis (374) so it is not clear if this is a non-specific inflammatory phenomenon (375). Incorporating measurement of these two alkanes may be useful in helping discriminate active and quiescent disease.

6.6.2.3 Future work

Large prospective studies are necessary to validate these findings and to underpin confidence in this technology and its application to IBD research. As with other metabonomic techniques, it is unlikely that a single breath biomarker will be discovered, more likely differences in combinations of relevant breath VOCs will be helpful in differentiating between diseases or monitoring inflammation.

As there is no single accepted methodology of this technique, the combination of GC-MS for VOC identification, with SIFT-MS for quantification would be a rigorous approach for

validation (375), similar to the approach of combining univariate and multivariate analysis in previous chapters of this thesis.

In future, the VOCs found to be altered by IBD in this study should be combined with ethane and pentane to test whether this improves the capability of SIFT-MS breath profiling to characterise disease activity in IBD.

6.7 Advances in Knowledge

- SIFT-MS breath profiles demonstrate specific patterns relating to Crohn's and UC to enable them to be differentiated from control and one another with good sensitivity and specificity.
- Significant candidate VOCs identified related to bacterial dysbiosis and oxidative stress – both mechanisms implicated in the pathogenesis of IBD.
- Exhaled breath analysis offers a valuable and innovative approach to further characterising, and understanding the meaning of the metabolome in IBD.

Chapter 7: General Discussion and Conclusion

'All diseases begin in the gut'

Hippocrates 460 – 370 BC

This thesis is composed of four metabonomic investigations of IBD, in exploration of refining the use of this science in the disease, to uncover pathogenic insights and better reveal the variation on the metabolic profile in different patient groups.

7.1 Exploring metabolic profiling in different ethnic groups

Despite South Asians being the second largest sized ethnic group in the UK, and IBD risk in the South Asian migrant population exceeding that of white Europeans, data in this group is sparse and lacking. In this chapter a large cohort of IBD patients and controls of these two ethnicities were profiled with urine NMR.

One of the key findings was that metabolic profiles differ between both IBD patients and healthy controls when comparing white Europeans and South Asians, demonstrating the significant effect of ethnicity on the metabolic profile, which may have been underestimated in other published metabonomic studies. Furthermore, metabolites identified as separating the ethnic groups, hippurate and p-cresol, are bacterial co-metabolites, and whilst differences were also seen between the diets of the two ethnic groups, it is most likely that the inter-connection of ethnicity, nutrition and dysbiosis are responsible for these alterations.

As most evidence-based guidelines are based on research from white participants, these results underline the importance of further research in IBD populations of different ethnicities, to improve an individualised approach to clinical decision making.

7.2 Metabolic profiling as a longitudinal technique

This chapter was to investigate changes in the urinary metabolic profile in IBD patients over several years, to try to detect changes in the baseline profile that may predict future disease risk, and also to look at how clinical outcomes including surgery and medical treatments affect the metabolome. Although only a relatively small cohort, analysis showed that intra-individual changes were strongly outweighed by inter-individual variation, and in fact no intervention except surgery altered the profile significantly. This was perhaps a surprising finding, given the anticipation that metabonomics may be a useful tool in longitudinal studies.

From these results, whilst only a pilot study, it is clear that further larger studies, specifically designed to examine changes over time in the individual IBD metabolome are necessary, if metabolic profiling is to be applied to monitor dynamic outcomes such as inflammatory resolution, or responses to treatment.

7.3 Pilot inter-omic exploration of IBD

Through examining different biofluids it has been seen that the metabolic signal in urine is stronger in Crohn's disease, and in faecal water UC has a stronger effect. This is consistent with data from other studies, and is likely due to the fact that faecal water metabolites are mostly colonic in origin and urinary metabolites reflect more systemic physiology. This may direct future study design in the choice of biofluid selected in exploring particular diseases.

Using several complementary 'omic' approaches together enables the direct relationship between microbial disruption and metabolic end products to be established, which previously was only hypothesized. An extrapolation of this design in a larger selective cohort could tie together specific microbial-metabolite mechanisms otherwise shown in ex-vivo work.

7.4 Novel method of analysing the IBD metabolic profile with breath VOC profiling

Moving forward alongside biofluid metabonomics the volatilome is a growing area of interest, and this chapter demonstrated distinct findings in breath related to IBD. Subsequently published studies have confirmed similar findings (173, 345) and endorsing SIFT-MS breath profiling as a potentially valuable research strategy, and a non-invasive and highly acceptable methodology.

7.5 Lessons learnt from this PhD

This thesis is the summation of several years of work, during which time the projects evolved and changed. In drawing together and analysing these results, aspects have been highlighted which could be improved upon – both practical and in scientific design, in order to maximise and improve results for future projects.

Patient recruitment was lengthy and time consuming and a large part of the time spent for all the studies. In future, a multi-site collaboration with shared data and sample collection would improve the efficiency of patient recruitment. IBD research specialist nurses in the clinic environment and publicising the study directly to patients may also help.

For a subsequent ethical approval it would be logical to include all studies in a combined application covering sampling of all biofluids in the same patient cohort, making a large integrated study much easier to perform. Similarly all patient phenotype data should be accumulated in a large database allowing stratified groups of each phenotype to be easily identified and selected.

The ethnicity study showed it has been difficult to extrapolate the effect of diet on the results found. To disentangle the relationship between diet, the microbiome and the metabolic

signature, in future even more detailed dietary data should be recorded from participants. Retrospective dietary recall is poor, and prospective food diaries are more accurate, however as these are time consuming it may be appropriate to apply this to specific subgroups only. This may further allow the host- nutrition-microbiome interaction to be explored.

7.6 Direction of future work

The ethnicity chapter has emphasised that there are important differences between IBD patients of white and ethnic minority groups, and there is a significant lack of research in this area. In further investigating this future studies should be powered to investigate first and second generation South Asians, to examine the effect of migration, environment and genetics.

In this group it would be interesting to perform genotyping, specifically for alleles associated with IBD in this ethnicity (for example HLA-DRB1*1502) to help distinguish the degree of influence of genetic and environmental pathogenic factors. Groups of patients of different ethnicities with sub-types of disease, such as isolated small bowel Crohn's disease should also be compared as these specific phenotypes clinically manifest differently and this may be detectable in the urinary metabolic profile.

The longitudinal metabonomic study unexpectedly did not shown significant alterations in the urinary metabolic profile over time. However, in further investigating this, a study with more IBD patients and the addition of controls, as well as more homogenous clinical groups, should be undertaken. An example would be treatment naïve patients newly initiated and monitored on biologics, with matched controls. The main limitation of the study in chapter 4 was that only 2 samples were available for each patient, but a prospectively recruited study would allow multiple samples to be measured longitudinally.

The inter-omic analysis has demonstrated that the integration of complex datasets is feasible but again requires larger cohorts for endorsement. As discussed in the chapter it may be appropriate here to perform full metagenomics, to not only identify and catalogue the bacteria present, but also allow more functional profiles which can be related to metabolic pathways. The incorporation of serum and/or breath metabonomics could also potentially add useful additional information to this.

The breath VOC findings are exciting and novel, but as with all pilot studies, larger studies are required to confirm and validate findings. The ideal study would involve a bigger, well phenotyped IBD population and matched controls. As in our study it would be essential to fast all participants pre-sampling, but in future studies, smokers should also be excluded until the specific effect of smoking on the breath metabolic profile is known.

Subsequent to our study being published, two further SIFT-MS studies were published, also demonstrating the ability of breath VOCs to distinguish IBD from controls (173, 345), corroborating our findings. By combining the subsequent findings of other SIFT-MS studies it would be interesting to measure all candidate compounds identified in the three studies as distinguishing IBD, as well as ethane and pentane from older published investigations. This may improve sensitivity and specificity of area under the ROC further.

All studies demonstrated alterations in VOCs of bacterial origin implicating dysbiosis as the cause, it would therefore be appropriate to demonstrate bacterial changes in the same group of patients with 16S profiling to show this relationship.

7.7 Perspective on metabolic profiling and its application in IBD research

The advantage of using a systems biology approach, such as metabonomics, is the powerful unbiased data extraction, which can measure multi-level physiological and pathological perturbations at one point in time. The changes seen in the metabolic profile in IBD have been shown here to be consistent and reproducible, but aside from distinguishing Crohns and colitis the nuances of the more subtle changes of the disease on the metabolic profile are yet to be fully understood.

One of the common criticisms of metabonomics is that it has yet to be translated into a valid clinical biomarker. Certainly with such complex data the challenge is to be able to decipher and explain the results to be able to render them clinically applicable.

Measuring a complex, dynamic and diverse system (the patient) along with a second integrated complex ecosystem (the microbiome) is incredibly difficult to perform experimentally. Moreover the extreme heterogeneity of humans makes proving mechanistic associations highly challenging. The results from this thesis show that the potential value of metabonomics may lie in enhancing our understanding of the complex relationship between internal (microbiome and ethnicity) and external factors in IBD pathogenesis and progression; rather than developing metabonomics as a simple biomarker or diagnostic test. In trying to extract the maximal information from metabonomic studies it is crucial to combine this with other 'omic' technologies to determine biological relationships, and / or with other complementary biofluid analysis such as breath.

The door is wide open for future studies to develop and build on the findings in this thesis to refine and enhance our knowledge of metabolic profiling in IBD, with the ultimate outcome of clinically relevant data that will improve the treatment of the disease for patients.

Publications arising from this thesis to date

Full papers

Hicks LC, Huang J, Kumar S, Powles ST, Orchard TR, Hanna GB, Williams HR. Analysis of Exhaled Breath Volatile Organic Compounds in Inflammatory Bowel Disease: A Pilot Study. *J Crohns Colitis*. 2015 Sep;9(9):731-7

Hicks LC, Ralphs SJ, Williams HR. Metabonomics and diagnostics. *Methods Mol Biol.* 2015;1277:233-44

Oral presentations

L. Hicks, et al **Exhaled volatile organic compound breath analysis in Inflammatory Bowel Disease**
Oral presentation: Digestive Diseases Federation: London IBD free papers section

STR Powles, **L. Hicks**, et al **Effect of Co-morbidities on urinary metabolic profiling in the characterisation of patients with inflammatory bowel disease**
Oral presentation: UEG week 2015 (UEG15-ABS-5251)

Abstracts

L. Hicks, et al A**ssessing the effect of ethnicity in urinary metabolic profiles in IBD**
Abstract: 25th UEG week Barcelona 2017
Abstract: 12th Congress of European Crohn's and Colitis Organisation, 2017

L. Hicks, et al **Effects of time on urinary metabolic signatures in IBD**
Abstract: 25th UEG week Barcelona 2017
Abstract: 12th Congress of European Crohn's and Colitis Organisation, 2017
Abstract: Digestive Diseases Week, Washington 2017

STR Powles, **L. Hicks**, et al **The Use of Rapid Evaporative Ionisation Mass Spectrometry (REIMS) in Faecal Samples to Identify Inflammatory Bowel Disease**
Abstract: 25th UEG week Barcelona 2017

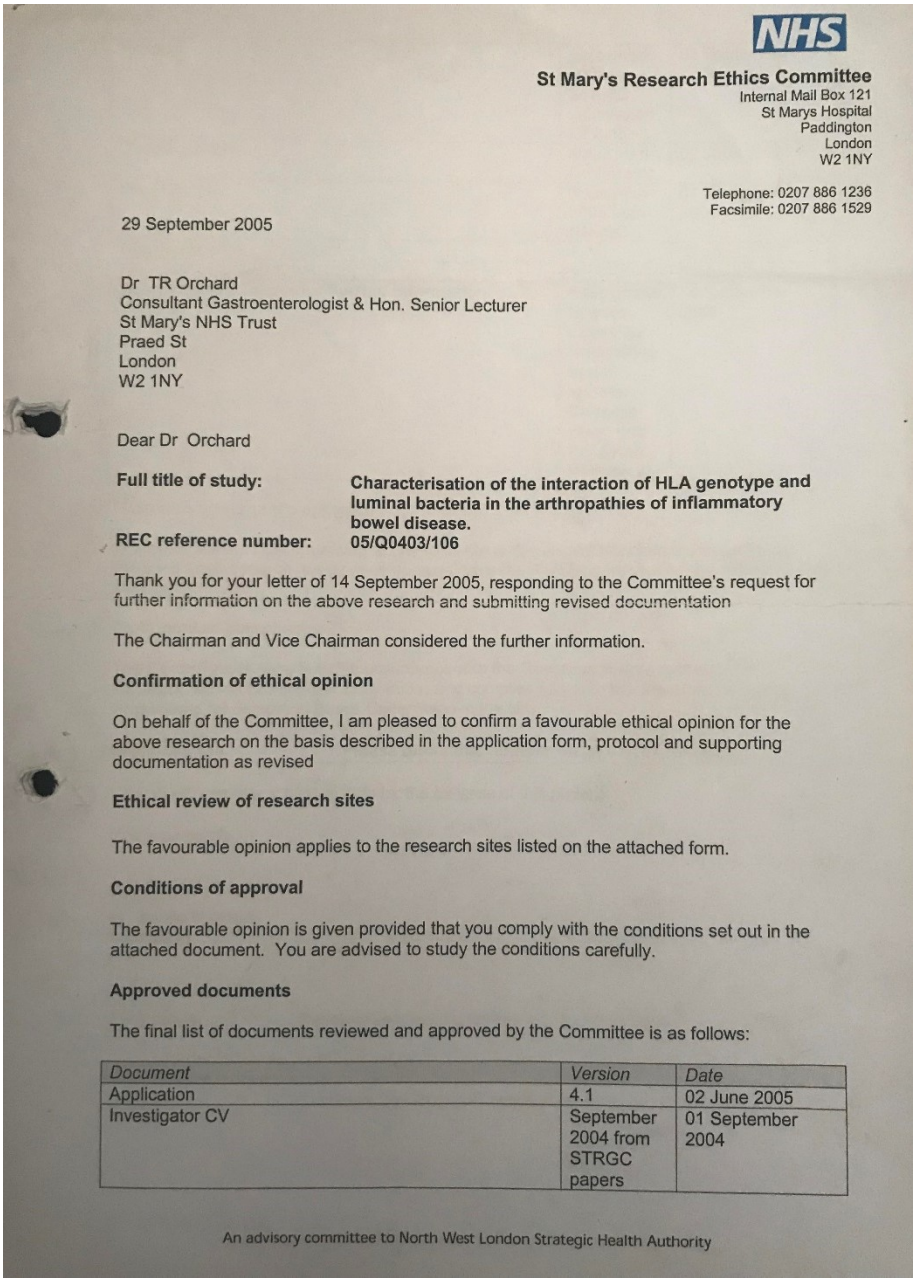
STR Powles, **L. Hicks**, et al. **Assessing the individual risk of acute severe colitis at diagnosis in a South Asian population**
Abstract: BSG Annual Meeting June 2017
Abstract: 12th Congress of European Crohn's and Colitis Organisation, 2017

STR Powles, **L. Hicks**, et al **Effect of Co-morbidities on urinary metabolic profiling in the characterisation of patients with inflammatory bowel disease**
Abstract: 2nd Digestive Disorders Federation Meeting, London 2015

L. Hicks, et al **Exhaled volatile organic compound breath analysis in Inflammatory Bowel Disease**
Abstract Number: A-1265 Abstract 10th Congress of ECCO - Inflammatory Bowel Diseases 2015, Barcelona 2015

L. Hicks, et al. **Urinary metabolic profiling of inflammatory bowel disease in a South Asian cohort**
Abstract: P617 8th Congress of European Crohn's and Colitis Organisation, 2013
Abstract 1601609. Digestive Diseases Week, Orlando 2013

Supplement 1



NHS

St Mary's Research Ethics Committee
Internal Mail Box 121
St Marys Hospital
Paddington
London
W2 1NY

Telephone: 0207 886 1236
Facsimile: 0207 886 1529

29 September 2005

Dr TR Orchard
Consultant Gastroenterologist & Hon. Senior Lecturer
St Mary's NHS Trust
Praed St
London
W2 1NY

Dear Dr Orchard

Full title of study: **Characterisation of the interaction of HLA genotype and luminal bacteria in the arthropathies of inflammatory bowel disease.**

REC reference number: **05/Q0403/106**

Thank you for your letter of 14 September 2005, responding to the Committee's request for further information on the above research and submitting revised documentation

The Chairman and Vice Chairman considered the further information.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised

Ethical review of research sites

The favourable opinion applies to the research sites listed on the attached form.

Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Application	4.1	02 June 2005
Investigator CV	September 2004 from STRGC papers	01 September 2004

An advisory committee to North West London Strategic Health Authority

Supplement 1: Ethical approval from Research Ethics Committee for study 05/Q0403/106
12 further amendments were approved from these original ethics.

Supplement 2

Imperial College London

Division of Medicine

CONTROL PARTICIPANT INFORMATION SHEET

RESEARCH INTO THE CAUSES OF INFLAMMATORY BOWEL DISEASE

BLOOD, URINE, BREATH AND STOOL SAMPLES

Dear Sir/Madam,

You are being invited to take part in a research study; please take time to read the following information carefully. Ask us if there is anything that is not clear. Participation is voluntary and you will be given a copy of this information sheet and your consent form. Thank you for reading on.

We are looking into the causes of inflammatory bowel disease (IBD) a condition which can cause debilitating symptoms and significantly impair quality of life. Previous research has suggested that genetic and metabolic factors may be important, and this research aims to study this in more detail, and to see whether the bacteria that live in the bowel may also affect who develops IBD. 1000 patients with and without IBD, and healthy individuals, will be studied using modern techniques.

If you agree to participate, this will involve simply giving some, or all of the following: a blood sample (usually at the same time as routine blood tests), a urine, a stool and a breath sample, plus a 5-10 minute chat to a doctor.

We will then analyse the samples to find out more about the disease links to genetic factors and the way in which the bacteria in the bowel play a part.

The results of this research will have no implications for individuals, and the samples will be analysed anonymously. However the research will give us a greater understanding of bowel disease, and why different people get different problems, and it may in future lead to the development of new treatments.

Thank you for reading this information sheet.

If you have any questions, please contact Dr Lucy Hicks or Sam Powles, who are contactable at the GI Unit, St Mary's Hospital, Praed St, London W2 1NY. Tel 0203 3126678 or 07766153195

Imperial College London

Division of Medicine

PATIENT INFORMATION SHEET

RESEARCH INTO THE CAUSES OF INFLAMMATORY BOWEL DISEASE

BLOOD, URINE, BREATH AND STOOL SAMPLES

Dear Sir/Madam,

You are being invited to take part in a research study; please take time to read the following information carefully. Ask us if there is anything that is not clear. Participation is voluntary and you will be given a copy of this information sheet and your consent form. Thank you for reading on.

We are looking into the causes of inflammatory bowel disease (IBD) a condition which can cause debilitating symptoms and significantly impair quality of life. Previous research has suggested that genetic and metabolic factors may be important, and this research aims to study this in more detail, and to see whether the bacteria that live in the bowel may also affect who develops IBD. 1000 patients with and without IBD, and healthy individuals, will be studied using modern techniques.

If you agree to participate, this will involve simply **giving some, or all of the following:** a blood sample (usually at the same time as routine blood tests), a urine, a stool and a breath sample, plus a 5-10 minute chat to a doctor.

We will then analyse the samples to find out more about the disease links to genetic factors and the way in which the bacteria in the bowel play a part.

Involvement in this research will in no way affect your treatment.

The results of this research will have no implications for individuals, and the samples will be analysed anonymously. However the research will give us a greater understanding of bowel disease, and why different people get different problems, and it may in future lead to the development of new treatments.

Thank you for reading this information sheet.

If you have any questions, please contact Dr Lucy Hicks or Sam Powles, who are contactable at the GI Unit, St Mary's Hospital, Praed St, London W2 1NY. Tel 0203 3126678 or 07766153195

Supplement 2: Control and patient information sheets

Supplement 3

all ethnicities	4-cresol sulphate	Acetate	Alanine	Citrate	Creatine	Creatinine	Dimethylamine	Formate	Glycine	Glycolate
Crohns	18.55	6.20	33.99	138.45	779.10	496.18	43.39	3.51	61.50	29.71
UC	20.36	6.77	35.43	153.19	819.19	522.02	45.65	3.07	50.41	31.54
Healthy controls	21.43	6.44	32.78	158.25	813.77	526.62	44.29	3.59	54.24	31.41
White Northern European										
Crohns	21.77	6.23	33.96	139.21	799.93	492.47	43.29	3.61	63.29	29.84
UC	26.21	6.59	33.36	170.96	806.46	511.83	44.21	3.35	53.39	30.34
controls	27.91	6.28	30.94	150.06	811.87	529.21	42.37	3.37	53.08	31.30
South Asian										
Crohns	10.76	6.83	35.48	147.47	818.49	565.04	44.63	3.38	55.12	31.29
UC	13.97	6.88	40.55	142.32	885.41	576.14	48.15	3.08	49.74	31.73
controls	11.30	6.67	38.55	172.78	844.76	537.35	45.00	3.85	59.58	32.21

all ethnicities	Guanidoacetate	Hippurate	Lactate	Methanol	Methyl-histidine	Succinate	Trans-aconitate	Trigonelline	Trimethylamine-N-oxide
Crohns	28.50	46.72	34.22	10.64	33.02	12.14	0.99	2.77	101.57
UC	25.08	61.51	35.62	10.41	34.02	13.46	1.07	2.32	84.56
Healthy controls	24.10	84.16	33.91	11.73	28.55	13.57	1.02	3.55	104.90
White Northern European									
Crohns	27.93	54.83	34.49	11.14	33.12	12.92	0.97	2.45	123.39
UC	26.76	82.09	34.55	10.99	38.31	14.67	1.02	2.23	94.17
controls	25.16	98.23	33.62	12.59	29.67	13.26	1.01	3.51	117.02
South Asian									
Crohns	31.14	17.74	31.81	10.25	32.47	9.52	1.04	4.18	44.92
UC	24.06	53.57	36.93	9.82	29.08	12.09	1.22	2.40	76.62
controls	21.80	75.96	36.54	10.56	26.42	14.12	1.08	3.75	86.66

Supplement 3: Median values for all metabolites in all groups

Supplement 4

All patients	p value across all three groups [a]	q value [b]	HC : CD [c]	HC : UC [c]	UC : CD [c]
Lactate	0.0532		0.8174	0.112	0.0779
Alanine	0.2287		0.2658	0.0984	0.5069
Acetate	0.4443		0.3814	0.5256	0.2532
4-cresol	0.5891		0.3795	0.4270	0.8370
Succinate	0.3412				
Citrate	0.0719				
Dimethylamine	0.2080				
Creatine	0.0974				
TMNO	0.0891				
Methanol	**0.0334***		**0.0074***	0.1430	0.4557
Glycine	**0.0105***	**0.0289***	**0.0156***	0.4864	**0.0069***
Methyl-histidine	0.2133				
Guanidoacetate	0.2025				
Glycolate	0.2794				
Creatinine	0.1688				
Trans-aconitate	0.9843				
Hippurate	**<0.0001***	**0.0447***	**<0.0001***	**0.0003***	0.1640
Formate	**0.0141***	**0.0474***	0.5671	**0.0056***	**0.0243***
Trigonelline	0.0588				

[a] As measured by Kruskall Wallis; [b] p value corrected post-multiple comparison testing

[c] As measured by Mann-Whitney *U* test; ***bold** statistically significant

White Northern Europeans	p value across all three groups [a]	q value [b]	HC : CD [c]	HC : UC [c]	UC : CD [c]
Lactate	0.4098				
Alanine	**0.0265***		**0.0100***	0.0517	0.7575
Acetate	0.8974				
4-cresol	0.3043				
Succinate	0.3695				
Citrate	**0.0353***		0.5140	0.0572	**0.0114***
Dimethylamine	0.7462				
Creatine	0.1750				
Trimethylamine-N-oxide	0.4648				
Methanol	**0.0134***	**0.0263***	**0.0031***	0.1027	0.3578
Glycine	**0.0154***	**0.0289***	**0.0056***	0.5198	0.0536
Methyl-histidine	0.2252				
Guanidoacetate	0.5549				
Glycolate	0.8482				
Creatinine	0.1485				
Trans-aconitate	0.9402				
Hippurate	**<0.0001***	**0.0447***	**<0.001***	0.1317	**0.0024***
Formate	0.2246				
Trigonelline	0.1677				

[a] As measured by Kruskall Wallis; [b] p value corrected post-multiple comparison testing

[c] As measured by Mann-Whitney *U* test; ***bold** statistically significant

South Asians	p value across all three groups [a]	q value [b]	HC : CD [c]	HC : UC [c]	UC : CD [c]
Lactate	0.5600				
Alanine	0.2067				
Acetate	0.7719				
4-cresol	0.2248				
Succinate	**0.0055***	**0.0132***	**0.0013***	0.0701	0.0644
Citrate	0.2937				
Dimethylamine	0.3265				
Creatine	0.3633				
Trimethylamine-N-oxide	**0.0046***	**0.0237***	**0.0007***	0.1127	0.0368*
Methanol	0.6230				
Glycine	0.4538				
Methyl-histidine	0.8000				
Guanidoacetate	0.3957				
Glycolate	0.6324				
Creatinine	0.3452				
Trans-aconitate	0.6727				
Hippurate	**0.0022***	**0.0447**	**0.0004***	0.0508	0.0459*
Formate	**0.0068***	0.0474*	0.0722	**0.0019***	0.6604
Trigonelline	0.3226				

[a] As measured by Kruskall Wallis; [b] p value corrected post-multiple comparison testing

[c] As measured by Mann-Whitney *U* test; *statistically significant; **bold** remains significant post multiple comparison testing

P values for each metabolite comparing between ethnicities			
	NE : South Asian Crohn's [a]	NE : South Asian UC [a]	NE : South Asian Healthy controls [a]
Lactate	0.5274	0.1162	**0.0216***
Alanine	0.9965	**0.0058***	**<0.0001***
Acetate	0.4146	0.4782	0.5996
4-cresol	**0.0019***	**0.0003***	**<0.0001***
Succinate	**0.0099***	**0.0114***	0.6939
Citrate	0.9133	0.1693	**0.0397***
Dimethylamine	0.5800	0.0677	0.1525
Creatine	0.5505	0.1004	0.7637
Trimethylamine-N-oxide	**0.0001***	**0.0336***	0.1167
Methanol	0.4401	0.6910	**0.0016***
Glycine	0.1648	0.7691	0.0808
Methyl-histidine	0.7113	0.1973	0.6227
Guanidoacetate	0.4827	0.6253	0.6979
Glycolate	0.8652	0.8141	0.3139
Creatinine	0.1846	0.0938	0.7637
Trans-aconitate	0.9341	0.0925	0.2390
Hippurate	**0.0016***	**0.0011***	**0.0032***
Formate	0.4046	0.6253	**0.0229***
Trigonelline	0.5622	0.5279	0.3964

[a] As measured by Kruskall Wallis

Supplement 4: All p values and corrected p values (q values) for comparisons across 3 groups, and between two diagnoses for all ethnicities and separately for white Europeans and South Asians. P values for each metabolite comparing groups of different ethnicities.

References

1. Van Assche G, Dignass A, Panes J, Beaugerie L, Karagiannis J, Allez M, et al. The second European evidence-based Consensus on the diagnosis and management of Crohn's disease: Definitions and diagnosis. Journal of Crohn's & colitis. 2010;4(1):7-27.
2. Dignass A, Eliakim R, Magro F, Maaser C, Chowers Y, Geboes K, et al. Second European evidence-based consensus on the diagnosis and management of ulcerative colitis part 1: definitions and diagnosis. Journal of Crohn's & colitis. 2012;6(10):965-90.
3. Matsuoka K, Kobayashi T, Ueno F, Matsui T, Hirai F, Inoue N, et al. Evidence-based clinical practice guidelines for inflammatory bowel disease. J Gastroenterol. 2018;53(3):305-53.
4. Silverberg MS, Satsangi J, Ahmad T, Arnott ID, Bernstein CN, Brant SR, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. Canadian journal of gastroenterology = Journal canadien de gastroenterologie. 2005;19 Suppl A:5a-36a.
5. Tremaine WJ. Is indeterminate colitis determinable? Curr Gastroenterol Rep. 2012;14(2):162-5.
6. Sica GS, Biancone L. Surgery for inflammatory bowel disease in the era of laparoscopy. World journal of gastroenterology : WJG. 2013;19(16):2445-8.
7. Larson DW, Pemberton JH. Current concepts and controversies in surgery for IBD. Gastroenterology. 2004;126(6):1611-9.
8. Mowat C, Cole A, Windsor A, Ahmad T, Arnott I, Driscoll R, et al. Guidelines for the management of inflammatory bowel disease in adults. Gut. 2011;60(5):571-607.
9. Eaden J. Review article: the data supporting a role for aminosalicylates in the chemoprevention of colorectal cancer in patients with inflammatory bowel disease. Alimentary pharmacology & therapeutics. 2003;18 Suppl 2:15-21.
10. Timmer A, McDonald JW, Tsoulis DJ, Macdonald JK. Azathioprine and 6-mercaptopurine for maintenance of remission in ulcerative colitis. Cochrane Database Syst Rev. 2012;9:CD000478.
11. Prefontaine E, Macdonald JK, Sutherland LR. Azathioprine or 6-mercaptopurine for induction of remission in Crohn's disease. Cochrane Database Syst Rev. 2009(4):CD000545.
12. Jones GR, Kennedy NA, Lees CW, Arnott ID, Satsangi J. Systematic review: The use of thiopurines or anti-TNF in post-operative Crohn's disease maintenance--progress and prospects. Alimentary pharmacology & therapeutics. 2014;39(11):1253-65.
13. Cohen BL, Sachar DB. Update on anti-tumor necrosis factor agents and other new drugs for inflammatory bowel disease. BMJ. 2017;357:j2505.
14. Jauregui-Amezaga A, Somers M, De Schepper H, Macken E. Next generation of biologics for the treatment of Crohn's disease: an evidence-based review on ustekinumab. Clin Exp Gastroenterol. 2017;10:293-301.
15. Duijvestein M, Battat R, Vande Casteele N, D'Haens GR, Sandborn WJ, Khanna R, et al. Novel Therapies and Treatment Strategies for Patients with Inflammatory Bowel Disease. Curr Treat Options Gastroenterol. 2018;16(1):129-46.
16. Rutgeerts P, Sandborn WJ, Feagan BG, Reinisch W, Olson A, Johanns J, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. The New England journal of medicine. 2005;353(23):2462-76.
17. Feagan BG, Panaccione R, Sandborn WJ, D'Haens GR, Schreiber S, Rutgeerts PJ, et al. Effects of adalimumab therapy on incidence of hospitalization and surgery in Crohn's disease: results from the CHARM study. Gastroenterology. 2008;135(5):1493-9.
18. Singh JA, Wells GA, Christensen R, Tanjong Ghogomu E, Maxwell L, Macdonald JK, et al. Adverse effects of biologics: a network meta-analysis and Cochrane overview. Cochrane Database Syst Rev. 2011(2):CD008794.
19. Peyrin-Biroulet L, Loftus EV, Jr., Colombel JF, Sandborn WJ. Long-term complications, extraintestinal manifestations, and mortality in adult Crohn's disease in population-based cohorts. Inflammatory bowel diseases. 2011;17(1):471-8.

20. D'Haens G, Baert F, van Assche G, Caenepeel P, Vergauwe P, Tuynman H, et al. Early combined immunosuppression or conventional management in patients with newly diagnosed Crohn's disease: an open randomised trial. Lancet. 2008;371(9613):660-7.
21. Hoekman DR, Stibbe JA, Baert FJ, Caenepeel P, Vergauwe P, De Vos M, et al. Long-Term Outcome of Early Combined Immunosuppression Versus Conventional Management in Newly Diagnosed Crohn's Disease. Journal of Crohn's & colitis. 2018.
22. Im JP, Ye BD, Kim YS, Kim JS. Changing treatment paradigms for the management of inflammatory bowel disease. Korean J Intern Med. 2018;33(1):28-35.
23. Norouzinia M, Chaleshi V, Alizadeh AHM, Zali MR. Biomarkers in inflammatory bowel diseases: insight into diagnosis, prognosis and treatment. Gastroenterol Hepatol Bed Bench. 2017;10(3):155-67.
24. Carroll MW, Hamilton Z, Gill H, Simkin J, Smyth M, Espinosa V, et al. Pediatric Inflammatory Bowel Disease Among South Asians Living in British Columbia, Canada: A Distinct Clinical Phenotype. Inflammatory bowel diseases. 2016;22(2):387-96.
25. Sofia MA, Rubin DT, Hou N, Pekow J. Clinical presentation and disease course of inflammatory bowel disease differs by race in a large tertiary care hospital. Digestive diseases and sciences. 2014;59(9):2228-35.
26. Ungar B, Haj-Natour O, Kopylov U, Yavzori M, Fudim E, Picard O, et al. Ashkenazi Jewish origin protects against formation of antibodies to infliximab and therapy failure. Medicine (Baltimore). 2015;94(18):e673.
27. Nishida A, Inoue R, Inatomi O, Bamba S, Naito Y, Andoh A. Gut microbiota in the pathogenesis of inflammatory bowel disease. Clin J Gastroenterol. 2018;11(1):1-10.
28. Zhou M, He J, Shen Y, Zhang C, Wang J, Chen Y. New Frontiers in Genetics, Gut Microbiota, and Immunity: A Rosetta Stone for the Pathogenesis of Inflammatory Bowel Disease. Biomed Res Int. 2017;2017:8201672.
29. Fiocchi C. Inflammatory bowel disease pathogenesis: where are we? Journal of gastroenterology and hepatology. 2015;30 Suppl 1:12-8.
30. Orholm M, Munkholm P, Langholz E, Nielsen OH, Sorensen TI, Binder V. Familial occurrence of inflammatory bowel disease. The New England journal of medicine. 1991;324(2):84-8.
31. Brant SR. Update on the heritability of inflammatory bowel disease: the importance of twin studies. Inflammatory bowel diseases. 2011;17(1):1-5.
32. Satsangi J, Welsh KI, Bunce M, Julier C, Farrant JM, Bell JI, et al. Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. Lancet. 1996;347(9010):1212-7.
33. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature. 2001;411(6837):599-603.
34. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119-24.
35. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. 2015;47(9):979-86.
36. Zhang YZ, Li YY. Inflammatory bowel disease: pathogenesis. World journal of gastroenterology : WJG. 2014;20(1):91-9.
37. Cammarota G, Ianiro G, Cianci R, Bibbo S, Gasbarrini A, Curro D. The involvement of gut microbiota in inflammatory bowel disease pathogenesis: potential for therapy. Pharmacology & therapeutics. 2015;149:191-212.
38. Ventham NT, Kennedy NA, Nimmo ER, Satsangi J. Beyond gene discovery in inflammatory bowel disease: the emerging role of epigenetics. Gastroenterology. 2013;145(2):293-308.
39. Petronis A, Petroniene R. Epigenetics of inflammatory bowel disease. Gut. 2000;47(2):302-6.

40. Cosnes J. Tobacco and IBD: relevance in the understanding of disease mechanisms and clinical practice. Best practice & research Clinical gastroenterology. 2004;18(3):481-96.
41. Mawdsley JE, Rampton DS. The role of psychological stress in inflammatory bowel disease. Neuroimmunomodulation. 2006;13(5-6):327-36.
42. Koutroubakis IE, Vlachonikolis IG, Kouroumalis EA. Role of appendicitis and appendectomy in the pathogenesis of ulcerative colitis: a critical review. Inflammatory bowel diseases. 2002;8(4):277-86.
43. Ananthakrishnan AN, Higuchi LM, Huang ES, Khalili H, Richter JM, Fuchs CS, et al. Aspirin, nonsteroidal anti-inflammatory drug use, and risk for Crohn disease and ulcerative colitis: a cohort study. Ann Intern Med. 2012;156(5):350-9.
44. Leslie WD, Miller N, Rogala L, Bernstein CN. Vitamin D status and bone density in recently diagnosed inflammatory bowel disease: the Manitoba IBD Cohort Study. The American journal of gastroenterology. 2008;103(6):1451-9.
45. Reese GE, Nanidis T, Borysiewicz C, Yamamoto T, Orchard T, Tekkis PP. The effect of smoking after surgery for Crohn's disease: a meta-analysis of observational studies. Int J Colorectal Dis. 2008;23(12):1213-21.
46. Lakatos PL, Szamosi T, Lakatos L. Smoking in inflammatory bowel diseases: good, bad or ugly? World journal of gastroenterology : WJG. 2007;13(46):6134-9.
47. 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. 2012;142(1):46-54.e42; quiz e30.
48. Lakatos PL. Recent trends in the epidemiology of inflammatory bowel diseases: up or down? World journal of gastroenterology : WJG. 2006;12(38):6102-8.
49. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, et al. Evolution of mammals and their gut microbes. Science. 2008;320(5883):1647-51.
50. Kitahora T, Utsunomiya T, Yokota A. Epidemiological study of ulcerative colitis in Japan: incidence and familial occurrence. The Epidemiology Group of the Research Committee of Inflammatory Bowel Disease in Japan. J Gastroenterol. 1995;30 Suppl 8:5-8.
51. Devkota S, Wang Y, Musch MW, Leone V, Fehlner-Peach H, Nadimpalli A, et al. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. Nature. 2012;487(7405):104-8.
52. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(33):14691-6.
53. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci Transl Med. 2009;1(6):6ra14.
54. Lee D, Albenberg L, Compher C, Baldassano R, Piccoli D, Lewis JD, et al. Diet in the pathogenesis and treatment of inflammatory bowel diseases. Gastroenterology. 2015;148(6):1087-106.
55. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, et al. Host-gut microbiota metabolic interactions. Science. 2012;336(6086):1262-7.
56. Dominguez-Bello MG, Blaser MJ, Ley RE, Knight R. Development of the human gastrointestinal microbiota and insights from high-throughput sequencing. Gastroenterology. 2011;140(6):1713-9.
57. Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, et al. Metagenomic Analysis of the Human Distal Gut Microbiome. Science. 2006;312(5778):1355-9.
58. Solga SF. Breath volatile organic compounds for the gut-fatty liver axis: promise, peril, and path forward. World journal of gastroenterology : WJG. 2014;20(27):9017-25.
59. Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. Cell. 2005;122(1):107-18.

60. Backhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, et al. The gut microbiota as an environmental factor that regulates fat storage. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(44):15718-23.
61. Lindenbaum J, Rund DG, Butler VP, Jr., Tse-Eng D, Saha JR. Inactivation of digoxin by the gut flora: reversal by antibiotic therapy. The New England journal of medicine. 1981;305(14):789-94.
62. Nicholson JK, Holmes E, Wilson ID. Gut microorganisms, mammalian metabolism and personalized health care. Nat Rev Microbiol. 2005;3(5):431-8.
63. Lynch SV, Pedersen O. The Human Intestinal Microbiome in Health and Disease. The New England journal of medicine. 2016;375(24):2369-79.
64. Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. Cell. 2012;148(6):1258-70.
65. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, et al. Gut microbiota composition correlates with diet and health in the elderly. Nature. 2012;488(7410):178-84.
66. Albenberg LG, Wu GD. Diet and the intestinal microbiome: associations, functions, and implications for health and disease. Gastroenterology. 2014;146(6):1564-72.
67. Brusca SB, Abramson SB, Scher JU. Microbiome and mucosal inflammation as extra-articular triggers for rheumatoid arthritis and autoimmunity. Curr Opin Rheumatol. 2014;26(1):101-7.
68. Sartor RB. Microbial influences in inflammatory bowel diseases. Gastroenterology. 2008;134(2):577-94.
69. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. Nature. 2012;489(7415):220-30.
70. Rose JD, Roberts GM, Williams G, Mayberry JF, Rhodes J. Cardiff Crohn's disease jubilee: the incidence over 50 years. Gut. 1988;29(3):346-51.
71. Martin-de-Carpi J, Rodriguez A, Ramos E, Jimenez S, Martinez-Gomez MJ, Medina E, et al. Increasing incidence of pediatric inflammatory bowel disease in Spain (1996-2009): the SPIRIT Registry. Inflammatory bowel diseases. 2013;19(1):73-80.
72. Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology. 2014;146(6):1489-99.
73. Jernberg C, Lofmark S, Edlund C, Jansson JK. Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. ISME J. 2007;1(1):56-66.
74. Jakobsson HE, Abrahamsson TR, Jenmalm MC, Harris K, Quince C, Jernberg C, et al. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. Gut. 2014;63(4):559-66.
75. Li J, Butcher J, Mack D, Stintzi A. Functional impacts of the intestinal microbiome in the pathogenesis of inflammatory bowel disease. Inflammatory bowel diseases. 2015;21(1):139-53.
76. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, et al. Moving pictures of the human microbiome. Genome Biol. 2011;12(5):R50.
77. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207-14.
78. Peterson DA, Frank DN, Pace NR, Gordon JI. Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases. Cell host & microbe. 2008;3(6):417-27.
79. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. Science. 2005;308(5728):1635-8.
80. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature. 2011;473(7346):174-80.
81. Lepage P, Seksik P, Sutren M, de la Cochetiere MF, Jian R, Marteau P, et al. Biodiversity of the mucosa-associated microbiota is stable along the distal digestive tract in healthy individuals and patients with IBD. Inflammatory bowel diseases. 2005;11(5):473-80.
82. Wang W, Chen L, Zhou R, Wang X, Song L, Huang S, et al. Increased proportions of Bifidobacterium and the Lactobacillus group and loss of butyrate-producing bacteria in inflammatory bowel disease. Journal of clinical microbiology. 2014;52(2):398-406.

83. DeGruttola AK, Low D, Mizoguchi A, Mizoguchi E. Current Understanding of Dysbiosis in Disease in Human and Animal Models. Inflammatory bowel diseases. 2016;22(5):1137-50.
84. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486(7402):222-7.
85. Lepage P, Hasler R, Spehlmann ME, Rehman A, Zvirbliene A, Begun A, et al. Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. Gastroenterology. 2011;141(1):227-36.
86. Sokol H, Lay C, Seksik P, Tannock GW. Analysis of bacterial bowel communities of IBD patients: what has it revealed? Inflammatory bowel diseases. 2008;14(6):858-67.
87. Dicksved J, Halfvarson J, Rosenquist M, Jarnerot G, Tysk C, Apajalahti J, et al. Molecular analysis of the gut microbiota of identical twins with Crohn's disease. ISME J. 2008;2(7):716-27.
88. Sepehri S, Kotlowski R, Bernstein CN, Krause DO. Microbial diversity of inflamed and noninflamed gut biopsy tissues in inflammatory bowel disease. Inflammatory bowel diseases. 2007;13(6):675-83.
89. Walker AW, Sanderson JD, Churcher C, Parkes GC, Hudspith BN, Rayment N, et al. High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. BMC Microbiol. 2011;11:7.
90. Andoh A, Tsujikawa T, Sasaki M, Mitsuyama K, Suzuki Y, Matsui T, et al. Faecal microbiota profile of Crohn's disease determined by terminal restriction fragment length polymorphism analysis. Alimentary pharmacology & therapeutics. 2009;29(1):75-82.
91. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(34):13780-5.
92. Frank DN, Robertson CE, Hamm CM, Kpadeh Z, Zhang T, Chen H, et al. Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. Inflammatory bowel diseases. 2011;17(1):179-84.
93. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012;13(9):R79.
94. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermudez-Humaran 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. 2008;105(43):16731-6.
95. Swidsinski A, Loening-Baucke V, Vaneechoutte M, Doerffel Y. Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. Inflammatory bowel diseases. 2008;14(2):147-61.
96. Joossens M, Huys G, Cnockaert M, De Preter V, Verbeke K, Rutgeerts P, et al. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. Gut. 2011;60(5):631-7.
97. Sokol H, Seksik P, Furet JP, Firmesse O, Nion-Larmurier I, Beaugerie L, et al. Low counts of Faecalibacterium prausnitzii in colitis microbiota. Inflammatory bowel diseases. 2009;15(8):1183-9.
98. Segain JP, Raingeard de la Bletiere D, Bourreille A, Leray V, Gervois N, Rosales C, et al. Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease. Gut. 2000;47(3):397-403.
99. Sellon RK, Tonkonogy S, Schultz M, Dieleman LA, Grenther W, Balish E, et al. Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. Infection and immunity. 1998;66(11):5224-31.
100. Flanagan PK, Campbell BJ, Rhodes JM. Lessons from diversion studies and antibacterial interventions. Dig Dis. 2012;30(4):347-50.
101. Janowitz HD, Croen EC, Sachar DB. The role of the fecal stream in Crohn's disease: an historical and analytic review. Inflammatory bowel diseases. 1998;4(1):29-39.

102. Casellas F, Borruel N, Papo M, Guarner F, Antolin M, Videla S, et al. Antiinflammatory effects of enterically coated amoxicillin-clavulanic acid in active ulcerative colitis. Inflammatory bowel diseases. 1998;4(1):1-5.
103. Sartor RB. Therapeutic manipulation of the enteric microflora in inflammatory bowel diseases: antibiotics, probiotics, and prebiotics. Gastroenterology. 2004;126(6):1620-33.
104. Biswas A, Petnicki-Ocwieja T, Kobayashi KS. Nod2: a key regulator linking microbiota to intestinal mucosal immunity. Journal of molecular medicine. 2012;90(1):15-24.
105. Tamboli CP, Neut C, Desreumaux P, Colombel JF. Dysbiosis as a prerequisite for IBD. Gut. 2004;53(7):1057.
106. Erickson AR, Cantarel BL, Lamendella R, Darzi Y, Mongodin EF, Pan C, et al. Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. PloS one. 2012;7(11):e49138.
107. Lenz EM, Bright J, Wilson ID, Hughes A, Morrisson J, Lindberg H, et al. Metabonomics, dietary influences and cultural differences: a 1H NMR-based study of urine samples obtained from healthy British and Swedish subjects. Journal of pharmaceutical and biomedical analysis. 2004;36(4):841-9.
108. Le Gall G. NMR spectroscopy of biofluids and extracts. Methods in molecular biology (Clifton, NJ). 2015;1277:29-36.
109. Fan TW, Lorkiewicz PK, Sellers K, Moseley HN, Higashi RM, Lane AN. Stable isotope-resolved metabolomics and applications for drug development. Pharmacology & therapeutics. 2012;133(3):366-91.
110. Lindon JC, Nicholson JK, Holmes E, Antti H, Bollard ME, Keun H, et al. Contemporary issues in toxicology the role of metabonomics in toxicology and its evaluation by the COMET project. Toxicology and Applied Pharmacology. 2003;187(3):137-46.
111. Johnson CH, Gonzalez FJ. Challenges and opportunities of metabolomics. Journal of cellular physiology. 2012;227(8):2975-81.
112. De Preter V. Metabolomics in the Clinical Diagnosis of Inflammatory Bowel Disease. Dig Dis. 2015;33 Suppl 1:2-10.
113. Bjerrum JT, Wang Y, Hao F, Coskun M, Ludwig C, Gunther U, et al. Metabonomics of human fecal extracts characterize ulcerative colitis, Crohn's disease and healthy individuals. Metabolomics. 2015;11:122-33.
114. 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 & colitis. 2013;7(2):e42-8.
115. Schicho R, Shaykhutdinov R, Ngo J, Nazyrova A, Schneider C, Panaccione R, et al. Quantitative metabolomic profiling of serum, plasma, and urine by (1)H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals. Journal of proteome research. 2012;11(6):3344-57.
116. Dawiskiba T, Deja S, Mulak A, Zabek A, Jawien E, Pawelka D, et al. Serum and urine metabolomic fingerprinting in diagnostics of inflammatory bowel diseases. World journal of gastroenterology : WJG. 2014;20(1):163-74.
117. 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. 2007;6(2):546-51.
118. Williams HR, Cox IJ, Walker DG, North BV, Patel VM, Marshall SE, et al. Characterization of inflammatory bowel disease with urinary metabolic profiling. The American journal of gastroenterology. 2009;104(6):1435-44.
119. Williams HR, Willsmore JD, Cox IJ, Walker DG, Cobbold JF, Taylor-Robinson SD, et al. Serum metabolic profiling in inflammatory bowel disease. Digestive diseases and sciences. 2012;57(8):2157-65.

120. Lenz EM, Bright J, Wilson ID, Morgan SR, Nash AF. A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. Journal of pharmaceutical and biomedical analysis. 2003;33(5):1103-15.
121. 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. Inflamm Res. 2011;60(9):831-40.
122. Hisamatsu T, Okamoto S, Hashimoto M, Muramatsu T, Andou A, Uo M, et al. Novel, objective, multivariate biomarkers composed of plasma amino acid profiles for the diagnosis and assessment of inflammatory bowel disease. PloS one. 2012;7(1):e31131.
123. Murdoch TB, Fu H, MacFarlane S, Sydora BC, Fedorak RN, Slupsky CM. Urinary metabolic profiles of inflammatory bowel disease in interleukin-10 gene-deficient mice. Analytical chemistry. 2008;80(14):5524-31.
124. Schicho R, Nazyrova A, Shaykhutdinov R, Duggan G, Vogel HJ, Storr M. Quantitative metabolomic profiling of serum and urine in DSS-induced ulcerative colitis of mice by (1)H NMR spectroscopy. Journal of proteome research. 2010;9(12):6265-73.
125. De Preter V, Verbeke K. Metabolomics as a diagnostic tool in gastroenterology. World J Gastrointest Pharmacol Ther. 2013;4(4):97-107.
126. Le Gall G, Noor SO, Ridgway K, Scovell L, Jamieson C, Johnson IT, et al. Metabolomics of fecal extracts detects altered metabolic activity of gut microbiota in ulcerative colitis and irritable bowel syndrome. Journal of proteome research. 2011;10(9):4208-18.
127. Jansson J, Willing B, Lucio M, Fekete A, Dicksved J, Halfvarson J, et al. Metabolomics reveals metabolic biomarkers of Crohn's disease. PloS one. 2009;4(7):e6386.
128. Santoru ML, Piras C, Murgia A, Palmas V, Camboni T, Liggi S, et al. Cross sectional evaluation of the gut-microbiome metabolome axis in an Italian cohort of IBD patients. Sci Rep. 2017;7(1):9523.
129. Chapman MA, Grahn MF, Boyle MA, Hutton M, Rogers J, Williams NS. Butyrate oxidation is impaired in the colonic mucosa of sufferers of quiescent ulcerative colitis. Gut. 1994;35(1):73-6.
130. Tedelind S, Westberg F, Kjerrulf M, Vidal A. Anti-inflammatory properties of the short-chain fatty acids acetate and propionate: a study with relevance to inflammatory bowel disease. World journal of gastroenterology : WJG. 2007;13(20):2826-32.
131. Nicholls AW, Mortishire-Smith RJ, Nicholson JK. NMR spectroscopic-based metabonomic studies of urinary metabolite variation in acclimatizing germ-free rats. Chem Res Toxicol. 2003;16(11):1395-404.
132. Bjerrum JT, Nielsen OH, Hao F, Tang H, Nicholson JK, Wang Y, et al. Metabonomics in ulcerative colitis: diagnostics, biomarker identification, and insight into the pathophysiology. Journal of proteome research. 2010;9(2):954-62.
133. Williams HR, Cox IJ, Walker DG, Cobbold JF, Taylor-Robinson SD, Marshall SE, et al. Differences in gut microbial metabolism are responsible for reduced hippurate synthesis in Crohn's disease. BMC gastroenterology. 2010;10:108.
134. Li M, Wang B, Zhang M, Rantalainen M, Wang S, Zhou H, et al. Symbiotic gut microbes modulate human metabolic phenotypes. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(6):2117-22.
135. Patel KP, Luo FJ, Plummer NS, Hostetter TH, Meyer TW. The production of p-cresol sulfate and indoxyl sulfate in vegetarians versus omnivores. Clin J Am Soc Nephrol. 2012;7(6):982-8.
136. Leonhartsberger S, Korsa I, Bock A. The molecular biology of formate metabolism in enterobacteria. J Mol Microbiol Biotechnol. 2002;4(3):269-76.
137. Seksik P, Rigottier-Gois L, Gramet G, Sutren M, Pochart P, Marteau P, et al. Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. Gut. 2003;52(2):237-42.
138. Yau Y, Leong RW, Zeng M, Wasinger VC. Proteomics and metabolomics in inflammatory bowel disease. Journal of gastroenterology and hepatology. 2013;28(7):1076-86.

139. Gruber L, Lichti P, Rath E, Haller D. Nutrigenomics and nutrigenetics in inflammatory bowel diseases. J Clin Gastroenterol. 2012;46(9):735-47.
140. Amann A, Costello Bde L, Miekisch W, Schubert J, Buszewski B, Pleil J, et al. The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. Journal of breath research. 2014;8(3):034001.
141. Probert CS, Jones PR, Ratcliffe NM. A novel method for rapidly diagnosing the causes of diarrhoea. Gut. 2004;53(1):58-61.
142. Probert CS, Ahmed I, Khalid T, Johnson E, Smith S, Ratcliffe N. Volatile organic compounds as diagnostic biomarkers in gastrointestinal and liver diseases. J Gastrointestin Liver Dis. 2009;18(3):337-43.
143. Walton C, Fowler DP, Turner C, Jia W, Whitehead RN, Griffiths L, et al. Analysis of volatile organic compounds of bacterial origin in chronic gastrointestinal diseases. Inflammatory bowel diseases. 2013;19(10):2069-78.
144. Ahmed I, Greenwood R, Costello Bde L, Ratcliffe NM, Probert CS. An investigation of fecal volatile organic metabolites in irritable bowel syndrome. PloS one. 2013;8(3):e58204.
145. De Preter V, Machiels K, Joossens M, Arijs I, Matthys C, Vermeire S, et al. Faecal metabolite profiling identifies medium-chain fatty acids as discriminating compounds in IBD. Gut. 2015;64(3):447-58.
146. Ahmed I, Greenwood R, Costello B, Ratcliffe N, Probert CS. Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease. Alimentary pharmacology & therapeutics. 2016;43(5):596-611.
147. Lewis VJ, Moss CW, Jones WL. Determination of volatile acid production of Clostridium by gas chromatography. Can J Microbiol. 1967;13(8):1033-40.
148. Phillips KD, Tearle PV, Willis AT. Rapid diagnosis of anaerobic infections by gas-liquid chromatography of clinical material. Journal of clinical pathology. 1976;29(5):428-32.
149. Garner CE, Smith S, de Lacy Costello B, White P, Spencer R, Probert CS, et al. Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. FASEB J. 2007;21(8):1675-88.
150. Zentek J, Buchheit-Renko S, Manner K, Pieper R, Vahjen W. Intestinal concentrations of free and encapsulated dietary medium-chain fatty acids and effects on gastric microbial ecology and bacterial metabolic products in the digestive tract of piglets. Arch Anim Nutr. 2012;66(1):14-26.
151. Arasaradnam RP, Ouaret N, Thomas MG, Quraishi N, Heatherington E, Nwokolo CU, et al. A novel tool for noninvasive diagnosis and tracking of patients with inflammatory bowel disease. Inflammatory bowel diseases. 2013;19(5):999-1003.
152. Pauling L, Robinson AB, Teranishi R, Cary P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. Proceedings of the National Academy of Sciences of the United States of America. 1971;68(10):2374-6.
153. Smith D, Turner C, Spanel P. Volatile metabolites in the exhaled breath of healthy volunteers: their levels and distributions. Journal of breath research. 2007;1(1):014004.
154. Amann A, Spanel P, Smith D. Breath analysis: the approach towards clinical applications. Mini reviews in medicinal chemistry. 2007;7(2):115-29.
155. Phillips M, Greenberg J. Ion-trap detection of volatile organic compounds in alveolar breath. Clinical chemistry. 1992;38(1):60-5.
156. Manolis A. The diagnostic potential of breath analysis. Clinical chemistry. 1983;29(1):5-15.
157. Dallinga JW, Smolinska A, van Schooten FJ. Analysis of volatile organic compounds in exhaled breath by gas chromatography-mass spectrometry combined with chemometric analysis. Methods in molecular biology (Clifton, NJ). 2014;1198:251-63.
158. Smith D, Spanel P. Selected ion flow tube mass spectrometry (SIFT-MS) for on-line trace gas analysis. Mass spectrometry reviews. 2005;24(5):661-700.
159. Boots AW, van Berkel JJ, Dallinga JW, Smolinska A, Wouters EF, van Schooten FJ. The versatile use of exhaled volatile organic compounds in human health and disease. Journal of breath research. 2012;6(2):027108.

160. Paschke KM, Mashir A, Dweik RA. Clinical applications of breath testing. F1000 medicine reports. 2010;2:56.
161. Miekisch W, Schubert JK, Noeldge-Schomburg GFE. Diagnostic potential of breath analysis—focus on volatile organic compounds. Clinica Chimica Acta. 2004;347(1–2):25-39.
162. Buszewski B, Kesy M, Ligor T, Amann A. Human exhaled air analytics: biomarkers of diseases. Biomed Chromatogr. 2007;21(6):553-66.
163. Kumar S, Huang J, Abbassi-Ghadi N, Mackenzie HA, Veselkov KA, Hoare JM, et al. Mass Spectrometric Analysis of Exhaled Breath for the Identification of Volatile Organic Compound Biomarkers in Esophageal and Gastric Adenocarcinoma. Annals of surgery. 2015.
164. Alkhouri N, Cikach F, Eng K, Moses J, Patel N, Yan C, et al. Analysis of breath volatile organic compounds as a noninvasive tool to diagnose nonalcoholic fatty liver disease in children. European journal of gastroenterology & hepatology. 2014;26(1):82-7.
165. Mazzatenta A, Pokorski M, Sartucci F, Domenici L, Di Giulio C. Volatile organic compounds (VOCs) fingerprint of Alzheimer's disease. Respiratory physiology & neurobiology. 2015;209:81-4.
166. Nakhleh MK, Badarny S, Winer R, Jeries R, Finberg J, Haick H. Distinguishing idiopathic Parkinson's disease from other parkinsonian syndromes by breath test. Parkinsonism & related disorders. 2015;21(2):150-3.
167. Guo L, Wang C, Chi C, Wang X, Liu S, Zhao W, et al. Exhaled breath volatile biomarker analysis for thyroid cancer. Translational research : the journal of laboratory and clinical medicine. 2015.
168. Hamer HM, De Preter V, Windey K, Verbeke K. Functional analysis of colonic bacterial metabolism: relevant to health? Am J Physiol Gastrointest Liver Physiol. 2012;302(1):G1-9.
169. Kokoszka J, Nelson RL, Swedler WI, Skosey J, Abcarian H. Determination of inflammatory bowel disease activity by breath pentane analysis. Diseases of the colon and rectum. 1993;36(6):597-601.
170. Sedghi S, Keshavarzian A, Klamut M, Eiznhamer D, Zarling EJ. Elevated breath ethane levels in active ulcerative colitis: evidence for excessive lipid peroxidation. The American journal of gastroenterology. 1994;89(12):2217-21.
171. Pelli MA, Trovarelli G, Capodicasa E, De Medio GE, Bassotti G. Breath alkanes determination in ulcerative colitis and Crohn's disease. Diseases of the colon and rectum. 1999;42(1):71-6.
172. Dryahina K, Spanel P, Pospisilova V, Sovova K, Hrdlicka L, Machkova N, et al. Quantification of pentane in exhaled breath, a potential biomarker of bowel disease, using selected ion flow tube mass spectrometry. Rapid communications in mass spectrometry : RCM. 2013;27(17):1983-92.
173. Dryahina K, Smith D, Bortlik M, Machkova N, Lukas M, Spanel P. Pentane and other volatile organic compounds, including carboxylic acids, in the exhaled breath of patients with Crohn's disease and ulcerative colitis. Journal of breath research. 2017;12(1):016002.
174. Koek GH, Verleden GM, Evenepoel P, Rutgeerts P. Activity related increase of exhaled nitric oxide in Crohn's disease and ulcerative colitis: a manifestation of systemic involvement? Respiratory medicine. 2002;96(7):530-5.
175. McKay LF, Eastwood MA, Brydon WG. Methane excretion in man--a study of breath, flatus, and faeces. Gut. 1985;26(1):69-74.
176. Nitta H, Kinoyama M, Teramoto F, Watanabe A, Koga H, Haruma K, et al. Exhaled carbon monoxide concentration is not elevated in patients with inflammatory bowel disease. Clinical and experimental medicine. 2007;7(2):77-81.
177. Patel N, Alkhouri N, Eng K, Cikach F, Mahajan L, Yan C, et al. Metabolomic analysis of breath volatile organic compounds reveals unique breathprints in children with inflammatory bowel disease: a pilot study. Alimentary pharmacology & therapeutics. 2014;40(5):498-507.
178. Arasaradnam RP, McFarlane M, Daulton E, Skinner J, O'Connell N, Wurie S, et al. Non-invasive exhaled volatile organic biomarker analysis to detect inflammatory bowel disease (IBD). Dig Liver Dis. 2016;48(2):148-53.

179. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet. 2018;390(10114):2769-78.
180. Hou JK, El-Serag H, Thirumurthi S. Distribution and manifestations of inflammatory bowel disease in Asians, Hispanics, and African Americans: a systematic review. The American journal of gastroenterology. 2009;104(8):2100-9.
181. Walker DG, Williams HR, Kane SP, Mawdsley JE, Arnold J, McNeil I, et al. Differences in inflammatory bowel disease phenotype between South Asians and Northern Europeans living in North West London, UK. The American journal of gastroenterology. 2011;106(7):1281-9.
182. Risch N, Burchard E, Ziv E, Tang H. Categorization of humans in biomedical research: genes, race and disease. Genome Biol. 2002;3(7):comment2007.
183. Schwartz RS. Racial profiling in medical research. The New England journal of medicine. 2001;344(18):1392-3.
184. Schmitt E. For 7 million people in census, one race category isn't enough. New York Times. 2001.
185. Office for National Statistics. Ethnicity and National Identity in England and Wales:2011 2011 [
186. Probert CS, Jayanthi V, Pinder D, Wicks AC, Mayberry JF. Epidemiological study of ulcerative proctocolitis in Indian migrants and the indigenous population of Leicestershire. Gut. 1992;33(5):687-93.
187. Jayanthi V, Probert CS, Pinder D, Wicks AC, Mayberry JF. Epidemiology of Crohn's disease in Indian migrants and the indigenous population in Leicestershire. Q J Med. 1992;82(298):125-38.
188. Bhopal RS, Cezard G, Bansal N, Ward HJ, Bhala N, researchers S. Ethnic variations in five lower gastrointestinal diseases: Scottish health and ethnicity linkage study. BMJ Open. 2014;4(10):e006120.
189. Park SJ, Kim WH, Cheon JH. Clinical characteristics and treatment of inflammatory bowel disease: a comparison of Eastern and Western perspectives. World journal of gastroenterology : WJG. 2014;20(33):11525-37.
190. Ng SC. Epidemiology of inflammatory bowel disease: focus on Asia. Best practice & research Clinical gastroenterology. 2014;28(3):363-72.
191. Makharia GK, Ramakrishna BS, Abraham P, Choudhuri G, Misra SP, Ahuja V, et al. Survey of inflammatory bowel diseases in India. Indian journal of gastroenterology : official journal of the Indian Society of Gastroenterology. 2012;31(6):299-306.
192. Goodhand JR, Kamperidis N, Joshi NM, Wahed M, Koodun Y, Cantor EJ, et al. The phenotype and course of inflammatory bowel disease in UK patients of Bangladeshi descent. Alimentary pharmacology & therapeutics. 2012;35(8):929-40.
193. Lee HJ, Jung ES, Lee JH, Hong SP, Kim TI, Kim WH, et al. Long-term clinical outcomes and factors predictive of relapse after 5-aminosalicylate or sulfasalazine therapy in patients with mild-to-moderate ulcerative colitis. Hepatogastroenterology. 2012;59(117):1415-20.
194. Bello C, Belaiche J, Louis E, Reenaers C. Evolution and predictive factors of relapse in ulcerative colitis patients treated with mesalazine after a first course of corticosteroids. Journal of Crohn's & colitis. 2011;5(3):196-202.
195. Prideaux L, Kamm MA, De Cruz PP, Chan FK, Ng SC. Inflammatory bowel disease in Asia: a systematic review. Journal of gastroenterology and hepatology. 2012;27(8):1266-80.
196. Thia KT, Loftus EV, Jr., Sandborn WJ, Yang SK. An update on the epidemiology of inflammatory bowel disease in Asia. The American journal of gastroenterology. 2008;103(12):3167-82.
197. Foster A, Jacobson K. Changing incidence of inflammatory bowel disease: environmental influences and lessons learnt from the South asian population. Front Pediatr. 2013;1:34.
198. Ng SC, Tsoi KK, Kamm MA, Xia B, Wu J, Chan FK, et al. Genetics of inflammatory bowel disease in Asia: systematic review and meta-analysis. Inflammatory bowel diseases. 2012;18(6):1164-76.

199. Yang SK, Loftus EV, Jr., Sandborn WJ. Epidemiology of inflammatory bowel disease in Asia. Inflammatory bowel diseases. 2001;7(3):260-70.
200. Mahurkar S, Banerjee R, Rani VS, Thakur N, Rao GV, Reddy DN, et al. Common variants in NOD2 and IL23R are not associated with inflammatory bowel disease in Indians. Journal of gastroenterology and hepatology. 2011;26(4):694-9.
201. Bernstein CN, Shanahan F. Disorders of a modern lifestyle: reconciling the epidemiology of inflammatory bowel diseases. Gut. 2008;57(9):1185-91.
202. Carr I, Mayberry JF. The effects of migration on ulcerative colitis: a three-year prospective study among Europeans and first- and second- generation South Asians in Leicester (1991-1994). The American journal of gastroenterology. 1999;94(10):2918-22.
203. Montgomery SM, Morris DL, Pounder RE, Wakefield AJ. Asian ethnic origin and the risk of inflammatory bowel disease. European journal of gastroenterology & hepatology. 1999;11(5):543-6.
204. Molodecky NA, Kaplan GG. Environmental risk factors for inflammatory bowel disease. Gastroenterol Hepatol (N Y). 2010;6(5):339-46.
205. Soon IS, Molodecky NA, Rabi DM, Ghali WA, Barkema HW, Kaplan GG. The relationship between urban environment and the inflammatory bowel diseases: a systematic review and meta-analysis. BMC gastroenterology. 2012;12:51.
206. Satsangi J, Silverberg MS, Vermeire S, Colombel JF. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. Gut. 2006;55(6):749-53.
207. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. Lancet. 1980;1(8167):514.
208. Walmsley RS, Ayres RC, Pounder RE, Allan RN. A simple clinical colitis activity index. Gut. 1998;43(1):29-32.
209. Hicks LC, Huang J, Kumar S, Powles ST, Orchard TR, Hanna GB, et al. Analysis of Exhaled Breath Volatile Organic Compounds in Inflammatory Bowel Disease: A Pilot Study. Journal of Crohn's & colitis. 2015;9(9):731-7.
210. Walsh MC, Brennan L, Malthouse JP, Roche HM, Gibney MJ. Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. The American journal of clinical nutrition. 2006;84(3):531-9.
211. Momozawa Y, Deffontaine V, Louis E, Medrano JF. Characterization of bacteria in biopsies of colon and stools by high throughput sequencing of the V2 region of bacterial 16S rRNA gene in human. PloS one. 2011;6(2):e16952.
212. Marchesi JR. Human distal gut microbiome. Environ Microbiol. 2011;13(12):3088-102.
213. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537-41.
214. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.
215. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Res. 2017.
216. Legendre PL, L,. Numerical Ecology 3rd ed. Oxford, UK. : Elsevier; 2012.
217. Lozupone C, Hamady M, Knight R. UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics. 2006;7:371.
218. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.
219. Tognarelli JM, Dawood M, Shariff MI, Grover VP, Crossey MM, Cox IJ, et al. Magnetic Resonance Spectroscopy: Principles and Techniques: Lessons for Clinicians. J Clin Exp Hepatol. 2015;5(4):320-8.
220. Bothwell JH, Griffin JL. An introduction to biological nuclear magnetic resonance spectroscopy. Biol Rev Camb Philos Soc. 2011;86(2):493-510.

221. Beckonert O, Keun HC, Ebbels TM, Bundy J, Holmes E, Lindon JC, et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat Protoc. 2007;2(11):2692-703.
222. Veselkov KA, Lindon JC, Ebbels TM, Crockford D, Volynkin VV, Holmes E, et al. Recursive segment-wise peak alignment of biological (1)h NMR spectra for improved metabolic biomarker recovery. Analytical chemistry. 2009;81(1):56-66.
223. Williams HR, Walker DG, North BV, Taylor-Robinson SD, Orchard TR. Metabonomics in ulcerative colitis. Journal of proteome research. 2010;9(5):2794-5.
224. Sands CJ, Coen M, Maher AD, Ebbels TM, Holmes E, Lindon JC, et al. Statistical total correlation spectroscopy editing of 1H NMR spectra of biofluids: application to drug metabolite profile identification and enhanced information recovery. Analytical chemistry. 2009;81(15):6458-66.
225. Novak P, Tepes P, Fistric I, Bratos I, Gabelica V. The application of LC-NMR and LC-MS for the separation and rapid structure elucidation of an unknown impurity in 5-aminosalicylic acid. Journal of pharmaceutical and biomedical analysis. 2006;40(5):1268-72.
226. Bales JR, Sadler PJ, Nicholson JK, Timbrell JA. Urinary excretion of acetaminophen and its metabolites as studied by proton NMR spectroscopy. Clinical chemistry. 1984;30(10):1631-6.
227. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Analytical chemistry. 2006;78(13):4281-90.
228. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics. 2006;7:142.
229. Euceda LR, Giskeodegard GF, Bathen TF. Preprocessing of NMR metabolomics data. Scand J Clin Lab Invest. 2015;75(3):193-203.
230. Parsons HM, Ludwig C, Gunther UL, Viant MR. Improved classification accuracy in 1- and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. BMC Bioinformatics. 2007;8:234.
231. Olivares A, Dryahina K, Navarro JL, Flores M, Smith D, Spanel P. Selected ion flow tube-mass spectrometry for absolute quantification of aroma compounds in the headspace of dry fermented sausages. Analytical chemistry. 2010;82(13):5819-29.
232. Worley B, Powers R. Multivariate Analysis in Metabolomics. Curr Metabolomics. 2013;1(1):92-107.
233. Mahadevan S, Shah SL, Marrie TJ, Slupsky CM. Analysis of metabolomic data using support vector machines. Analytical chemistry. 2008;80(19):7562-70.
234. Cloarec O, Dumas ME, Trygg J, Craig A, Barton RH, Lindon JC, et al. Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in 1H NMR spectroscopic metabonomic studies. Analytical chemistry. 2005;77(2):517-26.
235. Cloarec O, Dumas ME, Craig A, Barton RH, Trygg J, Hudson J, et al. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. Analytical chemistry. 2005;77(5):1282-9.
236. Dona AC, Kyriakides M, Scott F, Shephard EA, Varshavi D, Veselkov K, et al. A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. Comput Struct Biotechnol J. 2016;14:135-53.
237. National Office of Statistics. Ethnicity and National Identity in England and Wales: 2011 2011 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/ethnicityandnationalidentityinenglandandwales/2012-12-11#tab-Changing-picture-of-ethnicity-over-time.
238. Spekhorst LM, Severs M, de Boer NKH, Festen EAM, Fidder HH, Hoentjen F, et al. The Impact of Ethnicity and Country of Birth on Inflammatory Bowel Disease Phenotype: a Prospective Cohort Study. Journal of Crohn's & colitis. 2017.

239. Sewell JL, Yee HF, Jr., Inadomi JM. Hospitalizations are increasing among minority patients with Crohn's disease and ulcerative colitis. Inflammatory bowel diseases. 2010;16(2):204-7.
240. Hanauer SB, Feagan BG, Lichtenstein GR, Mayer LF, Schreiber S, Colombel JF, et al. Maintenance infliximab for Crohn's disease: the ACCENT I randomised trial. Lancet. 2002;359(9317):1541-9.
241. Prideaux L, Kang S, Wagner J, Buckley M, Mahar JE, De Cruz P, et al. Impact of ethnicity, geography, and disease on the microbiota in health and inflammatory bowel disease. Inflammatory bowel diseases. 2013;19(13):2906-18.
242. Weljie AM, Newton J, Mercier P, Carlson E, Slupsky CM. Targeted profiling: quantitative analysis of 1H NMR metabolomics data. Analytical chemistry. 2006;78(13):4430-42.
243. Zhang AH, Sun H, Qiu S, Wang XJ. NMR-based metabolomics coupled with pattern recognition methods in biomarker discovery and disease diagnosis. Magnetic resonance in chemistry : MRC. 2013;51(9):549-56.
244. Lees HJ, Swann JR, Wilson ID, Nicholson JK, Holmes E. Hippurate: the natural history of a mammalian-microbial cometabolite. Journal of proteome research. 2013;12(4):1527-46.
245. Zuppi C, Messana I, Tapanainen P, Knip M, Vincenzoni F, Giardina B, et al. Proton nuclear magnetic resonance spectral profiles of urine from children and adolescents with type 1 diabetes. Clinical chemistry. 2002;48(4):660-2.
246. Holmes E, Loo RL, Stamler J, Bictash M, Yap IK, Chan Q, et al. Human metabolic phenotype diversity and its association with diet and blood pressure. Nature. 2008;453(7193):396-400.
247. Calvani R, Miccheli A, Capuani G, Tomassini Miccheli A, Puccetti C, Delfini M, et al. Gut microbiome-derived metabolites characterize a peculiar obese urinary metabotype. Int J Obes (Lond). 2010;34(6):1095-8.
248. Clifford MN, Copeland EL, Bloxsidge JP, Mitchell LA. Hippuric acid as a major excretion product associated with black tea consumption. Xenobiotica. 2000;30(3):317-26.
249. Van Dorsten FA, Daykin CA, Mulder TP, Van Duynhoven JP. Metabonomics approach to determine metabolic differences between green tea and black tea consumption. J Agric Food Chem. 2006;54(18):6929-38.
250. Shiomi Y, Nishiumi S, Ooi M, Hatano N, Shinohara M, Yoshie T, et al. GCMS-based metabolomic study in mice with colitis induced by dextran sulfate sodium. Inflammatory bowel diseases. 2011;17(11):2261-74.
251. Tomassini A, Vitalone A, Marini F, Pratico G, Sciubba F, Bevilacqua M, et al. 1H NMR-based urinary metabolic profiling reveals changes in nicotinamide pathway intermediates due to postnatal stress model in rat. Journal of proteome research. 2014;13(12):5848-59.
252. Yap IK, Angley M, Veselkov KA, Holmes E, Lindon JC, Nicholson JK. Urinary metabolic phenotyping differentiates children with autism from their unaffected siblings and age-matched controls. Journal of proteome research. 2010;9(6):2996-3004.
253. Williams A, Sturman S, Steventon G, Waring R. Metabolic biomarkers of Parkinson's disease. Acta Neurol Scand Suppl. 1991;136:19-23.
254. Zheng P, Wang Y, Chen L, Yang D, Meng H, Zhou D, et al. Identification and validation of urinary metabolite biomarkers for major depressive disorder. Mol Cell Proteomics. 2013;12(1):207-14.
255. Stein TP, Koerner B, Schluter MD, Leskiw MJ, Gaprindachvilli T, Richards EW, et al. Weight loss, the gut and the inflammatory response in aids patients. Cytokine. 1997;9(2):143-7.
256. Dong F, Zhang L, Hao F, Tang H, Wang Y. Systemic responses of mice to dextran sulfate sodium-induced acute ulcerative colitis using 1H NMR spectroscopy. Journal of proteome research. 2013;12(6):2958-66.
257. McCole DF. The epithelial glycine transporter GLYT1: protecting the gut from inflammation. J Physiol. 2010;588(Pt 7):1033-4.

258. Balasubramanian K, Kumar S, Singh RR, Sharma U, Ahuja V, Makharia GK, et al. Metabolism of the colonic mucosa in patients with inflammatory bowel diseases: an in vitro proton magnetic resonance spectroscopy study. Magnetic resonance imaging. 2009;27(1):79-86.
259. Kuhls DA, Rathmacher JA, Musngi MD, Frisch DA, Nielson J, Barber A, et al. Beta-hydroxy-beta-methylbutyrate supplementation in critically ill trauma patients. J Trauma. 2007;62(1):125-31; discussion 31-2.
260. Nielsen HR, Killmann SA. Urinary excretion of beta-aminoisobutyrate and pseudouridine in acute and chronic myeloid leukemia. J Natl Cancer Inst. 1983;71(5):887-91.
261. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0--The Human Metabolome Database in 2013. Nucleic Acids Res. 2013;41(Database issue):D801-7.
262. Kassel DB, Martin M, Schall W, Sweeley CC. Urinary metabolites of L-threonine in type 1 diabetes determined by combined gas chromatography/chemical ionization mass spectrometry. Biomed Environ Mass Spectrom. 1986;13(10):535-40.
263. Diaz SO, Barros AS, Goodfellow BJ, Duarte IF, Carreira IM, Galhano E, et al. Following healthy pregnancy by nuclear magnetic resonance (NMR) metabolic profiling of human urine. Journal of proteome research. 2013;12(2):969-79.
264. Stamler J, Brown IJ, Yap IK, Chan Q, Wijeyesekera A, Garcia-Perez I, et al. Dietary and urinary metabonomic factors possibly accounting for higher blood pressure of black compared with white Americans: results of International Collaborative Study on macro-/micronutrients and blood pressure. Hypertension. 2013;62(6):1074-80.
265. Walsh MC, McLoughlin GA, Roche HM, Ferguson JF, Drevon CA, Saris WH, et al. Impact of geographical region on urinary metabolomic and plasma fatty acid profiles in subjects with the metabolic syndrome across Europe: the LIPGENE study. Br J Nutr. 2014;111(3):424-31.
266. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011;334(6052):105-8.
267. Smith EA, Macfarlane GT. Enumeration of human colonic bacteria producing phenolic and indolic compounds: effects of pH, carbohydrate availability and retention time on dissimilatory aromatic amino acid metabolism. J Appl Bacteriol. 1996;81(3):288-302.
268. Neuman MG, Nanau RM. Inflammatory bowel disease: role of diet, microbiota, life style. Translational research : the journal of laboratory and clinical medicine. 2012;160(1):29-44.
269. Mar JS, LaMere BJ, Lin DL, Levan S, Nazareth M, Mahadevan U, et al. Disease Severity and Immune Activity Relate to Distinct Interkingdom Gut Microbiome States in Ethnically Distinct Ulcerative Colitis Patients. MBio. 2016;7(4).
270. Lindsay KL, Hellmuth C, Uhl O, Buss C, Wadhwa PD, Koletzko B, et al. Longitudinal Metabolomic Profiling of Amino Acids and Lipids across Healthy Pregnancy. PloS one. 2015;10(12):e0145794.
271. Hester CM, Jala VR, Langille MG, Umar S, Greiner KA, Haribabu B. Fecal microbes, short chain fatty acids, and colorectal cancer across racial/ethnic groups. World journal of gastroenterology : WJG. 2015;21(9):2759-69.
272. Sheflin AM, Melby CL, Carbonero F, Weir TL. Linking dietary patterns with gut microbial composition and function. Gut microbes. 2017;8(2):113-29.
273. Donovan SM. Introduction to the special focus issue on the impact of diet on gut microbiota composition and function and future opportunities for nutritional modulation of the gut microbiome to improve human health. Gut microbes. 2017;8(2):75-81.
274. Orlich MJ, Singh PN, Sabate J, Jaceldo-Siegl K, Fan J, Knutsen S, et al. Vegetarian dietary patterns and mortality in Adventist Health Study 2. JAMA Intern Med. 2013;173(13):1230-8.
275. Wu GD, Compher C, Chen EZ, Smith SA, Shah RD, Bittinger K, et al. Comparative metabolomics in vegans and omnivores reveal constraints on diet-dependent gut microbiota metabolite production. Gut. 2016;65(1):63-72.
276. Chan Q, Loo RL, Ebbels TM, Van Horn L, Daviglus ML, Stamler J, et al. Metabolic phenotyping for discovery of urinary biomarkers of diet, xenobiotics and blood pressure in the INTERMAP Study: an overview. Hypertens Res. 2017;40(4):336-45.

277. Solanky KS, Bailey NJ, Beckwith-Hall BM, Bingham S, Davis A, Holmes E, et al. Biofluid 1H NMR-based metabonomic techniques in nutrition research - metabolic effects of dietary isoflavones in humans. J Nutr Biochem. 2005;16(4):236-44.
278. Xu J, Yang S, Cai S, Dong J, Li X, Chen Z. Identification of biochemical changes in lactovegetarian urine using 1H NMR spectroscopy and pattern recognition. Analytical and bioanalytical chemistry. 2010;396(4):1451-63.
279. Hwang C, Ross V, Mahadevan U. Popular exclusionary diets for inflammatory bowel disease: the search for a dietary culprit. Inflammatory bowel diseases. 2014;20(4):732-41.
280. Triggs CM, Munday K, Hu R, Fraser AG, Gearry RB, Barclay ML, et al. Dietary factors in chronic inflammation: food tolerances and intolerances of a New Zealand Caucasian Crohn's disease population. Mutat Res. 2010;690(1-2):123-38.
281. Franco-de-Moraes AC, de Almeida-Pititto B, da Rocha Fernandes G, Gomes EP, da Costa Pereira A, Ferreira SRG. Worse inflammatory profile in omnivores than in vegetarians associates with the gut microbiota composition. Diabetol Metab Syndr. 2017;9:62.
282. Slupsky CM, Rankin KN, Wagner J, Fu H, Chang D, Weljie AM, et al. Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. Analytical chemistry. 2007;79(18):6995-7004.
283. Kochhar S, Jacobs DM, Ramadan Z, Berruex F, Fuerholz A, Fay LB. Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabonomics. Anal Biochem. 2006;352(2):274-81.
284. Psihogios NG, Gazi IF, Elisaf MS, Seferiadis KI, Bairaktari ET. Gender-related and age-related urinalysis of healthy subjects by NMR-based metabonomics. NMR Biomed. 2008;21(3):195-207.
285. Rist MJ, Muhle-Goll C, Gorling B, Bub A, Heissler S, Watzl B, et al. Influence of Freezing and Storage Procedure on Human Urine Samples in NMR-Based Metabolomics. Metabolites. 2013;3(2):243-58.
286. Lauridsen M, Hansen SH, Jaroszewski JW, Cornett C. Human urine as test material in 1H NMR-based metabonomics: recommendations for sample preparation and storage. Analytical chemistry. 2007;79(3):1181-6.
287. Bernini P, Bertini I, Luchinat C, Nincheri P, Staderini S, Turano P. Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. J Biomol NMR. 2011;49(3-4):231-43.
288. Storr M, Vogel HJ, Schicho R. Metabolomics: is it useful for inflammatory bowel diseases? Curr Opin Gastroenterol. 2013;29(4):378-83.
289. Bibiloni R, Mangold M, Madsen KL, Fedorak RN, Tannock GW. The bacteriology of biopsies differs between newly diagnosed, untreated, Crohn's disease and ulcerative colitis patients. J Med Microbiol. 2006;55(Pt 8):1141-9.
290. Vasquez N, Mangin I, Lepage P, Seksik P, Duong JP, Blum S, et al. Patchy distribution of mucosal lesions in ileal Crohn's disease is not linked to differences in the dominant mucosa-associated bacteria: a study using fluorescence in situ hybridization and temporal temperature gradient gel electrophoresis. Inflammatory bowel diseases. 2007;13(6):684-92.
291. Mercado CI, Cogswell ME, Valderrama AL, Wang CY, Loria CM, Moshfegh AJ, et al. Difference between 24-h diet recall and urine excretion for assessing population sodium and potassium intake in adults aged 18-39 y. The American journal of clinical nutrition. 2015;101(2):376-86.
292. Dixon LJ, Kabi A, Nickerson KP, McDonald C. Combinatorial effects of diet and genetics on inflammatory bowel disease pathogenesis. Inflammatory bowel diseases. 2015;21(4):912-22.
293. Palau-Rodriguez M, Tulipani S, Marco-Ramell A, Minarro A, Jauregui O, Gonzalez-Dominguez R, et al. Characterization of Metabolomic Profile Associated with Metabolic Improvement after Bariatric Surgery in Subjects with Morbid Obesity. Journal of proteome research. 2018;17(8):2704-14.

294. Peyrin-Biroulet L, Reinisch W, Colombel JF, Mantzaris GJ, Kornbluth A, Diamond R, et al. Clinical disease activity, C-reactive protein normalisation and mucosal healing in Crohn's disease in the SONIC trial. Gut. 2014;63(1):88-95.
295. Walker DG, Williams HR, Bancil AS, Rai P, Pantelidis P, Chambers J, et al. Ethnicity differences in genetic susceptibility to ulcerative colitis: a comparison of Indian asians and white northern Europeans. Inflammatory bowel diseases. 2013;19(13):2888-94.
296. Polytarchou C, Koukos G, Iliopoulos D. Systems biology in inflammatory bowel diseases: ready for prime time. Curr Opin Gastroenterol. 2014;30(4):339-46.
297. Assfalg M, Bertini I, Colangiuli D, Luchinat C, Schafer H, Schutz B, et al. Evidence of different metabolic phenotypes in humans. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(5):1420-4.
298. Bernini P, Bertini I, Luchinat C, Nepi S, Saccenti E, Schafer H, et al. Individual human phenotypes in metabolic space and time. Journal of proteome research. 2009;8(9):4264-71.
299. Stanberry L, Mias GI, Haynes W, Higdon R, Snyder M, Kolker E. Integrative analysis of longitudinal metabolomics data from a personal multi-omics profile. Metabolites. 2013;3(3):741-60.
300. Pacchiarotta T, Derks RJ, Nevedomskaya E, van der Starre W, van Dissel J, Deelder A, et al. Exploratory analysis of urinary tract infection using a GC-APCI-MS platform. Analyst. 2015;140(8):2834-41.
301. Rhee EP. Metabolomics and renal disease. Curr Opin Nephrol Hypertens. 2015;24(4):371-9.
302. Nicholson JK, Wilson ID, Lindon JC. Pharmacometabonomics as an effector for personalized medicine. Pharmacogenomics. 2011;12(1):103-11.
303. Griffin JL, Wang X, Stanley E. Does our gut microbiome predict cardiovascular risk? A review of the evidence from metabolomics. Circ Cardiovasc Genet. 2015;8(1):187-91.
304. Gerich ME, McGovern DP. Towards personalized care in IBD. Nat Rev Gastroenterol Hepatol. 2014;11(5):287-99.
305. Sperisen P, Cominetti O, Martin FP. Longitudinal omics modeling and integration in clinical metabonomics research: challenges in childhood metabolic health research. Front Mol Biosci. 2015;2:44.
306. Jansen JJ, Hoefsloot HC, Boelens HF, van der Greef J, Smilde AK. Analysis of longitudinal metabolomics data. Bioinformatics. 2004;20(15):2438-46.
307. Becker C, Neurath MF, Wirtz S. The Intestinal Microbiota in Inflammatory Bowel Disease. ILAR J. 2015;56(2):192-204.
308. Sheehan D, Shanahan F. The Gut Microbiota in Inflammatory Bowel Disease. Gastroenterol Clin North Am. 2017;46(1):143-54.
309. Sartor RB, Wu GD. Roles for Intestinal Bacteria, Viruses, and Fungi in Pathogenesis of Inflammatory Bowel Diseases and Therapeutic Approaches. Gastroenterology. 2017;152(2):327-39 e4.
310. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 2014;42(Database issue):D643-8.
311. Wozniacka A, Wieczorkowska M, Gebicki J, Sysa-Jedrzejowska A. Topical application of 1-methylnicotinamide in the treatment of rosacea: a pilot study. Clin Exp Dermatol. 2005;30(6):632-5.
312. Chlopicki S, Swies J, Mogielnicki A, Buczko W, Bartus M, Lomnicka M, et al. 1-Methylnicotinamide (MNA), a primary metabolite of nicotinamide, exerts anti-thrombotic activity mediated by a cyclooxygenase-2/prostacyclin pathway. Br J Pharmacol. 2007;152(2):230-9.
313. Ruiz-Rodado V, Nicoli ER, Probert F, Smith DA, Morris L, Wassif CA, et al. 1H NMR-Linked Metabolomics Analysis of Liver from a Mouse Model of NP-C1 Disease. Journal of proteome research. 2016;15(10):3511-27.
314. Chhibber-Goel J, Gaur A, Singhal V, Parakh N, Bhargava B, Sharma A. The complex metabolism of trimethylamine in humans: endogenous and exogenous sources. Expert Rev Mol Med. 2016;18:e8.

315. Chhibber-Goel J, Singhal V, Parakh N, Bhargava B, Sharma A. The metabolite Trimethylamine-N-oxide is an emergent biomarker of human health. Curr Med Chem. 2016.
316. Wilson A, Teft WA, Morse BL, Choi YH, Woolsey S, DeGorter MK, et al. Trimethylamine-N-oxide: A Novel Biomarker for the Identification of Inflammatory Bowel Disease. Digestive diseases and sciences. 2015;60(12):3620-30.
317. Scanlan PD, Shanahan F, O'Mahony C, Marchesi JR. Culture-independent analyses of temporal variation of the dominant fecal microbiota and targeted bacterial subgroups in Crohn's disease. Journal of clinical microbiology. 2006;44(11):3980-8.
318. Sokol H, Seksik P, Rigottier-Gois L, Lay C, Lepage P, Podglajen I, et al. Specificities of the fecal microbiota in inflammatory bowel disease. Inflammatory bowel diseases. 2006;12(2):106-11.
319. Scheppach W, Weiler F. The butyrate story: old wine in new bottles? Curr Opin Clin Nutr Metab Care. 2004;7(5):563-7.
320. De Preter V, Bulteel V, Suenaert P, Geboes KP, De Hertogh G, Luypaerts A, et al. Pouchitis, similar to active ulcerative colitis, is associated with impaired butyrate oxidation by intestinal mucosa. Inflammatory bowel diseases. 2009;15(3):335-40.
321. Matsuoka K, Kanai T. The gut microbiota and inflammatory bowel disease. Semin Immunopathol. 2015;37(1):47-55.
322. Shin NR, Whon TW, Bae JW. Proteobacteria: microbial signature of dysbiosis in gut microbiota. Trends Biotechnol. 2015;33(9):496-503.
323. Fei N, Zhao L. An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. ISME J. 2013;7(4):880-4.
324. Willing B, Halfvarson J, Dicksved J, Rosenquist M, Jarnerot G, Engstrand L, et al. Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. Inflammatory bowel diseases. 2009;15(5):653-60.
325. 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. 2006;55(2):205-11.
326. Willing BP, Dicksved J, Halfvarson J, Andersson AF, Lucio M, Zheng Z, et al. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. Gastroenterology. 2010;139(6):1844-54 e1.
327. Miquel S, Martin R, Rossi O, Bermudez-Humaran LG, Chatel JM, Sokol H, et al. Faecalibacterium prausnitzii and human intestinal health. Curr Opin Microbiol. 2013;16(3):255-61.
328. De Cruz P, Kang S, Wagner J, Buckley M, Sim WH, Prideaux L, et al. Association between specific mucosa-associated microbiota in Crohn's disease at the time of resection and subsequent disease recurrence: a pilot study. Journal of gastroenterology and hepatology. 2015;30(2):268-78.
329. Haberman Y, Tickle TL, Dexheimer PJ, Kim MO, Tang D, Karns R, et al. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. J Clin Invest. 2014;124(8):3617-33.
330. Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell host & microbe. 2014;15(3):382-92.
331. Qi CJ, Zhang Q, Yu M, Xu JP, Zheng J, Wang T, et al. Imbalance of Fecal Microbiota at Newly Diagnosed Type 1 Diabetes in Chinese Children. Chin Med J (Engl). 2016;129(11):1298-304.
332. Cui Y, Wei H, Lu F, Liu X, Liu D, Gu L, et al. Different Effects of Three Selected Lactobacillus Strains in Dextran Sulfate Sodium-Induced Colitis in BALB/c Mice. PloS one. 2016;11(2):e0148241.
333. Ellegaard KM, Engel P. Beyond 16S rRNA Community Profiling: Intra-Species Diversity in the Gut Microbiota. Front Microbiol. 2016;7:1475.
334. Marco-Ramell A, Palau-Rodriguez M, Alay A, Tulipani S, Urpi-Sarda M, Sanchez-Pla A, et al. Evaluation and comparison of bioinformatic tools for the enrichment analysis of metabolomics data. BMC Bioinformatics. 2018;19(1):1.
335. Knight R, Jansson J, Field D, Fierer N, Desai N, Fuhrman JA, et al. Unlocking the potential of metagenomics through replicated experimental design. Nat Biotechnol. 2012;30(6):513-20.

336. Norman JM, Handley SA, Virgin HW. Kingdom-agnostic metagenomics and the importance of complete characterization of enteric microbial communities. Gastroenterology. 2014;146(6):1459-69.
337. Ursell LK, Haiser HJ, Van Treuren W, Garg N, Reddivari L, Vanamala J, et al. The intestinal metabolome: an intersection between microbiota and host. Gastroenterology. 2014;146(6):1470-6.
338. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-occurrence relationships in the human microbiome. PLoS Comput Biol. 2012;8(7):e1002606.
339. Peng G, Hakim M, Broza YY, Billan S, Abdah-Bortnyak R, Kuten A, et al. Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. Br J Cancer. 2010;103(4):542-51.
340. Hryniuk A, Ross BM. A preliminary investigation of exhaled breath from patients with celiac disease using selected ion flow tube mass spectrometry. J Gastrointestin Liver Dis. 2010;19(1):15-20.
341. Kumar S, Huang J, Abbassi-Ghadi N, Spanel P, Smith D, Hanna GB. Selected ion flow tube mass spectrometry analysis of exhaled breath for volatile organic compound profiling of esophago-gastric cancer. Analytical chemistry. 2013;85(12):6121-8.
342. King J, Kupferthaler A, Unterkofler K, Koc H, Teschl S, Teschl G, et al. Isoprene and acetone concentration profiles during exercise on an ergometer. Journal of breath research. 2009;3(2):027006.
343. Issaq HJ, Van QN, Waybright TJ, Muschik GM, Veenstra TD. Analytical and statistical approaches to metabolomics research. Journal of separation science. 2009;32(13):2183-99.
344. Spanel P, Dryahina K, Smith D. The concentration distributions of some metabolites in the exhaled breath of young adults. Journal of breath research. 2007;1(2):026001.
345. Rieder F, Kurada S, Grove D, Cikach F, Lopez R, Patel N, et al. A Distinct Colon-Derived Breath Metabolome is Associated with Inflammatory Bowel Disease, but not its Complications. Clin Transl Gastroenterol. 2016;7(11):e201.
346. Bos LD, Sterk PJ, Schultz MJ. Volatile metabolites of pathogens: a systematic review. PLoS Pathog. 2013;9(5):e1003311.
347. Mochalski P, Sponring A, King J, Unterkofler K, Troppmair J, Amann A. Release and uptake of volatile organic compounds by human hepatocellular carcinoma cells (HepG2) in vitro. Cancer Cell Int. 2013;13(1):72.
348. Pimentel M, Mathur R, Chang C. Gas and the microbiome. Curr Gastroenterol Rep. 2013;15(12):356.
349. Yoshimura M, Nakano Y, Yamashita Y, Oho T, Saito T, Koga T. Formation of methyl mercaptan from L-methionine by Porphyromonas gingivalis. Infection and immunity. 2000;68(12):6912-6.
350. Wilson AD. Advances in electronic-nose technologies for the detection of volatile biomarker metabolites in the human breath. Metabolites. 2015;5(1):140-63.
351. Banik GD, De A, Som S, Jana S, Daschakraborty SB, Chaudhuri S, et al. Hydrogen sulphide in exhaled breath: a potential biomarker for small intestinal bacterial overgrowth in IBS. Journal of breath research. 2016;10(2):026010.
352. Picton R, Eggo MC, Langman MJ, Singh S. Impaired detoxication of hydrogen sulfide in ulcerative colitis? Digestive diseases and sciences. 2007;52(2):373-8.
353. Aprea E, Morisco F, Biasioli F, Vitaglione P, Cappellin L, Soukoulis C, et al. Analysis of breath by proton transfer reaction time of flight mass spectrometry in rats with steatohepatitis induced by high-fat diet. J Mass Spectrom. 2012;47(9):1098-103.
354. Brannelly NT, Hamilton-Shield JP, Killard AJ. The Measurement of Ammonia in Human Breath and its Potential in Clinical Diagnostics. Crit Rev Anal Chem. 2016:0.
355. Dummer J, Storer M, Sturney S, Scott-Thomas A, Chambers S, Swanney M, et al. Quantification of hydrogen cyanide (HCN) in breath using selected ion flow tube mass spectrometry--HCN is not a biomarker of Pseudomonas in chronic suppurative lung disease. Journal of breath research. 2013;7(1):017105.

356. Neerincx AH, Mandon J, van Ingen J, Arslanov DD, Mouton JW, Harren FJ, et al. Real-time monitoring of hydrogen cyanide (HCN) and ammonia (NH(3)) emitted by Pseudomonas aeruginosa. Journal of breath research. 2015;9(2):027102.
357. Smith D, Spanel P, Gilchrist FJ, Lenney W. Hydrogen cyanide, a volatile biomarker of Pseudomonas aeruginosa infection. Journal of breath research. 2013;7(4):044001.
358. Gilchrist FJ, Bright-Thomas RJ, Jones AM, Smith D, Spanel P, Webb AK, et al. Hydrogen cyanide concentrations in the breath of adult cystic fibrosis patients with and without Pseudomonas aeruginosa infection. Journal of breath research. 2013;7(2):026010.
359. Boshier PR, Mistry V, Cushnir JR, Kon OM, Elkin SL, Curtis S, et al. Breath metabolite response to major upper gastrointestinal surgery. J Surg Res. 2015;193(2):704-12.
360. Fuchs P, Loeseken C, Schubert JK, Miekisch W. Breath gas aldehydes as biomarkers of lung cancer. International journal of cancer Journal international du cancer. 2010;126(11):2663-70.
361. Li J, Peng Y, Liu Y, Li W, Jin Y, Tang Z, et al. Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography-mass spectrometry. Clin Chim Acta. 2014;436:59-67.
362. Amal H, Shi DY, Ionescu R, Zhang W, Hua QL, Pan YY, et al. Assessment of ovarian cancer conditions from exhaled breath. International journal of cancer Journal international du cancer. 2015;136(6):E614-22.
363. Miekisch W, Kischkel S, Sawacki A, Liebau T, Mieth M, Schubert JK. Impact of sampling procedures on the results of breath analysis. Journal of breath research. 2008;2(2):026007.
364. Rezaie A, Parker RD, Abdollahi M. Oxidative stress and pathogenesis of inflammatory bowel disease: an epiphenomenon or the cause? Digestive diseases and sciences. 2007;52(9):2015-21.
365. Spitz DR, Azzam EI, Li JJ, Gius D. Metabolic oxidation/reduction reactions and cellular responses to ionizing radiation: a unifying concept in stress response biology. Cancer Metastasis Rev. 2004;23(3-4):311-22.
366. Huang J, Kumar S, Hanna GB. Investigation of C3-C10 aldehydes in the exhaled breath of healthy subjects using selected ion flow tube-mass spectrometry (SIFT-MS). Journal of breath research. 2014;8(3):037104.
367. Smith D, Spanel P. Pitfalls in the analysis of volatile breath biomarkers: suggested solutions and SIFT-MS quantification of single metabolites. Journal of breath research. 2015;9(2):022001.
368. Turner C, Spanel P, Smith D. A longitudinal study of ammonia, acetone and propanol in the exhaled breath of 30 subjects using selected ion flow tube mass spectrometry, SIFT-MS. Physiological measurement. 2006;27(4):321-37.
369. Le Marchand L, Wilkens LR, Harwood P, Cooney RV. Breath hydrogen and methane in populations at different risk for colon cancer. International journal of cancer Journal international du cancer. 1993;55(6):887-90.
370. Ajibola OA, Smith D, Spanel P, Ferns GA. Effects of dietary nutrients on volatile breath metabolites. J Nutr Sci. 2013;2:e34.
371. Jareno-Esteban JJ, Munoz-Lucas MA, Carrillo-Aranda B, Maldonado-Sanz JA, de Granda-Orive I, Aguilar-Ros A, et al. Volatile organic compounds in exhaled breath in a healthy population: effect of tobacco smoking. Arch Bronconeumol. 2013;49(11):457-61.
372. Szulejko JE, McCulloch M, Jackson J, McKee DL, Walker JC, Solouki T. Evidence for Cancer Biomarkers in Exhaled Breath. Sensors Journal, IEEE. 2010;10(1):185-210.
373. Markar SR, Wiggins T, Kumar S, Hanna GB. Exhaled breath analysis for the diagnosis and assessment of endoluminal gastrointestinal diseases. J Clin Gastroenterol. 2015;49(1):1-8.
374. Humad S, Zarling E, Clapper M, Skosey JL. Breath pentane excretion as a marker of disease activity in rheumatoid arthritis. Free Radic Res Commun. 1988;5(2):101-6.
375. Smith D, Spanel P. On the importance of accurate quantification of individual volatile metabolites in exhaled breath. Journal of breath research. 2017;11(4):047106.