University of Alberta

Environmental Exposures, Gut Microbiota, and Urinary Metabolomic Fingerprint of Crohn's Disease Patients Who Have Undergone Ileo-colonic Resection

by

Robert Tso

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of requirements for degree of

Master of Science
in
Experimental Medicine

Department of Medicine

Spring 2014
Edmonton, Alberta

Abstract

Inflammatory bowel disease (IBD) is an idiopathic disease that causes intestinal inflammation and lesions. Canada has one of the highest rates of IBD in the world. One of the major types of IBD is Crohn's Disease (CD) and a large portion of CD patients will undergo surgical resection in the course of their disease. However, up to 90% of these patients will have endoscopic recurrence by 5 years post-surgery. The cause of the recurrence is unknown, but is thought to follow the same course as the initial onset of Crohn's lesions. In this study CD patients in endoscopic remission and relapse that have undergone ileocolonic resection were compared. Results show there were specific changes in gut microbial composition, variations in urinary metabolites and unique environmental exposures, both current and childhood, between the remission and relapse patients.

Table of Contents

List of Tables

Chapter 1: Introduction

1.3. Pathogenesis

1.4. Treatment

Chapter 3: Comparison of colonic and ileal samples

3.1 Bacterial Composition

Chapter 4: Comparison of patients in remission with those suffering relapse

4.1 Patient Population

4.2 Bacterial Composition: RDP10 Taxonomic Identification

4.3 Bacterial Composition: QIIME Pipeline Taxonomic Identification

4.4 Metabolomics

4.5 Environmental Questionnaire

Chapter 5: Comparison of samples from patients with no inflammation with those with any inflammation

5.2 Bacterial Composition

5.3 Metabolomics

5.4 Environmental Questionnaire

Chapter 6: Comparison of males with females

6.1 Patient Population

6.2 Metabolomics

6.3 Metabolomics

List of Figures

Chapter 1: Introduction

Chapter 2: Methods

Chapter 3: Comparison of colonic and ileal samples

Chapter 4: Comparison of patients in remission with those suffering relapse

Chapter 5 Comparison of samples from patients with no inflammation with those with any inflammation

5.2 Bacterial Composition

5.3 Metabolomics

5.4 Environmental Questionnaire

Chapter 6 Comparison of males with females

6.2 Bacterial Composition

6.3 Metabolomics

List of Abbreviations

AIEC	Adherent-invasive E. coli
ASA	Aminosalicylate acid
ATG16L1	Autophagy related 16-like 1
CARD15	Caspase activation recruitment domain 15
CARD9	Caspase recruitment domain-containing protein 9
CD	Crohn’s disease
CEGIIR	Centre of Excellence for Gastrointestinal Inflammation and Immunity Research
DNA	Deoxyribonucleic acid
DSS	Dextran sodium sulfate
GWAS	Genome-wide association studies
IBD	Inflammatory bowel disease
IL	Interleukin
ITLN1	Intelectin-1
MAP	Mycobacterium avium subspecies paratuberculosis
MS	Mass spectrometry
MTPS	Multitag pyrosequencing
NF-κB	Nuclear factor kappa-B
NMR	Nuclear magnetic resonance
NOD2	Nucleotide-binding oligomerization domain-containing protein 2
PCA	Principle component analysis
PLS-DA	Partial least squares discriminant analysis
QIIME	Quantitative insights into microbial ecology
RDP	Ribosomal database project
RF	Radio frequency
ROC	Receiver operating characteristic
SNP	Single-nucleotide polymorphism
SVM	Support vector machines
TLR	Toll like receptor
TNF	Tumor necrosis factor
UC	Ulcerative colitis
Weka	Waikato environment for knowledge analysis

Chapter 1: Introduction

1.1 Inflammatory Bowel Disease

Inflammatory bowel disease (IBD) is a chronic relapsing inflammatory disorder mainly affecting the gastrointestinal tract. Common symptoms include abdominal pain, vomiting, diarrhea, weight loss, intestinal bleeding, malabsorption and fatigue [1]. IBD can arise at any age, but it most commonly occurs among individuals between the ages of 20 to 40.[2] Canada has one of the highest prevalence and incidence rates of IBD in the world with approximately 233000 cases and approximately 10000 new cases being diagnosed every year.[3] In other words, 1 in every 150 Canadians has some form of IBD, or a 0.7% prevalence rate of IBD in Canada [3].

There is a huge economic and personal burden associated with IBD. In Canada alone, economic costs are estimated to be approximately 2.8 billion dollars in 2012, 1.2 billion of which can be associated with direct medical costs including medication, hospitalization and physician visits [3]. That works out to almost $12000 per individual with IBD every year. The indirect costs of 1.6 billion can be mostly associated with long term work losses and out of pocket expenses [3]. In addition to the enormous financial cost to patients with IBD, their quality of life decreases with periods of disease worsening which lead to physical, emotional and social changes [4].

1.2 Disease Presentation

IBD is mainly comprised of Crohn’s disease (CD) and ulcerative colitis (UC) which have distinct characteristics. Inflammation in UC is a continuous mucosal inflammation that is restricted to the colon while inflammation in CD is discontinuous and can involve any part of the gastrointestinal tract from the mouth to the anus (Figure 1.2.1) [5, 6]. Up to 65% of CD patients have involvement of the terminal ileum, and up to 80% of these patients will undergo some form of surgical resection during the course of their disease [7, 8]. Unfortunately surgery is not curative and is a temporary solution as more than 50% of patients will have endoscopic recurrence by 1 year post surgery and up to 90% by 5 years [9]. Clinical recurrence rates have been reported as 20-30% at 1 year with an increase of 10% in each of the subsequent years [10]. The requirement for repeat surgery occurs in 15-45% of patients at 3 years, 26-65% at 10 years and 42-91% at 15 years [11]. In this thesis, I will concentrate on CD.

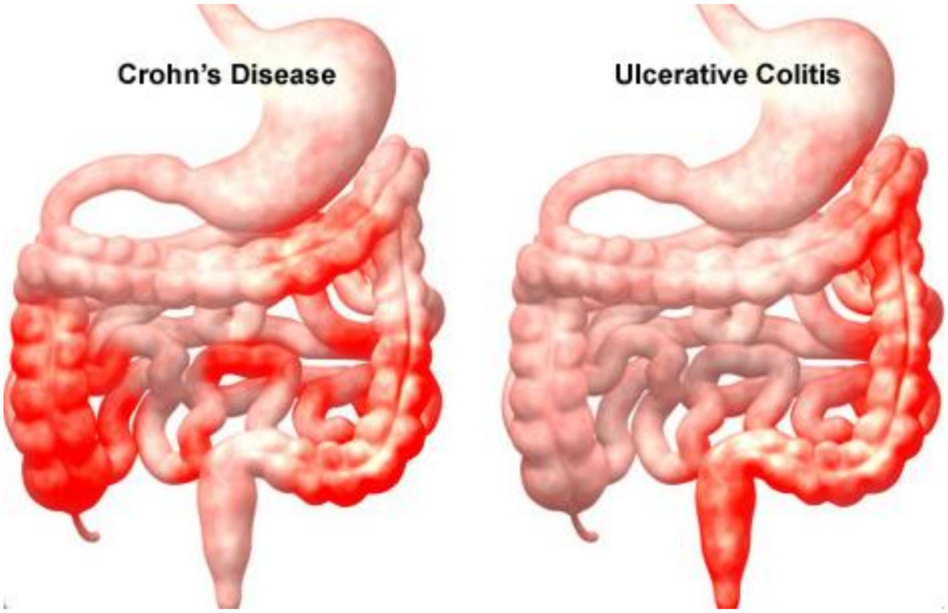


Figure 1.2.1 Inflammation in CD (left) can involve any part of the gastrointestinal tract while UC (right) is restricted to the colon [12].

1.3 Pathogenesis

The cause of CD is unknown. Current research suggests that it is a complex disease that is a result of a genetically predisposed individual being exposed to yet to be determined environmental factors which leads to a dysfunctional immune response [13].

1.3.1 Genetics

Genetic studies have implicated genes and genetic loci in having a critical role in CD pathogenesis. Genome-wide association studies(GWAS) have shown up to 71 CD loci may play a role in the pathogenesis of the disease (Table 1.3.1)[14].

The best studied risk loci is the NOD2 (nucleotide-binding oligomerization domain-containing protein 2)/ CARD15 (caspase activation recruitment domain 15), which is an intracellular pattern recognition receptor that is activated by muramyl dipeptide, a bacterial factor derived from the cell wall of both gram-negative and gram-positive bacteria [15]. 30-40% of CD patients show some NOD2/CARD15 mutation [16]. NOD2/CARD15 expression is high in Paneth cells and is up regulated though inflammatory cytokines such as tumor necrosis factor-α (TNF-α) [17]. The three main genetic variants associated with NOD2/CARD15 are located in or very close to the leucine rich repeat domain and are hypothesized to interfere with bacterial recognition [18]. The NOD2/CARD15 genetic variation has been shown to be the strongest genetic determinant of CD susceptibility with an odds ratio of 3.99 [19]. However, carrying the NOD2/CARD15 genetic variants is not sufficient for the development of CD, as 7% of healthy controls also have the SNP(Single-nucleotide polymorphism), suggesting a key role for environmental determinants [20].

Other loci that have been implicated in CD are the toll like receptor (TLR) genes. TLR2, TLR3, TLR4, and TLR9 have all been associated with CD but TLR4 has been studied the most [21]. TLR4 signaling has been shown to interact with NOD2/CARD15 activation, which strengthens its association with CD [22]. TLR4 is found on cell membranes and binds to numerous ligands such as Lipopolysaccharide (LPS) and uric acid [23].

Caspase recruitment domain-containing protein 9 (CARD9) is another gene implicated in CD. CARD9 has functional properties that link TLR, NOD2/CARD15, and C type lectins with the adaptive immune response [24].

Genes related to autophagy such as autophagy related 16-like 1 (ATG16L1) have also been thought to be relevant in CD. Autophagy has been thought to help with the containment of inflammation by getting rid of pathogens, controlling NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells) signaling and has been implicated in Paneth cell biology by disrupting the Paneth cell granule exocytosis pathway [25].

Genes involved in epithelial barrier functions such as intelectin-1 (ITLN1) are also thought to be important in CD. ITLN1 appears to be localized in the Paneth and goblet cells of the small intestine in the bottom of the epithelial crypts [26]. Because of its localization in the crypts it is predicted to serve as a protective role in innate immune response by serving as an organizer and stabilizer of the membrane by preventing the loss of digestive enzymes and protection from certain pathogens [26].

These genetic factors all lead to inflammatory immune responses, clearly indicating that the genes involved in the actual immune response are very important in CD [27]. In particular IL-10 (interleukin-10), which represents an anti-inflammatory and immunosuppressive cytokine, has been shown to be involved in very early onset pediatric inflammatory bowel disease. In one particular study, a decrease in expression due to some sort of IL-10 deficiency was found in almost one quarter (16/66) of patients with early onset IBD [28].

Table 1.3.1.1 Known CD loci

dbSNP ID	Chromosome	Risk Allele	Candidate Genes
rs11209026	1p31	G	*IL23R*
rs2476601	1p13	G	*PTPN22*
rs4656940	1q23	A	*CD244, ITLN1*
rs7517810	1q24	T	*TNFSF18, TNFSF4, FASLG*
rs7554511	1q32	C	*C1orf106, KIF21B*
rs3792109	2q37	A	*ATG16L1*
rs3197999	3p21	A	*MST1, GPX1, BSN*
rs11742570	5p13	C	*PTGER4*
rs12521868	5q31	T	*SLC22A4, SLC22A5, IRF1, IL3*
rs7714584	5q33	G	*IRGM*
rs6556412	5q33	A	*IL12B*
rs6908425	6p22	C	*CDKAL1*
rs1799964	6p21	C	*LTA, HLA-DQA2, TNF, LST1, LTB*
rs6568421	6q21	G	*PRDM1*
rs415890	6q27	C	*CCR6*
rs1456896	7p12	T	*IKZF1, ZPBP, FIGNL1*
rs4871611	8q24	A	
rs10758669	9p24	C	*JAK2*
rs3810936	9q32	C	*TNFSF15, TNFSF8*
rs12242110	10p11	G	*CREM*
rs10761659	10q21	G	*ZNF365*
rs4409764	10q24	T	*NKX2-3*
rs7927997	11q13	T	*C11orf30*
rs11564258	12q12	A	*MUC19*, *LRRK2*
rs3764147	13q14	G	*C13orf31*
rs2076756	16q12	G	*NOD2*
rs2872507	17q21	A	*GSMDL, ZPBP2, ORMDL3, IKZF3*
rs11871801	17q21	A	*MLX, STAT3*
rs1893217	18p11	G	*PTPN2*
rs740495	19p13	G	
rs1736020	21q21	C	
rs2838519	21q22	G	*ICOSLG*
rs2797685	1p36	A	VAMP3
rs3180018	1q22	A	SCAMP3, MUC1
rs1998598	1q31	G	DENND1B
rs3024505	1q32	T	IL10, IL19
rs13428812	2p23	G	DNMT3A
rs780093	2p23	T	GCKR
rs10495903	2p21	T	THADA

rs10181042	2p16b	T	C2orf74, REL
rs2058660	2q12c	G	IL18RAP, IL12RL2, IL18R1, IL1RL1
rs6738825	2q33	A	PLCL1
rs7423615	2q37	T	SP140
rs13073817	3p24	A	
rs7702331	5q13	A	
rs2549794	5q15	C	ERAP2, LRAP
rs11167764	5q31	C	NDFIP1
rs359457	5q35	T	CPEB4
rs17309827	6p25	T	
rs1847472	6q15	G	BACH2
rs212388	6q25	G	TAGAP
rs6651252	8q24	T	
rs4077515	9q34c	T	CARD9, SNAPC4
rs12722489	10p15	C	IL2RA
rs1819658	10q21	C	UBE2D1
rs1250550	10q22e	G	ZMIZ1
rs102275	11q12	C	FADS1
rs694739	11q13	A	PRDX5, ESRRA
rs2062305	13q14	G	TNFSF11
rs4902642	14q24	G	ZFP36L1
rs8005161	14q35	T	GALC, GPR65
rs17293632	15q22	T	SMAD3
rs151181	16p11e	G	IL27, SH2B1, EIF3C, LAT, CD19
rs3091315	17q12b	A	CCL2, CCL7
rs12720356	19p13	G	TYK2, ICAM1, ICAM3
rs736289	19q13e	T	
rs281379	19q13e	A	FUT2, RASIP1
rs4809330	20q13	G	RTEL1, TNFRSF6B, SLC2A4RG
rs181359	22q11	T	YDJC
rs713875	22q12e	C	MTMR3
rs2413583	22q13	C	MAP3K7IP1

1.3.2 Environmental Factors

Having a genetic predisposition to CD does not appear to be enough to cause the disease. In monozygotic twins the concordance level is only at 50% [29]. This suggests that some sort of trigger appears to be necessary and several environmental factors have been implicated in activating CD. Some of these factors may include smoking, diet, socioeconomic status, stress, childhood exposures to animals or pollution, medications, infectious agents, or microbial agents [30].

Smoking is the only well replicated environmental factor associated with of CD [31-33]. One meta-analysis showed that smoking increases the risk of having CD by an odds ratio of 2.0 (1.65-2.47) [34]. In addition, CD is more severe in smokers compared to non-smokers based on the number of flare-ups, penetrating complications, surgery, and postoperative recurrence [34]. For example, patients who smoke have more fistulas and abscesses than patients who have never smoked [35]. There is conflicting evidence whether there is a dose response effect for the number of cigarettes consumed daily or not [36, 37] Smoking has also been associated with a higher risk of CD that involves the ileum compared to involvement of the colon [38, 39]. Smokers also appear to have a lower rate of response and shorter durations of response to anti-inflammatory medicine such as infliximab [40]. In addition smoking cessation appears to be beneficial on the course of CD as the risk of flare ups, the need for steroids and need for surgical resection was similar to CD patients that never smoked [41]. Unfortunately, the mechanisms by which smoking influences CD are still unknown. There are hundreds of different substances in tobacco and there are numerous targets of smoking including the mucous layer, as well as innate and adaptive immune cells [42].

Several infectious agents have also been proposed to have a role in CD. One example is perinatal infections. Patients who had a perinatal infection noted in their medical file had an increased chance to develop CD with an odds ratio of 3.8 (2.6- 5.8) [43].

There have been many studies looking at the effect of dietary factors in the pathogenesis of CD. Some studies show a high dietary fat intake is associated with an increased CD risk [44], while others do not show any significant association [45]. The same inconsistencies are seen in studies that looked at carbohydrates, proteins, fruits, vegetables, fibre, and meat in relation to CD [44-50].

Finally, the prevalence of CD is extremely low in developing countries compared to westernized nations; therefore a popular theory is the “hygiene hypothesis” since westernized nations live a much cleaner lifestyle. More antibiotics, cleaner drinking water and vaccinations all limit the exposure of the intestine to enteric pathogens which may cause the immune system to be vulnerable to challenges the intestine may face [51]. In addition, one study showed infants were at 2.9 times the odds of having IBD compared to controls if they have used antibiotics in the first year of life suggesting the a role of the developing microbiome in IBD [52].

1.3.3 Microbiome

The microbial environment of the gut appears to be a significant factor in CD pathogenesis and recurrence. The current hypothesis is the intestinal microbiota or their by-products initiate inflammatory responses and disrupt the gastrointestinal epithelium. Evidence from both animal and human studies have confirmed that microbiota play some sort of role in the pathogenesis of CD [53, 54]. Animals with genetic pre-disposition to colitis will develop spontaneous colitis in a normal bacteria rich environment but generally, do not develop inflammation in a germ free environment [55]. Similarly, CD patients undergoing surgery who had a diverting loop ileostomy prevented disease recurrence in the segments of intestine that were no longer exposed to intestinal microbes, but recurrence of the disease occurred once fecal flow was restored [56].

Although it is accepted that the gut bacteria play some sort of essential role in the pathogenesis of CD, no specific organism or group of organisms have been positively identified. *A* dysbiosis in the gut microbiota including a decrease in microbial diversity and stability as well as changes in the abundance of specific microbes has been well described in CD. In addition, one study showed infants were at 2.9 times the odds of having IBD compared to controls if they have used antibiotics in the first year of life suggesting the a role of the developing microbiome in IBD patients, but whether these changes are causative or associative remains to be shown [50-53]. Overall, studies have found a 30-50% decrease in microbial diversity, significantly lower temporal stability of dominant microbial species, alterations in community composition, increased adherence to mucosa, invasiveness or virulence of select species, and alterations in functional and metabolic characteristics of microbes in CD patients compared with healthy individuals [53, 54, 57-59]. More specifically there have been consistent findings that CD patients have decreased abundance of *Firmicutes* including *Clostridium* and *Bacillus* species, but an increase in Proteobacteria including *Escherichia* species [60-64]. Other findings that relate to compositional changes in CD patients are inconsistent. For example, some studies show a significant increase in abundance of *Bacteroidetes* while others show a significant decrease [60, 61]. Figure 1.3.3.1 shows some of the reported compositional changes in the literature.

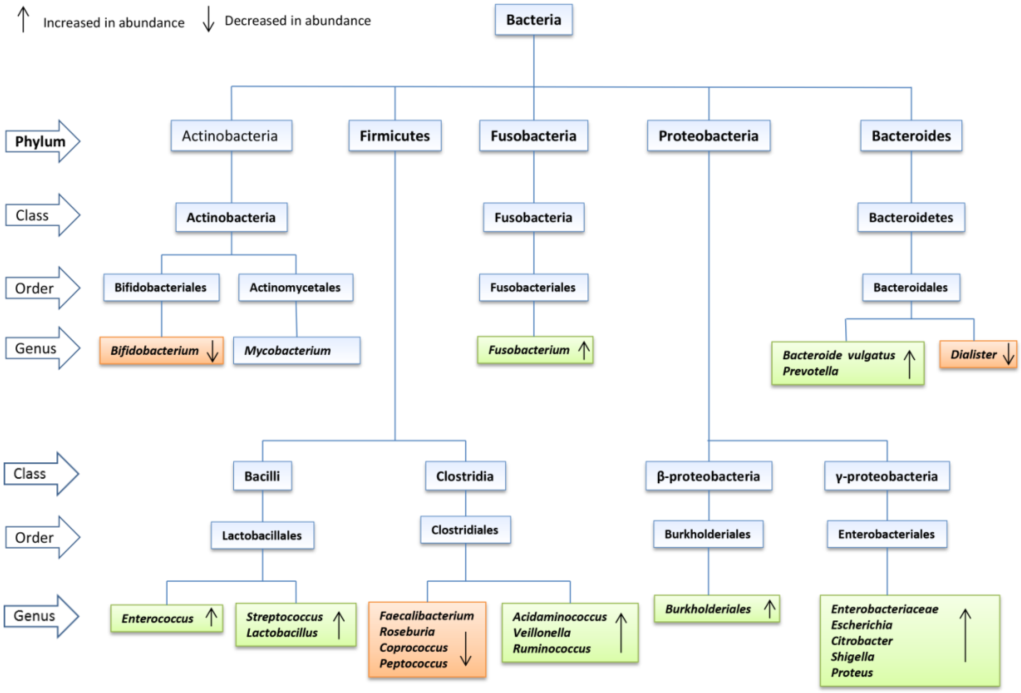


Figure 1.3.3.1 Suspected bacterial compositional changes in CD.

Mycobacterium avium subspecies *paratuberculosis* (MAP) is a specific bacteria thought to have a role in CD because it is an intracellular pathogen that causes a similar inflammatory condition in cattle called Johne's disease. One meta-analysis compared the presence of MAP in tissue samples of patients with CD to controls showed an odds ratio of 7.01 (3.95-12.4) [65]. Recently *Campylobacter concisus* has been implicated in CD. *C. concisus* has been shown to invade Caco-2 intestinal epithelial cells and alter barrier function [66]. 66% of children with CD tested positive for *C. concisus* while only 33% of healthy control has this bacteria suggesting this may be an early trigger of CD [67]. Another species that has consistently been implicated in CD is *Helicobacter pylori.* One study showed the presence of *Helicobacter pylori* and enterohepatic *Helicobacter* was significantly higher in children with CD (59%) compared to controls (9%) [68]. Adherent-invasive *E. coli* (AIEC) have also been extensively studied for their role in CD [69, 70]. These bacteria adhere to and invade the intestinal epithelial cells and they can be found in 30% of CD patients compared to 6% of controls [57]. Finally, *Faecalibacterium prausnitzii* has been constantly shown to be decreased in CD patients compared to controls [71-73]. As shown, there have been many bacterial associations of CD with specific bacteria, but none have really shown to be causative. It is unclear whether the onset of inflammation alters the environment of the gut to allow for these particular organisms to grow or whether these organisms contribute to the cause to inflammation or both. However it is clear that the gut microbiome is highly variable and individualized, and is hugely dependent on diet and other environmental factors.

There have been a number of different studies that have attempted to determine whether there is localized dysbiosis between inflamed and non-inflamed tissue. Two studies show there is a difference [74, 75] however, the majority do not[63, 76-79].

1.3.4 Immune system

Defects in the innate and adaptive immune system have been shown to contribute to the mechanism that causes CD. One theory of how the innate immune system contributes is that defects in the innate immune response results in a failure to properly contain luminal bacteria and therefore leads to an inappropriate adaptive inflammatory response [80]. The adaptive immune system in CD patients typically produces a T helper 1 response. Also, T-helper 17 cells that produce inflammatory cytokines and reduction in T regulatory cells in CD patients have been implicated in the pathogenesis of the disease [81]. In addition, defensins are another important class of antimicrobial peptides. They provide nonspecific defense against a variety of microrganisms and it has been shown CD patients have differntially expressed beta-defensins [82].

1.4 Treatment

There are several groups of medications to treat CD. The first group is aminosalicylates, which includes drugs that contain 5-aminoalicylate acid (5-ASA) such as sulfasalazine and mesalamine [83]. Aminosalicylates have anti-inflammatory properties and generally prescribed to patients with milder attacks of CD flares. One recent meta-analysis showed that the role of 5-ASAs in inducing remission of active CD and preventing relapse of CD is uncertain. Although there was a trend towards a benefit with sulfasalazine over placebo there was no benefit of mesalamine over placebo [83]. Overall that paper recommended that more clinical trials were necessary. The second main group of medications for CD are corticosteroids including prednisone and hydrocortisone. These medications are used to alleviate symptoms of moderate to severe CD by reducing inflammation and used when other medications have stopped working or to stop a sudden flare-up. In one clinical trial, approximately one third of patients on corticosteroids were able to achieve clinical remission [84]. Immunomodulators are another class of drugs used to control CD. This class of medications, including azathioprine and methotrexate, suppresses the body’s immune system to reduce inflammation. Typically immunomodulators are used when corticosteroids have been ineffective. One clinical trial showed approximately 30% of patients receiving azathioprine alone were in clinical remission at week 26 [85]. The latest generation of medications for relieving CD symptoms are biologics or anti-TNF (Tumor necrosis factor) agents. These medications target and block molecules involved in inflammation. Some examples include adalimumab and infliximab. In the same clinical trial that showed azathioprine to induce remission in 30% of patients, it showed infliximab to induce remission in 44% of patients [85]. Recently the topic of medical management of CD has been reviewed and recommendations have been made [86-89].

When conventional medication therapy fails, surgery becomes the next form of treatment for CD. Up to 70% of CD patients will require some sort of surgical resection in the course of their disease due to complications [7]. These complications may include stenosis, abscess, or fistula. The majority of CD cases involve the terminal ileum and surgery usually involves removal of the diseased bowel and anastomosis of the remaining healthy intestine. However, after ileocolonic resection and anastomosis, up to 60% of patients have endoscopic recurrence near the neoterminal ileum after 1 year and up to 90% after 5 years [90]. The cause of CD relapse is not fully understood but it is likely that it is caused by the same factors that cause the initial disease. The need for another operation after the initial resection range rises from 25-60% at 5years post operation to 42-91% after 15 years [11]. The Rutgeerts scoring system is the most widely used scoring system for endoscopic post-operative recurrence. The scoring system goes from I_0 to I_4 with I_4 being the worst (Figure 1.4.1) (Table 1.4.1).

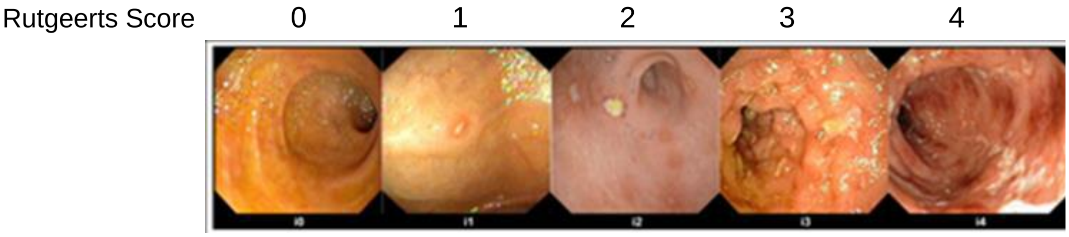


Figure 1.4.1 Visual representation of the Rutgeerts scoring system. I_0, represents no ulcers (left), while I_4 represents widespread inflammation with large ulcers (right) (image from Dr. Troy Perry).

Table 1.4.1 Rutgeerts scoring system

Rutgeerts Score	Description
I_0	no aphthous ulcers
I_1	less than 5 aphthous ulcers
I_2	more than 5 aphthous lesions with normal intervening mucosa, skip areas of larger lesions or lesions confined to ileocolonic anastomosis (i.e. less than 1 cm in length)
I_3	diffuse aphthous ileitis with diffusely inflamed mucosa
I_4	diffuse inflammation with larger ulcers, nodules and or narrowing

The prevention of post-operative recurrence has received a lot of attention lately. One meta-analysis that reviewed 23 studies showed there was no advantage of taking probiotics, but there was a reduced risk of endoscopic recurrence with mesalamine therapy (Risk ratio of 0.50 and 95%CI of 0.29-0.84) and with azathioprine/6-mercaptopurine therapy (Risk ratio of 0.59 and 95%CI of 0.38-0.92) relative to placebo [91]. In addition there have been recent studies that show treatment with infliximab prevents CD recurrence as nearly 85% of patients in the study that were on infliximab were in endoscopic remission after 1 year compared to 9% of patients in the placebo group [92].

1.5 Metabolomics

Metabolomics is one of the newest fields of study in biology and is defined as a non-biased identification and quantification of all metabolites in a biological system. Metabolites are the final downstream products of the genome and proteome and are defined as the small molecules that are the chemical products of metabolism [93].

Unlike the transcriptome, which is subject to epigenetic regulations or the proteome which is subject to post translational modifications, the metabolome is a direct measurement of the function or activity occurring in the biological system [94]. Metabolomics is more sensitive than other “omics” because metabolite concentrations change rapidly reflecting the continuous changes of various pathways compared to the slower changes in protein levels and the DNA (deoxyribonucleic acid) which does not change. (Figure 1.5.1) [95]. In addition, metabolite levels are greatly influenced by various factors such as environment, drugs, age, and diseases [95]. The metabolome is also the most diverse in terms of chemical and physical properties [93]. It is this diversity, and ability to produce a complete compositional representation of individual biological samples that may potentially lead metabolomics to become a powerful scientific discovery tool [96].

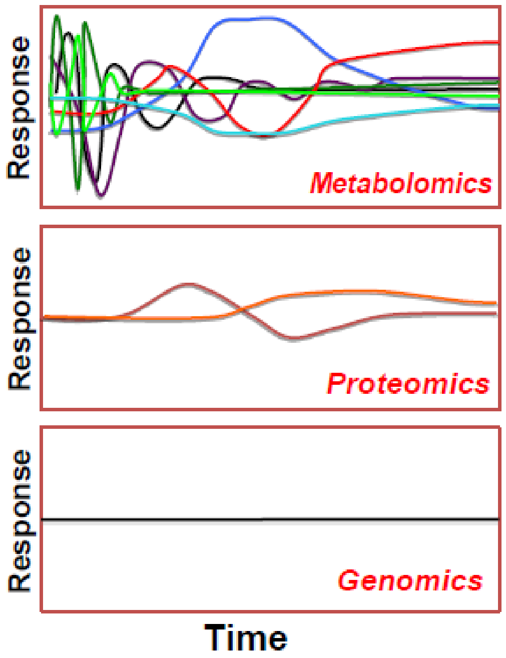


Figure 1.5.1 Metabolomic response (top) is influenced much more that proteomic response (middle) or genomic response (bottom) by various environmental factors.

1.5.1 Techniques

Unlike other “omics”, there is no single preferential method for metabolomic identification as each method has its advantages and disadvantages. The ideal method would be able to identify all metabolites from a single sample; it would be able to yield precise quantitative data; and lastly be high throughput. Currently there is no technique that is able to meet all of these criteria, therefore the choice of technique will depend of the experimental goals [95]. The two most common techniques are Nuclear Magnetic Resonance (NMR) and Mass Spectrometry (MS). MS usually includes a separation stage based on gas chromatography or liquid chromatography or optical spectroscopy techniques [97].

1.5.2 Nuclear Magnetic Resonance (NMR)

NMR is one of the most common techniques used in metabolomic studies [97]. NMR is able to take advantage of the spin properties of the nucleus of atoms, and the spin-spin coupling of the atom to indicate the number and properties of nearby nuclei along with their electromagnetic connectivity [93]. In addition to the spin properties of the atom, chemical shifts in the NMR spectrum allow the determination of the nature of the chemical environment the nucleus is located in [93]. The readout contains an array of peaks, and the heights and positions of these peaks allow the determination of the carbon-hydrogen framework and identification of the metabolites [95]. A typical NMR spectrometer is comprised of four major parts. The first part is a strong stable magnet. The second piece is a radio frequency transmitter that is used to emit a precise radio frequency (RF). The third part is a detector, which is used to measure the amount of RF that is absorbed by the sample and the final major component is a computer that is used to record the amount of energy absorbed (RF) as a function of the magnetic field strength (Figure 2) [98]. Although there are a handful of different nuclei that can be used in NMR such as the carbon 13 nucleus, the ^{1}H nucleus is the most commonly used [93].

One advantage to NMR is that it is specific but it is not selective. In other words, every resonance that is recorded is unique to a particular compound and it also helps to provide information about the structural components of the sample [93]. NMR is a systems approach in looking at the metabolome and in addition is also a non-destructive technology, which means the samples can be recovered after analysis [95]. NMR is an effective analytical technique since it is able to provide both qualitative and quantitative data. It can give information about the chemical structure of the sample and since the intensity of the signal is directly related to the amount of the substance it is also a quantitative technique [99].

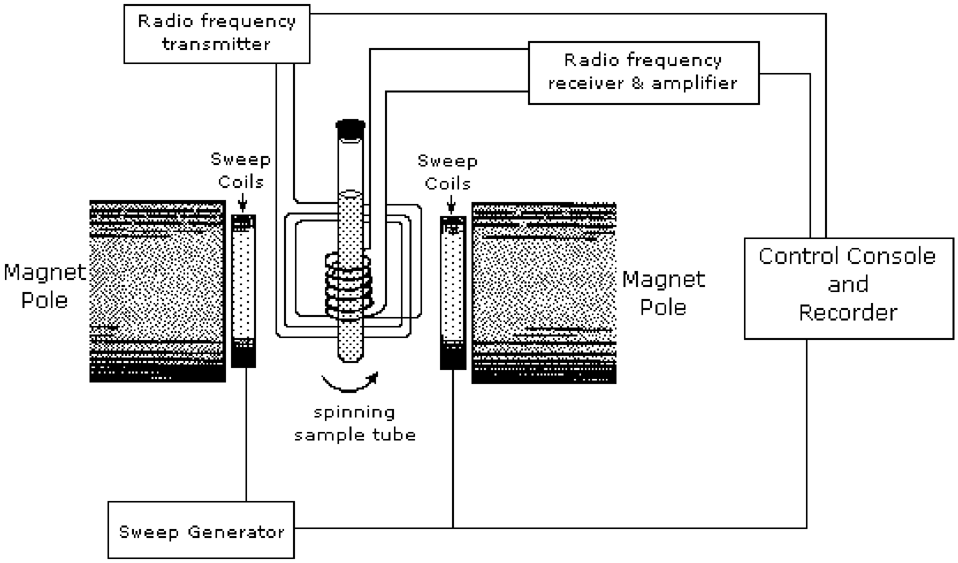

Radio frequency transmitter
Radio frequency receiver & amplifier
Sweep Coils
Sweep Coils
Magnet Pole
Magnet Pole
Control Console and Recorder
spinning sample tube
Sweep Generator

Figure 1.5.2.1 Schematic of a NMR [100].

1.5.3 Mass Spectrometry

Briefly, mass spectrometry (MS) is another technique that has been used to study metabolomics. MS can either be used by itself or in combination with another type of chromatography to separate the sample out before identifying metabolites. A mass spectrometer works by taking the sample and bombarding it with electrons to break the molecule apart. Then it takes the different fragments and measures the molecular weight to charge ratio of each fragment to identify the molecule [101]. An advantage of using MS for metabolomics is that metabolites present at concentrations orders of magnitude lower than are detectable by NMR can be identified using MS [101]. The techniques that are used to separate the sample before MS include gas chromatography, high performance liquid chromatography, and capillary electrophoresis, just to name a few. One main advantage to using these pre-separation techniques is that they can be focused to allow identification of specific metabolites [101].

1.6 Machine Learning

With the popularization of high throughput techniques and systems biology methods, the amount of data being generated is becoming more and more extensive. The amount of data being generated is almost overwhelming and the only way to comprehend it is through the aid of computers. One of the ways is through the utilization of machine learning. Machine learning can be defined as programing computers to optimize a performance task using example data or past experience [102]. In terms of a research point of view, machine learning is the use of computers that implement certain learning algorithms to attempt to learn from the data provided; perhaps trying to distinguish one group of patients from a different group or extracting only the most important information out of the sea of information.

Machine learning methods can be divided into two main categories; supervised and unsupervised learning. Supervised learning is where the class labels of the training data are known (input where the desired output is known) and the learning algorithm tries to build a model that best distinguishes or separates the classes [102]. For example, trying to distinguish a group of sick patients from healthy controls with training data when the status of the patients is known would be supersized learning method. On the other hand unsupervised learning is where the class labels are not known and the learning algorithm tries to build a model to separate the natural groups, or discover the structure within the data [102]. For example, clustering patients with similar characteristics together when the status (healthy or sick) of the patient is not known would be an unsupervised learning problem.

One of the most commonly used unsupervised learning techniques is principle component analysis (PCA). With the number of features commonly well exceeding the number of data points, dimensionality reduction is often a crucial step in machine learning. Mathematically PCA takes the all the data and re-expresses it as a set of new orthogonal variables called principle components and is also able to represent patterns of similar data points in plots or maps [103]. Its goal is to extract the most important information that explains the largest possible variance in the data, compress, and simplify the data set while preserving as much information as possible [103].

Another common unsupervised learning method is clustering. There are countless different clustering algorithms and the main goal of all of them is to group a collection of data points into subsets or “clusters”, such that each point within a cluster is more similar to one another than objects in other clusters [104]. One of the most common is hierarchical clustering. As the name suggests, a hierarchical representation is created by merging clusters together at each lower level and usually presented in a dendogram (Figure 1.6.1) [104].

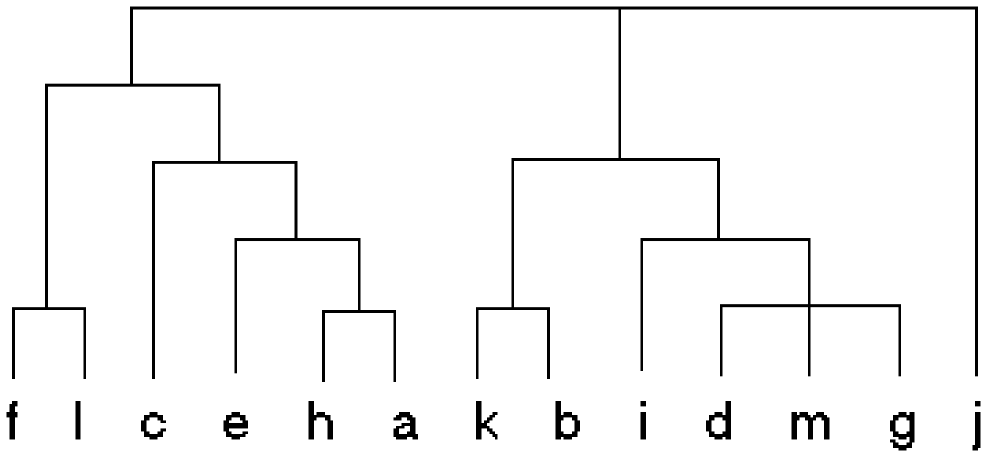


Figure 1.6.1 Sample dendogram. Each letter represents an object and they are successively combined based on similarity until there is only one group remaining (image from http://www2.cs.uregina.ca/~dbd/cs831/notes/clustering/clustering.html).

Supervised learning techniques generally follow a couple main steps. First there must be a labeled training data set that the learning algorithm builds its model on. This training set must be representative of the data in the real world. Secondly the learning algorithm builds the best possible model based on the training data. The best possible model may be based on a variety of factors such as probability, similarity, or regression. Then lastly, the model is tested for accuracy by applying the model to a test data set [104]. In many cases, two sets of data (training and testing) are not available and a technique called cross-validation is commonly used to estimate the accuracy of a predictive model. It works by splitting a dataset into partitions, then the model is built on one subset (training set) and then tested for accuracy using the other subset (testing set). This is then repeated multiple times using different partitions and the accuracy is based on the average over all the cycles [104].

Some of the major supervised learning techniques include decision trees, support vector machines (SVM), naive Bayes classifier, and partial least squares discriminant analysis (PLS-DA).

Decision trees are an example of a divide and conquer technique, where each branch of the tree is split based on the amount of information gained by making the split. One specific example is the C4.5 algorithm to generate a decision tree. Briefly this algorithm builds the tree based on information entropy and each additional split maximizes the normalized difference in entropy(information gain) [105].

The naïve Bayes classifier is commonly used because it is fast and easy. The naïve Bayes classifier is probability based classifier based on Bayes theorem, which makes the assumption that all variables are independent within the class. This independence assumption makes it very easy to model or train on most datasets and is commonly used as a baseline performance mark [106].

Another more powerful supervised learning technique is SVM. SVMs were first developed for binary classification and basically did this by looking for the optimal separating hyperplane between the two classes by maximizing the distance between the two closest points to the separating hyperplane(Figure 1.6.2) [107].

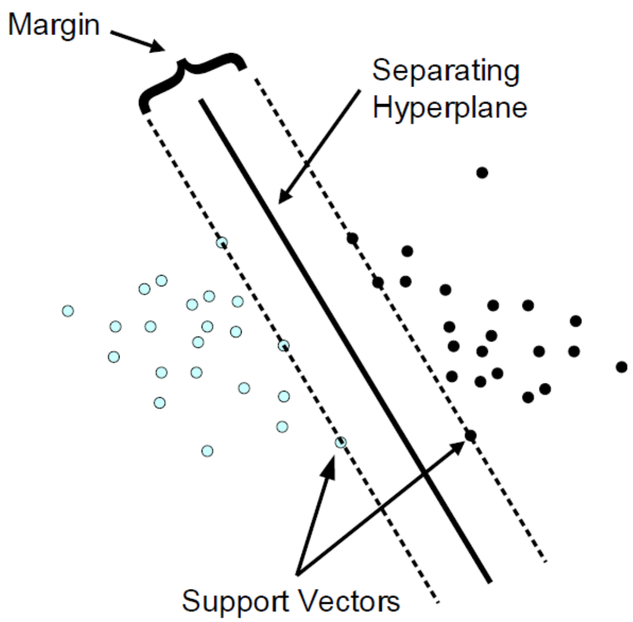


Figure 1.6.2 Support Vector Machine constructs a hyperplane to separate two classes which can be used for classification [108].

Finally partial least squares (PLS) regression combines features from PCA and Multiple Linear Regression [109]. PLS builds a model by adding data points sequentially so the parameters in the model are continuously being updated. PLS-DA is based on the PLS regression model, but the dependent variable is categorical. This approach has been shown to be useful for many classification tasks [110]. PLS-DA plots are evaluated with cross validation in which two values, R^2 and Q^2, are used to determine how well the model fits the data and the predictive error of the model respectively. A good PLS-DA model has values of R^2 and Q^2 close to each other and close to 1.

In some datasets, the number of features (e.g. metabolites or bacteria) greatly outnumbers the number of instances (patients). In this case it may be necessary to select the most important/ useful features before running each machine learning algorithm. Training a classifier using the maximum number of features is not always the best option, as irrelevant or redundant features can negatively influence the performance. Countless feature selection algorithms exist but two of the most common are Best First and Greedy search methods. A Greedy search may either start with all or no features and stops when the addition or subtraction of features no longer results in an improvement. Best First search is similar to Greedy search however it is able to backtrack along the selection path to explore different possibilities that may be better but whatever feature selection algorithm is used, care must be made to ensure the feature selection is done in-fold to ensure valid results.

1.7 Aims and Objectives

The primary objective of this project was to determine if specific environmental exposures or gut microbes were associated with disease relapse or remission in CD patients that have undergone ileocolonic resection.

The secondary objectives were:

1) To determine if a specific metabolomic profile in urine could differentiate between CD patients in remission or relapse.
2) To determine if bacterial and metabolomic features could differentiate between males and females.

1.8 Hypothesis

Crohn's disease patients in remission would have a different and unique microbial profile compared with Crohn's disease patients suffering from relapse, and this would be mirrored by a distinct urinary metabolomic profile

Chapter 2: Methods

2.1 Patient Population

Patient biopsies that satisfied the criteria listed below stored in the CEGIIR biobank were identified and selected by Dr. Levinus Dieleman. These biopsies were obtained from macroscopically healthy areas of the neo-terminal ileum (*n*=30) and colon (*n*=8) from Crohn's disease (CD) patients that had previously undergone an ileocolonic resection. Patients were assessed for endoscopic postoperative recurrence of CD based on the Rutgeerts scoring system. A patient was considered in endoscopic remission (*n*=20) if he or she were assigned a Rutgeerts score of 0, 1 or 2, and considered in endoscopic relapse (n=18) if they were assigned a score of 3 or 4. Urine samples were also collected from these patients and were also extracted from the biobank. The majority of the urine samples were taken on the day as the biopsy with the exception of 5 patients. These 5 patients had their urine collected an average of 4.4 (2-10) months prior to the biopsy. Lastly a patient questionnaire was collected from all the patients. This provided responses to hundreds of questions such as age, sex, smoking history, second hand smoke, alcohol history, family history of disease, early childhood environmental exposures, medications, and certain dietary exposures were collected. (See Appendix 5 for questions extracted from the CEGIIR questionnaire.)

2.2 DNA Extraction

Biopsies were thawed and DNA extracted using the QIAamp DNA stool mini kit. 1.0mL of TH150 buffer,(0.176g Tris-HCl, 0.106g Trisma base, 1.74g NaCl, and 700ml water) 50µl sodium dodecyl sulfate (10%w/v) and 10µl of Proteinase K (20mg/ml) was added to the tissue and incubated at 55˚C for 1-2hours to completely digest the tissue. The samples were resuspended to which 0.1mm zirconia beads were added, and tissue was disrupted three times using a bead beater for 30 seconds each with cooling on ice between each round. The tubes were then heated at 95˚C for 15 minutes and the rest of the protocol was followed according to the QIAamp DNA stool mini kit. (See Appendix 1 for complete protocol.)

2.3 Microbiome Analysis

The extracted DNA was sent to MicroBiome Analysis Center at George Mason University for sequencing and identification of microbial communities. The microbial composition of the biopsies was determined using the 16S rRNA gene.

10ng of extracted total genomic DNA was subject to Length Heterogeneity PCR (LH-PCR) as a quality control to ensure linear amplification of the DNA [111]. In short, the DNA was amplified by PCR using fluorescently labeled forward primer 27F (5'-(6FAM) AGAGTTTGATCCTGGCTCA G-3') and unlabeled reverse primer 355R' (5'-GCTGCCTCCCGTAGGAGT-3') then the products were diluted and separated on an ABI 3130xl fluorescent capillary sequencer. Finally peak areas were calculated and OTU's that made up less than 1% of the total composition were eliminated from the analysis to eliminate the variable low abundance components within the communities.   A proprietary multitag pyrosequencing(MTPS) process was then used to characterize the microbiota [112]. Briefly, a set of 96 emulsion PCR fusion primers were generated by combining the 454 emulsion PCR linkers and unique 8 base identifiers on either of the 27F or 355R universal 16S rRNA primers. As a result each sample was amplified with a uniquely barcoded set of primers that allows for up to 96 samples to be pooled. The samples were then sequenced using a GS-FLX pyrosequencer (Roche) and sorted using custom PERL scripts.

2.4 RDP10 Analysis

Identification of the taxa present in each of the samples was done using the Bayesian analysis tool in Version 10 of the Ribosomal Database Project. A 180 base pair cut-off was used to filter the data and the taxa present at greater than 1% abundance of the community were tabulated at the phylum, class, order, family and genus levels.

2.5 QIIME (Quantitative Insights into Microbial Ecology) Analysis

The QIIME pipeline was used as a secondary method to identify the bacterial taxa in the biopsies. A minimum quality score of 25, no ambiguous bases were allowed and no mismatches were allowed to assign samples to taxonomic groups. Operational taxonomic units were selected using the default QIIME settings of uclust[113] and a similarity sequence threshold of 0.97. Lastly, like the RDP10 analysis taxa a 1% abundance cut-off was used to filter the taxonomic groups.

2.6 Urine Sample Preparation and Metabolomic Analysis

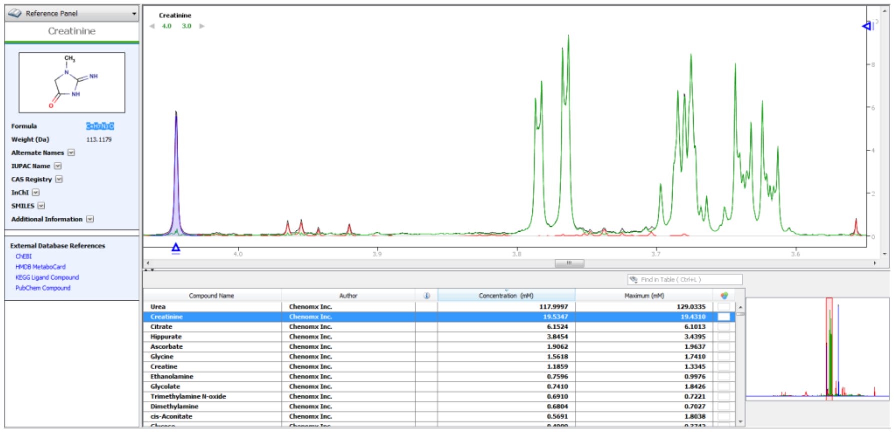
Urine samples were removed from the -80°C freezer and allowed to thaw. 720μl of the sample was added to 80μl of DSS (4, 4-dimethyl-4-silapentane-1-sulfonic acid) Chenomx Standard (IS2; 4.6485mM DSS) and then the samples were pH corrected to a pH of between 6.7-6.8. Solutions of 1.0M, 0.1M and 0.01M NaOH and HCl were used to obtain the consistent pH. Finally the 700µl of each samples were transferred into 4 inch long 5mm diameter NMR tubes and capped.

The samples were analyzed using an Oxford 600Hz NMR spectrometer with a Varian VNMRS two channel console and running VNMRJ software version 2.2C on a RHEL 4 host computer.   Before samples were inserted into the robot rack, the outside of the tubes were first cleaned with ethanol and kimwipes® to remove any debris or oils from handling. Shims were optimized until an optimal line width value was obtained at relative peaks heights of: 50% (< 1.0 Hz), 0.55% (< 12.0 Hz), and 0.11% (< 20.0 Hz) was achieved. All sample handling was done with a Varian 768 AS sample handling robot. Any spectra that did not meet acceptable line height values were discarded and the sample was re-run manually.

After spectra have been obtained, samples were removed from NMR tubes and the pH was rechecked to ensure that the pH had not shifted a significant amount.

2.7 Spectral analysis

The analysis of the NMR spectra was done with Chenomx Inc. NMR suite software version 7.6. An example of a spectrum is shown below (Figure 2.7.1) Spectra were pre-processes where the phasing of the spectra was aligned, the water peak was eliminated, the baseline was corrected and a final shim correction was done. Identification of the metabolites was done using the Chenomx chemical compound library and concentrations were quantified using the Chenomx standard that was added to all the samples. A total of 67 metabolites were identified, quantified, and also verified by Chenomx.



Formula $C_4H_7N_3O$
Weight (Da) 113.1179
Alternate Names
IUPAC Name
CAS Registry
InChI
SMILES
Additional Information

External Database References
ChEBI
HMDB MetaboCard
KEGG Ligand Compound
PubChem Compound

Compound Name	Author		Concentration (mM)	Maximum (mM)
Urea	Chenomx Inc.		117.9997	129.0335
Creatinine	Chenomx Inc.		19.5347	19.4310
Citrate	Chenomx Inc.		6.1524	6.1013
Hippurate	Chenomx Inc.		3.8454	3.4395
Ascorbate	Chenomx Inc.		1.9062	1.9637
Glycine	Chenomx Inc.		1.5618	1.7410
Creatine	Chenomx Inc.		1.1859	1.3345
Ethanolamine	Chenomx Inc.		0.7596	0.9976
Glycolate	Chenomx Inc.		0.7410	1.8426
Trimethylamine N-oxide	Chenomx Inc.		0.6910	0.7221
Dimethylamine	Chenomx Inc.		0.6804	0.7027
cis-Aconitate	Chenomx Inc.		0.5691	1.8038

Figure 2.7.1 An example spectra of a urine sample with aligned phasing, baseline correction and removal of the water peak using Chenomx NMR Suite 7.6 software.

2.8 Statistical analysis

Univariate and multivariate statistical analysis was done with the aid of the Waikato Environment for Knowledge Analysis (Weka)[114]. Weka is a collection of machine learning algorithms with tools for pre-processing, classification and clustering. More specifically Weka was used to build decision trees, SVM, and naïve Bayes classifiers. In addition Weka was used for all feature selection algorithms including best first (forward and backward) as well as greedy search (forward and backwards). N-fold cross validation was used to evaluate these models as well as ROC (Receiver operating characteristic) curves. A good ROC area is close to 1, while a random classifier would be 0.5. Statistical analysis was also done using the online tool MetaboAnalyst 2.0 [115]. MetaboAnalyst was used to perform PCA, PLS-DA and clustering. PLS-DA plots are evaluated with cross validation in which two values, R^2 and Q^2, are used to determine how well the model fits the data and the predictive error of the model respectively. A good PLS-DA model has values of R^2 and Q^2 close to each other and close to 1.

2.9 Correlation Analysis

The microbiota features at the class and genus levels along with the metabolites were correlated using a Pearson's correlation function and then filtered for correlations >0.45 and $p<0.05$. A correlation matrix was then generated to visualize the clusters of correlations. In addition correlation networks were created using Cytoscape software as another means to visualize the correlations.

Results

Chapter 3: Comparison of colonic and ileal samples

3.1 Bacterial Composition

As our sample biopsy set contained both colonic and ileal samples, I initially compared the bacterial communities of the samples to see if there were any differences between the two. Other studies have shown that the ileocecal valve acts as a barrier for microbial communities, and after resection and removal of the ileocecal valve, the bacterial communities in the colon and ileum become homogenous [116]. Based upon this, I hypothesized that the samples would not be different, and therefore I would be able to include the entire data set from the biopsies in further analyses.

There were a total of 38 biopsies collected for this study. 30 samples were ileal biopsies while 8 were colonic biopsies. There was no difference in bacterial composition between the colonic and ileal biopsies at the genus level. The PCA score plot and the PLS-DA score plot did not show any separation between the two sets of biopsies. (Figure 3.1.1) In addition, cross validation of the PLS-DA model reveals R^2 and Q^2 values of 0.27 and -0.18 respectively suggesting the model is fitting a random correlation in the data. (The values are not close to 1 and are not similar to each other)

The only sample that appears to be different from the rest is sample GR5621. (Figure 3.1.1) This biopsy was the only one to be stored in RNAlater which is not supposed to affect the DNA at all but our data suggest it had some sort of effect here. (See appendix for a detailed breakdown of biopsy location) Upon closer inspection this sample was missing some of the Firmicutes, more specifically *Leuconostocaceae*. Based on these findings, this sample was discarded from the rest of the samples in further analyses.

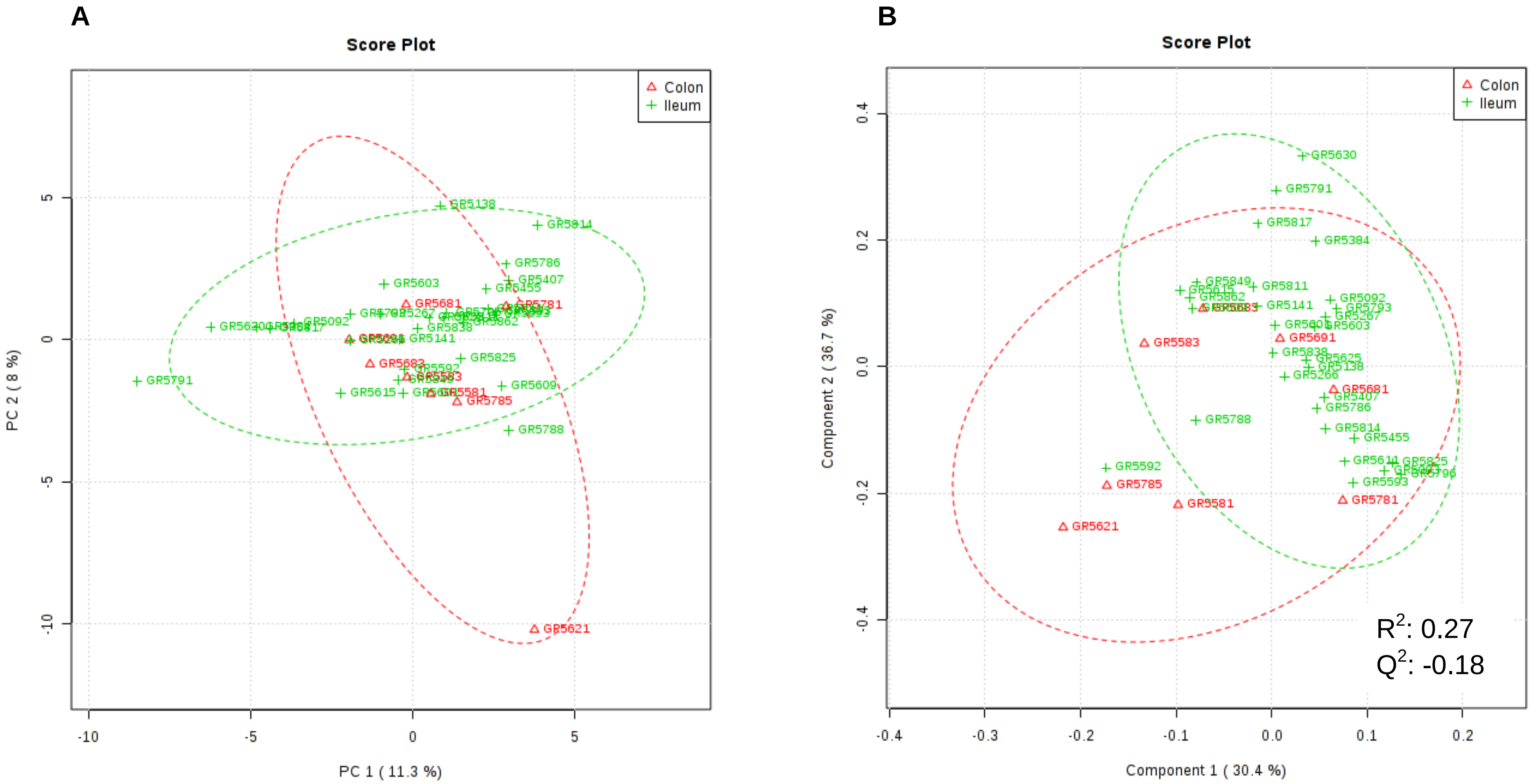


Figure 3.1.1 A) PCA Score plot and **B)** PLS-DA Score plot of the bacterial composition of colonic (red triangles) and ileal (green crosses) samples consisting of 70 different bacteria at the genus level.

As seen in Table 3.1.1, other machine learning algorithms were also unable to distinguish the colonic biopsies from the ileal biopsies. (The baseline accuracy starts at 78.9% since 30 out of the 38 samples are ileal biopsies.) In addition log transforming the data had little effect on the results.

Table 3.1.1 Machine learning performance on colonic and ileal biopsies

			Log Transformation	
Algorithm	Accuracy	ROC Area	Accuracy	ROC Area
J48 Decision Tree	71.1%	0.656	71.1%	0.652
Naïve Bayes Classifier	73.7%	0.480	73.7%	0.479
Support Vector Classifier	71.1%	0.496	73.7%	0.513

In fold feature selection (Tables 3.1.2) did not significantly improve results with any of the algorithms. The feature selection algorithms actually all selected the same features, which is why the results are the same regardless of what feature selection algorithm was used.

Table 3.1.2 Machine learning performance on colonic and ileal biopsies with feature selection

	Feature Selection			
	Best First		Greedy	
Algorithm	Accuracy	ROC Area	Accuracy	ROC Area
J48 Decision Tree	78.9%	0.363	78.9%	0.363
Naïve Bayes Classifier	78.9%	0.644	78.9%	0.644
Support Vector Classifier	73.7%	0.467	73.7%	0.467

Similar results were seen at the class, order and family taxonomic levels.

Therefore, for the remainder of the analyses the colonic and ileal biopsies were considered homogeneous and the assumption was made that any difference in bacterial composition was not due to location of the biopsy.

Chapter 4: Comparison of patients in remission with those suffering relapse

4.1 Patient Population

The main aim of this study was to determine if there were any differences between Crohn's disease patients currently in relapse and those currently in remission that had previously undergone an ileocolonic resection.

There were no significant differences in age, gender, years since last surgery, or current medications between CD patients suffering from post-operative endoscopic recurrence (n=18) compared with those in remission (n=20). (Table 4.1.1) A patient was considered in endoscopic remission if their Rutgeerts score was 0, 1, or 2 and relapse if their score was 3 or 4. (Figure 4.1.1)

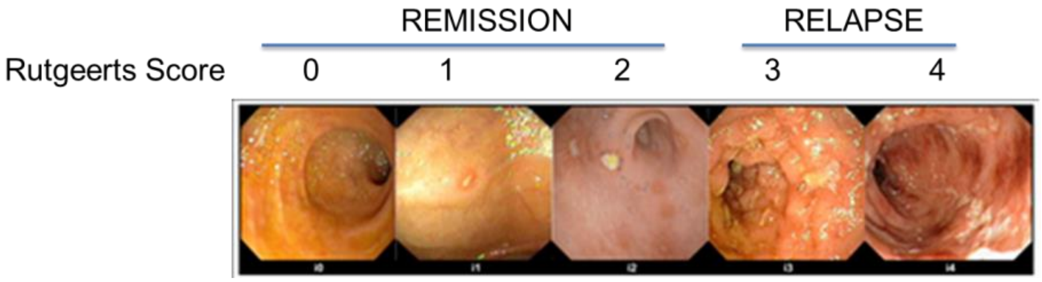


Figure 4.1.1 Remission patients had a Rutgeerts score of 0, 1 or 2, while patients in endoscopic recurrence had a score of 3 or 4. (Image from Dr. Troy Perry)

Table 4.1.1 Patient information

Category	Remission	Relapse
Number of Patients	20	18
Mean Age ± SD, years **(Range)**	45.5 ± 12.6 (23- 68)	49.0 ± 12.2 (21-69)
Years since last surgery **Mean ± SD (Range)**	9.0 ± 8.1 (3-35)	12 ± 10.3 (3-41)
Gender • **Male** • **Female**	 6 (30%) 14 (70%)	 6 (33%) 12 (67%)
Current Medications		
• 5-ASA	3(15%)	5(28%)
• Steroids	2 (10%)	3 (17%)
• Immuno-modulators	9 (45%)	5 (28%)
• Biologics	3 (15%)	5(28%)
• Antibiotics	2 (10%)	1 (6%)

4.2 Bacterial Composition: RDP10 Taxonomic Identification

The microbial environment of the gut has been shown to be a significant factor in CD pathogenesis and recurrence. Therefore the bacterial composition was compared between the two groups of patients to determine whether specific differences in bacterial composition could be identified in patients in post-operative endoscopic recurrence or remission.

As shown in Figure 4.2.1 the individual bacterial composition is highly individualized within each patient. For example the *Bacteroidia* class has a range from 0-63% depending on the individual.

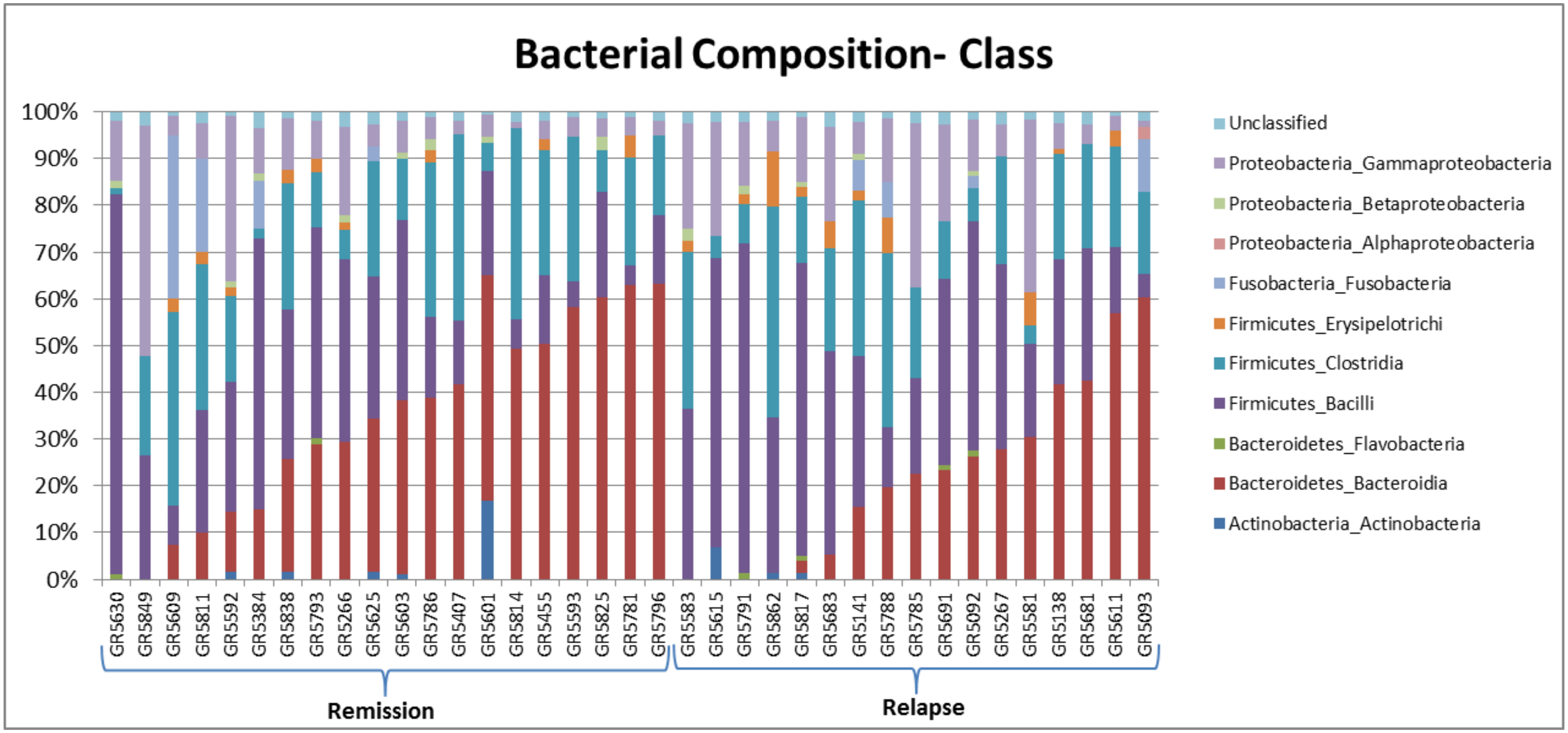


Figure 4.2.1 Individual bacterial composition at the class level of intestinal biopsies. Remission patients are on the left of the chart and relapse patients are on the right side. The percent abundance is on the y-axis and the patients are on the x axis.

After averaging the bacterial abundances for the remission patients and relapse patients there were some significant differences between the two classes. The remission patients had significantly more *Bacteroidia* (*p*= 0.04) and significantly less *Erysipelotrichi* (*p*=0.04) compared to the relapse patients using the students t-test. (Figure 4.2.2)

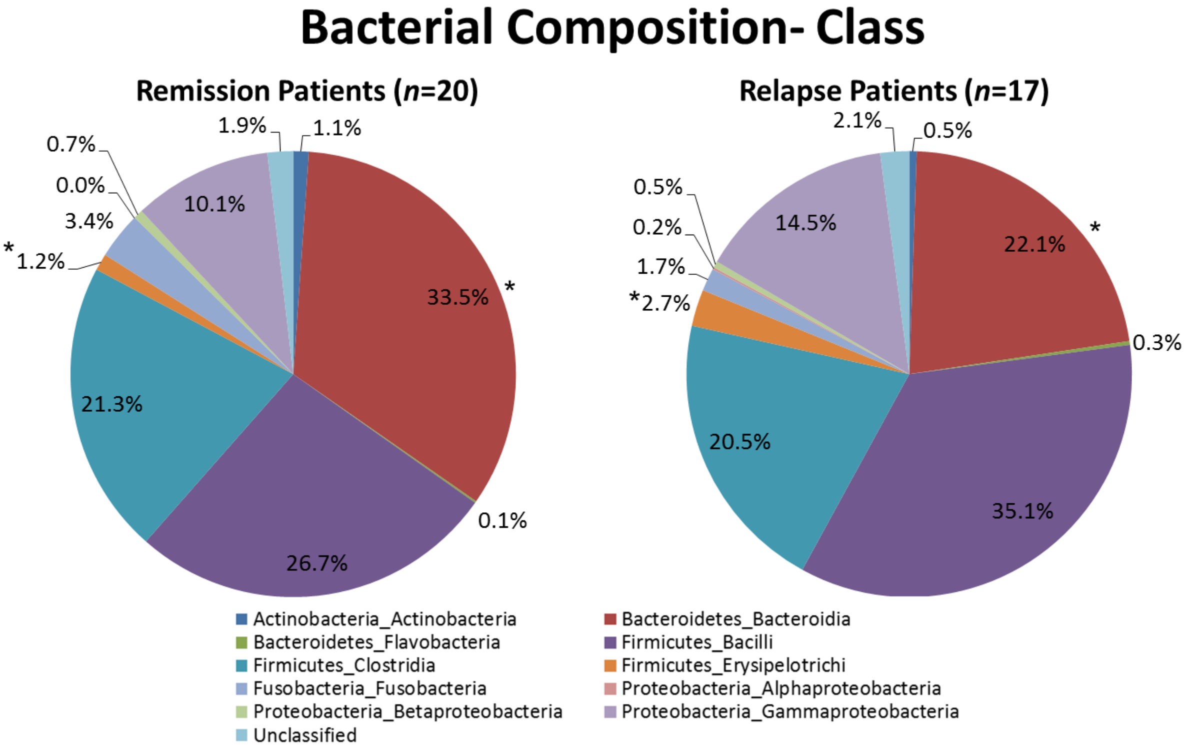


Figure 4.2.2 Averaged bacterial composition at the class level of remission patients (left) and relapse patients (right) Asterisk indicated a significant difference (p <0.05) using the students t-test in the relative amounts of *Bacteroidia* and *Erysipelotrichi* between the two groups.

Despite the individual differences in composition between the remission and relapse patients, no separation was achieved based on the average bacterial composition between the relapse and remission patients at any taxonomic level using PCA or PLS-DA. The PCA score plot (Figure 4.2.3A) did not show any sort of natural groups forming between the two sets of biopsies. Although the PLS-DA score plot (Figure 4.2.3B) shows a separation this is caused by overfitting. Overfitting occurs when a model fits the random error or noise in the data instead of the underlying relationship. Models that overfit the data normally perform very poorly during cross validation. Cross validation of this PLS-DA model reveals R^2 (how well the model fits the training data) and Q^2 (predictive ability of the model) values of 0.68 and -0.21 respectively suggesting the model is fitting a random correlation in the data. (The values are not close to 1 and are not similar to each other) R^2 is variation explained by the model and Q^2 is the prediction error measure.

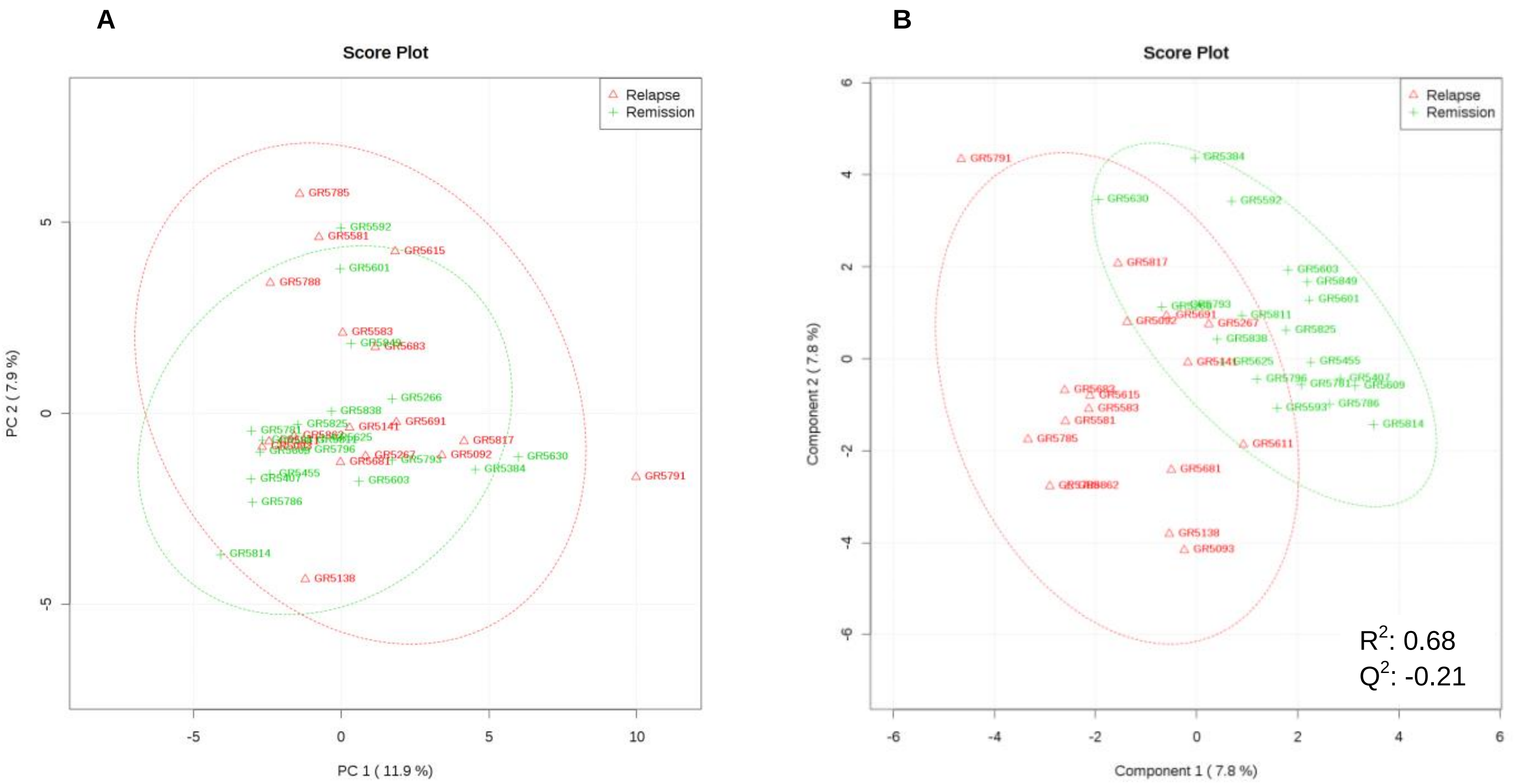


Figure 4.2.3 A) PCA score plot and **B)** PLS-DA score plot of the bacterial composition of relapse (red triangles) and remission (green crosses) patients consisting of 70 different bacteria at the genus level.

Next a variety of machine learning algorithms were used on the bacterial data. As seen in Table 4.2.1, other machine learning algorithms were also unable to separate the remission patients from the relapse patients at the genus level. All the algorithms tested using an n fold cross validation to ensure the most data was used for training as possible.

Lastly the in fold feature selection algorithms, best first (forward and backwards) and greedy (forward and backwards) completely failed because all the algorithms picked out a single feature and the resulting machine learning algorithms just classified all cases as remission .

Table 4.2.1 Machine learning performance on remission and relapse biopsies

Algorithm	Accuracy	ROC Area	Log Transformation	
			Accuracy	ROC Area
J48 Decision Tree	45.9%	0.428	51.4%	0.321
Naïve Bayes Classifier	51.4%	0.500	51.4%	0.500
Support Vector Classifier	54.1%	0.500	54.1%	0.513

Similar results were seen at the class, order and family levels.

4.3 Bacterial Composition: QIIME Pipeline Taxonomic Identification

The QIIME (Quantitative Insights into Microbial Ecology) pipeline was used to verify the taxonomic classification of the original RDP 10 classifier and to obtain diversity indices.

Overall the taxonomic classification was very similar between the two identification methods with the exception of two differences. The first difference was in the *Erysipelotrichi* bacteria, which were found to be significantly different between remission and relapse groups in the initial classification. In the QIIME pipeline, this was no longer a significant difference. Upon closer inspection the discrepancy was caused by less *Erysipelotrichi* being assigned to the relapse patients in the QIIME pipeline.

The second difference was the occurrence of a significant increase in *Gammaproteobacteria* in the relapse patients (*p*=0.02) (Figure 4.3.1). This increased abundance of *Gammaproteobacteria* in patients with active disease has been previously reported [117]. In addition the *Bacteroidia* class was still different (p=0.02) between the two groups as seen in the RDP10 classification.

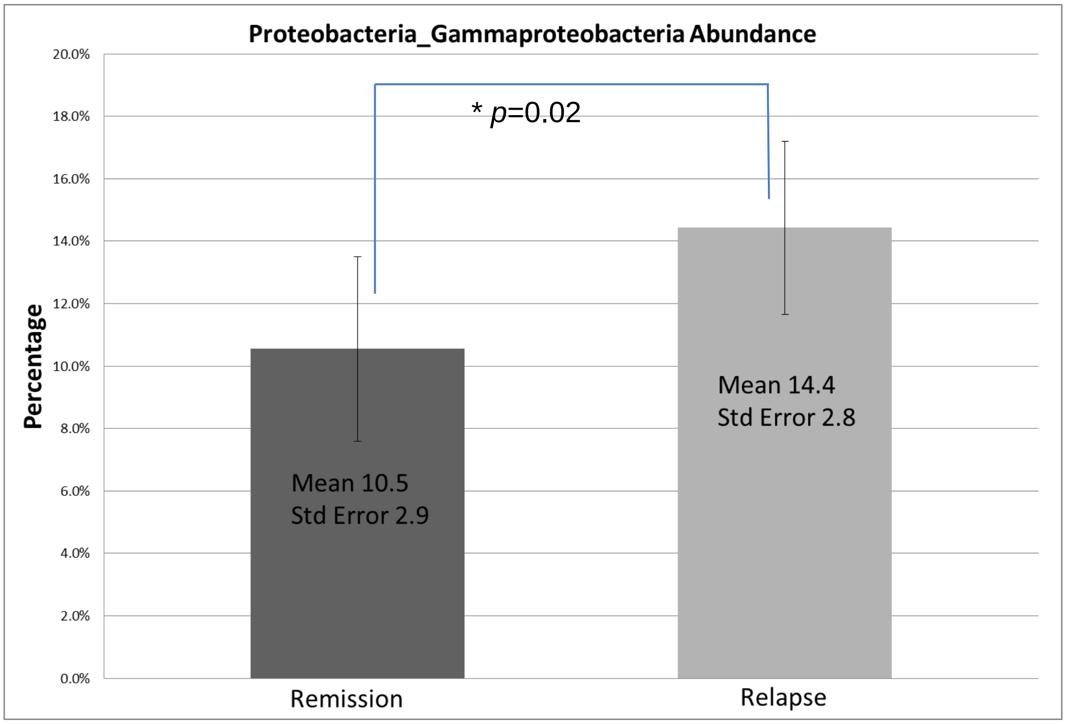


Figure 4.3.1 Average abundance of *Gammaproteobacteria* in remission patients (left) and relapse patients (right).

There were certain bacteria that each method identified in which the other did not as summarized in Table 4.3.1 but these were primarily the low abundance bacteria that were only identified in particular samples.

Table 4.3.1 Unique bacteria identified by each classification method that were not identified by the other method

RDP 10	QIIME
- *Flavobacteriaceae (Bacteroidete)*	- *Bacteroidales S24-7 (Bacteroidete)*
- *Eubacteriaceae (Firmicute)*	- *Barnesiellaceae (Bacteroidete)*
- *Incertae Sedis XIV (Firmicute)*	- *Paraprevotellaceae (Bacteroidete)*
- *Methylocystaceae (Proteobacteria)*	- *Turicibacteraceae (Firmicute)*
	- *Pseudomonadaceae (Proteobacteria)*

The chao alpha diversity is a measure of species diversity within a community. The alpha diversity (± SD) of the remission patients was 145 ± 33 and 149 ± 55 for the relapse patients indicating that the abundance of different taxa within each separate community was very similar.

Overall the taxonomic classification of the 16s rRNA sequences was very similar between the two identification methods (RDP 10 and QIIME) but the slight differences did result in some discrepancies in what the data shows. Each involved different pipelines which resulted in a slightly different assignment of bacteria. Possible reasons for this difference are the way each method deals with the filtering of sequences. For example the filtering of chimeras, which are organisms that are comprised of two or more different populations of genetically distinct cells, are handled differently in each pipeline resulting in different classifications of that specific sequence. The QIIME pipeline filters the chimeras completely out while the RDP10 classifier leaves them in and assigns the sequence to the most probable bacteria.

4.4 Metabolomics

The urinary metabolomics signature was looked at next in these patients. Metabolites are the final chemical products of metabolism and are a direct measurement of the function occurring in the biological system. A large number of urinary metabolites are produced by the gut bacteria and the measurement of these metabolites can be used to reflect the function of the gut bacteria. The bacteria composition itself was unable to distinguish the remission patients from the relapse patients but perhaps the functional aspect of these bacteria would be able to.

There was no difference based on the urinary metabolic signature between the relapse and remission patients. No metabolites were significantly different between the two groups of patients and in addition the PCA score plot (Figure 4.4.1) did not show any separation between the urine samples. Although the PLS-DA score plot (Figure 4.4.1) shows a separation this is caused by over fitting since cross validation of the PLS-DA model reveals R^2 and Q^2 values of 0.70 and -0.54 respectively suggesting the model is fitting a random correlation in the data. (The values are not close to 1 and are not similar to each other.)

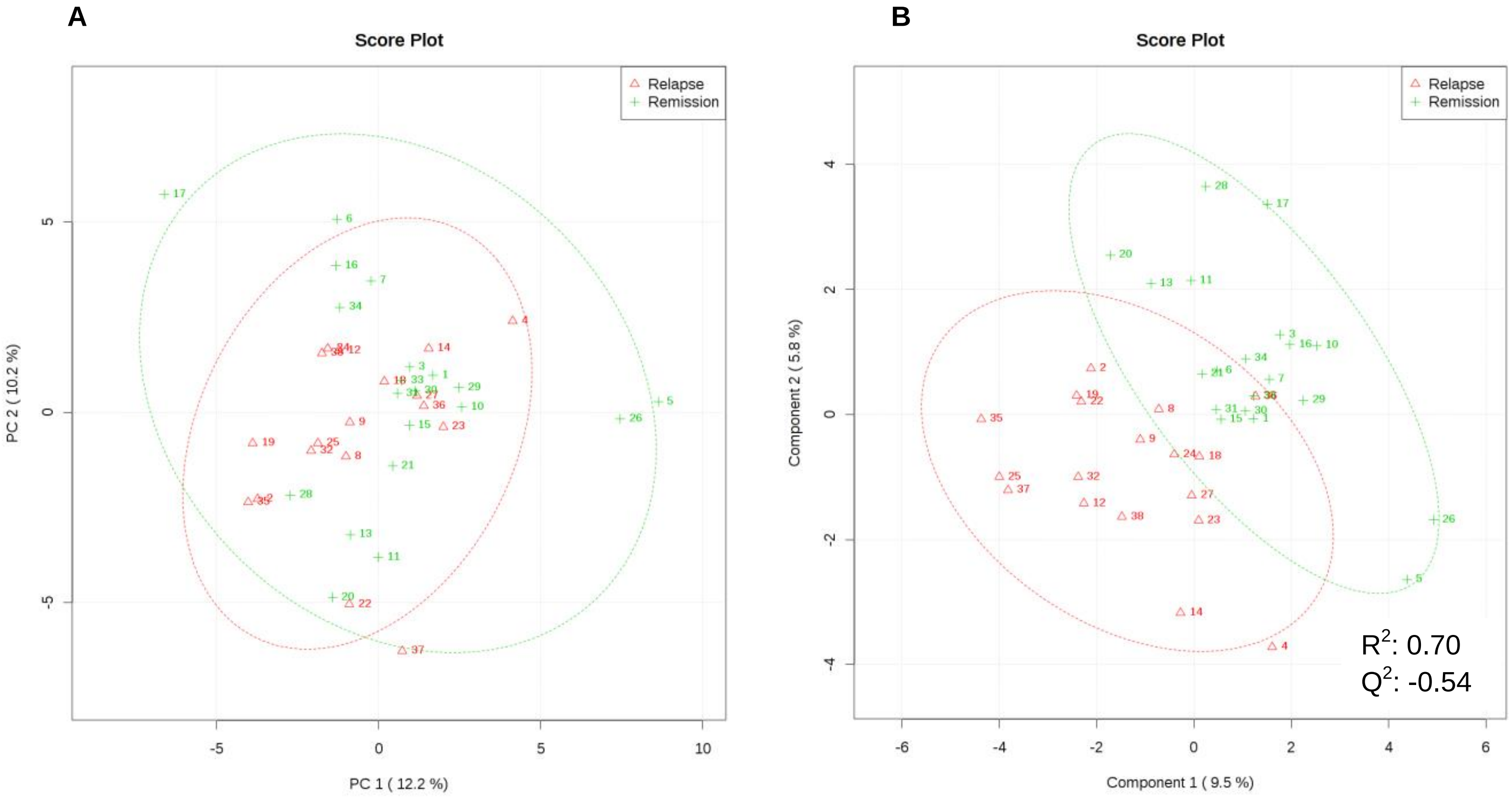


Figure 4.4.1 A) PCA Score plot and **B)** PLS-DA score plot of the urinary metabolomic composition of relapse (red triangles) and remission (green crosses) patients consisting of 67 different metabolites.

Table 4.4.1 Machine learning performance on remission and relapse patient's urinary metabolites

Algorithm	Accuracy	ROC Area	Log Transformation		Feature Selection	
			Accuracy	ROC Area	Accuracy	ROC Area
J48 Decision Tree	44.7%	0.421	44.7%	0.426	18.4%	0.018
Naïve Bayes Classifier	44.7%	0.409	47.4%	0.538	44.7%	0.522
Support Vector Classifier	44.7%	0.450	47.4%	0.450	47.4%	0.450

As seen in Table 4.4.1 none of the machine learning algorithms including infold feature selection were able to create a model to distinguish between the relapse and remission patients based on the urinary metabolites. All of the feature selection algorithms used selected the same features.

Although there were no significant differences in urinary metabolite concentrations, there were some metabolites that had a greater than 2 fold difference and/or p-value that trended towards significance (p>0.1) between the groups as shown in Table 4.4.2.

Table 4.4.2 Metabolites of interest (>2 fold difference or *p<0.1)*

Metabolite	Fold Change (Relapse/Remission)	P value
1, 6-Anhydro-β-D-glucose	3.8	0.08
2-Hydroxyisobutyrate	-2.0	0.09
3-Hydroxyisovalerate	-1.5	0.07
4-Hydroxyphenylacetate	2.1	0.08
Acetoacetate	-2.7	0.16
Acetone	-3.2	0.13
Creatine	-1.9	0.09
Fucose	1.6	0.06
Glycolate	1.3	0.08
Methanol	-2.4	0.10
Propylene glycol	-3.5	0.14
Quinolinate	1.6	0.05
Tyrosine	1.6	0.08

Table 4.4.3 provides a description of the metabolites that were found to be possibly different. Interestingly fucose is a metabolite that has been previously associated with intestinal disorders. More specifically fucose could be used as a marker for gastric ulcers [118].

Table 4.4.3 Description of metabolites of interest (>2 fold difference or *p<0.1).*

Metabolite	**Description**
1, 6-Anhydro-β-D-glucose	-formed from the pyrolysis of carbohydrates. -highly correlated with regional fired and biomarker for wood smoke exposure [119].
2-Hydroxyisobutyrate	-metabolite of methyl tert-butyl ether, and obtained through environmental exposure, usually rapidly eliminated from the body -derivative of butyrate which is produced by many bacteria in the *Clostridia* class and also by specific bacteria such as *Faecalibacterium prausnitizii* and *Roseburia sp*. [120] -associated in lung cancer [121].
3-Hydroxyisovalerate	-normal human metabolite -associated in many diseases including lung cancer and type 1 diabetes [121, 122].
4-Hydroxyphenylacetate	-metabolite of tyrosine via enteric bacteria -associated in hypertension, and preterm infants [123, 124].
Acetoacetate	-organic acid and can be produced in the human liver under certain conditions of poor metabolism leading to excessive fatty acid breakdown -associated in diabetes [125].
Acetone	-ketone bodies produced during ketoacidosis -by product of fermentation and produced by several bacteria in the *Clostridia* class *including Clostridium butylicum* and *Clostridium aurantibutyricum[126]* -associated in breast and lung cancer[121, 127]
Creatine	-amino acid that occurs in vertebrate tissues and in urine -associated in cirrhosis and lung cancer [128, 129].
Fucose	-monosaccharide that is a common component of many glycolipids produced by mammalian cells. -*Bacteroidetes* actively degrade fucose into useful monosaccharides for the body to use. -associated in cirrhosis and gastric ulcers [118].
Glycolate	-smallest alpha-hydroxy acid. Glycolic acid finds applications in skin care products. -associated in lung cancer and biliary atresia [121, 130].
Methanol	-simplest alcohol. Toxicity is due to the metabolic products of alcohol dehydrogenase. -associated in alcoholism [131].
Propylene glycol	-used as a solvent for intravenous, oral, and topical pharmaceutical preparation. -associated in lung cancer [121].
Quinolinate	-metabolite of tryptophan -associated in neurodegenerative disorders [132].
Tyrosine	-essential amino acid that readily passes the blood-brain barrier -associated in Alzheimer’s and cachexia [121, 133].

Information from the Human Metabolome Database

4.5 Environmental questionnaire

Previous research has suggested that some sort of environmental trigger is necessary for triggering a relapse after surgery. Several environmental factors have been implicated including smoking, diet, socioeconomic status, stress, childhood exposures, medications, and/or infectious agents [30].

The CEGIIR environmental questionnaire was developed to try to capture these suspected factors. Thus, by examining the responses between Crohn's disease patients currently in relapse and patients currently in remission, factors that are involved in post-operative recurrence may be identified.

A total of 881 true/false questions were extracted from the CEGIIR questionnaire. After removing the questions that were either all true or all false for all the patients, 262 potentially meaningful questions remained.

Using all of these meaningful questions to try to build a model to separate the remission patients from the relapse patients was not successful. As seen in Tables 4.5.1 and 4.5.2 none of the machine learning algorithms including infold feature selection were able to distinguish between the two groups of patients.

Table 4.5.1 Machine learning performance on remission and relapse questionnaire data

Algorithm	Accuracy	ROC Area
J48 Decision Tree	52.5%	0.518
Naïve Bayes Classifier	42.5%	0.363
Support Vector Classifier	45.0%	0.444

Table 4.5.2 Machine learning performance on remission and relapse questionnaire data with feature selection

Algorithm	Feature Selection			
	Best First Forward		Greedy Forward	
	Accuracy	ROC Area	Accuracy	ROC Area
J48 Decision Tree	40.0%	0.075	40.0%	0.075
Naïve Bayes Classifier	52.5%	0.020	52.5%	0.020
Support Vector Classifier	50.0%	0.476	50.0%	0.476

There were no significant differences in the environmental factors between the two groups using the Fisher's exact test, (See Appendix 3 for table of p-values.) However, looking at each question individually did reveal some interesting environmental factors that appeared to show a trend of being different between the remission and relapse patients with this small sample size. (Figure 4.5.1) Patients in relapse appeared to be smoking more marijuana, trying more herbal drugs, and more were currently on prednisone and aspirin. On the other hand, remission patients seemed to have more exposure to second hand smoke from their spouses and more had a high weekly sugar intake. More patients in relapse had previously been on asacol but more patients in remission had previously been on prednisone, Entocort capsules, azathioprine, and Pentasa. In addition more remission patients had previously taken birth control pills in their twenties. Lastly, early childhood exposures seemed to be different between the remission and relapse patients. More patients in relapse had had high fibre foods as a child, but more patients in remission had animal exposure on a farm, ate fish, and drank non tap water in their childhood.

Smoking cigarettes has been the only well replicated environmental factor associated with CD relapse [31-33], but this finding of smoking marijuana may be related and be a novel finding. On the other hand more patients in remission are being exposed to second hand smoke at home which somewhat contradicts previous literature. Other contradictions to literature include a higher sugar intake for remission patients [134].

In addition early childhood exposures of drinking non tap water, animal exposure on a farm supports the hygiene hypothesis for CD. (Figure 4.5.1)

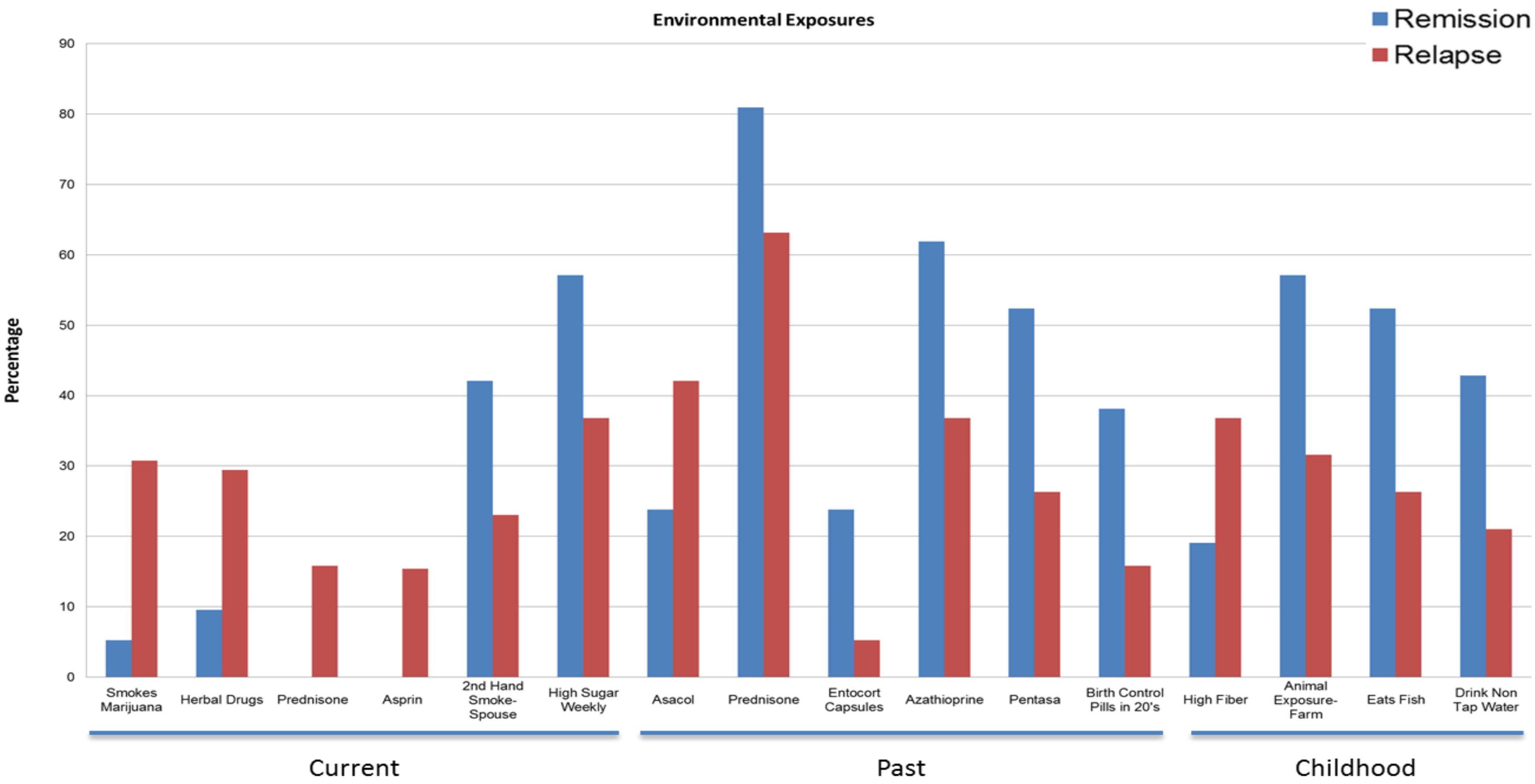


Figure 4.5.1 Percentage of remission patients (blue) and relapse patients (red) that are currently (left), previously (middle) or in their childhood (right) been exposed to certain environmental factors.

4.6 Correlations between bacteria and metabolites

A correlation analysis was done next in an attempt to link the bacterial species with the urinary metabolites.

Pearson correlations were calculated between the urinary metabolites and the bacteria at the class and genus levels and then filtered for correlations greater than 0.45 and p<0.05. A correlation matrix was then generated to visualize the clusters of correlations. (Figure 4.6.1 and 4.6.2) Both positive and negative correlations between urinary metabolites and microbes at the class and genus level were seen. Patients in endoscopic remission had positive correlations between metabolites involved in energy production such as citrate, fumerate, glucose, glutamate, pyruvate, and succinate with *Enterobacteriaceae*. Positive correlations were also apparent between metabolites involved in amino acid metabolism such as alanine, benzoate, histidine, glycine, taurine, and uracil with *Enterobacteriaceae*.

In contrast, patients in endoscopic relapse had positive correlations between energy production (citrate, aminoisobutyrate, creatine, fumerate, malonate, and pyruvate) and amino acid metabolism (phenylalanine, methyhistidine, and tyrosine) with *Bacteroides*. Additionally, the relapse patients had negative correlations between amino acid metabolism (alanine, leucine, glutamine, glycine, ethanolamine, and glycine) with *Clostridia.*

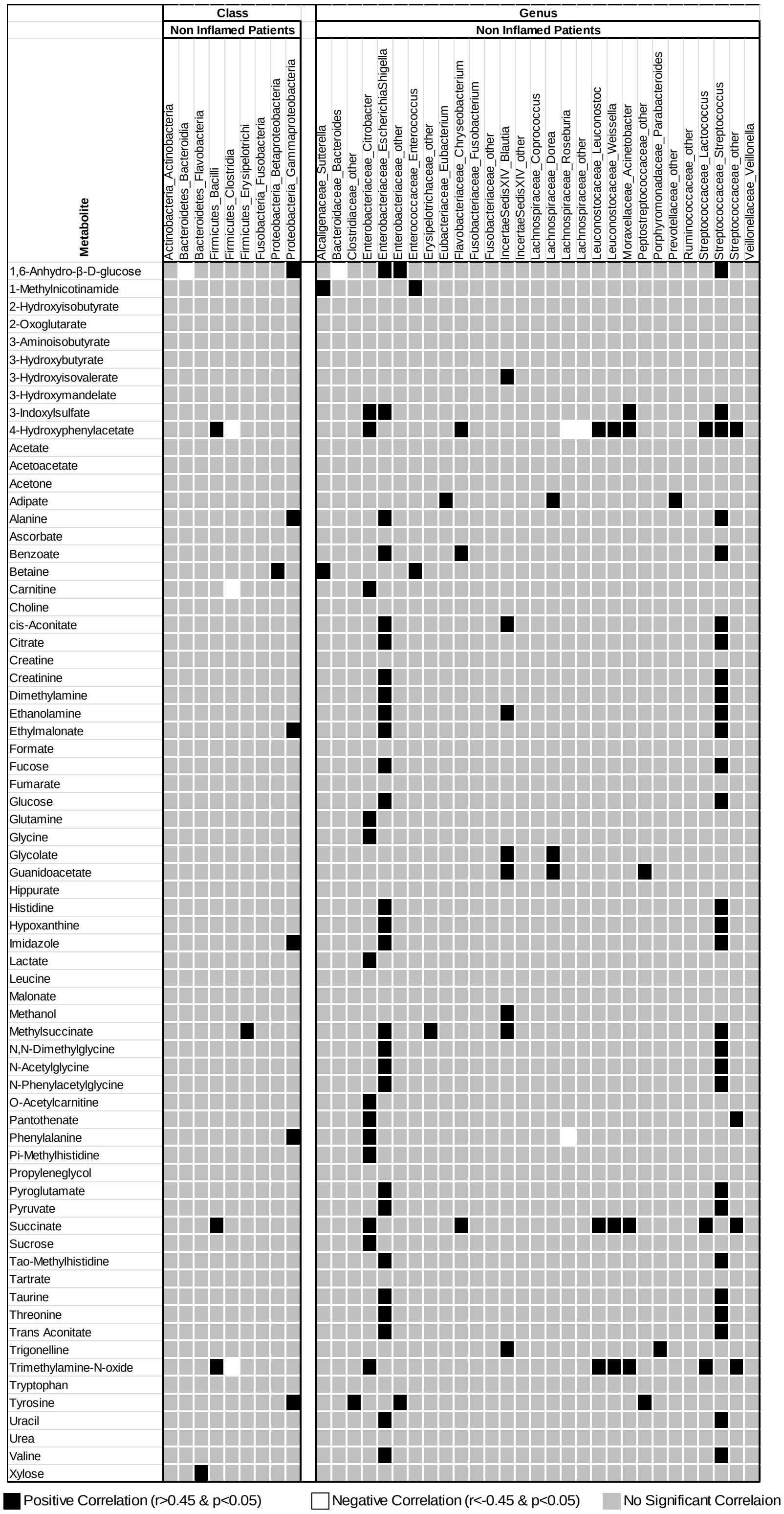


Figure 4.6.1 Significant Correlations in remission patients

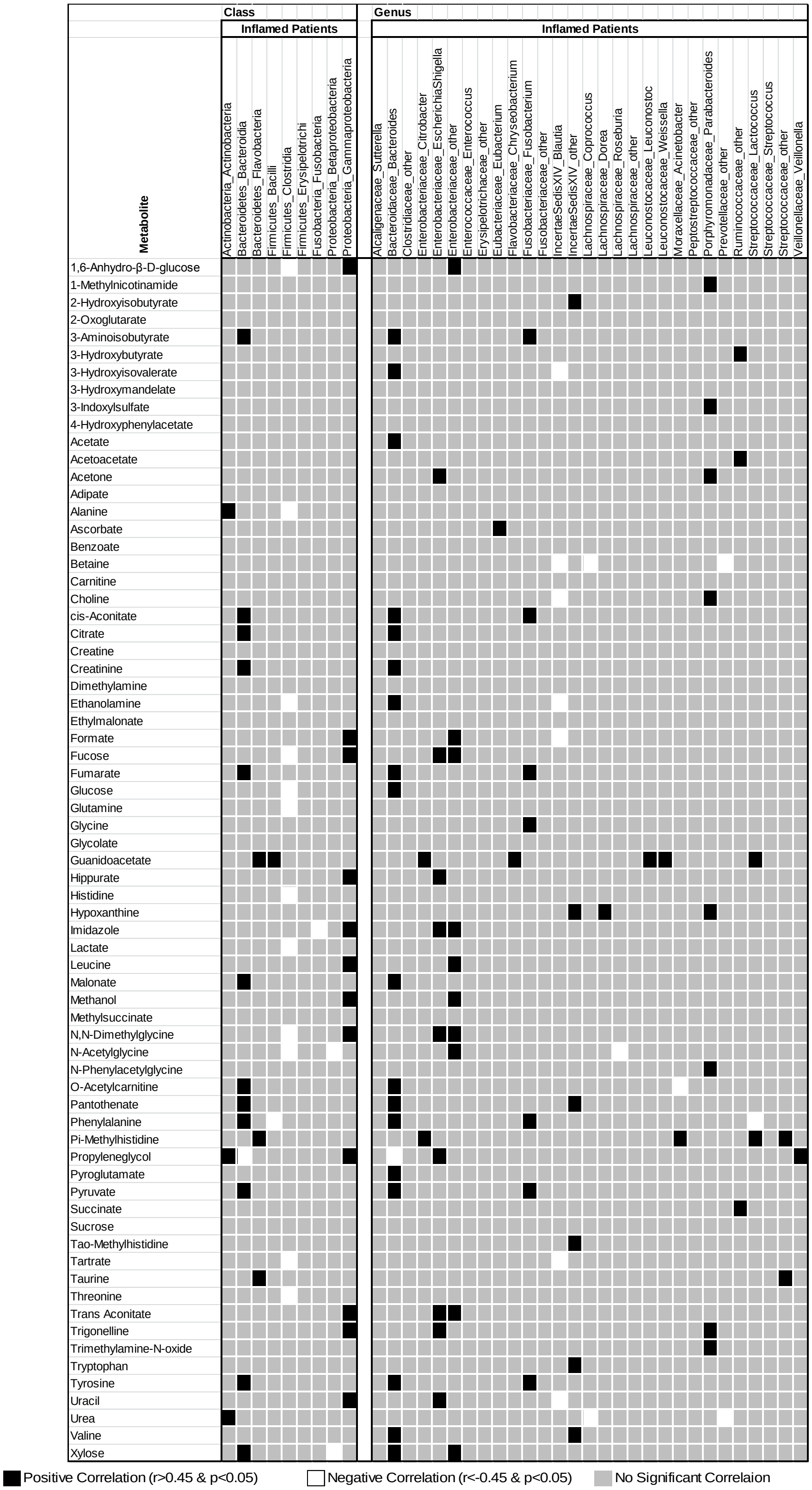


Figure 4.6.2 Significant correlations in relapse patients

Correlation networks were then performed separately for relapse and remission patients using spearman rank correlations with coefficients that were greater than 0.45 and p<0.05. They were imported into Cytoscape for visualization. These networks are more of a systems biology approach and most useful as a discovery tool where specific correlations of interest can be investigated further by importing and merging these correlations with known networks.

The correlations between the bacteria alone consist of 56 correlations between 6 nodes for the remission patients (Figure 4.6.3) and 45 correlations between 26 nodes for the relapse patients (Figure 4.6.4). The individual networks allow an alternate visual representation of the correlations. The clustering coefficient is a measure of degree to which the nodes in a network tend to cluster together. It is a value between 0 and 1. A value of 1 means the number of edges between neighboring nodes is the maximum possible number of edges. The clustering coefficients are 0.411 and 0.545 for the remission and relapse networks respectively suggesting they are similar.

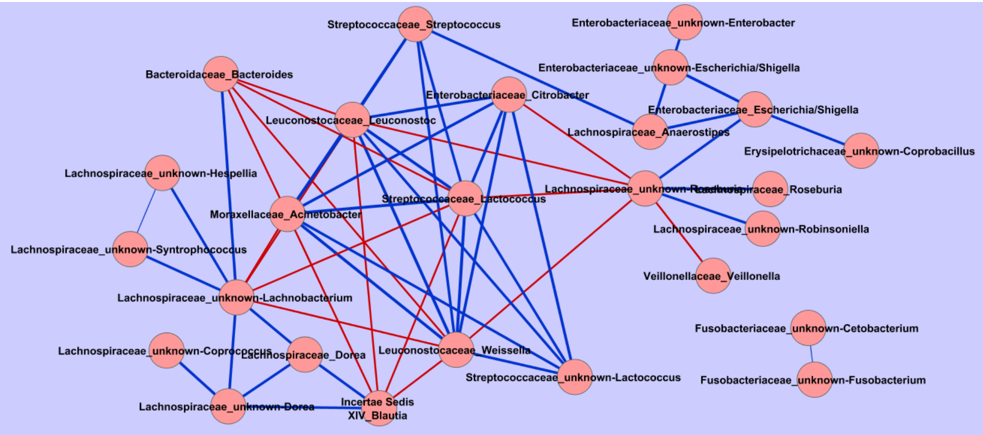

Streptococcaceae_Streptococcus
Enterobacteriaceae_unknown-Enterobacter
Enterobacteriaceae_unknown-Escherichia/Shigella
Bacteroidaceae_Bacteroides
Enterobacteriaceae_Citrobacter
Enterobacteriaceae_Escherichia/Shigella
Leuconostocaceae_Leuconostoc
Lachnospiraceae_Anaerostipes
Erysipelotrichaceae_unknown-Coprobacillus
Lachnospiraceae_unknown-Hespellia
Lachnospiraceae_unknown-Roseburia
Lachnospiraceae_Roseburia
Streptococcaceae_Lactococcus
Moraxellaceae_Acinetobacter
Lachnospiraceae_unknown-Robinsoniella
Lachnospiraceae_unknown-Syntrophococcus
Veillonellaceae_Veillonella
Lachnospiraceae_unknown-Lachnobacterium
Fusobacteriaceae_unknown-Cetobacterium
Lachnospiraceae_unknown-Coprococcus
Lachnospiraceae_Dorea
Leuconostocaceae_Weissella
Streptococcaceae_unknown-Lactococcus
Fusobacteriaceae_unknown-Fusobacterium
Lachnospiraceae_unknown-Dorea
Incertae Sedis
XIV_Blautia

Figure 4.6.3 Bacterial correlation network of remission patients with correlations greater than 0.45 and p value < 0.05. Red lines indicate negative correlations while blue lines represent positive correlations

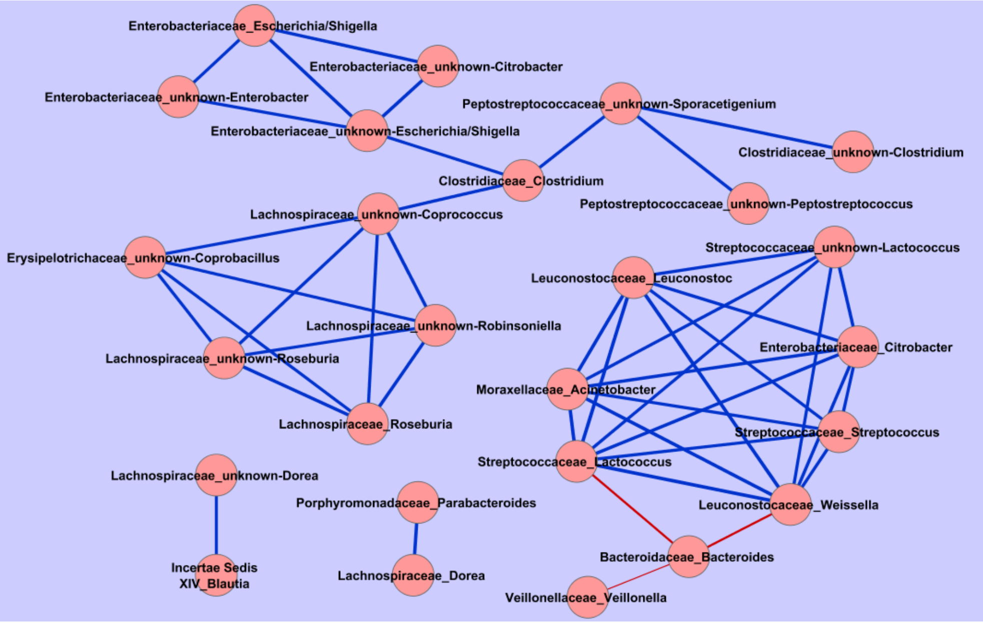

Enterobacteriaceae_Escherichia/Shigella
Enterobacteriaceae_unknown-Citrobacter
Enterobacteriaceae_unknown-Enterobacter
Enterobacteriaceae_unknown-Escherichia/Shigella
Peptostreptococcaceae_unknown-Sporacetigenium
Clostridiaceae_unknown-Clostridium
Clostridiaceae_Clostridium
Peptostreptococcaceae_unknown-Peptostreptococcus
Lachnospiraceae_unknown-Coprococcus
Erysipelotrichaceae_unknown-Coprobacillus
Streptococcaceae_unknown-Lactococcus
Leuconostocaceae_Leuconostoc
Lachnospiraceae_unknown-Robinsoniella
Lachnospiraceae_unknown-Roseburia
Enterobacteriaceae_Citrobacter
Moraxellaceae_Acinetobacter
Lachnospiraceae_Roseburia
Streptococcaceae_Streptococcus
Streptococcaceae_Lactococcus
Lachnospiraceae_unknown-Dorea
Porphyromonadaceae_Parabacteroides
Leuconostocaceae_Weissella
Bacteroidaceae_Bacteroides
Incertae Sedis
XIV_Blautia
Lachnospiraceae_Dorea
Veillonellaceae_Veillonella

Figure 4.6.4 Bacterial correlation network of relapse patients with correlations greater than 0.45 and p value < 0.05. Red lines indicate negative correlations while blue lines represent positive correlations

The correlations between the urinary metabolites and bacteria consists of 57 correlations between 46 nodes for the remission patients (Figure 4.6.5) and 71 correlations between 66 nodes for the relapse patients (Figure 4.6.6)

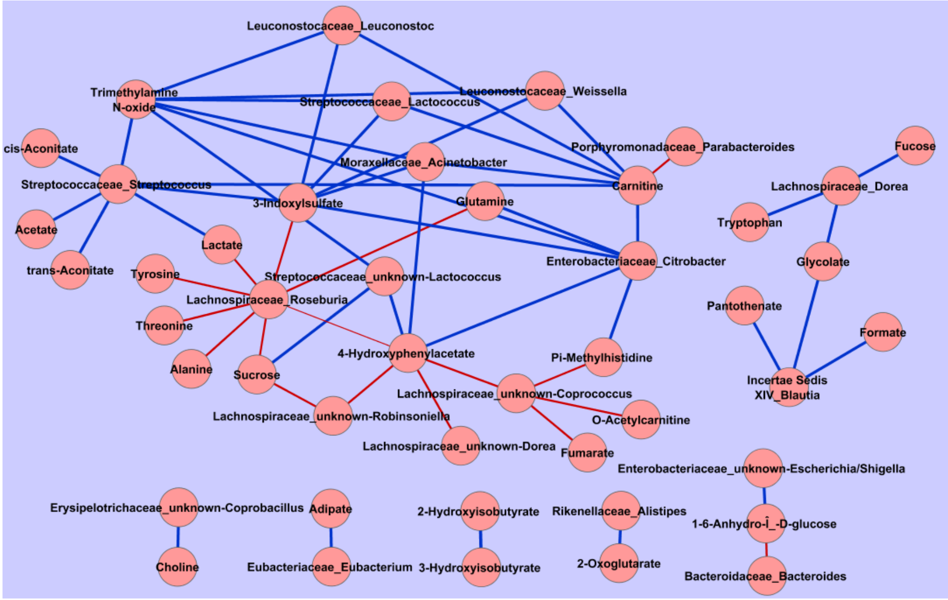

Leuconostocaceae_Leuconostoc
Trimethylamine N-oxide
Streptococcaceae_Lactococcus
Leuconostocaceae_Weissella
cis-Aconitate
Porphyromonadaceae_Parabacteroides
Fucose
Moraxellaceae_Acinetobacter
Streptococcaceae_Streptococcus
Carnitine
Lachnospiraceae_Dorea
3-Indoxylsulfate
Glutamine
Acetate
Tryptophan
Lactate
Enterobacteriaceae_Citrobacter
Glycolate
trans-Aconitate
Tyrosine
Streptococcaceae_unknown-Lactococcus
Lachnospiraceae_Roseburia
Pantothenate
Threonine
Formate
4-Hydroxyphenylacetate
Pi-Methylhistidine
Alanine
Sucrose
Incertae Sedis XIV_Blautia
Lachnospiraceae_unknown-Coprococcus
Lachnospiraceae_unknown-Robinsoniella
O-Acetylcarnitine
Lachnospiraceae_unknown-Dorea
Fumarate
Enterobacteriaceae_unknown-Escherichia/Shigella
Erysipelotrichaceae_unknown-Coprobacillus
Adipate
2-Hydroxyisobutyrate
Rikenellaceae_Alistipes
1-6-Anhydro-Î_-D-glucose
Choline
Eubacteriaceae_Eubacterium
3-Hydroxyisobutyrate
2-Oxoglutarate
Bacteroidaceae_Bacteroides

Figure 4.6.5 Urinary metabolomics and bacterial correlation network of remission patients with correlations greater than 0.45 and p value < 0.05. Red lines indicate negative correlations while blue lines represent positive correlations

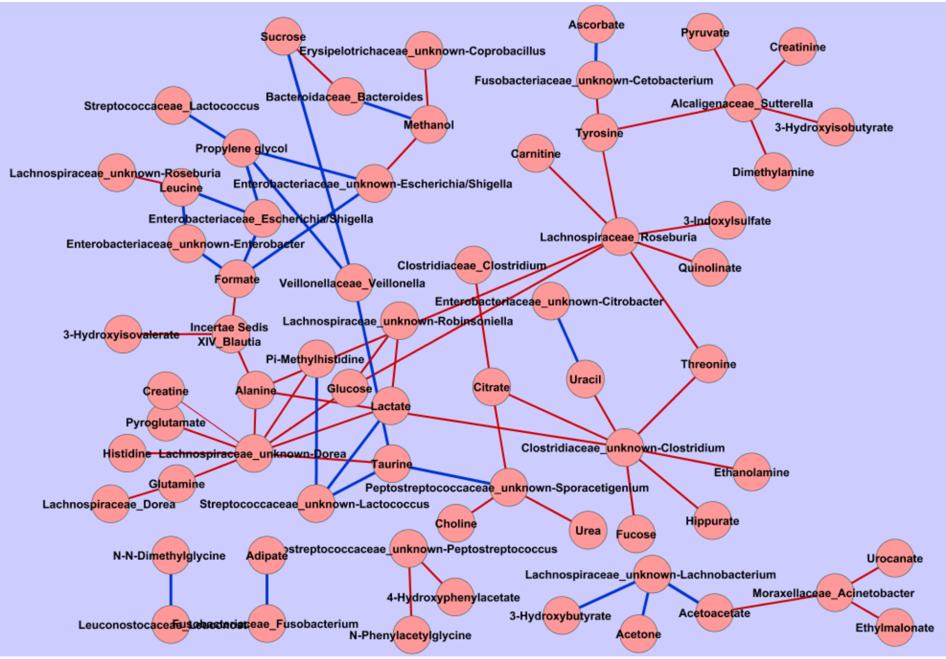

Ascorbate
Pyruvate
Sucrose
Creatinine
Erysipelotrichaceae_unknown-Coprobacillus
Fusobacteriaceae_unknown-Cetobacterium
Bacteroidaceae_Bacteroides
Alcaligenaceae_Sutterella
Streptococcaceae_Lactococcus
Methanol
3-Hydroxyisobutyrate
Tyrosine
Propylene glycol
Carnitine
Dimethylamine
Lachnospiraceae_unknown-Roseburia
Leucine
Enterobacteriaceae_unknown-Escherichia/Shigella
Enterobacteriaceae_Escherichia/Shigella
3-Indoxylsulfate
Lachnospiraceae_Roseburia
Enterobacteriaceae_unknown-Enterobacter
Quinolinate
Clostridiaceae_Clostridium
Formate
Veillonellaceae_Veillonella
Enterobacteriaceae_unknown-Citrobacter
3-Hydroxyisovalerate
Incertae Sedis XIV_Blautia
Lachnospiraceae_unknown-Robinsoniella
Pi-Methylhistidine
Threonine
Uracil
Creatine
Alanine
Glucose
Citrate
Lactate
Pyroglutamate
Clostridiaceae_unknown-Clostridium
Histidine
Lachnospiraceae_unknown-Dorea
Taurine
Ethanolamine
Glutamine
Peptostreptococcaceae_unknown-Sporacetigenium
Lachnospiraceae_Dorea
Streptococcaceae_unknown-Lactococcus
Choline
Urea
Fucose
Hippurate
N-N-Dimethylglycine
Adipate
ostreptococcaceae_unknown-Peptostreptococcus
Urocanate
Lachnospiraceae_unknown-Lachnobacterium
Moraxellaceae_Acinetobacter
4-Hydroxyphenylacetate
3-Hydroxybutyrate
Acetoacetate
Leuconostocaceae_Leuconost
Fusobacteriaceae_Fusobacterium
N-Phenylacetylglycine
Acetone
Ethylmalonate

Figure 4.6.6 Urinary metabolomics and bacterial correlation network of relapse patients with correlations greater than 0.45 and p value < 0.05. Red lines indicate negative correlations while blue lines represent positive correlations

Chapter 5: Comparison of samples from patients with no inflammation with those with any inflammation

5.1 Patient Population

The next comparison that was made was between patients in complete endoscopic remission with no inflammation (n=7; 4 females, 3 males) compared with those with any sort of inflammation at all (n=31; 21 females, 10 males). A patient was considered having no inflammation if their Rutgeerts score was 0 and any inflammation if their score was 1, 2, 3 or 4. This comparison was made with the rationale that any sort of inflammation may be caused by a unique bacterial fingerprint or metabolomic signature. The aim was to determine if the patients in complete remission with no inflammation were unique.

5.2 Bacterial Composition

As mentioned before, the microbial environment of the gut has been shown to be a significant factor in CD pathogenesis and recurrence. Therefore the bacterial composition was compared between patients with no inflammation to patients with any inflammation at all.

The patients with absolutely no inflammation had a significantly increased abundance of *Bacteroidia* (Figure 5.2.1) compared to the patients with any inflammation. Although this difference was also seen in the comparison between remission (Rutgeerts score 0, 1 and 2) versus relapse (Rutgeerts score 3 and 4) this difference between patients with no inflammation and patients with any inflammation was greater suggesting that some member of the *Bacteroidia* may have a protective function, or alternatively, that *Bacteroidia* are more sensitive to increased inflammatory mediators in the lumen.

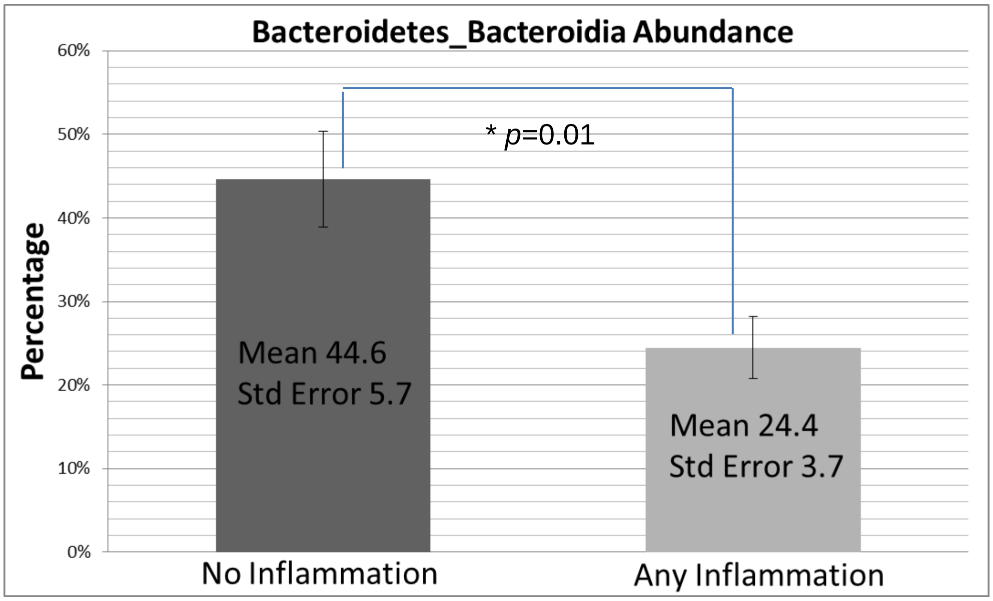


Figure 5.2.1 A) Average abundance of *Bacteroidia* in patients with no inflammation (left) and patients with any inflammation (right)

However, again despite the differences in *Bacteroidia* abundance, there was no separation based on the complete bacterial composition between the patients with no inflammation and the patients with any inflammation at the genus level. The PCA score plot (Figure 5.2.1) did not show any separation between the two sets of biopsies. Although the PLS-DA score plot (Figure 5.2.2) shows a separation this is caused by overfitting since cross validation of the model reveals R^2 and Q^2 values of 0.68 and -0.21 respectively suggesting the model is fitting a random correlation in the data. (The values are not close to 1 and are not similar to each other.)

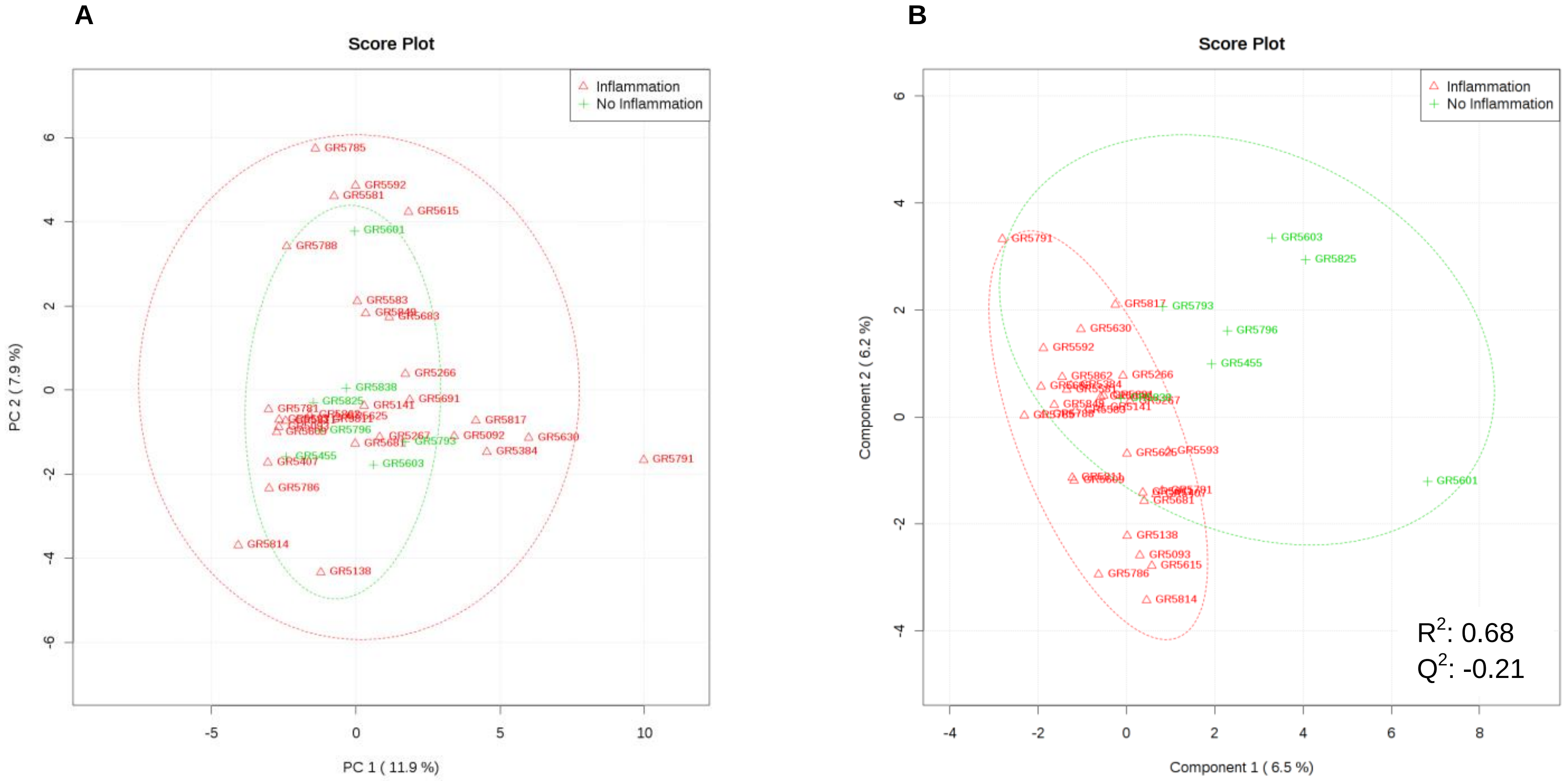


Figure 5.2.2 A) PCA score plot and **B)** PLS-DA score plot of the bacterial composition of patients with any inflammation (red triangles) and patients with no inflammation (green crosses) consisting of 70 different bacteria at the genus level.

In addition none of the machine learning algorithms were able to distinguish the two groups from each other (baseline was 30/37= 81 %). In addition feature selection was not helpful.

Table 5.2.1 Machine learning performance on biopsies with any inflammation and no inflammation

Algorithm	Accuracy	ROC Area	Log Transformation	
			Accuracy	ROC Area
J48 Decision Tree	75.7%	0.549	75.7%	0.307
Naïve Bayes Classifier	75.7%	0.657	70.3	0.488
Support Vector Classifier	70.3%	0.437	70.3%	0.433

As a whole there was not a large nor consistent enough difference between the patients with no inflammation and the patients with any inflammation to be able to build a model to distinguish the two populations even though there was a significant difference in the *Bacteroidia* between the two populations.

5.3 Metabolomics

The urinary metabolomics signature was compared next in these patients. As mentioned before, metabolites are a direct measurement of the function occurring in the biological system and a large number of urinary metabolites are produced by the gut bacteria. The bacteria composition itself was insufficient to distinguish the patients with no inflammation from patients with any inflammation but perhaps the functional aspect of these bacteria would be able to.

There were 5 urinary metabolites that were significantly different between the patients with no inflammation and patients with any inflammation. (Table 5.3.1) Methanol, propylene glycol, xylose, and acetate concentrations were increased in the no inflammation group while ATP and glycolate were increased in patients with any inflammation. Acetate is known to increase the epithelial barrier of the intestine [135] while methanol extracts has shown to have some anti-inflammatory effects in mice [136] so the results here agree with previous findings. The increased acetate in the patients with no inflammation may be coming from acetate producing bacteria such as the *Bacteroidia[137]* which makes sense since the patients with no inflammation have an increase abundance of this bacteria as previously shown. Another source of the acetate could be from fatty acid metabolism since acetate is one of the major end products.[138]

On the other hand xylose has been shown to be increased in CD patients compared to controls [139]. However, this study compared CD patients in remission compared with relapse.

Table 5.3.1 Significantly different metabolites

Metabolite (mM)	No Inflammation ±SEM (n=7)	Any Inflammation ±SEM (n=31)	P value
Methanol	0.047 ± 0.036	0.010 ± 0.001	0.02
ATP	0.010 ± 0.004	0.021 ± 0.002	0.02
Propylene glycol	0.706 ± 0.581	0.119 ± 0.047	0.02
Xylose	0.220 ± 0.131	0.097 ± 0.013	0.03
Acetate	0.331 ± 0.273	0.099 ± 0.014	0.04
Glycolate	0.332 ± 0.059	0.560 ± 0.062	0.048

A total of 6 metabolites were significantly different when comparing patients with no inflammation to patients with any inflammation whereas the previous comparison between relapse and remission patients did not result in any significantly different metabolites. Table 5.3.2 summarizes the metabolites. Methanol, propylene glycol, and glycolate were also metabolites of interest when comparing relapse patents to those in remission but were not significant in that case, whereas ATP xylose and acetate are new metabolites of possible interest.

Table 5.3.2 Description of significantly different metabolites

Metabolite	Description
Methanol	-simplest alcohol. Toxicity is due to the metabolic products of alcohol dehydrogenase. -produced by many bacteria as a product of various metabolic processes such as ammonia oxidation [140] -associated with alcoholism [131].
ATP	-extracellular signalling molecule -contributes to cellular energy charge and participates in overall energy balance, maintaining cellular homeostasis.
Propylene glycol	-used as a solvent for intravenous, oral, and topical pharmaceutical preparation. -associated in lung cancer [121].
Xylose	-a monosaccharide containing five carbon atoms -essential sugar for human nutrition -intestinal absorption of D-xylose has been shown to be useful in evaluating small intestinal function[141] -associated in lung cancer [121].
Acetate	-simple carboxylic acid. -central to the metabolism of carbohydrates and fats -shown to increase intestinal barrier function [142]. -produced by many gram negative, aerobic bacteria that derive their energy from the oxidation of ethanol into acetate such as A*acetobacteriium[143]* -associated in Propionic academia and phenylketonuria [144].
Glycolate	-smallest alpha-hydroxy acid, glycolic acid finds applications in skin care products. -associated in lung cancer and biliary atresia [121, 130].

Information from the Human Metabolome Database

The PCA score plot (Figure 5.3.1A) did not show any separation between the two groups. Although the PLS-DA score plot (Figure 5.3.1B) shows a separation this is caused by overfitting since cross validation of the PLS-DA model reveals R^2 and Q^2 values of 0.75 and -0.46 respectively suggesting the model is fitting a random correlation in the data.

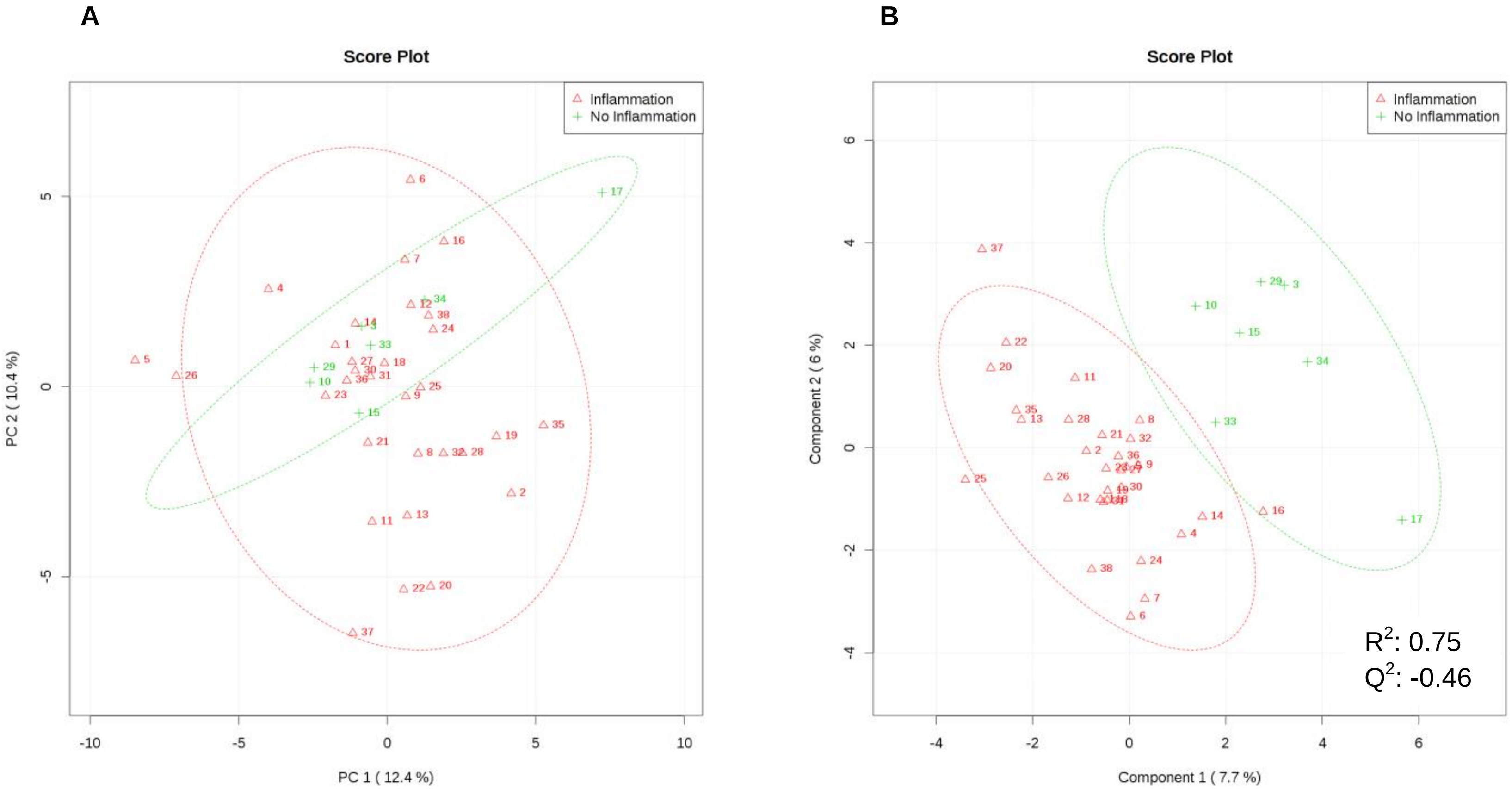


Figure 5.3.1 A) PCA Score plot and **B)** PLS-DA score Plot of the urinary metabolomic composition of patients with any inflammation (red triangles) and patients with no inflammation (green crosses) consisting of 67 different metabolites.

Lastly none of the machine learning algorithms (Table 5.3.3) were able to create a model to separate the patients with inflammation from the patients with no inflammation at all.

Table 5.3.3 Machine learning performance on urinary metabolites from patients with no inflammation and patients with any inflammation

Algorithm	Accuracy	ROC Area	Log transformation		With Feature Selection	
			Accuracy	ROC Area	Accuracy	ROC Area
J48 Decision Tree	76.3%	0.624	68.4%	0.558	78.9%	0.333
Naïve Bayes Classifier	73.7%	0.634	71.1%	0.557	76.3%	0.434
Support Vector Classifier	81.6%	0.538	81.6%	0.555	81.6%	0.462

As a whole there were individual metabolites that were different between CD patients without any inflammation compared to patients with any inflammation some of which agree with previous literature while others do not. However these individual differences were not enough to create a reliable model to separate or distinguish the two groups of patients

5.4 Environmental Questionnaire

The next thing that was looked at between the patients with and without any inflammation was environmental factors using the CEGIIR questionnaire. The goal was to highlight whether differences previously found between relapse and remission patients are also found when comparing patients of the extreme phenotype of no inflammation compared to any inflammation.

The number of patients with a Rutgeerts grade of 0 (n=7) was much lower than the number of patients with some sort of inflammation (n=31). As a result a single patient in first group represents a much larger percent in its group when compared to a single patient in the inflammation group. Therefore the results may be skewed therefore the only environmental factors looked at were the ones that were different between relapse and remission patients.

Figure 5.4.1 shows the environmental factors between the no inflammation and any inflammation groups that were looked at previously (between the relapse and remission patients.) In general all of the trends stayed the same although bigger differences appeared to be apparent in the current and childhood exposures. It is clear that patients with no inflammation are not smoking marijuana and not trying herbal medications. In addition there was a much greater difference in childhood exposures as patients with no inflammation were drinking non tap water, eating fish and having animal exposure compared to patients with any inflammation. However none of the comparisons were significantly different using the Fishers exact test. (See Appendix 3 for p-values.)

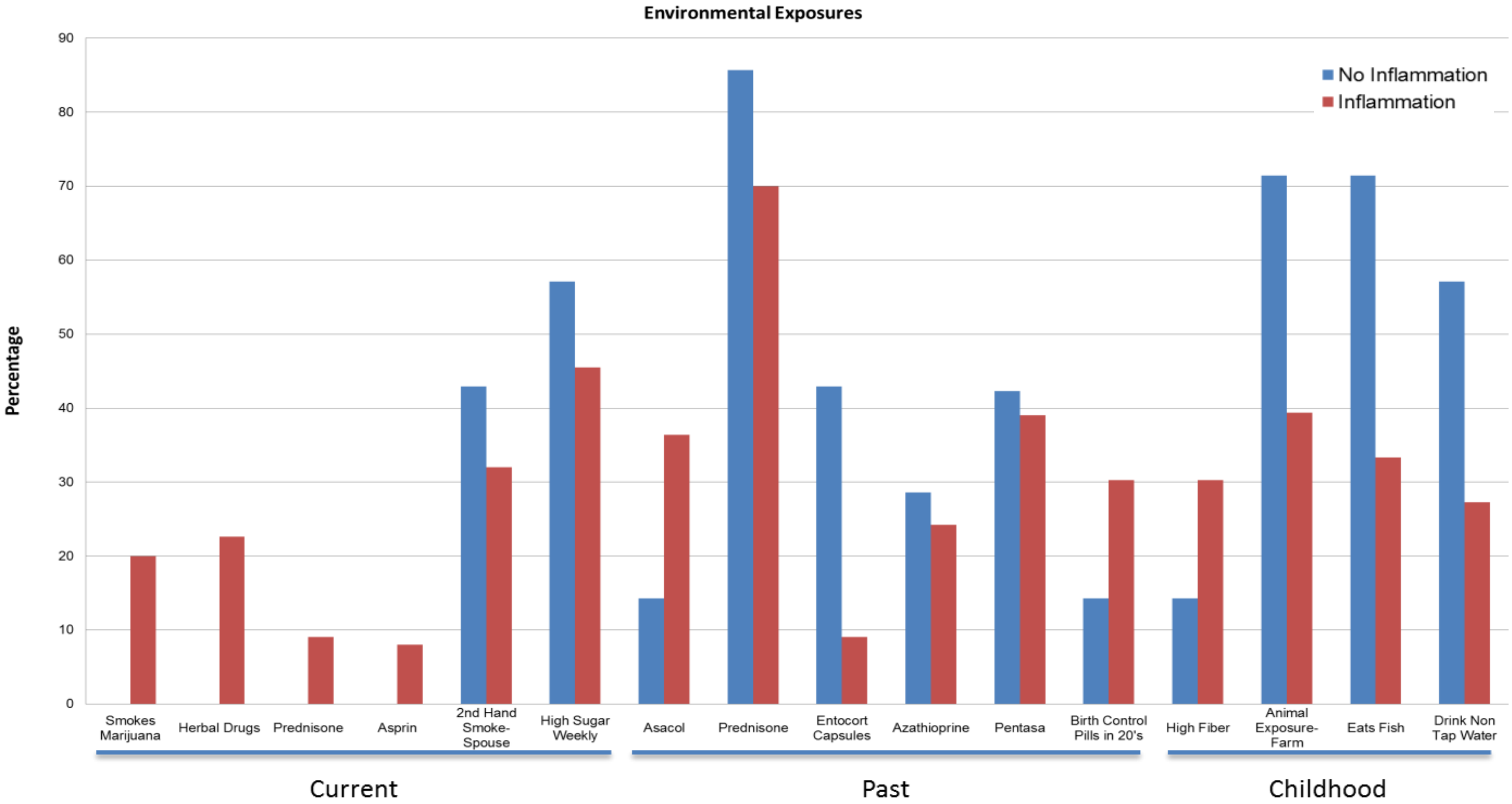


Figure 5.4.1 Percentage of patients with any inflammation (red) and patients with no inflammation (blue) that are currently (left) in the past (middle) or in their childhood (right) been exposed to certain environmental factors.

Again, despite the individual differences, a model could not be created to differentiate the patients with no inflammation from the patients with any inflammation using the questionnaire data. (Table 5.4.1)

Table 5.4.1 Machine learning performance on questionnaire data from patients with no inflammation and patients with any inflammation

Algorithm	Accuracy	ROC Area	Feature Selection	
			Accuracy	ROC Area
J48 Decision Tree	70.0%	0.091	77.5%	0.000
Naïve Bayes Classifier	82.5%	0.369	77.5%	0.208
Support Vector Classifier	85.0%	0.571	77.5%	0.470

Overall the environmental differences between remission and relapse patients were very similar to the comparison between patients with and without any inflammation. However the latter highlights a greater difference in childhood exposures especially drinking non tap water, eating fish and animal exposures.

Chapter 6: Comparison of males with females

6.1 Patient Population

A secondary objective was to compare the bacteria and urinary metabolomic signature between males (n=12) and females (n=25).

Table 6.1.1 Patient information

	Males	**Females**
Number of Patients	12 (32%)	25 (68%)
Mean Age ± SD years **(Range)**	52.7 ± 11.6 (23-68)	44.9 ± 11.7 (21-69)
Disease Status • **Endoscopic Remission** • **Endoscopic Relapse**	 6 (50%) 6 (50%)	 13 (52%) 12 (48%)
Years since last surgery **Mean ± SD (Range)**	10.1 ± 7.7 (3-29)	11.3 ± 9.7 (3-39)
Current Medications		
• 5 ASA	3 (25%)	5 (20%)
• Steroids	1 (8%)	4 (16%)
• Immuno-Modulators	2 (17%)	12 (48%)
• Biologics	2 (17%)	6 (24%)
• Antibiotics	1 (8%)	2 (8%)

6.2 Bacterial Composition

When averaging each sex regardless of disease state, females had a significant increase in abundance of *Gammaproteobacteia* (p=0.03) when compared to males at the class level.

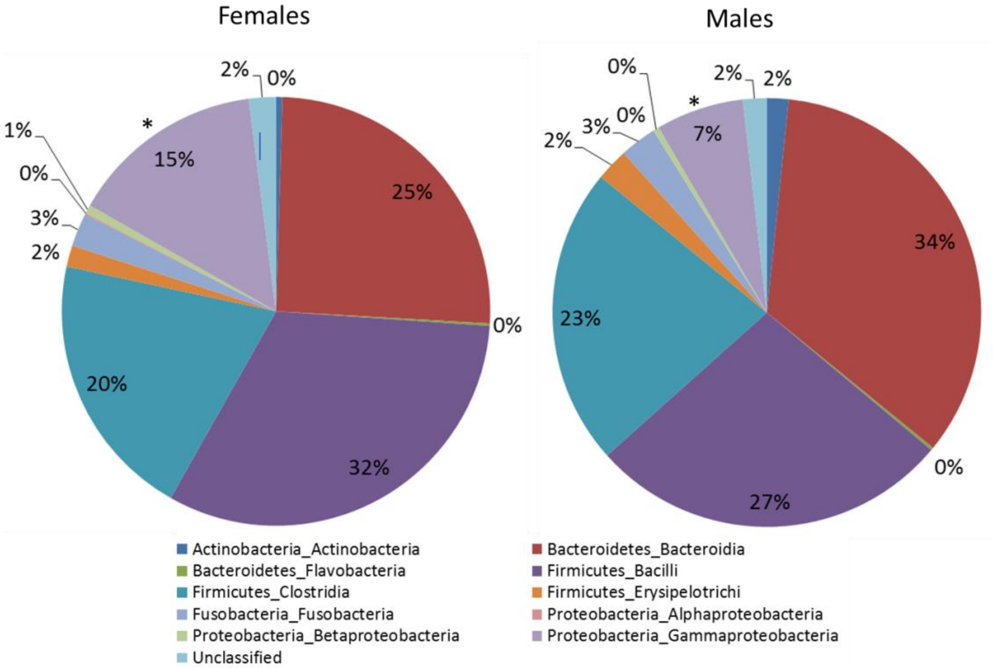


Figure 6.2.1 Averaged bacterial composition at the class level of Females (left) and males (right) Asterisk indicated a significant difference (p value <0.05) using the students t-test.

Despite the individual difference seen in abundance of *Gammaproteobacteia*, there was no separation based on the complete bacterial composition between males and females. The PCA score plot (Figure 6.2.6) when labeled with the patients sex did not show any natural clusters. Although the PLS-DA score plot (Figure 5.2.2) shows a separation this is caused by overfitting since cross validation of the PLS-DA model reveals R^2 and Q^2 values of 0.68 and -0.19 respectively suggesting the model is fitting a random correlation in the data. (The values are not close to 1 and are not similar to each other.)

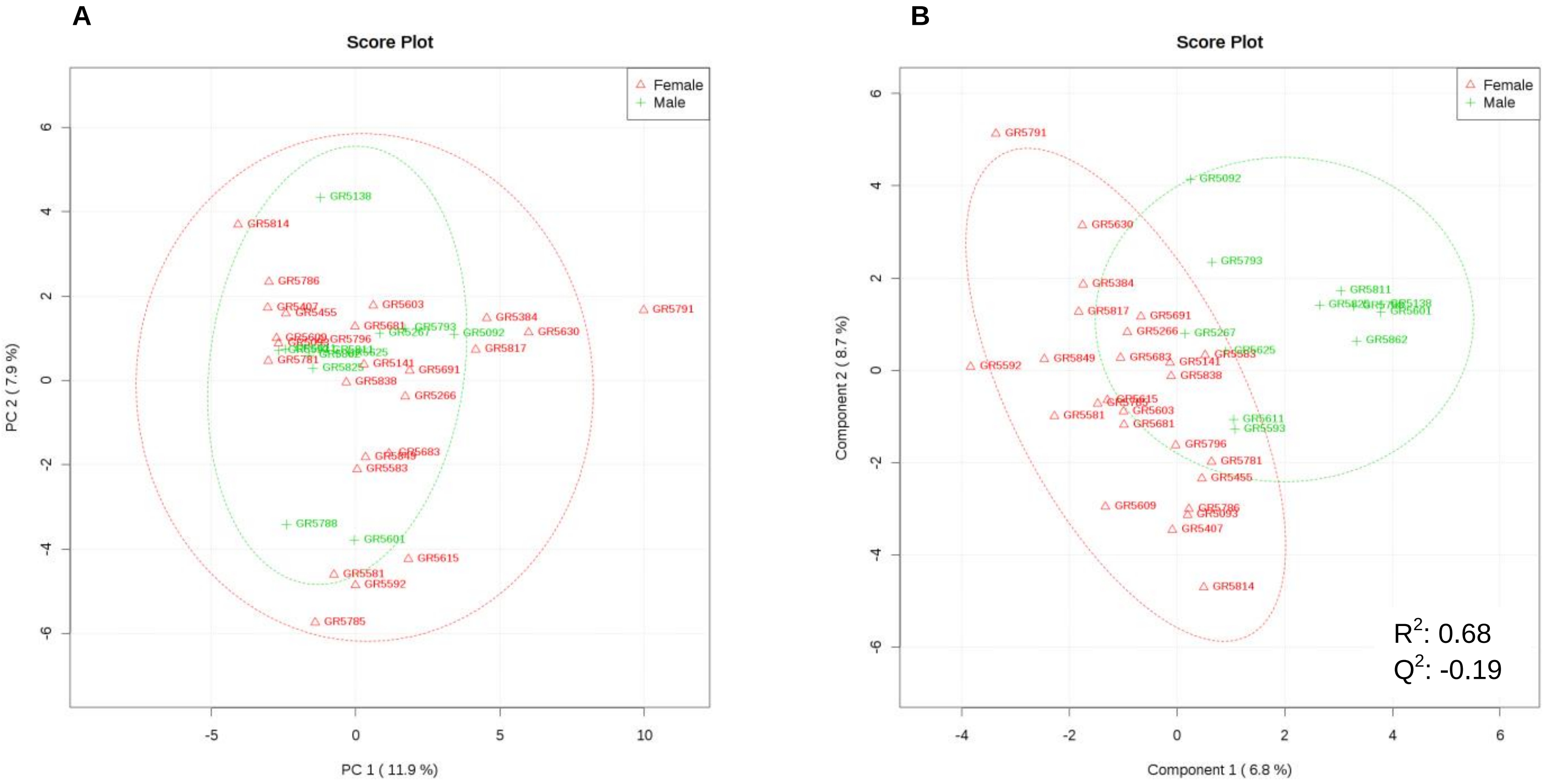


Figure 6.2.2 A) PCA Score plot and **B)** PLS-DA score plot of the bacterial composition of females (red triangles) and males (green crosses) patients consisting of 70 different bacteria at the genus level.

In addition none of the machine learning algorithms were able to distinguish the two groups from each other (baseline was 25/37= 68 %). In addition feature selection did not help at all.

Table 6.2.1 Machine learning performance on biopsies from male and female patients

Algorithm	Accuracy	ROC Area	Log Transformation	
			Accuracy	Accuracy
J48 Decision Tree	67.6%	0.475	67.6%	0.467
Naïve Bayes Classifier	62.2%	0.505	64.9%	0.583
Support Vector Classifier	64.9%	0.523	64.9%	0.480

In this patient population there was significant increase in the *Gammaproteobacteria* in the females compared to the males but the overall composition was not different enough to distinguish one sex from another based on the bacterial composition.

6.3 Metabolomics

The urinary metabolomics signature was compared next between the males and female patients. There was a significant difference in the metabolomic signature between the males and females in this study. Out of the 67 metabolites measured, 21 metabolites were significantly different between the two groups. This result was expected as males and females have very different metabolic processes. The results can be seen in Table 6.3.1.

Table 6.3.1 Significantly different metabolites

Metabolite	Female (n=25) conc ± SEM (mM)	Male(n=12) conc ± SEM (mM)	P-value
Pyroglutamate	0.48 ± 0.05	0.23 ± 0.04	0.003
Glutamine	0.64 ± 0.07	0.32 ± 0.04	0.005
Ethanolamine	0.72 ± 0.07	0.36 ± 0.07	0.01
trans-Aconitate	0.07 ± 0.01	0.03 ± 0.01	0.01
Threonine	0.19 ± 0.02	0.09 ± 0.01	0.01
Dimethylamine	0.73 ± 0.09	0.34 ± 0.06	0.01
Succinate	0.12 ± 0.02	0.03 ± 0.01	0.01
Glycine	2.71 ± 0.45	0.90 ± 0.15	0.01
Lactate	0.30 ± 0.05	0.10 ± 0.02	0.01
Urocanate	0.04 ± 0.01	0.01 ± 0.003	0.02
Glucose	0.57 ± 0.07	0.29 ± 0.05	0.02
cis-Aconitate	0.60 ± 0.10	0.21 ± 0.07	0.02
Adenosine	0.01 ± 0.002	0.00 ± 0.001	0.02
3-Hydroxyisobutyrate	0.13 ± 0.02	0.07 ± 0.01	0.02
Creatinine	22.48 ± 2.85	11.79 ± 2.22	0.02
Valine	0.06 ± 0.01	0.03 ± 0.01	0.03
Uracil	0.09 ± 0.01	0.05 ± 0.01	0.03
Alanine	0.34 ± 0.05	0.19 ± 0.03	0.04
Formate	0.19 ± 0.03	0.10 ± 0.01	0.04
Quinolinate	0.07 ± 0.01	0.03 ± 0.01	0.04
Leucine	0.03 ± 0.01	0.02 ± 0.003	0.05

The PCA score plot (Figure 6.3.1) of the metabolites did not show any separation between the males and females; however it appears that the males are in a much tighter group whereas the females are more widespread and variable. This could possibly be explained by the different stages of their menstrual cycle.

However the PLS-DA score plot (Figure 6.3.2) does show a clear separation between the sexes; however unlike previous comparison this model did cross validate with R^2 and Q^2 values of 0.91 and 0.32 respectively. This suggests that the model is fitting the data adequately.

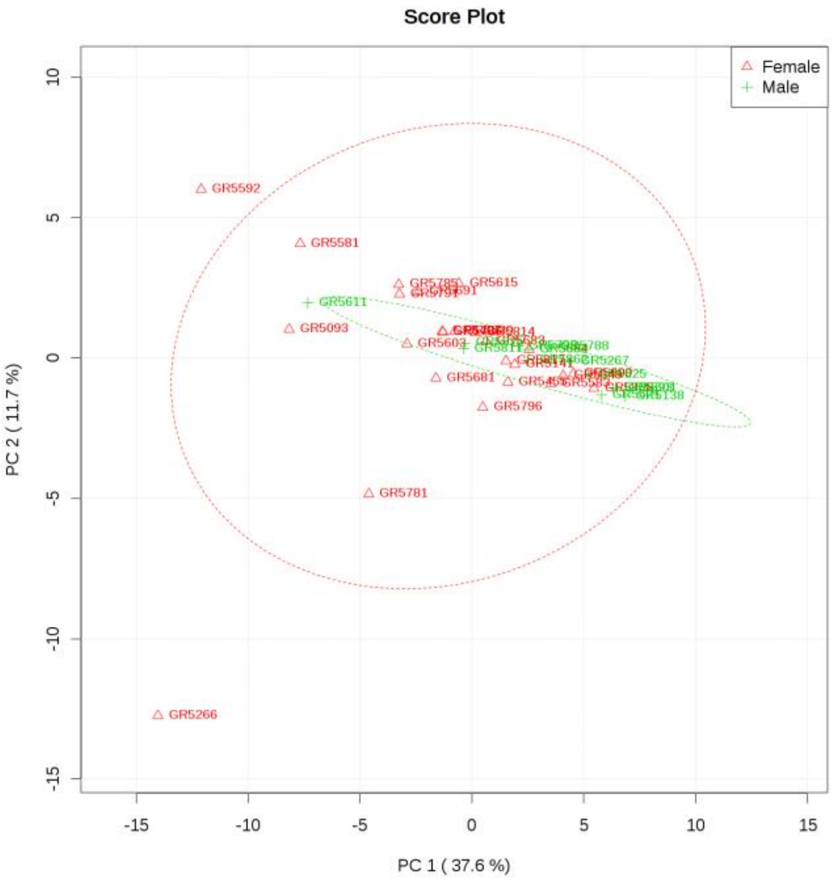


Figure 6.3.1 A) PCA Score plot of the urinary metabolomic composition of females (red triangles) and males (green crosses) consisting of 67 different metabolites.

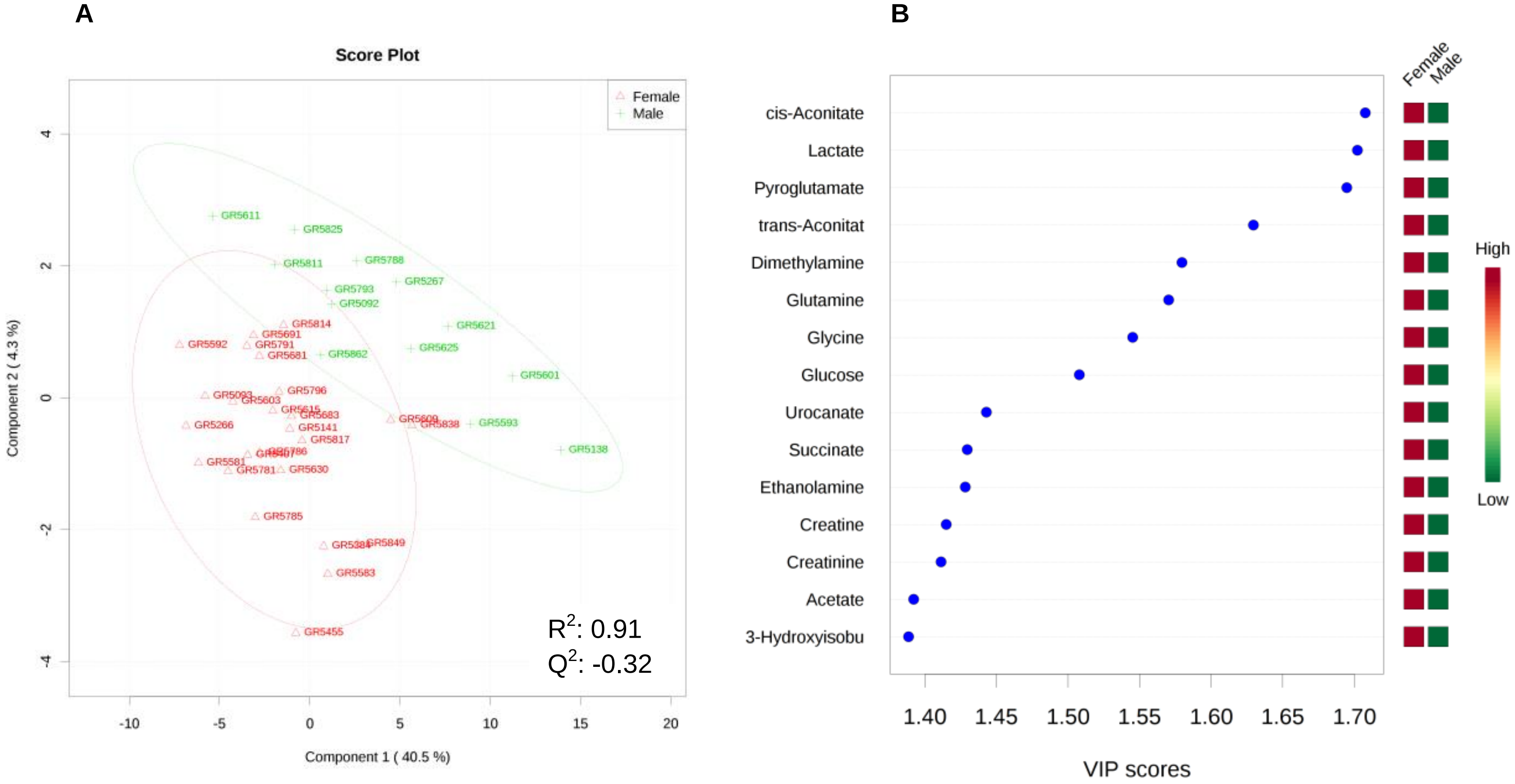


Figure 6.3.2 A) PLS-DA Score plot and **B)** Variable importance in projection (VIP) of the urinary metabolomic composition of females (red triangles) and males (green crosses) consisting of 67 different metabolites.

To confirm the results from PLS-DA model, machine learning algorithms (Table 5.4.0) were also run. They were able to perform much better in trying to create a model to separate the male patients from the female patients compared to the rest of the comparisons made. The best model I was able to make was using the best first feature selection algorithm along with the support vector classifier to come up with an accuracy of 81.6% and ROC area of 0.749.

Table 6.3.2 Machine learning performance on urinary metabolites from male and female patients

Algorithm	Accuracy	ROC Area	Log transformation		Feature Selection	
			Accuracy	ROC Area	Accuracy	ROC Area
J48 Decision Tree	65.8%	0.588	73.7%	0.626	71.1%	0.691
Naïve Bayes Classifier	71.1%	0.741	78.9%	0.729	71.1%	0.759
Support Vector Classifier	71.1%	0.688	76.3%	0.709	81.6%	0.749

In conclusion the urinary metabolomics signature was unique between men and women. There were a large number of individual metabolites that were different between the two sexes. In addition, these differences were unique enough to create a model to separate the two classes.

Chapter 7: Discussion

Post-operative recurrence of CD is thought to mimic the pathogenesis at onset of the disease and there is much evidence to suggest environmental exposures or gut microbes play a major role [11]. Therefore, the main aim of this project was to determine which specific environmental exposures or gut microbes may be associated with post-operative recurrence by comparing CD patients that had undergone ileocolonic resection that are in remission versus those in relapse. I hypothesized that Crohn's disease patients in remission would have a different and unique microbial profile compared with Crohn's disease patients suffering from relapse, and this would be mirrored by a distinct urinary metabolomic profile

To test this hypothesis, a retrospective study was carried out using samples obtained from the CEGIIR biobank to examine differences in bacterial composition, urinary metabolomic signature and environmental factors in CD patients that had previously undergone an ileocolonic resection. Comparisons were first made between patients in endoscopic remission (Rutgeerts score 0, 1, or 2) to patients in endoscopic relapse (Rutgeerts score 3 or 4); follow up comparisons were then made between patients with no inflammation (Rutgeerts score of 0) to patients with any inflammation (Rutgeerts score 1, 2, 3, or 4) with the rational that intestinal inflammation of any sort may be associated by a unique bacterial fingerprint or metabolomics signature.

The bacterial differences between remission and relapse patients was expected based upon previous reports in the literature [145]. The significant decrease in *Bacteroidetes* found in this project was supported by earlier studies that found a decrease in Bacteroidetes when comparing CD patients to controls [146]. However more recent studies have implicated an increase in the abundance of *Bacteroidetes* in CD patients [147]. In addition, another study that directly looked at changes in bacteria flora in the neo-terminal ileum of CD patients after ileocolonic resection found an increase in *Bacteroidetes* in CD patients with endoscopic recurrence at 3 months and 1 year after surgery compared to patients in endoscopic remission [148].

My finding of an increase in abundance of *Erysipelotrichi* has not been previously reported regarding CD but was reported to be increased in irritable bowel syndrome (IBS) patients [149]. Lastly, the increase in *Gammaproteobacteria* found in relapse patients has been reported in previous studies in CD patients compared to controls [145].

The bacterial composition between patients with no inflammation to patients with any sort of inflammation highlighted the decrease in *Bacteroidetes* in patients with any inflammation. As mentioned earlier the *Bacteroidetes* have previously been implicated as decreased in abundance in CD patients compared to controls however more recent studies contradict this finding and report an increase in abundance [146, 148].

The phylum *Bacteroidetes* is composed of many gram-negative, rod shaped bacteria that are commonly found in the environment. They are one of the most abundant species in the human intestine and perform the function of degrading a wide variety of polysaccharides into useful monosaccharides for the body to utilize [150]. A decrease in *Bacteroidetes* , more specifically the *Bacterodales* order, in the relapse patients may be significant since this group of bacteria hydrolyze the available fucose to provide the host a supply of nutrients and promote a gastrointestinal health [146, 150].

The *Erysipelotrichi* are part of the Firmicute phylum and include a well-known human pathogen, *Erysipelothrix rhusiopathiae* [151]. An increase of these pathogenic bacteria in relapse patients may suggest a role in CD recurrence after surgery. In addition *Gammaproteobacteria* is also comprised of many pathogens such as *Salmonella, Yersinia* and *Vibro* [152]. An increase in abundance of these *Gammaprotebacteria* in the relapse patients may also suggest they might potentially play a role in post-operative recurrence. Overall the bacterial results suggest that some of the bacteria that may play a role in the initial onset of CD may also play a vital role in the relapse of CD after an ileocolonic resection.

Despite the difference in bacterial abundances it is unclear whether these differences are driving the onset of inflammation or the result of the inflammation. During inflammation, there is an increase in reactive oxygen species (ROS) in the lumen due to immune cell activity in response to pathogen invasion and metabolic stress and also increased oxygen due to a breakdown in gut barrier. This could create an environment that allows for the growth of specific bacteria that are less sensitive to oxidative environments [153-155]. Therefore there would be a decrease in some of the more sensitive bacteria and an increase in others that are more adaptable to living under those types of conditions. The retrospective experimental design used in these studies, is unable to distinguish between these two possibilities. This study looked at a single snapshot of the patient’s microbiota and in order to potentially tell the difference between these possibilities, a longitudinal study would have to be done with multiple snapshots taken at regular times before and after the onset of inflammation.

The urinary metabolomics signature is a direct measurement of the function occurring in the biological system. The gut microbiota is responsible for a large number of urinary metabolites and the measurement of these metabolites can be used to reflect the function of the gut bacteria. The bacteria composition itself was unable to distinguish the remission patients from the relapse patients but perhaps the functional aspect of these bacteria would be able to since two distinct bacteria could possibly perform the same function.

Overall the urinary metabolomic composition between the relapse and remission patients was very similar. Although there was a trend of several metabolites being different between the two classes, there was no significant difference in any of the metabolites. However, the most interesting metabolite that was increased in relapse patients (1.6 fold change; p value 0.06)) was fucose. This coincides with the decreased abundance of *Bacteroidetes*, as these bacteria actively degrade fucose into useful monosaccharides for the body so the increase in fucose makes sense with the decreased abundance of *Bacteroidetes*. [150] In addition, specific bacteria may rely on sensing chemical signals for colonization and regulation of genes in the gastrointestinal tract [156]. In particular, fucose sensing has been shown to regulate the intestinal colonization of the gastrointestinal pathogen enterohemorrhagic *E.coli* [157].

When the urinary metabolites between patients with no inflammation were compared to the patients with any inflammation, a larger difference was found and some significantly different urinary metabolites were revealed.

ATP was found to be increased in urine from patients with inflammation. ATP is a signaling molecule that can be derived from commensal bacteria and can induce T_H17 cell differentiation; importantly, an abnormal T_H17 response has been implicated in the pathogenesis of IBD [158]. Acetate, as well as other short chain fatty acids has been shown to increase intestinal barrier function [142]. Interestingly acetate was increased in patients with no inflammation at all which supports the fact acetate may be protective and increase intestinal barrier function.

However the differences in metabolite concentrations in the urine were not sufficient to build any machine learning model that could differentiate between patients in relapse and remission, likely due to the small number of patients in each group. However, the systems biology approach of building correlation networks between metabolites and bacteria could be used as a tool to build future hypotheses and connections that could be explored further. For example, the relationship between *Streptococcus* and acetate could be investigated further since the remission correlation network showed acetate was positively correlated with the abundance of *Streptococcus.* Both *Streptococcus* and acetate have been implicated in CD as the abundance of *Streptococcus* has been shown to be increased in CD patients and acetate has been shown to increase intestinal barrier function [142, 159].

When examining the responses to the environmental questionnaire, one of the most interesting findings was that patients in relapse were smoking more marijuana than patients in remission. A possible explanation for this result was that patients in relapse may be self-medicating with marijuana. This practice of self-medication may also explain why patients in relapse were trying more herbal drugs and were on more aspirin. There is some evidence that smoking cannabis is effective in the induction of remission in patients with Crohn’s disease [160], suggesting that the patients may be obtaining some symptomatic relief through the use of marijuana. Another surprising result was that patients in remission were exposed to more second hand smoke than patients in relapse, as cigarette smoking is a well-known risk factor for recurrence of CD [161].

Lastly, early childhood exposures were another interesting result that came out of the study. There is much evidence to suggest that early childhood exposures influence the risk of CD through the exposure to microorganisms that may influence host immune function. One study suggests that the period between birth and the age of 5 as being particularly important [162]. Indeed, drinking non tap water, being exposed to animals on a farm, and eating fish as a child were all associated with remission patients. These environmental factors could possibly expose these patients to a diverse community of microorganisms which appear to be related to maintaining remission after ileocolonic resection. A previous study supported my finding of drinking non tap water as the study showed CD patients were less likely to drink non tap water than controls in their childhood [163]. The environmental factors initially looked at between the relapse patients and remission patients were very similar to the results seen here between patients with no inflammation and any inflammation. The same trends existed. An interesting result observed was the amplified difference between the two groups regarding the childhood exposures. The difference in the percentage of patients with no inflammation compared to patients with any inflammation was much larger when it came to drinking non tap water as a child, eating fish as a child and animal exposure on a farm as a child.

The main aim of this study was to determine what the bacterial, metabolomic and environmental differences were between CD patients in relapse and remission that have previously undergone an ileocolonic resection. Although I was able to show some differences in these parameters between the groups, the differences were not consistent enough (as seen with the large standard errors) and the groups not large enough to build any sort of machine learning model to differentiate the two groups.

A secondary objective of this project was to compare the bacteria and metabolism between the males and female patients. There has been a lot of interest on the characterization of the human gut such as the Human Microbiome Project and the American Gut study [164](www.americangut.org). As seen in this project and numerous others, the bacterial composition is unbelievably complex and variable within any population. Factors such as diet, antibiotics, work and home environment as well as a countless other factors influence bacterial composition. With that being said, in this project females had a significant increase in abundance of *Gammaproteobacteria* when compared to males at the class level

However, like the other comparisons made, the minor difference in individual taxa was insufficient to separate the males from the females using any machine learning algorithm.

The urinary metabolomics signature was a different story. There were a large number of metabolites that were significantly different between the males and females suggesting a difference in metabolism between the two sexes. This was an expected result as it is known that different sexes have different metabolisms and different processes that occur. PCA showed that the metabolomic signature of females were much more scattered/ variable compared to males which makes sense because of the natural cycles of females.

PLS-DA as well as all the machine learning algorithms were able to separate the sexes much better than any other comparison made based on the urinary metabolites.

Overall the bacterial composition was very similar between males and females and the slight differences that did exist were neither consistent nor large enough to clearly differentiate one group from another. This has also been reported by the Human Microbiome project and the American gut study [164]. (www.americangut.org) However, the difference in metabolism between males and females allowed us to be able to differentiate between the two based on metabolomics signature.

Chapter 8: Limitations

There were several limitations associated with this study including the fact that it was a retrospective study, the sample size was limited, and the samples came from a heterogeneous population of patients

With this project being a retrospective study, all of the disadvantages that come along with it are relevant such as selection bias and information bias. Selection bias is the error in choosing the individuals to take part in the study and the selection of patients in this study was a clear limitation. Patients in this study ranged from having their ileocolonic surgery 3 years ago all the way to 41 years ago. This introduced a lot of age dependent variability within the project such as the changing of metabolism over time. In addition a patient in endoscopic remission at the time of colonoscopy could have experienced numerous flares before while a patient classified as a relapse patient could have just had his/her first relapse. Also many patients have had multiple surgeries and have been on a variety of different post-surgical drugs. These factors were considered to be important from the beginning of the study but it was limited as to what biopsies were available. Information bias is bias arising from measurement error or a flaw in measuring variables. In this project possible information bias could have arose from human error from chart reviews, that were done by hand, or from the CEGIIR questionnaire where patients could have given the wrong information.

Another limitation could have been the biopsies themselves. Biopsies were taken from the macroscopically healthy areas of the intestine regardless of whether a patient had inflammation or not. Perhaps a biopsy taken from the actual inflamed area of the intestine would be more revealing of the bacteria that may be causing the inflammation. At minimum, it would show if there was a difference in bacterial composition between the inflamed and non-inflamed tissue.

Sample size is another big limitation of this project. All of the machine learning algorithms would probably perform much better given more data to train/ build a model on. With the limited number of samples and the hundreds of measurements/ features obtained for each sample it results in a very difficult task. Again this project was limited by the number of samples available. Throughout the duration of the project, additional patients were being recruited but it is not viable to process and run individual samples as they arrive. In addition the costs of sequencing and running samples on the NMR are limiting factors.

Finally there is a disconnect between endoscopic remission and clinical remission in evaluating CD patients. Some of these patients were in endoscopic relapse but showed no external symptoms, so perhaps a limitation is in classifying these patients into distinct groups. Classifying these patients as either relapse or remission may be naive when in fact CD is really a continuous spectrum with relapse and remission on the extreme ends and infinitely many states in between.

Chapter 9: Future Direction and Final Thoughts

From the results of this project I believe the most interesting results came out of the environmental factors such as smoking marijuana and various childhood exposures that were different between CD patients in relapse and remission. Since all the data has already been collected perhaps a future project could be done to extract all of the patients that fit these criteria and perform the same analysis on this much larger group of patients to see if the same trends occur.

The best way to go forward would be a large prospective study following CD patients from the day of surgery. This trial is currently ongoing in CEGIIR and results from this trial will likely be very interesting. This prospective study will eliminate many of the limitations from this study including the time from surgery until the time of biopsy. In addition the patients will have regular checkups and progress can be recorded at the same time for all the patients.

In conclusion I think this project did succeed in fulfilling its aims. Specific environmental exposure, gut microbes and urinary metabolites were highlighted as possibly associated with post-operative recurrence of CD. The results give insights and stress what a complex disease CD is.

Bibliography

1. Abraham, C. and J.H. Cho, *Inflammatory bowel disease.* N Engl J Med, 2009. **361**(21): p. 2066-78.

2. Cosnes, J., et al., *Epidemiology and natural history of inflammatory bowel diseases.* Gastroenterology, 2011. **140**(6): p. 1785-1794.

3. *CCFC Impact Report 2012. The Impact of Inflammatory Bowel Disease in Canada 2012: Final Report and Recommendations.*

4. Künsebeck, H.W., J. Körber, and H. Freyberger, *Quality of life in patients with inflammatory bowel disease.* Psychotherapy and Psychosomatics, 2010. **54**(2-3): p. 110-116.

5. Lunney, P.C. and R.W. Leong, *Review article: Ulcerative colitis, smoking and nicotine therapy.* Aliment Pharmacol Ther, 2012. **36**(11-12): p. 997-1008.

6. Goyette, P., et al., *Molecular pathogenesis of inflammatory bowel disease: genotypes, phenotypes and personalized medicine.* Ann Med, 2007. **39**(3): p. 177-99.

7. Bernell, O., A. Lapidus, and G. Hellers, *Risk factors for surgery and recurrence in 907 patients with primary ileocaecal Crohn's disease.* British Journal of Surgery, 2000. **87**(12): p. 1697-1701.

8. Baumgart, D.C. and W.J. Sandborn, *Crohn's disease.* The Lancet. **380**(9853): p. 1590-1605.

9. Buisson, A., et al., *Review article: the natural history of postoperative Crohn's disease recurrence.* Alimentary Pharmacology & Therapeutics, 2012. **35**(6): p. 625-633.

10. Becker, J.M., *Surgical therapy for ulcerative colitis and Crohn's disease.* Gastroenterol Clin North Am, 1999. **28**(2): p. 371-90, viii-ix.

11. Rutgeerts, P., *Strategies in the prevention of post-operative recurrence in Crohn's disease.* Best Pract Res Clin Gastroenterol, 2003. **17**(1): p. 63-73.

12. Marks, J.W. December 2, 2013]; Available from: http://www.medicinenet.com/crohns_disease_pictures_slideshow/article.htm.

13. Baumgart, D.C. and W.J. Sandborn, *Crohn's disease.* The Lancet, 2012.

14. Franke, A., et al., *Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci.* Nature genetics, 2010. **42**(12): p. 1118-1125.

15. Ogura, Y., et al., *A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease.* Nature, 2001. **411**(6837): p. 603-606.

16. Economou, M., et al., *Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis.* The American journal of gastroenterology, 2004. **99**(12): p. 2393-2404.

17. Lala, S., et al., *Crohn's disease and the NOD2 gene: a role for paneth cells.* Gastroenterology, 2003. **125**(1): p. 47-57.

18. van Heel, D.A., et al., *Detection of muramyl dipeptide-sensing pathway defects in patients with Crohn's disease.* Inflamm Bowel Dis, 2006. **12**(7): p. 598-605.

19. Barrett, J.C., et al., *Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease.* Nature genetics, 2008. **40**(8): p. 955-962.

20. McGovern, D., et al., *NOD2 (CARD15), the first susceptibility gene for Crohn's disease.* Gut, 2001. **49**(6): p. 752-754.

21. van Heel, D.A., et al., *Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs.* Hum Mol Genet, 2004. **13**(7): p. 763-70.

22. Hume, G.E., et al., *Novel NOD2 haplotype strengthens the association between TLR4 Asp299gly and Crohn's disease in an Australian population.* Inflammatory bowel diseases, 2008. **14**(5): p. 585-590.

23. Seong, S.-Y. and P. Matzinger, *Hydrophobicity: an ancient damage-associated molecular pattern that initiates innate immune responses.* Nature Reviews Immunology, 2004. **4**(6): p. 469-478.

24. Hsu, Y.M.S., et al., *The adaptor protein CARD9 is required for innate immune responses to intracellular pathogens.* Nature Immunology, 2007. **8**(2): p. 198-205.

25. Cadwell, K., et al., *A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells.* Nature, 2008. **456**(7219): p. 259-63.

26. Wrackmeyer, U., et al., *Intelectin: a novel lipid raft-associated protein in the enterocyte brush border.* Biochemistry, 2006. **45**(30): p. 9188-97.

27. Bouma, G. and W. Strober, *The immunological and genetic basis of inflammatory bowel disease.* Nature Reviews Immunology, 2003. **3**(7): p. 521-533.

28. Kotlarz, D., et al., *Loss of interleukin-10 signaling and infantile inflammatory bowel disease: implications for diagnosis and therapy.* Gastroenterology, 2012. **143**(2): p. 347-355.

29. Halfvarson, J., et al., *Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics.* Gastroenterology, 2003. **124**(7): p. 1767-1773.

30. Molodecky, N.A. and G.G. Kaplan, *Environmental risk factors for inflammatory bowel disease.* Gastroenterology & hepatology, 2010. **6**(5): p. 339.

31. Calkins, B.M., *A meta-analysis of the role of smoking in inflammatory bowel disease.* Digestive Diseases and Sciences, 1989. **34**(12): p. 1841-1854.

32. Cosnes, J., et al., *Smoking, azathioprine, and clinical course in Crohn's disease - Reply.* Gastroenterology, 1996. **111**(4): p. 1161-1162.

33. Reese, G.E., et al., *The effect of smoking after surgery for Crohn's disease: a meta-analysis of observational studies.* Int J Colorectal Dis, 2008. **23**(12): p. 1213-21.

34. Birrenbach, T. and U. Böcker, *Inflammatory bowel disease and smoking. A review of epidemiology, pathophysiology, and therapeutic implications.* Inflammatory bowel diseases, 2004. **10**(6): p. 848-859.

35. Lindberg, E., G. Jarnerot, and B. Huitfeldt, *Smoking in Crohn's disease: effect on localisation and clinical course.* Gut, 1992. **33**(6): p. 779-82.

36. Cosnes, J., et al., *Effects of current and former cigarette smoking on the clinical course of Crohn's disease.* Aliment Pharmacol Ther, 1999. **13**(11): p. 1403-11.

37. van der Heide, F., et al., *Effects of active and passive smoking on disease course of Crohn's disease and ulcerative colitis.* Inflammatory bowel diseases, 2009. **15**(8): p. 1199-1207.

38. Brant, S.R., et al., *Defining complex contributions of NOD2/CARD15 gene mutations, age at onset, and tobacco use on Crohn's disease phenotypes.* Inflammatory bowel diseases, 2003. **9**(5): p. 281-289.

39. Bustamante, M., et al., *Relationship between smoking and colonic involvement in inflammatory bowel disease.* Revista Espanola De Enfermedades Digestivas, 1998. **90**(12): p. 837-840.

40. Parsi, M.A., et al., *Predictors of response to infliximab in patients with Crohn's disease.* Gastroenterology, 2002. **123**(3): p. 707-713.

41. Cosnes, J., et al., *Smoking cessation and the course of Crohn's disease: an intervention study.* Gastroenterology, 2001. **120**(5): p. 1093-1099.

42. Carbonnel, F., et al., *Environmental risk factors in Crohn's disease and ulcerative colitis: an update.* Gastroenterol Clin Biol, 2009. **33 Suppl 3**: p. S145-57.

43. Carbonnel, F., et al., *Environmental risk factors in Crohn's disease and ulcerative colitis: an update.* Gastroenterologie Clinique Et Biologique, 2009. **33**: p. S145-S157.

44. Amre, D.K., et al., *Imbalances in dietary consumption of fatty acids, vegetables, and fruits are associated with risk for Crohn's disease in children.* Am J Gastroenterol, 2007. **102**(9): p. 2016-25.

45. Jantchou, P., et al., *Animal protein intake and risk of inflammatory bowel disease: The E3N prospective study.* Am J Gastroenterol, 2010. **105**(10): p. 2195-201.

46. Sakamoto, N., et al., *Dietary risk factors for inflammatory bowel disease: a multicenter case-control study in Japan.* Inflamm Bowel Dis, 2005. **11**(2): p. 154-63.

47. Tragnone, A., et al., *Dietary Habits as Risk-Factors for Inflammatory Bowel-Disease.* European Journal of Gastroenterology & Hepatology, 1995. **7**(1): p. 47-51.

48. Reif, S., et al., *Pre-illness dietary factors in inflammatory bowel disease.* Gut, 1997. **40**(6): p. 754-60.

49. Halfvarson, J., et al., *Environmental factors in inflammatory bowel disease: a co-twin control study of a Swedish-Danish twin population.* Inflamm Bowel Dis, 2006. **12**(10): p. 925-33.

50. Bernstein, C.N., et al., *A population-based case control study of potential risk factors for IBD.* Am J Gastroenterol, 2006. **101**(5): p. 993-1002.

51. Guarner, F., et al., *Mechanisms of disease: the hygiene hypothesis revisited.* Nature Clinical Practice Gastroenterology & Hepatology, 2006. **3**(5): p. 275-284.

52. Shaw, S.Y., J.F. Blanchard, and C.N. Bernstein, *Association between the use of antibiotics in the first year of life and pediatric inflammatory bowel disease.* The American journal of gastroenterology, 2010. **105**(12): p. 2687-2692.

53. Ott, S., et al., *Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease.* Gut, 2004. **53**(5): p. 685-693.

54. Schultsz, C., et al., *The intestinal mucus layer from patients with inflammatory bowel disease harbors high numbers of bacteria compared with controls.* Gastroenterology, 1999. **117**(5): p. 1089-1097.

55. Mizoguchi, A. and E. Mizoguch, *Inflammatory bowel disease, past, present and future: lessons from animal models.* Journal of Gastroenterology, 2008. **43**(1): p. 1-17.

56. Rutgeerts, P., et al., *Effect of faecal stream diversion on recurrence of Crohn's disease in the neoterminal ileum.* Lancet, 1991. **338**(8770): p. 771-4.

57. Darfeuille-Michaud, A., et al., *High prevalence of adherent-invasive< i> Escherichia coli</i> associated with ileal mucosa in Crohn’s disease.* Gastroenterology, 2004. **127**(2): p. 412-421.

58. Stephens, N.S., et al., *Urinary NMR metabolomic profiles discriminate inflammatory bowel disease from healthy.* Journal of Crohn's and Colitis, 2012.

59. Scanlan, P.D., et al., *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): p. 3980-3988.

60. Rehman, A., et al., *Transcriptional activity of the dominant gut mucosal microbiota in chronic inflammatory bowel disease patients.* J Med Microbiol, 2010. **59**(Pt 9): p. 1114-22.

61. Frank, D.N., et al., *Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases.* Proc Natl Acad Sci U S A, 2007. **104**(34): p. 13780-5.

62. Baumgart, M., et al., *Culture independent analysis of ileal mucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum.* ISME J, 2007. **1**(5): p. 403-18.

63. Gophna, U., et al., *Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis.* Journal of clinical microbiology, 2006. **44**(11): p. 4136-4141.

64. Martinez-Medina, M., et al., *Abnormal microbiota composition in the ileocolonic mucosa of Crohn's disease patients as revealed by polymerase chain reaction‐denaturing gradient gel electrophoresis.* Inflammatory bowel diseases, 2006. **12**(12): p. 1136-1145.

65. Shanahan, F. and J. O'Mahony, *The mycobacteria story in Crohn's disease.* Am J Gastroenterol, 2005. **100**(7): p. 1537-8.

66. Man, S.M., et al., *Host attachment, invasion, and stimulation of proinflammatory cytokines by Campylobacter concisus and other non-Campylobacter jejuni Campylobacter species.* J Infect Dis, 2010. **202**(12): p. 1855-65.

67. Man, S.M., et al., *Campylobacter concisus and other Campylobacter species in children with newly diagnosed Crohn's disease.* Inflamm Bowel Dis, 2010. **16**(6): p. 1008-16.

68. Kaakoush, N.O., et al., *Detection of Helicobacteraceae in intestinal biopsies of children with Crohn's disease.* Helicobacter, 2010. **15**(6): p. 549-57.

69. Martinez-Medina, M., et al., *Molecular diversity of Escherichia coli in the human gut: new ecological evidence supporting the role of adherent-invasive E. coli (AIEC) in Crohn's disease.* Inflammatory bowel diseases, 2009. **15**(6): p. 872-882.

70. Lapaquette, P., et al., *Crohn's disease-associated adherent-invasive E. coli are selectively favoured by impaired autophagy to replicate intracellularly.* Cellular microbiology, 2010. **12**(1): p. 99-113.

71. Fujimoto, T., et al., *Decreased abundance of Faecalibacterium prausnitzii in the gut microbiota of Crohn's disease.* Journal of Gastroenterology and Hepatology, 2013. **28**(4): p. 613-619.

72. Willing, B., 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): p. 653-660.

73. Sokol, H., 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, 2008. **105**(43): p. 16731-16736.

74. Mylonaki, M., et al., *Molecular characterization of rectal mucosa-associated bacterial flora in inflammatory bowel disease.* Inflammatory bowel diseases, 2005. **11**(5): p. 481-487.

75. Sepehri, S., et al., *Microbial diversity of inflamed and noninflamed gut biopsy tissues in inflammatory bowel disease.* Inflammatory bowel diseases, 2007. **13**(6): p. 675-683.

76. Seksik, P., et al., *Search for localized dysbiosis in Crohn's disease ulcerations by temporal temperature gradient gel electrophoresis of 16S rRNA.* Journal of clinical microbiology, 2005. **43**(9): p. 4654-4658.

77. Sokol, H., et al., *Molecular comparison of dominant microbiota associated with injured versus healthy mucosa in ulcerative colitis.* Gut, 2007. **56**(1): p. 152-154.

78. Vasquez, N., 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): p. 684-692.

79. Bibiloni, R., et al., *The bacteriology of biopsies differs between newly diagnosed, untreated, Crohn's disease and ulcerative colitis patients.* Journal of medical microbiology, 2006. **55**(8): p. 1141-1149.

80. Kobayashi, K.S., et al., *Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract.* Science Signaling, 2005. **307**(5710): p. 731.

81. Brand, S., *Crohn's disease: Th1, Th17 or both? The change of a paradigm: new immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn's disease.* Gut, 2009. **58**(8): p. 1152-1167.

82. Wehkamp, J., et al., *Inducible and constitutive β-defensins are differentially expressed in Crohn's disease and ulcerative colitis.* Inflammatory bowel diseases, 2003. **9**(4): p. 215-223.

83. Ford, A.C., et al., *Efficacy of 5-aminosalicylates in Crohn's disease: systematic review and meta-analysis.* The American journal of gastroenterology, 2011. **106**(4): p. 617-629.

84. Sandborn, W., et al., *Baseline Corticosteroid Use and Corticosteroid-Free Clinical Remission in Crohn's Disease Patients Treated With Certolizumab Pegol in the PRECiSE 2 Trial: P-25.* Inflammatory bowel diseases, 2012. **18**: p. S23-S24.

85. Colombel, J.F., et al., *Infliximab, azathioprine, or combination therapy for Crohn's disease.* New England Journal of Medicine, 2010. **362**(15): p. 1383-1395.

86. Burger, D. and S. Travis, *Conventional medical management of inflammatory bowel disease.* Gastroenterology, 2011. **140**(6): p. 1827-1837. e2.

87. Lichtenstein, G.R., S.B. Hanauer, and W.J. Sandborn, *Management of Crohn's disease in adults.* The American journal of gastroenterology, 2009. **104**(2): p. 465-483.

88. Garrick, V., et al., *A multidisciplinary team model of caring for patients with perianal Crohn's disease incorporating a literature review, topical therapy and personal practice.* Frontline Gastroenterology, 2013. **4**(2): p. 152-160.

89. Gassull, M.A., *Conventional Medical Management of Crohn's Disease: Sulfasalazine*, in *Crohn's Disease and Ulcerative Colitis*. 2012, Springer. p. 365-369.

90. Nos, P. and E. Domenech, *Postoperative Crohn's disease recurrence: a practical approach.* World J Gastroenterol, 2008. **14**(36): p. 5540-8.

91. Doherty, G., et al., *Interventions for prevention of post-operative recurrence of Crohn's disease.* Cochrane Database Syst Rev, 2009(4): p. CD006873.

92. Regueiro, M., et al., *Infliximab prevents Crohn's disease recurrence after ileal resection.* Gastroenterology, 2009. **136**(2): p. 441-450. e1.

93. Dunn, W.B. and D.I. Ellis, *Metabolomics: current analytical platforms and methodologies.* TrAC Trends in Analytical Chemistry, 2005. **24**(4): p. 285-294.

94. Patti, G.J., O. Yanes, and G. Siuzdak, *Innovation: Metabolomics: the apogee of the omics trilogy.* Nature reviews Molecular cell biology, 2012. **13**(4): p. 263-269.

95. Claudino, W.M., et al., *Metabolomics: available results, current research projects in breast cancer, and future applications.* Journal of Clinical Oncology, 2007. **25**(19): p. 2840-2846.

96. German, J.B., B.D. Hammock, and S.M. Watkins, *Metabolomics: building on a century of biochemistry to guide human health.* Metabolomics, 2005. **1**(1): p. 3-9.

97. Rochfort, S., *Metabolomics reviewed: a new "omics" platform technology for systems biology and implications for natural products research.* Journal of natural products, 2005. **68**(12): p. 1813-1820.

98. Frydman, L., A. Lupulescu, and T. Scherf, *Principles and features of single-scan two-dimensional NMR spectroscopy.* Journal of the American Chemical Society, 2003. **125**(30): p. 9204-9217.

99. Moco, S., et al., *Metabolomics technologies and metabolite identification.* TrAC Trends in Analytical Chemistry, 2007. **26**(9): p. 855-866.

100. Unkel, A. *Basic Knowledge of Nuclear Magnetic Resonance Spectroscopy (NMR)*. 2012; Available from: http://cnx.org/content/col11429/1.1/.

101. Dettmer, K., P.A. Aronov, and B.D. Hammock, *Mass spectrometry‐based metabolomics.* Mass spectrometry reviews, 2007. **26**(1): p. 51-78.

102. Alpaydin, E., *Introduction to machine learning*. 2004: MIT press.

103. Abdi, H. and L.J. Williams, *Principal component analysis.* Wiley Interdisciplinary Reviews: Computational Statistics, 2010. **2**(4): p. 433-459.

104. Hastie, T., R. Tibshirani, and J. Friedman, *The Elements of Statistical Learnin*. 2009, New York: Springer.

105. Quinlan, J.R., *C4. 5: programs for machine learning*. Vol. 1. 1993: Morgan kaufmann.

106. Rennie, J.D., et al. *Tackling the poor assumptions of naive bayes text classifiers*. in *ICML*. 2003: Washington DC).

107. Cortes, C. and V. Vapnik, *Support-vector networks.* Machine learning, 1995. **20**(3): p. 273-297.

108. Meyer, D., *Support Vector Machines.* The Interface to libsvm in package e1071. e1071 Vignette, 2012.

109. Wold, H., *Estimation of principal components and related models by iterative least squares.* Multivariate analysis, 1966. **1**: p. 391-420.

110. Barker, M. and W. Rayens, *Partial least squares for discrimination.* Journal of chemometrics, 2003. **17**(3): p. 166-173.

111. Sikaroodi, M. and P.M. Gillevet, *Quality control in multi-tag pyrosequencing of microbial communities.* Biotechniques, 2012. **53**(6): p. 381-3.

112. PM, G., *Multitag Sequencing and Ecogenomic Analysis*, EPO, Editor. 2006.

113. Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST.* Bioinformatics, 2010. **26**(19): p. 2460-2461.

114. Hall, M., et al., *The WEKA data mining software: an update.* ACM SIGKDD Explorations Newsletter, 2009. **11**(1): p. 10-18.

115. Xia, J., et al., *MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis.* Nucleic Acids Research, 2012. **40**(W1): p. W127-W133.

116. Borowiec, A.M., et al., *Small bowel fibrosis and systemic inflammatory response after ileocolonic anastomosis in IL-10 null mice.* Journal of Surgical Research, 2012. **178**(1): p. 147-154.

117. Mukhopadhya, I., et al., *IBD—what role do Proteobacteria play?* Nature Reviews Gastroenterology and Hepatology, 2012. **9**(4): p. 219-230.

118. Sakai, T., et al., *Rapid, simple enzymatic assay of free L-fucose in serum and urine, and its use as a marker for cancer, cirrhosis, and gastric ulcers.* Clinical chemistry, 1990. **36**(3): p. 474-476.

119. Migliaccio, C.T., et al., *Urinary levoglucosan as a biomarker of wood smoke exposure: observations in a mouse model and in children.* Environmental health perspectives, 2009. **117**(1): p. 74.

120. Louis, P. and H.J. Flint, *Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine.* FEMS microbiology letters, 2009. **294**(1): p. 1-8.

121. Wishart, D.S., et al., *HMDB: a knowledgebase for the human metabolome.* Nucleic Acids Research, 2009. **37**(suppl 1): p. D603-D610.

122. ŞTEFAN, L.I., et al., *1H-NMR URINE METABOLIC PROFILING IN TYPE 1 DIABETES MELLITUS.* Rev. Roum. Chim, 2010. **55**(11-12): p. 1033-1037.

123. van Dorsten, F.A., et al., *The metabolic fate of red wine and grape juice polyphenols in humans assessed by metabolomics.* Molecular nutrition & food research, 2010. **54**(7): p. 897-908.

124. Hoehn, T., et al., *Urinary excretion of the nitrotyrosine metabolite 3-nitro-4-hydroxyphenylacetic acid in preterm and term infants.* Neonatology, 2007. **93**(2): p. 73-76.

125. Bales, J.R., et al., *Use of high-resolution proton nuclear magnetic resonance spectroscopy for rapid multi-component analysis of urine.* Clinical chemistry, 1984. **30**(3): p. 426-432.

126. Monot, F., et al., *Acetone and butanol production by Clostridium acetobutylicum in a synthetic medium.* Applied and environmental microbiology, 1982. **44**(6): p. 1318-1324.

127. Silva, C.L., M. Passos, and J.S. Câmara, *Solid phase microextraction, mass spectrometry and metabolomic approaches for detection of potential urinary cancer biomarkers—A powerful strategy for breast cancer diagnosis.* Talanta, 2012. **89**: p. 360-368.

128. Marescau, B., et al., *Guanidino compounds in serum and urine of cirrhotic patients.* Metabolism, 1995. **44**(5): p. 584-588.

129. Stretch, C., et al., *Prediction of skeletal muscle and fat mass in patients with advanced cancer using a metabolomic approach.* The Journal of nutrition, 2012. **142**(1): p. 14-21.

130. Nittono, H., et al., *Sulfated and nonsulfated bile acids in urine of patients with biliary atresia: analysis of bile acids by high-performance liquid chromatography.* Journal of Pediatric Gastroenterology and Nutrition, 1986. **5**(1): p. 23-29.

131. Jones, A. and A. Helander, *Changes in the concentrations of ethanol, methanol and metabolites of serotonin in two successive urinary voids from drinking drivers.* Forensic science international, 1998. **93**(2): p. 127-134.

132. Smythe, G.A., et al., *ECNI GC-MS analysis of picolinic and quinolinic acids and their amides in human plasma, CSF, and brain tissue*, in

Developments in Tryptophan and Serotonin Metabolism. 2003, Springer. p. 705-712.

133. Fonteh, A., et al., *Free amino acid and dipeptide changes in the body fluids from Alzheimer's disease subjects.* Amino acids, 2007. **32**(2): p. 213-224.

134. Yamamoto, T., M. Nakahigashi, and A. Saniabadi, *Review article: diet and inflammatory bowel disease–epidemiology and treatment.* Alimentary Pharmacology & Therapeutics, 2009. **30**(2): p. 99-112.

135. Harig, J.M., et al., *Treatment of diversion colitis with short-chain-fatty acid irrigation.* N Engl J Med, 1989. **320**(1): p. 23-8.

136. Cho, E.-j., et al., *Anti-inflammatory effects of methanol extract of Patrinia scabiosaefolia in mice with ulcerative colitis.* Journal of ethnopharmacology, 2011. **136**(3): p. 428-435.

137. Maslowski, K.M., et al., *Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43.* Nature, 2009. **461**(7268): p. 1282-1286.

138. Macfarlane, S. and G.T. Macfarlane. *Regulation of short-chain fatty acid production*. in *PROCEEDINGS-NUTRITION SOCIETY OF LONDON*. 2003: Cambridge Univ Press.

139. Schicho, R., et al., *Quantitative metabolomic profiling of serum, plasma, and urine by 1H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals.* Journal of Proteome Research, 2012. **11**(6): p. 3344-3357.

140. Taher, E. and K. Chandran, *High-Rate, High-Yield Production of Methanol by Ammonia-Oxidizing Bacteria.* Environmental science & technology, 2013. **47**(7): p. 3167-3173.

141. Antunes, D.M.F., et al., *The serum D-xylose test as a useful tool to identify malabsorption in rats with antigen specific gut inflammatory reaction.* International journal of experimental pathology, 2009. **90**(2): p. 141-147.

142. Elamin, E.E., et al., *Short-Chain Fatty Acids Activate AMP-Activated Protein Kinase and Ameliorate Ethanol-Induced Intestinal Barrier Dysfunction in Caco-2 Cell Monolayers.* The Journal of nutrition, 2013: p. jn. 113.179549.

143. BALCH, W.E., et al., *Acetobacterium, a new genus of hydrogen-oxidizing, carbon dioxide-reducing, anaerobic bacteria.* International Journal of Systematic Bacteriology, 1977. **27**(4): p. 355-361.

144. Gronwald, W., et al., *Urinary metabolite quantification employing 2D NMR spectroscopy.* Analytical chemistry, 2008. **80**(23): p. 9288-9297.

145. Willing, B.P., et al., *A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes.* Gastroenterology, 2010. **139**(6): p. 1844-1854. e1.

146. Frank, D.N., et al., *Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases.* Proceedings of the National Academy of Sciences, 2007. **104**(34): p. 13780-13785.

147. Man, S.M., N.O. Kaakoush, and H.M. Mitchell, *The role of bacteria and pattern-recognition receptors in Crohn's disease.* Nature Reviews Gastroenterology and Hepatology, 2011. **8**(3): p. 152-168.

148. Neut, C., et al., *Changes in the bacterial flora of the neoterminal ileum after ileocolonic resection for Crohn's disease.* The American journal of gastroenterology, 2002. **97**(4): p. 939-946.

149. Salonen, A., W.M. de Vos, and A. Palva, *Gastrointestinal microbiota in irritable bowel syndrome: present state and perspectives.* Microbiology, 2010. **156**(11): p. 3205-3215.

150. Hooper, L.V., T. Midtvedt, and J.I. Gordon, *How host-microbial interactions shape the nutrient environment of the mammalian intestine.* Annual review of nutrition, 2002. **22**(1): p. 283-307.

151. Wang, Q., B. Chang, and T. Riley, *Erysipelothrix rhusiopathiae.* Veterinary microbiology, 2010. **140**(3-4): p. 405.

152. Sekirov, I., et al., *Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection.* Infection and immunity, 2008. **76**(10): p. 4726-4736.

153. Omatsu, T., et al., *Involvement of reactive oxygen species in indomethacin-induced apoptosis of small intestinal epithelial cells.* Journal of Gastroenterology, 2009. **44**(19): p. 30-34.

154. Kuwano, Y., et al., *Tumor necrosis factor alpha activates transcription of the NADPH oxidase organizer 1 (NOXO1) gene and upregulates superoxide production in colon epithelial cells.* Free Radic Biol Med, 2008. **45**(12): p. 1642-52.

155. Yu, L.C.-H., et al., *Host-microbial interactions and regulation of intestinal epithelial barrier function: from physiology to pathology.* World journal of gastrointestinal pathophysiology, 2012. **3**(1): p. 27.

156. Fischbach, M.A. and J.L. Sonnenburg, *Eating for two: how metabolism establishes interspecies interactions in the gut.* Cell host & microbe, 2011. **10**(4): p. 336-347.

157. Pacheco, A.R., et al., *Fucose sensing regulates bacterial intestinal colonization.* Nature, 2012. **492**(7427): p. 113-7.

158. Atarashi, K., et al., *ATP drives lamina propria TH17 cell differentiation.* Nature, 2008. **455**(7214): p. 808-812.

159. Mondot, S., et al., *Highlighting new phylogenetic specificities of Crohn's disease microbiota.* Inflammatory bowel diseases, 2011. **17**(1): p. 185-192.

160. Naftali, T., et al., *Cannabis Induces a Clinical Response in Patients with Crohn's Disease: a Prospective Placebo-Controlled Study.* Clinical Gastroenterology and Hepatology, 2013.

161. Gustavsson, A., et al., *Smoking is a risk factor for recurrence of intestinal stricture after endoscopic dilation in Crohn's disease.* Alimentary Pharmacology & Therapeutics, 2013. **37**(4): p. 430-437.

162. Montgomery, S., et al., *Siblings and the risk of inflammatory bowel disease.* Scandinavian Journal of Gastroenterology, 2002. **37**(11): p. 1301-1308.

163. Bernstein, C.N., et al., *Population-based case control study of seroprevalence of Mycobacterium paratuberculosis in patients with Crohn's disease and ulcerative colitis.* Journal of clinical microbiology, 2004. **42**(3): p. 1129-1135.

164. Turnbaugh, P.J., et al., *The human microbiome project.* Nature, 2007. **449**(7164): p. 804-810.

Appendix

1. DNA Extraction Protocol

- 1.0mL of TH150 buffer, (0.176g Tris-HCl, 0.106g Trisma base, 1.74g NaCl, and 700ml water) 50µl of sodium dodecyl sulfate (10%w/v) and 10µl of Proteinase K (20mg/ml) was added to the tissue and incubated at 55˚C for 1-2hours to completely digest the tissue.
- Samples were re-suspended and vortexed to which 0.1mm zirconia beads were added. Vortex continuously to homogenize the sample tissue.
- Process the sample three times using a bead beater for 30 seconds each with cooling on ice between each round. The tubes were then heated at 95˚C for 15 minutes
- Tubes were vortexed and centrifuged at 13600 rpm for 1 min,
- The supernatant was pipetted into a 2ml microcentrifuge tube to which 1 Inhibit EX Tablet is added and vortex until dissolved. Then the suspension was incubated for 1 minute at room temperature to allow inhibitors to adsorb to the Inhibit EX matrix.
- Samples were centrifuged at 13600rmp for 5 minutes to pellet inhibitors bound to the InhibitEX matrix
- The supernatant was transferred to a clean 1.5ml eppendorff tube and centrifuged again for 3 minutes at 13600 rpm to clean from the InhibitEX residues.
- 200µl of the supernatant, 15µl of Proteinase K, and 200µl Buffer AL was added to a new 1.5ml microcentrifuge tube and vortexed to mix. The samples were then incubated at 70˚C for 10 min.
- 200µl of ethanol (100%) was added to lysate the cells then vortexed.
- The samples were then added to a QIAamp spin column (provided with the kit) and centrifuged at 13600rpm for 1 min and the filtrate was discarded
- 500µl Buffer AW1 was added to the spin column, centrifuged at 13600rmp for 1min then repeated with 500µl Buffer AW2.
- The spin column was transferred to a new 1.5mL microcentrifuge tube to which 25-50µl sterile water was added directly to the QIAamp membrane. Then the tubes were incubated for 1 minute at room temperature, centrifuged at 13600rmp for 1 min to elute the DNA and repeated with another 25-50µl of sterile water.

2. Biopsy Locations

Biopsy #	Specimen Type
GR5266	Ileum
GR5384	Ileum
GR5407	Ileum
GR5455	Ileum
GR5592	Ileum
GR5593	Ileum
GR5601	Ileum
GR5603	Ileum
GR5609	Ileum
GR5625	Ileum
GR5630	Ileum
GR5781	D-Colon
GR5786	Ileum
GR5793	Ileum
GR5796	Ileum
GR5811	Ileum
GR5814	Ileum
GR5825	Ileum
GR5838	Ileum
GR5849	Ileum
GR5092	Ileum
GR5093	Ileum
GR5138	Ileum
GR5141	Ileum
GR5267	Ileum
GR5581	D-Colon
GR5583	A-Colon
GR5611	Ileum
GR5615	Ileum
GR5681	D-Colon
GR5683	A-Colon
GR5691	A-Colon
GR5785	A-Colon
GR5788	Ileum
GR5791	Ileum
GR5817	Ileum
GR5862	Ileum

3. P-values for Environmental Questions

P-values for the comparison between relapse and remission patients

Environmental Factor	**P-Value (Fishers exact test)**
Smokes marijuana	0.132
Herbal drugs	0.207
Prednisone	0.098
Aspirin	0.157
2nd hand smoke- spouse	0.450
High sugar- weekly	0.225
Asacol in the past	0.314
Prednisone in the past	0.293
Entocort capsules in the past	0.186
Azathioprine in the past	0.205
Pentasa in the past	0.117
Birth control pills in 20's	0.163
High fiber in childhood	0.293
Animal Exposure on a Farm in childhood	0.125
Eats Fish in childhood	0.117
Drink non tap water in childhood	0.186

P-values for the comparison between patients with any inflammation to patients with no inflammation

Environmental Factor	**P-Value (Fishers exact test)**
Smokes marijuana	0.560
Herbal drugs	0.309
Prednisone	0.552
Aspirin	0.605
2nd hand smoke- spouse	0.667
High sugar- weekly	0.689
Asacol in the past	0.393
Prednisone in the past	0.650
Entocort capsules in the past	0.055
Azathioprine in the past	0.407
Pentasa in the past	0.592
Birth control pills in 20's	0.650
High fiber in childhood	0.650
Animal Exposure on a Farm in childhood	0.211
Eats Fish in childhood	0.094
Drink non tap water in childhood	0.187

4. Raw Bacterial Data

PHYLUM

Relapse Patients

RDP10 PHYLUM	GR5092	GR5093	GR5138	GR5141	GR5267	GR5581	GR5583	GR5611	GR5615	GR5681	GR5683	GR5691	GR5785	GR5788	GR5791	GR5817	GR5862	**Average**	**Standard Deviation**
Actinobacteria	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.012	0.005	0.016
Bacteroidetes	0.275	0.605	0.424	0.163	0.284	0.309	0.000	0.572	0.012	0.428	0.063	0.245	0.227	0.200	0.016	0.038	0.000	0.227	0.196
Firmicutes	0.563	0.232	0.504	0.676	0.629	0.309	0.723	0.389	0.668	0.508	0.713	0.523	0.399	0.577	0.810	0.789	0.902	0.583	0.183
Fusobacteria	0.025	0.114	0.000	0.066	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.076	0.000	0.000	0.000	0.017	0.035
Proteobacteria	0.128	0.047	0.067	0.085	0.080	0.379	0.257	0.037	0.251	0.052	0.212	0.221	0.366	0.144	0.163	0.157	0.075	0.160	0.107

Remission Patients

RDP10 PHYLUM	GR5266	GR5384	GR5407	GR5455	GR5592	GR5593	GR5601	GR5603	GR5609	GR5625	GR5630	GR5781	GR5786	GR5793	GR5796	GR5811	GR5814	GR5825	GR5838	GR5849	**Average**	**Standard Deviation**
Actinobacteria	0.000	0.000	0.000	0.000	0.015	0.000	0.168	0.011	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.011	0.037
Bacteroidetes	0.305	0.160	0.420	0.505	0.131	0.582	0.487	0.380	0.075	0.331	0.013	0.629	0.389	0.303	0.634	0.103	0.494	0.604	0.243	0.000	0.339	0.208
Firmicutes	0.469	0.603	0.540	0.437	0.480	0.374	0.283	0.521	0.525	0.552	0.827	0.320	0.530	0.596	0.319	0.602	0.480	0.319	0.620	0.487	0.494	0.130
Fusobacteria	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.349	0.032	0.000	0.000	0.000	0.000	0.000	0.199	0.000	0.000	0.000	0.000	0.034	0.089
Proteobacteria	0.211	0.123	0.035	0.047	0.370	0.041	0.061	0.086	0.048	0.065	0.153	0.046	0.075	0.093	0.041	0.090	0.020	0.072	0.122	0.500	0.115	0.120

CLASS

Relapse Patients

RDP10 CLASS	GR5092	GR5093	GR5138	GR5141	GR5267	GR5581	GR5583	GR5611	GR5615	GR5681	GR5683	GR5691	GR5785	GR5788	GR5791	GR5817	GR5862	**Average**	**Standard Deviation**
Actinobacteria_Actinobacteria	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.012	0.005	0.016
Bacteroidetes_Bacteroidia	0.263	0.604	0.418	0.155	0.277	0.305	0.000	0.570	0.000	0.425	0.054	0.233	0.225	0.198	0.000	0.028	0.000	0.221	0.198
Bacteroidetes_Flavobacteria	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.013	0.010	0.000	0.003	0.005
Firmicutes_Bacilli	0.491	0.049	0.266	0.324	0.397	0.199	0.363	0.140	0.620	0.282	0.435	0.399	0.204	0.127	0.705	0.627	0.333	0.351	0.185
Firmicutes_Clostridia	0.071	0.175	0.226	0.333	0.231	0.039	0.337	0.214	0.048	0.225	0.219	0.123	0.194	0.374	0.085	0.142	0.452	0.205	0.118
Firmicutes_Erysipelotrichi	0.000	0.000	0.012	0.019	0.000	0.071	0.023	0.035	0.000	0.000	0.059	0.000	0.000	0.075	0.020	0.020	0.117	0.027	0.034
Fusobacteria_Fusobacteria	0.025	0.114	0.000	0.066	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.076	0.000	0.000	0.000	0.017	0.035
Proteobacteria_Alphaproteobacteria	0.000	0.027	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.006
Proteobacteria_Betaproteobacteria	0.012	0.000	0.000	0.013	0.000	0.000	0.027	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.010	0.000	0.005	0.008
Proteobacteria_Gammaproteobacteria	0.110	0.014	0.055	0.068	0.067	0.370	0.223	0.031	0.244	0.041	0.201	0.207	0.352	0.136	0.138	0.140	0.065	0.145	0.108
Unclassified	0.016	0.018	0.024	0.023	0.028	0.015	0.026	0.009	0.021	0.027	0.032	0.026	0.025	0.014	0.021	0.010	0.020	0.021	0.006

Remission Patients

RDP10 CLASS	GR5266	GR5384	GR5407	GR5455	GR5592	GR5593	GR5601	GR5603	GR5609	GR5625	GR5630	GR5781	GR5786	GR5793	GR5796	GR5811	GR5814	GR5825	GR5838	GR5849	**Average**	**Standard Deviation**
Actinobacteria_Actinobacteria	0.000	0.000	0.000	0.000	0.015	0.000	0.168	0.011	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.011	0.037
Bacteroidetes_Bacteroidia	0.295	0.150	0.418	0.504	0.130	0.582	0.483	0.371	0.074	0.327	0.000	0.628	0.387	0.289	0.633	0.099	0.494	0.602	0.242	0.000	0.335	0.209
Bacteroidetes_Flavobacteria	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004
Firmicutes_Bacilli	0.390	0.579	0.136	0.146	0.276	0.056	0.223	0.385	0.084	0.303	0.812	0.044	0.173	0.451	0.146	0.264	0.061	0.226	0.322	0.264	0.267	0.191
Firmicutes_Clostridia	0.062	0.022	0.398	0.269	0.184	0.310	0.059	0.133	0.414	0.247	0.013	0.231	0.332	0.117	0.171	0.310	0.410	0.089	0.269	0.214	0.213	0.128
Firmicutes_Erysipelotrichi	0.017	0.000	0.000	0.023	0.020	0.000	0.000	0.000	0.027	0.000	0.000	0.046	0.025	0.028	0.000	0.027	0.000	0.000	0.029	0.000	0.012	0.015
Fusobacteria_Fusobacteria	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.349	0.032	0.000	0.000	0.000	0.000	0.000	0.199	0.000	0.000	0.000	0.000	0.034	0.089
Proteobacteria_Alphaproteobacteria	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Proteobacteria_Betaproteobacteria	0.014	0.016	0.000	0.000	0.011	0.000	0.013	0.014	0.000	0.000	0.014	0.000	0.025	0.000	0.000	0.000	0.000	0.029	0.000	0.000	0.007	0.009
Proteobacteria_Gammaproteobacteria	0.190	0.097	0.029	0.040	0.355	0.040	0.047	0.066	0.043	0.047	0.128	0.040	0.047	0.081	0.031	0.076	0.012	0.041	0.110	0.492	0.101	0.120
Unclassified	0.032	0.036	0.018	0.019	0.009	0.012	0.007	0.020	0.008	0.027	0.021	0.011	0.011	0.020	0.018	0.024	0.023	0.013	0.013	0.029	0.019	0.008

ORDER

Relapse Patients

RDP10 ORDER	GR5092	GR5093	GR5138	GR5141	GR5267	GR5581	GR5583	GR5611	GR5615	GR5621	GR5681	GR5683	GR5691	GR5785	GR5788	GR5791	GR5817	GR5862	**Average**	**Standard Deviation**
Actinobacteria_Actinobacteria_Actinomycetales	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Actinobacteria_Actinobacteria_Bifidobacteriales	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.016
Actinobacteria_Actinobacteria_Coriobacteriales	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Bacteroidetes_Bacteroidia_Bacteroidales	0.263	0.604	0.418	0.155	0.277	0.305	0.000	0.570	0.000	0.153	0.425	0.054	0.233	0.225	0.198	0.000	0.028	0.000	0.217	0.193
Bacteroidetes_Flavobacteria_Flavobacteriales	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.013	0.010	0.000	0.003	0.005
Firmicutes_Bacilli_Lactobacillales	0.490	0.049	0.265	0.323	0.397	0.199	0.363	0.140	0.619	0.000	0.282	0.433	0.399	0.204	0.127	0.705	0.626	0.333	0.331	0.198
Firmicutes_Clostridia_Clostridiales	0.071	0.175	0.226	0.333	0.231	0.039	0.335	0.214	0.048	0.273	0.225	0.219	0.123	0.194	0.374	0.085	0.142	0.452	0.209	0.115
Firmicutes_Erysipelotrichi_Erysipelotrichales	0.000	0.000	0.012	0.019	0.000	0.071	0.023	0.035	0.000	0.047	0.000	0.059	0.000	0.000	0.075	0.020	0.020	0.117	0.028	0.034
Fusobacteria_Fusobacteria_Fusobacteriales	0.025	0.114	0.000	0.066	0.000	0.000	0.000	0.000	0.000	0.488	0.000	0.000	0.000	0.000	0.076	0.000	0.000	0.000	0.043	0.116
Proteobacteria_Alphaproteobacteria_Rhizobiales	0.000	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.006
Proteobacteria_Betaproteobacteria_Burkholderiales	0.000	0.000	0.000	0.000	0.000	0.000	0.024	0.000	0.000	0.036	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.004	0.010
Proteobacteria_Gammaproteobacteria_Enterobacteriales	0.055	0.010	0.017	0.026	0.025	0.350	0.190	0.018	0.202	0.000	0.000	0.164	0.170	0.331	0.116	0.053	0.064	0.024	0.101	0.110
Proteobacteria_Gammaproteobacteria_Pasteurellales	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Proteobacteria_Gammaproteobacteria_Pseudomonadales	0.051	0.000	0.036	0.030	0.040	0.017	0.031	0.012	0.031	0.000	0.024	0.031	0.032	0.019	0.013	0.079	0.065	0.035	0.030	0.020
Unclassified	0.032	0.023	0.026	0.048	0.030	0.020	0.034	0.011	0.032	0.004	0.045	0.041	0.032	0.027	0.020	0.035	0.044	0.039	0.030	0.012

Remission Patients

RDP10 ORDER	GR5266	GR5384	GR5407	GR5455	GR5592	GR5593	GR5601	GR5603	GR5609	GR5625	GR5630	GR5781	GR5786	GR5793	GR5796	GR5811	GR5814	GR5825	GR5838	GR5849	**Average**	**Standard Deviation**
Actinobacteria_Actinobacteria_Actinomycetales	0.000	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
Actinobacteria_Actinobacteria_Bifidobacteriales	0.000	0.000	0.000	0.000	0.000	0.000	0.062	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.014
Actinobacteria_Actinobacteria_Coriobacteriales	0.000	0.000	0.000	0.000	0.000	0.000	0.105	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.006	0.023
Bacteroidetes_Bacteroidia_Bacteroidales	0.295	0.150	0.418	0.504	0.130	0.582	0.483	0.371	0.074	0.327	0.000	0.628	0.387	0.289	0.633	0.099	0.494	0.602	0.242	0.000	0.335	0.209
Bacteroidetes_Flavobacteria_Flavobacteriales	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004
Firmicutes_Bacilli_Lactobacillales	0.388	0.579	0.136	0.145	0.276	0.055	0.223	0.385	0.084	0.300	0.812	0.043	0.173	0.449	0.146	0.264	0.061	0.224	0.321	0.260	0.266	0.191
Firmicutes_Clostridia_Clostridiales	0.062	0.022	0.398	0.269	0.184	0.310	0.059	0.133	0.414	0.247	0.013	0.231	0.332	0.117	0.171	0.310	0.410	0.089	0.269	0.214	0.213	0.128
Firmicutes_Erysipelotrichi_Erysipelotrichales	0.017	0.000	0.000	0.023	0.020	0.000	0.000	0.000	0.027	0.000	0.000	0.046	0.025	0.028	0.000	0.027	0.000	0.000	0.029	0.000	0.012	0.015
Fusobacteria_Fusobacteria_Fusobacteriales	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.349	0.032	0.000	0.000	0.000	0.000	0.000	0.199	0.000	0.000	0.000	0.000	0.034	0.089
Proteobacteria_Alphaproteobacteria_Rhizobiales	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Proteobacteria_Betaproteobacteria_Burkholderiales	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.010	0.000	0.000	0.000	0.000	0.022	0.000	0.000	0.000	0.000	0.029	0.000	0.000	0.004	0.008
Proteobacteria_Gammaproteobacteria_Enterobacteriales	0.145	0.045	0.018	0.025	0.331	0.029	0.030	0.025	0.000	0.022	0.040	0.037	0.026	0.029	0.014	0.051	0.000	0.026	0.080	0.469	0.072	0.118
Proteobacteria_Gammaproteobacteria_Pasteurellales	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.007
Proteobacteria_Gammaproteobacteria_Pseudomonadales	0.042	0.048	0.000	0.011	0.021	0.000	0.015	0.037	0.000	0.023	0.081	0.000	0.017	0.048	0.015	0.024	0.000	0.000	0.026	0.019	0.021	0.021
Unclassified	0.051	0.056	0.030	0.023	0.027	0.023	0.011	0.039	0.022	0.048	0.043	0.015	0.018	0.026	0.021	0.026	0.035	0.030	0.024	0.037	0.030	0.012

Family

Relapse Patients

RDP10 FAMILY	GR5092	GR5093	GR5138	GR5141	GR5267	GR5581	GR5583	GR5611	GR5615	GR5621	GR5681	GR5683	GR5691	GR5785	GR5788	GR5791	GR5817	GR5862	**Average**	**Standard Deviation**
Actinobacteria_Actinobacteria_Actinomycetales_Micrococcaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Actinobacteria_Actinobacteria_Bifidobacteriales_Bifidobacteriaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.016
Actinobacteria_Actinobacteria_Coriobacteriales_Coriobacteriaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Bacteroidetes_Bacteroidia_Bacteroidales_Bacteroidaceae	0.260	0.600	0.311	0.145	0.273	0.303	0.000	0.513	0.000	0.136	0.385	0.047	0.230	0.149	0.196	0.000	0.021	0.000	0.198	0.180
Bacteroidetes_Bacteroidia_Bacteroidales_Porphyromonadaceae	0.000	0.000	0.045	0.000	0.000	0.000	0.000	0.052	0.000	0.000	0.029	0.000	0.000	0.076	0.000	0.000	0.000	0.000	0.011	0.023
Bacteroidetes_Bacteroidia_Bacteroidales_Prevotellaceae	0.000	0.000	0.053	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.013
Bacteroidetes_Bacteroidia_Bacteroidales_Rikenellaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Bacteroidetes_Flavobacteria_Flavobacteriales_Flavobacteriaceae	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.013	0.010	0.000	0.003	0.005
Firmicutes_Bacilli_Lactobacillales_Enterococcaceae	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.017	0.000	0.000	0.000	0.002	0.005
Firmicutes_Bacilli_Lactobacillales_Lactobacillaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.269	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.063
Firmicutes_Bacilli_Lactobacillales_Leuconostocaceae	0.362	0.035	0.196	0.231	0.296	0.147	0.267	0.108	0.263	0.000	0.214	0.315	0.293	0.157	0.081	0.517	0.455	0.247	0.232	0.135
Firmicutes_Bacilli_Lactobacillales_Streptococcaceae	0.117	0.013	0.063	0.085	0.090	0.049	0.091	0.029	0.081	0.000	0.064	0.108	0.093	0.044	0.029	0.175	0.157	0.078	0.076	0.046
Firmicutes_Clostridia_Clostridiales_Clostridiaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.023	0.031	0.000	0.000	0.000	0.004	0.010
Firmicutes_Clostridia_Clostridiales_Eubacteriaceae	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
Firmicutes_Clostridia_Clostridiales_Incertae Sedis XIV	0.000	0.000	0.024	0.011	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.010	0.022	0.005	0.008
Firmicutes_Clostridia_Clostridiales_Lachnospiraceae	0.044	0.145	0.138	0.288	0.209	0.029	0.290	0.179	0.000	0.087	0.097	0.200	0.090	0.146	0.261	0.000	0.089	0.396	0.149	0.109
Firmicutes_Clostridia_Clostridiales_Peptostreptococcaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.160	0.000	0.000	0.012	0.016	0.044	0.056	0.000	0.000	0.016	0.039
Firmicutes_Clostridia_Clostridiales_Ruminococcaceae	0.000	0.000	0.039	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.106	0.000	0.000	0.000	0.000	0.000	0.012	0.016	0.010	0.026
Firmicutes_Clostridia_Clostridiales_Veillonellaceae	0.018	0.000	0.016	0.000	0.000	0.000	0.031	0.011	0.045	0.000	0.000	0.000	0.000	0.000	0.026	0.018	0.014	0.000	0.010	0.014
Firmicutes_Erysipelotrichi_Erysipelotrichales_Erysipelotrichaceae	0.000	0.000	0.012	0.019	0.000	0.071	0.023	0.035	0.000	0.047	0.000	0.059	0.000	0.000	0.075	0.020	0.020	0.117	0.028	0.034
Fusobacteria_Fusobacteria_Fusobacteriales_Fusobacteriaceae	0.025	0.114	0.000	0.065	0.000	0.000	0.000	0.000	0.000	0.487	0.000	0.000	0.000	0.000	0.076	0.000	0.000	0.000	0.043	0.116
Proteobacteria_Alphaproteobacteria_Rhizobiales_Methylocystaceae	0.000	0.022	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.005
Proteobacteria_Betaproteobacteria_Burkholderiales_Alcaligenaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.000	0.000	0.036	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.009
Proteobacteria_Gammaproteobacteria_Enterobacteriales_Enterobacteriaceae	0.055	0.010	0.017	0.026	0.025	0.350	0.190	0.018	0.202	0.000	0.000	0.164	0.170	0.331	0.116	0.053	0.064	0.024	0.101	0.110
Proteobacteria_Gammaproteobacteria_Pasteurellales_Pasteurellaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Proteobacteria_Gammaproteobacteria_Pseudomonadales_Moraxellaceae	0.050	0.000	0.036	0.030	0.039	0.016	0.030	0.012	0.028	0.000	0.023	0.031	0.032	0.018	0.011	0.079	0.063	0.034	0.030	0.020
Unclassified	0.058	0.060	0.051	0.074	0.057	0.036	0.062	0.031	0.045	0.012	0.081	0.077	0.057	0.040	0.026	0.070	0.084	0.067	0.055	0.020

Remission Patients

RDP10 FAMILY	GR5266	GR5384	GR5407	GR5455	GR5592	GR5593	GR5601	GR5603	GR5609	GR5625	GR5630	GR5781	GR5786	GR5793	GR5796	GR5811	GR5814	GR5825	GR5838	GR5849	**Average**	**Standard Deviation**
Actinobacteria_Actinobacteria_Actinomycetales_Micrococcaceae	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002
Actinobacteria_Actinobacteria_Bifidobacteriales_Bifidobacteriaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.062	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.014
Actinobacteria_Actinobacteria_Coriobacteriales_Coriobacteriaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.105	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.006	0.023
Bacteroidetes_Bacteroidia_Bacteroidales_Bacteroidaceae	0.293	0.148	0.386	0.491	0.127	0.559	0.195	0.263	0.073	0.295	0.000	0.563	0.385	0.283	0.631	0.098	0.427	0.597	0.237	0.000	0.302	0.198
Bacteroidetes_Bacteroidia_Bacteroidales_Porphyromonadaceae	0.000	0.000	0.026	0.000	0.000	0.020	0.017	0.088	0.000	0.025	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.024
Bacteroidetes_Bacteroidia_Bacteroidales_Prevotellaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.267	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.044	0.000	0.000	0.000	0.016	0.060
Bacteroidetes_Bacteroidia_Bacteroidales_Rikenellaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.000	0.002	0.005
Bacteroidetes_Flavobacteria_Flavobacteriales_Flavobacteriaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
Firmicutes_Bacilli_Lactobacillales_Enterococcaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.061	0.000	0.000	0.003	0.014
Firmicutes_Bacilli_Lactobacillales_Lactobacillaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.062	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.014
Firmicutes_Bacilli_Lactobacillales_Leuconostocaceae	0.286	0.422	0.104	0.104	0.170	0.044	0.120	0.290	0.059	0.221	0.603	0.032	0.124	0.334	0.107	0.191	0.049	0.118	0.230	0.200	0.190	0.143
Firmicutes_Bacilli_Lactobacillales_Streptococcaceae	0.095	0.146	0.029	0.040	0.098	0.010	0.038	0.087	0.023	0.069	0.197	0.000	0.044	0.104	0.038	0.068	0.000	0.045	0.080	0.055	0.063	0.049
Firmicutes_Clostridia_Clostridiales_Clostridiaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.001	0.004
Firmicutes_Clostridia_Clostridiales_Eubacteriaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.041	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.007	0.024
Firmicutes_Clostridia_Clostridiales_Incertae Sedis XIV	0.000	0.000	0.045	0.021	0.034	0.020	0.000	0.011	0.000	0.000	0.000	0.053	0.033	0.000	0.000	0.000	0.035	0.000	0.027	0.011	0.015	0.018
Firmicutes_Clostridia_Clostridiales_Lachnospiraceae	0.038	0.000	0.318	0.224	0.121	0.254	0.000	0.085	0.401	0.165	0.000	0.146	0.278	0.084	0.133	0.255	0.248	0.059	0.213	0.154	0.159	0.114
Firmicutes_Clostridia_Clostridiales_Peptostreptococcaceae	0.000	0.000	0.014	0.000	0.013	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.003	0.005
Firmicutes_Clostridia_Clostridiales_Ruminococcaceae	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.022	0.000	0.000	0.000	0.000	0.000	0.020	0.014	0.031	0.014	0.000	0.000	0.000	0.006	0.010
Firmicutes_Clostridia_Clostridiales_Veillonellaceae	0.013	0.018	0.000	0.000	0.000	0.000	0.058	0.011	0.000	0.015	0.012	0.000	0.000	0.000	0.000	0.000	0.011	0.014	0.000	0.000	0.008	0.014
Firmicutes_Erysipelotrichi_Erysipelotrichales_Erysipelotrichaceae	0.017	0.000	0.000	0.023	0.020	0.000	0.000	0.000	0.027	0.000	0.000	0.046	0.025	0.028	0.000	0.027	0.000	0.000	0.029	0.000	0.012	0.015
Fusobacteria_Fusobacteria_Fusobacteriales_Fusobacteriaceae	0.000	0.099	0.000	0.000	0.000	0.000	0.000	0.000	0.348	0.032	0.000	0.000	0.000	0.000	0.000	0.199	0.000	0.000	0.000	0.000	0.034	0.088
Proteobacteria_Alphaproteobacteria_Rhizobiales_Methylocystaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Proteobacteria_Betaproteobacteria_Burkholderiales_Alcaligenaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.000	0.027	0.000	0.000	0.003	0.008
Proteobacteria_Gammaproteobacteria_Enterobacteriales_Enterobacteriaceae	0.145	0.045	0.018	0.025	0.331	0.029	0.030	0.025	0.000	0.022	0.040	0.037	0.026	0.029	0.014	0.051	0.000	0.026	0.080	0.469	0.072	0.118
Proteobacteria_Gammaproteobacteria_Pasteurellales_Pasteurellaceae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.007
Proteobacteria_Gammaproteobacteria_Pseudomonadales_Moraxellaceae	0.042	0.048	0.000	0.011	0.021	0.000	0.015	0.037	0.000	0.023	0.079	0.000	0.017	0.044	0.015	0.023	0.000	0.000	0.023	0.018	0.021	0.021
Unclassified	0.072	0.076	0.061	0.061	0.055	0.043	0.021	0.062	0.039	0.092	0.057	0.061	0.047	0.064	0.048	0.057	0.061	0.054	0.072	0.058	0.058	0.015

GENUS	GR5092	GR5093	GR5138	GR5141	GR5267	GR5581	GR5583	GR5611	GR5615	GR5681	GR5683	GR5691	GR5785	GR5788	GR5791	GR5817	GR5862	**Average**	**Standard Deviation**
Relapse Patients																			
Alcaligenaceae_Parasutterella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Alcaligenaceae_Sutterella	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004
Alcaligenaceae_unknown-Sutterella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Bacteroidaceae_Bacteroides	0.255	0.590	0.305	0.144	0.270	0.300	0.000	0.506	0.000	0.380	0.046	0.227	0.148	0.194	0.000	0.021	0.000	0.199	0.182
Bacteroidaceae_unknown-Bacteroides	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
Bifidobacteriaceae_Bifidobacterium	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.016
Clostridiaceae_Clostridium	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.015	0.000	0.000	0.000	0.002	0.006
Clostridiaceae_unknown-Clostridium	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.000	0.001	0.004
Clostridiaceae_unknown-Thermotalea	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Coriobacteriaceae_Collinsella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Enterobacteriaceae_Citrobacter	0.015	0.000	0.000	0.012	0.012	0.000	0.010	0.000	0.000	0.000	0.012	0.010	0.000	0.000	0.024	0.011	0.000	0.006	0.007
Enterobacteriaceae_Escherichia/Shigella	0.015	0.000	0.000	0.000	0.000	0.139	0.094	0.000	0.109	0.000	0.018	0.043	0.172	0.025	0.000	0.022	0.000	0.037	0.055
Enterobacteriaceae_Raoultella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Enterobacteriaceae_unknown-Citrobacter	0.000	0.000	0.000	0.000	0.000	0.026	0.014	0.000	0.000	0.000	0.012	0.000	0.018	0.000	0.000	0.000	0.000	0.004	0.008
Enterobacteriaceae_unknown-Enterobacter	0.000	0.000	0.000	0.000	0.000	0.028	0.000	0.000	0.010	0.000	0.064	0.000	0.000	0.015	0.000	0.000	0.000	0.007	0.017
Enterobacteriaceae_unknown-Erwinia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.001	0.002
Enterobacteriaceae_unknown-Escherichia/Shigella	0.000	0.000	0.000	0.000	0.000	0.125	0.065	0.000	0.056	0.000	0.042	0.000	0.110	0.055	0.000	0.000	0.000	0.027	0.042
Enterobacteriaceae_unknown-Klebsiella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.056	0.000	0.000	0.000	0.000	0.000	0.003	0.014
Enterobacteriaceae_unknown-Raoultella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.001	0.004
Enterobacteriaceae_unknown-Salmonella	0.000	0.000	0.000	0.000	0.000	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.006
Enterococcaceae_Enterococcus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.001	0.004
Erysipelotrichaceae_Coprobacillus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.001	0.003
Erysipelotrichaceae_Turicibacter	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.001	0.004
Erysipelotrichaceae_unknown-Coprobacillus	0.000	0.000	0.000	0.016	0.000	0.066	0.021	0.031	0.000	0.000	0.053	0.000	0.000	0.069	0.000	0.016	0.100	0.022	0.031
Eubacteriaceae_Eubacterium	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
Fusobacteriaceae_Fusobacterium	0.025	0.108	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.075	0.000	0.000	0.000	0.012	0.031
Fusobacteriaceae_unknown-Cetobacterium	0.000	0.000	0.000	0.055	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.013
Fusobacteriaceae_unknown-Fusobacterium	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fusobacteriaceae_unknown-Ilyobacter	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Incertae Sedis XIV_Blautia	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.002	0.005
Incertae Sedis XIV_unknown-Blautia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
Lachnospiraceae_Anaerostipes	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lachnospiraceae_Coprococcus	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
Lachnospiraceae_Dorea	0.000	0.000	0.012	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004
Lachnospiraceae_Roseburia	0.010	0.000	0.018	0.020	0.011	0.000	0.050	0.019	0.000	0.000	0.019	0.000	0.000	0.034	0.000	0.000	0.029	0.012	0.015
Lachnospiraceae_unknown-Coprococcus	0.000	0.000	0.000	0.056	0.030	0.000	0.050	0.028	0.000	0.000	0.045	0.039	0.085	0.088	0.000	0.000	0.065	0.029	0.032
Lachnospiraceae_unknown-Dorea	0.000	0.000	0.036	0.000	0.056	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.008	0.017
Lachnospiraceae_unknown-Hespellia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.001	0.004
Lachnospiraceae_unknown-Lachnobacterium	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.014	0.000	0.011	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.003	0.006
Lachnospiraceae_unknown-Robinsoniella	0.000	0.000	0.000	0.086	0.038	0.000	0.030	0.013	0.000	0.000	0.018	0.000	0.010	0.020	0.000	0.000	0.028	0.014	0.022
Lachnospiraceae_unknown-Roseburia	0.013	0.107	0.038	0.079	0.053	0.014	0.149	0.082	0.000	0.017	0.095	0.038	0.035	0.099	0.000	0.070	0.153	0.061	0.048
Lachnospiraceae_unknown-Syntrophococcus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.074	0.004	0.018
Lactobacillaceae_Lactobacillus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.255	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.062
Lactobacillaceae_unknown-Lactobacillus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
Leuconostocaceae_Leuconostoc	0.178	0.014	0.093	0.111	0.149	0.066	0.128	0.054	0.146	0.119	0.158	0.137	0.081	0.040	0.244	0.231	0.105	0.121	0.062
Leuconostocaceae_Weissella	0.181	0.020	0.102	0.117	0.144	0.079	0.134	0.052	0.112	0.094	0.153	0.152	0.074	0.040	0.269	0.220	0.140	0.123	0.064
Methylocystaceae_unknown-Terasakiella	0.000	0.022	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.005
Micrococcaceae_Rothia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Moraxellaceae_Acinetobacter	0.043	0.000	0.029	0.024	0.031	0.012	0.023	0.000	0.024	0.019	0.025	0.028	0.014	0.000	0.062	0.052	0.028	0.024	0.017
Moraxellaceae_unknown-Acinetobacter	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.001	0.002
Pasteurellaceae_unknown-Haemophilus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Peptostreptococcaceae_unknown-Peptostreptococcus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.000	0.000	0.001	0.006
Peptostreptococcaceae_unknown-Sporacetigenium	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.015	0.044	0.029	0.000	0.000	0.006	0.013
Porphyromonadaceae_Barnesiella	0.000	0.000	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
Porphyromonadaceae_Odoribacter	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004
Porphyromonadaceae_Parabacteroides	0.000	0.000	0.024	0.000	0.000	0.000	0.000	0.051	0.000	0.000	0.000	0.000	0.075	0.000	0.000	0.000	0.000	0.009	0.022
Prevotellaceae_Prevotella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Prevotellaceae_unknown-Paraprevotella	0.000	0.000	0.052	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.013
Rikenellaceae_Alistipes	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ruminococcaceae_Faecalibacterium	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.005
Ruminococcaceae_Subdoligranulum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.079	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.019
Ruminococcaceae_unknown-Lactonifactor	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.001	0.002
Ruminococcaceae_unknown-Ruminococcus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Streptococcaceae_Lactococcus	0.094	0.000	0.046	0.063	0.073	0.040	0.077	0.023	0.071	0.049	0.081	0.071	0.037	0.024	0.140	0.124	0.062	0.063	0.036
Streptococcaceae_Streptococcus	0.012	0.000	0.012	0.017	0.000	0.000	0.000	0.000	0.000	0.012	0.020	0.000	0.000	0.000	0.018	0.019	0.000	0.006	0.008
Streptococcaceae_unknown-Lactococcus	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.017	0.013	0.000	0.003	0.006
Veillonellaceae_Dialister	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
Veillonellaceae_Megamonas	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Veillonellaceae_Phascolarctobacterium	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
Veillonellaceae_Veillonella	0.000	0.000	0.000	0.000	0.000	0.000	0.028	0.000	0.045	0.000	0.000	0.000	0.000	0.021	0.014	0.011	0.000	0.007	0.013

Remission Patients

RDP Family_Genus	GR5266	GR5384	GR5407	GR5455	GR5592	GR5593	GR5601	GR5603	GR5609	GR5625	GR5630	GR5781	GR5786	GR5793	GR5796	GR5811	GR5814	GR5825	GR5838	GR5849	**Average**	**Standard Deviation**
Incertae Sedis XIV_Blautia	0.000	0.000	0.036	0.014	0.032	0.013	0.000	0.000	0.000	0.000	0.000	0.045	0.016	0.000	0.000	0.000	0.026	0.000	0.018	0.000	0.010	0.014
Lachnospiraceae_unknown-Dorea	0.000	0.000	0.042	0.059	0.000	0.047	0.000	0.038	0.000	0.028	0.000	0.025	0.090	0.000	0.000	0.062	0.043	0.000	0.018	0.000	0.023	0.027
Bacteroidaceae_Bacteroides	0.291	0.146	0.380	0.487	0.126	0.552	0.192	0.259	0.072	0.293	0.000	0.557	0.380	0.280	0.626	0.097	0.421	0.593	0.236	0.000	0.299	0.196
Erysipelotrichaceae_unknown-Coprobacillus	0.015	0.000	0.000	0.020	0.017	0.000	0.000	0.000	0.025	0.000	0.000	0.039	0.022	0.024	0.000	0.025	0.000	0.000	0.024	0.000	0.010	0.013
Enterobacteriaceae_unknown-Citrobacter	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
Lachnospiraceae_unknown-Coprococcus	0.000	0.000	0.020	0.027	0.017	0.037	0.000	0.000	0.000	0.062	0.000	0.000	0.072	0.000	0.000	0.027	0.000	0.000	0.017	0.024	0.015	0.021
Clostridiaceae_unknown-Clostridium	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Clostridiaceae_Clostridium	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Peptostreptococcaceae_unknown-Peptostreptococcus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Peptostreptococcaceae_unknown-Sporacetigenium	0.000	0.000	0.013	0.000	0.011	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.002	0.005
Lachnospiraceae_unknown-Lachnobacterium	0.000	0.000	0.053	0.054	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.000	0.022	0.000	0.010	0.000	0.000	0.000	0.009	0.017
Rikenellaceae_Alistipes	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.000	0.002	0.005
Enterobacteriaceae_unknown-Escherichia/Shigella	0.045	0.000	0.000	0.000	0.088	0.000	0.011	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.014	0.000	0.000	0.021	0.033	0.011	0.022
Enterobacteriaceae_unknown-Enterobacter	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.001	0.005
Lachnospiraceae_unknown-Robinsoniella	0.000	0.000	0.086	0.010	0.000	0.027	0.000	0.011	0.278	0.000	0.000	0.022	0.032	0.037	0.000	0.016	0.041	0.000	0.051	0.065	0.034	0.063
Enterobacteriaceae_Citrobacter	0.015	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.007
Lachnospiraceae_Anaerostipes	0.000	0.000	0.000	0.000	0.027	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.002	0.006
Eubacteriaceae_Eubacterium	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.040	0.000	0.000	0.000	0.000	0.000	0.000	0.096	0.000	0.000	0.000	0.007	0.023
Moraxellaceae_Acinetobacter	0.034	0.041	0.000	0.000	0.017	0.000	0.012	0.030	0.000	0.018	0.062	0.000	0.012	0.040	0.013	0.018	0.000	0.000	0.022	0.014	0.017	0.017
Lachnospiraceae_unknown-Roseburia	0.032	0.000	0.065	0.023	0.058	0.083	0.000	0.012	0.075	0.033	0.000	0.067	0.037	0.023	0.082	0.090	0.078	0.029	0.085	0.030	0.045	0.031
Enterobacteriaceae_unknown-Erwinia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Moraxellaceae_unknown-Acinetobacter	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Bacteroidaceae_unknown-Bacteroides	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lachnospiraceae_Coprococcus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lactobacillaceae_unknown-Lactobacillus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Erysipelotrichaceae_Coprobacillus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Veillonellaceae_Dialister	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Veillonellaceae_Phascolarctobacterium	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Porphyromonadaceae_Odoribacter	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Erysipelotrichaceae_Turicibacter	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Methylocystaceae_unknown-Terasakiella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Enterobacteriaceae_unknown-Salmonella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fusobacteriaceae_unknown-Ilyobacter	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ruminococcaceae_Subdoligranulum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Streptococcaceae_unknown-Lactococcus	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004
Leuconostocaceae_Weissella	0.144	0.203	0.048	0.057	0.082	0.023	0.056	0.123	0.026	0.116	0.310	0.013	0.055	0.169	0.044	0.097	0.022	0.057	0.110	0.104	0.093	0.073
Enterobacteriaceae_Escherichia/Shigella	0.058	0.000	0.000	0.000	0.214	0.016	0.000	0.000	0.000	0.000	0.000	0.015	0.011	0.000	0.000	0.020	0.000	0.000	0.037	0.013	0.019	0.048
Micrococcaceae_Rothia	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002
Alcaligenaceae_Parasutterella	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002
Veillonellaceae_Megamonas	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002
Clostridiaceae_unknown-Thermotalea	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.001	0.004
Enterobacteriaceae_Raoultella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.001	0.004
Alcaligenaceae_unknown-Sutterella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.005
Ruminococcaceae_unknown-Ruminococcus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.022	0.000	0.000	0.000	0.000	0.001	0.005
Pasteurellaceae_unknown-Haemophilus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.006
Coriobacteriaceae_Collinsella	0.000	0.000	0.000	0.000	0.000	0.000	0.104	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.023
Fusobacteriaceae_unknown-Fusobacterium	0.000	0.014	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.007
Prevotellaceae_Prevotella	0.000	0.000	0.000	0.000	0.000	0.000	0.258	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.058
Leuconostocaceae_Leuconostoc	0.137	0.212	0.054	0.045	0.084	0.020	0.062	0.165	0.032	0.103	0.284	0.018	0.068	0.163	0.060	0.091	0.026	0.060	0.118	0.095	0.095	0.069
Lactobacillaceae_Lactobacillus	0.000	0.000	0.000	0.000	0.000	0.000	0.058	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.013
Streptococcaceae_Lactococcus	0.073	0.122	0.024	0.027	0.042	0.000	0.028	0.073	0.013	0.051	0.174	0.000	0.032	0.089	0.033	0.058	0.000	0.035	0.064	0.043	0.049	0.043
Enterobacteriaceae_unknown-Raoultella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.117	0.006	0.026
Lachnospiraceae_Dorea	0.000	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.002	0.006
Enterobacteriaceae_unknown-Klebsiella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.255	0.013	0.057
Alcaligenaceae_Sutterella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.026	0.000	0.000	0.001	0.006
Lachnospiraceae_unknown-Syntrophococcus	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.002	0.005
Enterococcaceae_Enterococcus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.059	0.000	0.000	0.003	0.013
Porphyromonadaceae_Barnesiella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.042	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.009
Veillonellaceae_Veillonella	0.013	0.000	0.000	0.000	0.000	0.000	0.056	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.004	0.013
Lachnospiraceae_Roseburia	0.000	0.000	0.015	0.000	0.000	0.019	0.000	0.000	0.029	0.020	0.000	0.014	0.000	0.000	0.011	0.038	0.013	0.017	0.021	0.000	0.010	0.012
Lachnospiraceae_unknown-Hespellia	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.001	0.004
Fusobacteriaceae_Fusobacterium	0.000	0.073	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000	0.000	0.198	0.000	0.000	0.000	0.000	0.015	0.046
Prevotellaceae_unknown-Paraprevotella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.042	0.000	0.000	0.000	0.002	0.009
Incertae Sedis XIV_unknown-Blautia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004
Fusobacteriaceae_unknown-Cetobacterium	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.309	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.069
Ruminococcaceae_Faecalibacterium	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004
Porphyromonadaceae_Parabacteroides	0.000	0.000	0.025	0.000	0.000	0.020	0.016	0.039	0.000	0.023	0.000	0.061	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.017
Streptococcaceae_Streptococcus	0.013	0.010	0.000	0.010	0.049	0.000	0.000	0.010	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.006	0.011
Bifidobacteriaceae_Bifidobacterium	0.000	0.000	0.000	0.000	0.000	0.000	0.060	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.014
Ruminococcaceae_unknown-Lactonifactor	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.001	0.002

5. Questions from the CEGIIR environmental questionnaire

1.6. Date of Birth

1.7. Gender

1.8. Number of Siblings

1.9. Postal Code

1.11. Which genetic (blood line or biological ancestry) population group(s) best describe you?

Family History- Disease

Has anyone in your immediate (birth, biological, related) family ever had the following disease(s)?

- ***Bowel Diseases* –** Celiac Disease, Crohn’s Disease, Irritable Bowel Disease, Ulcerative Colitis, Ulcer Disease, Unknown, Other.
- ***AutoImmune Diseases* –** Asthma, Grave Disease, Hypothyroidism, Juvenile Diabetes, Lupus Erythematosus, Multiple Sclerosis, Psoriasis, Rheumatoid Arthritis, Sjogren Syndrome, Unknown, Other.
- ***Liver Diseases*** - Alcohol Cirrhosis, Autoimmune Hepatitis, Fatty Liver NASH, Hemochromatosis, Hepatitis B, Hepatitis C, Primary Biliary Cirrhosis (PBC), Primary Sclerosing Cholangitis (PSC) Wilson Disease, Alcoholism, Psychiatric Condition (such as Depression**,** Anxiety Disorder, or Bipolar Disorder) Unknown, Other.

Smoking History

3.1a. Do you currently use any non-pharmaceutical tobacco products other than cigarette smoking? (e.g. pipe, smokeless tobacco products)

3.1b. In a typical week, during how much time, according to your best estimate, are you in the presence of other people smoking cigarettes or otherwise exposed to cigarette smoke from other people?

Have you ever smoked cigarettes regularly? (No means less than 20 packs of cigarettes or 12oz of tobacco in a lifetime or less than 1 cigarette a day for 1 year)

If yes, answer questions a-e:

a. Do you now smoke cigarettes (as of 1 month ago)?
b. On the average of the entire time you smoked, how many cigarettes did you smoke per day?
c. How old were you when you first started regular cigarette smoking?
d. If you have stopped smoking cigarettes completely, how old were you when you stopped?
e. When you were smoking, did you ever stop smoking for >6 months

If yes, answer question i-ii:

i. For how many years in total did you stop smoking cigarettes?
ii. Did you stop smoking within 6 months of being diagnosed with IBD?

3.2. Have you ever smoked a pipe regularly? (Yes means more than 12oz of tobacco in a lifetime.)

If yes, answer question a-e:

a. Do you now smoke a pipe (as of 1 month ago)?
b. On the average of the entire time you smoked a pipe how much pipe tobacco did you smoke per week?
c. How old were you when you first started to smoke a pipe?
d. If you have stopped smoking a pipe completely, how old were you when you stopped?
e. When you were smoking a pipe, did you ever stop smoking for >6 months?

If yes, answer question i-ii:

i. For how many years in total did you stop smoking a pipe?
ii. Did you stop smoking a pipe within 6 months of being diagnosed with IBD

3.3. Have you ever smoked cigars regularly? (Yes means more than 1 cigar/week for a year)

If yes, answer question a-e:

a. Do you now smoke cigars (as of 1 month ago)?
b. On the average of the entire time you smoked cigars, how many cigars did you smoke per week?
c. How old were you when you first started to smoke cigars regularly?
d. If you have stopped smoking cigars completely, how old were you when you stopped?
e. When you were smoking cigars, did you ever stop smoking for >6 months?

If yes, answer question i-ii:

i. For how many years in total did you stop smoking cigars?
ii. Did you stop smoking cigars within 6 months of being diagnosed with IBD?

3.4. Have you ever used smokeless tobacco (chew, snuff, snus and other tobacco products used in the mouth or nose without smoking) regularly? (Yes means more than once/day for a year)

If yes, answer question a-e:

a. Do you now use chewing tobacco (as of 1 month ago)?
b. On the average of the entire time you use chewing tobacco, how many chews did you use per day?
c. How old were you when you first started to use chewing tobacco regularly?
d. If you have stopped using chewing tobacco completely, how old were you when you stopped?
e. When you were using chewing tobacco, did you ever stop using for >6 months?

If yes, answer question i-ii:

i. For how many years in total did you stop using chewing tobacco? |__|__|
ii. Did you stop using chewing tobacco within 6 months of being diagnosed with IBD? Yes No

3.5. Have you ever smoked marijuana regularly? (Yes means more than 1 cigarette/week for a year)

If yes, answer question a-e:

a. Do you now smoke marijuana (as of 1 month ago)?
b. On the average of the entire time you smoked marijuana, how many cigarettes did you smoke per week?
c. How old were you when you first started to smoke marijuana regularly?
d. If you have stopped smoking marijuana completely, how old were you when you stopped?
e. When you were smoking marijuana, did you ever stop smoking for >6 months?

If yes, answer question i-ii:

i. For how many years in total did you stop smoking marijuana?
ii. Did you stop smoking marijuana within 6 months of being diagnosed with IBD?

3.6. As an adult, now or in the past, have you ever lived with a regular cigarette smoker who smoked in your home?

If yes, answer question a-b:

a. Spouse or Partner?
Years of exposure?
b. Others in household?
Years of exposure?
How many others?

3.7. When you are not at home, do you regularly spend time indoors where there are people smoking cigarettes?

If yes, answer question a-b:

a. At Work?
   Years of exposure?
b. Other than work?
   Years of exposure?

Environmental History

4.1. As a child (age 2-12), did you have any pets/animals you cared for or handled at home?

4.2. Currently, do you have any pets/animals you care for or handle at home/work

4.3. Did you ever live on a farm – or have regular exposure to a farm?

4.4. What type of farm was this?

Medication History

5.1. Have you ever taken Birth Control Pills for longer than 3 months?

If yes, between what ages?

5.2. During the past week, have you taken any medications not prescribed by a doctor (e.g. over the Counter)? (e.g. for fever, occasional headache, congestion, cough, allergy, stomach upset, indigestion, heartburn, body aches/pains/discomforts, skin problems)

If yes, please select from below the medication(s) you are currently taking:

Dyspepsia: Gaviscon, Maalox, Mylanta, Pepid AC, Zantac 75, Tums or Rolaids, Other.

Constipation: Chronulac, Colace, Dulcolax, Fleet enema, Glycerin supp, Metamucil, Milk of Magnesia, Prodiem, Senokot, Senokot – S, Other.

Diarrhea: Imodium, Lomotil, Pepto-Bismol, Other.

Pain, Discomfort, and/or fever*:* Acephen, Acetaminophen/Tylenol/Anacin, Acetylsalicylic Acid/ASA/Aspirin/Ecotrin, Aleve, Anacin, Endocet, Excedrin, FeverAll, Gelpirin, Genapap, Genebs. Goody's, Ibuprofen – Advil, Motrin, Liquiprin, Panadol, Percocet, Roxice, Supac, Tempra, Tylenol (any type), Tylox, Vanquish, Wygesic, Other.

***Vitamin Supplements on a weekly basis greater than 3 times per week**:* Multi-Vitamins, Vitamin A, Vitamin B, Vitamin C, Vitamin D, Vitamin E, Iron, Calcium, Fish oil supplements, Probiotic supplements, Other.

5.3. Have you ever taken aspirin at least twice per week regularly for longer than 3 months?

If yes, between what ages?

5.4. Have you ever used Arthritis drugs? At least twice per week regularly for longer than 3 months?

If yes, between what ages?

5.4. Continue... Have you ever used one of the following drugs for at least twice per week for longer than 3 months (Used in the Past or Currently)? Please select the drugs below (Please indicate all that apply).

Aspirin, Motrin or Advil (ibuprofen), Naprosyn (naproxen), Celebrex (celecoxib), Voltaren (diclofenac), Ultradol (etodolac), Indocid (indomethacin), Toradol (ketorolac), Clinoril (sulindac), Idarac (floctafenine), Ponstan (mefenamic acid), Relafen (nabumetone), Pirox (piroxicam), Mobiflex (tenoxicam), Metacam (meloxicam), Froben (flurbiprofen), Rhodis (ketoprofen), Daypro (oxaprozin), Surgam (tiaprofenic acid), Dolobid (diflunisal)

5.5. Have you ever used herbal or Alternative Therapy drugs?

If yes, How Many?

How Often?

5.6. Have you ever been told you have Crohn's Disease?

If yes, what medications have you ever used for Crohn's Disease treatment (Used in the Past or Currently)?

5-ASA: Mesasal, Olsalazine Sodium (Dipentum), Pentasa, Salofalk, Salofalk enema, Sulfasalazine (Azulfidine), Other

***Antibiotics**:* Ciprofloxacin (Cipro), Metronidazole (Flagyl) , Other

***Corticosteroids**:*

***Budesonide** (*Entocort): Entocort Capsules, Entocort Enema,
Cortenema, Methylprednisolone (Medrol),
Prednisone, Other: ,

***Immuno-modulators**:* 6-MP (Purinethanol), Azathioprine (Imuran),
Cyclosporine (Neoral), Methotrexate (Rheumatrex,
Trexall), Mycophenolate Mofetil (Cellcept), Tacrolimus
(Prograf), Other.

***Biologics**:* Adalimumab (Humira), Certolizumab (Cimzia), Infliximab
(Remicade), Other:

***Other than all of above**. Please specify*

N/A; I have NOT taken any medications for my Crohn's Disease treatment

5.7. Have you ever been told you have Colitis?

If yes, what medications have you ever used for Colitis treatment (Used in the Past or Currently)?

5-ASA: Asacol, Mesasal, Olsalazine Sodium (Dipentum), Pentasa,
Salofalk, Salofalk enema, Sulfasalazine (Azulfidine), Other.

Antibiotics: Ciprofloxacin (Cipro), Metronidazole (Flagyl), Other.

Corticosteroids:

Budesonide (Entocort) - (Entocort Capsules or Entocort Enema),
Cortenema, Methylprednisolone (Medrol), Prednisone, Other

Immuno-modulators: 6-MP (Purinethanol), Azathioprine (Imuran),
Cyclosporine (Neoral), Methotrexate (Rheumatrex,
Trexall), Mycophenolate Mofetil (Cellcept), Tacrolimus
(Prograf), Other.

Biologics: Adalimumab (Humira), Certolizumab (Cimzia), Infliximab (Remicade), Other:

Other than all of above. Please specify:

5.8. Have you ever been told that you have Liver Disease?

If yes, what medications have you used for Liver Disease Treatment (Used in the Past or Currently)?

Adefovir (Hepsera), Amiloride (Midamor), Baraclude (Entecavir), Ciprofloxacin, Cyclosporine (Neoral), Darbopoetin (Aranesp, Amgen)

G-CSF/Granulocyte Colony-Stimulating Factors: lenograstim (Granocyte), Filgrastim (Neupogen), Pegylated Filgrastim (Neulasta)

Interferon: alpha 2a (Roferon-A), alpha 2b (Intron), beta 1a (Avonex, Rebif), beta 1b (Bataseron)

Interferon and ribavirin: Interferon alfa 2a + Ribavirin (Roferon A + Ribavirin), Interferon alfa 2b + Ribavirin (Intron A + Rebetol)

Imuran (Azathioprine), Lactulose, Lamivudine (Heptovir, 3TC), Lasix (Furosemide), Metronidiazole (Flagyl), Mycophenolate Mofetil (Cellcept), Nadolol (Corgard), Naloxone (Narcan), Pegylated interferon(alfa 2a (Pegasys), or alfa 2b (PEG-Intron)), Pegylated interferon and ribavirin (Peginterferon alfa 2a + Ribavirin (Pegasys and Copegus) or Peginterferon alfa 2b + Ribavirin (PEG-Intron and Rebetol)), Prednisone, Propranolol (Inderal, Avlocardyl), Questran (Cholestyramine), Rifampin (Rifadin, Rimactin), Septra, Sirolimus (Rapamune), Spironolactone (Aldactone), Tacrolimus (Prograf), Tenofovir (Viread), Tyzeka (Telbivudine), Ursodiol (Urso)

Other than above

History of Consuming Alcoholic Beverages

5.9a. Have you ever consumed alcoholic beverages (beer, wine, or liquor)?

5.9b. If yes at what age did you start drinking alcoholic beverages?

5.9c. If yes when did you last have an alcoholic beverage?

5.9d. If yes, please fill in below the most recent 3 periods and the average number of drinks per week, where 1 drink is equal to a 4 ounce glass of wine, 1 ounce of hard liquor or a 12 ounce beer.

5.9e. Have you ever attended alcohol rehabilitation or a detox center?

Diet

7.1. Were you breast fed as an infant

7.2. As a child (ages 2-12), did you regularly drink water from non-tap sources such as streams, lakes, barrels, at recreational locations (e.g. on vacations, camping, sports facilities, etc.)?

7.3. Currently, do you regularly drink water from non-tap sources such as streams, lakes, barrels, at recreational locations (e.g. on vacations, camping, sport facilities, etc

7.4. As a child (ages 2-12), how often did you drink unpasteurized cow's milk?

7.5. Currently, how often do you normally drink unpasteurized cow's milk?

7.6. As a child (ages 2-12), how often did you normally consume unpasteurized dairy products*?

7.7. Currently, how often do you normally consume unpasteurized dairy products*?

7.8. As a child (ages 2-12), how often did you usually have sugary foods (baked goods, candy,

chocolate) or drinks (pop, iced capucccino, iced tea, lemonade, fruit juices, fruit cocktail drinks)?

7.9. Currently, how often do you usually have sugary foods (baked goods, candy, chocolate) or

drinks (pop, iced cappucino, iced tea, lemonade, fruit juices, cocktail drinks - with or without alcohol)?

7.10. As a child (ages 2-12), how often did you usually eat "high fiber" foods? For example: bran,

oats, vegetables (e.g. peas, broccoli, beans), fruits, nuts, or fiber supplements (e.g. Metamucil)?

7.11. Currently, how often do you usually eat "high fiber" foods? For example: bran, oats, vegetables

(e.g. peas, broccoli, beans), fruits, nuts, or fiber supplements (e.g. Metatmucil)?

7.12. As a child (ages 2-12), how often did you usually eat fish, seafood or fish oil (omega-3 and/or cod liver oil) supplements?

7.13. Currently, how often do you usually eat fish, seafood or fish oil (omega-3 and/or cod liver oil) supplements?