Effects of Therapy on Host-Microbe Interactions in Pediatric Inflammatory Bowel Diseases

by

Stephanie Dijk

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Physiology

University of Alberta

Abstract

Introduction

By 2030, 1/100 Canadians are predicted to be diagnosed with inflammatory bowel diseases (IBD); approximately 25% of cases are diagnosed in children. IBD include Crohn disease (CD) and ulcerative colitis (UC), and manifest with chronic gastrointestinal inflammation. Pathogenesis remains poorly understood but IBD are associated with alterations in microbial recognition and clearance, an altered intestinal microbiome, and certain dietary exposures. Treatments include immune suppression and dietary therapy. Delineating how the intestinal microbiome is associated with treatment outcomes is critical in understanding disease pathogenesis.

Methods

Multi-centre, prospective IFX studies in 33 paediatric Crohn disease (pCD) and 32 pediatric ulcerative colitis (pUC) patients were conducted with bio-specimen collection before and after treatment. Stool was analyzed for short chain fatty acids (SCFA), fecal calprotectin (FCP), and microbial composition; urine metabolomics was performed. In addition, a cohort of 103 prospectively followed pCD patients initiating dietary therapy completed food frequency questionnaires (FFQs) at the time of diagnosis that were analyzed for dietary patterns using principal component analysis.

Results

IFX treatment was associated with increased SCFA ($p=0.01$), altered tryptophan and *Clostridia*-associated metabolites ($p<0.05$), and increased SCFA producing bacteria, such as *Pseudobutyrivibrio* ($p=0.07$), *Akkermansia* ($p=0.004$), and *Ruminococcus* ($p=0.07$). FCP and

disease severity decreased with treatment ($p<0.0001$ and <0.001, respectively), while microbial diversity and richness increased ($p<0.022$). In pUC, SCFA decreased in males but showed an average increase in females ($p=0.0095$), with similar sex-dependent associations and changes with microbial diversity and richness ($p<0.05$). Disease severity and location in pCD were also associated with changes in SCFA, metabolites, and microbiota.

The top four dietary patterns named after foods with high factor loadings were: "Vegetarian", "Meat", "Pre-packaged", and "Mature". The patterns characterized 28.9% of dietary variability in the pCD EEN-FFQ cohort, and varied with age, sex, and disease location.

Conclusion

Increases in SCFA, tryptophan and *Clostridia* metabolites, as well as SCFA-producing bacteria with therapy may suggest beneficial alterations in the microbiome as an important part of treatment response. Differential responses between males and females suggest sex-specific mechanisms may be present. Sex-specific variations in the dietary therapy cohort may identify a potential mechanism. Further investigation is required to examine the causal associations between microbial metabolites and host health in pediatric IBD. These findings may contribute to understanding disease etiology and development of novel biomarkers, microbial therapeutic targets, or dietary guidelines.

Preface

Informed consent was obtained by research coordinators from parents and/or patients as applicable, and participants were free to withdraw from the studies listed in this thesis at any time. The IDeaL and EEN-FFQ studies utilized the CIDsCaNN research infrastructure (cidscann.ca) and were approved by the University of Alberta Health Research Ethics Board (Pro00042980). PUCIFX was approved by the University of Calgary Health Research Ethics board (E-25172).

ImageKids (Appendix 1: ImageKids) received ethics approval in each of the participating centres internationally. ImageKids was supported financially by AbbVie (www.abbvie.com).

Acknowledgements

I would like to start by thanking my graduate supervisor, Dr. Eytan Wine; thank you for providing me with such an excellent graduate school environment, and for so many exciting opportunities to explore my interests. Your kind support, mentorship, and guidance have helped me learn so much about myself and encouraged me to pursue my goals. A special thanks to my husband, Lennart Dijk, who has always supported me through my academic and life endeavours – thank you for being my number one fan and believing in me. I also would like to thank my committee members Dr. Catherine Field and Dr. Karen Madsen for their advice and guidance throughout my degree. Your direction helped me to focus my efforts and gave me the tools to accomplish my goals. I am also personally thankful for the Physiology Department, including Dr. Gregory Funk, Dancy Bogdanovic, and Kim Sawada who were very helpful to me early on in organizing a leave. During a difficult time, thank you for your kindness and support. Thank you to all the past and present Wine lab member for your help and for making the lab environment what it is. A special thank you to Dr. Rosica Valcheva for your endless patience in teaching me lab techniques. Your honesty and trustworthiness kept me going in the right direction, and I will miss speaking with and learning from you while we work. Thank you to Naomi Hotte for always knowing where everything is and troubleshooting every technique under the sun. Without Dr. Terry Zhang & Dr. Juan Jovel I would still be fumbling in the statistical dark, so thank you for your kindness, humor, and patience in answering my endless questions. Thank you to Dr. Megan Jarman for being such a wonderful resource for dietary analysis. Thank you to the Canadian Institutes of Health Research (CIHR), the Women & Children's Health Research Institute (WCHRI) with the Stollery Children's Hospital Foundation, the Faculty of Medicine & Dentistry, and the Faculty of Graduate Studies and Research (FGSR) for funding me in my research. Finally, thank you to the rest of my friends and family for all the love, and supporting me no matter how crazy the endeavour. Thank you.

Table of Contents

List of Tables

List of Figures

List of Abbreviations (in alphabetical order)

5-ASA: Aminosalicylates

AIEC: Adherent Invasive *Escherichia coli*

ASCA: Anti-*Saccharomyces cerevisiae* antibody

ASV: Amplicon Sequence Variants

CARD: Caspase Recruitment Domain Family Member

CAC: Citric Acid Cycle

CCD: Colonic and Ileocolonic Crohn Disease

CD: Crohn Disease

CDAI: Crohn's Disease Activity Index

CDED: Crohn's Disease Exclusion Diet

CDI: *Clostridioides difficile* Infection

CIDsCaNN: Canadian Children IBD Network

CRP: C-Reactive Protein

DNA: Deoxyribonucleic Acid

DSS: Dextran Sulfate Sodium

EEN: Exclusive Enteral Nutrition

ELISA: Enzyme Linked Immunosorbent Assay

EPIC: Edmonton Pediatric IBD Clinic

ESR: Erythrocyte Sedimentation Rate

FFQ: Food Frequency Questionnaire

FCP: Fecal Calprotectin

GC: Gas Chromatography

GC-MS: Gas Chromatography – Mass Spectroscopy

GCS: Glucocorticoids

HDAC: Histone Deacetylase

HPHPA: (3-hydroxyphenyl)-3-hydroxypropionic acid

IBD: Inflammatory Bowel Disease

ICD: Ileal Crohn Disease

IL: Interleukin

IFX: Infliximab

QIIME2: Quantitative Insights Into Microbial Ecology 2

LPS: Lipopolysaccharide

LNA: Locked Nucleic Acids

LOD: Limit of Detection

MCT: Monocarboxylate Transporter

MMP: Matrix Metalloproteinases

MRE: Magnetic Resonance Enterography

NF-κB: Nuclear Factor – κB

NOD: Nucleotide-binding Oligomerization Domain

OmpC: Outer Membrane Porin C

OTU: Operational Taxonomic Unit

PCA: Principal Component Analysis

PCR: Polymerase Chain Reaction

pCD: Pediatric Crohn Disease

PCoA: Principal Coordinate Analysis

PERMANOVA: Permutational Analysis of Variance

PGA: Physician Global Assessment

pIBD: Pediatric Inflammatory Bowel Disease

PLS-DA: Partial Least Squares – Discriminant Analysis

pUC: Pediatric Crohn Disease

PUCAI: Pediatric Ulcerative Colitis Activity Index

QEA: Quantitative Enrichment Analysis

qPCR: Quantitative Polymerase Chain Reaction

RS: Resistant Starch

SCFA: Short Chain Fatty Acids

SES-CD: Simple Endoscopic Score for Crohn's disease

TAGC: The Applied Genomics Centre

TMIC: The Metabolomics Innovation Centre

TNF: Tumour Necrosis Factor

UC: Ulcerative Colitis

VIP: Variable Important in Projection

wPCDAI: Weighted Pediatric Crohn Disease Activity Index

Chapter 1. INTRODUCTION

1.1 Inflammatory Bowel Diseases

Inflammatory bowel diseases (IBD) are projected to affect 1 in 100 Canadians by the year 2030;[1] one of the highest rates worldwide. Approximately 25% of cases are diagnosed in children.[2] The two main subtypes of IBD include Crohn disease (CD) and ulcerative colitis (UC), while those with IBD-unclassified have mixed features. IBD consists of chronic intestinal inflammation, in the case of CD with skip lesions occurring anywhere along the gastrointestinal tract from mouth to anus, while in UC inflammation is restricted to the colon in a contiguous fashion. There is currently no cure, and the etiology is not completely understood; however, there is evidence that genetics, environment, immune response, and the intestinal microbiome interact to play a role in disease progression.[3] Unique to pediatric IBD (pIBD) is the growth and development component: failure to thrive is a common presentation that leads to an IBD diagnosis, and pIBD carries a risk of linear growth failure, delayed puberty, and reduced peak bone density.[4, 5] Children with IBD are also at an increased risk of psychosocial difficulties including depression and anxiety.[6]

Younger age of diagnosis in pediatric CD (pCD) is associated with more complicated disease.[7, 8] The Paris classification is used by clinicians to classify phenotypes of pIBD; complicated disease is defined as stricturing and/or penetrating disease in pCD, and is associated with poorer outcomes and increased need for surgery.[9] pCD patients are also more likely to have small bowel involvement than adults, and a younger age at diagnosis is linked with an increased need for surgical intervention.[7] Pediatric UC (pUC) patients are more likely to have extensive inflammation in the form of pancolitis.[10] Almost half of children diagnosed with pCD, and

roughly 16% of children diagnosed with pUC require surgery in the ten years following diagnosis,[11] and children with IBD have on average a lower health related quality of life.[12] In follow-up studies, diagnosis with IBD in childhood has been associated with increased risk of cancer and mortality in adulthood.[13] Therefore, there is clearly a need to better define causes and predictors of response to therapy in IBD, especially in children.

1.2 Tools to Assess IBD Activity

There are numerous methods to assess disease activity in pIBD, including: fecal calprotectin (FCP), clinical scores, and endoscopy. The most accurate, and invasive, is endoscopy to directly visualize disease activity and collect biopsies for microscopic evaluation. To minimize the need for endoscopy and its associated risks and challenges, non-invasive methods are used wherever possible to quantify disease activity after initial diagnosis. Measured in stool, FCP is better correlated with the simple endoscopic score in Crohn's Disease (SES-CD) than C-reactive protein (CRP) or the Crohn's disease activity index (CDAI) clinical score.[14] Calprotectin is a protein abundant in neutrophil granulocytes that displays antimicrobial properties.[15] FCP can differentiate between inactive disease, mild, moderate and high disease activity, as identified on endoscopy.[14] In children with IBD, FCP has been found to be a sensitive indicator of intestinal inflammation in both pUC and pCD.[16]

Clinical scores include the pediatric CDAI (PCDAI), the pediatric ulcerative colitis activity index (PUCAI), and the physician global assessment of disease activity (PGA). These scores allow clinicians to non-invasively quantify disease activity. Numerous versions of the PCDAI are available, with the weighted PCDAI (wPCDAI) providing the best correlation with

FCP; a score of ≤12.5 accurately and non-invasively identifies mucosal healing.[17] Components of the clinical scores are described below in the Methods section.

1.3 Etiology and Pathophysiology of IBD

While the precise etiology of IBD remains unclear, intestinal microbiome alterations, host immunity, environment, and genetic predispositions are suspected to contribute to an underlying dysfunctional host-microbe interaction.

1.3.1 Genetics

It is thought that genetics may play more of a role in pediatric vs adult IBD, but the majority of new cases are diagnosed without a family history,[18] suggesting that genetics plays a minor role in the pathogenesis of disease. Many of the IBD genetic associations are in genes involved in intestinal epithelial barrier integrity, microbial recognition and clearance, and regulation of host-immune response.[3, 19] The caspase recruitment domain family member 15 gene (*CARD15*), more commonly known as nucleotide-binding oligomerization domain two (NOD2), is the strongest genetic association with IBD, and is specifically correlated with CD.[20] NOD2 is a pattern recognition molecule in the cell cytosol that can recognize bacterial peptidoglycan.[21, 22] Individuals with the IBD associated NOD2 variants have decreased intestinal antimicrobial peptides (e.g. α-defensins)[23], and the most common IBD-associated variant has no response *in vitro* to bacterial peptidoglycan and lipopolysaccharide (LPS).[24] Intestinal epithelial cells with this variant have drastically reduced ability to prevent intracellular invasion with *Salmonella typhimirium* compared to wildtype,[25] and patients with these variants were also found to have

lower circulating levels of anti-TNF monoclonal antibody therapy (discussed below) between doses than patients with the wild-type gene during biologic therapy, which may result in reduced efficacy of treatment.[26]

There are also IBD-associated *JAK2* variants, which are linked to increased intestinal permeability in CD.[27] *JAK2* is involved in cytokine signalling, and the IBD associated variant has a gain of function mutation.[28] Inhibition of JAK signalling can decrease T cell mediated inflammation,[29] and JAK inhibition by the small molecule Tofacitinib has proven effective in some IBD populations and was developed for use in other immune-mediated diseases, such as rheumatoid arthritis and psoriasis.[30]

Another gene, *CARD9* has certain variants with increased representation in both CD and UC, while other more rare *CARD9* variants appear to be protective.[31, 32] *CARD9* encodes for a protein involved in downstream signalling for pattern recognition molecules that recognize bacterial and fungal antigens, such as NOD2.[33] Defective CARD9 is associated with impaired activation of innate immunity in response to intracellular pathogens,[34] while absence is associated with potentially severe immune-compromise.[32] The IBD-associated variant also correlates with increased nuclear factor-κB (NF-κB) signalling.[32] NF-κB signalling results in an increased production of proinflammatory cytokines, including tumour necrosis factor α (TNFα) and interleukin 6 (IL-6).[32, 35, 36]

Although a significant number of genetic associations have been identified in IBD, only 26% and 19% of the heritability of CD and UC, respectively, can be explained by genetics.[37] Genetic associations also cannot explain the recent rise in incidence of IBD,[1] highlighting the importance of assessing environmental factors as contributors to the pathophysiology of IBD.

1.3.2 Microbiome

The human gut microbiome is composed of bacteria, viruses, archaea, and eukaryotes, with approximately as many bacterial cells are there are human cells in the body.[38] Alterations in the gut microbiome are thought to play an important role in numerous human diseases. The microbiome is also critical to health, and is essential for healthy immune development and certain metabolic functions such as the fermentation of dietary fibers and generation of certain vitamins important in human health.[39, 40]

Alterations in the intestinal microbiota are a well-documented hallmark of IBD, with somewhat different alterations in CD vs UC, and more significant differences with more active disease, relative to controls.[41] Dysbiosis is a frequently used, but nonspecific term, that describes alterations from a variably defined "healthy" microbiome. The term dysbiosis has been used for a range of microbial shifts, from those associated with antibiotics to those associated with disease. Microbial changes found in pIBD include lower levels of mucosal bacteria species richness in pCD (measured by Chao1), reduced fecal microbiota diversity,[42] and alterations in certain mucosal and bacterial taxa, including higher numbers of *Enterobacteriaeceae* and lower abundances of Clostridiales.[43] Diversity measures, such as Shannon H and Simpson provide an index to quantify the number of different microbe present, their evenness of distribution, as well as the abundances of each. Richness measures, such as OTU numbers and Chao1 indicate the numbers of different microbes detected or predicted to exist in a given specimen. Although increased bacterial diversity has been often associated with health,[44] there are many instances where an inverse relationship is true, such as the lower microbial diversities seen in exclusively breastfed vs. formula or mixed fed infants.[45] Reduced levels of *Clostridium* species, *Faecalibacterium prausnitzii*, and *Bifidobacterium* species are seen in IBD, with relative

abundances reduced more so in CD than UC.[41] Reduced *Bacteroides* species are also seen in IBD, with further reductions in active disease vs remission.[46] Gevers *et al*. (2014) found significant differences between pCD and controls that were more numerous in mucosal tissue samples than stool, including some changes in mucosal microbiota that weren't reflected in stool microbiota.[43] Microbial dysbiosis has been found to correlate strongly with disease severity, as measured by the pediatric Crohn disease activity index (PCDAI).[43] Levels of *Enterobacteriaceae*, *Fusobacterium*, and *Haemophilus*, along with age of disease onset and PCDAI at diagnosis have been found to be significantly predictive of future PCDAI.[43] Disease location is also associated with variations in microbial abundance: patients with predominantly ileal CD (ICD) had lower abundances of *F. prausnitzii* and higher levels of *Escherichia coli* in intestinal mucosa specimens compared to patients with predominantly colonic/ileocolonic CD (CCD).[47]

Although UC has shown less microbial distinction from controls than CD,[48] Shah *et al*. (2015) found that patient response to initial therapy was associated with differential abundances of certain microbial taxa.[49] Shah *et al.* also found *Clostridium* and *Bacteroides* OTUs were only detectable at baseline in the UC patients who went on to have no clinical response to treatment.[49] In severe pUC, phylum level changes compared to controls included lower levels of Firmicutes and higher levels of Proteobacteria, with lower microbial richness in pUC that was more severe in the pUC children who failed to respond to glucocorticoid therapy.[50] Additionally, fluctuations in the UC microbiome are associated with disease severity and need for colectomy.[51] A more severe depletion of Ruminococcaeae and Lachnospiraceae with an increased abundance of *Streptococcus anginosus* were found in patients with more severe disease, and longitudinal

follow-up linked increased microbiota variability with an increased likelihood of colectomy in pUC.[51]

Impaired barrier integrity is another feature of IBD. pUC patients have shown a reduction in mucous barrier thickness, and fewer mucin-containing goblet cells, with bacteria colonizing closer to intestinal epithelial cells than in controls.[52] Similarly, in pCD bacterial colonization was found to be in closer proximity to intestinal epithelial cells than in controls.[52] Confocal laser endomicroscopy has shown increased leakage of IV administered fluorescein through the intestinal epithelial barrier in pCD and pUC compared to controls.[53] A similar study in adults showed increased leakage of IV administered fluorescein detected by confocal laser endomicroscopy, to be associated with increased probability of relapse in the following 12 months in both CD and UC.[54]

These alterations in the intestinal microbiota seen in IBD carry a number of potential consequences for the patient.[55] Microbes can directly impact health by increasing or decreasing resistance to colonization by pathogens, producing vitamins and nutrients, and training the development of a balanced immune system.[55, 56] A balance host-microbe interaction is essential for health; germ-free mice show aberrations in a number of body systems including: cardiovascular, respiratory, gastrointestinal, and immunity. [56] Germ-free animals also show significant morphological and functional intestinal changes.[56] Interestingly, some of these aberrances can be corrected by colonization with specific bacteria, including improving intestinal barrier integrity with the addition of certain *Lactobacillus* species to germ-free animals. [57-59] An unhealthy relationship between gut microbiota and the host in IBD may perpetuate or trigger disease; elucidation of the importance of this relationship can improve our understanding of the etiologies of these diseases.

1.3.3 Metabolites

Metabolites are compounds necessary for or formed during metabolism, and are composed of both host and microbe factors that can reflect the real-time microenvironment of the gut. Identification and quantification of metabolites can provide a concrete assessment of microbial function as well as human metabolism. In addition to altered microbial abundances seen in IBD, variations in urinary metabolites suggest altered microbial function in the IBD microbiome. In adults with IBD, serum and urine metabolites can distinguish active disease from remission.[10] Urine metabolites can discriminate adult IBD patients from healthy controls, with a varying ability to distinguish CD from UC,[60, 61] while serum and fecal metabolites have been shown capable of distinguishing pCD from pUC and healthy controls.[62] A more limited number of studies suggest that urinary metabolomics can also differentiate pIBD from healthy controls.[63] Metabolites commonly altered in pediatric and adult IBD belong to metabolic pathways that suggest variations in glutathione metabolism, the citric acid cycle (CAC), and importantly, microbial activity.[60, 63, 64] Hippurate is an example of a metabolite marker of microbial metabolism, and is defined as a co-metabolite, as it is produced by a combination of host and microbial metabolism.[65] Hippurate is notably absent in germ-free animals,[65] and significantly reduced in IBD.[66] Pallister *et al.* found hippurate to be associated positively with microbiome diversity, as well as increased fruit and whole grain intake.[67] When potential dietary influences were excluded, Williams *et al.* (2009) found that reduced urinary hippurate persisted to be a marker for IBD vs controls.[68] Microbial metabolism as a source for hippurate variation was supported in an additional study by Williams *et al*. (2010), where supplementation with the hippurate precursor sodium benzoate did not increase urinary hippurate in CD or controls, and

CD patients had persistently lower levels of urinary hippurate.[69] In a longitudinal study of patients at risk for developing metabolic syndrome, hippurate levels were found to have positive associations with Shannon diversity and numerous taxa in fecal microbiota.[67] In humans, hippurate is considered a uremic toxin and may play a role in metabolic acidosis by stimulating renal ammoniagenesis.[70]

Other microbe associated metabolites include short chain fatty acids (SCFA), which are produced from microbial fermentation of dietary fibers and other substrates and are significantly reduced in both adult and pediatric IBD. [60, 61, 71, 72] Stool from IBD patients has less acetate, butyrate, and propionate, while lactate and pyruvate are increased.[71] Effects of SCFA include suppressing intestinal permeability and increasing the population of regulatory T cells that can help to prevent chronic inflammation.[73, 74] SCFA can also activate anti-inflammatory signalling cascades by binding to intestinal G-protein coupled receptors (GPCR) such as GPR109A, GPR43, and GPR41.[75] Impaired metabolism of SCFA has been implicated in IBD; for example, dextran sulfate sodium (DSS) is characterized by reduced butyrate oxidation while glucose oxidation increases.[76] Additionally, GPR43 expression in the ileum was significantly reduced in CD compared to controls, regardless of disease activity.[77] This could explain why exogenous administration of SCFAs has not consistently resulted in clinical improvement in the treatment of IBD.[78] Previous studies examining direct administration of SCFA has shown mixed results in the use of butyrate enemas, despite promising *in vitro, ex vivo*, and animal studies showing decreased inflammatory markers and improved intestinal barrier.[79, 80]

Butyrate produced by bacterial fermentation is the primary fuel source for colonocytes,[81] and colonocytes from germ-free mice have fewer citric acid cycle (CAC) intermediate enzymes.[76] Butyrate has multiple physiologic effects in the gut, including reducing bacterial

translocation *in vitro*;[82] butyrate acts as both an energy source and a histone deacetylase (HDAC) inhibitor.[76] Butyrate concentrations are increased by some resistant starches (RS), and could explain how dietary RS is associated with a reduced risk of colorectal cancer.[83] SCFA production varies in response to diet and intestinal transit time.[83] Measurement of stool SCFA content does not always reflect production, as longer transit time is associated with reduced butyrate.[83] Impacts of transmit time of SCFA concentrations could be due to increased opportunity for absorption and utilization during slower transit, or related to delivery time of RS and dietary fibre to the large bowel and time for microbial metabolism and fermentation.[83] A number of other factors are known to impact SCFA concentrations: antibiotics reduce SCFA and different dietary fibers can increase SCFAs in varying amounts.[84] Additionally, SCFA production can modify the intestinal microbiome: their acidity can select for more acid-tolerant bacteria.[84] SCFA may play a role in patient response to treatment; Wang *et al*. found a significant reduction in SCFA producing bacteria prior to Infliximab (IFX) therapy in pIBD, with a shift to a profile more similar to healthy controls after therapy.[80] The abundance of SCFA-producing species has been shown to be predictive of sustained remission in pIBD.[80]

1.3.4 Diet

A "Western" diet that is low in dietary fibre and high in processed and fatty foods has been proposed to partially account for the recent rise in the incidence in IBD, particularly in the western/developed world.[85] Although a number of dietary risk factors and associations have been identified, the majority have been identified through case-control or cohort studies. It is difficult to design and conduct experimental dietary studies that can prove causality, and in animal studies it can be challenging to approximate a human diet and consumption patterns. Assessing for

associations with individual foods or nutrients does not capture the complexity of micro and macronutrient interactions. Additional complexities include the associations of dietary intake with culture and genetics, as well as numerous reporting biases.

Diet has been identified to have associations with both increased and decreased risk in adult and pediatric IBD. CD has especially strong associations with dietary intake.[86] Foods associated with increased risk of IBD have included fast food[87] and increased protein intake that is potentially specific to animal protein.[86] Protective factors include breast-feeding,[86] and specifically for pCD: increased vegetables, fruits, fish, olive oil, grains, and nuts.[88] Foods associated with increased risk for pCD in females included meat, sugary foods like dessert, and high fat foods.[88] Diet is proposed to exert effects through alterations in host immunity, host barrier integrity, and alterations in the microbiota.[86] In the absence of dietary fibers, some microbes display enhanced mucolytic activity – digesting and depleting the host intestinal mucous layer, compromising intestinal barrier integrity and resulting in increased immune activation.[86] Epithelial barrier disruption, as evidenced by increased gap density, correlated positively with increased inflammation and disease activity.[53] Increased dietary fibers and RS can increase SCFA production, potentially improving production of anti-microbial peptides and decreasing intestinal permeability.[86] Diet may also impact the levels of SCFA receptors; in mice fed a high fat and sugar diet there was reduced GPR43 receptor expression.[77]

Diet is one of the main determinants of the intestinal microbiome.[89] Previous studies analyzing the correlation between diet and the microbiome have found that dietary patterns are significantly more predictive of the intestinal microbiome than individual nutrients – likely as the complex interaction of whole foods cannot be predicted from isolated components.[90] In animal models, certain diets will increase colonization with pathogenic organisms such as adherent

invasive *E. coli* (AIEC),[86] which is a key pathobiont implicated in the pathophysiology of CD.[91] Targeting these associations between host, diet, and microbe, a number of dietary therapies have been developed, and are described in the Treatment section.

1.4 Host – Microbe Interactions

Together, the genetic, microbial, dietary, and metabolic associations in IBD suggest a dysregulated host-microbe relationship as important in the pathogenesis of IBD. This is further supported by evidence that microbiota-targeting therapies like antibiotics and diet can result in clinical improvement in IBD.[82, 86] In adult CD, fecal microbiota transplant (FMT) has shown some promise, but results still vary based on dosing, frequency, donor profile, and preparation techniques.[92] As another example, diversion of the fecal stream in pIBD through an ileostomy can result in clinical improvement, with disease severity often increasing again when the fecal stream is returned to the colon.[93] Additionally, almost all colitis animal models, such as Interleukin-10 (IL-10) knockout or $TNF^{deltaARE}$ mice, do not develop intestinal inflammation under germ-free conditions.[94, 95] Roy *et al.* (2017) have shown that specific variations in microbial dysbiosis are associated with specific disease phenotypes and pathologies unique to the microbial community introduced.[96] *F. prausnitzii* has been shown to have anti-inflammatory effects, and is known to be reduced in CD; its absence is predictive of disease relapse.[97] Beneficial microbes such as *F. prausnitzii* that are reduced in IBD may improve mucosal barrier function through tight junction proteins, inducible heat shock proteins, and immunomodulation.[82]

Increased leakage of luminal antigens across the intestinal barrier in IBD, presumably due to genetic and environmental factors, result in an excessive cytokine-dependent

inflammatory response in mucosal tissue.[98] These antigens can include bacterial products or components such as LPS or peptidoglycan. Microbial antigens can increase NF-κB signalling through increased production of TNFα, which is recognized as an integral part of protection against invading pathogens.[99] The main sources of TNFα in the intestinal mucosa are macrophages, dendritic cells, and T cells.[98] TNFα is thought to play a role in the pathogenesis of IBD via fibroblast activation, enhanced production of pro-inflammatory cytokines, and enhanced T cell resistance to apoptosis.[98] TNFα effects through NF-κB also result in Paneth cell death, damage to intestinal epithelial cells, and increased production of matrix metalloproteinases (MMPs) by myofibroblasts.[98] Paneth cells are specialized secretory intestinal epithelial cells located in intestinal crypts of the small bowel (but also can be found in the large bowel in IBD) that are important for producing antimicrobial and immune-stimulating molecules.[100] They've been shown to be important in defense against intestinal pathogens and play a role in regulating the abundance of certain intestinal microbiota such as Bacteroidetes and Firmicutes, two phyla significantly altered in IBD.[100 50 101] *In vitro*, the addition of TNFα to intestinal epithelial cell monolayers also results in increased permeability.[82] Environmental factors such as diet and the intestinal microbiome along with genetic predispositions may play a role in perpetuating a dysregulated immune response that results in the chronic inflammation found in IBD.[98]

Another supporting factor for the role of an imbalanced host-microbe interactions is the ability of increased anti-microbial antibodies in patient serum to predict disease phenotype in CD years before diagnosis.[102] Patients with more antimicrobial antibodies were more likely to have complicated CD, with higher titers predicting an earlier progression to complicated disease.[102] In pCD, the antimicrobial antibodies anti *Saccharomyces cerevisiae* (ASCA), anti-outer membrane porin C [(OmpC), an *E. coli* antigen], and anti-Cbir1 (a flagellin protein), are also predictive of

complicated disease.[103] The more antimicrobial antibodies a child with pCD was positive for, the higher the likelihood of an earlier progression to complicated disease and an earlier need for surgical intervention.[103] Antimicrobial antibodies are generally less common in pUC,[104] but associations with FCP levels in pUC have been found.[105] The reduced associations between antimicrobial antibodies in pUC compared with pCD may be reflective of the degree of dysbiosis, with the CD microbiome typically further distinguished from healthy controls than the UC microbiome.[48]

1.5 Treatments in pIBD

The majority of current treatments for IBD focus on reducing inflammation, mostly through suppression of the immune system. Treatment of pIBD has some variations between pCD and pUC, but both can include glucocorticoids (GCS), thiopurines, methotrexate, aminosalicylates (5-ASA), and biologic therapy.[106, 107] Dietary therapy, which is unique in that it does not suppress the immune system, is a recommended therapy currently specific to pCD, with exclusive enteral nutrition (EEN) recognized as the first line therapy for luminal pCD (this will be discussed in detail below).[106] 5-ASAs are anti-inflammatory drugs and are suggested as a first line therapy to induce and maintain remission in mild or moderate pUC; GCS are the second choice for therapy to induce remission in both pCD and pUC.[106, 107] GCS carry the risk of steroid dependence, and have many side effects including increased risk for infections, diabetes, osteoporosis, and cataracts.[108] Biologic therapy with IFX is recommended for chronically active or steroid dependent/resistant pUC when 5-ASA and thiopurines do not control disease, and is used for both induction and maintenance of remission.[107] Biologics are compounds purified from another organism, and in the case of IFX is a recombinant antibody produced *in vitro* by a mouse

myeloma cell line.[109] In pUC, FMT and antibiotics are not currently recommended; but probiotics like VSL#3 or *E. coli* Nissle 1917 are approved as adjuvant therapy in mild pUC.[107] VSL#3 includes several *Lactobacilli* and *Bifidobacteria* species, along with *Streptococcus salivarius* subspecies *thermophilus*; ongoing studies are needed to prove if probiotics are more effective than placebo.[110]

Even among therapies targeting microbiota, effects on host-microbe interactions remain poorly elucidated. Intestinal microbiome-altering effects of GCS in animals in both normal and inflammatory conditions include increased *Bifidobacterium* and *Lactobacillus* species with elimination of mucin degrader *Mucispirillum*.[110] In pUC patients GCS use was associated with increased *Actinomyces* abundance in patients who sustained remission, and an accompanying decrease in a *Clostridium* OTU.[51] In another immune-mediated disease, systemic lupus erythematosus, modifying the microbiota with bromofuranone (a known disruptor of quorum sensing within biofilms)[111] was associated with enhanced effectiveness of prednisone treatment.[112] Thiopurines may increase mucosal bacteria numbers and adherence, decrease fecal bacterial diversity and richness, and *in vitro* inhibit *Mycobacterium avium* subspecies *paratuberculosis;*[110] which is thought by some to play a causative role in CD due to similarities with mycobacterial enteritis.[113] FMT, most commonly used in recurrent *Clostridioides difficile* infections (CDI), is currently being investigated for utility in IBD and is showing some promise in UC and CD, albeit at a lower success rate than in CDI.[114] Complexity in the host-microbe relationship of IBD when compared to CDI may explain the mixed results and weaker relationships so far observed between FMT and remission in IBD.[110]

1.5.1 Anti-TNF Therapy

As TNFα is a major immune factor involved in the pathogenesis of IBD, anti-TNFα therapy is recommended in pCD when, despite immunomodulatory therapy, a patient has persistent active luminal disease, or when there is active steroid-refractory disease, or active perianal fistulizing disease.[106] In pUC IFX is recommended for chronically active or steroid-dependent disease that is not controlled by 5-ASA and thiopurines.[107] IFX is a monoclonal antibody that can recognize both membrane-bound and secreted TNFα.[115] Membrane bound TNF may play more of a role than secreted soluble TNF; and agents that preferentially block soluble TNF have not been shown to be effective therapeutically in IBD.[98] Blocking of TNF effects by IFX has been shown to result in T cell apoptosis and reduction in cytokines, chemokines, and adhesion molecules that promote the influx of immune cells into the gut.[116] The relationship between TNFα and mucosal healing is complex, as TNFα knockout mice are more likely to develop colitis with DSS exposure.[82] Furthermore, the ability of the probiotic VSL#3 to decrease intestinal permeability in inflamed ileal tissue *ex vivo* was shown to be TNF-dependent in mice.[117] Regardless, *in vivo* anti-TNFα therapy in pIBD patients has been shown to improve the intestinal epithelial barrier and promote mucosa healing.[118, 119] Risks and side-effects of anti-TNFα include an increased risk of infection, paradoxical psoriasis in pCD, and infusion reactions.[5]

Lewis *et al*. in 2015 examined microbiome changes in pCD patients over eight weeks of anti-TNF therapy, and found microbiota shifts that after therapy more closely resembled healthy controls.[120] pCD patients with active disease compared to controls were found to have reduced relative abundance of almost one third of genera compared to healthy controls, including: *Roseburia*, *Ruminococcus*, *Akkermansia*, *Prevotella*, *Coprococcus*, and *Eubacterium*.[120]

Kowalska-Duplaga *et al*. (2020) found that Shannon diversity and OTU numbers increased with IFX therapy in pCD, with fewer differences in microbiota taxa comparing pCD and controls after IFX treatment than at baseline.[121] Wang *et al*. (2018) found that pCD patients who would have a sustained response to IFX therapy (defined here as a PCDAI score of $\leq$10 throughout follow-up), saw increases in levels of SCFA-producing genera with IFX therapy, including: *Blautia*, *Faecalibacterium*, *Lachnospira*, and *Roseburia*.[80] Adalimumab is another anti-TNF biologic therapy used in pIBD.[5] Although a number of other biologics are used and under development for pIBD, such as the integrin blocking Vedolizumab,[122] IFX is the most commonly prescribed biologic in pIBD and only IFX will be explored in detail here. Because IFX therapy is administered via intravenous infusion, there is no direct interaction with the host intestinal microbiome. Resultant microbial changes may be due to modified host-microbe interactions in the form of decreased inflammation through mediators such as reactive oxygen species or antimicrobial peptides, or shifts may reflect other changes concomitant with therapy such as diet.

1.5.2 Dietary Therapy

Currently a first line therapy for induction of remission in pCD, EEN therapy entails liquid meal replacement beverage and the exclusion of all other food and beverages except water for a period of typically six to eight weeks. The EEN formulas most commonly used include Modulen IBD®, Ensure/Ensure Plus®, Pediasure®, or Nutren/Nutren Junior®.[123] Although the safety of this treatment is excellent, with no concern for side effects, EEN poses considerable challenges to patients. The most commonly described challenges involve the palatability and monotony of the diet, as well as cost.[123] The exact mechanism of action of EEN remains

unknown, but theories include immune and/or microbial modulation, and improved intestinal barrier function with decreased inflammation.[124]

EEN therapy significantly impacts the microbiome within one week of initiation,[120] and preliminary studies have found that there are differential changes in the microbiota of responders vs non-responders to EEN. Kaakoush *et al*. (2015) found that responders had a greater decrease in OTUs than non-responders (response was defined as a PCDAI<10 after eight to 12 weeks of therapy).[125] Leach *et al*. (2008) found that patients with the greatest decreases in *Bacteroides* and *Prevotella* species with EEN treatment saw the greatest reduction in PCDAI, with microbiota changes persisting even four months after EEN completion.[126] In a study by Quince *et al*. (2015) EEN resulted in decreases in *Bifidobacterium*, *Ruminococcus*, and *Faecalibacterium*; Shannon diversity also decreased on EEN but returned to pre-treatment levels two months after EEN completion.[127] A number of other dietary therapies exist in pCD, including the Crohn's Disease Exclusion diet (CDED), which has also shown significant impacts on the microbiome.[128] The importance of these microbial changes in response to therapy remain unclear.

1.6 Gaps and Remaining Challenges

Although there are a number of effective treatments for pIBD, all carry risks of serious side effects or other challenges. Appropriately selecting therapy for patients is based on broad classifications of disease phenotypes rather than assessing in each patient for the contributions of the intestinal microbiome, diet, and genotype. A more in depth understanding of these factors in IBD pathophysiology could allow ***personalized medicine*** to address the imbalances in the host-microbe interaction unique to each patient. Better mechanistic insight into how existing therapies such as IFX and EEN are affected by and impact the host-microbe relationship can help guide

precision medicine in pIBD. The potential role of host intestinal microbiota in mediating response to therapy remains to be elucidated, and clarification can identify additional therapeutic targets as well as provide indicators of treatment effectiveness. Prospective data collected through the IDeaL and PUCIFX studies, together with repeated sample collection as described below, can help us to start filling in some of these gaps. Understanding these relationships can provide novel therapeutic targets and biomarkers for treatment candidacy and success.

Although dietary therapy can induce remission in up to 85% of pCD patients,[129] characteristics evident at baseline that can distinguish between those likely to respond and who are not can ensure that patients receive the most appropriate therapies as quickly as possible. Given the described microbiota changes associated with successful therapy, identifying patients most likely to exhibit these microbiota shifts with EEN may allow better identification of ideal EEN candidates. In animal models, intestinal microbial shifts in response to dietary therapy are best predicted by previous diet.[130] Identifying the role of previous diet and the baseline microbiome in prediction of response to EEN could improve identification of therapy candidates and provide novel dietary and microbial targeted therapies.

1.7 Hypothesis and Objectives

I hypothesized that **patient microbiota and metabolites are altered in treatment differentially in responders vs non-responders to therapy in pIBD.** My first objective was to ***identify the patterns of metabolites and microbiota associated with clinical outcomes in pIBD patients receiving IFX therapy***. The second objective was to ***identify previous dietary patterns that could predict patient response to dietary therapy in pCD patients.***

Partial correction of the dysbiosis characteristic of pIBD with IFX therapy identifies alterations in the host-microbe interface.[120] The differential response in the microbiome and metabolites in responders vs. non-responders to therapy suggests that the host-microbe relationship is either directly affected by successful therapy, or plays a role in mediating a positive response to IFX.[80] Prospective data collection from the IFX cohorts (IDeaL and PUCIFX) can identify baseline microbial abundances or metabolites that may be predictive of therapy response, as well as microbial abundances and metabolites that are characteristic of treatment success. Changes in these abundances or metabolites with therapy can help elucidate how the host-microbe relationship is shifting with treatment in pIBD.

Given the differential response to therapy in the microbiome in responders vs. non-responders that is also present in EEN, the microbiome also likely plays an important role in determining response to therapy.[125, 126] Shifts in the microbiome may be critical for successful treatment by improving the host-microbe relationship. As previous diet has been established as an important predictor of microbial response to dietary therapy,[130] we will examine how a child's long term dietary pattern is associated with their response to dietary therapy in newly diagnose pCD patients.

Chapter 2. METHODS

In order to fill the gaps in understanding the pathogenesis of pIBD and provide the ability to predict response to therapy, I was able to take advantage of robust, prospectively collected data and samples. These data have allowed me to examine the host-microbe relationship and how it relates to response to therapy in pIBD.

Two prospective pharmacokinetic studies, **PUCIFX** and **IDeaL** (described in detail below), provided an excellent opportunity to conduct ancillary studies to investigate correlations suggestive of host-microbe interactions during IFX therapy and possibly identify microbes or metabolites that may be predictive of response to therapy. Prospective data collection with repeated sample collection allowed for analysis of changes over time. Specimen collection was performed prior to each IFX infusion as detailed below, with specimens from before the first infusion referred to as baseline, those before the third infusion referred to as "pre-dose three", and those before the fifth referred to as "pre-dose five".

The Canadian Children Inflammatory Bowel Disease Network (CIDsCaNN, https://cidscann.ca/) is a CIHR-funded inception cohort, which has facilitated the collection of data and samples from over 1,400 newly diagnosed pIBD cases. Through this study we prospectively collected food frequency questionnaires (FFQ) at the time of diagnosis for children with IBD. Prospective longitudinal treatment and outcome data were available for children from across Canada; a subset of children newly diagnosed with pCD and receiving EEN who had completed FFQs were selected to examine the ability of previous diet to predict EEN outcomes. This cohort comprises what is referred to here as the **EEN-FFQ** study; the pCD patients from the EEN-FFQ study are a different population than the pCD patients enrolled in the IDeaL study.

2.1 Ethics

Informed consent was obtained by research coordinators from parents and/or patients as applicable, and participants were free to withdraw from the studies at any time. The IDeaL and EEN-FFQ studies utilized the CIDsCaNN research infrastructure and were approved by the University of Alberta Health Research Ethics Board (Pro00042980). PUCIFX was approved by the University of Calgary Health Research Ethics board (E-25172).

2.2 Study Design – IFX Cohorts

2.2.1 IDeaL – pCD

A prospective cohort of 33 patients with pCD were recruited for a multi-centre observational IFX pharmacokinetic study (IDeaL; PI: Dr. Hien Huynh; co-I Dr. Eytan Wine). The study utilized the CIDsCaNN network of participating children's hospital sites across Canada, including the Edmonton Pediatric IBD Clinic (EPIC, www.kidsibd.ca) at the Stollery Children's Hospital. Patients who provided personal/parental consent with a known diagnosis of pCD, were age two to 17 years, and had no previous history of receiving anti-TNF therapy were eligible for inclusion. Participants could withdraw from the study at any time. Patients eligible and consented for the study who contributed specimens for analysis came from CIDSCANN sites at the Stollery Children's Hospital in Edmonton, BC Children's Hospital in Vancouver, and the Hospital for Sick Children in Toronto, as detailed in **Table 1**.

Table 1 *IDeaL patient recruitment sites.*

Location of recruitment	Number of Participating Patients
Toronto – Hospital for Sick Children	8
Edmonton – Stollery Children's Hospital	23
Winnipeg – Manitoba Children's Hospital	2

IFX was prescribed at 5mg/kg, rounded up to the next 100mg; physicians were free to make dosing adjustments throughout the study based on clinical judgment. The first five doses of IFX were administered over a range of 13-22 weeks (accounting for clinical decisions to shorten duration between infusions as needed), following the dosing schedule shown in **Table 2**. Data and specimens were collected according to the schedule outlined in **Table 2**, with collected blood analysed for albumin, erythrocyte sedimentation rate (ESR), and hematocrit for use in calculating PCDAI and wPCDAI scores as described below.

Baseline endoscopy was performed and used to generate a Paris Classification and simple endoscopic score for Crohn's disease (SES-CD) for each patient.[8] A PCDAI score was generated at each visit, and is based on symptoms such as abdominal pain, stool consistency and number per day, and general well-being along with objective measures such as change in height and weight, abdominal tenderness, hematocrit, ESR, albumin, and the presence of perianal disease.[131] Physician global assessment of disease activity (PGA) was also documented at each visit.

Table 2 *IDeaL schedule for IFX infusion, and data and specimen collection.*

	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7
IFX Infusion:	Screening & Consent	Baseline/ First Dose	Dose Two	Dose Three		Dose Four	Dose Five
Visit Timing:	-	-	2 weeks ± 5 days from visit 2	4 weeks ± 5 days from visit 3	4 weeks ± 7 days from visit 4	2 weeks ± 7 days from visit 5	
Specimen Collection:		Blood Urine Stool	Blood	Blood	Blood Urine Stool	Blood Urine Stool	
PCDAI, PGA:		x	x	x	x		
Medication Use:	x	x	x	x	x	x	x
Anthropometrics:		x	x	x			x
SES-CD:		x					
Paris Classification:		x					
	Adverse events & protocol deviations as needed throughout.						

Stool specimens were collected by patients/families within four days prior to their IFX infusion, or within one day following the infusion. Patients were instructed not to include any stool that came into contact with urine or toilet water and were given a sterile collection tube and collection kit. If stool was not collected within two hours of delivering sample to coordinator, patients/families placed stool in the refrigerator or freezer until bringing it to their next IFX infusion. Once received by coordinators, samples were transferred to -80°C. One patient supplied stool approximately one month after each IFX infusion; these specimens were excluded from baseline analysis but included in the pre-dose five analyses. Two patients did not provide stool at their fifth IFX infusion, but provided stool collected within one day of receiving IFX dose four that were used in the analysis in place of pre-dose five. For patients where pre-dose four

specimens were used, pre-dose four patient characteristics, such as wPCDAI and disease activity were also used when assessing correlations with clinical data.

Urine was collected by patients/families within one week prior to their IFX infusion, or within 2 days following the infusion and then frozen. Patients were instructed to collect mid-stream urine and were supplied with sterile collection tubes. Samples, once received by coordinators, were transferred to -80°C. One patient supplied urine approximately one month after each IFX infusion; these specimens were excluded from baseline analysis but included in the pre-dose five analyses. Two patients did not provide urine specimens pre-dose five but provided urine within one day of receiving IFX dose four that were used in the analysis in place of pre-dose five. For patients where pre-dose four specimens were used, pre-dose four patient characteristics such as wPCDAI and disease activity were also used when assessing correlations with clinical data. Specimens collected at collaborating sites (**Table 1**) were sent to the Wine lab at the University of Alberta on dry ice and immediately transferred to -80ºC upon arrival; all specimens were received still frozen with intact dry ice.

2.2.2 PUCIFX – pUC

A prospective cohort of 22 patients with pUC were recruited for a multi-centre observational IFX pharmacokinetic study (PUCIFX; PI: Dr. Jennifer deBruyn, University of Calgary; co-I: Dr. Eytan Wine). Eligible patients were included at the time of starting IFX induction, were less than 18 years old, had a stool negative for *Clostridioides difficile* on polymerase chain reaction (PCR) and/or toxin assay, and had a diagnosis of UC established with clinical, endoscopic, and histological criteria.[132] Patients were excluded if they had a known concomitant gastrointestinal disorder such as celiac disease, or had previously received

immunoglobulin therapy for any reason. Parents and/or patients provided informed consent to participate in the study; specimens were provided by patients from Edmonton's Stollery Children's Hospital, Vancouver's BC Children's Hospital, and Manitoba Children's Hospital as detailed in **Table 3**.

Table 3 *PUCIFX patient recruitment sites.*

Location of recruitment	**Number of Participating Patients**
Vancouver – BC Children's Hospital	5
Edmonton – Stollery Children's Hospital	11
Winnipeg – Manitoba Children's Hospital	6

IFX was prescribed at 5mg/kg, rounded up to the next 100mg; physicians were free to adjust dosing and infusion timing based on clinical judgment. IFX was administered IV at weeks 0, 2, and 6, as illustrated in **Table 4**, followed by infusions every eight weeks thereafter for maintenance, or as deemed appropriate by physician clinical judgment. Stool and urine specimens were collected at baseline and prior to administration of dose three ("pre-dose three"). Endoscopy was conducted prior to starting therapy and was used to define disease location and extent using the Paris classification.[8] A pediatric ulcerative colitis activity index score (PUCAI) was generated at baseline and each subsequent visit, assessing: rectal bleeding, abdominal pain, activity limitations, and stool frequency, consistency, and timing.[133] Clinical and anthropometric data were collected as shown in **Table 4**.

Table 4 *PUCIFX IFX Infusion schedule. Urine and stool specimens were collected at the dose one visit and after four to eight weeks at the dose three visit.*

	Visit 1	Visit 2	Visit 3	Visit 4
IFX Infusion:	Screening & Consent	Baseline/ First Dose	Dose Two	Dose Three
Visit Timing:	-	-	2 weeks from visit 2	4 weeks from visit 3
Specimen Collection:		Blood Urine Stool	Blood	Blood
PUCAI, PGA:		x	x	x
Medication Use:	x	x	x	x
Anthropometrics:		x	x	x
Paris Classification:		x		

Stool specimens were collected by patients/families within five days before or after their IFX infusions. Patient instructions and supplies, and research coordinator handling of specimens were as described above for the IDeaL protocol.

2.3 Stool Analyses – IFX cohorts

Stools were kept in -80ºC until use, and were then thawed on ice; approximately one gram was aliquoted for SCFA extraction and 180-220 mg portioned for DNA extraction. Portions were refrozen at -80°C until ready for extraction. To measure FCP in the pUC and pCD multi-omics cohorts, research associate Dr. Rosica Valcheva performed FCP extraction and quantification utilizing BÜHLMANN fCAL® enzyme-linked immunosorbent assay (ELISA) kit according to provided instructions (BÜHLMANN Diagnostics Corporation, Amherst, USA. www.buhlmannlabs.com).

2.3.1 DNA Extraction

The pre-weighed stool was thawed on ice in two millilitre microcentrifuge tubes. QIAGEN® QIAamp® Fast DNA Stool Mini Kits (Product #51604; Qiagen, Hilden, Germany) were used to extract DNA according to the supplied protocol, with modifications to add a bead beating step as described here. One millilitre of InhibitEX® (Qiagen) buffer was added to stool samples and centrifuged until completely homogenized. Homogenized suspensions were heated for 15 minutes on a 95°C heating block. Eight hundred microlitres of homogenate was transferred into a two ml tube with 400mg of 0.1mm glass beads. A FastPrep-24 (M.P Biomedicals, Irvine, CA, USA) bead beater was run three times for 30 seconds at four metres per second; between each beating step samples were incubated on ice for five minutes to keep cool. After bead-beating, samples were centrifuged at 14,000 rpm for one minute. Five hundred microlitres of supernatant was transferred into a clean microcentrifuge tube. Twenty-five microlitres of Qiagen proteinase K was added to each tube and 500 µL of Qiagen Buffer AL was added before inverting samples to mix. Samples were then incubated for 10 minutes on a heating block at 70°C. Five hundred microlitres of 95% ethanol was then added to samples, which were inverted to mix. Sample lysate was then added 600 µl at a time to QIAamp spin columns and centrifuged at 14,000 rpm for one minute. Filtrate and tube were discarded, and spin column centrifugation repeated until all lysate was centrifuged. Five hundred microlitres of Qiagen Buffer AW1 was added to the spin columns and centrifuged at 14,000 rpm for one minute. Filtrate and tube were discarded. Five hundred microliters of Qiagen Buffer AW2 were added to the spin columns and centrifuged for three minutes at 14,000 rpm. Tube and filtrate were discarded; the spin columns were placed in new microcentrifuge tubes and centrifuged for three minutes at 14,000 rpm to remove all buffers. Spin columns were transferred into new tubes and

100 µl of double-distilled sterile water was added to each for one minute prior to centrifuging for one minute at 14,000 rpm. A NanoDrop spectrophotometer (ThermoFisher Scientific Canada, Waltham, USA. www.thermofisher.com\ca) was used to analyze DNA concentration. Three samples with DNA concentrations below 30 ng/µl were re-extracted from stool using the describe protocol. Six specimens with inadequate fecal material to repeat extraction were concentrated by research associate Dr. Valcheva. DNA concentration was performed by initially precipitating DNA by adding twice the sample volume of isopropanol and incubating at -20°C for 30 minutes. Samples were then centrifuged at 10,000 G for 15 minutes. The supernatant was discarded, and the DNA pellet was washed with 70% ethanol, followed by centrifugation at 10,000 G for two minutes. The supernatant was discarded, and DNA pellets were dried at room temperature prior to being resuspended in 25 µl of sterile double distilled water. All samples were diluted to achieve a final DNA concentration between 35 and 90 ng/µl; RNAse and DNAse free equipment and filtered pipette tips were used throughout the extraction, dilution, and concentration procedures. After achieving final concentrations DNA specimens were frozen at -80°C until shipment to Génome Québec for sequencing.

2.3.2 16S Sequencing

DNA extractions were priority shipped overnight in an insulated cooler of dry ice to McGill University. Library preparation and Illumina Miseq sequencing were performed by Génome Québec. Library preparation PCR was done with the following primers: forward "ACACTGACGACATGGTTCTACAGTGYCAGCMGCCGCGGTAA", and reverse "TACGGTAGCAGAGACTTGGTCTGGACTACNVGGGTWTCTAAT". Twenty-four µl of PCR mastermix containing 2.5 µl Roche PCR buffer 10x with 18 mM $MgCl_2$, 1.25 µl or Roche DMSO, 0.5 µl dNTP mix 10 nM (New England BioLabs, Ipswich, MA, USA. www.neb.ca), 0.1

µl of TAW 5U- µl Roche FastStart High Fi, 19.35 µl water, and 0.15 µl each of the forward and reverse primer were added to 1 µl DNA extract. PCR amplification was done starting with 2 minutes at 94°C followed by 33 cycles of: 94°C for 30 seconds, 58°C for 30 seconds, and then 72°C for 30 seconds. Once cycling was complete the thermocycler rested at 72°C for 7 minutes prior to completion. Amplification was then verified on a 2% agarose gel. PCR indexation was then performed where a barcoding step adds an index (barcode) to each sample and the sequence of Illumina adaptors required for DNA to bind the flow cell during sequencing. A PCR indexation master mix of 2 µl Roche PCR 10x buffer without $MgCl_2$, 1.44 µl of Roche 25 mM $MgCl_2$, 1 µl Roche DMSO, 0.4 µl dNTP 10 mM mix (New England BioLabs), 0.1 µl of TAQ 5U0 µl Roche FastStart High Fi and 12.06 µl of water added for a total of 17 µl mastermix to each sample. Added to the mastermix with 1 µl of PCR amplified DNA was 2 µl of index/barcode. The indexation PCR was then run on a Torpedo thermocycler with a ten-minute 95°C initial incubation followed by 15 cycles of: 95°C for 15 seconds, 60°C for 30 seconds, and 72°C for 60 seconds. After the 15 cycles the thermocycler rested at 72°C for three minutes prior to completion. Barcode incorporation was then verified for each sample on a 2% agarose gel. Amplicons were then quantified with Quanti-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies of ThermoFisher Scientific, Carlsbad, CA, USA. www.ThermoFisher.com\ca). Libraries were then generated by pooling 50 ng of each sample and were then cleaned up with sparQ PureMag Beads (Quantabio, Beverly, MA, USA. www.quantabio.com). Quality control of the library was initiated with quantification using Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit [Kapa Biosystems (Roche), Basel, Switzerland. www.roche.com]. Average fragment size was determined using a LabChip GX nucleic acid analyzer (PerkinElmer, Waltham, MA, USA. www.perkinelmer.com). Before sequencing, 8% of Phix control library

(derived from a PhiX bacteriophage genome) was added to the amplicon pool for a final concentration of 8 pM, serving as a sequencing control. Sequencing was then performed on the Illumina MiSeq using the MiSeq Reagent Kit v2 (Illumina, San Diego, CA, USA. www.illumina.com) with the following Locked Nucleic Acid™ (LNA™) modified custom primers (Exiqon [Qiagen], Aarhus, Denmark. www.exiqon.com):

Read one: LNA-CS1: ACACTGACGACATGGTTCTACA

Read two: LNA-CS2: TACGGTAGCAGAGACTTGGTCT

Index read primer: LNA-CS2rc: AGACCAAGTCTCTGCTACCGTA.

Sequencing was performed in Illumina MiSeq PE250, with each sample receiving 0.006 sequencing units.

2.3.3 Bioinformatics

Bioinformatics and calculation of diversity scores were done with the assistance of Dr. Juan Jovel, manager of The Applied Genomics Centre (TAGC) at the University of Alberta. Sequences received from Génome Québec were analyzed utilizing Quantitative Insights into Microbial Ecology (QIIME2).[134] On average, 60,022 raw reads per sample were obtained (with a minimum of 36,716 reads). Initially, sequences were filtered according to Q-score, using the quality-filter q-score script with default parameters [Q score <= 4 in a maximal stretch of three nucleotides; uncalled (N) nucleotides were not allowed; trimmed sequences were kept only if its length was at least 75% of the original sequence length]. Quality of trimmed sequences was visually inspected to determine a maximal sequence length to allow during processing. Sequences were then denoised using the Deblur workflow[135] with a trimming length of 220bp. Deblur is a workflow specific for Illumina data, which conducts error correction and uses the

16S Greengenes reference sequence as positive filter for removal of non 16S sequences.[135-138] Deblur provided a set of representative error-corrected sequences at the sub-operational taxonomic unit (OTU) level, referred to as amplicon sequence variants (ASV). Phylogenetic, diversity, and ordination analyses were based on Deblur-corrected representative sequences. For taxonomical annotation, the V3-V4 region was cropped out of the Greengenes database with OTUs clustered at 99% of similarity. Cropped sequences were used to train a classifier that assigned taxonomic levels to each sequence based on the taxonomy profiles available in the reference data base. Finally, the classifier was applied to conduct taxonomical classification of the Deblur-denoised representative sequences. A detailed workflow provided by Dr. Juan Jovel is included in **Appendix** 2: Bioinformatics Workflow.

Alpha diversity and richness calculations for the Chao1, Shannon H, and Simpson indexes were done on using Deblur-denoised sequences random samples to the minimum sequencing depth of 36,716 reads.

2.3.4 qPCR Verification

To verify the 16S sequencing results, quantitative polymerase chain reaction (qPCR) was done assessing for the following genera that had high mean relative abundances identified in 16S sequencing from Génome Québec: *Bifidobacterium, Akkermansia,* and *Faecalibacterium*. Total bacteria were also assessed to quantify the total bacterial DNA present. DNA concentrations ranged from 40-80 ng/µl.

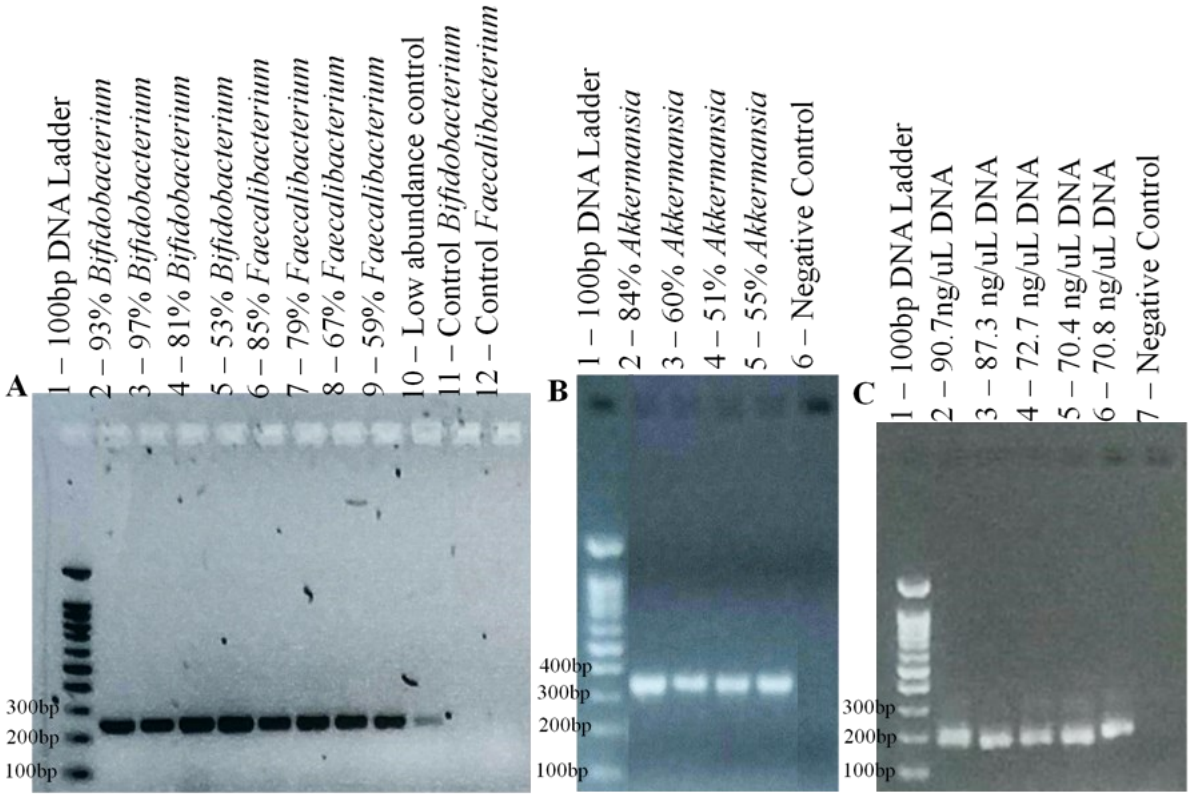
Table 5 *Primers for qPCR*

Target Genera	**Sequences**	**Fragment Size**	**Annealing Temperature**	**Reference**
Faecalibacterium	FWD: GGA GGA AGA AGG TCT TCG G REV: AAT TCC GCC TAC CTC TGC ACT	247 bp	60ºC	Fitzgerald *et al*. (2018)[139]
Bifidobacterium	FWD: TCG CGT CYG GTG TGA AAG REV: CCA CAT CCA GCR TCC AC	243 bp	60ºC	Rinttila *et al*. (2004)[140]
Akkermansia	FWD: CAG CAC GTG AAG GTG GGG AC REV: CCT TGC GGT TGG CTT CAG AT	327 bp	60ºC	Collado *et al*. (2007)[141]
Total Bacteria	FWD: CGG YCC AGA CTC CTA CGG G REV: TTA CCG CGG CTG CTG GCA C	200 bp	60ºC	Lee *et al*. (1996)[142]

To generate standards of known concentrations for use in qPCR, PCR amplification was run on aliquots of samples with high relative abundance of each taxon. For total bacteria PCR amplification was done on samples with high DNA concentrations after original stool extraction. For PCR, one microlitre of each primer, 12.5 µTop Taq Master Mix (Product #200403, Qiagen), and 8.5 µl of sterile double distilled water was added with two microlitres of DNA to sterile PCR tubes. PCR conditions were as follows: initial three minutes at 95ºC, followed by 35 cycles of 30 seconds at 95ºC, 30 seconds at 60ºC, and 30 seconds at 72ºC, followed by a final five minutes at 72ºC prior to completion. One negative control was run without the addition of DNA for each primer set. Following PCR amplification, DNA was run on a 2% agarose gel with five microlitres of SYBR Safe DNA Gel Stain (Product #S33102, Invitrogen, Carlsbad, CA, USA.). Five microlitres of a 100bp DNA ladder (Invitrogen, product #15628050) was run in the first

lane, and five microlitres of each PCR product was run with 3-5 µL of 10X Blue Juice Gel loading buffer (Invitrogen, product #10816015) (**Error! Reference source not found.**) to c onfirm a pure product of the appropriate fragment size.

Figure 1 *PCR amplification of select bacterial genera*



Legend*: Samples with high percentage relative abundance of each microbe were amplified to generate standards. % = relative abundance in each specimen.* ***A*** *Bifidobacterium and Faecalibacterium,* ***B*** *Akkermansia, and* ***C*** *Total Bacteria*

Remaining PCR products were combined and purified using QiaQuick PCR Purification Kit (Qiagen, Product #28104). Protocols were followed as supplied with the exception that DNA was eluted with water rather than buffer EB (10mM Tris-Cl, pH 8.5). DNA concentration was

then quantified using a Nanodrop spectrophotometer (ThermoFisher). Copy numbers were calculated using the following formula:

$$\frac{6.02x10^{23}\ x\ DNA\ concentration\ in\ \mu g/\mu L}{660\ x\ fragment\ size}$$

The fragment copy numbers for standards were as follows: *Bifidobacterium* $7.8x10^{16}$ counts, *Faecalibacterium* $8.09x10^{16}$ counts, *Akkermansia* $9.26x10^{16}$ counts, and Total Bacteria $2.42x10^{17}$ counts. For qPCR standards were run at dilutions of 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, and 10^{-8} concentrations. For qPCR 10 µL of Power SYBR green master mix (Applied Biosystems, Foster City, CA, USA. Product #4367659), 1 µL of forward primer, 1 µL of reverse primer, 8 µL of water, and 2 µL of DNA at 40-80ng/µL concentration were added per well. One standard for one organism were assessed per 96 well plate, with two negative controls per plate.

Standards were run at 10^{-3} to 10^{-8} concentrations in duplicate on each plate for qPCR, all plates were run with the follow temperature and cycle settings: 95ºC for 20 seconds followed by 30 cycles of 95ºC for 30 seconds, 60ºC for 20 seconds, and 72ºC for 30 seconds. After 30 cycles a melting curve was run to ensure purity of amplification with an escalation of 0.5ºC/minute up to 95ºC. Melting curves were assessed visually at the end of each qPCR plate to ensure purity of product amplification.

2.3.5 SCFA Extraction and Gas Chromatography

Weighed stool was thawed on ice; 800 µl of distilled water and 200 µl of 25% phosphoric acid were added to each 1 g of stool. Samples were thoroughly vortexed until completely homogenized and centrifuged for ten minutes at 4000 rpm. The supernatant was carefull6y

pipetted off into a fresh Eppendorf tube and centrifuged for 30 minutes at 17,000xG. Supernatant removal into a fresh Eppendorf tube and centrifugation/supernatant distillation was repeated until the supernatant was translucent. Supernatants were frozen at -80°C and transferred on dry ice to the University of Alberta Faculty of Agricultural, Life, & Environmental Science chromatography core facility for gas chromatography (GC) SCFA analysis. Samples were prepared by chromatography technician Lisa Nikolai using 800 µl of acid extraction with 200 µl of isocaproic acid as an internal standard and 200 µl of 5% phosphoric acid prior to running GC. Samples were run on a Varian 430 gas chromatographer. A Stabilwax-DA 30 mm by 0.53 mm by 0.5 µm column was used with a 7.5 psi head pressure and 20 mL/min split vent flow. Injector temperature was 170°C and detector temperature was190°C. The following temperature program was used: column at 90°C held for 0.1 min, increased to 170°C at a rate of 10°C/min and held for 2 min for a total run time of 10 minutes. Results were divided by original stool weight to give µg/g of stool.

2.4 Metabolomics

Urine samples were thawed on ice and 250 µl aliquots collected and transferred to The Metabolomics Innovation Centre (TMIC; https://www.metabolomicscentre.ca/) at the University of Alberta. Urine underwent gas chromatography mass spectroscopy organic acids assay, following methods as per the protocol outlined in Bouatra *et al*. (2013).[143] pUC and pCD organic acids gas chromatography-mass spectroscopy (GC-MS) assays were conducted in separate batches, and were analyzed separately to exclude compounds where ≥50% of patients had concentrations below the LOD. Three compounds that were included in pCD analyses were not included in pUC analyses due to LOD discrepancies, and these three compounds were also

excluded from analysis of IFX effects where pUC and pCD were combined, as well as when assessing for differences between pUC and pCD.

To adjust for individual patient hydration level, urine creatinine was quantified by TMIC using the Jaffe assay, and used to adjust urinary metabolite concentrations. Urine samples were filtered through IRA-4oo resin at an alkaline pH and diluted 1:10. The Jaffe assay was then performed on the filtered and diluted urine using yellow alkaline picric acid; absorbance was measured at 492 nm. All statistical analyses that follow were performed on the µMol/mmol creatinine concentrations. Metabolites where ≥50% of the specimens had below the limit of detection (LOD) were excluded from any analyses. Remaining specimens with metabolite concentrations below the LOD, the LOD concentration value was used.

2.5 Clinical Data Collection and Analysis

Clinical data were collected and managed by research coordinators using Research Electronic Data Capture (REDCap).[144, 145] Continuous clinical variables in both pUC and pCD cohorts were made binary or ordinal for use in permutational multivariate analysis of variance (PERMANOVA). Disease activity was quantified by PGA (considered the gold standard for assessing disease severity when recorded by an experienced specialist), with the following categories of disease activity: none, mild, moderate, and severe. Ages were converted to tertiles, and due to small samples sizes disease activity was made binary by combining no and mild disease severity and combining moderate and severe disease severity. A wPCDAI score of ≤12.5 has been shown to be the most accurate non-invasive cut-off to identify mucosal healing, and has the best correlation with fecal calprotectin.[17] wPCDAI uses similar measures as PCDAI but excludes height velocity, abdominal tenderness on exam, and hematocrit.[17] Remission was

defined as a wPCDAI of ≤12.5 by dose five for pCD and a PUCAI <10 by dose three for pUC. Clinical indices are summarized in

Table 6. Complicated pCD includes either of stricturing or penetrating disease or both, as defined by the Paris classification (B2 and/or B3) and was identified by the physician on assessment.[8] Disease location for pCD was broadly classified into ileal/upper Crohn disease (ICD) or colonic/ileocolonic Crohn (CCD) disease; a Paris classification of any colonic disease (L2 or L3) with or without ileal/upper involvement resulted in a CCD classification.[8]

Table 6 *Clinical indices used.*

Clinical Score	Components	Cut-Off to Define Remission
PCDAI – for use in pCD	Abdominal pain, stool consistency and number per day, general well-being, height and weight, abdominal tenderness, hematocrit, ESR, albumin, and the presence of perianal disease or extra-intestinal manifestations	NA
wPCDAI – for use in pCD	Abdominal pain, stool consistency and number per day, general well-being, weight, ESR, albumin, and the presence of perianal disease or extra-intestinal manifestations	≤12.5
PUCAI – for use in pUC	Rectal bleeding, abdominal pain, activity limitations, and stool frequency, consistency, and timing	<10

Disease location in pUC was divided based on Paris classification into either pancolitis (E4) or non-pancolitis, which included proctitis (E1), left sided colitis (E2), or extensive colitis(E3).[8] To define antimicrobial exposure, a cut-off of antimicrobial use within the 30 days prior to sample collection was used; glucocorticoid (GCS) use was defined as taking GCS at the time of sample collection at any dose, regardless of dosage tapering. Other concomitant

medication use data were collected, but to avoid over-inflation of Type I error, only GCS and antimicrobials were examined for effects.

2.6 Statistics and Data Analyses for IFX cohorts

To assess for correlation of clinical data with 16S, metabolomics, and SCFA, data sets were analyzed separately with PERMANOVAs to identify variables of interest. P.A.S.T.3 was used to perform PERMANOVAs with 9,999 permutations utilizing Euclidean distances.[146] Two-way PERMANOVAs were performed where sample size permitted to assess for significant interactions between variables, namely: age, sex, antimicrobial use, or disease location. Results from PERMANOVAs done with untransformed metabolite and SCFA data were compared to results from log-transformed and pareto scaled data. Nearly all significant findings were lost with data transformation and scaling, so we decided to proceed with analyzing metabolite and SCFA data using untransformed data and nonparametric methods to increase the likelihood of identifying potential relationships, despite small sample sizes.

PERMANOVAs with 16S data was done using centred log ratio transformed data; data were transformed in collaboration with Dr. Terry Zhang using R (R Foundation for Statistical Computing, Vienna, Austria, www.R-project.org).[147] Two filtering techniques were trialled on 16S sequencing data prior to analysis to remove very low relative abundance taxa and allow a reduced number of comparisons. A generalized filter combining all pCD and pUC specimens at all timepoints and filtering taxa with an average relative abundance <0.1%, or filtering taxa only if they had an average relative abundance <0.1% at both baseline and after treatment (combining pUC and pCD), allowing for taxa that were very low at a single timepoint to remain within the analysis if an adequate relative abundance was found at another timepoint. PERMANOVAs were

done with unfiltered data and results were compared to PERMANOVAs repeated with data filtered by each of the described techniques to ensure results were not extensively modified. Similar results were obtained with both filtered and unfiltered data, so filtering taxa if average relative abundances were <0.1% at both timepoints was used as it would theoretically support our main outcome of interest in detecting changes with treatment. At the phylum level this resulted in a minimum of 97.4% total abundance remaining after filtering; on average only 0.08% of abundance was filtered; three of the original nine phyla were removed leaving six in the analysis. At the genus level a minimum of 91.9% of total abundance remained after filtering, on average only 0.6% of abundance was filtered; 18 of the original 95 general were removed leaving 57 for analysis. Significance for statistical testing was defined as $p<0.05$, with a trending result defined as $p<0.1$.

PERMANOVAs were done on pCD and pUC separately, together, and contrasting pCD and pUC. Analysis was done on baseline data, post-treatment data, as well as changes over time (deltas). Additionally, baseline and post-treatment data were combined to examine for potential variable effects. Although this technique breaches the statistical assumption of independence among samples, a doubling of sample size could allow identification of important associations, allowing for additional exploratory analyses. Results from analyses combining data from both timepoints are reported as "all timepoints"; this was done with both multivariate (PERMANOVA) and univariate analysis where indicated.

Principal component analysis (PCA) of changes over time were done on urinary metabolites using Metaboanalyst (www.metaboanalyst.ca).[148] Metabolite concentrations were log transformed and Pareto scaled to adjust for higher concentration metabolites prior to PCA. Principal component (PC) scores for each sample were obtained.

After PERMANOVAs, univariate analysis was done to identify taxa, metabolites, or SCFA that may be responsible for driving the differences between groups. Variables that were significant or trending on the multivariate PERMANOVA analysis were assessed using univariate nonparametric techniques. Additionally, all data sets were assessed in pair-wise analysis using Wilcoxon-signed rank tests. Kruskal-Wallis tests were used to assess univariate relationships with age, and Mann-Whitney U tests were used to assess differences between groups. Non-parametric tests were completed using Metaboanalyst[148] and Stata 14 (Stata Statistical Software: Release 14. College Station, TX: StataCorp LP).[149] Stata 14 was also used to performed ordered logistic regression to compare changes in disease activity from baseline to after treatment.[149] Adjustments for multiple comparisons in univariate testing were done using Benjamini-Hochberg methods, with a significant finding defined as FDR<0.05 after adjustment, and a trend defined as FDR<0.1 after adjustment.

Spearman's rank correlation was used to assess for correlation of significant findings between datasets, and Fisher's exact and Mann-Whitney U-tests were used to compare categorical clinical features between groups, using in Stata 14.[149] Data was visualized using GraphPad Prism version 8.4.2 for Windows (GraphPad Software, San Diego, CA, USA, www.graphpad.com).

2.7 pCD EEN-FFQ Study

A separate cohort of pCD patients receiving EEN therapy was available for dietary analysis (diet data were not available for IDeaL and PUCIFX). Two hundred and eleven patients newly diagnosed with pCD who were beginning induction therapy with EEN had dietary

information collected using food frequency questionnaires (FFQ). For younger children, parents completed the FFQs. FFQs were collected from CIDsCaNN sites throughout Canada; those completed within 90 days before or after EEN initiation were included in analysis, resulting in a sample size of 103 children (**Table 7**) .

Table 7 *Centre of data collections for EEN-FFQ cohort.*

CIDsCaNN Site	**EEN-FFQ Participants**
Calgary – Alberta Children's Hospital	6
Edmonton – Stollery Children's Hospital	16
Halifax – Halifax Children's Hospital	24
Hamilton – McMaster Children's Hospital	13
Montreal – Montreal Children's Hospital	1
Ottawa – The Children's Hospital of Eastern Ontario	7
Ste Justine – Centre Hospitalier Universitaire Sainte-Justine	2
Toronto – The Hospital for Sick Children	25
Vancouver – BC Children's Hospital	6
Winnipeg – The Children's Hospital of Winnipeg	3

This specific FFQ was previously validated for use in pediatric and adolescent populations, assessed dietary intakes over the previous 12 months, and had questions regarding physical activity and parental education that were not utilized in this study.[150] The FFQ semi-quantitatively assessed dietary intake by providing sample portion sizes and allowing patients to select if their portions were larger than, smaller than, or approximately equal to the given example portion size. Questions were arrange in the following food groups (with the number of associated questions indicated): breads, cereals, and related foods: nine questions; soups, potatoes, and rice: seven questions; meat and poultry: eight questions; fish and seafood: three questions; mixed dishes: three questions; vegetable protein foods and eggs: five questions; vegetables: 11 questions; milk products: four questions, fruits: six questions; baked goods,

sweets, and salty snacks: six questions; juices and drinks: four questions; milk and similar beverages: four questions; coffee, alcoholic drinks & related items: six questions, and three questions clarifying oil type and frequency and cooking as well as food modifications. Food modification questions allowed patients to report consumption of: fried chicken, chicken skin, lean or extra-lean ground beef, breaded fried fish, low-calorie salty snacks, low-fat or fat free cheese/yogurt/margarine/mayonnaise/salad dressings, calcium fortified juices, decaffeinated coffee, meal substitute bars of beverages, and higher energy or high protein bars of beverages. Food frequency options for each question ranged from two times or more per day down to one to three times monthly or never/rarely.

For each food item on the questionnaire a caloric value was assigned. Three typical options for each food were selected and averaged from the Canadian Nutrient File (CNF) online database (Health Canada, 2015. https://food-nutrition.canada.ca/). Patient frequency selections were converted into a daily value, using the mid-range of the selection when applicable (i.e. three to five times per week was computed as 4/7 servings per day). Kilocalories were adjusted based on the reported servings sizes; 50% of the calories were assigned for selections of “less than” the given serving size, and 150% of the calories were assigned for selections of “greater than”. Modifier questions were used to adjust caloric intake utilizing additional questions; for example: as indicated chicken with/without skin or fried/unfriend, and low/no fat or regular fat dairy options. Meal, energy, and protein bars were added at three servings per month if selected by patients. Using the averaged CNF calorie counts total daily servings across all food groups was used to calculate a daily kilocalorie value for each patient. Estimated daily kilocalories were assessed for unusual values and outliers were tested by converting values into Z-scores. No

unrealistic values were identified, and no Z-scores were greater than 2 or less than -2 so no patients were excluded on this basis.

Next, FFQ responses to the total 82 questions were collapsed into 43 food and beverage groups based on nutrient similarity and common patterns of consumption (**Table 8**). Food groups were defined in consultation with two dietitians specializing in pediatric gastroenterology: Jessica Wu (RD) and Min Chen (RD, PhD).

Table 8 *FFQ Food Groups used for multivariate analysis.*

Food Group Name	FFQ Items Included in Food Group
High Fibre Cereals	High Fibre Cereals
Other Cereals	Other Cereals
Sugary Condiments	Sugar added to cereals; Jam, Honey, Sweet Spreads; Sugar in coffee/tea
Refined Grains	Commercial Sliced White Bread, Other White Breads
Whole Grains	Commercial Sliced Whole Wheat Breads, Other Whole Wheat Breads
Rice etc.	Rice, Rice Noodles, Couscous
Salty Snacks	Salty snacks (chips, crackers, pretzels, popcorn)
Nuts/Seeds	Peanut Butter; Sunflower seeds; Nuts, peanuts, other seeds
Vegetable Soup	Tomato or Vegetable Soups
Other	Other Soups
High fat Potato	French Fries or pan-fried potatoes
Potato	Boiled, Mashed, or baked potato
Margarine	Margarine
Butter	Butter, on bread or on cooked vegetables
Red Meat	Beef Other Meats (veal, lamb, game…)
Pork	Pork
Liver/Organ Meats	Liver, other organ meats
White Meat	Chicken, turkey
Processed Meat	Ham, cold cuts Sausages hot dogs
Meat Sauce	Sauces (brown, white, BBQ, gravy…)
Fish	Salmon, trout, sardines, herring, tuna Other fish (sole, cod, fish sticks…)
Seafood	Seafood
Pasta	Pasta with tomato sauce with or without meat, Pasta with creamy sauce
Pizza	Pizza
Legumes & Pulses (Excluding soy)	Beans, peas, lentils, hummus
Vegetable Protein Soy	Tofu, Soya
Eggs	Eggs
Vegetables	Green/yellow beans, green peas, corn; Carrots; Tomatoes; green, red, yellow sweet peppers; Other Vegetables; Leafy Greens; Cruciferous Vegetables
Vegetable Juice	Tomato or vegetable juices
Salad dressings, mayonnaise, dips	Salad dressings, mayonnaise, dips

Cheese	Cheese
Yogurt	Yogurt
Fruit	apples, pears; Bananas; Melons; Oranges, grapefruits, tangerines; Other fruits Berries
Sugary Desserts	Cakes, pies, doughnuts, pastries; muffins; cookies; Candies, chocolate; Milk-Based Desserts; Ice Cream
Pure Fruit Juice	Pure Fruit Juice – 100%
Sugar Sweetened Beverages	fruit drinks (sugar added), Soda
Diet Soda	Diet Soft drinks
Full Fat Milk	Whole milk, Milk or cream in coffee/tea
Reduced Fat Milk	1%, 2% or skim milk
Milk Alternatives	Soya Milk
Coffee	Coffee
Alcohol	Beer; Table wine, aperitifs; Spirits
Granola bars, chewy bars, cereal bars	Granola bars, chewy bars, cereal bars

2.7.1 Dietary Pattern Analysis

Data-driven approaches, such as principal component analysis (PCA), to guide dietary analysis rely on multivariate methods to establish patterns inherent in a population's reported dietary intake.[151] PCA captures in each principle component as much variation as possible in the data, and in the context of diet can be interpreted as dietary patterns within the population.[152] Each patient receives a principal component (PC) score for each principal component/dietary pattern that can be interpreted as their adherence to that particular diet. PC scores can then be assessed for correlation with outcomes or characteristics of interest when assessing for dietary associations. Avoiding multiple comparisons that would be generated if assessing one food group at a time can prevent over-inflation of type I error. Multivariate analysis can also account for increases in some food groups that are accompanied by decreases in others, presenting both as associations with potential variables of interest.

Energy adjustment is another important factor when analyzing dietary patterns. Adjustment reduces the risk of confounding caused by variations in total energy intake, allowing investigators to examine the effects of specific nutrients more clearly.[153] Energy adjusting intake also helps control for body size, rate of metabolism, and physical activity,[153] particularly important when working in the pediatric population, as patients in our study cohort ranged from age two to 18. Adjusting for energy intake when utilizing PCA to establish dietary patterns can help ensure proportional changes between dietary intakes are driving the differences between participants.[154]

Food group servings per day were kilocalorie-adjusted using the Willett residuals method as described by Willet *et al* (1997)[153] in Stata 14.[149] Adjusted servings were used to conduct PCA that was orthogonally rotated to minimize between the different dietary factors.[152] Each patient's score for each PC reflected their adherence to that particular dietary pattern and was converted into Z-scores for correlation with clinical outcomes (individuals are adherent to all patterns but to different extents). Z scores greater than 3 or less than -3 were deemed outliers and were excluded from further analysis. Based on a sharp drop off in Eigenvalues, a cut off of four principal components/ dietary patterns was used. Dietary factor loadings contributing more than 0.2 were considered to be of importance in the first four components and used to define and name the dietary patterns. PC scores were assessed for correlation with clinical features such as age, sex, and disease location as defined by Paris classification (ICD vs CCD).[8] Mann-Whitney U tests and spearman correlation were completed using Stata 14.[149]

Chapter 3. RESULTS

3.1 IFX Cohorts: Patient and Sample Characteristics

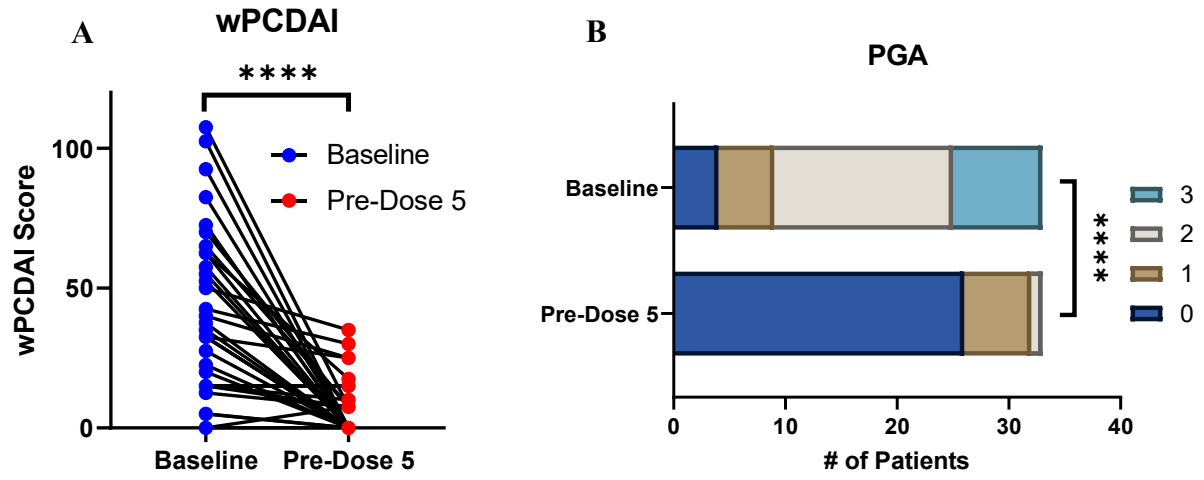
3.1.1 pCD Cohort

Patient characteristics are summarized in **Table 9**. No pCD patients had a history of surgical resection, nor did any patients require resection during the study. No patients progressed to complicated disease throughout the study duration. Sample numbers available for the various analyses are summarized in **Table 10**. Four patients had no disease activity at baseline; these patients were excluded from pairwise and changes over time/with treatment analyses, as well as baseline pCD vs pUC comparisons. These four patients had received a course of EEN to induce remission, three of four had IFX initiated during or immediately after EEN (while they were already in remission) as planned maintenance therapy. The last of these four patients completed the course of EEN and initiated IFX maintenance therapy two months later. As shown in **Figure 1**, wPCDAI and disease activity as assessed by PGA were significantly reduced after IFX therapy. Twenty six of 32 (81%) patients achieved a wPCDAI score of $\leq$12.5 by their dose five IFX infusion; characteristics of responders vs. non-responders is provided in **Table 11**. As shown in **Figure 2**, there was no significant difference in wPCDAI or PGA at baseline between responders vs. non responders (highlighting the need for better predictors of response). Concomitant medication and treatment use by patients are summarized in **Table 12**.

Table 9 *pCD Baseline patient clinical characteristics.*

Clinical Characteristic	Mean/Count
Mean Age (range)	12.4 years (5.3 - 17.9)
Sex	48% Male (16/33)
Mean Disease Duration (range)	0.6 years (0-7.1)
Blood Per Rectum	61% (20/33)
Perianal Disease	48% (16/33)
Mean SES-CD at Baseline (95% C.I.)	15.5 (12.1-27.5)
Colonic Disease	76% (25/33)
Complicated Disease	12% (4/33)

Figure 1 *wPCDAI (**A**) and PGA (**B**) significantly reduced with IFX therapy in pCD.*



Legend*: **A** wPCDAI score was significantly lower after IFX therapy in pCD. **B** Patients had significantly less disease activity on PGA after IFX therapy in pCD. ****=p<0.0001. n =33.*

Table 10 *pCD specimens available for analysis.*

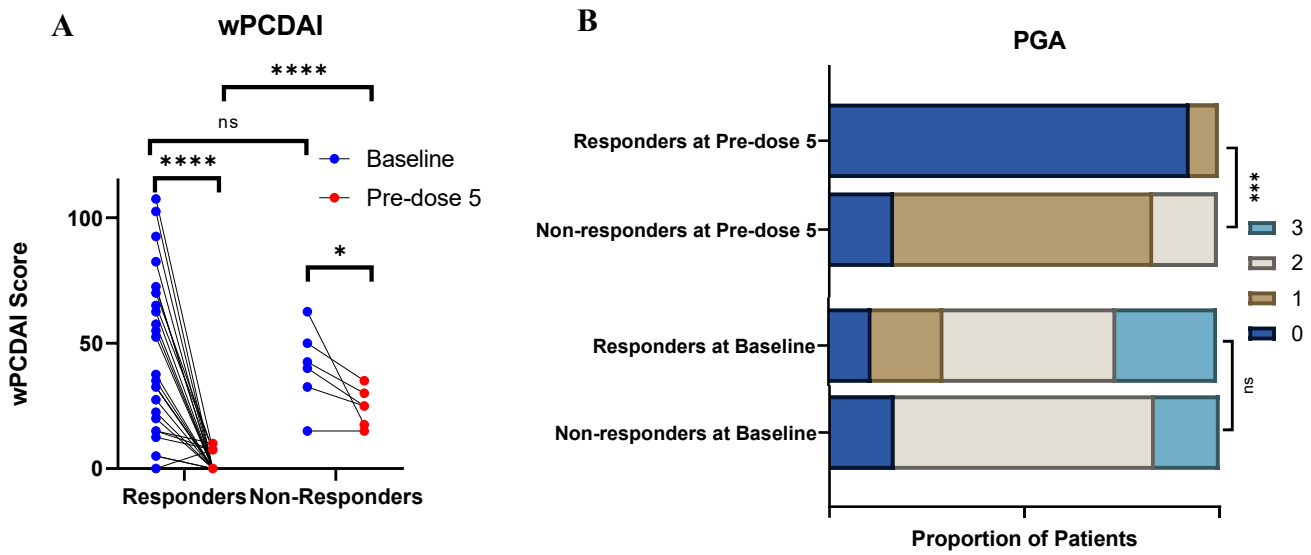
Analysis:	**Baseline**	**Pre-Dose 5**	**Complete Pairs**
Urine	30	31	30
Stool for SCFA	25	27	25
Stool for 16S	27	28	27
Stool for FCP	27	30	24

Table 11 *Characteristics of responders vs non-responders*

	Responders n=27	**Non-responders** n=6	**p-value**
Sex – Male	48% (13/27)	50% (3/6)	0.642
Mean Disease Duration (range)	0.63 years (0-7.1)	0.29 years (0.01-0.6)	0.47
Colonic Disease	74% (20/27)	83% (5/6)	0.544
Perianal Disease	52% (14/27)	33% (2/6)	0.358
Mean Age (range)	12.1 years (5.3-17.0)	13.9 years (11.6-15.0)	0.1946

Legend*: Responders were defined by a pre-dose five wPCDAI ≤12.5*

Figure 2 *No significant difference in disease severity at baseline between pCD responders & non-responders.*



Legend*: **A** There was no significant different between responders and non responders before IFX treatment in pCD. Response was defined by a pre-dose five wPCDAI≤12.5. **B** There was no significant difference in baseline PGA between responders and non-responders. *=p<0.05, ***=p<0.001, ****=p<0.0001, ns = not significant. Responders n=27, Non-responders n=6.*

Table 12 *Concomitant medication/treatment use at each timepoint.*

Treatment	Use at Baseline	Use at Pre-dose 5
EEN	6/33	0/33
Antimicrobials	11/33	4/33
Glucocorticoids (Methylprenisolone, Prednisone)	5/33	1/33
Methotrexate	13/33	25/33
Aminosalicylates (Sulfasalazine)	3/33	2/33
Azathioprine	6/33	6/33

Legend: *Antimicrobial use was defined as any antimicrobial taken within the 30 days prior to sample collection; other medications were included if concomitant at the time of the specific IFX infusion. Medications were included regardless of dose, including some patients tapering GCS.*

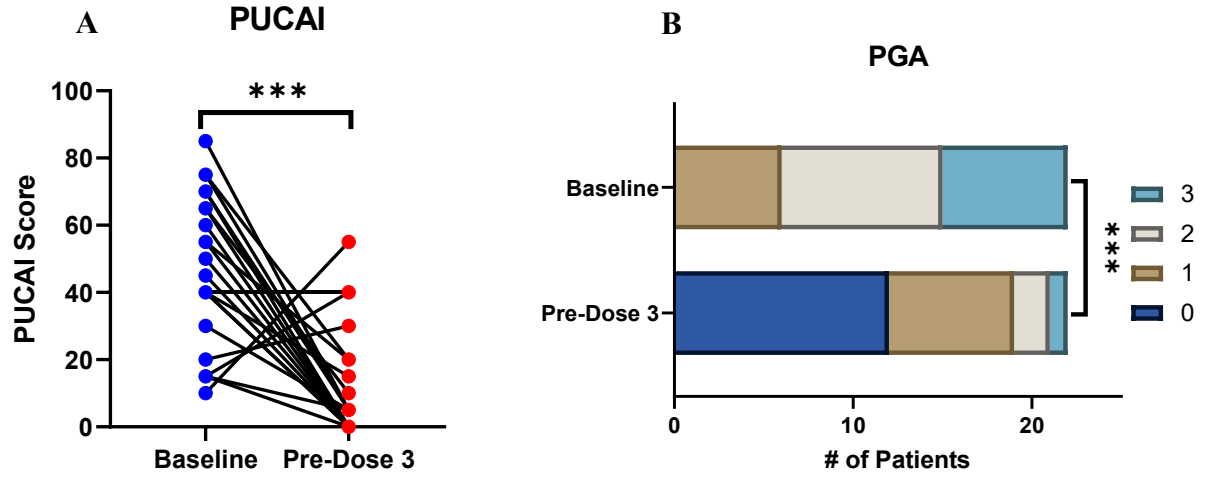
3.1.2 pUC Cohort

Patient characteristics for the pUC cohort are summarized in **Table 13**. No pUC patients had a history of surgical resection, nor did any patients require resection during the study. Sample numbers available for the various analyses are summarized in **Table 14**. Unlike pCD, all patients had active (mild to severe) disease at baseline, so all were included in pairwise and changes over time/with treatment analyses. As shown in **Figure 3**, PUCAI score and disease activity as assessed by PGA were significantly reduced with IFX therapy in pUC. Fourteen of 22 (64%) patients achieved a PUCAI score of <10 by their third IFX infusion; characteristics of responders vs. non-responders are provided in **Table 15**. Concomitant medication use by patients is summarized in **Table 16**.

Table 13 *pUC patient clinical characteristics.*

Clinical Characteristic	Mean/Count		P-value
Mean Age (range)	12.56 years (7.64 – 16.56)		
Sex	8/22 Males 36%		
Mean Disease Duration (range)	1.22 years (0 – 8.57)		
Blood Per Rectum	**Baseline**: 18/22	**Pre-Dose 3**: 7/21	0.002
Pancolitis	55% (12/22)		

Figure 3 *PUCAI score and PGA significantly reduced with IFX therapy in pUC.*



Legend: ***A*** *PUCAI score is significantly reduced after IFX therapy in pairwise analysis in pUC.* ***B*** *PGA is significantly lower after IFX therapy in ordered logistic regression in pUC. *** = p<0.001. n=22*

Table 14 *pUC specimens available for analysis.*

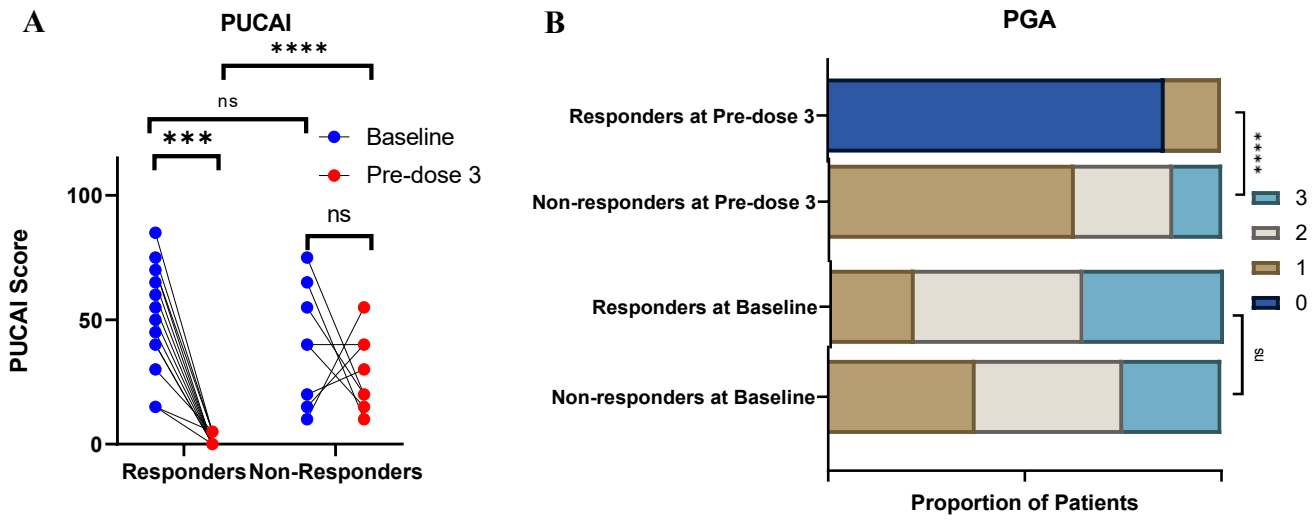
	Baseline	Pre-Dose 3	Pairs
Urine	20	20	20
Stool for SCFA	12	10	10
Stool for 16S	13	10	10
Stool for FCP	12	16	10

Table 15 *Characteristics of responders vs non-responders.*

	Responders n=14	**Non-responders** n=8	**p-value**
Sex	4/14 male	4/8 male	0.19
Mean Disease Duration (range)	1.7 years (0-8.6)	0.5 years (0-1.5)	0.19
Pancolitis	8/14	5/8	1
PUCAI Mean	Dose 1: 51.07 Dose 3: 1.79	Dose 1: 40 Dose 3: 28.75	0.29
Mean Age (range)	12.3 years (7.6-16.6)	12.79 years (8.6-15.7)	0.73
PGA	**Baseline:** None – 0 Mild – 21% (n=3) Moderate – 43% (n=6) Severe – 36% (n=5) **Pre-Dose 3:** None – 86% (n=12) Mild – 14% (n=2) Moderate – 0 Severe – 0	**Baseline:** None – 0 Mild – 38% (n=3) Moderate – 38% (n=3) Severe – 25% (n=2) **Pre-Dose 3:** None – 0 Mild – 63% (n=5) Moderate – 25% (n=2) Severe – 13% (n=1)	**Baseline:** 0.43 **Pre-Dose 3:** <0.0001

Legend*: Responders were defined by a pre-dose three PUCAI <10.*

Figure 4 *No significant difference in disease severity at baseline between pUC responders & non-responders.*



Legend*: **A** There was no significant different between responders and non responders before IFX treatment in pUC. Response was defined by a pre-dose three PUCAI<10. **B** There was no significant difference in baseline PGA between responders and non-responders. ***=p<0.001, ****=p<0.0001, ns = not significant. Responders n=14, Non-responders n=8.*

Table 16 *Concomitant medication use at each timepoint.*

Medication	**Use at Baseline**	**Use at Pre-dose 3**
Antimicrobials	0/22	0/22
Glucocorticoids (Methylprenisolone, Prednisone)	12/22	7/22
Methotrexate	3/22	1/22
5-Aminosalicylate (Sulfasalazine, Mesalazine)	12/22	7/22
Azathioprine	12/22	7/22

Legend: *Antimicrobial use was defined as any antimicrobial taken within the 30 days prior to sample collection; other medications were included if concomitant at the time of the specific IFX infusion. Medications were included regardless of dose, including some patients tapering GCS.*

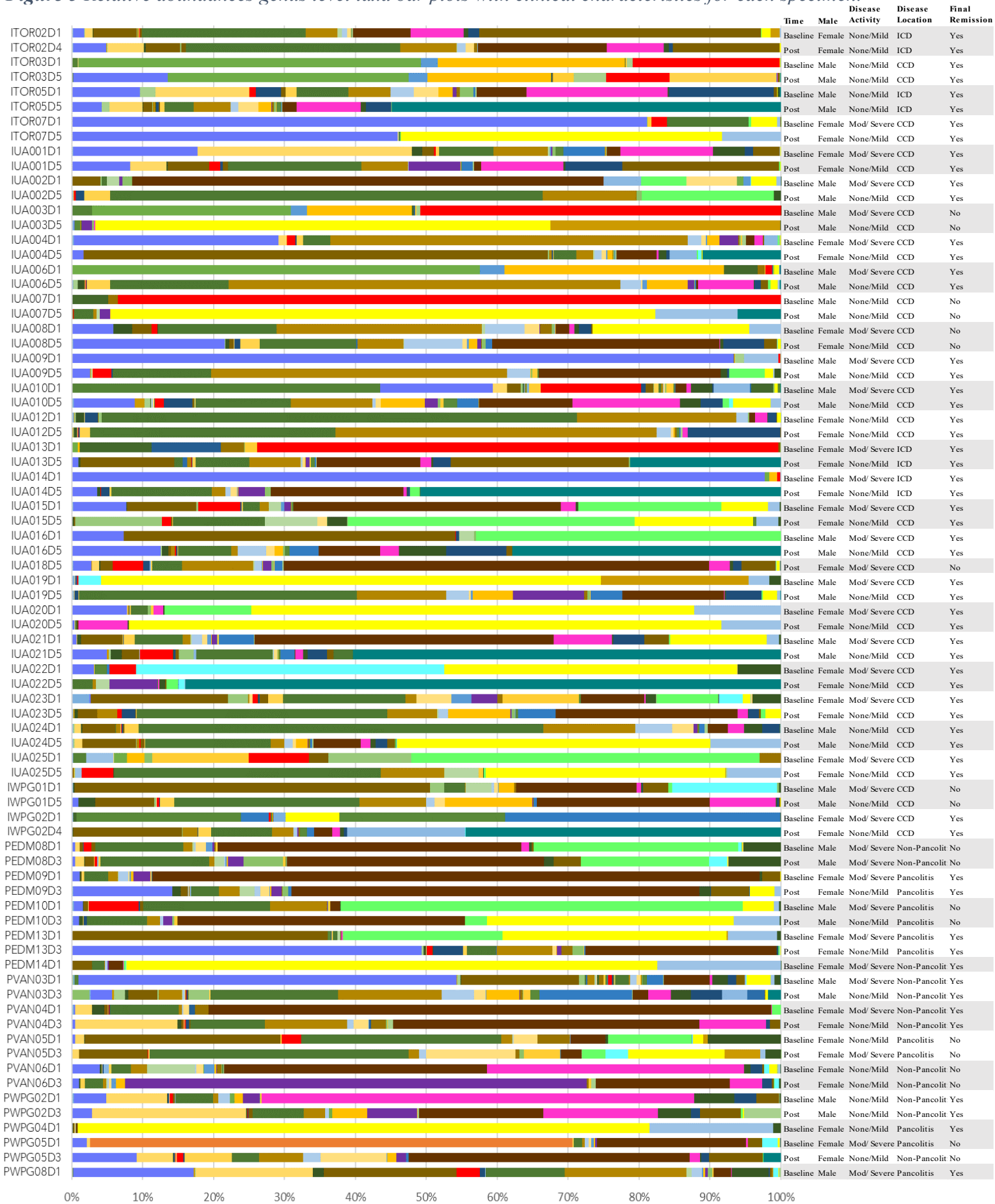
3.2 Microbiota Composition

pCD (IDeaL) and pUC (PUCIFX) were sequenced as one batch at Génome Québec. On average, 60,022 raw reads per sample were obtained (with a minimum of 36,716 reads). A total of 9069 unique OTUs were identified, and average of 334 OTUs per specimen were identified, with a minimum of 83 and a maximum of 581 in a single specimen. Taxonomy was assigned down to the genus level, with analysis at the genus and phylum level reported here. Nine taxa at the phylum level were identified, six of which remained after filtered those with less than 0.1% average abundance both before and after IFX treatment. On average, only 0.08% of total abundance was filtered, with a maximum of 2.6% abundance lost per specimen at the phylum level. Ninety-five taxa were identified at the genus level, 57 of which remained after filtering those with less than 0.1% average abundance both before and after IFX treatment. On average only 0.6% of abundance was filtered, with a maximum of 8.1% of abundance lost per specimen at the genus level.

After initial analysis, it became evident that certain samples had exceptionally high relative abundances of specific taxa (**Figure 5**). The highest relative abundances included: 98% *Bifidobacterium* in a baseline specimen from pCD patient, 84% *Akkermansia* after treatment in a pCD patient, and 86% *Faecalibacterium* in a baseline specimen from a pUC patient, as shown in **Figure 5**. No clear pattern was evident as relatively homogeneous specimens were from both the pCD and pUC cohorts, as well as before and after IFX treatment. Clinical data were assessed, and no clear relationships could be identified with features such as sex, disease severity/presence of blood in stool, medication use, antimicrobial use, or disclosed probiotic use. Troubleshooting of high relative abundances was performed to assess for any bio-informatics sources. Taxonomy

assignments were re-attempted utilizing SILVA[155, 156] in the QIIME2 pipeline rather than Greengenes[135-138] with little differences in relative abundances observed. Assigning OTUs based on 97% sequence similarity with the Greengenes taxonomic database was also attempted, with resulting relative abundances highly similar to when 99% similarity was used.

Figure 5 *Relative abundances genus level taxa bar plots with clinical characteristics for each specimen.*



Legend*: Specimens arranged in pairs where possible, baseline on top, after IFX on bottom.*

3.2.1 qPCR Validation Results

To validate 16S results, qPCR was conducted utilizing *Akkermansia, Bifidobacterium*, and *Faecalibacterium* specific primers with Eubacterium 16S primers as DNA concentration controls. These three taxa had exceptionally high abundance in several specimens so were selected for validation with total bacteria sequences to provide relative abundances. qPCR was unable to detect quantifiable amounts of DNA in approximately one third of stool DNA extractions. The specimens with exceptionally high relative abundances in the 16S analysis, i.e. the previously mentioned: 98% *Bifidobacterium* in a baseline specimen from pCD patient, 84% *Akkermansia* after treatment in a pCD patient, and 86% *Faecalibacterium* in a baseline specimen from a pUC patient, had undetermined DNA concentrations in some or all four validation plates.

The consistent inability to amplify 16S sequences across qPCR plates, along with the exceptionally high relative abundances identified on 16S sequencing suggests poor quality DNA after stool extraction, perhaps due to a chemical contaminant. As described by Pollock *et al*. (2018) a number of organic and inorganic inhibitors can exist in stool that will interfere with 16S rRNA sequencing results.[157] Although our stool extraction kit was modified by the addition of a bead-beating step, which has been shown to increase DNA yield, this may not have been enough to overcome substantial inhibitors that remained present after extraction.[157]

While the preliminary results obtained from original unvalidated 16S sequencing are shown here, stool will require DNA to be re-extracted with techniques capable of purification from inhibitors such as humic acids or polyphenols that may present prior to publication of this work. Repeated 16S sequencing and qPCR validation will be done to provide verified microbial abundances for use in analyses and integrative models.

3.3 Changes with IFX Therapy

3.3.1 pCD

3.3.1.1 Urine Metabolites

Sixty-two compounds were identified on GC-MS organic-acids assay by TMIC. Eighteen out of these 62 compounds were excluded from analysis as ≥50% of subjects had metabolite concentrations below the LOD. Two-way PERMANOVAs were run where sample size permitted; there were no significant interactions between age, sex, or antibiotic use with any other variable that altered the significance or trending result. For example, age and perianal disease had significant interactions at all timepoints, but the effects of both variables remained insignificant after adjusting for the effects of the other so unadjusted p-values are shown here. As shown in **Table 17**, disease location (ICD vs CCD) was the only clinical characteristic associated with an altered multivariate metabolite profile. Associations with IFX therapy will be discussed here; associations with other variables are discussed below in respective sections. Although differences between baseline and pre-dose five were not significant on multivariate comparison, on pair-wise analysis several metabolites were significantly altered (**Table 18**).

Table 17 *CD Urine Metabolite PERMANOVA Analyses Results.*

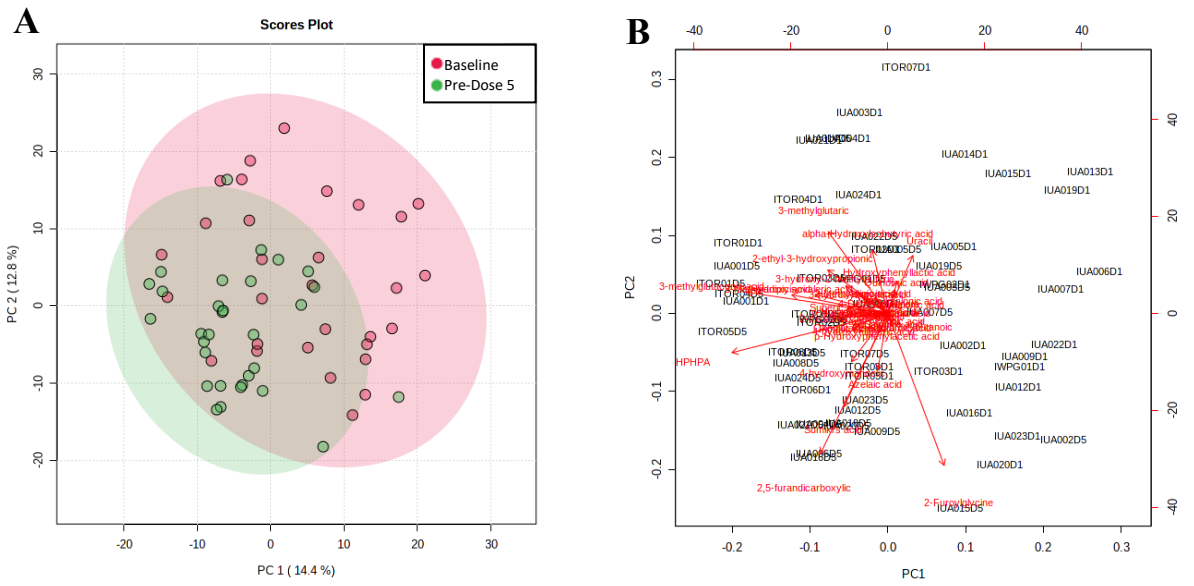
Variable	Baseline	Pre-Dose 5	Both Timepoints	Deltas
Baseline vs. Pre-Dose 5			0.1949^{+}	
ICD vs CCD	0.65	**0.0004**$^{+}$	**0.0018**	**0.0019**
Perianal Disease	0.3018	0.5885	0.2956	0.8066
Antimicrobial Use	0.214	0.8671	0.1526	**Baseline**: 0.99 **Pre-Dose 5**: 0.38
Disease Severity$^{\blacklozenge}$	0.46	NA	**0.099**	**Baseline:** 0.14 **Pre-Dose 5:** NA
Remission by Dose 5 (wPCDAI $\leq$12.5)	0.56	0.06		0.10
Pre-Dose 5 FCP<250	0.20	0.73		0.98
Pre-dose 5 FCP<100	0.29	0.79		0.90
$\geq$75% drop in FCP	0.38	0.73		0.96
GCS	0.2276		0.131	**Baseline:** 0.1781 **Pre-Dose 5**: NA
Sex	0.61	0.71	0.56	0.40
Age	0.89	0.82	0.97	0.67

Legend*: P-values for associations of each variable with urinary metabolite profile in multivariate analysis. ICD = Ileal Crohn Disease, CCD = Colonic or Ileocolonic Crohn Disease, wPCDAI = weight Pediatric Crohn Disease Activity Index, FCP = Fecal Calprotectin, GCS = glucocorticoids.* $^{+}$ *= significant compounds on univariate analysis. NA = insufficient numbers to perform statistical analysis.* $^{\blacklozenge}$ *= no or mild disease activity vs moderate or severe disease activity. Bold = p<0.1*

Unsupervised analysis of urinary metabolites at both timepoints shown in **Figure 6A** PCA showed no distinct clustering in pCD. Principal component one accounted for 14.4% of variability and principal component two accounted for 12.8% of variability. Factor loadings for the top three contributing metabolites for the first two principal components is shown. The biplot

visualizing these factor loadings is shown in **Figure 6B** and demonstrates metabolite influences on specimen PCA plotting.

***Figure 6**. Urine Metabolites do not separate by time in the pCD cohort.*



C

Principal Component 1		**Principal Component 2**	
Metabolite	**Factor Loading**	**Metabolite**	**Factor Loading**
(3-hydroxyphenyl)-3-hydroxypropionic acid (HPHPA)	-0.52	2-furoylglycine	-0.54
3-methylglutaconic acid	0.47	2,5-furandicarboxylic acid	-0.5
3-methyladipic acid	-0.32	Sumiki's acid	-0.33

*Legend: **A**. PCA of pCD baseline and pre-dose five urinary metabolites using Euclidian distances. Red = baseline, green = pre-dose five. **B.** Biplot of Euclidean distance PCA, red text shows metabolites important in accounting for sample variability. Black text are specimens plotted on PCA coordinates. **C.** Factor loadings of top three metabolites contributing to each principal component as visualized in biplot.*

***Table 18**. Urinary metabolites significantly altered with IFX therapy in pCD.*

Metabolite	Baseline	upper 95%	lower 95%	Pre-Dose 5	upper 95%	lower 95%	FDR
Quinolinic acid	27.79	35.38	20.20	15.30	19.40	11.20	0.00022
Hydroxyphenyllactic acid ⬇	2.23	3.55	0.91	0.76	1.17	0.34	0.022
Oxoglutaric acid ⬆	35.38	47.78	22.97	51.94	67.24	36.64	0.022
2,5-furandicarboxylic acid ⬆	1.90	2.83	0.97	6.52	9.70	3.35	0.023
4-hydroxyhippuric acid ⬆	6.52	9.70	3.35	82.28	116.88	47.68	0.023
3-methylglutaconic acid ⬆	0.96	1.49	0.44	1.27	1.89	0.65	0.023
1H-Indole-3-acetic acid ⬆	10.34	16.36	4.32	35.33	62.76	7.89	0.023
Ethylmalonic acid ⬇	4.21	6.12	2.30	2.29	3.11	1.47	0.023
3-hydroxyphenylacetic Acid ⬆	0.21	0.34	0.09	1.85	3.65	0.06	0.023
Glycolic acid ⬆	99.60	124.07	75.14	137.95	173.66	102.24	0.023
4-hydroxymandelic acid	2.65	3.88	1.43	4.03	5.10	2.97	0.040
HPHPA ⬆	11.06	22.65	-0.53	23.49	34.46	12.53	0.045

***Legend:** Mean concentrations in µmol/mmol creatinine of urinary metabolites significantly altered (FDR<0.05) with IFX therapy, in paired analysis, ranked by ascending FDR. Only significant compounds shown.*

3.3.1.2 SCFA

Two-way PERMANOVAs were run where sample size permitted; there were no significant interactions between age, sex, or antibiotic use with any other variable that altered the significance or trending result. IFX therapy was associated with a significantly altered SCFA profile on multivariate analysis (**Table 19**). Associations with IFX therapy are discussed here; associations with other variables are discussed below in their respective sections.

Table 19 *IFX treatment, antimicrobials, and disease severity are associated with an altered SCFA profile in pCD.*

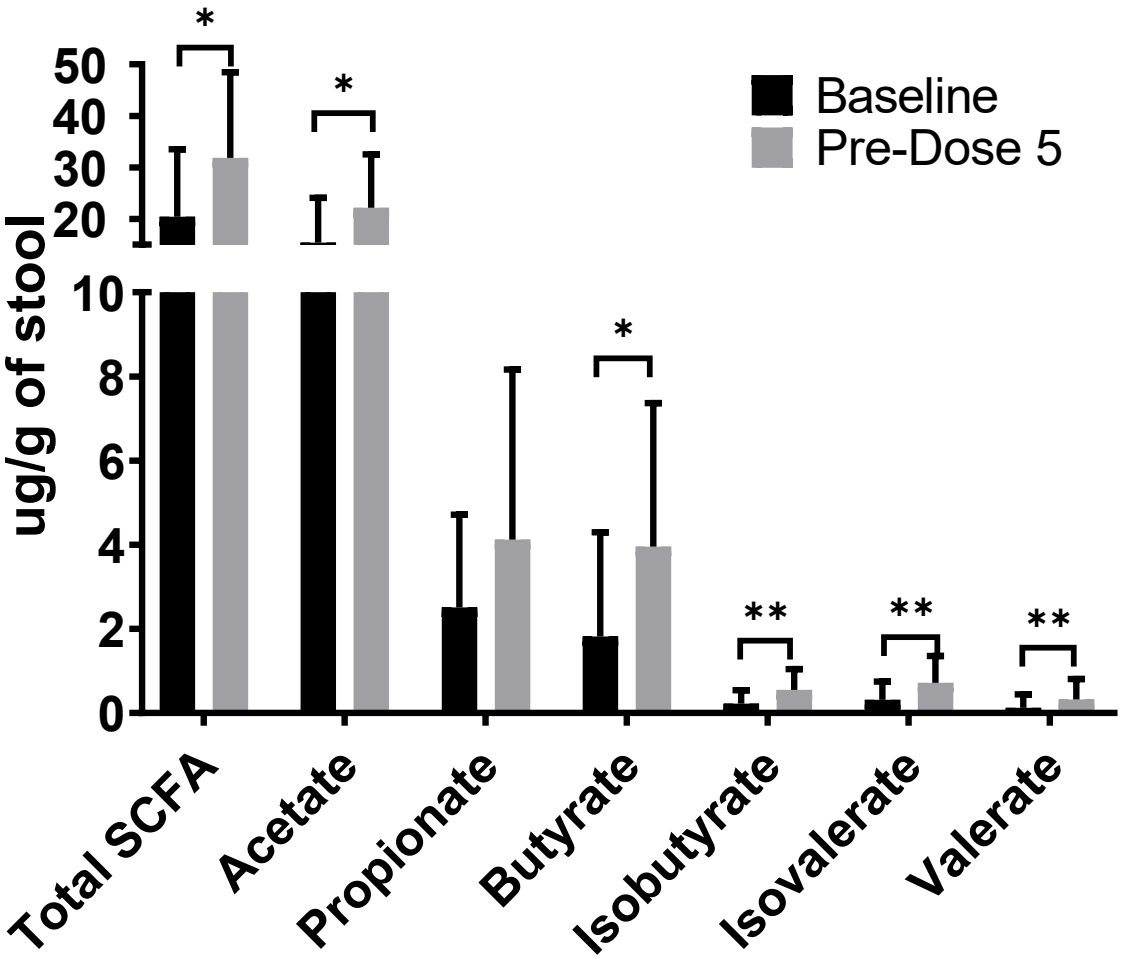
Variable	Baseline	Pre-Dose 5	Both Timepoints	SCFA Deltas
Baseline vs. Pre-Dose 5			**0.0179**$^{+}$	
ICD vs CCD	0.36	0.37	0.69	0.3766
Perianal Disease	0.9971	0.1287	0.2882	0.9236
Antimicrobials	0.16	0.71	**0.0821**$^{+}$	**At Baseline**: 0.3109 **At Dose 5**: 0.6687
Disease Severity$^{\blacklozenge}$	0.77	NA	**0.0708**$^{+}$	**At Baseline**: 0.9901 **Pre-Dose 5:** NA
Remission by Dose 5 (wPCDAI $\leq$12.5)	0.88	0.92		0.9289
Pre-Dose 5 FCP<250	0.4661	0.2824		0.4771
Pre-Dose 5 FCP<100	0.8723	0.4158		0.285
$\geq$75% drop in FCP	0.695	0.5354		0.8207
GCS	0.5443		0.4992	**Baseline:** 0.9898 **Pre-Dose 5:** NA
Sex	0.53	0.41	0.94	0.5421
Age	0.88	0.91	0.77	0.9911

Legend*: pCD Stool SCFA PERMANOVA analysis results.* $^{+}$ *= significant compounds on univariate analysis. NA = insufficient numbers to perform statistical analysis.* $^{\blacklozenge}$ *= no or mild disease activity vs moderate or severe disease activity. Bold = p<0.1*

Univariate analysis of fecal SCFA concentrations revealed all but propionate had a significantly increased concentration after IFX treatment in pCD (**Figure 7**). Several outliers were identified by a Z-score greater than three or less than negative three, so univariate analysis

was repeated with outliers excluded. Significant findings persisted after the exclusion of outliers, so for further analysis statistical outliers were included (**Table 20**).

***Figure** 7 SCFAs increase with IFX therapy in pCD*



*Legend: *=p<0.05, **=p<0.01*

***Table 20**. SCFA increases persist with exclusion of outliers.*

SCFA	P-value
Acetate	**0.0496**
Propionate	0.1138
Isobutyrate	**0.0048**
Butyrate	**0.0285**
Isovalerate	**0.0082**
Valerate	**0.0024**
	Outliers Excluded **0.0042**
Total SCFA	**0.019**
	Outliers Excluded **0.0333**

***Legend:** Pairwise analysis of pCD stool SCFA changes over time, n=21 pairs, with outliers excluded; n=20 for Valerate and Total SCFA.Outliers were detected by Z-scores greater than ±3; analysis was repeated with outliers excluded to validate results.*

3.3.1.3 16S

Genus Level

Two-way PERMANOVAs were run where sample size permitted; there were no significant interactions between age, sex, or antibiotic use with any other variable that altered the significance or trending result. As shown in **Table 21**, at the genus level, IFX treatment, antimicrobial use prior to baseline, and disease severity were all associated with significantly different microbial profiles when specimens from both timepoints are combined. Antimicrobial use prior to baseline also had a trending association with microbiota changes over time with IFX therapy. Changes associated with IFX therapy are discussed here; associations with other variables are discussed in their respective sections below.

Table 21 *Altered microbiota associated with IFX therapy, antimicrobial use, and disease severity.*

Variable	Baseline	Pre-Dose 5	Both Timepoints	Baseline to Dose 5 Changes
Baseline vs. Pre-Dose 5			**0.0005**$^{+}$	
ICD vs CCD	0.5282	0.097	0.0885	NA
Perianal Disease	0.7522	0.7045	0.3607	0.9377
Antimicrobial Use	**0.0304**$^{+}$	0.327	**0.0004**$^{+}$	**Baseline: 0.0513**$^{+}$ **Pre-Dose 5: 0.0616**
Remission by Dose 5 (wPCDAI ≤12.5)	0.8813	0.3182		0.4756
Disease Severity	0.0665	NA	**0.0017**$^{+}$	**Baseline:** 0.6312 **Pre-Dose 5:** NA
Remission by Dose 5 (wPCDAI ≤12.5)	0.8813	0.3182		0.4756
Final FCP <100	0.7463	0.8801		0.3602
Final FCP <250	0.2717	0.9391		0.3965
FCP >75% Drop	0.7439	0.876		0.1203
Glucocorticoids	0.8766	0.9674	0.5741	**Baseline:** 0.9154 **Pre-Dose 5:** NA
Sex	0.464	0.8091	0.6485	0.2122
Age	0.3052	0.8103	0.1785	0.9089

Legend: *Genus level pCD stool microbiota PERMANOVAs, Wilcoxon rank sum used for paired analysis of baseline vs. pre dose five.* $^{+}$ *= significantly different or trending differences in taxa on univariate analysis. NA = insufficient numbers of patients to perform analysis. Bold = p<0.1*

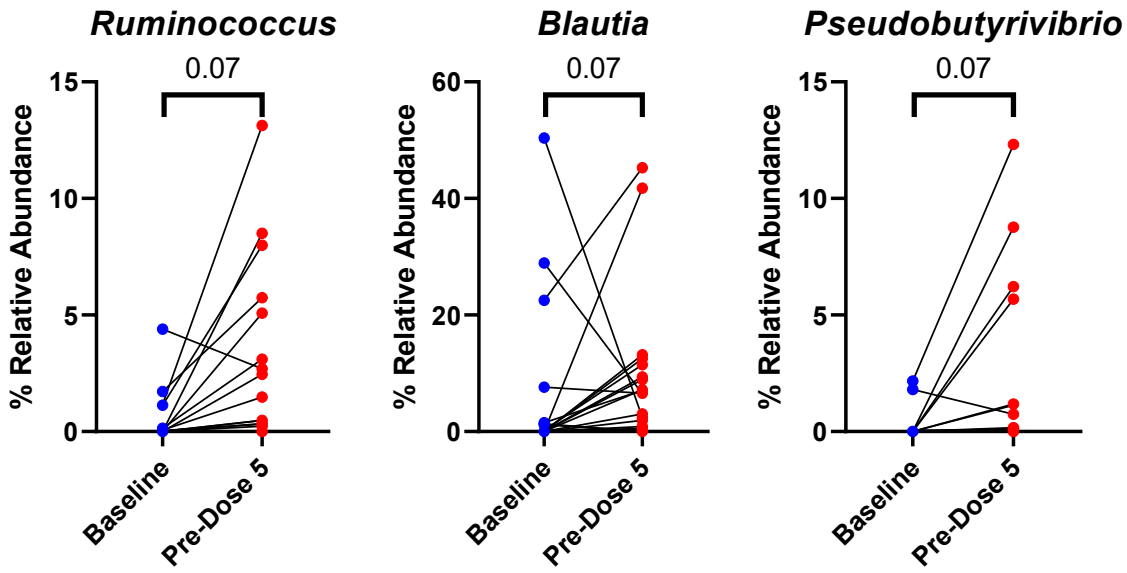
On pair-wise univariate analysis, after adjusting for multiple comparisons, there were three genera with trending differences after IFX treatment in pCD, shown in **Table 22** and **Figure 8**.

Table 22 *Three genera with trending differences with IFX treatment*

Genus	P Value	FDR
Ruminococcus ⬆	0.004528	0.07475
Blautia ⬆	0.004875	0.07475
Pseudobutyrivibrio ⬆	0.005246	0.07475

Legend*: Genera with trending differences in pCD stool microbiome between baseline and pre-dose five. Only genera with significant or trending differences are shown here.*

Figure 8*. Three genera with trending differences with IFX treatment*



Legend*: Relative abundance of genera trending different in paired analysis of pCD stool microbiome between baseline and pre-dose five.*

Phylum Level

Two-way PERMANOVAs were run where sample size permitted; there were no significant interactions between age, sex, or antibiotic use with any other variable that altered the significance or trending result. At the phylum level, IFX treatment, disease severity, and achieving a final FCP <250 were associated with significantly different microbiota profiles on multivariate analysis. Phylum level microbiota associations were identified with IFX therapy,

disease severity, and FCP, as shown in **Table 23**. Changes associated with IFX therapy are discussed here; associations with other variables are discussed in their respective sections below.

Table 23 *IFX treatment, disease severity, and a final FCP <250 associated with phylum level differences in stool microbiota in pCD.*

Variable	**Baseline 16S**	**Pre-Dose 5 16S**	**Both Timepoints 16S**	**Baseline to Dose 5 Changes**
Baseline vs. Pre-Dose 5			**0.0369**$^{+}$	
ICD vs CCD	0.5279	0.0444	0.0478	NA
Perianal Disease	0.5461	0.2487	0.1236	0.9017
Antimicrobial Use	0.8458	0.7933	0.5964	**Baseline**: 0.35 **Pre-Dose 5**: 0.7105
Disease Severity	0.54	0.97	**0.0145**$^{+}$	**Baseline:** 0.5638 **Pre-Dose 5**: NA
Remission by Dose 5 (wPCDAI ≤12.5)	0.807	0.6765		0.4292
Final FCP <100	0.1652	0.3444		0.3282
Final FCP <250	**0.0066**	0.2972		0.1779
FCP >75% Drop	0.121	0.3238		0.4316
Glucocorticoids	0.9277	0.9615	0.8248	**Baseline:** 0.7918 **Pre-Dose 5**: NA
Sex	0.1222	0.8221	0.6687	0.1265
Age	0.9653	0.9649	0.9897	0.8794

Legend: *Phylum level pCD stool microbiota PERMANOVAs, Wilcoxon rank sum used for paired analysis of baseline vs. pre dose five.* $^{+}$ *= significantly different or trending differences in taxa on univariate analysis. NA = after exclusion of patients in remission at baseline, there were insufficient numbers to analyze associations between disease activity, or dose five GCS use and phylum changes over time.*

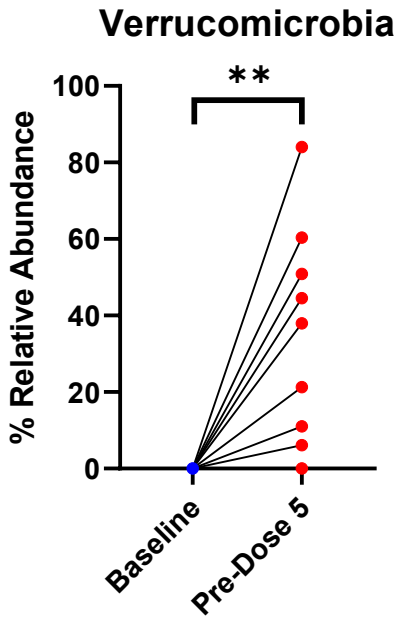
Univariate pair-wise analysis of phylum level changes over time identified significant changes in Verrucomicrobia relative abundance after adjusting for multiple comparisons (**Table 24** and **Figure 9**).

Table 24 *An increase in Verrucomicrobia was associated with IFX treatment in pCD.*

Phylum	Baseline	Upper 95%	Lower 95%	Pre-Dose 5	Upper 95%	Lower 95%	P Value	FDR
Verrucomicrobia ⬆	0%	0.004	0%	13.76%	24.27%	3.25%	8.22E-04	0.004
Actinobacteria ⬇	19.96%	33.96%	5.96%	6.13%	10.72%	1.54%	0.407	0.91
Bacteroidetes ⬆	6.47%	12.76%	0.19%	5.93%	12.05	0%	0.586	0.91
Firmicutes ⬆	54.47%	70%	38.94%	56.25%	71.38%	41.12%	0.711	0.91
Fusobacteria ⬇	0.4%	1.21%	0%	0.01%	0.02%	0%	0.879	0.91
Proteobacteria ⬇	18.62%	32.39%	4.84%	17.88%	32%	3.76%	0.913	0.91

Legend*: Relative abundance phylum level associated changes in stool microbiota with IFX treatment in pCD.*

Figure 9 *An increase in Verrucomicrobia was associated with IFX treatment in pCD.*



Legend*: Changes in the phylum Verrucomicrobia relative abundance with IFX treatment in pCD. **=p<0.01*

3.3.1.4 Microbial Diversity & Richness

As shown in **Table 25**, IFX treatment and antimicrobial use were associated with significant and/or trending changes in microbial diversity and richness. Changes with IFX treatment are discussed here; associations with antimicrobial use are discussed in variable specific sections below.

Table 25 *IFX Treatment and antimicrobial use associated with microbial diversity and richness in pCD.*

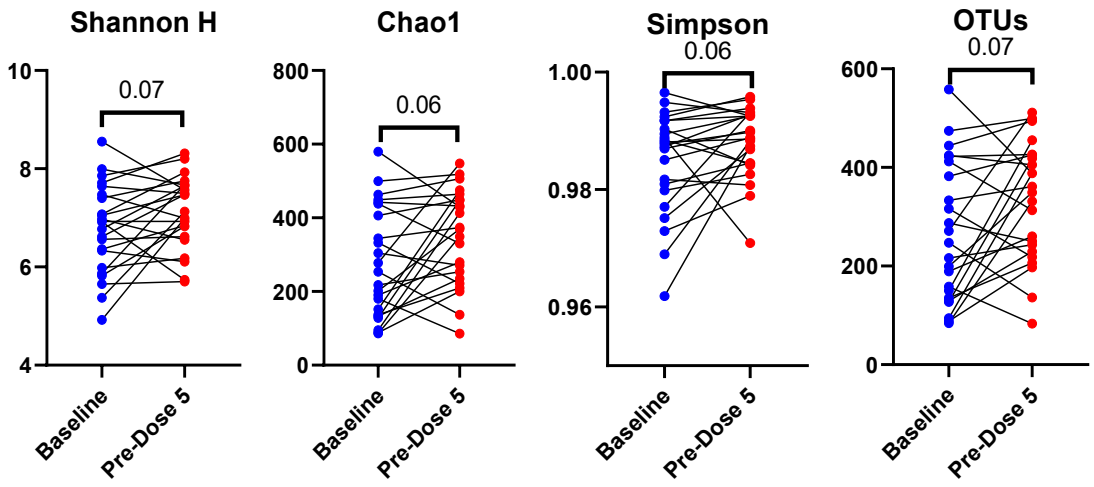
Variable	Baseline	Pre-Dose 5
Baseline vs. Pre-Dose 5	Paired Analysis: **Chao1 – 0.0636** **Shannon – 0.0727** **Simpson – 0.0553**	

	OTUs – 0.0680	
ICD vs CCD	Chao1 – 1.0 Shannon – 0.8378 Simpson – 0.7848 OTUs – 0. 9728	Chao1 – 0.1680 Shannon – 0.2122 Simpson – 0.2935 OTUs – 0.1310
Antimicrobials	**Chao1 – 0.0484** **Shannon – 0.0679** Simpson – 0.1034 **OTUs -0.0430**	Chao1 – 0.928 Shannon – 0.1936 Simpson – 0.3160 OTUs – 0.3159
Disease Severity	Chao1 – 0.4840 Shannon – 0.5598 Simpson – 0.5598 OTUs -0.5402	NA
Remission by Dose 5 (wPCDAI $\leq$12.5)	Chao1 – 0.4528 Shannon – 0.8378 Simpson – 0.5390 OTUs -0.5390	Chao1 – 0.5288 Shannon – 0.7882 Simpson – 0.9761 OTUs -0.4902
Final FCP <100	Chao1 – 0.6733 Shannon – 1.0 Simpson – 0.7518 OTUs – 0.6541	Chao1 – 0.9600 Shannon – 0.9600 Simpson – 0.6879 OTUs – 0.9800
Final FCP <250	Chao1 – 0.1910 Shannon – 0.3428 Simpson – 0.3679 OTUs -0.1823	Chao1 – 0.7709 Shannon – 0.7709 Simpson – 0.7341 OTUs – 0.8271
FCP >75% Drop	Chao1 – 0.5005 Shannon – 0.9785 Simpson – 0.8928 OTUs – 0.5353	Chao1 – 0.4218 Shannon – 0.2482 Simpson – 0.2917 OTUs – 0.3660
Glucocorticoids	Chao1 – 0.5851 Shannon – 0.6822 Simpson – 0.4528 OTUs – 0.5850	Chao1 – 0.2396 Shannon – 0.5775 Simpson – 0.2396 OTUs – 0.2826
Sex	Chao1 – 0.7709 Shannon – 0.8843 Simpson – 0.8461 OTUs -0.7159	Chao1 – 0.7646 Shannon – 0.9816 Simpson – 0.8 OTUs -0.8178
Age	Chao1 – 0.7237 Shannon – 0.7386 Simpson – 0.6934 OTUs -0.6821	Chao1 – 0.4749 Shannon – 0.7447 Simpson – 0.7315 OTUs -0.5311

***Legend**: Diversity analyses in pCD fecal microbiota. Chao1 and OTU numbers indicate microbial richness while Shannon H and Simpson are indices of microbial diversity accounting for richness, abundance, and evenness. Bold = $p<0.1$*

There was a trending increase in Shannon H and Simpson diversity indices as well as richness as measured by Chao1 and OTU numbers from baseline to pre-dose five fecal microbiota (**Figure 10**). Shannon H increased from a mean of 6.8 (95% C.I. 6.4-7.2) to a mean of 7.1 (95% C.I. 6.8-7.4) after IFX treatment. Chao1 increased from a mean of 278.2 (95% C.I. 213.5-343) to a mean of 347.8 (95% C.I. 292.1-403.6) after IFX treatment. Simpson diversity increased from a mean of 0.985 (95% C.I. 0.981-0.989) to 0.988 (95% C.I.0.986-0.991) after IFX treatment. OTUs increased from a mean of 266.9 per specimen (95% C.I. 205.8-328.1) to a mean of 329.8 per specimen (95% C.I. 277.9-381.7).

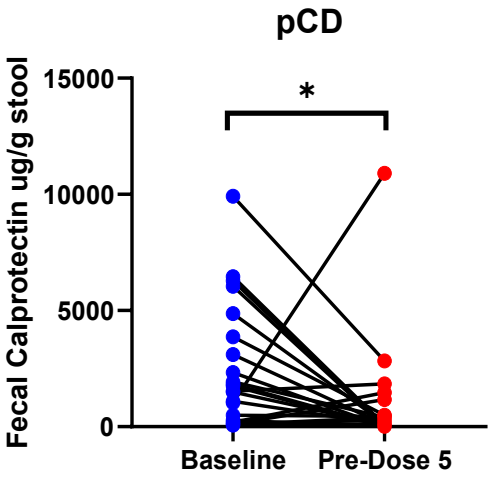
Figure 10 *Trending increases in microbiota diversity with IFX treatment in pCD.*



3.3.1.5 FCP

Fecal calprotectin decreased significantly from baseline to pre-dose five in pCD (p=0.0225), from an average of 2965.8 µg/g stool (C.I. 1169.3-4762.3) at baseline to 981.1 µg/g stool (C.I. 111.1-1851.1) at dose five (**Figure 11**).

Figure 11 *FCP decreases with IFX treatment in pCD.*



Legend: *Pairwise analysis, n=24 pairs*

3.3.2 pUC

3.3.2.1 Urine Metabolites

Sixty-two compounds were identified on GC-MS organic-acids assay by TMIC. Twenty-one out of 62 compounds were below the LOD in ≥50% of specimens, so these compounds were excluded from analysis, leaving 41 compounds. Two-way PERMANOVAs were run where sample size permitted, there were no significant interactions between age or sex and any other variable that altered the significance or trending result. As shown in **Table 26**, disease location, glucocorticoid use, and patient age were associated with significantly altered metabolite profiles. Associations with variables are discussed in respective sections below; IFX therapy was not associated with an altered urinary metabolite profile in pUC.

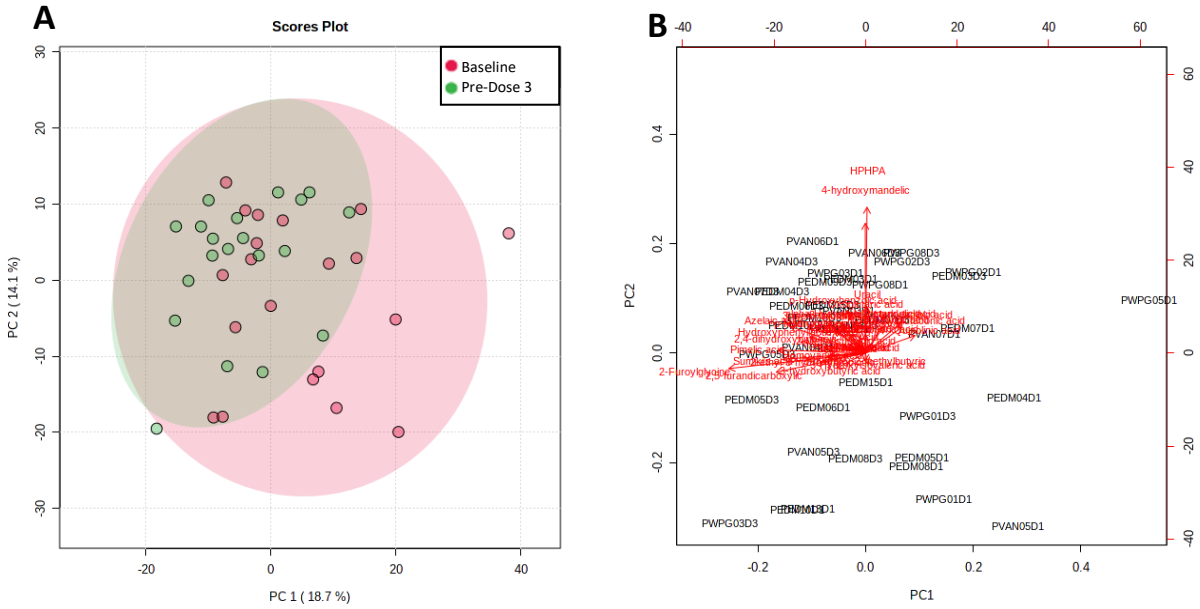
Table 26 *Disease location and glucocorticoid use are associated with significantly different urinary metabolite profiles*

Variable	Baseline	Pre-Dose 3	Both Timepoints	Deltas
Baseline vs. Pre-Dose 3			0.6934	
Pancolitis vs Non-pancolitis	**0.0497**	**0.0007**$^{+}$	**0.0002**$^{+}$	0.3006
Disease Severity$^{\blacklozenge}$	0.8671	0.9065	0.5806	0.7815
Glucocorticoids	**0.0184**	0.8216	0.2116	**Baseline: 0.0091** Pre-dose 3: 0.2149
Remission by Dose 3 (PUCAI <10)	0.0792	0.897		0.0751
PUCAI <10 or drop >20 by Dose 3	0.3756	0.6841		0.7014
Pre-Dose 3 FCP<250	0.1308	0.505		0.2838
≥75% drop in FCP	0.86	0.1239		0.3363
Sex	0.2377	0.574	0.1032	0.94
Age	0.1181	0.2046	**0.0034**$^{+}$	0.99

Legend: *pUC urine metabolite PERMANOVA analysis results.* $^{+}$ *= significant compounds on univariate analysis.* $^{\blacklozenge}$ *= no or mild disease activity vs moderate or severe disease activity. Bold = p<0.1*

Unsupervised analysis of urinary metabolites from both timepoints shown in **Figure 12A** PCA showed no distinct clustering in pUC. Principal component one accounted for 18.7% of variability, and principal component two accounted for 14.1% of variability. Factor loadings for the top three contributing metabolites are shown in **Figure 12C**. The biplot visualizing these factor loadings is shown in **Figure 12B** and demonstrates metabolite influence on specimen PCA plotting.

Figure 12 *Urine Metabolites do not separate by time in the pUC cohort.*



C

Principal Component 1		**Principal Component 2**	
Metabolite	**Factor Loading**	**Metabolite**	**Factor Loading**
2-furoylglycine	-0.52	HPHPA	0.62
2,5-furandicarboxylic acid	-0.34	4-hydroxymandelic acid	0.55
Pimelic acid	-0.33	Uracil	0.20

Legend: ***A****. PCA of pUC urinary metabolites using Euclidean distances. Red = baseline, green = urine collected at IFX dose three.* ***B****. Biplot of Euclidean distance PCA, red text shows metabolites important in accounting for sample variability. Black text are specimens plotted on PCA coordinates.* ***C.*** *Factor loadings of top three metabolites contributing to each principal component as visualized in biplot.*

3.3.2.2 SCFA

Two-way PERMANOVAs were run where sample size permitted; there were no significant interactions between age or sex with any other variable that altered the significance or trending result. Multivariate analysis of fecal SCFA concentrations in pUC showed pre-dose three FCP levels <250, GCS use, and patient sex were associated with significantly different SCFA profiles. FCP <250 and patient sex were significantly associated with SCFA changes over time, while GCS use was only significant when specimens from both timepoints were combined (**Table 27**). Associations with FCP, GCS, and sex are discussed in their respective sections below; IFX therapy was not associated with an altered SCFA profile in pUC.

Table 27 *Final FCP <250, GCS use, and patient sex associated with significantly different SCFA profiles in pUC.*

Variable	Baseline SCFA	Pre-Dose 3 SCFA	Both Timepoints SCFA	SCFA Changes (Deltas)
Baseline vs. Pre-Dose 3			0.2336	
Pancolitis	0.8901	0.7256	0.5964	0.6674
Disease Severity⬧	0.2111	NA	0.973	**Baseline**: 0.4493 **Pre-Dose 3:** NA
Remission by Dose 3 (PUCAI <10)	0.5186	0.2892		0.4564
PUCAI <10 or drop >20 by D3	0.9475	0.6716		0.4612
Pre-Dose 5 FCP<250	0.0752	0.4124		**0.0399**$^{+}$
≥75% drop in FCP	0.3922	0.7258		**0.0411**
Glucocorticoids	0.1534	0.1476	**0.0177**$^{+}$	**Baseline:** 0.3681 **Pre-Dose 3:** 0.5632
Sex	0.**0708**	0.1891	0.8624	**0.0095**$^{+}$
Age	0.3244	0.67	0.8862	0.8735

Legend: *pUC Stool SCFA PERMANOVA analysis results.* $^{+}$ *= significant compounds on univariate analysis. NA = insufficient numbers to perform statistical analysis.* ⬧ *= no or mild disease activity vs moderate or severe disease activity. Bold = p<0.1*

3.3.2.3 16S

Genus Level

Two-way PERMANOVAs were run where sample size permitted; there were no significant interactions between age or sex with any other variable that altered the significance or trending result. Multivariate analysis identified pancolitis, a drop in PUCAI >20, and patient sex to be associated with significantly different microbiota at the genus level in pUC (**Table 28**). IFX

therapy was not associated with altered microbiota at the genus level; associations with other variables are discussed in their respective sections below.

Table 28 *Disease location, drop in PUCAI score, and patient sex significantly associated with microbiota profile in pUC at the genus level.*

Variable	Baseline 16S	Pre-Dose 3 16S	Both Timepoints 16S	Baseline to Dose 3 Changes
Baseline vs. Pre-Dose 3			0.6057	
Pancolitis vs non-pancolitis	0.9109	**0.0669**	**0.047**$^{+}$	0.9934
Disease Severity	0.8727	NA	0.2	**Baseline:** 0.5499 **Pre-Dose 3:** NA
Remission by Dose 3 (PUCAI <10)	0.6454	0.116		0.3436
PUCAI Drop ≥20	0.6418	**0.0466**		0.3127
Final FCP <250	0.6404	0.9304		0.1957
FCP >75% Drop	0.3561	0.4655		0.7862
Glucocorticoids	0.6952	0.7035	0.2267	**Baseline:** 0.8205 **Pre-Dose 3:** NA
Sex	0.1227	0.7973	**0.0918**	0.6239
Age	0.7396	0.6008	0.4412	0.2041

Legend: *Genus level pUC stool microbiota PERMANOVAs. + = significantly different or trending differences in taxa on univariate analysis. NA = too few patients still had moderate or severe disease or were taking GCS to compare groups at the pre-dose three timepoint. Bold = p<0.1.*

Phylum Level

Two-way PERMANOVAs were run where sample size permitted; there were no significant interactions between age or sex with any other variable that altered the significance or trending result. Trending and significant differences on multivariate analysis of phylum levels changes in pUC are detailed in **Table 29**, with significant associations with disease location and severity identified. There were no significant or trending associations identified on univariate analysis at the phylum level. IFX therapy was not associated with altered microbiota at the phylum level; associations with other variables are discussed in their respective sections below.

Table 29 *Disease location and severity associated phylum level microbiota in pUC.*

Variable	Baseline 16S	Pre-Dose 3 16S	Both Timepoints 16S	Baseline to Dose 3 Changes
Baseline vs. Pre-Dose 3			0.1282	
Pancolitis vs Non-pancolitis	0.22	0.3703	**0.0647**	0.3042
Disease Severity	0.2598	NA	**0.0193**	**Baseline:** 0.4489 **Pre-Dose 3:** NA
Remission by Dose 3 (PUCAI <10)	0.7742	0.49		0.8722
PUCAI Drop ≥20	0.3251	0.199		0.9838
Final FCP <250	0.517	0.8779		0.5788
FCP >75% Drop	0.905	0.9184		0.9876
Glucocorticoids	0.803	0.1673	0.5434	**Baseline:** 0.6095 **Pre-Dose 3:** NA
Sex	0.3856	0.4843	0.6113	0.5605
Age	0.1617	0.2747	0.4309	0.1374

Legend: *Phylum level pUC stool microbiota PERMANOVAs. NA = too few patients still had moderate or severe disease or were taking GCS to compare groups at the pre-dose three timepoint.*

3.3.2.4 Microbial Diversity & Richness

As shown in **Table 30**, IFX treatment, patient remission, FCP changes, and patient sex were associated with significantly different stool microbial diversity in pUC. Associations with IFX therapy are discussed here; associations with other variables are discussed in their respective sections below.

Table 30 *IFX treatment, patient remission, FCP changes, and patient sex associated with stool microbial diversity in pUC.*

Variable	Baseline	Pre-Dose 3
Baseline vs. Pre-Dose 3	Paired Analysis: Chao1 – 0.2917 Shannon – 0.1069 **Simpson – 0.0940** OTUs – 0.2263	
Pancolitis vs Non-pancolitis	Chao1 – 0.6682 Shannon – 0.8864 Simpson – 0.5677 OTUs – 0.5677	Chao1 – 0.2008 Shannon – 0.2864 Simpson – 0.2864 OTUs -0.2395
Disease Severity	Chao1 – 0.6434 Shannon – 1 Simpson – 0.8774 OTUs -0.7576	NA
Remission by Dose 3 (PUCAI<10)	Chao1 – 0.4642 Shannon – 0.1432 **Simpson – 0.0404** OTUs -0.4642	Chao1 – 0.4647 Shannon – 0.6015 Simpson – 0.3772 OTUs -0.4633
PUCAI Drop >20	Chao1 – 0.8774 Shannon – 0.7576 Simpson – 0.5371 OTUs – 0.8774	Chao1 – 0.1385 Shannon – 0.9093 Simpson – 0.7324 OTUs – 0.1373
Final FCP <250	Chao1 – 0.1228 Shannon – 0.1649 Simpson – 0.2801 OTUs – **0.0896**	Chao1 – 0.2100 Shannon – 0.2100 Simpson – 0.3051 OTUs – 0.2086
FCP >75% Drop	Chao1 – 0.4751 Shannon – 0.3914 Simpson – 0.3914 OTUs – 0.3914	Chao1 – 0.1356 Shannon – **0.0881** Simpson – 0.2008 OTUs – 0.1645
Glucocorticoids	Chao1 – 0.7751 Shannon – 1.0 Simpson – 0.8864 OTUs – 0.6682	Chao1 – **0.0527** Shannon – 0.4250 Simpson – 0.9093 OTUs – **0.0396**
Sex	**Chao1 – 0.0281** **Shannon – 0.0281** **Simpson – 0.079** **OTUs -0.0192**	Chao1 – 0.6698 Shannon – 0.5224 Simpson – 0.2008 OTUs -0.7484
Age	Chao1 – 0.9391 Shannon – 0.6926 Simpson – 0.4352 OTUs -0.9580	Chao1 – 0.7338 Shannon – 0.7016 Simpson – 0.2767 OTUs -0.8576

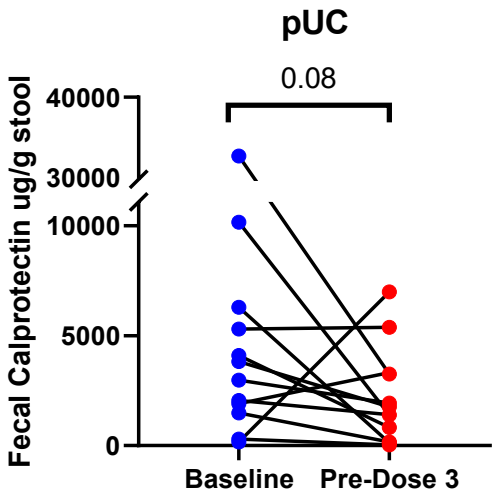
Legend: *Diversity analyses in pUC fecal microbiota. Bold = p<0.1*

There was a trending increase in the Simpson diversity with IFX treatment in pUC from a score of 0.60 at baseline (95%C.I. 0.47-0.72), to 0.78 at pre-dose three (95%C.I. 0.71-0.85) (p value = 0.0940).

3.3.2.5 FCP

There was a non-significant decrease in FCP in pUC from baseline to pre-dose three (p=0.0844), from an average of 5939.9 µg/g stool (C.I. 297.1-11582.7) at baseline to 2216.2 µg/g stool (C.I. 834.7-3597.7) at dose three (**Figure 13**).

Figure 13 *FCP decreases with IFX treatment in pUC.*



Legend: *FCP changes associated with IFX therapy. n=12 pairs. *=p<0.05*

FCP had no significant or trending associations with age, sex, or disease location in pUC.

3.3.3 Comparing and Combining pCD and pUC

3.3.3.1 Urine Metabolites

Two-way PERMANOVAs were run where sample size permitted; there were no significant interactions between age, sex, or antibiotic use with any other variable that altered the significance or trending result. pCD vs pUC status was not found to have a significant interaction with changes over time on two-way PERMANOVA. As shown in **Table 31**, pCD and pUC were associated with significantly different metabolite profiles at baseline, after treatment, and when all timepoints were combined. Urinary metabolites significantly different between pCD and pUC were identified on univariate analysis at baseline, dose 3/5, and when all time points are combined. (**Table 32**, **Table 33**, and **Table 34**, respectively). Although changes with treatment did not reach statistical significance (p=0.0796), when pUC and pCD were combined, pairwise analysis revealed several metabolites significantly different at baseline vs after treatment (**Table 35**).

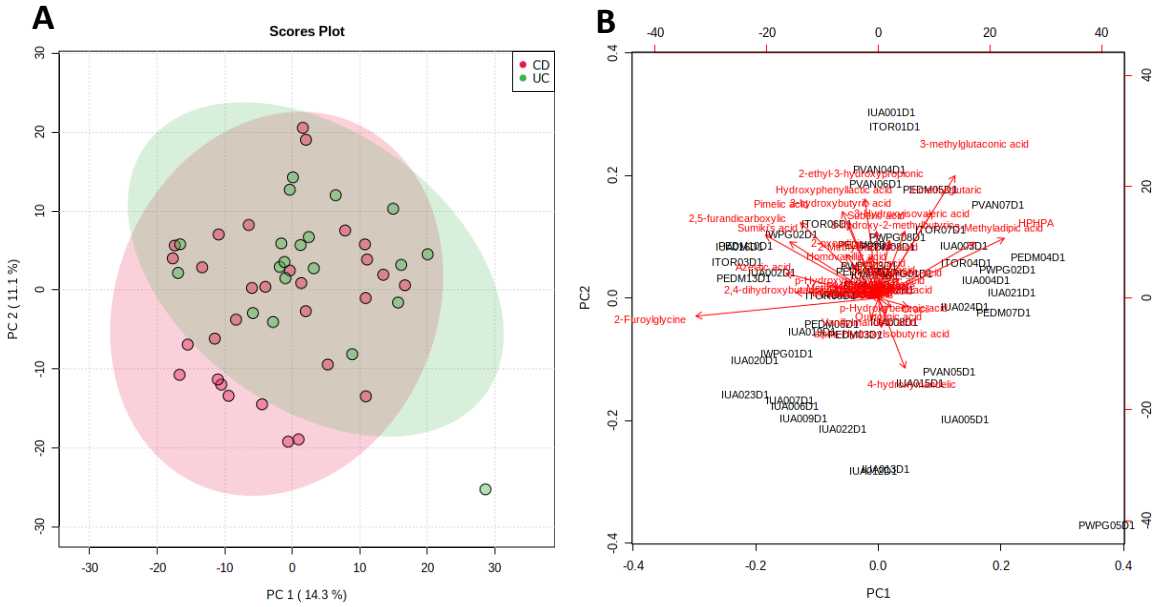
Table 31 *pCD and pUC associated with significantly different urine metabolites and microbiota, while IFX therapy in pIBD is associated with changes in urine metabolites, SCFA, and microbiota.*

Variable	**CD vs UC**			**Baseline vs Dose 3/5 CD & UC Combined**
	Baseline	**Dose 3/5**	**All timepoints combined**	
Urine Metabolites	**0.0027**$^{+}$	**0.0058**$^{+}$	**0.0001**$^{+}$	0.0796^{+}
Urine Metabolite Changes with Treatment (Deltas)			0.8475	
Stool SCFA	0.7296	0.8869	0.7875	**0.0065**$^{+}$
Stool SCFA Changes with Treatment (Deltas)			0.6804	
Fecal Microbiota Genus Level	**0.0129**	**0.0949**	**0.0005**$^{+}$	**0.0002**$^{+}$
Fecal Microbiota Changes with Treatment (Delta) Genus Level			0.6542	
Fecal Microbiota Phylum Level	0.1963	0.195	0.1806	**0.0081**$^{+}$
Fecal Microbiota Changes with Treatment (Delta) Phylum Level			0.1364	

Legend: *Multivariate PERMANOVA, and pairwise analysis p-values comparing pCD and pUC and assessing for IFX associated effects. pCD pre-dose five and pUC pre-dose three were combined to assess the treatment associated effects. No significant interactions were found on two-way PERMANOVA when adjusting for age, sex, and antimicrobial use.* $^{+}$ *= significant compounds on univariate analysis. Bold = p<0.1*

Unsupervised analysis (PCA) of baseline urine metabolites of combined pIBD shown in **Figure 14A** demonstrated no distinct clustering of pUC and pCD at baseline. Principal component one accounted for 14.3% of variability, and principal component two accounted for 11.1% of variability. Factor loadings for the top three contributing metabolites are shown in **Figure 14C.** The biplot visualizing these factor loadings is shown **Figure 14B** and demonstrates metabolite influence on specimen PCA plotting.

Figure 14 *Urine metabolites do not separate pCD and pUC before IFX treatment.*



C

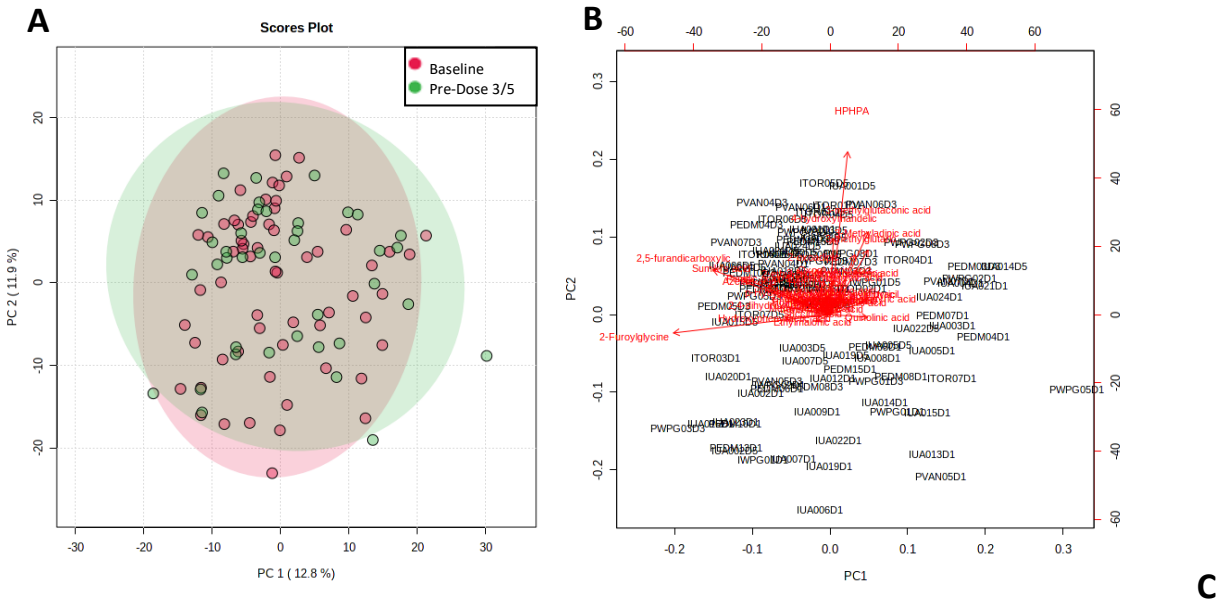
Principal Component 1		Principal Component 2	
Metabolite	**Factor Loading**	**Metabolite**	**Factor Loading**
2-furoylglycine	0.53	3-methylglutaconic acid	0.40
HPHPA	0.37	2-ethyl-3-hydroxypropionic acid	0.33
2,5-furandicarboxylic acid	-0.33	Hydroxyphenyllactic acid	-0.29

Legend*: **A** PCA of pCD and pUC urine metabolites at baseline using Euclidean distances. Red = pCD, Green = pUC. **B**. Biplot of Euclidean distance PCA, red text shows variables important in accounting for sample variability. Black text are specimens plotted on PCA coordinates. **C.** Factor loadings of top three metabolites contributing to each principal component as visualized in biplot.*

Unsupervised analysis (PCA) of urine metabolites at both timepoints of combined pIBD showed no distinct clustering before and after treatment, shown in **Figure 15A**. Principal

component one accounted for 12.8% of variability, and principal component two accounted for 11.9% of variability. Factor loadings for the top three contributing metabolites are shown in **Figure 15.** The biplot visualizing these factor loadings is shown in **Figure 15B** and demonstrates metabolite influence on specimen PCA plotting.

Figure 15 *Urine Metabolites do not separate by time in the combined pIBD cohort.*



Principal Component 1		Principal Component 2	
Metabolite	**Factor Loading**	**Metabolite**	**Factor Loading**
2-Furoylglycine	0.60	HPHPA	0.65
2,5-furandicarboxylic acid	-0.45	3-methylglutaconic acid	0.33
Sumiki's acid	-0.34	4-hydroxymandelic acid	0.31

Legend: ***A*** *PCA of pCD and pUC urine metabolites at both timepoints using Euclidean distances. Red = pCD, Green = pUC.* ***B*** *Biplot of Euclidean distance PCA, red text shows variables important in accounting for sample variability. Black text are specimens plotted on PCA coordinates.*

As shown in **Table 32** and **Table 33**, univariate analysis identified five urinary metabolites at baseline and five after treatment that were significantly different in pUC vs pCD. Citric acid, quinolinic acid, and vanillylmandelic acid were significantly reduced in pUC at both timepoints. 2,4-dihydroxybutanoic acid and 4-deoxythreonate were only reduced in pUC at baseline. 3-hydroxybutyric acid was only increased in pUC after treatment and 4-hydroxymandelic acid was only reduce d in pUC after treatment.

Table 32 *Five urinary metabolites reduced in pUC compared to pCD at baseline.*

Metabolite	pCD Baseline	Upper 95%	Lower 95%	pUC Baseline	Upper 95%	Lower 95%	FDR
Quinolinic acid	27.79	35.38	20.20	6.09	8.88	3.29	1.40E-06
Vanillylmandelic acid	9.47	11.45	7.49	4.44	6.40	2.49	0.0036
2,4-dihydroxybutanoic acid	34.31	45.18	23.44	10.54	15.51	5.56	0.0036
Citric Acid	670.43	870.52	470.33	259.26	355.04	163.48	0.0062
4-Deoxythreonate	30.08	36.78	23.38	17.50	21.54	13.46	0.0196

Legend: *Mean concentrations in µmol/mmol creatinine of urinary metabolites significantly different between pCD and pUC at baseline, only compounds with FDR<0.05 shown here.*

Table 33 *Five urinary metabolites significantly different between pUC and pCD after treatment.*

Metabolite	pCD Pre-Dose 5	Upper 95%	Lower 95%	pUC Pre-Dose 3	Upper 95%	Lower 95%	FDR
Citric Acid	887.17	1164.88	609.47	383.12	532.60	233.64	0.026
Quinolinic acid	14.80	18.89	10.72	6.35	9.39	3.30	0.026
Vanillylmandelic acid	8.97	10.89	7.06	4.81	6.22	3.40	0.026
3-hydroxybutyric acid	8.40	11.79	5.01	33.00	65.06	0.95	0.030
4-hydroxymandelic acid	3.95	4.99	2.92	2.09	3.00	1.18	0.041

Legend: *Mean concentrations in µmol/mmol creatinine of urinary metabolites significantly different between pCD pre-dose five and pUC pre-dose three.*

When combining urine specimens from both timepoints (**Table 34**), citric acid, quinolinic acid, vanillylmandelic acid, 2,4-dihydroxybutanoic acid, 4-hydroxymandelic acid, and 3-deoxythreonate persisted in being significantly different between pCD and pUC. An additional metabolite, 2-ethyl-3-hydroxypropionic acid, was identified as being significantly increased in pUC, that was not identified on separate analysis of each timepoint.

Table 34 *Seven urinary metabolites significantly different between pCD and pUC.*

Metabolite	pCD	Upper 95%	Lower 95%	pUC	Upper 95%	Lower 95%	FDR
Quinolinic acid	21.19	25.66	16.72	6.22	8.19	4.24	2.90E-07
Citric Acid	780.53	949.89	611.17	321.16	408.15	234.17	4.80E-05
Vanillylmandelic acid	9.22	10.55	7.88	4.63	5.78	3.47	4.80E-05
2,4-dihydroxybutanoic acid	30.15	36.38	23.92	14.67	18.33	11.02	0.003
2-ethyl-3-hydroxypropionic acid	4.44	5.66	3.22	9.16	14.43	3.89	0.014
4-hydroxymandelic acid	3.31	4.11	2.52	2.45	4.09	0.81	0.020
4-Deoxythreonate	29.01	32.89	25.13	21.37	25.18	17.55	0.020

Legend*: Mean concentrations in µmol/mmol creatinine of urinary metabolites significantly different between pCD and pUC when specimens from both timepoints were combined for analysis.*

Pairwise analysis of changes with therapy identified three compounds that were significantly changed with IFX therapy in pIBD (**Table 35**).

Table 35 *Three urinary metabolites significantly altered with IFX therapy in pIBD.*

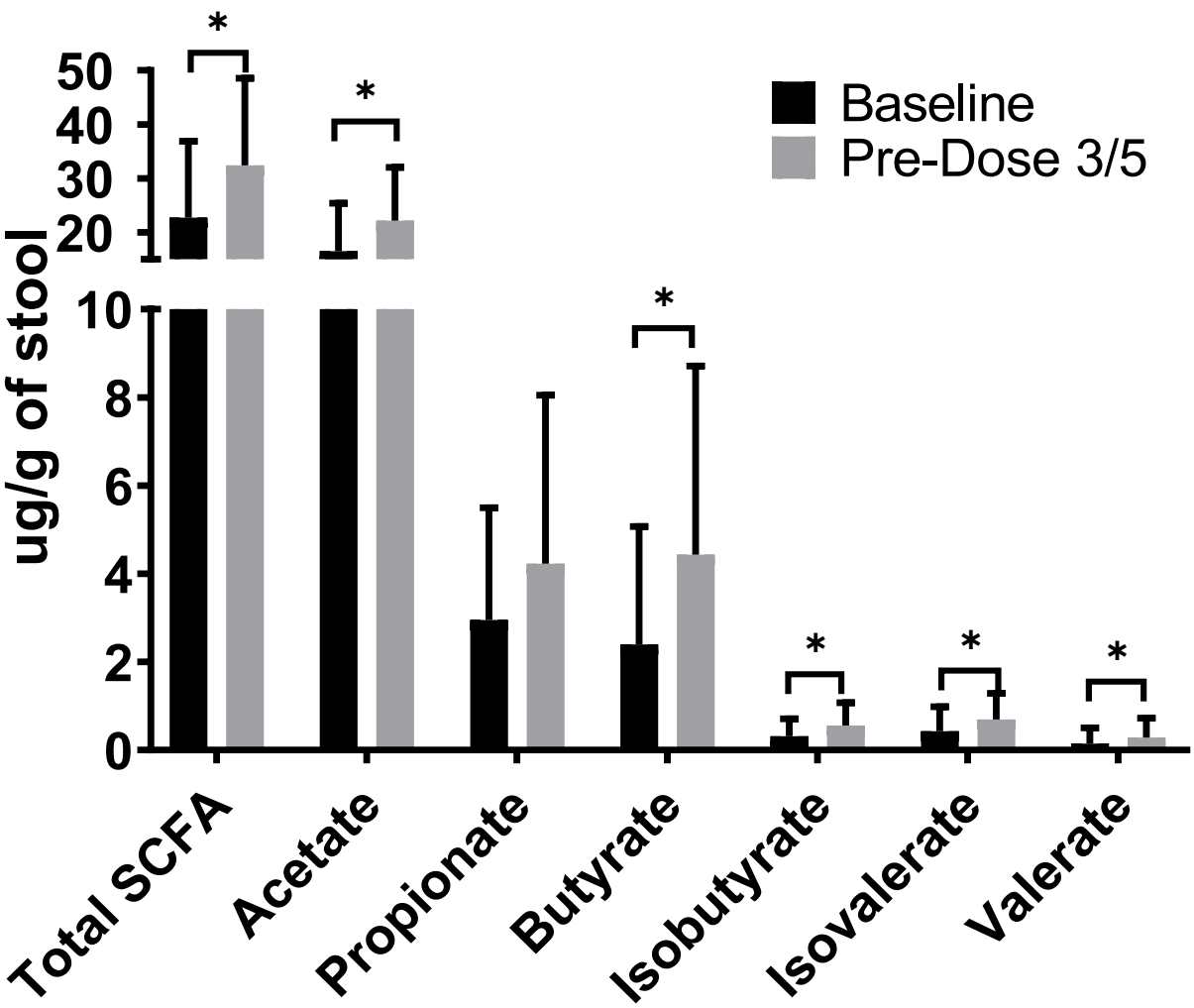
Metabolite	Baseline	Upper 95%	Lower 95%	Pre-Dose 3/5	Upper 95%	Lower 95%	FDR
Oxoglutaric acid ⬆	32.56	43.83	21.29	49.10	65.30	32.90	0.021
Quinolinic acid ⬇	18.50	24.36	12.64	11.47	14.40	8.55	0.021
2,5-furandicarboxylic acid ⬆	1.77	2.44	1.09	4.72	6.70	2.75	0.056

Legend*: Mean concentrations in µmol/mmol creatinine of urine metabolites that changed significantly in pair-wise analysis (n=46 pairs) during IFX therapy in pUC and pCD. pUC pre-dose three and pCD pre-dose five were combined in the after-treatment group as pre-dose 3/5. Only significantly altered (FDR<0.05) metabolites shown.*

3.3.3.2 SCFA

Two-way PERMANOVAs were run where sample size permitted; there were no significant interactions between age, sex, or antibiotic use with any other variable that altered the significance or trending result. pCD or pUC status was not found to have a significant interaction with changes over time on two-way PERMANOVA. Multivariate analysis, shown in **Table 31**, showed before and after IFX treatment in pIBD was associated with a significantly different SCFA profile. As shown in **Figure 16**, SCFA were significantly increased with IFX therapy with the exception of propionate, which showed a non-significant increase.

Figure 16 *SCFA concentration increases with IFX in pIBD.*



Legend*: *=p<0.05.*

3.3.3.3 16S

Genus Level

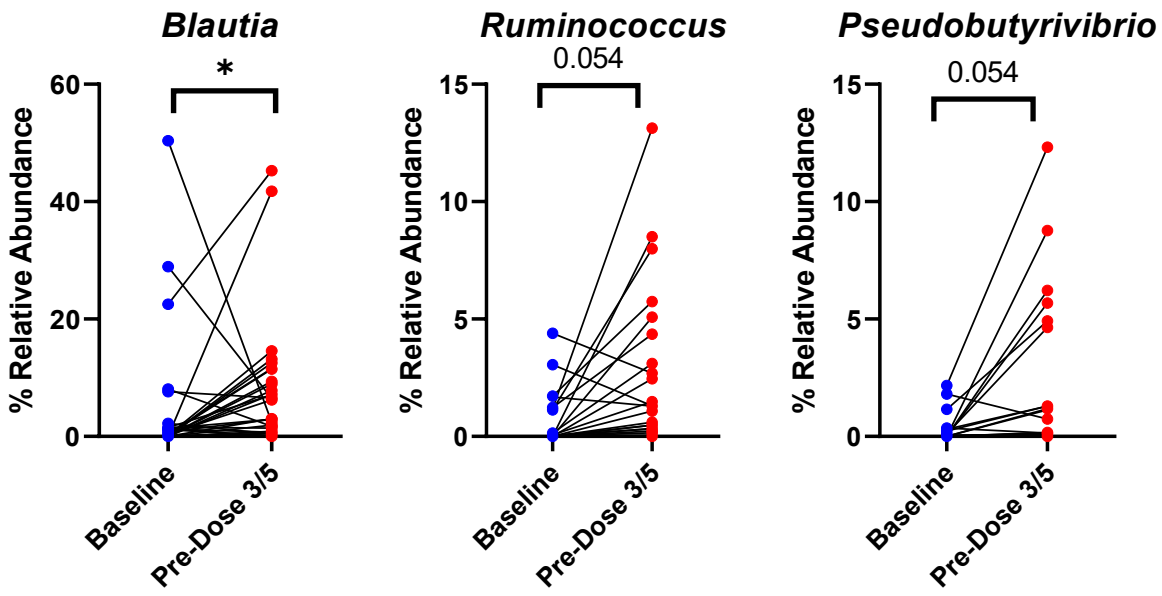
Two-way PERMANOVAs were run where sample size permitted; there were no significant interactions between age, sex, or antibiotic use with any other variable that altered the significance or trending result. pCD or pUC status was not found to have a significant interaction with changes over time on two-way PERMANOVA. There were significantly different

microbiota profiles between pUC and pCD at all timepoints and when timepoints combined (**Table 31**). There was also a significant difference in microbiota at baseline compared to after treatment with pCD and pUC were combined.

Combining specimens from both timepoints in analysis showed one significantly altered taxa at the genus level between pCD and pUC: *Faecalibacterium* had an average relative abundance of 24.74 % (95%C.I. 13.68-35.79%) in pUC and only 8.58% (95%C.I. 4.54-12.62%) in pCD (p value = 0.00016, FDR = 0.0089).

Genera significantly associated or trending changed with IFX therapy when pCD and pUC are combined are visualized in **Figure 17**.

Figure 17 *Microbiota changes with IFX treatment in pIBD.*



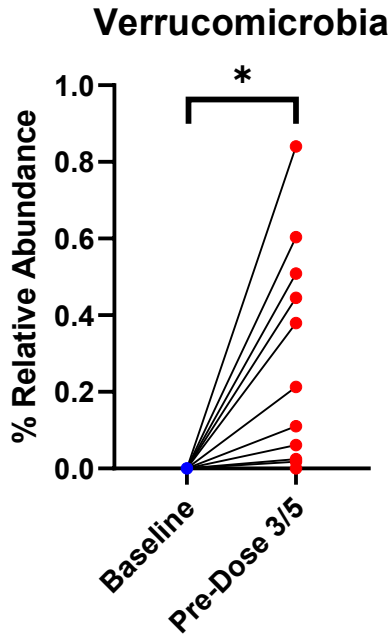
Legend: *Genera with significant or trending changes on Wilcoxon rank sum paired analysis, associated with IFX therapy when pUC and pCD are combined. pUC dose three and pCD dose five were combined in the paired analysis.*

Phylum Level

No significant interactions were found on two-way PERMANOVA with age, sex, and Antimicrobial use and timepoint or pCD vs pUC. No significant interactions were found on two-way PERMANOVA between baseline vs Dose 3/5 and pCD vs pUC status. IFX treatment was associated with significant microbiota changes at the phylum level, but no significant microbiota differences were seen between pCD and pUC at any timepoint (**Table 31**).

Verrucomicrobia was the only significantly altered phylum on univariate pairwise analysis. Shown in **Figure 18**, only one pCD and one pUC patient had Verrucomicrobia detectable in their stool at baseline. Average relative abundance increased from 0.00003% at baseline to an average of 0.097% after treatment with IFX (p-value = 0.0028, FDR = 0.017).

Figure 18 *Verrucomicrobia increased with IFX treatment in pIBD patients.*



Legend*: Phylum level changes with IFX therapy when pUC and pCD are combined. *=p<0.05.*

3.3.3.4 Microbial Diversity & Richness

There were no significant or trending differences pCD vs pUC in microbial diversity of richness (**Table 36**). There was a significant association with diversity and richness in IFX therapy when pCD and pUC were combined for analysis (**Table 36**).

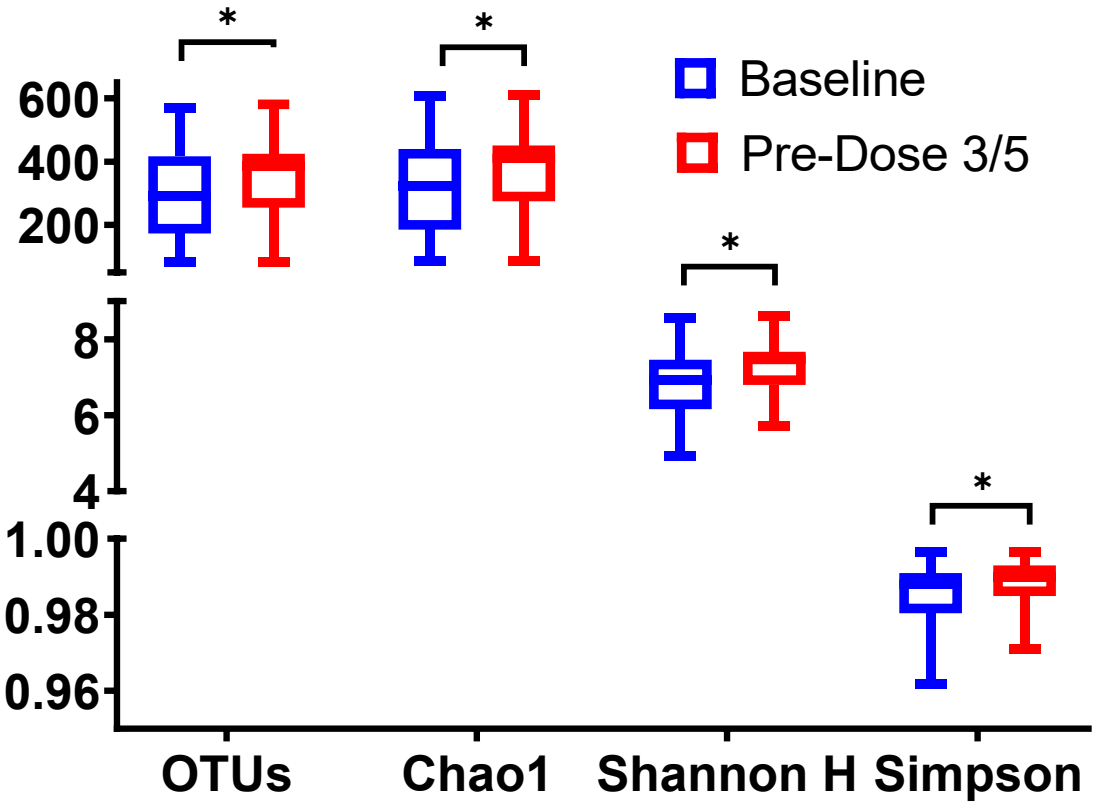
Table 36 *Stool microbial diversity changes with IFX therapy in pIBD.*

Variable	**CD vs UC**		**Baseline vs Dose 3/5 CD & UC Combined**
	Baseline	**Dose 3/5**	
Fecal Microbiota	Chao1 – 0.1102 Shannon – 0.8822 Simpson – 0.8536 OTUs – 0.1102	Chao1 – 0.4073 Shannon – 0.1358 Simpson – 0.1961 OTUs -0.3199	**Chao1 – 0.0217** **Shannon – 0.0121** **Simpson – 0.0155** **OTUs – 0.0183**

Legend*: Diversity comparisons in pIBD. Bold = p<0.1*

Pairwise analysis of stool microbiota diversity measures, shown in **Figure 19**, demonstrate an increase with IFX treatment in pIBD when pUC and pCD are combined. Mean OTUs increased from 296.8 (95% C.I. 247.9-345.8) per specimen at baseline to 352.3 (95% C.I. 311.2-393.4) per specimen after IFX treatment. Richness as measured by Chao1 increased from a mean of 312.1 (95% C.I.259.9-364.3) at baseline to a mean of 371.3 (95% C.I. 327.6-415.1) after treatment. Shannon H diversity increased from a mean of 6.8 (95% C.I.6.5-7.1) at baseline to a mean of 7.2 (7.0-7.5) after treatment. Simpson diversity increased from a mean of 0.985 (95% C.I. 0.982-0.988) at baseline to a mean of 0.989 (95% C.I.0.987-0.991) after IFX treatment.

Figure 19 *Stool microbial diversity increases with IFX therapy in pIBD.*



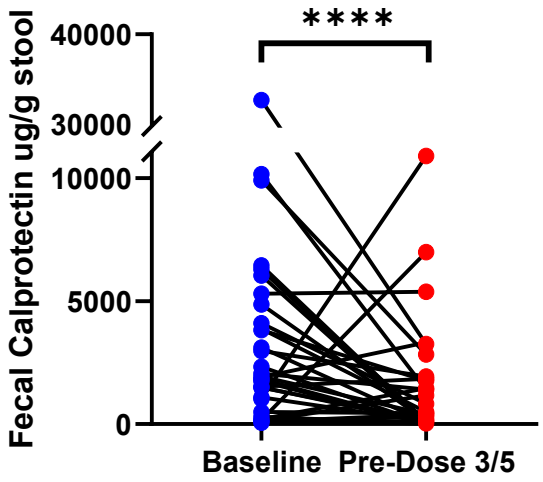
Legend: *Pairwise analysis of changes in diversity measures with IFX therapy in pUC and pCD. *=p<0.05*

3.3.3.5 FCP

Combining pCD and pUC, there was a significant decrease in FCP with IFX therapy from baseline (5255 µg/g stool) to pre-dose three/five (1480 µg/g stool), p<0.0001 (**Figure 20**). FCP levels were not significantly different between pCD and pUC at baseline but were significantly different when comparing pCD pre-dose five to pUC pre-dose three (pCD 1001 µg/g stool vs pUC 2436 µg/g stool, p=0.0081). Differences may represent the timing of IFX treatment, with pCD patients having received four IFX doses and pUC only two. When all

timepoints were combined there was no significant difference between pCD and pUC FCP (p=0.1442).

Figure 20 *FCP decreases with IFX treatment in pIBD.*



Legend: *FCP changes associated with IFX therapy. pCD n=24 pairs, pUC n=12 pairs. ****=p<0.0001.*

3.3.4 Preliminary pCD Integrative Models

Integrative modelling can attempt to distinguish potentially microbially related metabolites from endogenous metabolites, by assessing for correlations between 16S, SCFA, FCP, and urine metabolomics datasets. These associations can then be examined to help identify the role of the intestinal microbiome in response to IFX treatment. The addition of the dietary data for the pCD IFX cohort is in progress and currently undergoing ethics approval; FFQs were completed for some patients as they were recruited through CIDsCaNN but dietary data was not included on the original IDeaL ethics approval.

There were no significant or trending associations between SCFA concentrations and FCP at any timepoint in either pUC or pCD. FCP was assessed for correlations with pCD urinary

metabolites that were significantly altered with IFX therapy. As shown in **Table 37**, of the 12 metabolites altered with IFX therapy in pCD (**Table 18**), changes with treatment (deltas) for six of them had significant negative associations with changes in FCP with treatment (FCP delta). There were no significant correlations with FCP concentrations when baseline or pre-dose five metabolites were assessed independently.

Table 37 *Changes in pCD urinary metabolites correlate significantly with FCP changes.*

Urinary Metabolite	Correlation Coefficient	p-value
3-methylglutaconic acid	-0.03883	0.0053
4-hydroxyhippuric acid	-0.323	0.0221
4-hydroxymandelic acid	-0.2937	0.0384
Glycolic acid	-0.3083	0.0294
HPHPA	-0.2937	0.0355
Oxoglutaric acid	-0.4707	0.0006

Legend*: Significant spearman correlations between pCD urinary metabolites that were significantly altered with IFX therapy. Six of the metabolites significantly associated with IFX therapy had significant associations with FCP.*

Further integrative modelling will be completed when 16S validation studies are completed, and as FFQs become available for the IFX pCD cohort.

3.4 Remission Associations in pCD

Patients that achieved remission as defined by a wPCDAI score of less than or equal to 12.5 did not have distinct metabolite, SCFA, or microbial profiles at baseline. After treatment there was a trending difference in urine metabolites on multivariate analysis as shown in **Table 38**; there were no significantly associated metabolites were identified on univariate analysis. There were also no significant differences in diversity of richness scores between patients that achieved a wPCDAI (**Table 39**) Trending metabolites are summarized in **Table 40**.

Table 38 *CD Urine Metabolite PERMANOVA Analyses Results.*

Variable	Baseline	Pre-Dose 5	Both Timepoints	Deltas
Remission by Dose 5 (wPCDAI ≤12.5) Urine Metabolites	0.56	**0.06**		0.10
Remission by Dose 5 (wPCDAI ≤12.5) SCFA	0.88	0.92		0.9289
Remission by Dose 5 (wPCDAI ≤12.5) Genus level	0.8813	0.3182		0.4756
Remission by Dose 5 (wPCDAI ≤12.5) Phylum Level	0.807	0.6765		0.4292

Legend*: Bold = p<0.1*

There were no significant or trending associations with microbial diversity or richness and pCD patient remission (**Table 39**).

Table 39 *Diversity analyses in pCD fecal microbiota.*

Variable	Baseline	Pre-Dose 5
Remission by Dose 5 (wPCDAI ≤12.5)	Chao1 – 0.4528 Shannon – 0.8378 Simpson – 0.5390 OTUs -0.5390	Chao1 – 0.5288 Shannon – 0.7882 Simpson – 0.9761 OTUs -0.4902

There were no statistically significant differences in urinary metabolite concentrations in patients who achieved remission by dose five compared to those who did not achieve remission. Eight metabolites demonstrated trending but non-significant increases in concentration, as shown in **Table 40**.

Table 40 *Eight urinary metabolites trending increased in patients in pCD patients in remission.*

Metabolite	No Remission n=6	Upper 95%	Lower 95%	Remission n=25	Upper 95%	Lower 95%	FDR
2-Methylsuccinic acid	1.07	1.53	0.62	2.28	2.86	1.69	0.061
Homovanillic acid	3.04	4.36	1.73	5.94	7.39	4.50	0.061
trans-Aconitic acid	19.60	31.61	7.59	38.71	46.70	30.72	0.061
Oxoglutaric acid	21.52	36.17	6.87	58.84	75.91	41.77	0.061
Glycolic acid	78.71	105.34	52.08	149.64	191.05	108.23	0.061
Citric Acid	399.25	883.00	-84.50	1004.28	1323.59	684.96	0.061
Ethylmalonic acid	0.88	1.43	0.33	2.57	3.52	1.62	0.062
Methylmalonic acid	1.24	1.88	0.59	2.41	3.022	1.79	0.065

Legend: *Mean concentrations in µmol/mmol creatinine of pre-dose five urinary metabolites trending an increase in pCD patients who achieved remission by pre-dose five (wPCDAI ≤12.5). Only metabolites with FDR<0.1 shown here.*

When remission is defined as a pre-dose FCP <250, there was an association on multivariate analysis with 16S sequencing at the phylum level in pCD **(Table 23)**. No significant differences in relative abundances at the phylum level were identified on univariate analysis.

3.5 Remission Associations in pUC

There were no significant associations with metabolites, SCFA, or microbial abundances when remission was defined as a PUCAI score of less than 10 in pUC. When remission was assessed as an FCP <250 by pre-dose three there was a significant association in SCFA changes over time, as shown in **Table 41**. There was also a trend for patients that would achieve an FCP <250 by pre-dose three to have a higher number of OTUs in baseline stool, shown in **Table 42**.

Table 41 *Remission in pUC is associated with a significantly different SCFA profile.*

Variable	Baseline	Pre-Dose 3	Both Timepoints	Deltas
Pre-Dose 3 FCP<250 Urinary Metabolites	0.1308	0.505		0.2838
Pre-Dose 5 FCP<250 SCFA	0.0752	0.4124		**0.0399**$^{+}$
Final FCP <250 genus level	0.6404	0.9304		0.1957
Final FCP <250 phylum level	0.517	0.8779		0.5788

Legend: *pUC urine metabolite PERMANOVA analysis results.* $^{+}$ *= significant compounds on univariate analysis. Bold = p<0.1*

Table 42 *IFX treatment, patient remission, FCP changes, and patient sex associated with stool microbial diversity in pUC.*

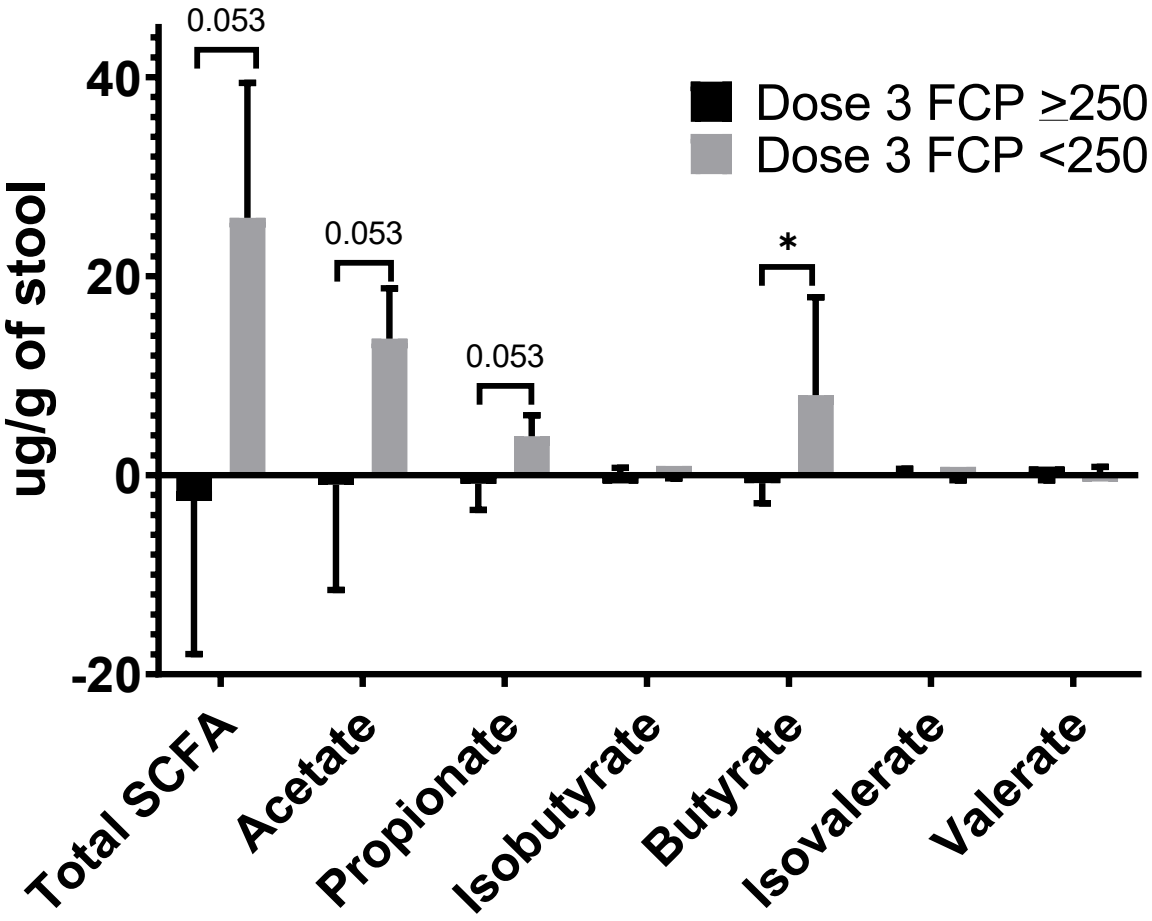
Variable	Baseline	Pre-Dose 3
Final FCP <250	Chao1 – 0.1228 Shannon – 0.1649 Simpson – 0.2801 OTUs – **0.0896**	Chao1 – 0.2100 Shannon – 0.2100 Simpson – 0.3051 OTUs – 0.2086

Legend: *Bold = p<0.1*

Ten pUC patients had stool available for SCFA analysis, four of seven with a final FCP ≥250 were male, while all three with a final FCP <250 were female. This sex difference did not

achieve statistical significance, p=0.167 on Fisher's Exact test. The changes in SCFA associated with achieving an FCP <250 are shown in **Figure 21**.

Figure 21 *Increases in stool SCFA in pUC patients who achieved an FCP<250.*



Legend: *Changes in SCFA with IFX treatment comparing patients who achieve a final FCP <250 (n=3) with those who did not (n=7). *=p<0.05.*

The eight patients who later achieved a PUCAI <10 by dose three demonstrated at baseline a statistically significant but very small reduction in Simpson diversity index, with an average of 0.98 (95%C.I. 0.97-0.99) compared to an average of 0.99 (95%C.I. 0.987-0.994) in the five patients who had a PUCAI score >10 at dose three. The four patients with an FCP <250

μg/g stool by dose three had an average of 266 (95%C.I. 142-390) identified OTUs at baseline, compared with an average of 380 (95%C.I. 267-492) in the baseline stool of the nine patients with an FCP ≥250 μg/g stool at dose three (p value = 0.0896).

Thirteen pUC patients had stool available for 16S analysis, five of nine patients with a final FCP ≥250 were male, while all four with a final FCP M250 were female. This sex difference had a trending significance, p=0.098.

An association with microbial diversity was seen in dose three stool, as the six patients where FCP dropped ≥ 75% from baseline had a trending lower Shannon H diversity at pre-dose three, with an average of 7.3 (95%C.I. 6.6-8.1) compared to an average of 7.8 (95%C.I. 7.5-8.1) in the four patients with a less than 75% drop in FCP (P-value = 0.0881). There was not a similar association between sex and FCP drop ≥75%; three of six patients with a drop <75% were male while two of seven with a drop ≥75% were male (p=0.5).

3.6 Antimicrobial Use Associations in pCD

As expected, antibiotic use was associated with significantly altered SCFA as summarized in **Table 43**, as well as significantly reduced microbial diversity and richness at baseline (**Table 44**).

Table 43 *CD Urine Metabolite PERMANOVA Analyses Results.*

Variable	Baseline	Pre-Dose 5	Both Timepoints	Deltas
Antimicrobial Use Urinary Metabolites	0.214	0.8671	0.1526	**Baseline**: 0.99 **Pre-Dose 5**: 0.38
Antimicrobials SCFA	0.16	0.71	**0.0821**$^{+}$	**At Baseline**: 0.3109 **At Dose 5**: 0.6687
Antimicrobial Use	**0.0304**$^{+}$	0.327	**0.0004**$^{+}$	**Baseline: 0.0513**$^{+}$ **Pre-Dose 5: 0.0616**
Antimicrobial Use Phylum Level	0.8458	0.7933	0.5964	**Baseline**: 0.35 **Pre-Dose 5**: 0.7105

Legend*: $^{+}$ = significant compounds on univariate analysis. Bold = p<0.1*

Table 44 *Antimicrobials are associated with reduced microbial diversity and richness in pCD.*

Variable	Baseline	Pre-Dose 5
Antimicrobials	**Chao1 – 0.0484** **Shannon – 0.0679** Simpson – 0.1034 **OTUs -0.0430**	Chao1 – 0.928 Shannon – 0.1936 Simpson – 0.3160 OTUs – 0.3159

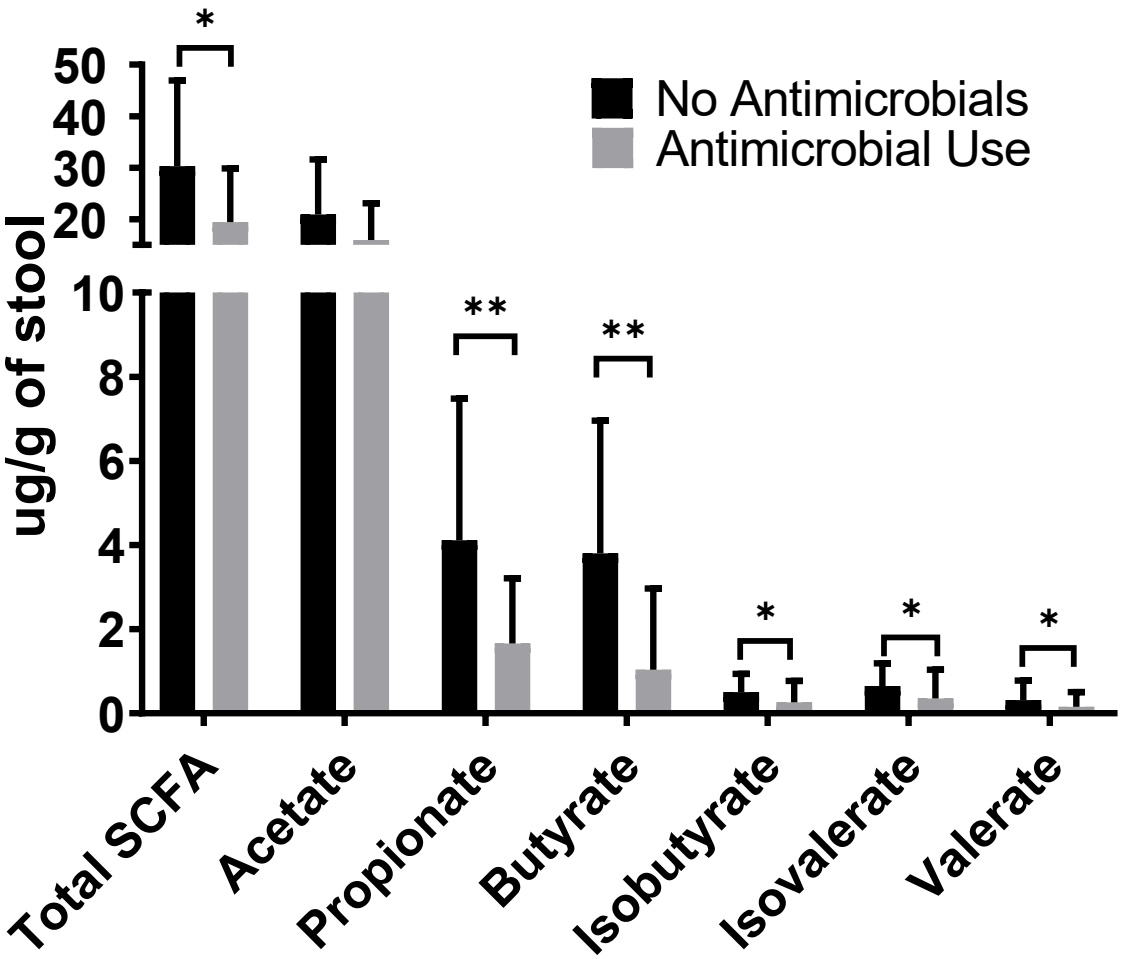
As shown in **Table 45** and **Figure 22**, antimicrobial use was associated with a significant reduction in all SCFA except acetate, which was non-significantly lower in patients with antimicrobial use in pCD.

***Table 45**. Antimicrobial use is associated with lower SCFA concentrations in pCD.*

	SCFA ug/g stool						
	Acetate	**Propionate**	**Isobutyrate**	**Butyrate**	**Isovalerate**	**Valerate**	**Total SCFA**
Antimicrobial Use	0.1645	**0.0054**	**0.0260**	**0.0017**	**0.0163**	**0.0347**	**0.0253**

***Legend**: P-values from univariate analysis of SCFA combining clinical specimens from both timepoints.*

***Figure 22 Lower** fecal SCFA concentrations in pCD associated with antimicrobial use.*



***Legend:** When specimens from both timepoints are combined. Antimicrobial use: n =12, no antimicrobial use: n=40. *=p<0.05. **=p<0.01*

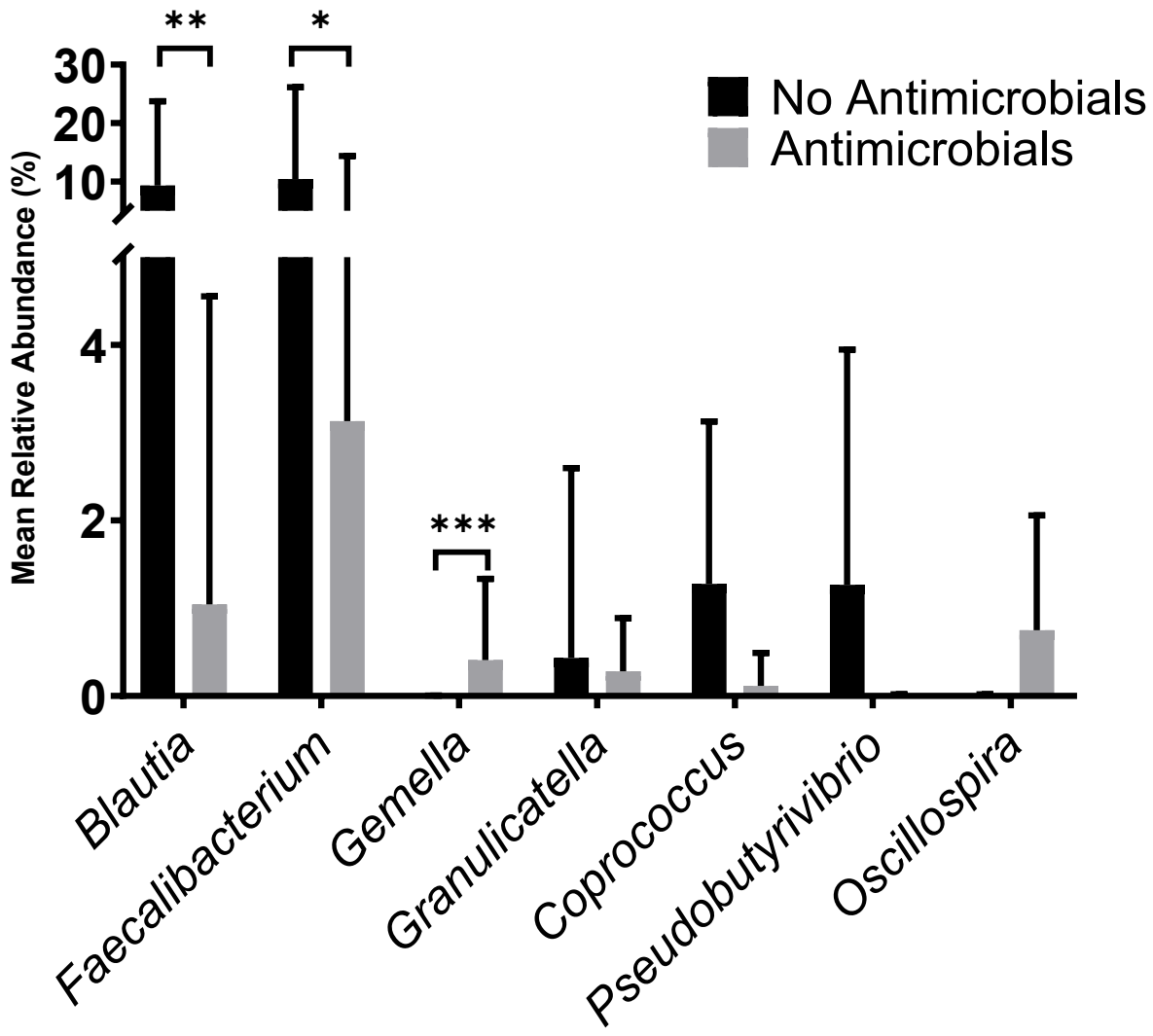
Univariate analysis of the associations between antimicrobial use and relative abundance at the genus level in pCD revealed that, at baseline, *Gemella* had a significantly lower relative abundance in the patients that had taken antimicrobials in the previous 30 days. Patients that had not taken antimicrobials in the previous 30 days had a mean relative abundance of 0.001% (0-0.004% C.I.), while those that had taken antimicrobials had a mean relative abundance of 0.52% (0-1.21% C.I.) (p=5.68E-05, FDR=0.0032352). Analysis of changes over time showed these patients that had taken antimicrobials prior to baseline showed a significant reduction in *Gemella* with IFX treatment, while the no antimicrobial group did not (p=0.0003, FDR=0.019331). Combining specimens from both timepoints reveals six genera that were significantly or trending lower in patients that had taken antimicrobials in the previous 30 days, and one, *Gemella*, that was significantly higher after adjusting for multiple comparisons (**Table 46** and **Figure 23**).

Table 46 *Antimicrobial use linked with genus level changes in pCD.*

Genus	**No Antimicrobials (mean relative abundance, 95%C.I.)**	**Antimicrobials (mean relative abundance, 95% C.I.)**	**P Value**	**FDR**
Gemella	0.001% (0-0.002)	0.41% (0-0.94)	1.63E-05	0.00093
Blautia	9.3% (4.8-13.9)	1.04% (0-3.1)	0.00023	0.0065
Faecalibacterium	10.4% (5.5-15.4)	3.1% (0-9.6)	0.0027	0.031
Granulicatella	0.43% (0-1.1)	0.28% (0-0.63)	0.013	0.084
Coprococcus	1.28% (0.69-1.9)	0.12% (0-0.33)	0.013	0.084
Pseudobutyrivibrio	1.27% (0.42-2.1)	0.004% (0-0.01)	0.013	0.084
Oscillospira	0.008% (0.003-0.01)	0.75% (0-1.5)	0.013	0.084

Legend*: Genera significantly different or trending different for relative abundances when patient had received antimicrobials anytime in the previous 30 days; samples from both timepoints combined for analysis. Only genera with significant or trending differences are shown here.*

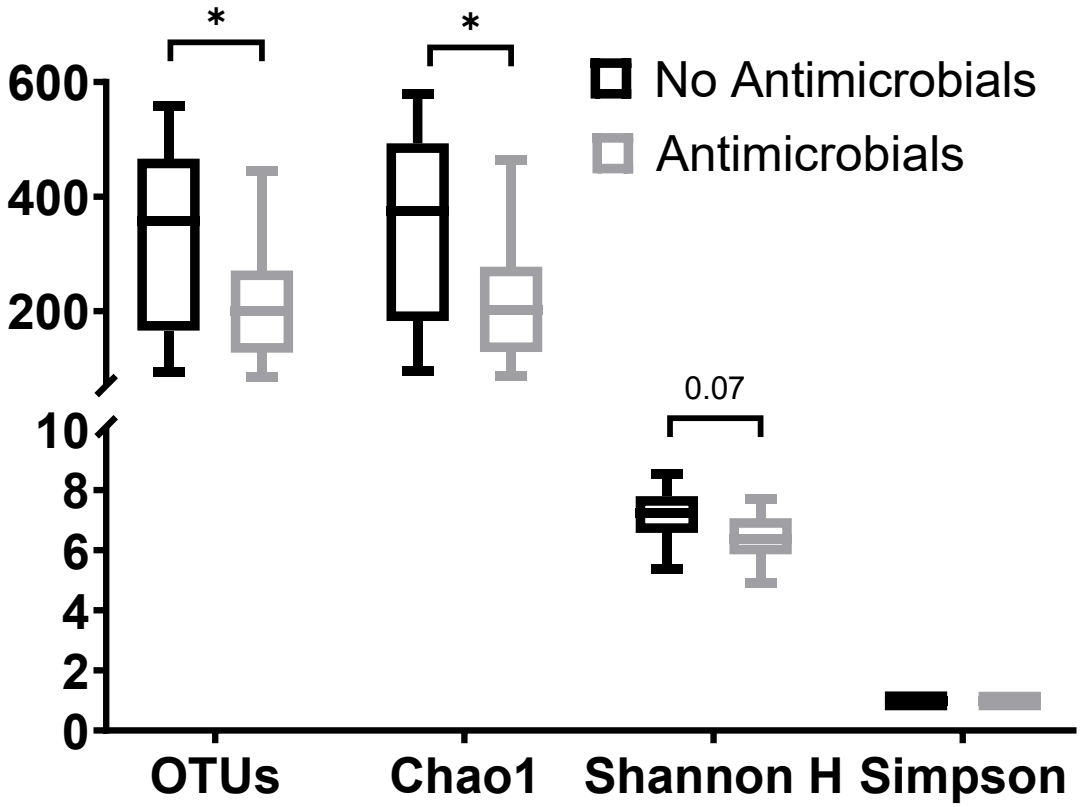
Figure 23 *Seven genera significantly or trending associated with antimicrobial use in pCD.*



Legend*: Genera significantly different or trending different in relative abundances when patients had received antimicrobials anytime in the previous 30 days; samples from both timepoints were combined for analysis. Recent antimicrobial use n=14, no antimicrobial use n=41. *=p<0.05.**=p<0.01. ***=p<0.001.*

Antimicrobial use prior to baseline was associated with significantly reduced microbial diversity and richness in baseline stool specimens, as shown in **Figure 24**.

Figure 24 *Baseline diversity indices significantly reduced with antimicrobial use in pCD.*



Legend: *Baseline stool specimen microbial diversity indices associated with antimicrobial use prior to baseline in pCD. Patients with antimicrobial use n=11, no antimicrobial use n=16.*

3.7 GCS Use Associations in pUC

In pUC, GCS use was significantly associated with urinary metabolites and SCFA in multivariate analysis (**Table 26**). Univariate analysis identified significantly altered SCFA associates with GCS use, summarized in **Table 49** and **Figure 25**. GCS use was assessed for associations at the genus and phylum level as there were consistent findings in metabolites, SCFA and diversity despite there being no significance or trend on multivariate analysis; results are discussed below.

Table 47 *Glucocorticoid use in pUC is associated with significantly different urinary metabolite and SCFA profiles.*

Variable	Baseline	Pre-Dose 3	Both Timepoints	Deltas
GCS Use Urinary Metabolites	**0.0184**	0.8216	0.2116	**Baseline: 0.0091** Pre-dose 3: 0.2149
GCS Use SCFA	0.1534	0.1476	**0.0177**$^{+}$	**Baseline:** 0.3681 **Pre-Dose 3:** 0.5632
GCS Use Genus Level	0.6952	0.7035	0.2267	**Baseline:** 0.8205 **Pre-Dose 3:** NA
GCS Use Phylum Level	0.803	0.1673	0.5434	**Baseline:** 0.6095 **Pre-Dose 3:** NA

Legend: *pUC urine metabolite PERMANOVA analysis results.* $^{+}$ *= significant compounds on univariate analysis. Bold = p<0.1*

A lower number of OTUs were observed in pre-dose three stool of patients on glucocorticoids, and as shown below in **Figure 26** there was also trending lower level of richness as measured by Chao1.

Table 48 *IFX treatment, patient remission, FCP changes, and patient sex associated with stool microbial diversity in pUC.*

Variable	Baseline	Pre-Dose 3
Glucocorticoids	Chao1 – 0.7751 Shannon – 1.0 Simpson – 0.8864 OTUs – 0.6682	Chao1 – **0.0527** Shannon – 0.4250 Simpson – 0.9093 OTUs – **0.0396**

Legend: *Bold = p<0.1*

Clinical characteristics of pUC patients on vs off glucocorticoids at baseline showed no differences in age, sex, likelihood to achieve remission, or disease duration, but the patients on glucocorticoids at baseline did have significantly higher PUCAI scores at baseline compared to those not on glucocorticoids at baseline [54.50 vs 37.50 (p=0.0437)].

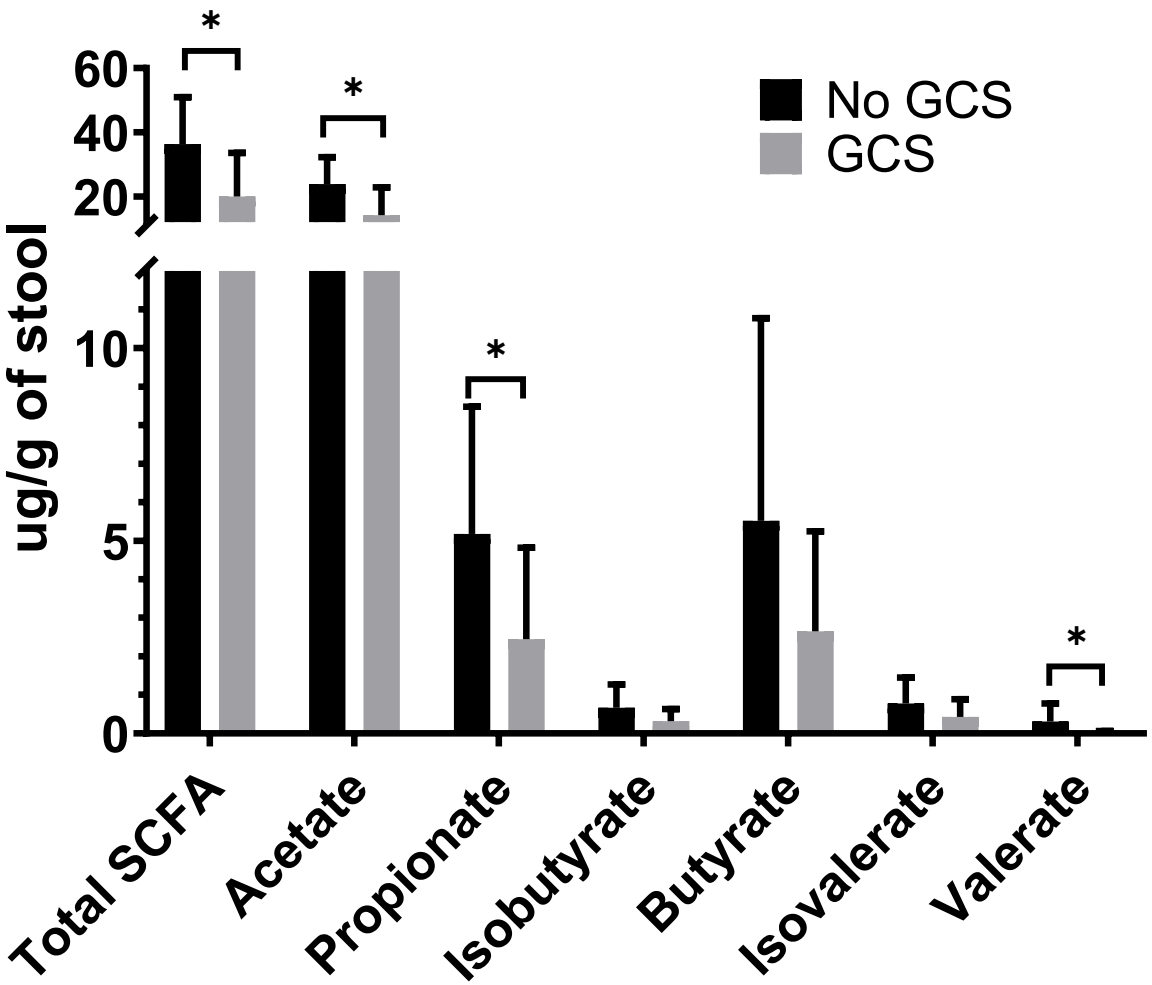
Univariate analysis of SCFA associations with GCS use showed a significant reduction in total SCFA, acetate, propionate, and valerate in patients taking GCS (**Table 49** and **Figure 25**).

Table 49 *GCS use is associated with lower SCFA concentrations in pUC.*

	No GCS	Upper 95%	Lower 95%	GCS	Upper 95%	Lower 95%	P value
Total SCFA	36.33	45.61	27.05	20.07	29.72	10.41	**0.017**
Acetate	23.88	29.18	18.58	14.18	20.42	7.95	**0.021**
Propionate	5.18	7.27	3.08	2.45	4.15	0.76	**0.025**
Valerate	0.66	1.05	0.28	0.32	0.55	0.09	**0.036**
Butyrate	5.52	8.86	2.17	2.65	4.51	0.80	0.16
Isobutyrate	0.78	1.21	0.35	0.43	0.76	0.10	0.22
Isovalerate	0.31	0.61	0.02	0.03	0.06	0.01	0.22

Legend*: Fecal SCFA differences associated with GCS use when specimens from both timepoints are combined for analysis. Samples from patients not taking GCS: n=12, samples from patients taking GCS: n=10. Bold = p<0.1*

Figure 25 *Significantly lower SCFA concentrations in pUC patients taking GCS.*



Legend: *GCS use in pUC is associated with lower fecal SCFA concentrations (when specimens from both timepoints are combined). Samples from patients not taking GCS n=12, samples from patients taking GCS n=10. *=p<0.05.*

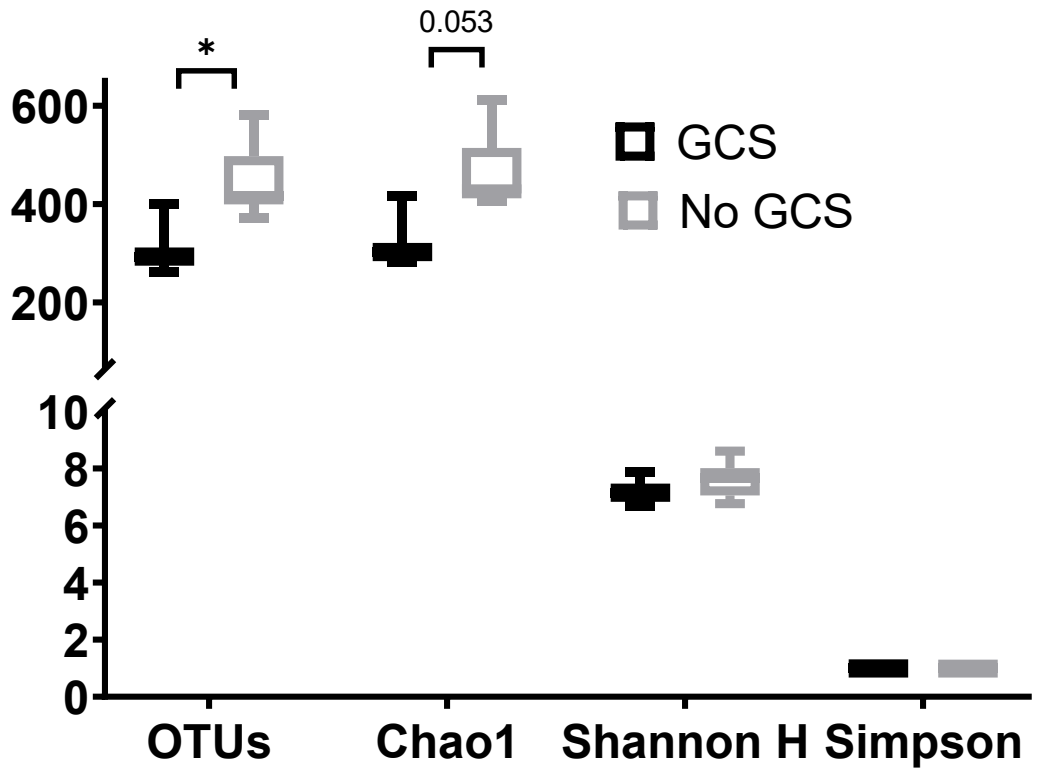
Among patients who submitted specimens analyzed for SCFA, baseline GCS use was not associated with any difference in sex, age, disease duration, average PUCAI scores at either dose, disease location, or likelihood to achieve remission by dose three.

Due to consistent findings associated with GCS in SCFA, urine metabolites, and diversity, GCS use was examined for associations with microbiota at the genus and phylum level

despite not being significant on multivariate analysis. A significantly higher mean relative abundance of *Proteus* in pUC with GCS use was identified when all doses were combined for analysis. The mean relative abundance in patients without GCS use was 0.001% relative abundance (95% C.I. 0-0.004), with only when patient having detectable *Proteus*. In patients with GCS use, *Proteus* was present in three patients, four total specimens, at a mean relative abundance of 0.58% (95% C.I. 0-1.74%). p=0.000862 FDR=0.049. No associations with GCS use were found at the phylum level.

As described above and shown in **Figure 26**, GCS use was associated with a significant reduction in microbial richness as measured by OTU numbers, and a trending reduction in richness as measured by Chao1. There were also non significant reductions in Shannon H and Simpson diversity indices with GCS use.

Figure 26 *Lower levels of richness and diversity associated with GCS use in pUC.*



Legend: *Pre-dose three stool microbiota diversity index comparisons between pUC patients on or off glucocorticoids at the time of sample collection. GCS = glucocorticoids. pUC patients on GCS n=3, off GCS n=7.*

3.8 Disease Severity Associations in pCD

Significant associations between disease severity in pCD were found with urinary metabolites, SCFA, and microbiota at the genus and phylum level (**Table 50**). Univariate analysis identified significantly associated individual SCFA, as well as taxa at the genus and phylum levels.

Table 50 *CD Urine Metabolite PERMANOVA Analyses Results.*

Variable	Baseline	Pre-Dose 5	Both Timepoints	Deltas
Disease Severity♦ Urinary Metabolites	0.46	NA	**0.099**	**Baseline:** 0.14 **Pre-Dose 5:** NA
Disease Severity♦ SCFA	0.77	NA	**0.0708**$^{+}$	**At Baseline**: 0.9901 **Pre-Dose 5:** NA
Disease Severity Genus Level	**0.0665**	NA	**0.0017**$^{+}$	**Baseline:** 0.6312 **Pre-Dose 5:** NA
Disease Severity Phylum Level	0.54	0.97	**0.0145**$^{+}$	**Baseline:** 0.5638 **Pre-Dose 5**: NA

Legend: $^{+}$ *= significant compounds on univariate analysis. NA = insufficient numbers to perform statistical analysis.* ♦ *= no or mild disease activity vs moderate or severe disease activity. Bold = p<0.1*

There were no associations between disease severity in pCD and microbial diversity or richness at baseline, as shown in **Table 51**. There were insufficient numbers of patients with moderate or severe disease after IFX therapy to conduct statistical analysis.

Table 51 *Diversity analyses in pCD fecal microbiota. Bold = p<0.1*

Variable	Baseline	Pre-Dose 5
Disease Severity	Chao1 – 0.4840 Shannon – 0.5598 Simpson – 0.5598 OTUs -0.5402	NA

Univariate analysis including samples from both timepoints identified four metabolites that were significantly lower in pCD patients that had moderate or severe disease (**Table 52**).

Table 52. *Four metabolites significantly lower with moderate or severe disease in pCD.*

Metabolite	None /Mild n=39	Upper 95%	Lower 95%	Moderate /Severe n=22	Upper 95%	Lower 95%	FDR
Glycolic acid	140.84	170.23	111.45	77.66	98.22	57.10	0.0041
3-hydroxyphenylacetic acid	1.51	2.89	0.13	0.14	0.27	0.011	0.022
4-hydroxymandelic acid	3.89	4.81	2.97	2.29	3.79	0.80	0.037
2,5-furandicarboxylic acid	5.67	8.17	3.16	1.52	2.46	0.57	0.043

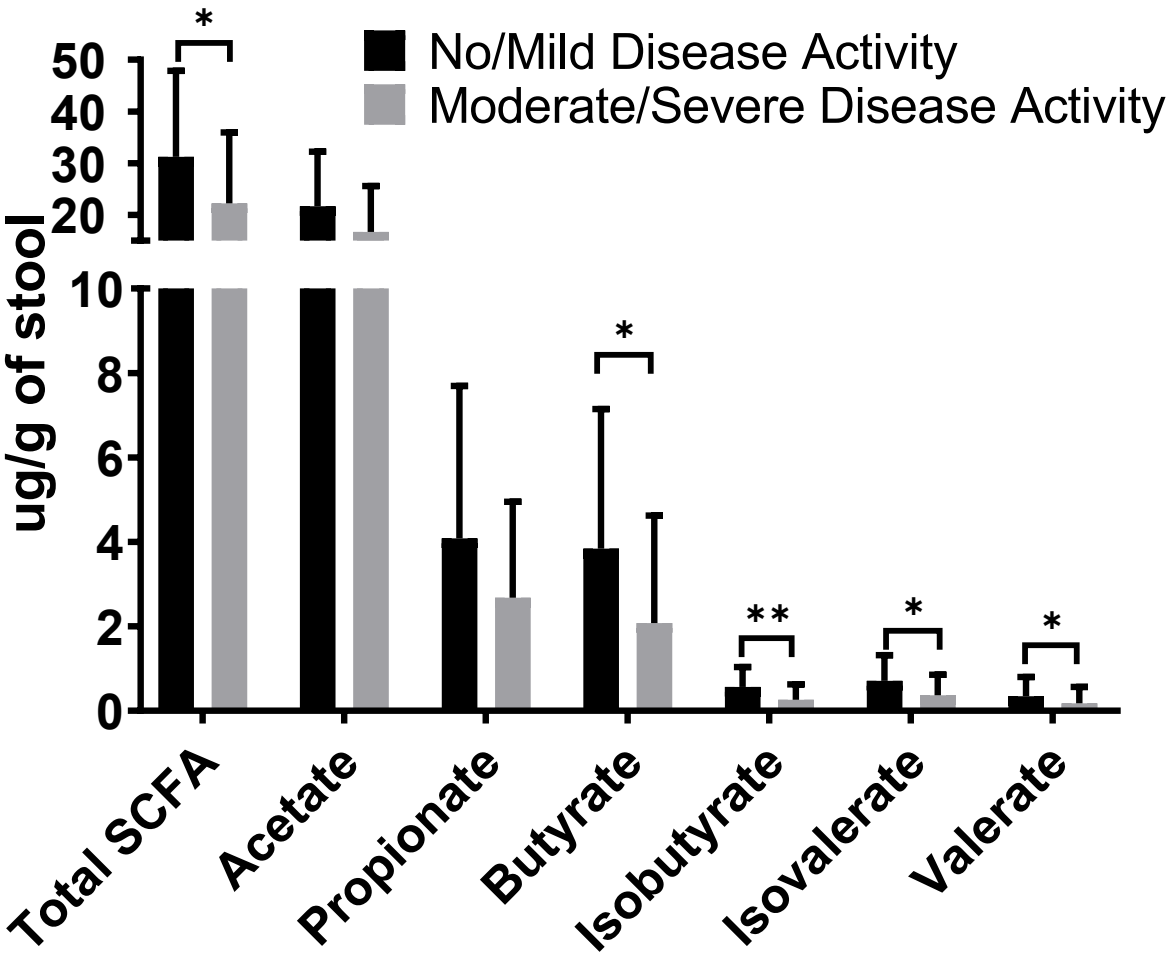
Legend: *Mean concentrations in µmol/mmol creatinine of urinary metabolites, only metabolites found to be significantly different (FDR<0.05) on univariate analysis are shown. Results when specimens from both timepoints were combined.*

Mean concentrations of SCFA were reduced in stool from patients with moderate or severe disease activity when specimens from both timepoints were combined for analysis (**Table 53**, **Figure 27**).

Table 53*. Increased disease severity is associated with lower SCFA concentrations in pCD.*

	SCFA ug/g stool						
	Acetate	**Propionate**	**Isobutyrate**	**Butyrate**	**Isovalerate**	**Valerate**	**Total SCFA**
Disease Severity	0.0575	0.1583	**0.0074**	**0.0253**	**0.0176**	**0.0252**	**0.0442**

Figure 27 *SCFA are reduced in pCD patients with increased disease activity.*



Legend*: When specimens from both timepoints are combined. Moderate or severe disease activity: n=20, mild or no disease activity: n=32. *=p<0.05. **=p<0.01*

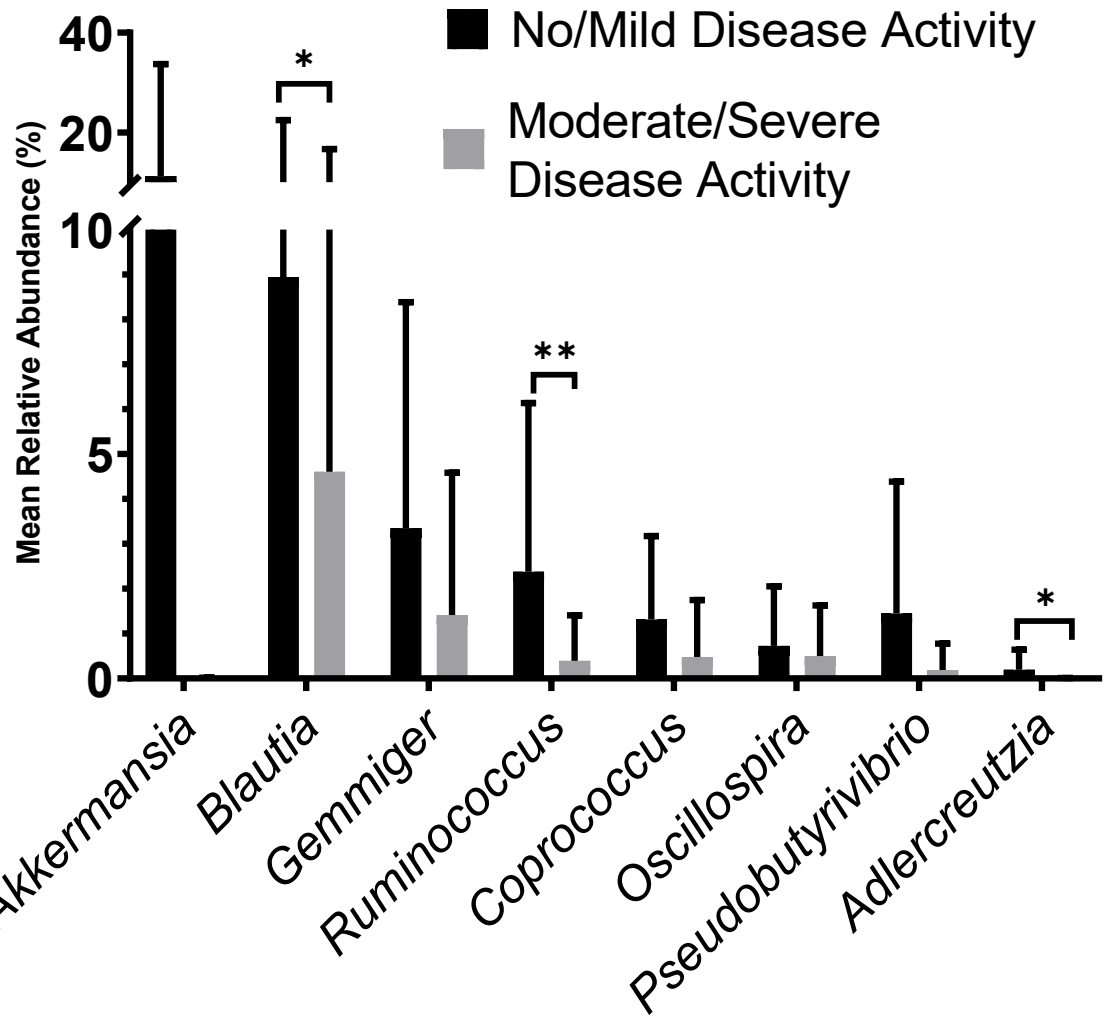
The results of univariate analysis of the association between disease severity and genera abundances are shown in **Table 54** and **Figure 28**, with all significant or trending genera showing a lower mean relative abundance.

Table 54 *Disease activity in pCD patients is associated with fecal genera changes.*

Genus	**P Value**	**FDR**
Ruminococcus	0.00103	0.0210
Blautia	0.00111	0.0210
Adlercreutzia	0.00323	0.0460
Coprococcus	0.00459	0.0524
Oscillospira	0.00939	0.0849
Pseudobutyrivibrio	0.01043	0.0849
Akkermansia	0.01417	0.0991
Gemmiger	0.01565	0.0991

Legend*: Genera significantly different or trending comparing fecal microbiota from pCD patients with no or mild disease activity vs those with moderate or severe disease activity, samples from both timepoints included in analysis. Only genera with significant or trending differences are shown here.*

Figure 28 *Genera with lower relative abundances with increased disease activity in pCD.*



Legend: *Relative abundances of significantly different or trending fecal microbiota from pCD patients with no or mild disease activity (n=33) vs those with moderate or severe disease activity (n=22), samples from both timepoints included in analysis. *=p<0.05. **=p<0.01.*

Phylum level univariate analysis identified Verrucomicrobia as significantly higher in patients with no or mild disease activity (**Table 55**). Verrucomicrobia was absent in all but one baseline specimen with 0.03% relative abundance that had moderate or severe disease. Closer analysis revealed *Akkermansia* as accounting for 100% of the identified Verrucomicrobia; it was

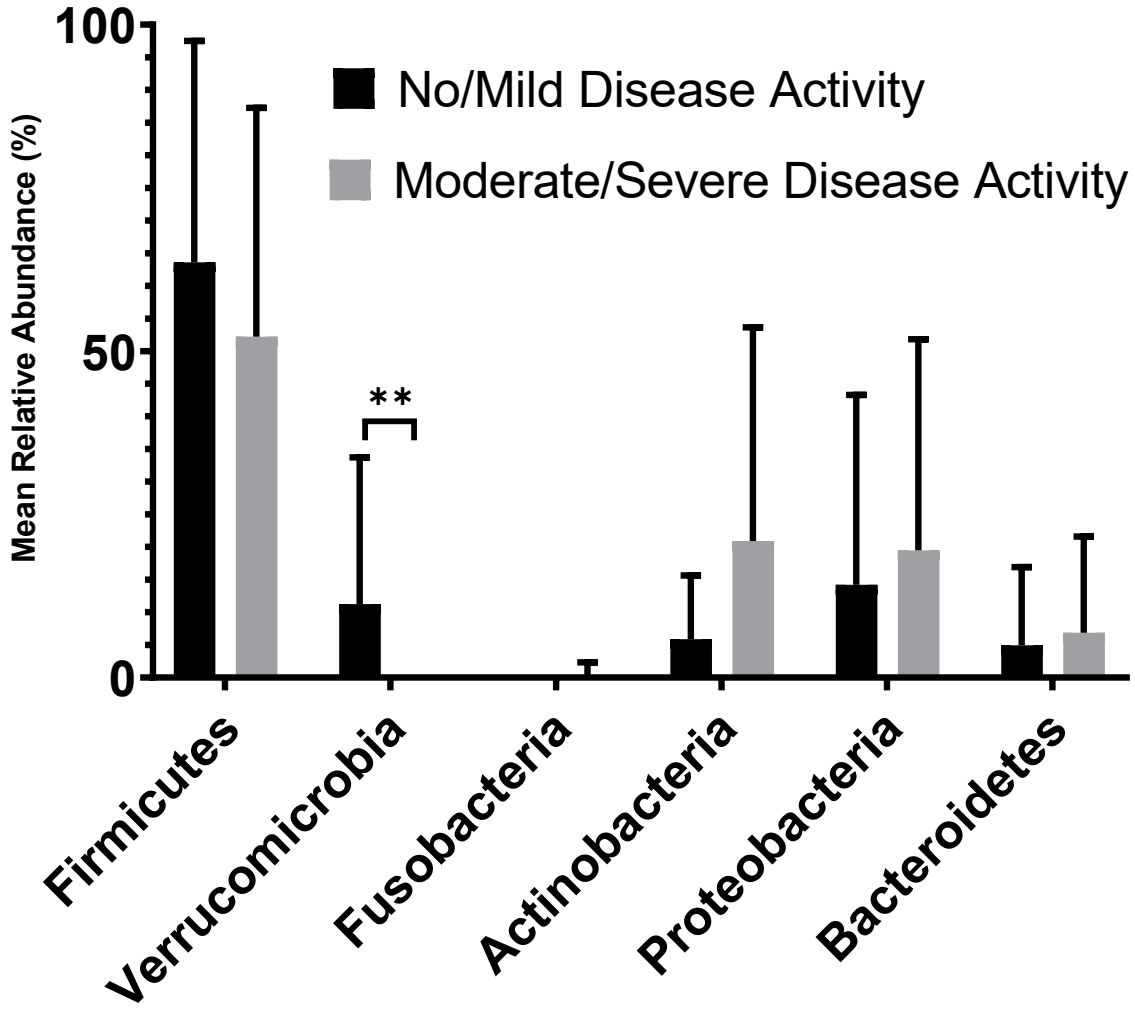
not significant at genus level univariate analysis after adjusted for FDR (p-value = 0.014, FDR = 0.099); *Akkermansia* shown in **Figure 28**, phylum level changes shown in **Figure 29**.

Table 55 *Significantly lower relative abundance of Verrucomicrobia in patients with moderate or severe pCD.*

Phylum	**No/ Mild**	**Upper 95%**	**Lower 95%**	**Moderate/ Severe**	**Upper 95%**	**Lower 95%**	**P Value**	**FDR**
Verrucomicrobia	0.08%	0.14%	0.02%	0%	0%	0%	0.00052	0.0031
Firmicutes	0.65%	0.75%	0.56%	0.55%	0.67%	0.42%	0.129	0.388
Fusobacteria	0%	0%	0%	0%	0.01%	0%	0.286	0.545
Actinobacteria	0.08%	0.11%	0.04%	0.17%	0.27%	0.07%	0.448	0.545
Proteobacteria	0.14%	0.23%	0.06%	0.19%	0.29%	0.08%	0.490	0.545
Bacteroidetes	0.05%	0.08%	0.01%	0.09%	0.15%	0.03%	0.545	0.545

Legend*: Relative abundance phylum level associations with disease severity in pCD when samples from both timepoints were included in analysis.*

Figure 29 *Significantly lower relative abundance of Verrucomicrobia in patients with moderate or severe pCD.*



Legend: *Relative abundance changes at the phylum level associated with disease severity in pCD when samples from both timepoints included in analysis. No or mild disease n=33, moderate or severe disease n=22. **=p<0.01*

As expected, a moderate or severe disease activity in pCD was associated with an increased FCP of 5288 µg/g stool compared to 1194 µg/g stool from patients with no or mild disease activity (p=0.0001). FCP was not significantly different between ICD and CCD in pCD.

3.9 Disease Severity Associations in pUC

Disease severity in pUC was associated significantly with the microbiota at the phylum level on multivariate analyses, but there were no significantly different taxa identified on univariate analysis (**Table 56**).

Table 56 *Disease severity associated with phylum level microbiota changes in pUC.*

Variable	Baseline	Pre-Dose 3	Both Timepoints	Deltas
Disease Severity Urinary Metabolites	0.8671	0.9065	0.5806	0.7815
Disease Severity SCFA	0.2111	NA	0.973	**Baseline**: 0.4493 **Pre-Dose 3:** NA
Disease Severity 16S Genus Level	0.8727	NA	0.2	**Baseline:** 0.5499 **Pre-Dose 3:** NA
Disease Severity 16S Phylum Level	0.2598	NA	**0.0193**	**Baseline:** 0.4489 **Pre-Dose 3:** NA

Legend: *pUC urine metabolite PERMANOVA analysis results. Bold = significant or trending associations. Disease severity classified as no or mild disease activity vs moderate or severe disease activity.*

There were also no significant or trending associations with microbial diversity or richness and disease severity in pUC (**Table 57**).

Table 57 *No diversity or richness associations with disease severity in pUC.*

Variable	Baseline	Pre-Dose 3
Disease Severity	Chao1 – 0.6434 Shannon – 1 Simpson – 0.8774 OTUs -0.7576	NA

Legend*: P-values from Mann-Whitney U test comparisons of diversity and richness by disease severity in pUC.*

3.10 Disease Location Associations in pCD

Multivariate and univariate analyses revealed significant associations with disease location in pCD. Multivariate significance is summarized in **Table 58**, with univariate results below. As shown in **Table 59**, there were no variations in microbial diversity and richness with disease location in pCD. Univariate analysis only found significantly associated metabolites after IFX therapy (pre-dose 5, **Figure 30**). There were no significantly different taxa at the genus or phylum level by disease location in pCD.

Table 58 *Disease location associated with altered metabolites & microbiota in pCD.*

Variable	Baseline	Pre-Dose 5	Both Timepoints	Deltas
ICD vs CCD Urine Metabolites	0.65	**0.0004**$^{+}$	**0.0018**	**0.0019**
ICD vs CCD SCFA	0.36	0.37	0.69	0.3766
ICD vs CCD 16S Genus level	0.5282	**0.097**	**0.0885**	NA
ICD vs CCD 16S Phylum level	0.5279	**0.0444**	**0.0478**	NA

Legend*:* $^{+}$ *= significant compounds on univariate analysis. NA = insufficient numbers to perform statistical analysis. Bold =* $p<0.1$

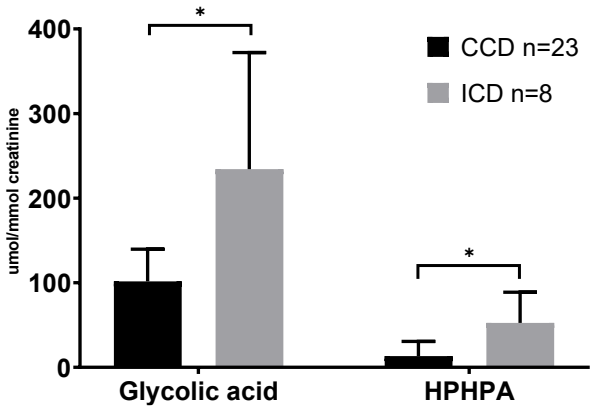
Table 59 *No differences in microbiota diversity or richness with disease location in pCD.*

Variable	Baseline	Pre-Dose 5
ICD vs CCD	Chao1 – 1.0 Shannon – 0.8378 Simpson – 0.7848 OTUs – 0. 9728	Chao1 – 0.1680 Shannon – 0.2122 Simpson – 0.2935 OTUs – 0.1310

Legend*: P-values from Mann-Whitney U test comparisons of diversity and richness by disease location in pCD.*

Two urinary compounds were significantly lower pre-dose five in pediatric patients with colonic disease (CCD) compared to patients with upper or ileal disease (ICD). Glycolic acid was 101.7 µmol/mmol creatinine in CCD vs 234.3 µmol/mmol creatinine in ICD (FDR=0.043). HPHPA was 52.7 µmol/mmol in ICD patients compared with 13.2 µmol/mmol creatinine in CCD patients (FDR=0.048), as shown in **Figure 30**.

Figure 30 *Higher concentrations of glycolic acid and HPHPA in ICD patients.*



Legend: *HPHPA = (3-hydroxyphenyl)-3-hydroxypropionic acid. * = p<0.05*

3.11 Disease Location Associations in pUC

Disease location in pUC had associations with urinary metabolites, as well as microbial taxa at the genus and phylum levels (**Table 60**). Significantly different urinary metabolites and genera were identified on univariate analysis; metabolites are shown in **Table 63** and **Table 64**, microbiota are discussed below.

Table 60 *Disease location and glucocorticoid use are associated with significantly different urinary metabolite profiles ($p<0.05$)*

Variable	Baseline	Pre-Dose 3	Both Timepoints	Deltas
Pancolitis vs Non-pancolitis Urinary Metabolties	**0.0497**	**0.0007**$^{+}$	**0.0002**$^{+}$	0.3006
Pancolitis SCFA	0.8901	0.7256	0.5964	0.6674
Pancolitis vs non-pancolitis genus level	0.9109	**0.0669**	**0.047**$^{+}$	0.9934
Pancolitis vs Non-pancolitis phylum level	0.22	0.3703	**0.0647**	0.3042

Legend: *pUC urine metabolite PERMANOVA analysis results.* $^{+}$ *= significant compounds on univariate analysis.* $^{\blacklozenge}$ *= no or mild disease activity vs moderate or severe disease activity.*

There were no significant or trending associations with disease location and microbial richness or diversity (**Table 61**).

Table 61 *IFX treatment, patient remission, FCP changes, and patient sex associated with stool microbial diversity in pUC.*

Variable	Baseline	Pre-Dose 3
Pancolitis vs Non-pancolitis	Chao1 – 0.6682 Shannon – 0.8864 Simpson – 0.5677 OTUs – 0.5677	Chao1 – 0.2008 Shannon – 0.2864 Simpson – 0.2864 OTUs -0.2395

Legend*: P-values from Mann-Whitney U test comparisons of diversity and richness by disease location in pUC.*

Comparing patients with pancolitis to patients without, there were no significant or trending differences in age, sex, PUCAI score, disease severity or duration, or likelihood to achieve a remission PUCAI score of less than ten (**Table 62**).

Table 62 *Associations in clinical characteristics and disease location in pUC.*

	Pancolitis	Non-pancolitis	p-value
Pre-Dose 3 PUCAI <10	7/12	5/8	0.612
PUCAI Mean	**Baseline**: 48.3 **Pre-dose 3**: 12.5	**Baseline**: 42.5 **Pre-dose 3**: 11.9	**Baseline**: 0.5091 **Pre-dose 3**: 0.8712
Mean Age (range)	12.9 years (11.0-14.8)	11.9 years (9.4-14.5)	0.5371
Sex	5/12 male	3/8 male	0.515
Mean Disease Duration (range)	1.44 years (0-3.0)	1.14 years (0.2-2.1)	0.6433
Moderate or Severe Disease	**Baseline**: 9/12 **Pre-Dose 3:** 2/12	**Baseline:** 6/8 **Pre-dose 3:** 1/8	**Baseline:** 0.693 **Pre-dose 3:** 0.656

Legend*: Comparison of clinical characteristics based on disease location in pUC, including only pUC patients who provided urine specimens for analysis. Significance testing using Mann-Whitney U tests or Fisher's exact.*

Urinary levels of oxoglutaric acid were significantly lower at baseline in patients with pancolitis, with an average of 11.45µmol/mmol creatinine compared to an average of 54.80µmol/mmol creatinine in those with non-pancolitis (FDR=0.02).

Table 63 *Nine metabolites significantly higher in non-pancolitis pUC vs pancolitis after IFX treatment.*

Metabolite	Pancolitis	Upper 95%	Lower 95%	Non-Pancolitis	Upper 95%	Lower 95%	FDR
Pyroglutamic acid	15.16	21.91	8.42	39.88	51.22	28.54	0.015
Uracil	3.64	5.94	1.34	13.65	20.32	6.98	0.015
Methylmalonic acid	1.62	2.10	1.15	3.31	4.51	2.12	0.031
Vanillylmandelic acid	3.36	4.71	2.00	6.98	9.41	4.55	0.031
Glycolic acid	68.64	94.04	43.24	152.03	206.45	97.62	0.037
Fumaric Acid	2.66	6.19	0	12.31	24.96	0	0.037
Homovanillic acid	3.58	4.84	2.32	6.77	9.03	4.51	0.037
p-Hydroxybenzoic acid	1.26	1.71	0.81	5.72	13.09	0	0.037
3-hydroxy-2-methylbutyric acid	9.58	19.24	0	13.04	17.78	8.30	0.044

Legend: *Mean concentrations in µmol/mmol creatinine of significantly altered urine metabolites at pre-dose three based on disease location in pUC.*

When combining specimens from both timepoints, there are 13 metabolites significantly lower in patients with pancolitis, as shown in **Table 64**.

Table 64 *Thirteen metabolites at significantly higher concentrations in non-pancolitis pUC vs pancolitis.*

Metabolites	Pancolitis	Upper 95%	Lower 95%	Non-Pancolitis	Upper 95%	Lower 95%	FDR
Pyroglutamic acid	19.08	24.12	14.04	39.30	48.46	30.13	0.0027
Oxoglutaric acid	15.98	20.72	11.24	66.94	109.58	24.31	0.0053
p-Hydroxybenzoic acid	1.12	1.49	0.76	5.15	9.01	1.29	0.014
Uracil	3.52	5.18	1.85	10.97	15.80	6.13	0.014
Fumaric Acid	2.32	4.22	0.41	11.78	19.24	4.32	0.014
Homovanillic acid	3.56	4.27	2.84	6.24	8.13	4.34	0.014
Methylmalonic acid	1.51	1.84	1.19	2.90	3.76	2.05	0.020
Citric Acid	225.79	300.95	150.63	464.29	639.01	289.58	0.021
Glycolic acid	68.26	83.48	53.05	122.94	156.14	89.75	0.021
HPHPA	8.29	13.83	2.76	32.36	49.63	15.09	0.027
3-methylglutaric acid	0.83	1.16	0.50	1.99	2.86	1.12	0.035
3-hydroxy-2-methylbutyric acid	8.96	14.14	3.78	15.08	21.11	9.06	0.035
Succinic acid	12.38	19.04	5.71	21.75	30.34	13.16	0.042

Legend*: Mean concentrations in µmol/mmol creatinine of pUC urinary metabolites significantly different based on disease location when all timepoints combined. Pancolitis n=24, non-pancolitis n=16.*

Univariate analysis of 16S data showed two trending associations with disease location at the genus level. Comparing samples from patients with pancolitis vs non-pancolitis, when analyzing all samples from both timepoints there was a trending difference in *Gemmiger* mean

relative abundance from an average of 11.13% relative abundance in non-pancolitis disease to 0.043% relative abundance in patients with pancolitis (p value = 0.0017, FDR = 0.062). There was also a trending difference in *Lactobacillus* mean relative abundance, which was absent at both timepoints in patients with non-pancolitis, but was present in two samples from two different patients with pancolitis (in one patient at baseline and the other at pre-dose three) for a mean relative abundance of 0.011% (p-value = 0.0022, FDR = 0.062).

3.12 Age Associations in pUC

Age was identified as significantly associated with urinary metabolites on multivariate analysis in pUC (**Table 65**). There were no other significant associations with SCFA, microbiota abundance, or microbial diversity or richness (**Table 66**).

Table 65 *Age associated significantly associated with urinary metabolites in pUC.*

Variable	Baseline	Pre-Dose 3	Both Timepoints	Deltas
Age Urinary Metabolites	0.1181	0.2046	**0.0034**$^{+}$	0.99
Age SCFA	0.3244	0.67	0.8862	0.8735
Age Genus Level	0.7396	0.6008	0.4412	0.2041
Age Phylum Level	0.1617	0.2747	0.4309	0.1374

Legend: *pUC urine metabolite PERMANOVA analysis results. + = significant compounds on univariate analysis. Bold = p<0.1*

Table 66 *No associations with age and bacterial diversity/richness in pUC.*

Variable	Baseline	Pre-Dose 3
Age	Chao1 – 0.9391 Shannon – 0.6926 Simpson – 0.4352 OTUs -0.9580	Chao1 – 0.7338 Shannon – 0.7016 Simpson – 0.2767 OTUs -0.8576

On univariate analysis, ten metabolites were significantly different across age tertiles in pUC when specimens from both timepoints were combined (**Table 67**). Age ranges for each tertile are shown in **Figure 31**.

Table 67*. Ten metabolites significantly changed across age tertiles in pUC.*

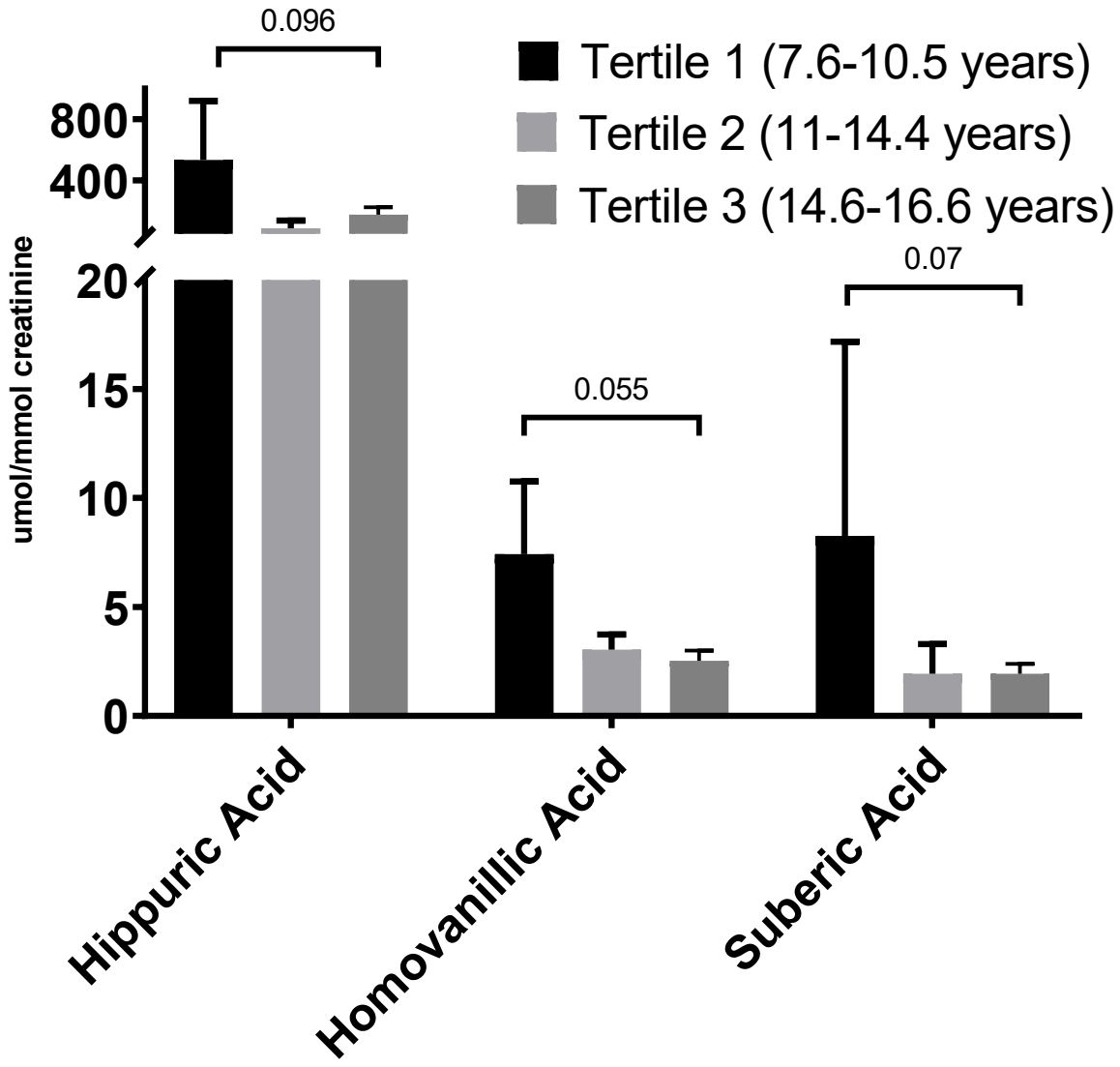
Metabolite	Tertile 1	Upper 95%	Lower 95%	Tertile 2	Upper 95%	Lower 95%	Tertile 3	Upper 95%	Lower 95%	FDR
Succinic acid	31.32	41.11	21.52	9.37	16.01	2.73	6.72	10.37	3.08	**0.0023**
Homovanillic acid	7.23	8.97	5.49	3.47	4.38	2.57	3.02	3.98	2.05	**0.0023**
Suberic acid	7.14	10.89	3.39	3.02	5.10	0.95	2.81	4.12	1.49	**0.0057**
2-ethyl-3-hydroxypropionic acid	17.46	32.54	2.37	3.18	4.03	2.33	5.98	8.00	3.95	**0.0067**
Hippuric acid	605.13	919.68	290.57	169.74	280.01	59.47	207.27	330.37	84.18	**0.0068**
3-hydroxy-2-methylbutyric acid	21.26	30.38	12.13	6.19	9.02	3.36	6.03	8.20	3.87	**0.0094**
Lactic acid	113.14	257.41	-31.13	14.05	19.64	8.45	35.44	53.10	17.79	**0.028**
Fumaric acid	13.75	22.32	5.18	1.60	2.94	0.26	2.31	3.84	0.78	**0.045**
Glycolic acid	122.09	154.36	89.82	86.68	117.59	55.78	61.14	85.91	36.37	**0.045**
Quinolinic acid	5.55	10.18	0.91	9.82	12.73	6.90	3.80	5.88	1.71	**0.045**

Legend: *Mean concentrations in µmol/mmol creatinine of pUC urinary compounds significant on Kruskal-Wallis between age tertiles when clinical specimens from both timepoints combined.*

Univariate analysis of age related metabolites was repeated for baseline values to further investigate age-related changes despite age only approaching a trending difference in metabolite

profile (**Table 67**). At baseline, urinary concentrations of homovanillic acid, suberic acid, and hippuric acid were trending significantly different between age tertiles, FDR = 0.055, 0.070, and 0.096, respectively (**Figure 31**); however, they did not reach statistical significance.

Figure 31 *pUC baseline urinary metabolites trending differences with age.*



Creatinine concentrations were used to adjust for patient hydration status, giving urine metabolites in µmol/mmol creatinine for analysis. Given the associations identified with age and urinary metabolite concentrations in pUC, creatinine was assessed for correlation with age in the pUC

cohort. On spearman correlation there was a significant positive correlation between age and creatinine concentration at baseline, p=0.0245, and after treatment, p=0.0040 (Spearman's rho = 0.5008, and 0.6135, respectively).

3.13 Sex Associations in pUC

Males and females had significantly different SCFA and genus level microbiota on multivariate analysis (**Table 68**). Univariate analysis identified that females had significantly lower microbial richness at baseline, and saw a significant increase associated with treatment that was not seen in males (**Table 69**).

Table 68 *Sex is significantly associated with SCFA and microbiota in pUC.*

Variable	Baseline	Pre-Dose 3	Both Timepoints	Deltas
Sex Urinary Metabolites	0.2377	0.574	0.1032	0.94
Sex SCFA	0.**0708**	0.1891	0.8624	**0.0095**$^{+}$
Sex 16S Genus Level	0.1227	0.7973	**0.0918**	0.6239
Sex 16S Phylum Level	0.3856	0.4843	0.6113	0.5605

Legend: *pUC urine metabolite PERMANOVA analysis results.* $^{+}$ *= significant compounds on univariate analysis.*

Table 69 *Males and females have significantly different microbial diversity and richness at baseline in pUC.*

Variable	Baseline	Pre-Dose 3	Delta
Sex	**Chao1 – 0.0281** **Shannon – 0.0281** **Simpson – 0.079** **OTUs - 0.0192**	Chao1 – 0.6698 Shannon – 0.5224 Simpson – 0.2008 OTUs - 0.7484	**Chao1 – 0.0330** Shannon – 0.6698 Simpson – 0.6698 **OTUs – 0.0330**

Legend: *Bold = p<0.1. Males had significantly higher diversity and richness at baseline, but females saw a significant increase in richness with IFX treatment that was not seen in males, shown below in* ***Figure 35****.*

Comparison of patient characteristics between all male and female pUC patients, summarized in **Table 70**, identified significantly longer disease duration in males. There were no

significant or trending differences in age, concomitant medication use, or disease location or severity.

Table 70 *Male pUC patients have significantly longer disease duration.*

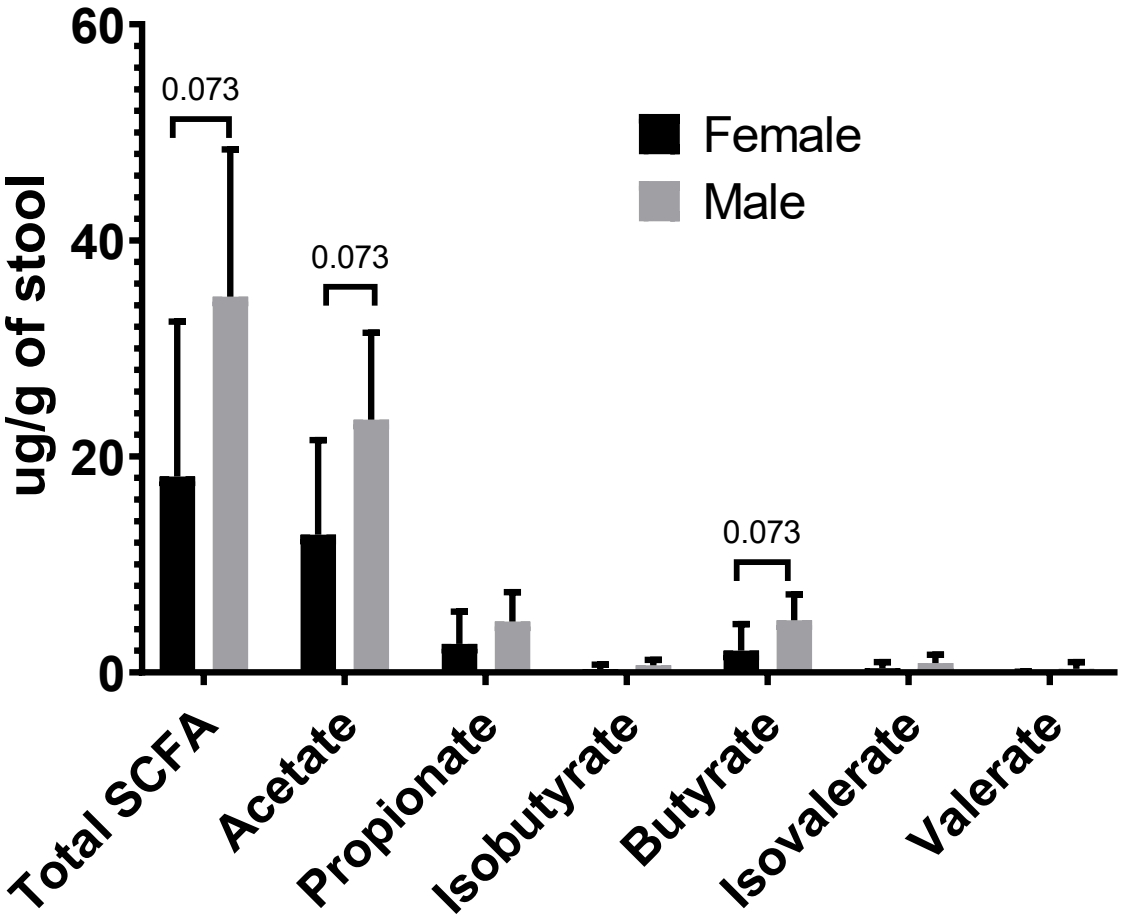
	Male n=8	**Female** n=14	**p-value**
Mean Disease Duration (range)	1.76 years (0.01-8.6)	0.29 years (0-0.66)	**0.0185**
Pancolitis	5/8	8/14	0.584
PUCAI Mean (range)	**Baseline**: 45 (10-85) **Pre-Dose 3**: 12.14 (0-55)	**Baseline**: 50.6 (15-75) **Pre-Dose 3**: 10.6 (0-40)	**Baseline**: 0.6311 **Pre-Dose 3**: 0.8309
Mean Age (range)	11.9 years (7.6-16.6)	13.8 years (15.7-11.0)	0.1722
FCP (95%C.I)	**Baseline:** 5407 (0-10890) **Pre-Dose 3:** 1800 (0-4724) Pairwise p=0.3980	**Baseline:** 6668 (0-17469) **Pre-Dose 3:** 2149 (0-4333) Pairwise p=0.1441	**Baseline**: p=0.4649 **Pre-Dose 3**: p=0.9578
Medication Use	**Baseline:** Azathioprine: 4/8 GCS:5/8 Methotrexate:2/8 5ASA:4/8 **Pre-Dose 3:** Azathioprine:2/8 GCS:3/8 Methotrexate:1/8 5ASA:2/8	**Baseline:** Azathioprine:8/14 GCS:7/14 Methotrexate:1/14 5ASA:8/14 **Pre-Dose 3:** Azathioprine:5/14 GCS:4/14 Methotrexate:0/14 5ASA:5/14	**Baseline:** Azathioprine: 0.546 GCS =: 0.454 Methotrexate: 0.291 5ASA: 0.546 **Pre-Dose 3:** Azathioprine: 0.490 GCS: 0.510 Methotrexate: 0.364 5ASA: 0.490

Legend*: bold = p<0.1*

To investigate the potential influence of baseline stool SCFA concentrations univariate analysis was done to compare male vs female stool SCFA concentration at baseline. There were no significant differences between male and female stool SCFA concentrations, but there was a trend for males to have higher acetate, butyrate, and total SCFA (p=0.073) (**Figure 32**).

Propionate, isobutyrate, isovalerate, and valerate showed nonsignificant increases; p=0.5224, p=0.6698, p=0.8312, and p=0.9148, respectively.

Figure 32 *At baseline, males with pUC have higher concentrations of acetate, butyrate, and total SCFA.*

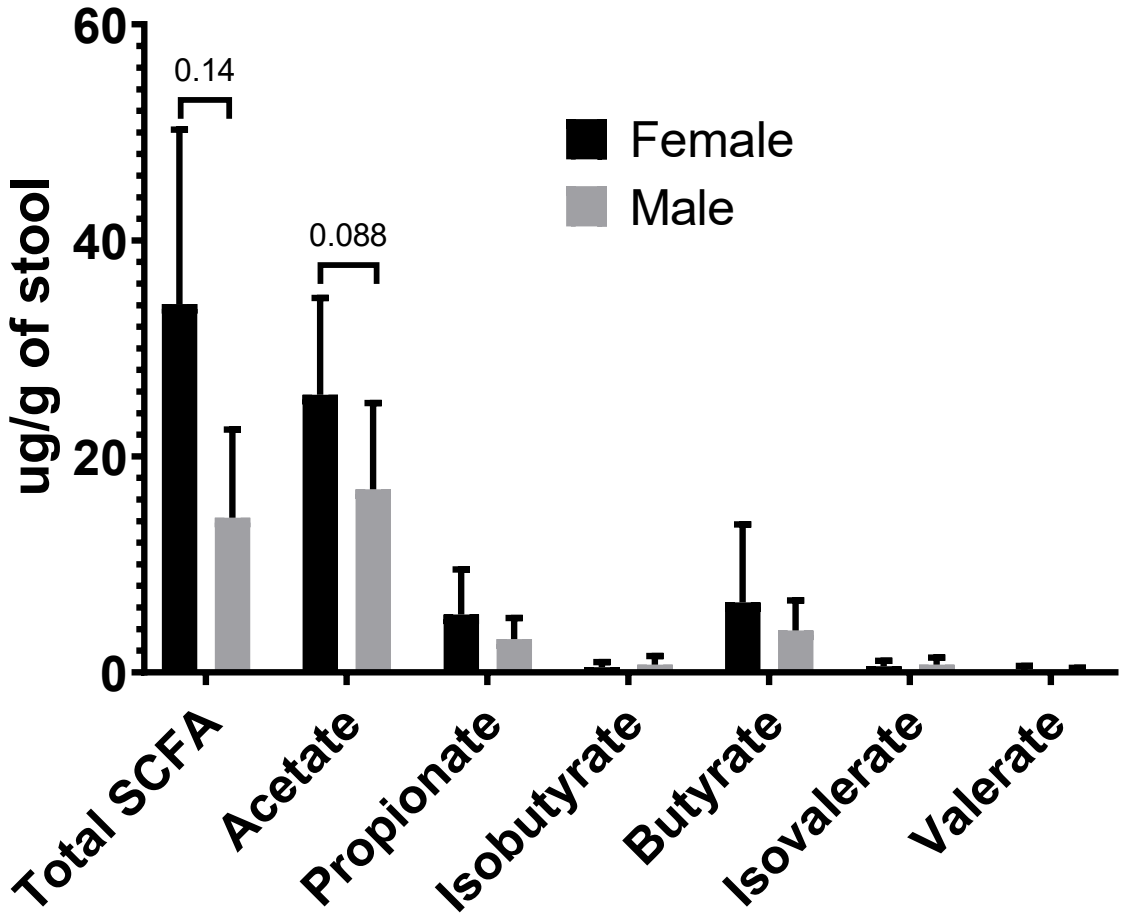


Legend: *Male vs female baseline stool SCFA concentration in pUC. One female and one male patient each contributed a baseline stool specimen but not a pre-dose three specimen, so could not be analyzed for changes over time but are included in the baseline analysis here. Females n=7, males n=5.*

Shown in **Figure 33** there was a trend for females to have higher acetate after treatment (p=0.088). Total SCFA and individual SCFAs were non significantly increased in females

compared to males after treatment: total SCFA p=0.1356, propionate p=0.5224, isobutyrate p=0.6698, butyrate p=0.6698, isovalerate p=0.8312, and valerate p=0.9148. Results comparing sexes after treatment showed a reversal of the trend observed in stool SCFA concentrations in pUC patients at baseline (**Figure 35**).

Figure 33 *Females have trending higher stool acetate after IFX treatment.*

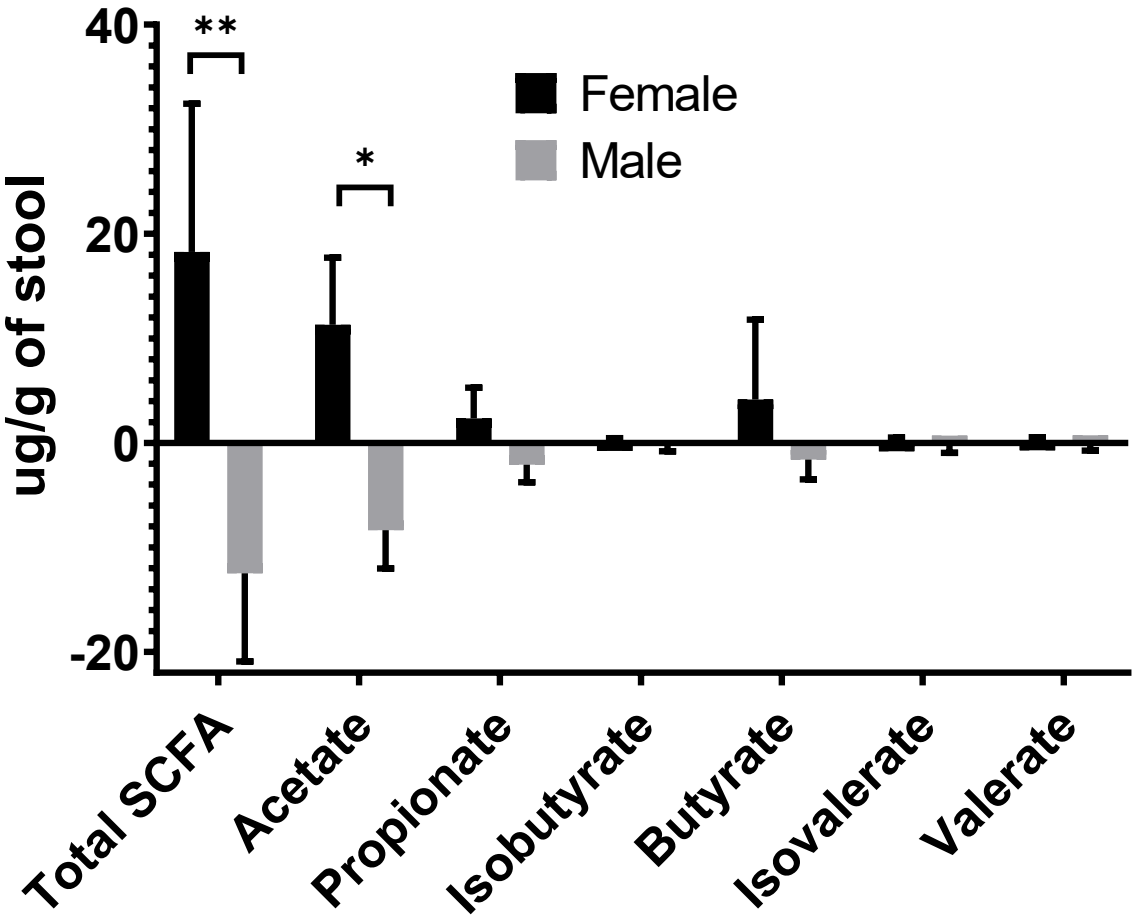


Legend: *Male vs female pre-dose three stool SCFA concentration in pUC. Females n=6, males n=4.*

Univariate analysis of SCFA changes over time for males vs females with pUC revealed two significant differences (**Figure 34**). All four male patients showed a decrease in acetate,

propionate, butyrate, and total SCFA. Two of four males also had a decrease in isobutyrate and three of four had decreases in isovalerate and valerate. All females saw an increase in acetate, valerate, and total SCFA (**Figure 34**). Male vs female clinical characteristic analysis is shown in **Table 71**, only for patients with stool available for SCFA analysis. Although among all male pUC patients in this study there was on average a significantly longer disease duration (**Table 70**), this was not significant among the subset of patients who submitted stool for SCFA analysis (**Table 71**). No significant or trending correlation was found between disease duration and changes in SCFA concentration with IFX therapy on spearman correlation analysis.

Figure 34 *Sex differences in changes in SCFA concentrations with IFX treatment.*



Legend: *Changes in stool SCFA concentrations (ug/g stool) with IFX treatment in pUC. Females n=6, males n=4. *=p<0.05.**=p<0.01.*

Analysis of male vs female patient characteristics in pUC assessing only patients who provided stool for SCFA analysis revealed no significant differences in age, disease duration, likelihood to achieve remission, disease location, medication use, FCP, or PUCAI score (**Table** 71).

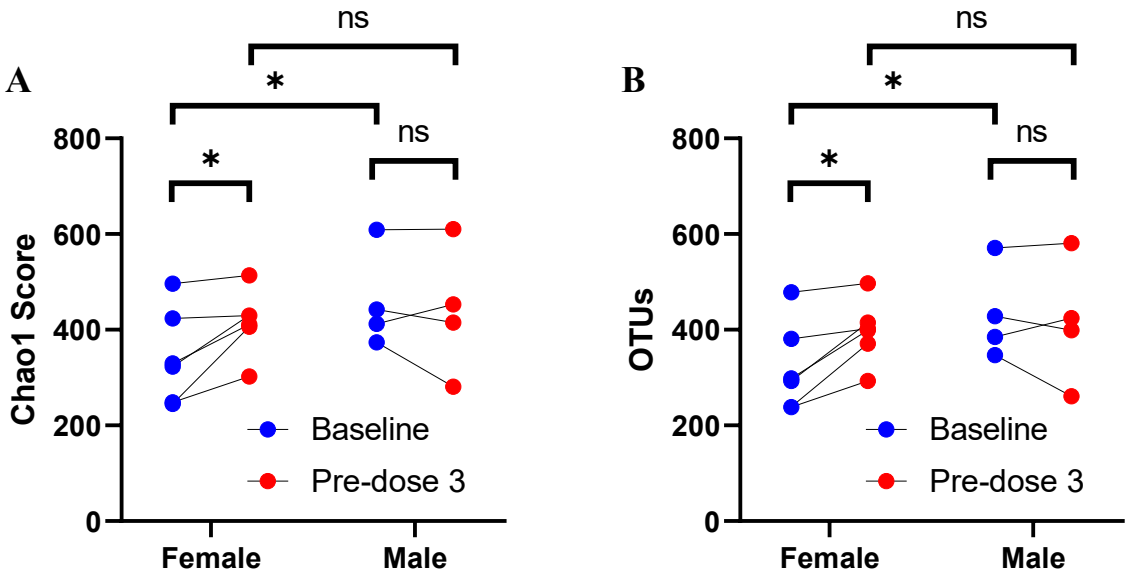
Table 71 *No significant differences in male vs female patient characteristics in pUC among patients assessed for stool SCFA.*

	Male	**Female**	**p-value**
Age (range)	12.6 years (12.4-16.6)	12.4 years (8.6-16.6)	0.6698
Disease Duration (range)	0.3 years (0-0.6)	0.7 years (0-1.6)	0.1995
Remission by Dose 3 PUCAI <10	2/4	3/6	0.652
Pancolitis	1/4	3/6	0.452
PUCAI Average	**Baseline**: 33.75 **Pre-Dose 3**: 13.75	**Baseline**: 43.33 **Pre-Dose 3**: 15.83	**Baseline**: 0.3758 **Pre-Dose 3**: 0.7406
GCS	**Baseline:** 3/5 **Pre-Dose 3:** 2/4	**Baseline:** 4/7 **Pre-Dose 3:** 1/6	**Baseline:** 0.689 **Pre-Dose 3:** 0.333
FCP (95%C.I.)	**Baseline:** 5407.4 (0-10889.7) **Pre-Dose 3:** 2248.6 (0-5598.7) Pairwise p=0.1441	**Baseline:** 6668.2 (0-17469.1) **Pre-Dose 3:** (2237.5 (0-4583.9) Pairwise p=0.3980	**Baseline:** 0.2568 **Pre-Dose 3:** 0.8501 **Delta:** 0.7055

Legend*: pUC patient characteristics, including only patients with stool available for SCFA for analysis at both time points to enable changes over time analysis. Males n=4, females n=6.*

Comparing changes over time, females vs males had significantly different changes in Chao1 scores and OTUs identified. Females saw a significant increase in richness as measured by Chao1 from baseline in paired analysis (p=0.0277), with an increase over time that was significantly greater, (p=0.0330), from the non-significant average decrease seen in males. A similar trend was seen for OTU numbers with females showing a significant increase in OTUs from baseline in paired analysis (p=0.0277), with an increase over time that was significantly greater from the non-significant decrease in OTUs seen in all four males (p=0.0330). As shown in **Table 69** and **Figure 35**, males had significantly higher average Chao1 and Shannon H diversity scores at baseline, and significantly higher OTUs at baseline. There was a trending higher average Simpson diversity score at baseline for males (**Table 69**).

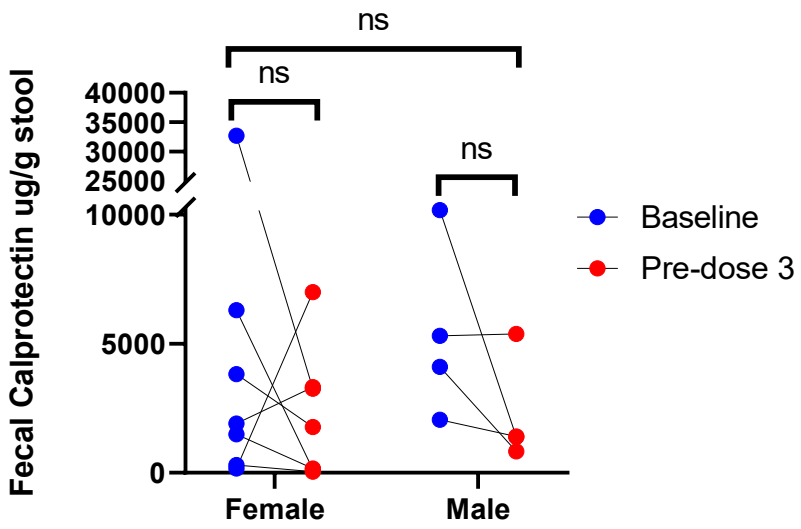
Figure 35 *Significantly different changes in microbial diversity and richness with IFX therapy in pUC.*



Legend: ***A*** *Chao1 richness index differences between males and females with pUC.* ***B*** *Differences in total OTU numbers identified in fecal 16S sequencing in males vs females with pUC. Males n=4 pairs, females n=6 pairs.*

Among the patients with pre-dose three stool available for FCP analysis (regardless of baseline availability), one of four patients with a pre-dose three FCP <250 was male, while six of 12 with a pre-dose three FCP ≥250 were male. Fisher's exact analysis of male vs female likelihood of a pre-dose three FCP <250 is non-significant, p=0.392. Among paired specimens, there was no statistically significant or trending difference between male and female FCP levels at baseline (0.2568), after treatment (p=0.8501), or in changes over time (p=0.7055), as shown in **Figure 36**.

Figure 36 *No sex differences in pUC FCP*



Legend: *pUC FCP by sex. Females n=7 pairs, males n=4 pairs, ns = not significant*

3.14 GCS Use Associations in pCD & pUC

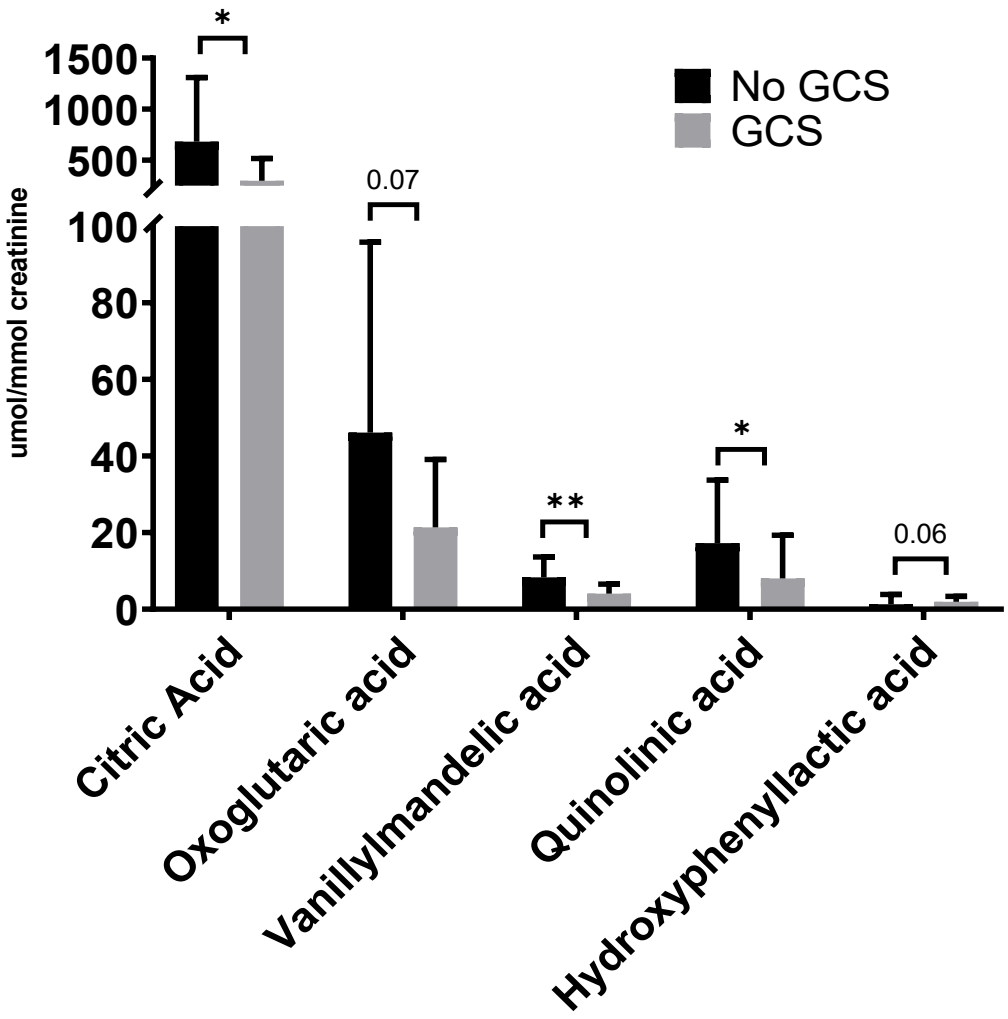
Glucocorticoid use was assessed in multivariate analysis for associations with urinary metabolite profiles when pUC and pCD were combined, and was found to be associated with a significantly different urinary metabolite profile at baseline and when all timepoints were combined in the combined, as well as a significantly different SCFA profile when specimens from all timepoints were combined (**Table 72**). Univariate analysis revealed significantly different metabolites only when specimens from all timepoints were combined (**Figure 37**).

Table 72 *Glucocorticoid use was associated with a significantly different metabolite profile.*

Variable	**pCD & pUC**		
	Baseline	**Dose 3/5**	**All timepoints combined**
Glucocorticoids Urine Metabolites	**0.0095**	0.147	**0.0047***
GCS SCFA	0.2426	0.3995	**0.0489**

Legend: *Multivariate PERMANOVA analysis-values comparing urine metabolites of pCD & pUC patients on vs off glucocorticoids at baseline and dose three/five.*

Figure 37 *Metabolites altered pIBD in patients taking GCS.*



Legend: *Mean concentrations in µmol/mmol creatinine of significantly different or trending urine metabolites comparing pCD and pUC patients on vs off glucocorticoids when specimens from all timepoints were included in analysis. Patients on GCS n=22, patients off GCS n=79. *=p<0.05, **=p<0.01*

Assessing baseline urine metabolites only in a combined analysis of pUC and pCD, quinolinic acid also had a trending lower concentration (p=0.0019, FDR = 0.073) in those samples collected from patients on glucocorticoids. Average concentration of quinolinic acid was 23.16 µmol/mmol creatinine (95% 16.16-30.17) in those not taking glucocorticoids, compared to

9.65 µmol/mmol creatinine (95% 2.64-16.66) in those taking glucocorticoids. Clinical characteristics of patients on GCS compared with those off at baseline were not significantly different when assessing age, sex, disease duration, proportion of patients with moderate or severe disease at baseline or after treatment, or likelihood to achieve remission by dose three for pUC or dose five for pCD.

Glucocorticoid use was assessed for an impact on SCFA profile in pIBD, as shown in **Table 72**, and was found to be associated with a significantly different SCFA profile only when samples from all timepoints were combined. On univariate analysis, total SCFA and individual SCFA contents were not significantly different or trending when pUC and pCD were combined to analyze associations with glucocorticoids in univariate analysis.

3.15 IFX Cohort Summary

Associations with multiple factors were explored for urinary metabolites, fecal SCFA, and fecal microbiota. IFX therapy had significant associations in each data set, with the majority of the significant findings in the pCD patients. Additionally, a number of other clinical characteristics were found to have consistent associations, including remission status, antimicrobial and GCS use, disease severity and location, and age and sex in pUC. While only associations can be identified here due to the observational nature of this study, potential relationships with diet and disease pathophysiology are examined below and in the Discussion section.

3.16 Dietary Analysis in pCD EEN Cohort

To assess for the impacts of previous diet on treatment response, a separate cohort of pCD patients that were receiving EEN induction therapy and had completed FFQs, referred to as the EEN-FFQ study, was analyzed for dietary patterns. Future analysis will include EEN treatment outcomes and baseline microbiome analysis for a subset of patients where available. PCA allows for establishment of dietary patterns with each patient receiving a score for adherence to each pattern. Principal components represent dietary patterns within the analyzed population. PCA allows for multivariate analysis, and data-driven dietary patterns that reflect ways of eating within the sampled population.

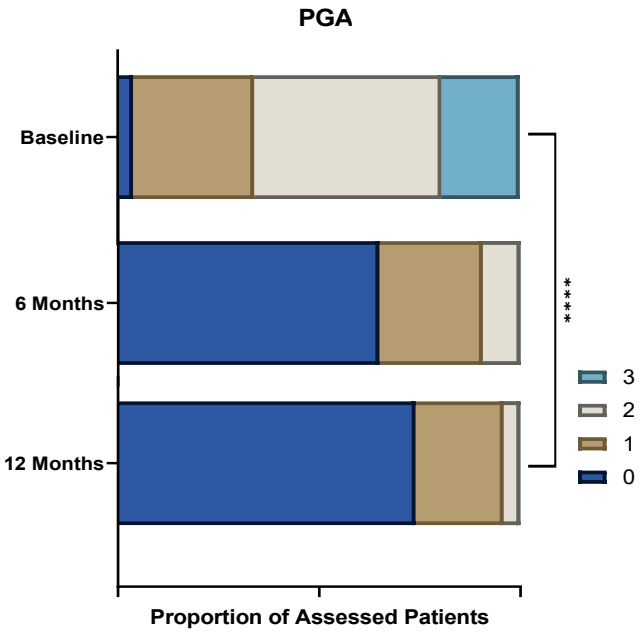
3.16.1 pCD EEN-FFQ Cohort

Of the 211 patients newly diagnosed with pCD who were beginning induction therapy with EEN, 103 had completed FFQs within 90 days before or after EEN initiation and had baseline and follow-up data collected. Patient characteristics are summarized in **Table 73**. As shown in **Figure 38**, disease activity, as assessed by PGA, was reduced in follow-up at six and 12 months after diagnosis and EEN initiation ($p<0.0001$).

Table 73 *pCD EEN cohort clinical characteristics.*

Clinical Characteristic	Mean/Count (%)
Mean Age (range)	12.5 (4.4-16.8)
Sex	61/103 male (59.2%)
Perianal Disease	9/103 (8.7%)
Colonic Disease	76/103 (73.8%)
Complicated Disease	4/103 (3.9%)

Figure 38 *Disease activity as assessed by PGA was significantly reduced after EEN therapy in the pCD EEN-FFQ cohort.*



Legend*: Patients were assessed at baseline when initiating EEN, and six and twelve months after EEN initiation. Baseline n=103, 6 Months n=97, 12 months n=96. **** = p<0.0001*

3.16.2 Dietary Patterns

As shown in **Table 74**, the PC eigenvalues obtained from the FFQ kilocalorie-adjusted PCA accounted for a maximum of 7.4% of dietary variance. The range of reported daily kilocalories was 835 to 4325 kilocalories/day. Given the sharp reduction in eigenvalues after component four, four PCs were used as dietary patterns to assess for clinical correlations.

Table 74 *PC Eigenvalues and % of dietary variance explained.*

Component	Eigenvalue	% of Variance
Principal Component 1	3.4051	7.4
Principal Component 2	3.38308	7.35
Principal Component 3	3.29477	7.16
Principal Component 4	3.2256	7.01
Principal Component 5	2.35795	5.13
Principal Component 6	2.16593	4.71
Principal Component 7	1.96347	4.27
Principal Component 8	1.90962	4.15
Principal Component 9	1.78143	3.87
Principal Component 10	1.5154	3.29

The factor loadings for each PC/dietary pattern are shown in **Table 75**. Only factors with loading greater than 0.2 were included as characteristic of each pattern. PC one, named the ***Vegetarian dietary pattern*** due to its positive association with soy and tofu and lack of meat positive associations, accounted for 7.4% of dietary variability. PC two, named the ***Meat dietary pattern*** due to its abundance of meat associations, accounted for 7.35% of dietary variability. PC three was named the ***Pre-packaged dietary pattern*** as the positive associations were all foods that could be found pre-packaged or as processed foods in grocery stores; this pattern accounted for 7.16% of dietary variability. The fourth PC accounted for 7.01% of dietary variability and was named the ***Mature dietary pattern*** due to the positive associations with coffee and alcohol,

which would not typically be expected in a younger pediatric population. The food groups characterizing each dietary pattern are detailed in **Table 75**; altogether these four dietary patterns explain 28.9% of the dietary variability within our pCD EEN-FFQ sample population.

Table 75 *Four dietary patterns in pCD.*

Dietary Pattern:	Vegetarian	Meat	Pre-Packaged	Mature
Food Groups Positively Associated	Whole Grains	Rice, Rice Noodles, Couscous	High Fiber Cereals	Chicken, Turkey without skin/fried
	Vegetable Soup	Other Soups	Sugary Condiments	Fish
	Soy/Tofu	Red Meat	Breaded Fish	Seafood
	Salad Dressing	Pork	Diet Soda	Vegetables
	Fruit	Liver Organs		Fruit
	Full Fat Dairy	Chicken, Turkey with skin or fried		Coffee
	Butter			Alcohol
				Milk Alternatives
Food Groups Negatively Associated	Chicken, Turkey with skin or fried	Chicken, Turkey without skin/fried	Lean Red Meat	Pizza
		Granola Bars	Processed Meat	

Legend*: Dietary patterns obtained as the first four PCs from FFQ PCA analysis. Green shading represents food groups with a positive association with dietary adherence scores; red shading indicates food groups with a negative association with dietary adherence scores.*

Once adherence scores for each patient were converted to Z-scores, outliers greater than 3 and less than minus three were excluded. Identified outliers included: one patient high on

vegetarian adherence, two patients high on meat adherence, two patients high on pre-packaged adherence, and one patient high and one patient low on mature adherence. The clinical characteristics associated with adherence to each dietary pattern are detailed in **Table 76**. Significant and trending associations with sex, age, and disease location were identified.

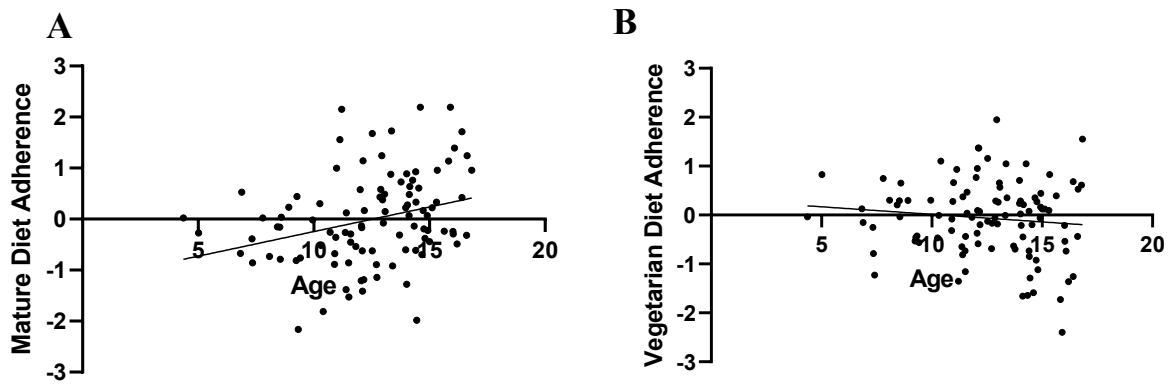
Table 76 *Dietary patterns vary with age, sex, and disease location in pCD.*

Dietary Pattern	**Vegetarian**	**Meat**	**Pre-Packaged**	**Mature**
Male vs Female	0.4586	0.3798	0.6996	**0.0407**
CCD vs ICD	0.7703	0.3201	0.3002	0.1071
CCD vs ICD in Males	0.6057	0.7470	0.3397	0.8718
CCD vs ICD in Females	0.8661	0.1995	**0.0031**	**0.0042**
Age	**0.0996**	0.5684	0.6467	**0.0006**

Legend*: Association of clinical characteristics with dietary pattern adherence among pCD EEN cohort. Bold = p<0.1*

There was no significant difference in the rate of colonic disease in males and females: 33/42 females had CCD, and 43/61 males had CCD, p=0.360 (χ^2). There was also no significant or trending age difference between males and females, p=0.3578. Assessing correlations between diet adherence and age, Spearman's rho for age and vegetarian diet adherence was -0.1640 and for age and mature diet adherence was 0.3355, as shown in **Figure 39**.

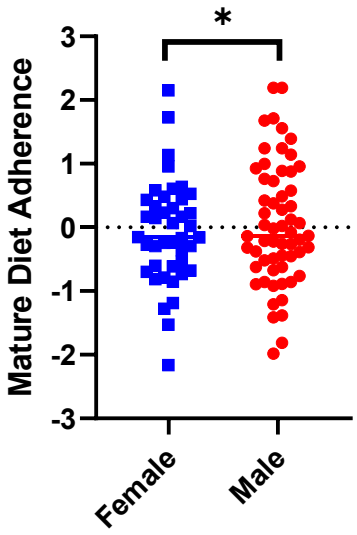
Figure 39 *Older pCD patients have higher mature diet adherence, and lower vegetarian diet adherence.*



Legend: *Significant and trending correlations between mature (**A**) and vegetarian (**B**) dietary patterns and age.*

Males had a significantly higher adherence to the Mature diet as shown in **Figure 40**. p=0.0407. The average Z-score in males was 0.044 (95% C.I. -0.20 – 0.29), while the average Z-score in females was -0.078 (95%C.I. -0.33 – 0.18).

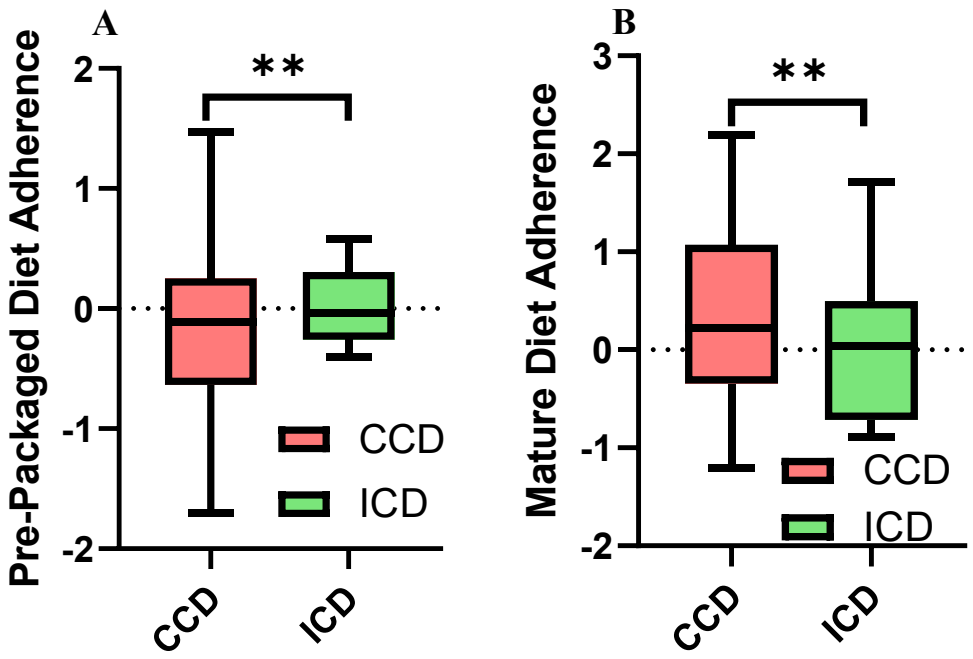
Figure 40 *Males are more adherent to a mature diet than females in pCD.*



Legend*: *=p<0.05. Blue = females mature diet adherence Z-scores, Red = male mature diet adherence Z-scores.*

In females only, an association was identified with disease location and adherence to the "Pre-packaged" and "Mature" diets. As shown in **Figure 41**, females with colonic CD were significantly less adherent to the "Pre-packaged" diet and significantly more adherent to the "Mature" diet ($p<0.01$).

Figure 41 *Females with CCD are less adherent to pre-packaged diet and more adherent to mature diet.*



Legend: *Adherence to Pre-packaged (**A**), and Mature (**B**) diets among females comparing ileal (ICD) and ileocolonic/colonic Crohn disease. **=p<0.01.*

3.16.3 pCD EEN-FFQ Summary

Associations with diet and age, sex, and disease location were identified in the pCD EEN-FFQ cohort. The predominant dietary patterns accounting for variation within our sample population were defined as “Vegetarian”, “Meat”, “Pre-packaged”, and “Mature” diets. The associations between these diets and EEN treatment outcomes will be examined in the future. Select baseline specimens will also be analyzed for microbiota with 16S sequencing and correlated with dietary intake.

Chapter 4. DISCUSSION

pUC and pCD are chronic inflammatory conditions with associated changes in metabolites, microbiota, and barrier integrity. Alterations in the host-microbe relationship are evident and may provide a target for novel therapeutics or serve as biomarkers for therapy candidacy or success. Understanding how current treatments, such as IFX and EEN, alter host-microbe interactions can improve our understanding of the pathophysiology of these diseases, enhance our ability to select patients for therapy, and could lead to production of novel therapies.

4.1 IFX Cohorts

This research identified a number of significant changes in microbiota, urinary metabolites, FCP, and SCFA associated with IFX treatment in pIBD. Baseline profiles that could distinguish patients likely to achieve remission with IFX may be particularly useful as biomarkers to better identify therapy candidates. Baseline predictive individual microbiota or metabolites were not identified, although there were some significant multivariate and diversity associations. Profiles identified here as associated with remission included phylum level microbiota profiles linked to a final FCP of <250 in pCD, as well as baseline Simpson stool microbiota diversity in pUC, discussed below. Markers that can distinguish successful patient response to treatment were not found in pCD based on wPCDAI score or FCP, but several potential markers were associated with increased disease severity. In pUC, phylum level associations may help distinguish disease severity. Other clinical characteristics discussed below

that were associated with significantly different metabolites or microbiota included: antimicrobial and glucocorticoid use, disease severity, disease location, age, and sex.

4.1.1 Markers of Host-Microbe Relationship Change with Treatment

4.1.1.1 pCD

The pCD cohort had significant changes in SCFA, urine metabolites, and phylum and genus level microbiota with IFX treatment. Urine metabolites that changed significantly in association with IFX treatment included: quinolinic, hydroxyphenyllactic, oxoglutaric, 2,5-furandicarboxylic, 4-hydroxyhippuric, 3-methylglutaconic, 1H-Indole-3-acetic, ethylmalonic, 3-hydroxyphenylacetic, glycolic, 4-hydroxymandelic acids, and HPHPA. Some of these metabolites have interesting links to IBD and gut microbes, based on available literature, which could be relevant to this study population.

Quinolinic acid is a metabolite of tryptophan, and IBD has been associated with reduced circulating levels of tryptophan metabolites;[159] lower levels of tryptophan metabolites are associated with increased disease activity.[159] pIBD is associated with a lower microbial diversity, which is associated with reduced tryptophan metabolites, but pCD patients had a trending increase in microbial richness and diversity, suggesting the quinolinic acid increase may be at least partly microbially sourced.[42, 43] Quinolinic acid may play a role in NAD+ regeneration in leukocytes, is produced by many immune cells (as well as in hepatocytes), and has increased production during inflammation.[160] Moffett & Namboodiri's "Tryptophan Depletion Hypothesis", theorize quinolinic acid is rapidly consumed by immune cells to regenerate NAD+;[160] it is possible that a reduction in quinolinic acid here reflects decreased immune cell

production after resolution of inflammation. Nikolaus *et al.* (2017) found quinolinic acid was higher in the serum of IBD patients compared to controls, suggesting shunting of tryptophan metabolites into the kynurenine pathway for NAD+ regeneration during inflammation.[161] Reduced quinolinic acid with treatment here suggests quinolinic acid has decreased production with resolving inflammation, resulting in increased tryptophan available for other pathways.

Hydroxyphenyllactic acid is known to be produced by *Bifidobacterium* and *Lactobacillus* species.[162] Some studies have found an increased abundance of *Bifidobacterium* and *Lactobacillus* in CD patients relative to controls; specific species and strains within these groups can be differentially associated with inflammation, and their relationships with IBD remain unclear.[163] Their expansion may replace butyrate-producing microbes that could otherwise suppress inflammation and improve gut barrier integrity.[79] Hydroxyphenyllactic acid is also found in a number of other microbes, including: *Actinetobacter*, *Clostridium*, *Enterococcus*, *Escherichia*, *Eubacterium*, *Klebsiella*, *Pseudomonas*, and *Staphylococcus*.[164] The relative abundances of these genera were not significantly altered in pCD with IFX therapy; increased hydroxyphenyllactic acid could also reflect changes in microbial function with IFX therapy.

Oxoglutaric acid was negatively associated with FCP in our study, and has previously been associated with refeeding in malnourished children and is an intermediate in the CAC.[165] As patient status improves with therapy, it is likely that dietary intake increases, together with their ability to absorb nutrients, as diarrhea and stool frequency decrease. 2,5-furandicarboxylic acid is thought to be a marker of both cooked food and/or fructose intake, and may also reflect improved dietary intake; it is also found in some bacteria, such as *Cupriavidus basilensis*, a Gram-negative soil bacterium that was not identified here.[166, 167] 4-hydroxyhippuric acid was also negatively associated with FCP, and like oxoglutaric acid is increased in refeeding malnourished

children.[165] It is also thought to be produced by microbial metabolism of dietary polyphenols,[168] and several *Bacteroides, Enterococcus,* and *Eubacterium* species are known to produce it.[169] 3-methylglutaconic was one of the main metabolites accounting for variability in the pCD cohort on PCA, was negatively associated with FCP, and significantly increased with IFX treatment. It is an intermediate in leucine metabolism[170] and may reflect dietary changes.

1H-Indole-3-acetic acid is also a tryptophan metabolite; it was increased with IFX treatment. 1H-Indole-3-acetic acid is produced by many *Clostridia*[171] species as well as *Bacillus,*[172] *Enterobacter*,[173] and *Pseudomonas*.[174-176] It is also an endogenous metabolite, and produced in plants that may be consumed in the diet.[177] The increase in 1H-Indole-3-acetic acid may result from decreased kynurenine pathway activity (which generates NAD+), increasing the availability of tryptophan that is no longer being used to produce quinolinic acid during inflammation.[161]

Ethylmalonic acid was significantly decreased after IFX treatment; it is implicated in a rare, fatal genetic disorder called ethylmalonic encephalopathy,[178] where it may inhibit cerebral mitochondrial creatinine kinase activity.[179] The implications of a decrease in pCD are unclear.

3-hydroxyphenylacetic acid saw a nine-fold increase with treatment in pCD and is thought to have a protective activity; it is also positively associated with *Clostridia*[180] and has also been found in *Klebsiella*.[181] It has been proposed to be produced by colonic bacteria from metabolism of dietary quercetins found in orange juice, and may be responsible for the bioactive effects that make dietary polyphenols beneficial for health.[182] Glycolic acid was significantly increased with IFX treatment and negatively correlated with FCP in pCD. Glycolic acid is increased with high protein diets, especially consumption of hydroxyproline such as in gelatin.[183, 184] Glycolic acid is also a marker for a number of Gram negative bacilli and Gram

positive cocci, including: *E. coli*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Enterobacter*, *Acinetobacter*, *Proteus mirabilis*, *Citrobacter frundii*, *Enterococcus faecalis*, Group B *Streptococci*, and *Staphylococcus saprophyticus* from metabolism of carbohydrates.[185] Endogenously, glycolic acid is an oxalic acid precursor and is metabolized in liver peroxisomes.[186] The increased glycolic acid seen in the pCD cohort here could be due to increased protein intake, as well as shifting microbial abundances and function with IFX treatment, which this study was not designed to assess. Changes in dietary intake with therapy were not assessed here.

4-hydroxymandelic acid was significantly increased with IFX treatment in pCD; in PCA analysis it was one of the top metabolites accounting for variation in pCD and was negatively correlated with FCP. It has been considered a strong urinary marker for polyphenols from wine and grape juice intake and is thought to be produced by metabolism of these polyphenols by the intestinal bacteria such as *Lactobacillus plantarum*.[187] HPHPA was also significantly increased with IFX treatment in pCD and negatively correlated with FCP. It is another metabolite associated with *Clostridia* species.[188, 189] HPHPA has been shown to decrease with antibiotic use; but in rats HPHPA decreased when they were supplemented with resistant starch over a low fibre diet.[188, 190]

Many of the metabolites identified as increased with IFX therapy in pCD are associated with microbial production, *Clostridia* in particular are associated with many of these metabolites such as HPHPA, 1H-Indole-3-acetic acid, and hydroxyphenyllactic acid, but had no corresponding increase in relative abundance with IFX therapy. Such changes could possibly reflect changing microbial function, where microbial metabolism is altered in association with IFX or its therapeutic consequences, such as decreased inflammation.

Despite the fact that all SCFA increased with IFX treatment in pCD, among microbiota only three genera significantly increased: *Ruminococcus*, *Blautia*, and *Pseudobutyrivibrio.* At the phylum level there was significantly increased Verrucomicrobia but this was entirely accounted for by increased *Akkermansia*. Increases in *Akkermansia* mean relative abundance was not significant at the genus level due to a greater number of multiple comparisons, (p=0.03, FDR = 0.17). These relative abundance changes may partially explain the SCFA increases, as *Ruminococcus* spp. can produce acetate and propionate, and *Blautia* and *Pseudobutyrivibrio* produce butyrate.[191-193] Taxa such as certain *Ruminococcus gnavus* strains have been found to be enriched in IBD, and has been found to transiently bloom in some CD patients.[194] Our findings here are in contradiction to previous findings; typically increased *R. gnavus* has been associated with increased disease activity, and *R. gnavus* can degrade mucin and produce inflammatory polysaccharides.[194, 195] *R. gnavus* is present in healthy individuals but with a relative abundance typically below 1%; here the mean relative abundance in CD patients was 2.3% after treatment, with a relative abundance as high as 13.1%.[194, 195] It is possible that these findings reflect another species of *Ruminococcus*, as 16S sequencing lacks the depth to accurately classify taxa to the species level. It is also likely that the relationship between *R. gnavus* and disease severity is complex, and this microbe may exhibit altered function, dependent on many variables. *Blautia* and *Pseudobutyrivibrio* are generally thought to be beneficial due to the significant butyrate production.[48]

Akkermansia muciniphila has been previously identified as greatly reduced in IBD patients compared to healthy controls.[196] When given as a probiotic to mice with DSS-induced colitis, *A. muciniphila* resulted in improved intestinal barrier function and histology scores.[197] *A. muciniphila* metabolizes colonic mucin and its mechanisms of beneficial effects are unknown,

but its administration results in decreased pro-inflammatory cytokines, such as TNF-α.[198] Such effects could enhance the ability of treatments like IFX to induce remission. As discussed in the introduction, increased SCFA could help explain the efficacy of IFX therapy as they can result in improved barrier integrity, decreased inflammation, and increased regulatory T cells.[73, 74]

The significant reduction in FCP with IFX therapy in pCD is a sensitive indicator of mucosal healing.[14] The reduction in inflammation associated reactive oxygen species may be a significant part of the resulting changes in metabolites and microbiota. *Ruminococcus, Blautia, Pseudobutyribrio,* and *Akkermansia*, are all anaerobic species, as are Clostridia, which are associated with many of the altered metabolites.[194, 195, 198]

4.1.1.2 pUC

The pUC cohort had a trending association of increased microbial diversity with treatment but no other significant or trending associations with IFX therapy were found. Given that the pUC cohort is approximately half the size of the pCD cohort and their IFX treatment period was approximately half as long, it is not surprising that fewer associations were identified with treatment, as more time may be required to have more significant changes, and a larger cohort may be necessary to detect statistically significant changes.

4.1.2.3 pCD and pUC Combined and Contrasted

Analysis of changes when pCD and pUC were combined identified some shifts in metabolites with IFX treatment, including: oxoglutaric acid, quinolinic acid, and 2,5-furandicarboxlyic acid. These metabolites were identified as significantly increased in the pCD cohort and are discussed above.

SCFA concentrations increased with treatment in pCD and when pCD and pUC were combined for analysis. As described in pCD, these compounds may partially explain the mechanism of IFX treatment success. An increase in potential SCFA-producing bacteria such as *Blautia* and *Pseudobutyrivibrio* may be responsible for the increased metabolites, as well as altered function of other microbes that did not exhibit significant increases in abundance. The increases in diversity and richness seen with treatment are consistent with those found by others examining the effects of IFX treatment in pIBD.[80] Generally, increased diversity is associated with health,[44] and intestinal microbiome diversity measures are consistently decreased in IBD.[199] As IFX results in decreased inflammation and reactive oxygen species, it is possible that more oxygen-sensitive organisms can survive in the gut environment, potentially explaining the increases in diversity and richness.

Metabolites that distinguished pCD and pUC at every timepoint included quinolinic acid, vanillylmandelic acid, and citric acid, which were all consistently higher in pCD. Vanillylmandelic acid can be traced to both endogenous and bacterial production, and citric acid is part of the CAC and can be linked to dietary intake, most commonly citrus fruits .[169] Quinolinic acid, as discussed above, may be related to tryptophan metabolism, specific bacterial species, and inflammation. Increased dietary metabolites after treatment could be reflective of the longer duration of therapy in pCD. This is demonstrated by the findings that, with an addition of ~14 weeks patients have likely further recovered and improved dietary intake, which may also explain a greater drop in quinolinic acid if inflammation has further dampened. As discussed in the introduction, the dysbioses of pCD and pUC have some differing characteristics; differences in microbially-associated metabolites at baseline may account for these differences.

4.1.2 Associations with Antimicrobial and Glucocorticoid Use

Antimicrobial and GCS use were associated with significant changes in the host-microbe interface in pCD and pUC, respectively. In the pUC cohort there was no antimicrobial use, which is in line with treatment recommendations.[107] Lower concentrations of SCFA with antimicrobial use in pCD are consistent with previous findings,[200, 201] and consistent with the 16S finding of lower relative abundances of *Faecalibacterium* and *Blautia*; *Faecalibacterium prausnitzii* and *Blautia faecis* are known SCFA producing bacteria.[202, 203]

Prior to starting IFX, *Gemella* was not detectable in the stool of any pCD patients that had not taken antimicrobials in the previous 30 days, but was present in the stool of six out of 11 pCD patients that had, with a mean relative abundance of 0.52%. Increased G*emella* has been seen in colorectal cancer and in the pulmonary microbiome during exacerbations of cystic fibrosis.[204, 205] *Gemella haemolysans* has also been associated with Kawasaki disease in children, and has been known to cause a number of serious infections in patients, including: endocarditis, meningitis, and abscesses.[206-209] Normally present in the upper respiratory and gastrointestinal tracts, in a proposed linkage, increased intestinal permeability in IBD may allow *Gemella* access to the bloodstream where these systemic infections can occur.[210] It is possible that the higher relative abundances seen here are related to decreased relative abundances in species susceptible to the antimicrobials, or an increased dysbiosis with potential *Gemella* expansion after antimicrobial reduction in competing taxa.

Glucocorticoid use was present in both cohorts, but significant associations were only found in the pUC cohort, where use was more prevalent. Twelve of 22 pUC patients that provided stool were using GCS at baseline compared with 11 out of 33 pCD ($p=0.118$, χ^2), and at dose three, seven out of 22 pUC patients were taking GCS, while only one out of 33 pCD

patientswere taking GCS at dose five ($p=0.003$, χ^2). GCS are not used as maintenance therapy in pIBD, but instead to induce remission during active disease flares.[107] The associations with GCS use in pUC found here included lower SCFA concentrations and microbiota diversity; a similar association as was found with increased disease severity. The metabolites reduced in those taking GCS include: citric acid, oxoglutaric acid, vanillylmandelic acid, and quinolinic acid, with trending higher concentrations of hydroxyphenyllactic acid (**Figure 37**). GCS use has been associated in animals with increased *Bifidobacterium* and *Lactobacillus* species,[110] which, as described above, are known to produce hydroxyphenyllactic acid, likely from dietary compounds.[162] Potentially, this highlights a additional mechanism by-which GCS may result in remission in IBD, as hydroxyphenyllactic acid can reduce reactive oxygen species in neutrophils and may function as an anti-oxidant.[162] Although there was no significant increase in *Bifidobacterium* or *Lactobacillus* identified here, there was a higher relative abundance of *Proteus* associated with GCS use in pUC. *Proteus* spp. have been implicated as pathogens enriched in IBD, with the ability to induced colitis in animal models; it is possible that increased *Proteus* is part of the reason for increased need of GCS in these patients.[211]

Of the metabolites reduced in those taking GCS (citric acid, oxoglutaric acid, vanillylmandelic acid, and quinolinic acid), oxoglutaric acid increased significantly with IFX treatment, and oxoglutaric and citric acids were at trending higher concentrations with patients in remission in pCD. As GCS are used in flares, these metabolites may indicate flaring status with increased disease severity in these patients, likely accompanied by reduced dietary intake as hypothesized above. Quinolinic acid, as described above, may be involved in inflammation. The anti-inflammatory effects of GCS may result in reduced quinolinic acid synthesis. GCS are known to influence tryptophan metabolism, which is the proposed mechanism by which they can

cause sleep and mood disturbances in patients,[212] and may also impact quinolinic acid production. A reduction in vanillylmandelic acid with GCS use and an associated increase with IFX treatment suggests that it may be a marker of decreasing disease activity.

4.1.3 Associations with Disease Activity and Remission

In pCD the four metabolites significantly lower in patients with moderate or severe disease activity included: glycolic acid, 3-hydroxyphenylacetic acid, 4-hydroxymandelic acid, and 2,5-furandicarboxylic acid. Given that only when specimens from both timepoints were combined for analysis were significant compounds identified, and as disease activity significantly decreased with treatment (**Table 9**), it is likely that these metabolites are related to the decrease with disease severity seen with IFX treatment. Urinary metabolites that had trending higher concentrations in pCD patients with established remission (**Table 40**) may serve as biomarkers for treatment success, and include those that were significantly increased with treatment such as – oxoglutaric acid and citric acid as previously discussed. Other trending metabolites that can indicate dietary intake include trans-aconitic acid, which has been considered a biomarker for soy consumption, and methylmalonic acid, an intermediate in fat and protein metabolism.[169] Indicators of altered microbial abundances or function in remission include 2-methylsuccinic acid, which may be produced by *Clostridium* species.[213] Homovanillic acid had a trending higher concentration in remission patients, and has many dietary associations, including: olives, avocados, dairy milk, and red wine, and is associated with *Bifidobacterium*.[214] Glycolic acid as previously mentioned has numerous dietary and bacterial associations, which are likely also impacted by the inflammation accompanying increased disease severity.

SCFA concentrations were significantly lower in pCD patients with moderate or severe disease, but this was also identified when specimens from all timepoints were combined for analysis, and likely also reflects the decreasing disease severity with IFX treatment. However, SCFA concentrations have been shown to be reduced in adult IBD with increased disease activity.[203]

The differential microbial abundances in pCD with moderate or severe disease include lower relative abundances in: *Akkermansia*, *Blautia*, *Gemmiger*, *Ruminococcus*, *Coprococcus*, *Oscillospira*, *Pseudobutyrivibrio*, and *Adlercreutzia*. Lower relative abundances seen at the phylum level in Verrucomicrobia with moderate or severe disease were entirely accounted for by lower numbers of *Akkermansia*. Of the noted changes, only *Gemmiger* and *Adlercreutzia* were not also significantly increased or trending increased with IFX therapy in pCD. *Gemmiger* has been found to be lower in relative abundance in UC patients during a disease flare compared to controls.[215] *Gemmiger* is an anaerobe that can produce formate and butyrate through carbohydrate fermentation,[216] which may partially explain the lower SCFA concentrations in patients with increased disease severity. *Adlercreutzia* may be involved with equol production, a biologically active form of isoflavone that requires microbial fermentation of soy and is thought to explain at least part of the health benefits associated with soy consumption.[217, 218] *Adlercreutzia* are also anaerobes, so may be decreased in the presence of increased reactive oxygen species with inflammation.[219]

In pUC there was multivariate significance or trends with urine metabolites and phylum and genus level microbial changes, but no significance identified on univariate analysis. The significant decreases in SCFA changes associated with a final FCP <250 (**Figure 21**), may be attributable to sex related changes, as discussed below.

4.1.4 Associations with Disease Location

Associations with disease location were found consistently in pCD and pUC. It has been previously found that microbiome variations are significant between ICD, CCD, and UC.[220] Urinary metabolites and microbial shifts identified here were significantly altered in ICD vs CCD in pCD, as well as in pancolitis vs non-pancolitis in pUC.

Glycolic acid and HPHPA were at elevated concentrations in ICD compared to CCD (**Figure 30**). While there was an altered microbial profile identified on multivariate analysis with disease location in pCD, no taxa were identified as significantly or trending different on univariate analysis. This suggests that alterations in the *Clostridia* linked metabolite HPHPA may reflect altered microbial function. The colon is home to more anaerobes, such as *Clostridia* species,[221] than elsewhere in the gut, and disease activity here may impact microbial metabolism, and hence, production of HPHPA.[188, 189] The changes in glycolic acid may reflect dietary and/or microbial differences given its broad range of associations. As described, other studies have shown significant fecal microbial variations based on disease location in pCD,[47] but such changes were only seen on multivariate analysis here and individual taxa changes were not found. The presence of colonic vs upper disease in CD may differentially influence the metabolism of microbes living in those intestinal environments.

Pancolitis in pUC showed consistent alterations in urine metabolites and stool microbiota. Impacted metabolites included lower levels of pyroglutamic acid, uracil, methylmalonic acid, vanillylmandelic acid, glycolic acid, fumaric acid, homovanillic acid, p-hydroxybenzoic acid, and 3-hydroxy-2-methylbutyric acid. Pancolitis in UC may predict an increased rate of complications, more severe disease, and an increased likelihood of colectomy.[222] Metabolomic

changes specific to pancolitis in UC have not been well studied. Pyroglutamic acid may be a biomarker for cheese consumption, and elevation may be associated with glutamine and/or glutathione regeneration problems.[169] Regeneration of glutathione is important in controlling damage caused by inflammation; a more extensive inflammation, such as that found in pancolitis, may explain altered metabolites involved in glutathione regeneration. Uracil also has dietary associations, including: yellow wax beans, daikon radish, and hazelnuts.[169] Fumarate is an intermediate in the CAC, and has previously been found to be at lower concentrations in UC patient serum compared to controls.[223] Fumarate is also a metabolite of the fungus *Aspergillus*, a pathogen whose overgrowth has been associated with rare complications in IFX therapy, although none of these complications were seen here.[224, 225] p-hydroxybenzoic acid is a metabolite in all living species; it is involved in ubiquinone synthesis in humans, and is associated with certain foods such as coriander, onions, soy beans, and rye bread; it is also associated with *Pseudomonas* species.[169] 3-hydroxy-2-methylbutyric acid is a metabolite from isoleucine metabolism.[169] The increased extent of inflammation in pancolitis vs non-pancolitis in pUC could result in a more dramatic shift of metabolites that may already by affected in pIBD. Given that, as mentioned, pancolitis patients are more likely to have worse outcomes, although it has not been studied in pUC, it is possible that these patients experience worse symptoms, and have more limited dietary intake, hence the reduced levels of diet associated metabolites.

The microbiome associations with pancolitis in pUC included lower relative abundances of *Gemmiger* and a trending higher relative abundance of *Lactobacillus*. Given previously established associations between increased disease severity and pancolitis, and lower relative abundances of *Gemmiger* in flaring UC,[215] a pancolitis phenotype may show more extreme

variations of the same UC microbial alterations due to the extent and associated severity of disease.

4.1.5 Associations with Age

A number of significant associations were found in urinary metabolites with patient age in pUC. These associations included decreasing with age of: succinic, homovanillic, suberic, 3-hydroxy-2-methylbutyric, glycolic, and quinolinic acids. Variable associations with age were: suberic, hippuric, lactic, and fumaric acids. Urinary creatinine exhibited a strong correlation with age at both timepoints in the pUC cohort; this is consistent with findings by others in pediatric populations.[226, 227] Although it is likely that these findings result from confounding by the associations with increasing creatinine and age, Gu *et al.* (2009) also found succinate levels to increase with age in children up to 12 years when concentrations were not normalized to creatinine levels.[228]

4.1.6 Associations with Sex

There were no significant or trending associations with sex in pCD, but sex presented with interesting associations across SCFA, microbiota abundances, and microbial diversity in pUC. There was a trend for males with pUC to have higher acetate, butyrate, and total SCFA content at baseline (p=0.073, **Figure 32**). This was accompanied by an increased microbial diversity and richness in males at baseline. Females saw a mean increase in acetate and SCFA that were of higher magnitude than the decreases in acetate, butyrate, and total SCFA that were seen in all four male pUC patients (**Figure 34**). This resulted in a reversal of the baseline trend,

with females having a trending higher acetate (p=0.88) and non-significantly higher total SCFA, propionate, and butyrate after treatment (**Figure 33**). Females saw significant increases in microbial richness, as measured by Chao1 and OTU numbers, while males did not, and after treatment, there were no significant differences in these measures between males and females (**Figure 35**). The associated changes in microbial diversity and abundance mirror those seen in SCFA concentration. These sex specific associations in SCFA with IFX therapy have not been previously identified in the literature.

In response to oligofructose (a dietary fibre) supplementation, male rats had an increase in butyrate that was not seen in female rats.[229] After eating a high-fibre meal, male human patients (a combination of healthy controls, celiac disease, and microscopic colitis patients) had higher concentrations of serum SCFAs, including: acetate, isovalerate, propionate, and butyrate.[230] McOrist *et al*. (2010) found that men had higher SCFAs on study entry and in response to resistant starch or nonstarch polysaccharide supplementation.[231] Significant differences in male and female microbiota and differential response to diet based on sex have been documented in a number of other studies.[231-233] With an average age of 12.6 years in pUC, many of the patients have entered puberty, resulting in significant differences in sex hormone levels between male and female patients, which could explain some of these effects.[234, 235]

In terms of outcomes based on sex, a large systematic review looking at 994 pCD and 416 pUC patients, assessing for sex differences in pIBD treatment response found no associations with treatment outcomes or medication risks.[236] In another study on treatments in pIBD, there was a non significant trend for IFX sustaining remission longer in males than females.[237] Methotrexate was also associated with significantly more success with IFX therapy in males, but was only trending the same association in females.[237] There is also a rare complication

of anti-TNF therapy, hepatosplenic T-cell lymphoma that occurs significantly more frequently in male patients, especially younger than 35 years of age.[238] Additionally, in pharmacokinetic studies, males have shown higher clearance of anti-TNF antibodies.[239]

Another possible explanation for difference concentration in SCFA in males vs females is differential utilization of SCFA. SCFA can diffuse across the epithelial layer at varying rates,[203] but absorption, particularly of butyrate, is enhanced by the monocarboxylate transporter (MCT) 1. During inflammation MCT1 expression is decreased in IBD patients and *in vitro* with the addition of TNF-α.[240] Sex hormone effects on MCT1 expression have been found in the livers of rats,[241] and sex had significant interactions with breed of pig on the levels of MCT2 intestinal colon expression in pigs.[242] Future studies assessing the colonic mucosa of pIBD patients for MCT expression could help account for an important variable on the health impacts of SCFA.

The differential responses to dietary fibre in SCFA production between sexes could explain the trends for males to have higher fecal SCFA as baseline. The findings by others of differential responses to therapy between the sexes could help explain the different changes over time between males and females. Although in a separate cohort, and in pCD opposed to pUC, associations with sex were also found in the EEN-FFQ pCD cohort, as discussed below. Diet variation between males and females could explain the variation in SCFA content.

4.2 EEN-FFQ Cohort

Sex differences in reported dietary intake were also found in the pCD EEN-FFQ cohort. Inclusion of these findings here provides a possible driver of the sex-differences identified in SCFA concentrations in the pUC IFX cohort, although this is completely speculative given that

these are different patients. Variations in diet in pediatric populations by sex have also been found by others. Dietary pattern analysis by D'Souza *et al*. (2008) in children and adolescents at risk for developing pCD also identified different dietary patterns in males vs females.[88] Variability in meat, fried/fast foods, and dessert consumption was important in females, and was defined as a Western pattern.[88] A "Prudent" pattern was important in characterizing dietary variability in both males and females.[88] The "Prudent" dietary pattern identified by D'Souza *et al*. is highly similar to the "Mature" dietary pattern we identified here; both are characterized by increased intake of vegetables, fruits, fish, and seafood.[88] Increasing fruits and vegetables is recommended in CD with the aim of reducing bacterial mucin metabolism and increasing SCFA production by providing dietary fibre for microbial fermentation.[243, 244] The "Prudent" diet's negative association with pCD risk in boys and girls[88] makes the similar "Mature" dietary pattern in our cohort of particular interest in examining for associations with EEN treatment outcomes. The higher fruit and vegetable intake with higher adherence to the mature diet could result in increased dietary fibre intake. If the higher adherence of male pCD patients to this diet identified here was also present in our pUC cohort, it could explain increased SCFA content at baseline due to the increase dietary fibre substrate for microbial fermentation (however, we did not have access to diet data in our IFX cohorts).

Within our pCD EEN-FFQ cohort, an association between the "Pre-packaged" and "Mature" dietary patterns and disease location was only present in females. It is possible that dietary impacts may be sex specific. In some cohorts, pCD patients with ICD have been found more likely to succeed with EEN than those with CCD;[245] therefore, it is possible that previous diet may play a role in determining therapy response. While these diets are not continued during

EEN therapy, the established microbiome may mediate EEN response. This will be examined in future analyses of response to therapy in the EEN-FFQ cohort.

4.4 Limitations

These two unique, prospective cohorts, where protocolized and detailed clinical data and biospecimens were collected, offered an excellent opportunity to longitudinally examine the host-microbe relationship in response to therapy in pIBD. However, there are some challenges inherent to this type of research, which need to be considered when analyzing the study results, some of which are related to inclusion criteria for patients in the studies. The addition of healthy controls could assist in identifying variations specific to the pathology of pIBD. Patients were not randomly assigned to receive treatment, and no control group was available. As both the IFX and EEN-FFQ cohorts were in observational studies, all of findings of both studies are correlations and associations that cannot prove causation (although treatment did follow a protocol). Due to the small sample sizes, especially for pUC, many important associations may have not reached statistical significance despite being potentially of clinical significance. Although there were relatively few patients, there was a rich clinical dataset and in many cases more variables than samples. Effects of other medications, age at diagnosis, race, diet, and a number of other variables could have been assessed, but this would have substantially increased the risk of inflating Type I error. This also creates another limitation of the study that adjusting for other variables was often impossible with other diminishing group sizes too low for statistical analysis.

Due to the limited accuracy of 16S sequencing, we could not identify microbiota to the species level, and in the absence of metagenomic sequencing we could not assess microbial function (although some of this could be implied by the SCFA and metabolite findings, as discussed above). Our 16S sequencing results have also not been verifiable by qPCR so will require re-extraction of stool and repeat sequencing to confirm validity. Inhibition by stool

contaminants such as humic acid or polyphenols are not uncommon,[157] but will require appropriate mediation in order to obtain verifiable results. These are some of the challenges of working with human patient samples, especially in a multi-centre setting.[246] Utilizing a pre-established organic acids urine panel for metabolomics allows accurate identification of each metabolite but does not capture many of the metabolites that could otherwise be identified. Additionally, urinary metabolites, as previously identified, are highly variable with diet and other external factors, and represent many endogenous metabolites, making it difficult to identify variations that may be due to microbial metabolism. Patient diet was not assessed as part of the IFX cohort, and has been shown to be an important influence in SCFA, microbiota, and metabolites, and as mentioned earlier, likely changed in response to therapy and could account for some of the changes associated with IFX therapy.

Despite all these limitations, some solid findings and links, contributing to our limited understanding of the host-microbe interface in pIBD, have been gained throughout this study. With the large number of variables examined in the described cohorts, results are not able to be summarized into a single unifying hypothesis explaining the physiology of treatment response in pIBD. However, these findings can contribute to understanding the complex host-microbe interplay in IBD.

4.3 Conclusion

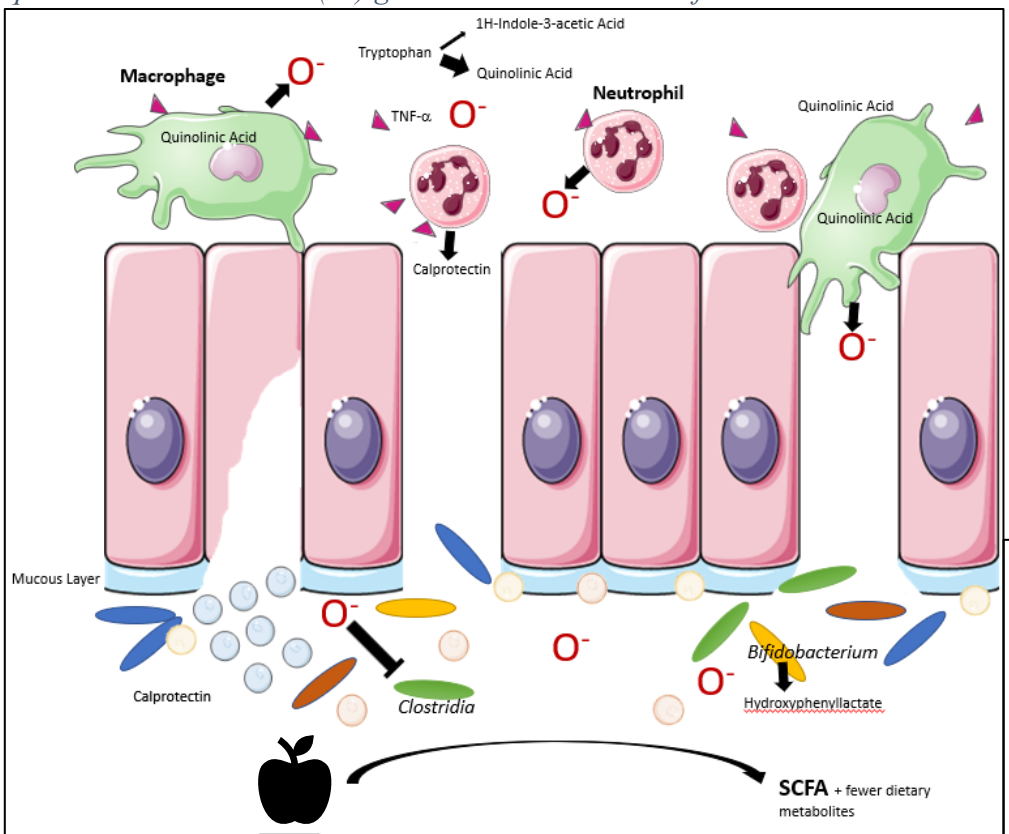
We hypothesized that a patient's microbiome, metabolites, and dietary patterns are associated with their response to therapy in pIBD. For our first objective we identified microbial and diet-associated metabolites associated with IFX therapy. For our second objective we identified dietary patterns that were associated with patient sex, age, and disease phenotype. Better understanding of the dysregulated host-microbe relationship in pIBD can allow us to improve current therapies and develop biomarkers for therapy candidacy and success. Development of novel therapeutics that specifically target the host-microbe relationship could improve outcomes for pIBD patients.

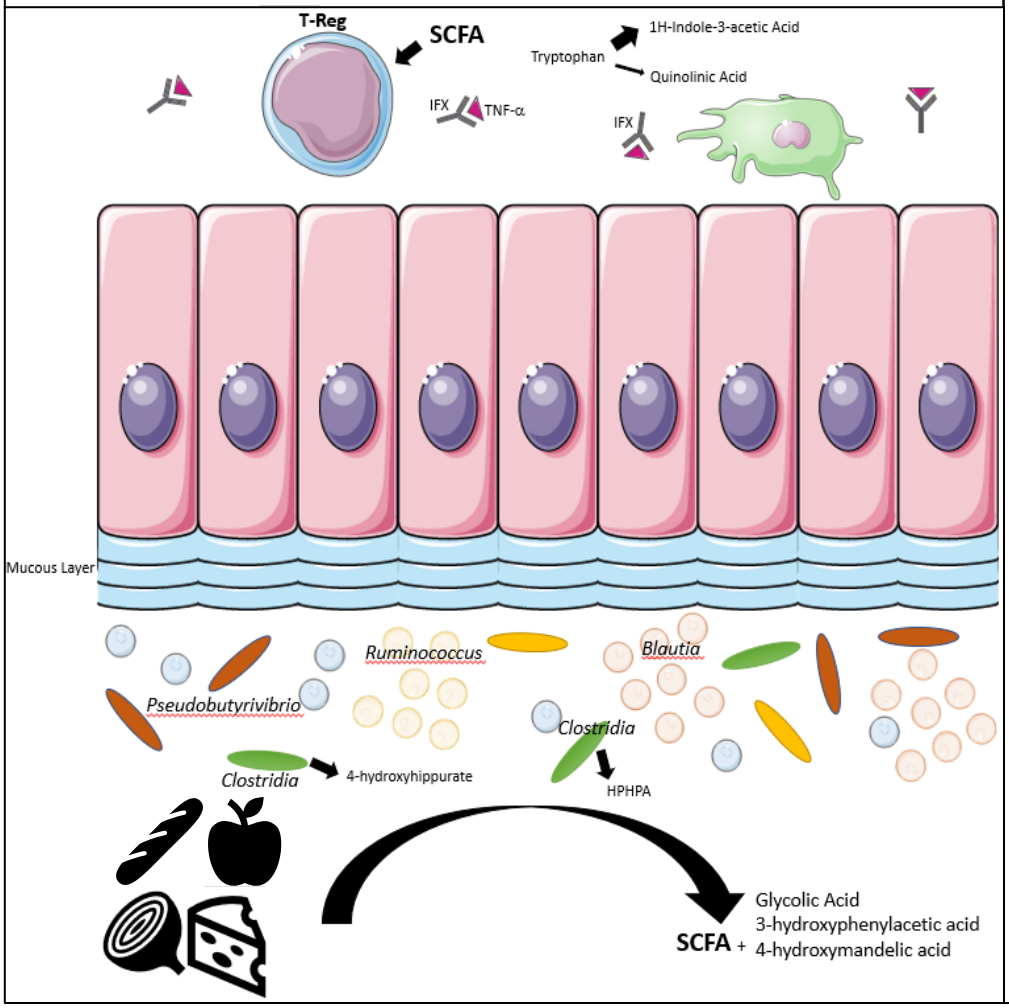
Within the IFX cohorts, final remission could not be predicted by any baseline SCFA, metabolite, or microbe identified on univariate analysis, suggesting that these could not be used to better identify IFX therapy candidates (although there was a trend for high OTU numbers at baseline in pUC patients who would achieve a final FCP <250). There was a significant number of metabolites, microbes, and SCFA that were altered with treatment, but no profile could identify the patients that successfully achieved remission vs. those who did not, suggesting that these could not be used to determine therapy success. Trending increases in metabolites in pCD patients in remission suggest that there may be changes that can define patients in remission, but a larger sample size may be required to determine this. It is possible that with a longer duration of therapy in either pCD or pUC that any microbes or metabolites identifying patients in remission may become more distinct but based upon results here we could not offer an early prediction of a patient likely to fail current therapy. The increases in SCFA and SCFA-producing microbiota with IFX therapy could identify the improved host-microbe interaction that helps explain IFX treatment effects. The proposed models mentioned above in the discussion section

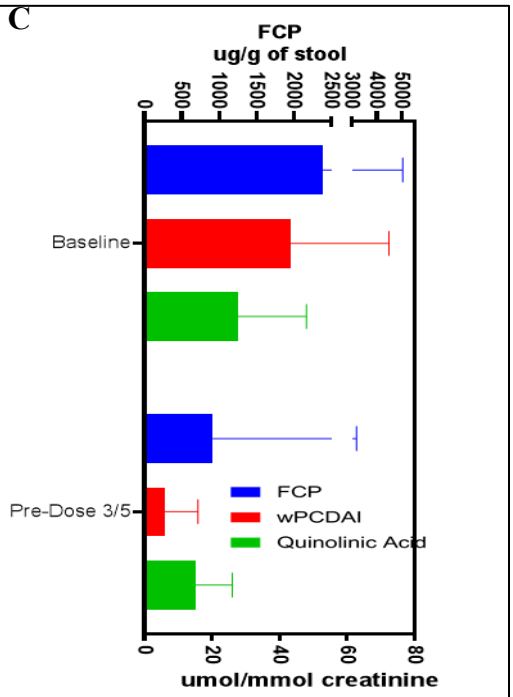
are visualized below in **Figure 42**. The host-microbe relationship may serve as an additional target for enhancing therapy or understanding mechanisms of disease. The sex differences in pUC highlight the importance of assessing sex-specific mechanisms of disease, as well as response to therapies.

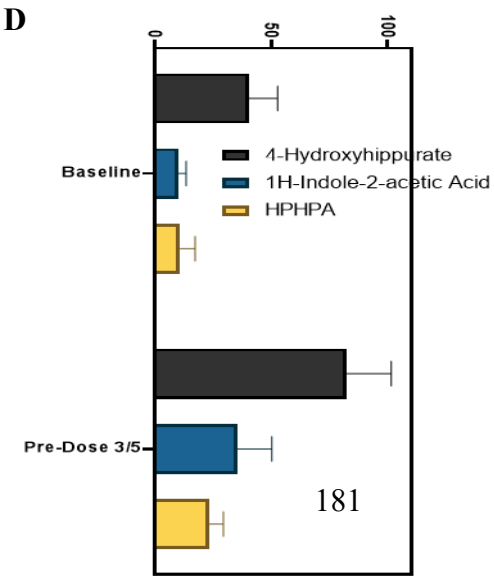
Figure 42 *Proposed model of IFX effect in pCD.* ***A*** *Uncontrolled TNFα signalling perpetuating neutrophil infiltration, quinolinic acid and ROS (O^-) generation → inhibition of anaerobic metabolism. Restricted dietary intake results in fewer diet related metabolites.* ***B*** *IFX inhibition of TNFα results in fewer ROS, Clostridia sp. exhibit increased function. Increased Blautia et al. produce more SCFA.* ***C*** *Decreased FCP, wPCDAI and quinolinic acid reflect decreasing inflammation.* ***D*** *Increased microbial metabolites in response to decreased inflammation.*

Cell images: www.smart.servier.com









4.5 Future Directions

Validation of 16S sequencing results is currently underway by qPCR assessing the abundances of total bacteria, *Akkermansia¸ Faecalibacterium*, and *Bifidobacterium*. Comparisons will be made between sequencing performed at Génome Québec and TAGC.

After 16S validation is performed, PCA will be run on 16S data at each timepoint to obtain PC scores. PC scores from metabolite, SCFA, and 16S PCAs will be correlated to assess for potential associations. Significant relationships will then be investigated by correlation by Spearman rank analysis of variables with high factor loadings in the first two principal components of each PCA.

Future analysis to assess for sex-related differences that may account for SCFA and 16S changes may be explored. Serum specimens collected for pharmacokinetic purposes may be utilized to assess steroid hormone levels for correlation with SCFA and 16S changes.

PICRUST microbial function predictions will be assessed to provide some estimations of microbial metabolism based on sequencing, with the understanding that these are lower accuracy estimations and genomic sequencing, mRNA, or protein analysis would provide more accurate information about microbial function.

Diet has an acute and important influence on urine metabolomics.[247] Urinary metabolomics has been shown able to differentiate dietary intakes such as vegetables and red meat.[248] Prospective FFQs collected as apart of the CIDsCaNN cohort were not included in the IDeaL ethics, and ethics approval has been submitted to include longitudinal dietary data as a variable potentially related to the above described findings in the pCD cohort. When these results

are available, dietary intake will be assessed for correlation with SCFA, metabolite, and microbiome data.

References:

1. Coward S, Clement F, Benchimol EI, et al. Past and Future Burden of Inflammatory Bowel Diseases Based on Modeling of Population-Based Data. *Gastroenterology*. Apr 2019;156(5):1345-1353.e4. doi:10.1053/j.gastro.2019.01.002
2. Kuenzig E, Benchimol EI, Bernstein CN, et al. The Rising Prevalence of Inflammatory Bowel Diease in Canada: Analyzing the Past to Predict the Future. *JCAG*. 2018;1(suppl_2):47-48. doi:10.1093/jcag/gwy009.029 %J Journal of the Canadian Association of Gastroenterology
3. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature*. 2011;474(7351):307-317. doi:10.1038/nature10209
4. Mamula P, Telega GW, Markowitz JE, et al. Inflammatory bowel disease in children 5 years of age and younger. *The American Journal of Gastroenterology*. 2002/08/01/ 2002;97(8):2005-2010. doi:https://doi.org/10.1016/S0002-9270(02)04272-7
5. Corica D, Romano C. Biological therapy in pediatric inflammatory bowel disease. *Journal of clinical gastroenterology*. 2017;51(2):100-110.
6. MacKner LM, Crandall WV, Szigethy EM. Psychosocial functioning in pediatric inflammatory bowel disease. *Inflammatory bowel diseases*. 2006;12(3):239-244.
7. Polito JM, Childs B, Mellits ED, Tokayer AZ, Harris ML, Bayless TM. Crohn's disease: Influence of age at diagnosis on site and clinical type of disease. *Gastroenterology*. 1996/09/01/ 1996;111(3):580-586. doi:https://doi.org/10.1053/gast.1996.v111.pm8780560
8. Levine A, Griffiths A, Markowitz J, et al. Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. *Inflammatory bowel diseases*. 2011;17(6):1314-1321.
9. Vernier–Massouille G, Balde M, Salleron J, et al. Natural history of pediatric Crohn's disease: a population-based cohort study. *Gastroenterology*. 2008;135(4):1106-1113.
10. Sawczenko A, Sandhu BK. Presenting features of inflammatory bowel disease in Great Britain and Ireland. *Archives of Disease in Childhood*. 7 January 2003 2003;88:995-1000.
11. Frolkis AD, Dykeman J, Negrón ME, et al. Risk of surgery for inflammatory bowel diseases has decreased over time: a systematic review and meta-analysis of population-based studies. *Gastroenterology*. 2013;145(5):996-1006.
12. Ross SC, Strachan J, Russell RK, Wilson SL. Psychosocial functioning and health-related quality of life in paediatric inflammatory bowel disease. *Journal of pediatric gastroenterology and nutrition*. 2011;53(5):480-488.
13. Malham M, Jakobsen C, Paerregaard A, Virta LJ, Kolho KL, Wewer V. The incidence of cancer and mortality in paediatric onset inflammatory bowel disease in Denmark and Finland during a 23-year period: a population-based study. *Alimentary pharmacology & therapeutics*. 2019;50(1):33-39.
14. Schoepfer AM, Beglinger C, Straumann A, et al. Fecal calprotectin correlates more closely with the Simple Endoscopic Score for Crohn's disease (SES-CD) than CRP, blood leukocytes, and the CDAI. *American Journal of Gastroenterology*. 2010;105(1):162-169.
15. Bunn SK, Bisset WM, Main MJ, Golden BE. Fecal calprotectin as a measure of disease activity in childhood inflammatory bowel disease. *Journal of pediatric gastroenterology and nutrition*. 2001;32(2):171-177.
16. Kostakis ID, Cholidou KG, Vaiopoulos AG, Vlachos IS, Perrea D, Vaos G. Fecal calprotectin in pediatric inflammatory bowel disease: a systematic review. *Digestive Diseases and Sciences*. 2013;58(2):309-319.
17. Turner D, Levine A, Walters TD, et al. Which PCDAI version best reflects intestinal inflammation in pediatric Crohn disease? *Journal of pediatric gastroenterology and nutrition*. 2017;64(2):254-260.

18. Kugathasan S, Judd RH, Hoffmann RG, et al. Epidemiologic and clinical characteristics of children with newly diagnosed inflammatory bowel disease in Wisconsin: a statewide population-based study. *The Journal of pediatrics*. 2003;143(4):525-531.
19. Knights D, Lassen KG, Xavier RJ. Advances in inflammatory bowel disease pathogenesis: linking host genetics and the microbiome. *Gut*. Oct 2013;62(10):1505-10. doi:10.1136/gutjnl-2012-303954
20. Cleynen I, González JR, Figueroa C, et al. Genetic factors conferring an increased susceptibility to develop Crohn's disease also influence disease phenotype: results from the IBDchip European Project. *Gut*. 2013;62(11):1556-1565. doi:10.1136/gutjnl-2011-300777
21. Al Nabhani Z, Dietrich G, Hugot J-P, Barreau F. Nod2: The intestinal gate keeper. *PLoS pathogens*. 2017;13(3)
22. Bonen DK, Ogura Y, Nicolae DL, et al. Crohn's disease-associated NOD2 variants share a signaling defect in response to lipopolysaccharide and peptidoglycan. *Gastroenterology*. 2003;124(1):140-146.
23. Tan G, Zeng B, Zhi F-C. Regulation of human enteric α-defensins by NOD2 in the Paneth cell lineage. *European Journal of Cell Biology*. 2015/01/01/ 2015;94(1):60-66. doi:https://doi.org/10.1016/j.ejcb.2014.10.007
24. Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*. 2001;411(6837):603-606.
25. Hisamatsu T, Suzuki M, Reinecker H-C, Nadeau WJ, McCormick BA, Podolsky DK. CARD15/NOD2 functions as an antibacterial factor in human intestinal epithelial cells. *Gastroenterology*. 2003;124(4):993-1000.
26. Schäffler H, Geiss D, Gittel N, et al. Mutations in the NOD2 gene are associated with a specific phenotype and lower anti-tumor necrosis factor trough levels in Crohn's disease. *Journal of Digestive Diseases*. 2018;19(11):678-684. doi:10.1111/1751-2980.12677
27. Prager M, Büttner J, Haas V, et al. The JAK2 variant rs10758669 in Crohn's disease: altering the intestinal barrier as one mechanism of action. *International Journal of Colorectal Disease*. 2012/05/01 2012;27(5):565-573. doi:10.1007/s00384-011-1345-y
28. Hedl M, Proctor DD, Abraham C. JAK2 disease-risk variants are gain of function and JAK signaling threshold determines innate receptor-induced proinflammatory cytokine secretion in macrophages. *The Journal of Immunology*. 2016;197(9):3695-3704.
29. Ghoreschi K, Jesson MI, Li X, et al. Modulation of innate and adaptive immune responses by tofacitinib (CP-690,550). *The Journal of Immunology*. 2011;186(7):4234-4243.
30. Feagan B. Update on tofacitinib for inflammatory bowel disease. *Gastroenterology & hepatology*. 2016;12(9):572.
31. Zhernakova A, Festen EM, Franke L, et al. Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. *Am J Hum Genet*. May 2008;82(5):1202-10. doi:10.1016/j.ajhg.2008.03.016
32. Leshchiner ES, Rush JS, Durney MA, et al. Small-molecule inhibitors directly target CARD9 and mimic its protective variant in inflammatory bowel disease. *Proceedings of the National Academy of Sciences of the United States of America*. 2017;114(43):11392-11397. doi:10.2307/26488594
33. Ruland J. CARD9 signaling in the innate immune response. *Annals of the New York Academy of Sciences*. 2008;1143(1):35-44.
34. Hsu Y-MS, Zhang Y, You Y, et al. The adaptor protein CARD9 is required for innate immune responses to intracellular pathogens. *Nature Immunology*. 2007/02/01 2007;8(2):198-205. doi:10.1038/ni1426
35. Zaidi D, Wine E. Regulation of nuclear factor kappa-light-chain-enhancer of activated B Cells (NF-κβ) in inflammatory bowel diseases. *Frontiers in pediatrics*. 2018;6:317.

36. Zaidi D, Huynh HQ, Carroll MW, Baksh S, Wine E. Tumor necrosis factor α-induced protein 3 (A20) is dysregulated in pediatric Crohn disease. *Clinical and experimental gastroenterology*. 2018;11:217.
37. Peters LA, Perrigoue J, Mortha A, et al. A functional genomics predictive network model identifies regulators of inflammatory bowel disease. *Nature genetics*. 2017;49(10):1437.
38. Sender R, Fuchs S, Milo R. Revised estimates for the number of human and bacteria cells in the body. *PLoS biology*. 2016;14(8):e1002533.
39. Shreiner AB, Kao JY, Young VB. The gut microbiome in health and in disease. *Current opinion in gastroenterology*. 2015;31(1):69.
40. Kamada N, Núñez G. Regulation of the immune system by the resident intestinal bacteria. *Gastroenterology*. 2014;146(6):1477-1488.
41. Prosberg M, Bendtsen F, Vind I, Petersen AM, Gluud LL. The association between the gut microbiota and the inflammatory bowel disease activity: a systematic review and meta-analysis. *Scandinavian journal of gastroenterology*. 2016;51(12):1407-1415.
42. Lane ER, Zisman TL, Suskind DL. The microbiota in inflammatory bowel disease: current and therapeutic insights. *Journal of inflammation research*. 2017;10:63.
43. Gevers D, Kugathasan S, Denson LA, et al. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe*. Mar 12 2014;15(3):382-392. doi:10.1016/j.chom.2014.02.005
44. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012;489(7415):220-230.
45. Ho NT, Li F, Lee-Sarwar KA, et al. Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. *Nature communications*. 2018;9(1):1-13.
46. Zhou Y, Zhi F. Lower level of bacteroides in the gut microbiota is associated with inflammatory bowel disease: a meta-analysis. *BioMed research international*. 2016;2016
47. Willing B, Halfvarson J, Dicksved J, et al. Twin studies reveal specific imbalances in the mucosaassociated microbiota of patients with ileal Crohn's disease. *Inflammatory bowel diseases*. 2009;15(5):653-660.
48. Nishino K, Nishida A, Inoue R, et al. Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. *Journal of gastroenterology*. 2018;53(1):95-106.
49. Shah R, Cope JL, Nagy-Szakal D, et al. Composition and function of the pediatric colonic mucosal microbiome in untreated patients with ulcerative colitis. *Gut Microbes*. 2016/09/02 2016;7(5):384-396. doi:10.1080/19490976.2016.1190073
50. Michail S, Durbin M, Turner D, et al. Alterations in the gut microbiome of children with severe ulcerative colitis. *Inflammatory bowel diseases*. 2012;18(10):1799-1808. doi:10.1002/ibd.22860
51. Schirmer M, Denson L, Vlamakis H, et al. Compositional and temporal changes in the gut microbiome of pediatric ulcerative colitis patients are linked to disease course. *Cell host & microbe*. 2018;24(4):600-610. e4.
52. Alipour M, Zaidi D, Valcheva R, et al. Mucosal Barrier Depletion and Loss of Bacterial Diversity are Primary Abnormalities in Paediatric Ulcerative Colitis. *Journal of Crohn's & colitis*. 2016;10(4):462-471. doi:10.1093/ecco-jcc/jjv223
53. Zaidi D, Bording-Jorgensen M, Huynh HQ, et al. Increased Epithelial Gap Density in the Noninflamed Duodenum of Children With Inflammatory Bowel Diseases. *Journal of pediatric gastroenterology and nutrition*. Dec 2016;63(6):644-650. doi:10.1097/mpg.0000000000001182
54. Kiesslich R, Duckworth CA, Moussata D, et al. Local barrier dysfunction identified by confocal laser endomicroscopy predicts relapse in inflammatory bowel disease. *Gut*. 2012;61(8):1146-1153. doi:10.1136/gutjnl-2011-300695
55. Brown EM, Sadarangani M, Finlay BB. The role of the immune system in governing host-microbe interactions in the intestine. *Nature immunology*. 2013;14(7):660.

56. Smith K, McCoy KD, Macpherson AJ. Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. Elsevier; 2007:59-69.
57. Haverson K, Rehakova Z, Sinkora J, Sver L, Bailey M. Immune development in jejunal mucosa after colonization with selected commensal gut bacteria: a study in germ-free pigs. *Veterinary immunology and immunopathology*. 2007;119(3-4):243-253.
58. Troy EB, Kasper DL. Beneficial effects of Bacteroides fragilis polysaccharides on the immune system. *Frontiers in bioscience: a journal and virtual library*. 2010;15:25.
59. Kozakova H, Schwarzer M, Tuckova L, et al. Colonization of germ-free mice with a mixture of three lactobacillus strains enhances the integrity of gut mucosa and ameliorates allergic sensitization. *Cellular & molecular immunology*. 2016;13(2):251-262.
60. Schicho R, Shaykhutdinov R, Ngo J, et al. Quantitative metabolomic profiling of serum, plasma, and urine by (1)H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals. *J Proteome Res*. Jun 1 2012;11(6):3344-57. doi:10.1021/pr300139q
61. Stephens NS, Siffledeen J, Su X, Murdoch TB, Fedorak RN, Slupsky CM. Urinary NMR metabolomic profiles discriminate inflammatory bowel disease from healthy. *Journal of Crohn's & colitis*. Mar 2013;7(2):e42-8. doi:10.1016/j.crohns.2012.04.019
62. Kolho KL, Pessia A, Jaakkola T, de Vos WM, Velagapudi V. Faecal and Serum Metabolomics in Paediatric Inflammatory Bowel Disease. *Journal of Crohn's & colitis*. Mar 1 2017;11(3):321-334. doi:10.1093/ecco-jcc/jjw158
63. Martin F-P, Ezri J, Cominetti O, et al. Urinary metabolic phenotyping reveals differences in the metabolic status of healthy and inflammatory bowel disease (IBD) children in relation to growth and disease activity. *International journal of molecular sciences*. 2016;17(8):1310.
64. Martin F-P, Su M-M, Xie G-X, et al. Urinary metabolic insights into host-gut microbial interactions in healthy and IBD children. *World J Gastroenterol*. 2017;23(20):3643-3654. doi:10.3748/wjg.v23.i20.3643
65. Lees HJ, Swann JR, Wilson ID, Nicholson JK, Holmes E. Hippurate: the natural history of a mammalian–microbial cometabolite. *Journal of proteome research*. 2013;12(4):1527-1546.
66. Sarosiek I, Schicho R, Blandon P, Bashashati M. Urinary metabolites as noninvasive biomarkers of gastrointestinal diseases: A clinical review. *World journal of gastrointestinal oncology*. 2016;8(5):459.
67. Pallister T, Jackson MA, Martin TC, et al. Hippurate as a metabolomic marker of gut microbiome diversity: Modulation by diet and relationship to metabolic syndrome. *Scientific Reports*. 2017/10/20 2017;7(1):13670. doi:10.1038/s41598-017-13722-4
68. Williams HR, Cox JI, Walker DG, et al. Characterization of inflammatory bowel disease with urinary metabolic profiling. *American Journal of Gastroenterology*. 2009;104(6):1435-1444.
69. Williams HR, Cox IJ, Walker DG, et al. Differences in gut microbial metabolism are responsible for reduced hippurate synthesis in Crohn's disease. *BMC gastroenterology*. 2010;10(1):108.
70. Dzúrik R, Spustová V, Krivošíková Z, Gazdíková K. Hippurate participates in the correction of metabolic acidosis. *Kidney International*. 2001;59:S278-S281.
71. Huda-Faujan N, Abdulamir AS, Fatimah AB, et al. The impact of the level of the intestinal short chain Fatty acids in inflammatory bowel disease patients versus healthy subjects. *Open Biochem J*. 2010;4:53-58. doi:10.2174/1874091X01004010053
72. Treem WR, Ahsan N, Shoup M, Hyams JS. Fecal short-chain fatty acids in children with inflammatory bowel disease. *Journal of pediatric gastroenterology and nutrition*. Feb 1994;18(2):159-64. doi:10.1097/00005176-199402000-00007
73. Suzuki T, Yoshida S, Hara H. Physiological concentrations of short-chain fatty acids immediately suppress colonic epithelial permeability. *Br J Nutr*. Aug 2008;100(2):297-305. doi:10.1017/s0007114508888733

10.1017/S0007114508888733. Epub 2008 Mar 18.
74. Smith PM, Howitt MR, Panikov N, et al. The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science*. 2013;341(6145):569-573.
75. Venegas DP, Marjorie K, Landskron G, et al. Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. *Frontiers in immunology*. 2019;10
76. Ahmad M, Krishnan S, Ramakrishna B, Mathan M, Pulimood A, Murthy S. Butyrate and glucose metabolism by colonocytes in experimental colitis in mice. *Gut*. 2000;46(4):493-499.
77. Agus A, Denizot J, Thevenot J, et al. Western diet induces a shift in microbiota composition enhancing susceptibility to Adherent-Invasive E. coli infection and intestinal inflammation. *Scientific reports*. 2016;6:19032.
78. Cavaglieri CR, Nishiyama A, Fernandes LC, Curi R, Miles EA, Calder PC. Differential effects of short-chain fatty acids on proliferation and production of pro- and anti-inflammatory cytokines by cultured lymphocytes. *Life Sci*. Aug 15 2003;73(13):1683-90. doi:10.1016/s0024-3205(03)00490-9
79. Segain JP, Raingeard de la Blétière D, Bourreille A, et al. Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease. *Gut*. 2000;47(3):397-403. doi:10.1136/gut.47.3.397
80. Wang Y, Gao X, Ghozlane A, et al. Characteristics of Faecal Microbiota in Paediatric Crohn's Disease and Their Dynamic Changes During Infliximab Therapy. *Journal of Crohn's & colitis*. Feb 28 2018;12(3):337-346. doi:10.1093/ecco-jcc/jjx153
81. Roediger W. Utilization of nutrients by isolated epithelial cells of the rat colon. *Gastroenterology*. 1982;83(2):424-429.
82. Sánchez de Medina F, Romero-Calvo I, Mascaraque C, Martínez-Augustin O. Intestinal Inflammation and Mucosal Barrier Function. *Inflammatory Bowel Diseases*. 2014;20(12):2394-2404. doi:10.1097/mib.0000000000000204
83. Topping DL, Clifton PM. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiological reviews*. 2001;81(3):1031-1064.
84. Pang T, Leach ST, Katz T, Day AS. Fecal biomarkers of intestinal health and disease in children. *Frontiers in pediatrics*. 2014;2:6.
85. Hou JK, Abraham B, El-Serag H. Dietary intake and risk of developing inflammatory bowel disease: a systematic review of the literature. *American Journal of Gastroenterology*. 2011;106(4):563-573.
86. Levine A, Boneh RS, Wine E. Evolving role of diet in the pathogenesis and treatment of inflammatory bowel diseases. *Gut*. 2018;67(9):1726-1738.
87. Niewiadomski O, Studd C, Wilson J, et al. Influence of food and lifestyle on the risk of developing inflammatory bowel disease. *Internal medicine journal*. 2016;46(6):669-676.
88. D'Souza S, Levy E, Mack D, et al. Dietary patterns and risk for Crohn's disease in children. *Inflammatory bowel diseases*. 2008;14(3):367-373.
89. Singh RK, Chang H-W, Yan D, et al. Influence of diet on the gut microbiome and implications for human health. *Journal of translational medicine*. 2017;15(1):73.
90. Johnson AJ, Vangay P, Al-Ghalith GA, et al. Daily sampling reveals personalized diet-microbiome associations in humans. *Cell host & microbe*. 2019;25(6):789-802. e5.
91. Martinez-Medina M, Garcia-Gil LJ. Escherichia coli in chronic inflammatory bowel diseases: an update on adherent invasive Escherichia coli pathogenicity. *World journal of gastrointestinal pathophysiology*. 2014;5(3):213.
92. Colman RJ, Rubin DT. Fecal microbiota transplantation as therapy for inflammatory bowel disease: a systematic review and meta-analysis. *Journal of Crohn's & colitis*. Dec 2014;8(12):1569-81. doi:10.1016/j.crohns.2014.08.006

93. Maxwell EC, Dawany N, Baldassano RN, et al. Diverting ileostomy for the treatment of severe, refractory, pediatric inflammatory bowel disease. *Journal of pediatric gastroenterology and nutrition*. 2017;65(3):299-305.
94. Keubler LM, Buettner M, Häger C, Bleich A. A multihit model: colitis lessons from the interleukin-10–deficient mouse. *Inflammatory bowel diseases*. 2015;21(8):1967-1975.
95. Schaubeck M, Clavel T, Calasan J, et al. Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence. *Gut*. 2016;65(2):225-237.
96. Roy U, Gálvez EJ, Iljazovic A, et al. Distinct microbial communities trigger colitis development upon intestinal barrier damage via innate or adaptive immune cells. *Cell reports*. 2017;21(4):994-1008.
97. Sokol H, Pigneur B, Watterlot L, 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):16731-16736.
98. Neurath MF. Cytokines in inflammatory bowel disease. *Nature Reviews Immunology*. 2014;14(5):329-342.
99. Rahman MM, McFadden G. Modulation of NF-κB signalling by microbial pathogens. *Nature Reviews Microbiology*. 2011;9(4):291-306.
100. Bevins CL, Salzman NH. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nature Reviews Microbiology*. 2011;9(5):356-368.
101. Comito D, Cascio A, Romano C. Microbiota biodiversity in inflammatory bowel disease. *Italian journal of pediatrics*. 2014;40(1):32.
102. Choung R, Princen F, Stockfisch T, et al. Serologic microbial associated markers can predict Crohn's disease behaviour years before disease diagnosis. *Alimentary pharmacology & therapeutics*. 2016;43(12):1300-1310.
103. Dubinsky MC, Kugathasan S, Mei L, et al. Increased immune reactivity predicts aggressive complicating Crohn's disease in children. *Clinical Gastroenterology and Hepatology*. 2008;6(10):1105-1111.
104. Birimberg-Schwartz L, Wilson DC, Kolho K-L, et al. pANCA and ASCA in children with IBD-unclassified, Crohn's colitis, and ulcerative colitis—a longitudinal report from the IBD Porto Group of ESPGHAN. *Inflammatory bowel diseases*. 2016;22(8):1908-1914.
105. Spencer EA, Davis SM, Mack DR, et al. Serologic reactivity reflects clinical expression of ulcerative colitis in children. *Inflammatory bowel diseases*. 2018;24(6):1335-1343.
106. Ruemmele FM, Veres G, Kolho KL, et al. Consensus guidelines of ECCO/ESPGHAN on the medical management of pediatric Crohn's disease. *Journal of Crohn's & colitis*. Oct 2014;8(10):1179-207. doi:10.1016/j.crohns.2014.04.005
107. Turner D, Ruemmele FM, Orlanski-Meyer E, et al. Management of paediatric ulcerative colitis, part 1: ambulatory care—an evidence-based guideline from European Crohn's and Colitis Organization and European Society of Paediatric Gastroenterology, Hepatology and Nutrition. *Journal of pediatric gastroenterology and nutrition*. 2018;67(2):257-291.
108. Saag K, Furst D. Major side effects of systemic glucocorticoids. In: Matteson E, ed. *UpToDate*. 2020.
109. Agency EM. Remicade. Accessed June 15, 2020, 2020. https://www.ema.europa.eu/en/medicines/human/EPAR/remicade
110. Nishida A, Inoue R, Inatomi O, Bamba S, Naito Y, Andoh A. Gut microbiota in the pathogenesis of inflammatory bowel disease. *Clinical journal of gastroenterology*. 2018;11(1):1-10.
111. Sivakumar K, Scarascia G, Zaouri N, Wang T, Kaksonen AH, Hong P-Y. Salinity-mediated increment in sulfate reduction, biofilm formation, and quorum sensing: a potential connection between quorum sensing and sulfate reduction? *Frontiers in microbiology*. 2019;10:188.

112. He Z, Kong X, Shao T, Zhang Y, Wen C. Alterations of the Gut Microbiome Associated with Promoting Efficacy of Prednisone by Bromofuranone in MRL/lpr Mice. *Frontiers in microbiology*. 2019;10:978.
113. Naser SA, Ghobrial G, Romero C, Valentine JF. Culture of Mycobacterium avium subspecies paratuberculosis from the blood of patients with Crohn's disease. *The Lancet*. 2004;364(9439):1039-1044.
114. Sunkara T, Rawla P, Ofosu A, Gaduputi V. Fecal microbiota transplant–a new frontier in inflammatory bowel disease. *Journal of inflammation research*. 2018;11:321.
115. Danese S. Mechanisms of action of infliximab in inflammatory bowel disease: an anti-inflammatory multitasker. *Digestive and Liver disease*. 2008;40:S225-S228.
116. Pasparakis M. Regulation of tissue homeostasis by NF-κB signalling: implications for inflammatory diseases. *Nature Reviews Immunology*. 2009;9(11):778-788.
117. Corridoni D, Pastorelli L, Mattioli B, et al. Probiotic bacteria regulate intestinal epithelial permeability in experimental ileitis by a TNF-dependent mechanism. *PloS one*. 2012;7(7)
118. Nobile S, Gionchetti P, Rizzello F, Calabrese C, Campieri M. Mucosal healing in pediatric Crohn's disease after anti-TNF therapy: a long-term experience at a single center. *European journal of gastroenterology & hepatology*. 2014;26(4):458-465.
119. Hyams J, Damaraju L, Blank M, et al. Induction and maintenance therapy with infliximab for children with moderate to severe ulcerative colitis. *Clinical Gastroenterology and hepatology*. 2012;10(4):391-399. e1.
120. Lewis James D, Chen Eric Z, Baldassano Robert N, et al. Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. *Cell Host & Microbe*. 2015/10/14/ 2015;18(4):489-500. doi:https://doi.org/10.1016/j.chom.2015.09.008
121. Kowalska-Duplaga K, Kapusta P, Gosiewski T, et al. Changes in the Intestinal Microbiota Are Seen Following Treatment with Infliximab in Children with Crohn's Disease. *Journal of clinical medicine*. 2020;9(3):687.
122. Amiot A, Peyrin-Biroulet L. Current, new and future biological agents on the horizon for the treatment of inflammatory bowel diseases. *Therapeutic advances in gastroenterology*. 2015;8(2):66-82.
123. Lawley M, Wu JW, Navas-López VM, et al. Global variation in use of enteral nutrition for pediatric Crohn disease. *Journal of pediatric gastroenterology and nutrition*. 2018;67(2):e22-e29.
124. Ashton JJ, Gavin J, Beattie RM. Exclusive enteral nutrition in Crohn's disease: Evidence and practicalities. *Clinical nutrition*. 2019;38(1):80-89.
125. Kaakoush NO, Day AS, Leach ST, Lemberg DA, Nielsen S, Mitchell HM. Effect of exclusive enteral nutrition on the microbiota of children with newly diagnosed Crohn's disease. *Clinical and translational gastroenterology*. 2015;6(1):e71.
126. Leach S, Mitchell H, Eng W, Zhang L, Day A. Sustained modulation of intestinal bacteria by exclusive enteral nutrition used to treat children with Crohn's disease. *Alimentary pharmacology & therapeutics*. 2008;28(6):724-733.
127. Quince C, Ijaz UZ, Loman N, et al. Extensive Modulation of the Fecal Metagenome in Children With Crohn's Disease During Exclusive Enteral Nutrition. *The American journal of gastroenterology*. 2015;110(12):1718-1730. doi:10.1038/ajg.2015.357
128. Levine A, Wine E, Assa A, et al. Crohn's disease exclusion diet plus partial enteral nutrition induces sustained remission in a randomized controlled trial. *Gastroenterology*. 2019;157(2):440-450. e8.
129. Critch J, Day AS, Otley A, et al. Use of enteral nutrition for the control of intestinal inflammation in pediatric Crohn disease. *Journal of pediatric gastroenterology and nutrition*. 2012;54(2):298-305.
130. Griffin NW, Ahern PP, Cheng J, et al. Prior dietary practices and connections to a human gut microbial metacommunity alter responses to diet interventions. *Cell host & microbe*. 2017;21(1):84-96.

131. Hyams JS, Ferry GD, Mandel FS, et al. Development and validation of a pediatric Crohn's disease activity index. *Journal of pediatric gastroenterology and nutrition*. 1991;12(4):439-447.
132. Levine A, Koletzko S, Turner D, et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. *Journal of pediatric gastroenterology and nutrition*. 2014;58(6):795-806.
133. Turner D, Otley AR, Mack D, et al. Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study. *Gastroenterology*. 2007;133(2):423-432.
134. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*. 2019/08/01 2019;37(8):852-857. doi:10.1038/s41587-019-0209-9
135. Amir A, McDonald D, Navas-Molina JA, et al. Deblur rapidly resolves single-nucleotide community sequence patterns. *MSystems*. 2017;2(2)
136. McDonald D, Price MN, Goodrich J, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal*. 2012;6(3):610.
137. Werner JJ, Koren O, Hugenholtz P, et al. Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. *The ISME journal*. 2012;6(1):94-103.
138. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*. 2010;26(2):266-267.
139. Fitzgerald CB, Shkoporov AN, Sutton TD, et al. Comparative analysis of Faecalibacterium prausnitzii genomes shows a high level of genome plasticity and warrants separation into new species-level taxa. *BMC genomics*. 2018;19(1):931.
140. Rinttilä T, Kassinen A, Malinen E, Krogius L, Palva A. Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *Journal of applied microbiology*. 2004;97(6):1166-1177.
141. Collado MC, Derrien M, Isolauri E, de Vos WM, Salminen S. Intestinal integrity and Akkermansia muciniphila, a mucin-degrading member of the intestinal microbiota present in infants, adults, and the elderly. *Appl Environ Microbiol*. 2007;73(23):7767-7770.
142. Lee D-H, Zo Y-G, Kim S-J. Nonradioactive method to study genetic profiles of natural bacterial communities by PCR-single-strand-conformation polymorphism. *Appl Environ Microbiol*. 1996;62(9):3112-3120.
143. Bouatra S, Aziat F, Mandal R, et al. The Human Urine Metabolome. *PLOS ONE*. 2013;8(9):e73076. doi:10.1371/journal.pone.0073076
144. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-81.
145. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *Journal of biomedical informatics*. 2019;95:103208.
146. Hammer Ø, Harper DA, Ryan PD. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia electronica*. 2001;4(1):9.
147. Team RC. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. 2013;doi:http://www.R-project.org/
148. Pang Z, Chong J, Li S, Xia J. MetaboAnalystR 3.0: Toward an Optimized Workflow for Global Metabolomics. *Metabolites*. May 7 2020;10(5)doi:10.3390/metabo10050186
149. *Stata Statistical Software: Release 15/*. StataCorp LP.; 2015.
150. Shatenstein B, Amre D, Jabbour M, Feguery H. Examining the relative validity of an adult food frequency questionnaire in children and adolescents. *Journal of pediatric gastroenterology and nutrition*. 2010;51(5):645-652.

151. Previdelli Á N, de Andrade SC, Fisberg RM, Marchioni DM. Using Two Different Approaches to Assess Dietary Patterns: Hypothesis-Driven and Data-Driven Analysis. *Nutrients*. Oct 2016;8(10)doi:10.3390/nu8100593
152. Reedy J, Wirfält E, Flood A, et al. Comparing 3 dietary pattern methods--cluster analysis, factor analysis, and index analysis--With colorectal cancer risk: The NIH-AARP Diet and Health Study. *American journal of epidemiology*. 2010;171(4):479-487. doi:10.1093/aje/kwp393
153. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. *The American journal of clinical nutrition*. 1997;65(4):1220S-1228S.
154. Northstone K, Ness A, Emmett P, Rogers I. Adjusting for energy intake in dietary pattern investigations using principal components analysis. *European journal of clinical nutrition*. 2008;62(7):931-938.
155. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*. 2012;41(D1):D590-D596.
156. Yilmaz P, Parfrey LW, Yarza P, et al. The SILVA and "all-species living tree project (LTP)" taxonomic frameworks. *Nucleic acids research*. 2014;42(D1):D643-D648.
157. Pollock J, Glendinning L, Wisedchanwet T, Watson M. The madness of microbiome: attempting to find consensus "best practice" for 16S microbiome studies. *Appl Environ Microbiol*. 2018;84(7):e02627-17.
158. Heyes MP, Saito K, Markey SP. Human macrophages convert L-tryptophan into the neurotoxin quinolinic acid. *Biochemical Journal*. 1992;283(3):633-635.
159. Nikolaus S, Schulte B, Al-Massad N, et al. Increased tryptophan metabolism is associated with activity of inflammatory bowel diseases. *Gastroenterology*. 2017;153(6):1504-1516. e2.
160. Moffett JR, Namboodiri MA. Tryptophan and the immune response. *Immunology and cell biology*. 2003;81(4):247-265.
161. Savitz J. The kynurenine pathway: a finger in every pie. *Molecular psychiatry*. 2019:1-17.
162. Beloborodova N, Bairamov I, Olenin A, Shubina V, Teplova V, Fedotcheva N. Effect of phenolic acids of microbial origin on production of reactive oxygen species in mitochondria and neutrophils. *Journal of biomedical science*. 2012;19(1):89.
163. Wang W, Chen L, Zhou R, et al. Increased proportions of Bifidobacterium and the Lactobacillus group and loss of butyrate-producing bacteria in inflammatory bowel disease. *Journal of clinical microbiology*. 2014;52(2):398-406.
164. Beloborodova N, Khodakova A, Bairamov I, Olenin AY. Microbial origin of phenylcarboxylic acids in the human body. *Biochemistry (Moscow)*. 2009;74(12):1350-1355.
165. Daniel AI, Kvissberg MEA, Senga E, et al. Urinary Organic Acids Increase After Clinical Stabilization of Hospitalized Children With Severe Acute Malnutrition. *Food and nutrition bulletin*. 2019;40(4):532-543.
166. Pettersen JE, Jellum E. The identification and metabolic origin of 2-furoylglycine and 2, 5-furandicarboxylic acid in human urine. *Clinica chimica acta*. 1972;41:199-207.
167. Koopman F, Wierckx N, de Winde JH, Ruijssenaars HJ. Identification and characterization of the furfural and 5-(hydroxymethyl) furfural degradation pathways of Cupriavidus basilensis HMF14. *Proceedings of the National Academy of Sciences*. 2010;107(11):4919-4924.
168. Rechner AR, Smith MA, Kuhnle G, et al. Colonic metabolism of dietary polyphenols: influence of structure on microbial fermentation products. *Free Radical Biology and Medicine*. 2004;36(2):212-225.
169. Wishart D, Tzur D, Knox C. HMDB: the Human Metabolome Database. Nucleic Acids Res. *Database*. 2007;(D521-6)
170. Kelley RI, Kratz L. 3-methylglutaconic acidemia in Smith-Lemli-Opitz syndrome. *Pediatric research*. 1995;37(5):671-674.

171. Finegold SM, Molitoris D, Song Y, et al. Gastrointestinal microflora studies in late-onset autism. *Clinical Infectious Diseases*. 2002;35(Supplement_1):S6-S16.
172. Idris EE, Iglesias DJ, Talon M, Borriss R. Tryptophan-dependent production of indole-3-acetic acid (IAA) affects level of plant growth promotion by Bacillus amyloliquefaciens FZB42. *Molecular plant-microbe interactions*. 2007;20(6):619-626.
173. Nutaratat P, Monprasit A, Srisuk N. High-yield production of indole-3-acetic acid by Enterobacter sp. DMKU-RP206, a rice phyllosphere bacterium that possesses plant growth-promoting traits. *3 Biotech*. 2017;7(5):305.
174. Buell CR, Joardar V, Lindeberg M, et al. The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000. *Proceedings of the National Academy of Sciences*. 2003;100(18):10181-10186.
175. Kochar M, Upadhyay A, Srivastava S. Indole-3-acetic acid biosynthesis in the biocontrol strain Pseudomonas fluorescens Psd and plant growth regulation by hormone overexpression. *Research in microbiology*. 2011;162(4):426-435.
176. Patten CL, Glick BR. Role of Pseudomonas putida indoleacetic acid in development of the host plant root system. *Appl Environ Microbiol*. 2002;68(8):3795-3801.
177. Weissbach H, King W, Sjoerdsma A, Udenfriend S. Formation of indole-3-acetic acid and trytamine in animals. A method for estimation of indole-3-acetic acid in tissues. *Journal of Biological Chemistry*. 1959;234:81-86.
178. Grosso S, Mostardini R, Farnetani MA, et al. Ethylmalonic encephalopathy. *Journal of neurology*. 2002;249(10):1446-1450.
179. Leipnitz G, Schuck PF, Ribeiro CA, et al. Ethylmalonic acid inhibits mitochondrial creatine kinase activity from cerebral cortex of young rats in vitro. *Neurochemical research*. 2003;28(5):771-777.
180. Xiong X, Liu D, Wang Y, Zeng T, Peng Y. Urinary 3-(3-hydroxyphenyl)-3-hydroxypropionic acid, 3-hydroxyphenylacetic acid, and 3-hydroxyhippuric acid are elevated in children with autism spectrum disorders. *BioMed research international*. 2016;2016
181. Martín M, Gibello A, Fernández J, Ferrer E, Garrido-Pertierra A. Catabolism of 3-and 4-hydroxyphenylacetic acid by Klebsiella pneumoniae. *Microbiology*. 1991;137(3):621-628.
182. Del Rio D, Costa L, Lean M, Crozier A. Polyphenols and health: what compounds are involved? *Nutrition, Metabolism and Cardiovascular Diseases*. 2010;20(1):1-6.
183. Knight J, Jiang J, Assimos D, Holmes R. Hydroxyproline ingestion and urinary oxalate and glycolate excretion. *Kidney international*. 2006;70(11):1929-1934.
184. Holmes RP, Goodman HO, Hart LJ, Assimos DG. Relationship of protein intake to urinary oxalate and glycolate excretion. *Kidney international*. 1993;44(2):366-372.
185. Becker J, Lange A, Fabarius J, Wittmann C. Top value platform chemicals: bio-based production of organic acids. *Current opinion in biotechnology*. 2015;36:168-175.
186. Poore RE, Hurst CH, Assimos DG, Holmes RP. Pathways of hepatic oxalate synthesis and their regulation. *American Journal of Physiology-Cell Physiology*. 1997;272(1):C289-C294.
187. Tabasco R, Sánchez-Patán F, Monagas M, et al. Effect of grape polyphenols on lactic acid bacteria and bifidobacteria growth: resistance and metabolism. *Food microbiology*. 2011;28(7):1345-1352.
188. Shaw W. Increased urinary excretion of a 3-(3-hydroxyphenyl)-3-hydroxypropionic acid (HPHPA), an abnormal phenylalanine metabolite of Clostridia spp. in the gastrointestinal tract, in urine samples from patients with autism and schizophrenia. *Nutritional neuroscience*. 2010;13(3):135-143.
189. Keşli R, Gökçen C, Buluğ U, Terzi Y. Investigation of the relation between anaerobic bacteria genus clostridium and late-onset autism etiology in children. *Journal of Immunoassay and Immunochemistry*. 2014;35(1):101-109.

190. Kieffer DA, Piccolo BD, Vaziri ND, et al. Resistant starch alters gut microbiome and metabolomic profiles concurrent with amelioration of chronic kidney disease in rats. *American Journal of Physiology-Renal Physiology*. 2016;310(9):F857-F871.
191. Koh A, De Vadder F, Kovatcheva-Datchary P, Bäckhed F. From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell*. 2016;165(6):1332-1345.
192. Kopečný J, Zorec M, Mrazek J, Kobayashi Y, Marinšek-Logar R. Butyrivibrio hungatei sp. nov. and Pseudobutyrivibrio xylanivorans sp. nov., butyrate-producing bacteria from the rumen. *International Journal of Systematic and Evolutionary Microbiology*. 2003;53(1):201-209.
193. Liu C, Finegold SM, Song Y, Lawson PA. Reclassification of Clostridium coccoides, Ruminococcus hansenii, Ruminococcus hydrogenotrophicus, Ruminococcus luti, Ruminococcus productus and Ruminococcus schinkii as Blautia coccoides gen. nov., comb. nov., Blautia hansenii comb. nov., Blautia hydrogenotrophica comb. nov., Blautia luti comb. nov., Blautia producta comb. nov., Blautia schinkii comb. nov. and description of Blautia wexlerae sp. nov., isolated from human faeces. *International Journal of Systematic and Evolutionary Microbiology*. 2008;58(8):1896-1902.
194. Hall AB, Yassour M, Sauk J, et al. A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome medicine*. 2017;9(1):103.
195. Henke MT, Kenny DJ, Cassilly CD, Vlamakis H, Xavier RJ, Clardy J. Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proceedings of the National Academy of Sciences*. 2019;116(26):12672-12677.
196. Png CW, Lindén SK, Gilshenan KS, et al. Mucolytic Bacteria With Increased Prevalence in IBD Mucosa AugmentIn VitroUtilization of Mucin by Other Bacteria. *American Journal of Gastroenterology*. 2010;105(11):2420-2428.
197. Bian X, Wu W, Yang L, et al. Administration of Akkermansia muciniphila ameliorates dextran sulfate sodium-induced ulcerative colitis in mice. *Frontiers in microbiology*. 2019;10:2259.
198. Zhai R, Xue X, Zhang L, Yang X, Zhao L, Zhang C. Strain-specific anti-inflammatory properties of two Akkermansia muciniphila strains on chronic colitis in mice. *Frontiers in cellular and infection microbiology*. 2019;9:239.
199. Ni J, Wu GD, Albenberg L, Tomov VT. Gut microbiota and IBD: causation or correlation? *Nature reviews Gastroenterology & hepatology*. 2017;14(10):573.
200. Holota Y, Dovbynchuk T, Kaji I, et al. The long-term consequences of antibiotic therapy: Role of colonic short-chain fatty acids (SCFA) system and intestinal barrier integrity. *PloS one*. 2019;14(8)
201. Zarrinpar A, Chaix A, Xu ZZ, et al. Antibiotic-induced microbiome depletion alters metabolic homeostasis by affecting gut signaling and colonic metabolism. *Nature communications*. 2018;9(1):1-13.
202. Zhou L, Zhang M, Wang Y, et al. Faecalibacterium prausnitzii produces butyrate to maintain Th17/Treg balance and to ameliorate colorectal colitis by inhibiting histone deacetylase 1. *Inflammatory bowel diseases*. 2018;24(9):1926-1940.
203. Venegas DP, Marjorie K, Landskron G, et al. Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. *Frontiers in immunology*. 2019;10
204. Wirbel J, Pyl PT, Kartal E, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nature medicine*. 2019;25(4):679-689.
205. Carmody LA, Zhao J, Schloss PD, et al. Changes in cystic fibrosis airway microbiota at pulmonary exacerbation. *Annals of the American Thoracic Society*. 2013;10(3):179-187.
206. Lee MR, Lee S-O, Kim S-Y, Yang SM, Seo Y-H, Cho YK. Brain abscess due to Gemella haemolysans. *Journal of clinical microbiology*. 2004;42(5):2338-2340.
207. Anil M, Ozkalay N, Helvaci M, et al. Meningitis due to Gemella haemolysans in a pediatric case. *Journal of clinical microbiology*. 2007;45(7):2337-2339.

208. Kaufhold A, Lütticken R, Franzen D. Endocarditis caused by Gemella haemolysans. *Infection*. 1989;17(6):385-387.
209. Esposito S, Polinori I, Rigante D. The gut microbiota-host partnership as a potential driver of Kawasaki syndrome. *Frontiers in pediatrics*. 2019;7:124.
210. FitzGerald SF, Moloney AC, Maurer BJ, Hall WW. Gemella endocarditis: consider the colon. *JOURNAL OF HEART VALVE DISEASE*. 2006;15(6):833.
211. Hamilton AL, Kamm MA, Ng SC, Morrison M. Proteus spp. as putative gastrointestinal pathogens. *Clinical microbiology reviews*. 2018;31(3):e00085-17.
212. Clark MS, Russo AF. Tissue-specific glucocorticoid regulation of tryptophan hydroxylase mRNA levels. *Molecular brain research*. 1997;48(2):346-354.
213. Eck R, Simon H. Preparation of both enantiomers of malic and citramalic acid and other hydroxysuccinic acid derivatives by stereospecific hydrations of cis or trans 2-butene-1, 4-dioic acids with resting cells of Clostridium formicoaceticum. *Tetrahedron*. 1994;50(48):13641-13654.
214. Boto-Ordóñez M, Urpi-Sarda M, Queipo-Ortuño MI, Tulipani S, Tinahones FJ, Andres-Lacueva C. High levels of Bifidobacteria are associated with increased levels of anthocyanin microbial metabolites: a randomized clinical trial. *Food & function*. 2014;5(8):1932-1938.
215. Ryan FJ, Ahern A, Fitzgerald R, et al. Colonic microbiota is associated with inflammation and host epigenomic alterations in inflammatory bowel disease. *Nature communications*. 2020;11(1):1-12.
216. GOSSLING J, Moore W. Gemmiger formicilis, n. gen., n. sp., an anaerobic budding bacterium from intestines. *International Journal of Systematic and Evolutionary Microbiology*. 1975;25(2):202-207.
217. Setchell KD, Clerici C. Equol: history, chemistry, and formation. *The Journal of nutrition*. 2010;140(7):1355S-1362S.
218. Nakatsu CH, Armstrong A, Clavijo AP, Martin BR, Barnes S, Weaver CM. Fecal bacterial community changes associated with isoflavone metabolites in postmenopausal women after soy bar consumption. *PLoS One*. 2014;9(10)
219. Maruo T, Sakamoto M, Ito C, Toda T, Benno Y. Adlercreutzia equolifaciens gen. nov., sp. nov., an equol-producing bacterium isolated from human faeces, and emended description of the genus Eggerthella. *International Journal of Systematic and Evolutionary Microbiology*. 2008;58(5):1221-1227.
220. Halfvarson J, Brislawn CJ, Lamendella R, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nature microbiology*. 2017;2(5):1-7.
221. McCracken VJ, Lorenz RG. The gastrointestinal ecosystem: a precarious alliance among epithelium, immunity and microbiota: Microreview. *Cellular microbiology*. 2001;3(1):1-11.
222. Farmer RG, Easley KA, Rankin GB. Clinical patterns, natural history, and progression of ulcerative colitis. *Digestive diseases and sciences*. 1993;38(6):1137-1146.
223. Kohashi M, Nishiumi S, Ooi M, et al. A novel gas chromatography mass spectrometry-based serum diagnostic and assessment approach to ulcerative colitis. *Journal of Crohn's and Colitis*. 2014;8(9):1010-1021.
224. Alderson JW, Van Dinter Jr TG, Opatowsky MJ, Burton EC. Disseminated aspergillosis following infliximab therapy in an immunosuppressed patient with Crohn's disease and chronic hepatitis C: a case study and review of the literature. *Medscape General Medicine*. 2005;7(3):7.
225. Bourne E-L, Dimou J. Invasive central nervous system aspergillosis in a patient with Crohn's disease after treatment with infliximab and corticosteroids. *Journal of Clinical Neuroscience*. 2016;30:163-164.
226. Bennett MR, Ma Q, Ying J, Devarajan P, Brunner H. Effects of age and gender on reference levels of biomarkers comprising the pediatric Renal Activity Index for Lupus Nephritis (p-RAIL). *Pediatric Rheumatology*. 2017;15(1):74.

227. Barr DB, Wilder LC, Caudill SP, Gonzalez AJ, Needham LL, Pirkle JL. Urinary creatinine concentrations in the US population: implications for urinary biologic monitoring measurements. *Environmental health perspectives*. 2005;113(2):192-200.
228. Gu H, Pan Z, Xi B, et al. 1H NMR metabolomics study of age profiling in children. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In vivo*. 2009;22(8):826-833.
229. Shastri P, McCarville J, Kalmokoff M, Brooks SP, Green-Johnson JM. Sex differences in gut fermentation and immune parameters in rats fed an oligofructose-supplemented diet. *Biology of sex differences*. 2015;6(1):13.
230. Jakobsdottir G, Bjerregaard JH, Skovbjerg H, Nyman M. Fasting serum concentration of short-chain fatty acids in subjects with microscopic colitis and celiac disease: no difference compared with controls, but between genders. *Scandinavian journal of gastroenterology*. 2013;48(6):696-701.
231. McOrist AL, Miller RB, Bird AR, et al. Fecal butyrate levels vary widely among individuals but are usually increased by a diet high in resistant starch. *The Journal of nutrition*. 2011;141(5):883-889.
232. Bolnick DI, Snowberg LK, Hirsch PE, et al. Individual diet has sex-dependent effects on vertebrate gut microbiota. *Nature communications*. 2014;5:4500.
233. Mueller S, Saunier K, Hanisch C, et al. Differences in fecal microbiota in different European study populations in relation to age, gender, and country: a cross-sectional study. *Appl Environ Microbiol*. 2006;72(2):1027-1033.
234. Biro FM, Pinney SM, Huang B, Baker ER, Walt Chandler D, Dorn LD. Hormone changes in peripubertal girls. *The Journal of Clinical Endocrinology & Metabolism*. 2014;99(10):3829-3835.
235. August G, Grumbach MM, Kaplan S. Hormonal changes in puberty: III. Correlation of plasma testosterone, LH, FSH, testicular size, and bone age with male pubertal development. *The Journal of Clinical Endocrinology & Metabolism*. 1972;34(2):319-326.
236. Lee GJ, Kappelman MD, Boyle B, et al. Role of sex in the treatment and clinical outcomes of pediatric patients with inflammatory bowel disease. *Journal of pediatric gastroenterology and nutrition*. 2012;55(6):701-706.
237. Grossi V, Lerer T, Griffiths A, et al. Concomitant use of immunomodulators affects the durability of infliximab therapy in children with Crohn's disease. *Clinical Gastroenterology and Hepatology*. 2015;13(10):1748-1756.
238. Shah ED, Coburn ES, Nayyar A, Lee KJ, Koliani-Pace JL, Siegel CA. Systematic review: hepatosplenic T-cell lymphoma on biologic therapy for inflammatory bowel disease, including data from the Food and Drug Administration Adverse Event Reporting System. *Alimentary Pharmacology & Therapeutics*. 2020;51(5):527-533.
239. Ordás I, Feagan BG, Sandborn WJ. Therapeutic drug monitoring of tumor necrosis factor antagonists in inflammatory bowel disease. *Clinical Gastroenterology and Hepatology*. 2012;10(10):1079-1087.
240. Thibault R, De Coppet P, Daly K, et al. Down-regulation of the monocarboxylate transporter 1 is involved in butyrate deficiency during intestinal inflammation. *Gastroenterology*. 2007;133(6):1916-1927.
241. Cao J, Ng M, Felmlee MA. Sex differences in hepatic MCT1 and CD147 expression. *The FASEB Journal*. 2017;31(1_supplement):1062.4-1062.4.
242. Sepponen K, Ruusunen M, Pakkanen J, Pösö A. Expression of CD147 and monocarboxylate transporters MCT1, MCT2 and MCT4 in porcine small intestine and colon. *The Veterinary Journal*. 2007;174(1):122-128.
243. Levine A, Rhodes JM, Lindsay JO, et al. Dietary Guidance for Patients With Inflammatory Bowel Disease from the International Organization for the Study of Inflammatory Bowel Disease. *Clinical Gastroenterology and Hepatology*. 2020;

244. Levine A, Rhodes JM, Lindsay JO, et al. Dietary Guidance From the International Organization for the Study of Inflammatory Bowel Diseases. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*. 2020/05// 2020;18(6):1381-1392. doi:10.1016/j.cgh.2020.01.046
245. de Bie C, Kindermann A, Escher J. Use of exclusive enteral nutrition in paediatric Crohn's disease in The Netherlands. *Journal of Crohn's and Colitis*. 2013;7(4):263-270.
246. Thomas V, Clark J, Doré J. Fecal microbiota analysis: an overview of sample collection methods and sequencing strategies. *Future microbiology*. 2015;10(9):1485-1504.
247. Walsh MC, Brennan L, Malthouse JPG, Roche HM, Gibney MJ. Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. *The American journal of clinical nutrition*. 2006;84(3):531-539.
248. O'Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *The American journal of clinical nutrition*. 2011;93(2):314-321.
249. Chong J, Soufan O, Li C, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic acids research*. Jul 2 2018;46(W1):W486-w494. doi:10.1093/nar/gky310
250. *Stata Statistical Software: Release 15*. StataCorp LP.; 2015.
251. Frolkis A, Knox C, Lim E, et al. SMPDB: The Small Molecule Pathway Database. *Nucleic acids research*. Jan 2010;38(Database issue):D480-7. doi:10.1093/nar/gkp1002
252. Faure M, Moënnoz D, Montigon F, Mettraux C, Breuille D, Ballevre O. Dietary threonine restriction specifically reduces intestinal mucin synthesis in rats. *The Journal of nutrition*. 2005;135(3):486-491.
253. Tang Y, Tan B, Xiong X, et al. Methionine deficiency reduces autophagy and accelerates death in intestinal epithelial cells infected with enterotoxigenic Escherichia coli. *Amino acids*. 2015;47(10):2199-2204.
254. Chen Y, Li D, Dai Z, et al. L-methionine supplementation maintains the integrity and barrier function of the small-intestinal mucosa in post-weaning piglets. *Amino Acids*. 2014;46(4):1131-1142.

Appendix 1: ImageKids

Appendix 1 includes my original contributions to the metabolomics portion of the ImageKids study, described below. These findings and cohort are not directly related to the main theme of my thesis so are included here as an appendix. In the future, this work will likely be part of a larger study I will help to co-author that will include microbiome analyses conducted at another centre.

Methods

ImageKids is a multicentre, prospective cohort observational study that was conducted to develop magnetic resonance enterography (MRE) indices for pCD. The study was conducted over 18 months with serum and urine specimens collected as study initiation (Visit One) and completion (Visit Four) for 56 pCD patients. A diverse population from 15 international centres in Europe, North America, Asia, and Australia were recruited. Patients in remission and flaring, aged eight to 18, and with a disease duration ranging from 0-13 years were assessed. Metabolomics were conducted to assess for correlation with disease activity and MRE findings. Metabolites were also assessed here for potential use as biomarkers or indicators of longitudinal changes in the host-microbe relationship. MRE was used to define three groups based on inflammation and damage: 1 – neither inflammation nor damage, 2 – inflammation only, and 3 – inflammation and damage.

Collected specimens were sent to the Wine lab at the University of Alberta, and further submitted to The Metabolomics Innovation Centre (TMIC) for metabolomics analysis. Urine underwent a gas chromatography mass spectroscopy organic acids assay. Serum underwent liquid chromatography followed by mass spectroscopy in the TMIC Prime Assay. Approximately 80 compounds were identified in urine specimens and 131 were identified in

serum. When the concentration of a metabolite was below the limit of detection (LOD), the value was replaced with the LOD value. Metabolites that had ≥50% of their concentrations below the LOD were excluded from the analysis.

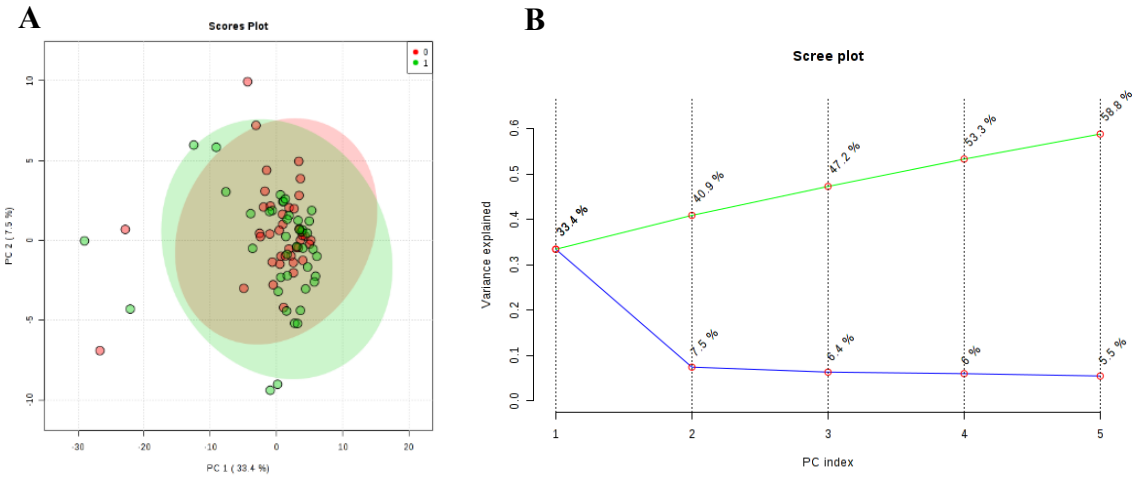
Serum was analyzed using Metaboanalyst.[249] Metabolite concentrations (μMol) were log transformed and Pareto scaled to normalize their distributions. Paired *t*-tests were conducted to compare changes in metabolite concentrations over time. PCA was used to conduct unsupervised analysis, visualize data, and identify potential patterns or clusters of subjects based upon metabolite concentration. Identified clusters were then analyzed using supervised multivariate analysis: partial least squares discriminate analysis (PLS-DA), to identify metabolites most responsible for cluster definition. PLS-DA models were subjected to 10-fold cross validation to rule-out overfitting and model accuracy and the ideal number of components to retain. PLS-DA models were also used to determine if MRE identified inflammation and damage state correlated with metabolite concentrations. Stata 14[250] was used to perform further statistical analysis for differences in clinical features of disease between patient groups. Breusch-Pagan tests were used to assess for homoscedasticity prior to performing simple linear regression, variables identified as heteroscedastic were log transformed and Pareto scaled prior to further analysis. Logistic and ordinal logistic regression were used for binary and ordinal variables such as presence of colonic disease or PGA score. Quantitative enrichment analysis (QEA) was conducted using Metaboanalyst,[249] referencing the Small Molecule Pathway Database.[251]

Results

PCA of urine metabolomics at baseline and after 18 months is shown in **Figure 43A.** As shown in the scree plot in **Figure 43B,** the first two principal components accounted for 33.4%, and 7.5% of metabolite variability, respectively. The PCA plot showed no significant clustering

of metabolites at either timepoint. Supervised analysis was not conducted due to the lack of evident clustering.

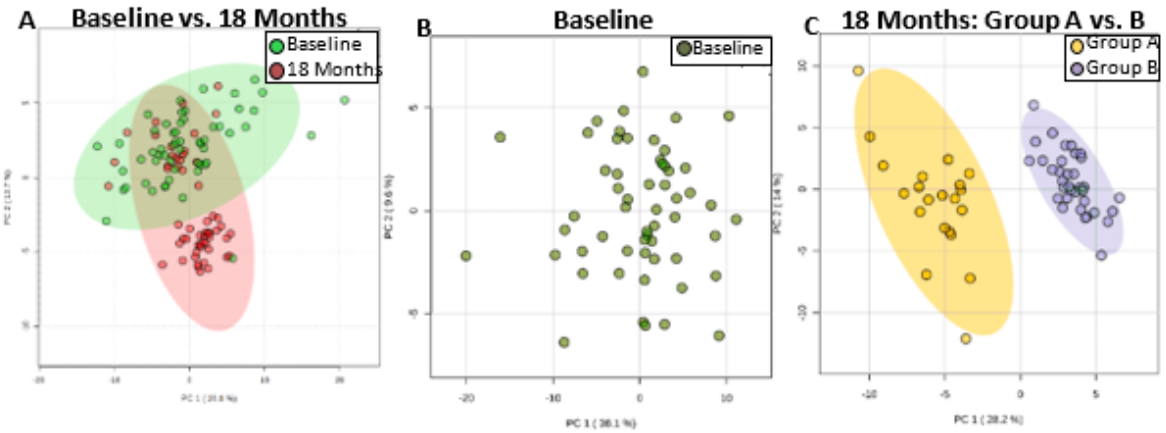
Figure 43 *Urine metabolites do not identify any clusters of patterns in ImageKids pCD cohort.*



Legend: ***A*** *Pareto scaled and log transformed urine metabolites in PCA* ***B*** *Scree plot depicts eigenvalues for the first five principal components.*

PCA of serum metabolites at baseline and after 18 months is shown in **Figure 44A;** concentrations with the first two principal components account for 24.9% and 14.4% of variability, respectively. No significant clustering was evident at baseline (**Figure 44B)**, but at the 18-month time-point two distinct clusters were evident (**Figure 44C**). PCA was repeated for visualization at baseline and 18 months separately. Clustering at 18-months was mainly attributed to variation in principal component one, which accounted for 28.2% of metabolite variation. Principal one scores were used to define the two clusters: Group A (yellow), and group B (purple).

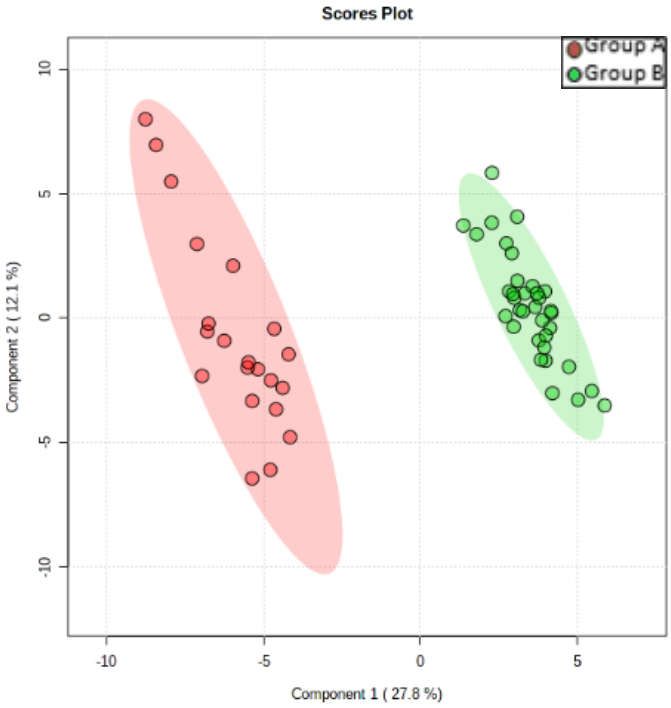
Figure 44 *Serum metabolites exhibit distinct clusters after 18 months in the ImageKids pCD cohort.*



Legend*: **A** PCA of serum metabolites from both timepoints. **B** PCA of baseline metabolites only. **C** PCA of 18 months serum metabolites only with clusters labelled Group A and Group B.*

PLS-DA was conducted on Group A and Group B as defined by PCA cluster, and is visualized in **Figure 45**. Ten-fold cross validation of the model showed $Q^2 = 0.97663$, $R^2 = 0.9925$, and Accuracy = 1.0 with 5 components retained. Components one and two accounted for 27.8% and 12.1% of variability, respectively. Variables important in projection in the PLS-DA model (VIPs) are summarized with VIP scores in **Table 77** with corresponding coefficient scores. Unpaired *t*-tests, with p-values adjusted for FDR, were conducted to determine metabolite concentrations significantly different between groups, resulting in 79 significantly altered compounds, detailed in **Table 78**.

Figure 45 *PLS-DA models metabolites defining Group A (red) and Group B (green) at 18 months.*



Thirty-four serum metabolites were identified as VIPS in distinguishing Group A and Group B in PLS-DA (**Table 77**).

***Table** 77 VIP serum metabolites distinguishing Group A from Group B in PLS-DA.*

	Component Score					Coefficient Score	Biological Significance
Compound	**1**	**2**	**3**	**4**	**5**		
Spermidine	3.65	3.58	3.57	3.56	3.56	100	Polyamine
Spermine	3.47	3.41	3.40	3.39	3.39	95.21	Polyamine
Putrescine	2.94	2.91	2.90	2.90	2.89	80.58	Biogenic Amine
Total dimethylarginine	2.53	2.48	2.47	2.47	2.47	69.36	Biogenic Amine
Threonine	2.46	2.41	2.40	2.39	2.39	67.28	Amino Acid
LYSO PC 20:3	2.05	2.03	2.03	2.02	2.02	56.22	Glycerophospholipid
Sarcosine	2.03	2.00	2.00	2.00	2.00	55.63	Biogenic Amine
Acetyl-carnitine	2.01	1.97	1.96	1.96	1.96	55.05	Acylcarnitine
Hydroxybutryl-carnitine	1.57	1.57	1.56	1.56	1.56	43.1	Acylcarnitine
Choline	1.55	1.53	1.54	1.54	1.53	42.52	
Tetradecanoyl-carnitine	1.49	1.46	1.46	1.46	1.46	40.87	Acylcarnitine
LYSO PC 28:1	1.47	1.46	1.46	1.45	1.45	40.32	Glycerophospholipid
Sphingomyelin 24:1 OH	1.40	1.40	1.39	1.39	1.39	38.41	Shingomyelin derivative
Hydroxyhexadecadienyl-carnitine	1.40	1.39	1.39	1.38	1.38	38.35	Acylcarnitine
Sphingomyelin 22:1 OH	1.40	1.39	1.38	1.38	1.38	38.18	Shingomyelin derivative
Butyric acid	1.39	1.36	1.36	1.37	1.37	37.98	Short Chain Fatty Acid
Hexenoylcarnitine	1.36	1.33	1.33	1.33	1.32	37.12	Acylcarnitine
PC aa 38:0	1.35	1.35	1.35	1.35	1.34	36.87	Glycerophospholipid
Sphingomyelin 22:2 OH	1.22	1.23	1.23	1.22	1.22	33.26	Shingomyelin derivative
PC aa 36:0	1.21	1.20	1.20	1.20	1.20	33.01	Glycerophospholipid
Nonaylcarnitine	1.20	1.18	1.17	1.17	1.17	32.89	Acylcarnitine
Sphingomyelin 20:2	1.16	1.17	1.16	1.16	1.16	31.83	Shingomyelin

							derivative
Asymmetric dimethylarginine	1.14	1.16	1.15	1.15	1.15	31.24	Biogenic Amine
Tiglylcarnitine	1.14	1.12	1.12	1.12	1.12	31.21	Acylcarnitine
PC ae 36:0	1.12	1.13	1.12	1.12	1.12	30.51	Glycerophospholipid
PC aa 40:2	1.09	1.11	1.10	1.10	1.10	29.82	Glycerophospholipid
Octadecenoylcarnitine	1.07	1.08	1.08	1.08	1.08	29.34	Acylcarnitine
Hydroxytetradecenoyl carnitine	1.07	1.05	1.04	1.05	1.05	29.17	Acylcarnitine
181 Sphingomyelin	1.05	1.06	1.06	1.06	1.06	28.82	Shingomyelin derivative
Proline	1.03	1.01	1.01	1.01	1.01	28.17	Amino Acid
PC ae 40:6	1.03	1.05	1.05	1.05	1.05	28.01	Glycerophospholipid
Betaine	1.01	1.00	1.00	1.00	1.00	27.66	Amino Acid
16:1 Sphingomyelin	1.01	1.04	1.03	1.03	1.03	27.54	Shingomyelin derivative
16:0 Sphingomyelin	1.00	1.02	1.02	1.01	1.01	27.25	Shingomyelin derivative

Student's T-tests identified 79 compounds with significantly different concentrations between Group A and Group B at the 18 months timepoint (**Table 78**).

Table 78 *Seventy-nine compounds with significantly different concentrations between Group A and Group B serum metabolites at the 18-month timepoint.*

	Group A Mean	**Group A Upper 95% CI**	**Group A Lower 95% CI**	**Group B Mean**	**Group B Upper 95% CI**	**Group B Lower 95% CI**	**FDR**
LYSO PC 20:3	34.06	40.76	27.36	29.41	34.5	24.32	8.4E-27
Total dimethylarginine	5.86	6.83	4.89	1.18	1.29	1.07	2.1E-26
Threonine	549.05	614.47	483.62	119.73	135.05	104.4	9.9E-24
Spermine	0.03	0.04	0.01	0.33	0.37	0.28	4.1E-22
Hexenoylcarnitine	0.06	0.07	0.05	0.05	0.06	0.05	1.2E-19

Acetylcarnitine	0.04	0.04	0.03	0.04	0.05	0.03	3.3E-13
Spermidine	0.05	0.09	0	0.3	0.33	0.27	6.8E-13
Hydroxyhexadecadienyl-carnitine	0.05	0.05	0.04	0.09	0.1	0.08	9.9E-13
Tetradecanoyl-carnitine	0.03	0.03	0.02	0.01	0.02	0.01	2.7E-12
Hydroxybutryl-carnitine	0.03	0.04	0.02	0.03	0.04	0.02	3.2E-12
Choline	34.22	39.08	29.36	30.15	33.78	26.53	3.8E-12
Putrescine	0.01	0.03	0	0.05	0.06	0.04	8.5E-11
Tiglylcarnitine	0.02	0.02	0.02	0.01	0.01	0.01	1.3E-10
Sphingomyelin 24:1 OH	2.1	2.51	1.69	1.08	1.19	0.98	4.8E-09
Sphingomyelin 22:1 OH	14.46	16.86	12.06	7.93	8.98	6.89	7.2E-09
LYSO PC 28:1	0.27	0.33	0.2	0.22	0.24	0.2	1.6E-08
PC aa 38:0	2.17	2.63	1.71	1.08	1.2	0.96	8.1E-08
Octadecenoylarnitine	0.14	0.16	0.12	0.21	0.23	0.2	1.6E-07
nonaylcarnitine	0.03	0.03	0.02	0.01	0.02	0.01	2.3E-07
Hydroxytetradecenoyl-carnitine	0.01	0.01	0.01	0.01	0.01	0.01	2.9E-07
PC aa 36:0	5.07	5.93	4.22	3.08	3.54	2.63	4.5E-07
Sphingomyelin 22:2 OH	11	12.94	9.06	6.51	7.46	5.55	5.4E-07
PC ae 36:0	1.03	1.2	0.87	0.68	0.76	0.59	1.0E-06
PC aa 40:2	0.32	0.37	0.26	0.21	0.23	0.18	1.3E-06
Valerylcarnitine	0.14	0.16	0.12	0.19	0.2	0.18	1.7E-06
Sphingomyelin 20:2	1.33	1.52	1.14	0.88	1.02	0.73	2.1E-06
16:0 Sphingomyelin	133.91	152.75	115.07	92.21	101.33	83.08	4.3E-06
Glutarylcarnitine	589.05	664.59	513.51	447.95	478.43	417.47	4.4E-06
18:1 Sphingomyelin	12.34	14.18	10.5	8.1	9.16	7.05	6.6E-06
Hexadecadienyl-carnitine	363.14	445.81	280.48	293.24	336.41	250.07	9.6E-06
Proline	290.95	336.41	245.49	178.57	196.31	160.84	1.0E-05
16:1 Sphingomyelin	17.01	19.37	14.65	11.43	12.83	10.04	1.2E-05
Asymmetric dimethylarginine	22.51	31.22	13.81	23.94	28.48	19.39	1.2E-05
PC ae 40:6	2.87	3.42	2.32	1.83	2.03	1.63	2.1E-05
Hydroxyhexadecenoyl-carnitine	0.01	0.01	0.01	0.01	0.01	0.01	2.4E-05
Octadecanoyl-carnitine	0.04	0.04	0.03	0.06	0.06	0.05	2.9E-05
Sarcosine	0.49	0.7	0.29	0.93	1.08	0.78	3.2E-05
Sphingomyelin 18:0	28.87	33.08	24.65	20.17	22.58	17.76	3.2E-05
Glucose	0.85	0.97	0.73	0.95	1.07	0.83	3.3E-05

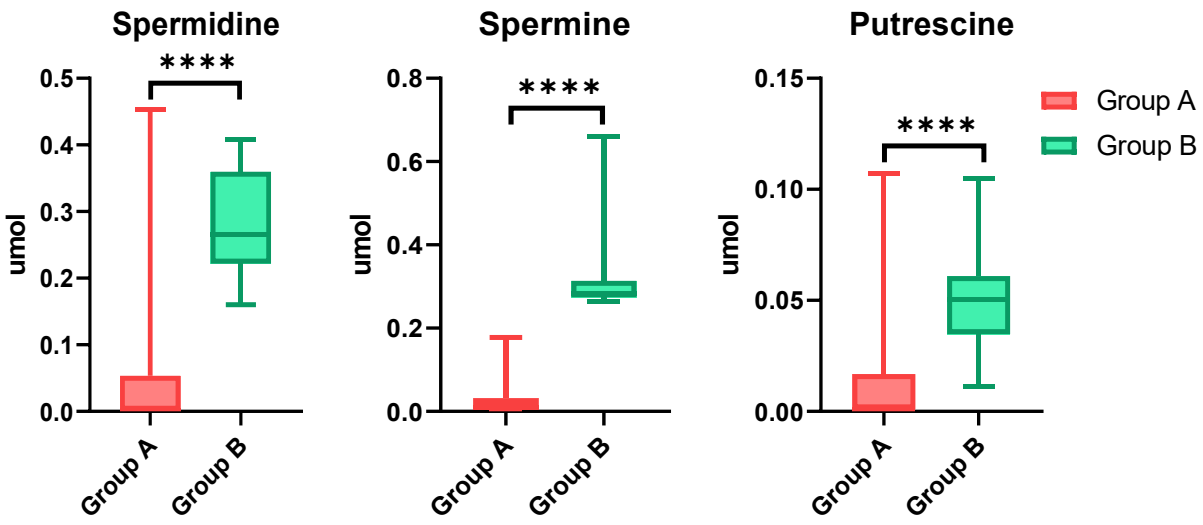
Leucine	1753.81	1950.32	1557.3	1812.38	2019.95	1604.81	6.0E-05
Hexadecenoyl-carnitine	0.09	0.1	0.08	0.12	0.13	0.11	1.0E-04
Betaine	56.37	87.62	25.11	123.3	214.17	32.44	1.6E-04
Glutamine	52.22	65.25	39.19	61.9	75.05	48.74	1.8E-04
Butyric acid	0.01	0.01	0.01	0.01	0.01	0.01	2.1E-04
PC aa 38:6	43.11	51.17	35.06	27.27	32.64	21.9	2.6E-04
Valine	202.11	225.49	178.74	140.4	152.5	128.3	3.2E-04
PC aa 40:6	14.15	17.26	11.04	8.62	10.07	7.17	4.0E-04
Octadecadienyl-carnitine	0.07	0.08	0.06	0.09	0.09	0.08	5.7E-04
Decadienylcarnitine	63.37	75.15	51.59	67.6	74.26	60.93	5.8E-04
Asparagine	134.93	156.91	112.96	96.07	107.44	84.7	7.1E-04
Tetradecadienylcarnitine	0.05	0.06	0.03	0.07	0.08	0.05	8.3E-04
Ornithine	82.31	91.01	73.62	63.4	71.2	55.59	1.2E-03
Sphingomyelin 16:1OH	3.63	4.3	2.95	2.67	3	2.34	1.3E-03
LYSO PC 26:0	0.15	0.18	0.12	0.14	0.15	0.14	1.4E-03
Hexadecanoyl-carnitine	0.01	0.01	0.01	0.01	0.01	0.01	1.5E-03
PC aa 40:1	0.24	0.28	0.2	0.17	0.19	0.16	1.5E-03
Tyrosine	69.53	79.48	59.59	51.08	57.05	45.1	1.8E-03
Alanine	6.63	7.8	5.45	1.97	2.28	1.65	2.2E-03
Arginine	6.14	6.77	5.52	6.29	7.13	5.45	2.4E-03
LYSO PC 26:1	0.31	0.39	0.22	0.19	0.22	0.16	2.7E-03
Glutaconyl-carnitine	5237.86	5979.41	4496.3	3258.57	3558.29	2958.85	3.6E-03
Octanoylcarnitine	0.1	0.13	0.08	0.06	0.07	0.06	3.6E-03
Serine	152.95	176.06	129.85	117.1	125.57	108.63	7.2E-03
Methylglutaryl-carnitine	24.69	29.78	19.6	20.74	23.02	18.45	7.5E-03
Benzoic acid	0.31	0.38	0.24	0.47	0.53	0.42	7.6E-03
Decenoylcarnitine	0.25	0.31	0.19	0.24	0.31	0.16	7.9E-03
Tryptophan	58.41	67.49	49.32	42	48.96	35.04	8.0E-03
Pyruvic acid	60.69	69.19	52.18	47.17	59.11	35.24	8.5E-03
Dodecanoylcarnitine	0.01	0.01	0.01	0.01	0.01	0.01	1.1E-02
Isoleucine	3.08	3.91	2.25	2.84	3.78	1.89	1.3E-02
Dodecanedioyl-carnitine	0.3	0.35	0.26	0.22	0.25	0.19	1.8E-02
Citrulline	87.43	95.39	79.47	95.94	106.6	85.28	3.0E-02
Trans-hydroxyproline	26.69	33.92	19.46	17.65	21.68	13.61	3.3E-02
Histidine	1.73	3.01	0.46	1.61	2.34	0.88	3.4E-02
Butyrylcarnitine	2.12	2.29	1.95	9.84	11.99	7.69	3.5E-02
Glutamic acid	0.01	0.02	0.01	0.02	0.02	0.02	3.8E-02
Propionylcarnitine	0.34	0.38	0.29	0.38	0.41	0.35	3.8E-02
Trimethylamine N-oxide	1.13	1.5	0.77	2.09	3.32	0.87	4.0E-02

Lysine	179.1	200.23	157.96	111.42	123.72	99.11	4.8E-02

***Legend**: Metabolite concentrations were log transformed and pareto scaled prior to conducting unpaired students T-tests. Only significant different metabolites are shown here. All metabolite concentrations are in µMol.*

Spermidine, spermine, and putrescine were identified as the top three compounds important in distinguishing the groups, and were also identified as significantly increased in Group B (**Table 78** and **Figure 46**)

***Figure 46** Spermidine, spermine, and putrescine are significantly higher in Group B serum at Visit 4.*



***Legend**: Student t-tests of log transformed and pareto scaled serum metabolite concentrations comparing Group A and Group B. ****=p<0.0001*

Regression identified several clinical characteristics associated with the PC1 score that defined Group A vs Group B membership. After adjusting for sex and previous history of bowel resection, the following independent predictors were identified that accounted for 35% of the variation between metabolite defined groups, with no significant interaction effects:

Patients in metabolite group A at Visit 4 were more likely have been on Antimicrobials at Visit 1 (p=0.004), as well as 6-mercaptopurine (either visit 1 or 4 – no change in this medication between timepoints, p=0.01). These patients also had a higher Visit 1 CRP (37.0 vs 12.7, p=0.009).

Although a significant interaction between PC1 and 5ASA use at Visit 4 was originally identified, after adjusting for antimicrobial and 6-mercaptopurine use it was no longer significant, and inclusion of 5ASA use at Visit 4 showed worsening predictable ability of the statistical model, from p=0.0002 to p=0.0001. Given these results 5ASA use at Visit 4 was not included as a significant predictor of PC1 score.

Adjusting the model for BMI Z-score showed worsening model predictive power, as well as no significant interactions, so was excluded. There was no significant different in BMI Z-score between Group A and Group B at either Visit 1 or 4 (p>0.4 for all). After adjusting for sex and history of resection, baseline CRP and 6-mercaptopurine use, anti-TNF and 5-ASA therapy were not significantly correlated with metabolite group. Additionally, there were no associations with disease location (colonic vs. ileal), or complicated vs. inflammatory disease. Unadjusted group comparisons are shown in **Table 79**.

Table 79 *Demographic and clinical characteristics of groups A & B.*

	Group A	**Group B**	**Unadjusted p-value**
Sex (male)	13/21	20/35	0.859
Age at Visit 1 Years (95%CI)	14.8 (13.8, 15.8)	14.0 (13.0, 15.1)	0.1432
Years disease duration	2.96 (1.72, 4.20)	2.78 (1.78, 3.77)	0.8078
wPCDAI Visit 1 Visit 4	 26.3 (15.5, 37.2) 5.5 (3.2, 7.8)	 34.0 (23.6, 44.5) 8.0 (4.2, 11.8)	 0.2778 0.2146
Fecal Calprotectin			

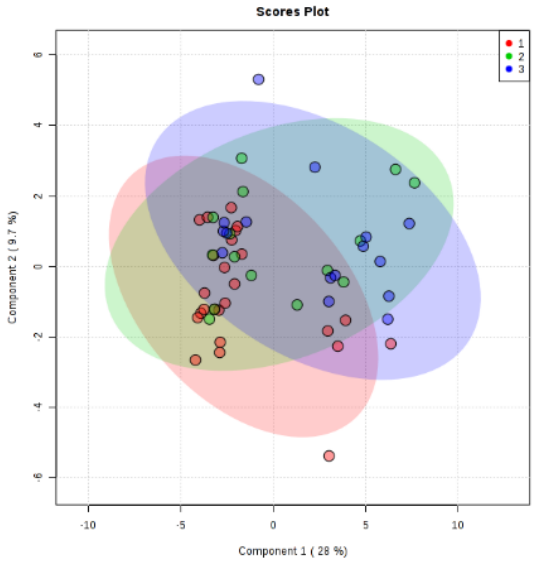
Visit 1	858.2 (551.9, 1164.6)	899.1 (408.9, 1389.2)	0.8722
Visit 4	537.1 (220.7, 853.6)	621.5 (372.7, 870.2)	0.6944
C-reactive Protein			
Visit 1	37.0 (13.6, 60.4)	12.7 (3.0, 22.4)	**0.057**
Visit 4	6.34 (3.7, 9.0)	8.6 (0, 18.5)	0.59
Bowel Resection			
Prior to Visit 1	1/21	4/35	0.435
Prior to Visit 4	3/21	8/35	0.518
Colonic Disease			
Visit 1	18/21	27/35	0.798
Visit 4	15/18*	22/29*	0.834
Complicated Disease			
Visit 1	5/21	7/35	0.788
Visit 4	5/18*	6/29*	0.662
Progression to Complicated Disease	2/15	1/26	0.302
Medications			
Steroids			
Visit 1	1/21	6/35	0.224
Visit 4	1/21	5/35	0.297
Steroids Initiated V1→V4?	3/20	8/29	0.404
Methotrexate (Oral or SC)	2/21	7/35	0.374
Visit 1	2/21	11/34*	0.118
Visit 4			
Azathioprine			
Visit 1	5/21	5/21	0.949
Visit 4	5/21	8/34*	0.985
Azathioprine initiated V1→V4?	7/16	9/27	0.647
6-mercaptopurine			
Visit 1	2/21	0/35	0.076
Visit 4	2/21	0/35	0.076
Anti-TNF			
Visit 1	4/21	17/35	0.124
Visit 4	13/21	24/35	0.817
Anti-TNF initiated V1→V4?	9/17	8/18	0.768
5-ASA			
Visit 1	1/21	0/35	0.203

<table>
<tr><td>Visit 4</td><td>3/21</td><td>0/35</td><td>0.032</td></tr>
<tr><td>Antimicrobials
Visit 1</td><td>4/21</td><td>0/35</td><td>0.014</td></tr>
<tr><td>PGA
Visit 1</td><td><table><tr><th>PGA Baseline</th><th># of patients</th></tr><tr><td>0</td><td>3</td></tr><tr><td>1</td><td>5</td></tr><tr><td>2</td><td>5</td></tr><tr><td>3</td><td>6</td></tr><tr><td>4</td><td>2</td></tr></table></td><td><table><tr><th>PGA Baseline</th><th># of patients</th></tr><tr><td>0</td><td>4</td></tr><tr><td>1</td><td>5</td></tr><tr><td>2</td><td>13</td></tr><tr><td>3</td><td>9</td></tr><tr><td>4</td><td>4</td></tr></table></td><td>0.6435</td></tr>
<tr><td>Visit 4</td><td><table><tr><th>PGA 18 months</th><th># of patients</th></tr><tr><td>0</td><td>5</td></tr><tr><td>1</td><td>10</td></tr><tr><td>2</td><td>4</td></tr><tr><td>3</td><td>1</td></tr><tr><td>4</td><td>0</td></tr></table></td><td><table><tr><th>PGA 18 months</th><th># of patients</th></tr><tr><td>0</td><td>11</td></tr><tr><td>1</td><td>8</td></tr><tr><td>2</td><td>13</td></tr><tr><td>3</td><td>2</td></tr><tr><td>4</td><td>0</td></tr></table></td><td>0.5417</td></tr>
</table>

***Legend**: Unpaired student t-tests were used for continuous variables, chi2 test was used for binary variables. *=medication data was missing for one patient for several medications at Visit 4. Initiation of medication between Visit 1 and Visit 4, termed "initiated V1→V4?" was also assessed.*

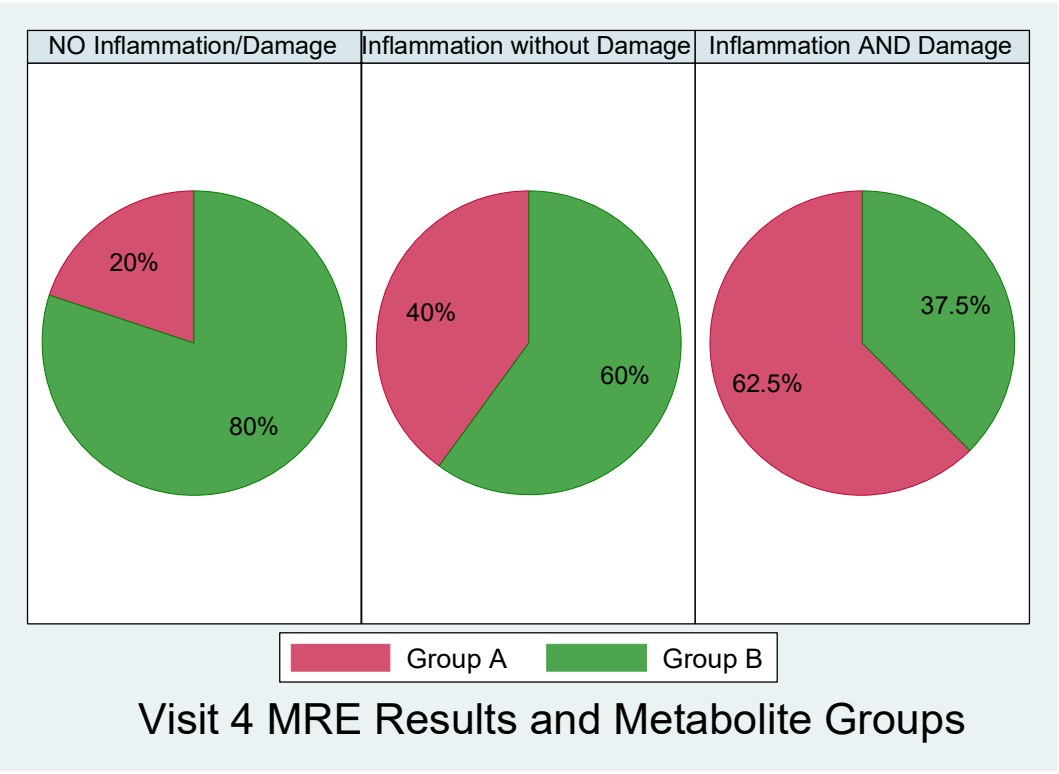
MRE Visit 1 and Visit 4 inflammation and/or damage status did not directly correlate with metabolite profile, as PLS-DA was unable to generate a predictive model at either time-point (**Figure 47**). Additionally, visit 1 serum metabolite profile was not predictive of visit 4 MRE findings on PERMANOVA analysis.

Figure 47 *PLS-DA showed non-significant separation between MRE and metabolite identified groups 1-3. Cross-validation of the model resulted in negative Q^2 values (poor fit).*



Visit 4 metabolite group status (A vs. B), did however, correlate with MRE inflammation/damage status at Visit 4 (**Figure 48**) as assessed by MRE findings correlated with proportion of group A vs. B membership ANOVA p-value = 0.021 (**Figure 48**). Tukey's post-hoc comparison revealed patients with no MRE identified inflammation/damage were more likely to belong to metabolite group B compared to those identified as having both inflammation and damage (p=0.016).

Figure 48 *Group A has more patients with inflammation and damage identified on MRE at Visit 4.*



Legend: *Patients identified as having inflammation AND damage on MRE at Visit 4 were significantly more likely to belong to metabolite Group A, p= 0.016. There was a non-statistically significant increase between no inflammation/damage to inflammation without damage (p=0.389), and from inflammation without damage to both inflammation and damage (p=0.373).*

Detailed Metabolite Changes:

Quantitative Enrichment Analysis was done (**Figure 49** and

Table 80) to identify variations in metabolite pathways between Groups A and B using metabolite concentrations with group labelling. Thirty-four phospholipids and 16 acylcarnitines could not be traced to single compounds within the Human Metabolome Database (HMDB.ca), that is utilized by the SPMDB, so were excluded from the enrichment analysis. **Table 78** details the concentrations significantly altered single compounds identified on univariate analysis.

Figure 49 *Metabolite pathways altered in Group B in comparison to Group A.*

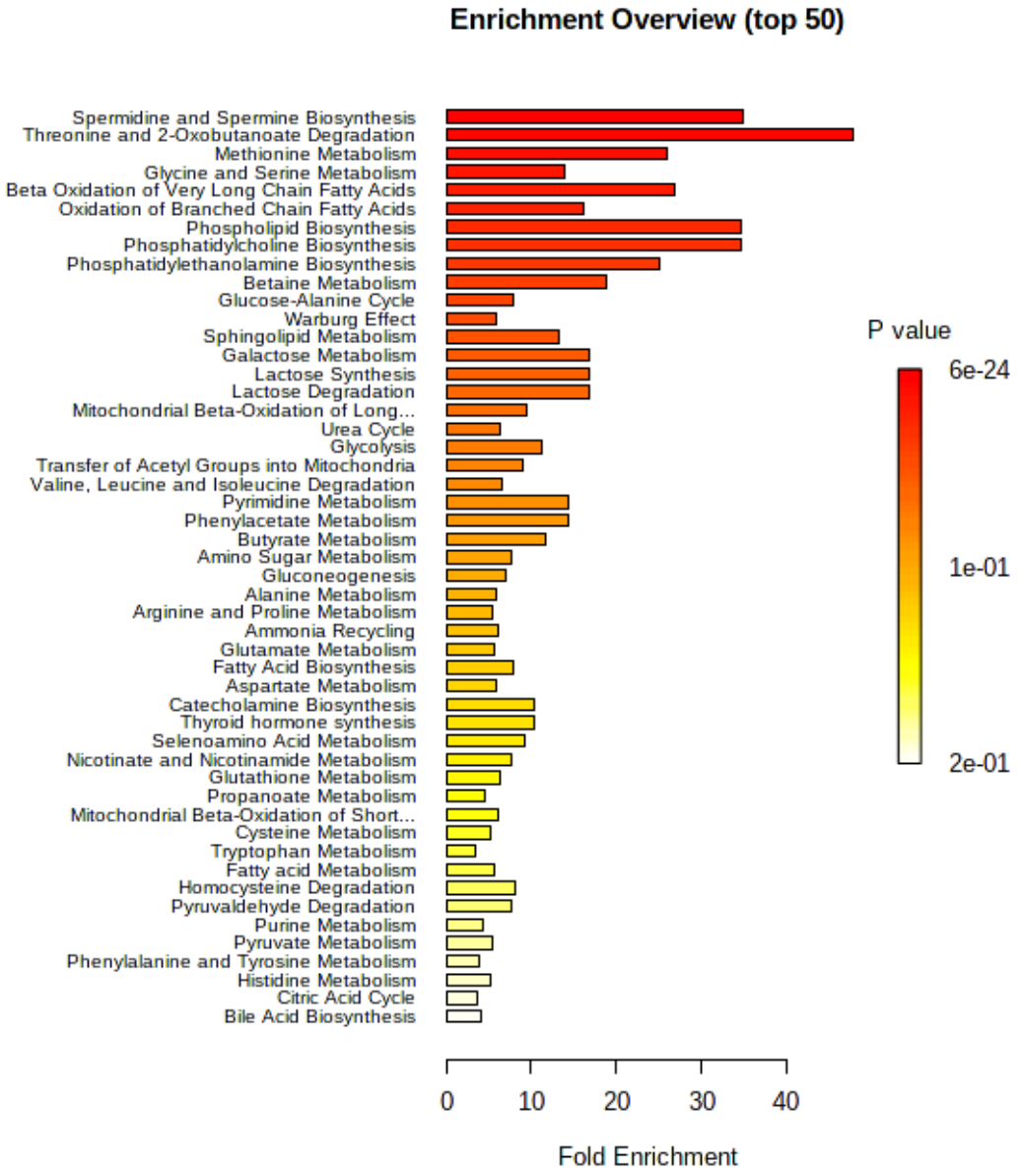


Table 80 Top ten metabolite pathways identified as altered in Group B compared to Group A.

	Metabolite Set	Total	Hits	P value	Holm P	FDR
	Spermidine and Spermine Biosynthesis	18	5	9.0418E-26	5.88E-24	5.88E-24
	Threonine and 2-Oxobutanoate Degradation	20	1	2.2939E-25	1.47E-23	7.46E-24
	Methionine Metabolism	43	8	1.4615E-23	9.21E-22	3.14E-22
	Glycine and Serine Metabolism	59	13	4.0078E-22	2.48E-20	6.51E-21
	Beta Oxidation of Very Long Chain Fatty Acids	17	2	2.2416E-14	1.37E-12	2.90E-13
	Oxidation of Branched Chain Fatty Acids	26	5	2.6791E-14	1.61E-12	2.90E-13
	Phospholipid Biosynthesis	29	1	3.1725E-13	1.87E-11	2.58E-12
	Phosphatidylcholine Biosynthesis	14	1	3.1725E-13	1.87E-11	2.58E-12
	Phosphatidylethanolamine Biosynthesis	12	2	2.2086E-11	1.26E-9	1.60E-10
	Betaine Metabolism	21	3	2.022E-10	1.13E-8	1.31E-9

***Legend**: Quantitative Enrichment Analysis Results. Details of significantly altered metabolite pathways as described in the Small Molecule Pathway Database (SMPDB.ca). Total = number of metabolites in pathway, Hits = number of metabolites in ImageKids serum belonging in pathway.*

Discussion

Although a baseline metabolite profile was not able to predict future inflammation and/or damage status on MRE, it was found to differentiate groups of patients over time. Patient featuress such as CRP and need for treatment are additional features associated with significant differences in metabolites.

Investigating the metabolic pathways that may be altered between these groups of patients suggests these groups may have differences related to pathophysiology of barrier integrity and immune function. Spermidine and spermine are polyamines made endogenously in many tissues.[169] They are involved in regulating DNA synthesis, cell migration, proliferation, apoptosis, and differentiation.[251] Group B had significantly higher ornithine, L-methionine, spermine, spermidine, and putrescine, all involved in spermidine and spermine biosynthesis.

The second most affected metabolic pathway was threonine and 2-oxobutanoate degradation. In this pathway only threonine was significantly impacted and was lower in Group B, suggesting that this degradation pathway is enriched in this group. Threonine may be related to dietary intake such as meat, cottage cheese or wheat germ,[252] but also may be involved in promoting cell mediated immune defenses.[251]

The third most significantly impacted metabolic pathway, methionine metabolism, may be related to gut barrier integrity. L-methionine supplementation in feed improved barrier integrity in pigs, and may enhance intestinal epithelial cell survival in the face of pathogens such as enterotoxigenic *E. coli*.[253, 254] Compounds involved in methionine metabolism include putrescine and spermidine, which as previously described are increased in group B.

The changes in these metabolic pathways along with the finding of increased inflammation and damage on MRE in Group A suggest that these metabolites reflect less severe disease in these pCD patients. PGA and wPCDAI scores were not significantly different between the two groups, suggesting these non-invasive measures may be insensitive to subtle metabolic differences associated with the MRE identified damage and inflammation. Patient body mass index Z-score did not correlate with metabolite clusters, and groups were a mix of both newly

diagnosed and previously known CD patients, suggesting no relationship between these metabolites and disease duration or anthropometrics.

Overall, these findings suggest that serum metabolomics may provide a sensitive indicator of metabolic changes associated with damage and inflammation that are not detectable using non-invasive clinical assessment but may be visualized on MRE.

Appendix 2: Bioinformatics Workflow

```
#!/bin/sh

#### PROJECT: MICROB727
#### DESCRIPTION: Microbiome in CD and UC patients under UDO treatment
#### DATE: 190812
#### RESEARCHER: Stephanie Dijk
#### PROGRAMMER: Juan Jovel (jovel@ualberta.ca)



########## STEP 1. Import FASTQ data into QIIME2 artifact

echo "Importing data into QIIME artifact"

qiime tools import --type 'SampleData[PairedEndSequencesWithQuality]' --
input-path manifest.csv --output-path paired-end-demux.qza --input-format
PairedEndFastqManifestPhred33

########## STEP 2. Create a summary for the demultiplexed libraries,
including
# quality scores plots

echo "Creating a summary for the demultiplexed libraries, including quality
scores plots"

qiime demux summarize --i-data paired-end-demux.qza --o-visualization
demux.qzv

########## STEP 3. Denoise (depurate) sequences using the Deblur algorithm
# (alternatively, the DADA algorithm can be used)

echo "Denoising (depurating) sequences using the Deblur algorithm"

### 3.1 Quality filtering

echo "Quality filtering"

qiime quality-filter q-score --i-demux paired-end-demux.qza --o-filtered-
sequences demux_filtered.qza --o-filter-stats demux-filter-stats.qza

# By inspecting artifact demux.qzv in https://view.qiime2.org, it can be seen
that # quality of sequences drops around position 220 nt.
# This was used as a trimming quality threshold.

### 3.2 Apply Deblur workflow

echo "Applying Deblur workflow"

qiime deblur denoise-16S --i-demultiplexed-seqs demux_filtered.qza --p-trim-
length 220 --o-representative-sequences rep-seqs-deblur.qza --o-table table-
deblur.qza --p-sample-stats --o-stats deblur-stats.qza
```

```
### 3.3 Generate summary statistics for Deblur results

echo "Generating summary statistics for Deblur results"

qiime metadata tabulate --m-input-file demux-filter-stats.qza --o-
visualization demux-filter-stats.qzv

qiime deblur visualize-stats --i-deblur-stats deblur-stats.qza --o-
visualization deblur-stats.qzv
########## STEP 4. Create visual summaries of data

# Here, a metadata file is required, which describes the samples and the
groups
# they belong to.

# Generate a Feature table that provides information on how many sequences
are
# associated with each sample and with each feature
# It also provides histograms of such distributions and summary statistics.

echo "Creating visual summaries"

qiime feature-table summarize --i-table table-deblur.qza --o-visualization
table.qzv --m-sample-metadata-file mapFile.txt

# The feature-table tabulate-seqs command will provide a mapping of feature
IDs
# to sequences, and provide links to easily BLAST each sequence against the
NCBI nt # database.

echo "Creating tabular summaries"

qiime feature-table tabulate-seqs --i-data rep-seqs-deblur.qza --o-
visualization rep-seqs.qzv

########## STEP 5. Generate a tree for phylogenetic diversity analysis

# The following step will generate a Phylogeny QIIME2 artifact

# First a multiple sequence alignment is conducted using the
FeatureData[Frequency]
# created in the previous step. Results will be stored in a
# FeatureData[AlignedSequence] artifact

echo "Conducting alignments"

qiime alignment mafft --i-sequences rep-seqs-deblur.qza --o-alignment
aligned-rep-seqs.qza

# The alignment is masked (or filtered) to remove positions that are highly
# variable, to remove noise

echo "Masking alignments"

qiime alignment mask --i-alignment aligned-rep-seqs.qza --o-masked-alignment
masked-aligned-rep-seqs.qza
```

```
# Apply FastTree to generate a phylogenetic tree from the masked alignment

echo "Building unrooted tree"

qiime phylogeny fasttree --i-alignment masked-aligned-rep-seqs.qza --o-tree
unrooted-tree.qza

# The FastTree program creates an unrooted tree, so now we apply midpoint
rooting
# to place the root of the tree at the midpoint of the longest tip-to-tip
distance # in the unrooted tree.

echo "Rooting tree"

qiime phylogeny midpoint-root --i-tree unrooted-tree.qza --o-rooted-tree
rooted-tree.qza
########## STEP 6. ALPHA and BETA diversity analysis

### 6.1 Calculate core diversity (sub-sampling 10,000 sequences)

echo "Calculating core diversity metrics"

qiime diversity core-metrics-phylogenetic --i-phylogeny rooted-tree.qza --i-
table table-deblur.qza --p-sampling-depth 10000 --m-metadata-file mapFile.txt
--output-dir core-metrics-results

### 6.2 Calculate Alpha diversity

echo "Calculating alpha diversity"

qiime diversity alpha-group-significance --i-alpha-diversity core-metrics-
results/faith_pd_vector.qza --m-metadata-file mapFile.txt --o-visualization
core-metrics-results/faith-pd-group-significance.qzv

########## STEP 7. Conduct PERMANOVA analysis

echo "Conducting PERMANOVA analysis"

qiime diversity beta-group-significance --i-distance-matrix core-metrics-
results/unweighted_unifrac_distance_matrix.qza --m-metadata-file mapFile.txt
--m-metadata-column group –o-visualization
core-metrics-results/unweighted-unifrac-group-significance.qzv --p-pairwise

########## STEP 8. ALPHA rarefaction plotting

echo "Plotting alpha rarefaction"

qiime diversity alpha-rarefaction --i-table table-deblur.qza --i-phylogeny
rooted-tree.qza --p-max-depth 10000 --m-metadata-file mapFile.txt --o-
visualization alpha-rarefaction.qzv

########## STEP 9. Taxonomical classification of sequences

### 9.1 Train a classifier

mkdir ./training-feature-classifiers
```

```
cd ./training-feature-classifiers

# As per September 10, 2018, the last version of the Greengenes database can
be
# downloaded at: http://greengenes.secondgenome.com/downloads/database/13_5

# Define the path of the database to use to train the classifier

db99=/data/tagc1/juan/projects/16S/db/gg_13_8_otus/rep_set/99_otus.fasta

echo "Importing OTUs for taxonomical classification"

qiime tools import --type 'FeatureData[Sequence]' --input-path $db99 --
output-path 99_otus.qza




# Add taxonomy
# Define the path of the database containing the taxonomy of sequences

db99tax=/data/tagc1/juan/projects/16S/db/gg_13_8_otus/taxonomy/99_otu_taxonom
y.txt

echo "Importing taxonomy"

qiime tools import --type 'FeatureData[Taxonomy]' --input-format
HeaderlessTSVTaxonomyFormat --input-path $db99tax --output-path ref-
taxonomy.qza

# Trim only the region of interest from the reference sequences, in this case
V3-V4
# regions, using the gene specific primers
# --p-f-primer CCTACGGGNGGCWGCAG --p-r-primer GACTACHVGGGTATCTAATCC
# If other regions of 16S are used, the primers sequences have to be adjusted
# accordingly

echo "Trimming V3-V4 regions from the database"

qiime feature-classifier extract-reads --i-sequences 99_otus.qza --p-f-primer
CCTACGGGNGGCWGCAG --p-r-primer GACTACHVGGGTATCTAATCC --p-trunc-len 220 --o-
reads ref-seqs.qza

# Now we can train the classifier

echo "Training the classifier"

qiime feature-classifier fit-classifier-naive-bayes --i-reference-reads ref-
seqs.qza --i-reference-taxonomy ref-taxonomy.qza --o-classifier
classifier.qza

# Now that the classifier has been trained we can use it to classify our
sequences

cd ..
```

```
# Perform taxonomical classification

echo "Performing taxonomical classification"

qiime feature-classifier classify-sklearn --i-classifier ./training-feature-
classifiers/classifier.qza --i-reads rep-seqs-deblur.qza --o-classification
taxonomy.qza

# Summarize taxonomy results

echo "Summarizing taxonomy results"

qiime metadata tabulate --m-input-file taxonomy.qza --o-visualization
taxonomy.qzv

# Generate taxonomy barplots

echo "Creating taxonomy barplots"

qiime taxa barplot --i-table table-deblur.qza --i-taxonomy taxonomy.qza --m-
metadata-file mapFile.txt --o-visualization taxa-bar-plots.qzv

########## STEP 10. Testing differential abundance with ANCOM

echo "Testing differential abundance with ANCOM"

# Since ANCOM does not handle zeroes, a pseudo-counts table will be generated
to eliminate zeroes

qiime composition add-pseudocount --i-table table-deblur.qza --o-composition-
table comp-table.qza

# Now run ANCOM

qiime composition ancom --i-table comp-table.qza --m-metadata-file
mapFile.txt --m-metadata-column group --o-visualization ancom-group.qzv

# To run ANCOM at a specific taxonomic level (L6 for example):

qiime taxa collapse --i-table table-deblur.qza --i-taxonomy taxonomy.qza --p-
level 6 --o-collapsed-table table-l6.qza

### STEP 10.1 Alternatively, differential abundance can be tested with GNEISS

echo "Testing differential abundance with GNEISS"

# Cluster samples according to their correlation

qiime gneiss correlation-clustering --i-table composition_gneiss.qza --o-
clustering hierarchy_gneiss.qza

# Build a linear regression model using balances

qiime gneiss ilr-hierarchical --i-table table-deblur-yg2.qza --i-tree
hierarchy_gneiss-yg2.qza --o-balances balances-yg2.qza

# Run regression analysis
```

```
qiime gneiss ols-regression --p-formula "color" --i-table balances_gneiss.qza
--i-tree hierarchy_gneiss.qza  --m-metadata-file mapFile.txt --o-
visualization regression_summary_gneiss.qzv

########## STEP 11. Conduct metabolic profiling with PICRUSt
# PICRUSt only works with Greengenes, not with SILVA

echo "Conducting PICRUSt analysis"

# Extract OTU table
# define path of reference OTUs file:

db99=/data/tagc1/juan/projects/16S/db/gg_13_8_otus/rep_set/99_otus.fasta

# Import reference OTUs sequences (99% in this case) into a qiime2 (qza)
artifact

qiime tools import --input-path $db99 --output-path 99_otus.qza --type
'FeatureData[Sequence]'

# cluster sequences according to closed references

qiime vsearch cluster-features-closed-reference --i-table table-deblur.qza --
i-sequences rep-seqs-deblur.qza --i-reference-sequences 99_otus.qza --p-perc-
identity 0.99 --o-clustered-table table-cr-99.qza --o-clustered-sequences
rep-seqs-cr-99.qza --o-unmatched-sequences unmatched-cr-99.qz



# Export to biom file

qiime tools export --input-path table-cr-99.qza --output-path exported-
feature-table_99

cd exported-feature-table_99

# Normalize data with PICRUSt script

~/miniconda2/bin/normalize_by_copy_number.py -i feature-table.biom -o
normalized_otus.biom

# Predict microbiome functionality with PICRUSt script

~/miniconda2/bin/predict_metagenomes.py -f -i normalized_otus.biom -o
metagenome_predictions.tsv -a nsti_per_sample.tab

echo “The pipeline was completed!”
```