Understanding the Etiology of Inflammatory Complications Following Ileal Pouch-Anal Anastomosis

A Study of the Genetic, Serological and Microbial Factors Associated with Ileal Inflammation

by

Andrea Dayna Tyler

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Institute of Medical Science
University of Toronto

© Copyright by Andrea D. Tyler, 2013
Understanding the Etiology of Inflammatory Complications Following Ileal Pouch-Anal Anastomosis

A Study of the Genetic, Serological and Microbial Factors Associated with Ileal Inflammation

Andrea Dayna Tyler
Doctor of Philosophy
Institute of Medical Science
University of Toronto
2013

Abstract

Introduction: Inflammatory pouch complications, including pouchitis, chronic pouchitis (CP) and a Crohn’s disease-like phenotype (CDL) of the pouch following ileal pouch-anal anastomosis (IPAA), are relatively common, and arise via unknown mechanisms. The phenotypic similarities between pouch inflammation and inflammatory bowel disease (IBD) suggest there may be common pathways involved in both disorders. The aim of this thesis is to investigate the serological, genetic and microbial factors contributing to the development of pouch inflammation in a large, well characterized patient cohort.

Methods: Subjects with IPAA were recruited, and clinical and demographic information was obtained through medical chart review and patient questionnaire, allowing patients to be grouped based on post-surgical phenotype. Blood and tissue was collected for genetic, serological and microbial analyses. Anti-microbial antibodies were detected using enzyme-linked immunosorbent assay (ELISA), genotyping was carried out using the
Illumina Goldengate custom SNP assay and Sequenome iPLEX platform, and tissue-associated microbial communities were assessed using 454 pyrosequencing.

Results: Among our cohort, smoking was associated with CDL ($P=0.003$) and Ashkenazi Jewish heritage with CP ($P<0.008$). $NOD2$insC (rs2066847) ($P=7.4\times10^{-5}$), anti-CBir1 ($P<0.0001$) and ASCA (IgG) ($P=0.03$) were significantly associated with inflammatory pouch outcomes. Additional SNPs in $NOX3$, $DAGLB$, and $NCF4$ were also marginally associated with pouch outcome. A multi-variable risk model combining clinical, serologic and genetic markers was constructed and could differentiate between chronic pouch inflammatory phenotypes and no pouchitis. Genus level microbial analysis demonstrated that several organisms ($Bacteroides$, $Parabacteroides$, $Blautia$ and $Moryella$) were detected less frequently among the inflammatory outcome groups ($P<0.05$). These associations remained significant even following adjustment for antibiotic use, smoking, country of birth and gender.

Conclusions: CD-associated anti-microbial antibodies and genetic markers are associated with chronic inflammatory pouch phenotypes. Additionally, changes in the composition of the pouch associated microbiome are associated with inflammation. These observations suggest that similar mechanisms may be involved in non-surgical IBD and pouchitis.
Acknowledgements

I would like to thank my Supervisor, Dr. Mark Silverberg and my thesis advisory committee (Dr. Zane Cohen, Dr. Nicola Jones, Dr. Ken Croitoru, Dr. David Guttman) for their support and encouragement throughout my graduate studies. Their guidance and advice was instrumental in the completion of this work. I was so lucky to have had the opportunity to have learned from the expertise of each of these individuals.

I would also like to thank the members of the Silverberg lab, most especially Boyko Kabakchiev and Joanne Stempak for their assistance with these studies, their willingness to teach, and their friendship over the years.

All of the work done as part of this thesis could not have been performed without the enthusiastic support of the patients included in these studies. These individuals happily shared their blood, tissue and time, in hopes of contributing to this work. I thank them for their unwavering support and for being a constant reminder of why I chose to pursue a degree in inflammatory bowel disease research.

I am also grateful for the financial support provided to me to complete this work by the Crohn’s and Colitis Foundation of Canada, and both The Department of Medicine and the Samuel Lunenfeld Research Institute at Mount Sinai Hospital.

Finally, I would like to thank my parents (Maureen and Kip Tyler), my Grandmothers (Dr. Marvis Tutiah, and Wilma Mitchell) and my Toronto ‘family’ for their support, guidance, availability for counselling, home cooking, laughter, and for inspiring me to reach higher than I might otherwise have imagined possible.
Table of Contents

Chapter 1: Introduction and Background................................................................. 1
  1.1 Inflammatory bowel disease ............................................................................. 1
    1.1.2 Risk factors for IBD ................................................................................. 2
    1.1.3 Ileal pouch-anal anastomosis and pouch inflammatory complications .......... 3
  1.2 Serology ........................................................................................................... 6
    1.2.1 A description of the anti-microbial antibodies identified in IBD ................. 7
    1.2.2 Association of serological markers with CD and UC ................................ 10
    1.2.3 IBD serology in pouch inflammatory complications ..................................... 12
    1.2.4 The role of anti-microbial antibodies in inflammatory disease pathogenesis ...... 12
  1.3 Genetics ........................................................................................................... 15
    1.3.1 Analyzing the human genome .................................................................... 15
    1.3.2 SNP genotyping: platforms and analysis .................................................... 17
    1.3.3 - Genetic Associations with IBD ................................................................. 23
    1.3.4 IBD genetic association studies in non-Caucasian populations ..................... 27
  1.4 Bacteria .......................................................................................................... 29
    1.4.1 Phylogenetic organization and taxonomic classification of bacteria ............. 30
    1.4.2 Human microbiome analysis techniques ..................................................... 35
    1.4.3 Culture-independent microbial community analysis ...................................... 36
      1.4.3.1 Sample procurement .......................................................................... 36
      1.4.3.2 Microbial DNA extraction from samples .............................................. 37
      1.4.3.3 Overview of Sequence-based microbial analysis tools ............................ 38
      1.4.3.4 A more detailed examination of next-generation sequencing and how it is applied to the analysis of the microbiome. .................................................. 41
      1.4.3.5 Sequence quality trimming, alignment and taxonomic assignment ........ 44
      1.4.3.6 Statistical analysis of microbial data .................................................... 50
    1.4.4 The Human Microbiome – lessons from community level analysis .............. 53
    1.4.5 Location-specific microbiome heterogeneity .............................................. 56
    1.4.7 Genes, serology and microbiome ............................................................... 59
    1.4.8 Microbiome in IBD .................................................................................... 60
    1.4.9 Microbiome and pouchitis ....................................................................... 64
  1.5 Overarching theme - bringing it all together.................................................... 65
1.6 Hypothesis and Aims ............................................................................................................ 65

Chapter 2: Antimicrobial Antibodies Are Associated with a Crohn's Disease-Like Phenotype following Ileal Pouch-Anal Anastomosis .......................................................... 67
  
  2.1 Abstract: .................................................................................................................. 70
  
  2.2 Introduction ............................................................................................................. 71
  
  2.3 Materials and Methods ............................................................................................ 71
  
  2.4 Results ..................................................................................................................... 74
  
  2.5 Discussion ............................................................................................................... 81

Chapter 3: The NOD2insC polymorphism is associated with worse outcome following ileal pouch-anal anastomosis for ulcerative colitis .......................................................... 85
  
  3.1 Abstract: ................................................................................................................ 87
  
  3.2 Introduction: ............................................................................................................ 89
  
  3.3 Methods: ................................................................................................................. 89
  
  3.4 Results: .................................................................................................................... 93
  
  3.5 Discussion ............................................................................................................. 105

Chapter 4: Characterization of the gut-associated microbiome in inflammatory pouch complications following ileal pouch-anal anastomosis ................................................. 109
  
  4.1 Abstract: ................................................................................................................. 111
  
  4.2 Introduction ........................................................................................................... 112
  
  4.3 Materials and Methods .......................................................................................... 113
  
  4.4 Results ................................................................................................................... 115
  
  4.5 Discussion ............................................................................................................. 128
  
  4.6 Chapter 4, Supplementary Methods ...................................................................... 134

5.1 Conclusions and general discussion .......................................................................... 136
  
  5.2 Future directions ...................................................................................................... 147

Appendix .......................................................................................................................... 150

References ....................................................................................................................... 185
List of tables

Table 1: Serological markers included in the Prometheus antibody panel and associated with IBD. ............................................................................................................................. 9
Table 2: Summary of the different technologies available for genotyping. ....................... 19
Table 3: Sample coding for genetic models typically used in genetic association analyses. ........................................................................................................................................... 21
Table 4: Advantages and disadvantages of using 16S rRNA gene-based analyses in construction of bacterial phylogenies and taxonomic assignment. ................................. 34
Table 5: Summary of the next-generation sequencing technologies currently available. 41
Table 6: Outcome groups following IPAA ....................................................................... 73
Table 7: Relationship between demographic characteristics and outcome following IPAA. ............................................................................................................................... 77
Table 8: Odds Ratios and Confidence intervals associated with increasing antibody positivity. ........................................................................................................................................ 80
Table 9: Clinical and phenotypic characteristics of the study population. ......................... 95
Table 10: Single nucleotide polymorphisms (SNPs) associated with outcome. .................. 98
Table 11: Odds ratios (OR) and confidence intervals (CI) for the five single nucleotide polymorphisms (SNPs) which are associated with outcome. ........................................ 99
Table 12: Single nucleotide polymorphisms (SNPs) previously associated with pouch outcome. ................................................................................................................. 101
Table 13: Factors significantly associated with outcome by multi-variate analysis. .......... 103
Table 14: Phenotypic characteristics of individuals among the four outcome groups. .. 117
Table 15: Association between outcome and bacterial positivity in multivariate analysis with smoking, birth country and gender included as co-variates. ......................... 126
List of figures

Figure 1: Hierarchical organization of taxonomic levels used for classifying organisms 32
Figure 2: Regions of 16S rRNA molecule typically used for evaluating the human microbiome. .................................................................................................................. 42
Figure 3: Proportions of patients (n=399) in each of the three outcome groups .......... 76
Figure 4: Overall levels of antibody positivity among three outcome groups.......... 79
Figure 5: Proportion of individuals in each specific outcome group.......................... 96
Figure 6: ROC curves generated for the risk score analysis................................. 104
Figure 7: Phylum level comparisons between four outcome groups ...................... 119
Figure 8: Proportion of patients positive for genera which were significantly associated with outcome................................................................. 121
Figure 9: Proportion of patients positive for genera in individuals with inflamed vs not-inflamed pouches................................................................. 123
Figure 10: Hypothesized model of the pathogenesis of ileal inflammation ............ 143
Abbreviations

IBD – Inflammatory bowel disease
CD – Crohn’s disease
UC – Ulcerative colitis
IBDU – IBD type unclassified
IC – Indeterminate colitis
IPAA – Ileal pouch anal anastomosis
PSC - Primary sclerosing cholangitis
FAP – Familial adenomatous polyposis
TNF – Tumor necrosis factor
ELISA - Enzyme linked immunosorbent assay
ASCA – Anti-*Saccharomyces cerevisiae* antibodies
pANCA - Perinuclear antineutrophil cytoplasmic antibodies
OmpC – Outer membrane porin C
SNP - Single nucleotide polymorphism
TDT – Transmission disequilibrium test
HWE - Hardy-Weinberg equilibrium
GWA – Genome-wide association
MAF - Minor allele frequency
FDR – False discovery rate
ROS – Reactive oxygen species
eQTL – Expression quantitative trait loci
HGT – Horizontal gene transfer
RDP – Ribosomal database project
T-RFLP – Terminal restriction fragment length polymorphism
SSCP - Single-strand conformation polymorphism
qPCR – Quantitative polymerase chain reaction
FISH – Fluorescence in situ hybridization
OTU – Operational taxonomic unit
BLAST – Basic local alignment search tool
NCBI – National centre for biotechnology information
MSA – Multiple sequence alignment
5-ASA – 5-aminosalicylic acid
MAMP – Microbial-associated molecular pattern
PSA – Polysaccharide A
TLR – Toll-like receptor
IL – Interleukin
NOD - Nucleotide-binding oligomerization domain-containing protein
IFN – Interferon
MHC – Major histocompatibility complex
LPS - Lipopolysaccharide
CRP – C-reactive protein
ESR – Erythrocyte sedimentation rate
SLiME - Synthetic learning in microbial ecology
AIEC – Adherent-invasive E. coli
FT – Fecal transplant
NP/NCP – No pouchitis/no chronic pouchitis
CP – Chronic pouchitis
P – Pouchitis
CDL – Crohn’s disease-like phenotype
MSH - Mount Sinai Hospital
HMC - Penn State Milton S. Hershey Medical Center
IBD-U – IBD unclassified
OR – Odds ratio
CI – Confidence interval
ROC - Receiver operating characteristic
AUC - Area under the curve
PDAI – Pouchitis disease activity index
PAS - Pouchitis activity score
Chapter 1: Introduction and Background

1.1 Inflammatory bowel disease

The Inflammatory Bowel Diseases (IBD), comprising Crohn’s disease (CD) and ulcerative colitis (UC) are chronic inflammatory conditions of the digestive tract of unknown etiology. In Canada, where rates of IBD are among the highest in the world, between 0.67 and 0.9% of the population is affected, with an approximately equal proportion of patients with a diagnosis of UC and CD\textsuperscript{1-3}. It is conservatively estimated, in a report provided by the Crohn’s and Colitis Foundation of Canada, that 28 Canadians are diagnosed with IBD daily. Symptoms are varied, depending on the location of disease, however commonly include abdominal cramping, nausea or vomiting, increased stool frequency, bloody stool, fatigue, anemia and weight loss\textsuperscript{2}. Children diagnosed with IBD may experience delays in physical and psychological development. Medical management of both CD and UC are similar with anti-inflammatories (5-ASA), immunosuppressants (methotrexate, azathioprine), steroids and biological agents (anti-TNFα) the current mainstays of treatment. However, IBD has no cure, and represents a significant healthcare burden\textsuperscript{2}.

Historically, CD and UC have been considered separate entities with different phenotypic characteristics and outcomes. CD can discontinuously affect any part of the digestive tract from mouth to anus and is often characterized by transmural inflammation. Additionally, patients may go on to develop structuring or fistulizing phenotypes which frequently require surgical management. As a result of such complications, up to 80% of CD patients will have to undergo some form of surgical resection during the course of their disease\textsuperscript{4}. While CD surgery is typically a temporary solution, with disease often recurring in nearby tissue, benefits include the obviation of obstructive and penetrating symptoms, decreased need for immunosuppressive medication and an increase in patient quality of life. UC, on the other hand, affects mainly the superficial layers of mucosa of the large bowel, and extends in a continuous fashion proximally from the rectum. Patients with UC do not develop structuring or fistulizing phenotypes and seldom have inflammation extending into the small bowel. In UC cases where surgical intervention is
necessary, subtotal colectomy or procto-colectomy is performed, and is considered curative in intent as all UC-susceptible tissue (colon) is removed. An additional subset of patients display traits characteristic of both disorders and are referred to as IBD type unclassified (IBDU). The Montreal classification further identifies patients with diagnostic features of both CD and UC specifically at the time of surgical resection as indeterminate colitis (IC)\(^5\). However, evidence suggests that the IBDs exist more as a spectrum, with overlapping phenotypic characteristics, disease mechanisms and responses to therapeutic and surgical management.

1.1.2 Risk factors for IBD

A great deal of study has been devoted to identifying risk factors for the development of IBD. Evidence of heritable risk factors include the increased incidence of IBD among individuals of Ashkenazi Jewish heritage\(^6\), and among individuals with a family history of IBD\(^7\). This relationship is slightly stronger in CD, but important in both diseases. Smoking and exposure to second hand smoke is among the best studied environmental risk factors, having been associated with an increased risk of CD, and a decreased risk of UC\(^8,9\). Additional environmental factors which have been suggested to play a role in the recent dramatic increase in IBD prevalence through much of the developed world include reduced breastfeeding, increased hygiene and sanitation, reduction in physical activity, exposure to pollution, consumption of a Western diet\(^10\), and use of antibiotics among children\(^11\). However, studies examining these factors have been associative in nature and are complicated by the difficulties in obtaining information on patients prior to the onset of symptoms or disease.

There is also evidence that multiple auto-inflammatory processes may occur via similar mechanisms. Individuals with IBD are at increased risk of comorbitides including primary sclerosing cholangitis (PSC), pyoderma gangrenosum, arthritis, and uveitis among others\(^12\). Furthermore, the presence of such co-morbidities may also signify a more severe disease course.

IBD is predominantly a Western disease with significantly higher incidence and prevalence in North America and northern Europe compared to other more equatorial
locations. This has led to speculation that environmental factors specific to Western nations, in conjunction with increased genetic susceptibility, leads to disease. Evidence supporting this hypothesis includes changes in the prevalence of IBD among first generation Canadians. While newly arrived immigrants, particularly those coming from equatorial locations, tend to have lower rates of IBD than do their Canadian counterparts, the risk of IBD among their children is equal to or exceeds the Canadian average\textsuperscript{13,14}. Further, a recent study examining the risk of IBD among South Asian immigrant populations within Canada observed that rates of IBD among second generation Canadians (children) was significantly higher than non-South Asians\textsuperscript{4,15}. Disease phenotypes tended to be more severe among this group as well, with extensive colonic disease more common than among non-south Asian individuals\textsuperscript{13}. Similar results were obtained in a study conducted in the United Kingdom\textsuperscript{15}.

Prevailing hypotheses regarding IBD pathogenesis take into account both genetic and environmental factors. It is believed that IBD susceptibility is partially genetically mediated with specific genetic loci modulating disease risk. However, in the absence of a specific environmental trigger, which is commonly believed to be microbial in origin, IBD may not develop. Thus, IBD is thought to result from a complex interaction between host genetics and microbial environment, with ultimate phenotype determined by the specific gene-microbe interactions.

1.1.3 Ileal pouch-anal anastomosis and pouch inflammatory complications

Individuals with severe or medication refractory UC, or who develop dysplasia, are candidates for total procto-colectomy and ileostomy or ileal pouch-anal anastomosis (IPAA). These patients are considered surgically ‘cured’ of their disease, due to the localized disease pattern characteristic of this form of IBD. Rather than receiving a conventional ileostomy, individuals undergoing colectomy for UC may elect to undergo IPAA, which allows transanal passage of stool and eliminates the requirement of an ileostomy. This procedure involves the surgical construction of a pouch or reservoir from the terminal ileum which is then connected to the rectal cuff or anus. IPAA is only rarely performed on individuals with CD, due to an increased risk of post-surgical complications, such as recurrence of chronic inflammation in the pouch or afferent limb,
or development of CD-related fistulas\textsuperscript{16}. Individuals with familial adenomatous polyposis (FAP), a condition characterized by proliferation of polyps throughout the large bowel, also often undergo IPAA, in order to prevent the development of colorectal cancer\textsuperscript{17}. When performed by experienced physicians, colectomy with IPAA has a relatively low rate of surgical complication and has been shown to enhance patient experience and quality of life when compared to an ileostomy\textsuperscript{18}.

Despite the surgical removal of UC susceptible tissue, many patients experience pouchitis, defined as inflammation of the ileal reservoir. Pouchitis is the most common complication following IPAA, affecting up to 50% of UC patients at some time following surgery, while occurring only rarely among individuals with FAP. Patients with pouchitis experience increased stool frequency, rectal bleeding, fecal urgency, cramping, malaise and occasionally fever. The endoscopic and histological picture is similar to that of IBD, with edema, erythema, loss of vascular pattern friability and ulceration seen on pouchoscopy and polymorphonuclear/mononuclear leukocyte infiltration, microscopic ulceration, and villus atrophy observed on histological examination\textsuperscript{19, 20}. Unlike IBD, most cases of pouchitis respond rapidly to antibiotic therapy and are transient in nature\textsuperscript{21}. However, a subset of patients with pouchitis will require long term antibiotics in order to maintain healthy pouches, or will require the use of additional therapeutic agents in order to remain healthy. Medications which may be successfully used in treating antibiotic refractory pouchitis include those which are typically used to treat IBD, such as corticosteroids, immunomodulators and anti-tumor necrosis factor (TNF\textsubscript{a}) agents. An additional subset of patients with IPAA for UC, will develop what has been described as a CD-like phenotype, with inflammation proximal to the pouch in either the afferent limb or upper digestive tract, or the development of a stricture, fistula or abscess which is not related to surgical complications\textsuperscript{21}. Such chronic inflammatory phenotypes have been suggested to have been misdiagnosed pre-colectomy CD, yet thorough chart review often demonstrates no evidence for this diagnosis prior to surgery or following pathological review of the surgical specimen\textsuperscript{22}.

Physicians often successfully treat patients presenting with symptoms of pouchitis empirically with antibiotics. However, it is often unclear whether pouch inflammation is indeed the cause of these symptoms. Several clinical tools for diagnosing pouchitis, and
differentiating this outcome from others have been developed. The pouchitis disease activity index (PDAI) and pouchitis activity score (PAS) make use of a combination of clinical, endoscopic and histological data in order to diagnose patients with pouchitis\textsuperscript{20}. In a recent analysis by Ben Bassat et al\textsuperscript{23}, poor correlation was shown between the clinical and both endoscopic and histological components of each of these scores, as well as between the endoscopic and histological subscores. This suggests that clinical symptoms are non-specific, and highlights the importance of using more objective measures of pouch inflammation, especially in the context of studies specifically analyzing etiological factors contributing to inflammation. Objective measures for accurately assessing inflammation which rely on biomarkers rather than invasive procedures such as endoscopy, would clearly be beneficial.

Several factors have been associated with an increased risk of pouchitis development following IPAA. The most important risk factor is a pre-colectomy diagnosis of IBD. Individuals with FAP rarely develop pouch inflammation, whereas those with UC are much more susceptible. Individuals with CD are only rarely offered this procedure due to the likelihood of disease recurrence. IC has been associated with an increased risk of complications compared to UC in some studies, although not all, suggesting that the risk of pouch inflammation for these individuals may be similar to that seen for those with severe UC\textsuperscript{24}. Some evidence also suggests that more extensive UC prior to surgery or the presence of backwash ileitis increases the risk of inflammatory complications, although findings regarding these traits have shown mixed results\textsuperscript{25-27}. Among UC patients undergoing IPAA, a diagnosis of PSC, presence of other extra-intestinal manifestations of IBD, a family history of CD among first degree relatives and smoking history have all been associated with an increased risk of pouch inflammation\textsuperscript{25, 28-30}.

As with IBD in general, the etiology of pelvic pouch inflammatory complications is poorly understood. The phenotypic similarities between pouch inflammation and both CD and UC suggest that common disease pathways are involved in these inflammatory processes. As such, the pelvic pouch represents a useful human model for evaluating \textit{de novo} ileal inflammation and may be a useful proxy for gaining insight into general mechanisms of IBD. Furthermore, the importance of describing predictive factors which
accurately stratify patients’ risk of inflammatory complications, will have broad applications in determining which individuals are good candidates for IPAA surgery.

1.2 Serology

The current gold-standard for the diagnosis and classification of patients with IBD is ileocolonoscopy with biopsy. While highly accurate and relatively safe, this procedure is invasive and costly. Further, in complicated cases, colonoscopy may not allow differentiation between IBD subtypes. As such, the development of non-invasive, inexpensive tests which are capable of differentiating between different forms of IBD and assessing disease activity, would be useful for physicians trying to diagnose patients. Currently available and widely used clinical blood tests which are helpful in managing IBD include measurements of C-reactive protein (CRP), and erythrocyte sedimentation rate (ESR) for example, each of which provide a broad measure of inflammatory activity. CRP is perhaps the most well studied blood-based biomarker, and has been shown to provide an effective estimation of disease severity in CD. There is also some evidence that this marker may help to predict relapse or need for colectomy in UC. However, these blood markers can also be elevated in other inflammatory conditions, and thus are not specific enough on their own to provide a definitive diagnosis of IBD. Several other studies have also shown poorer correlations between CRP and bowel disease activity, and have suggested that it does not provide sufficient resolution to distinguish between CD and UC, highlighting the need for additional markers.

A newer marker, fecal calprotectin, has been shown to correlate well with various measures of IBD disease activity especially in colonic disease. Several studies have demonstrated that this marker is at least as useful as CRP in measuring disease activity, yet it too is unable to differentiate between CD and UC. Furthermore, the necessity of obtaining stool samples to carry out this test has prevented its more widespread utility.

The value of biomarkers clinically, and in IBD research is apparent. However, none of the above tests, on their own, are capable of accurately distinguishing between phenotypes, or predicting which patients are at risk of a more severe outcome. Several
newer biomarkers with interesting applications for IBD research and clinical management, both as diagnostic tools, and to assist in the sub-classification of patients with a diagnosis of CD or UC, are anti-microbial antibodies which can be detected in the serum of patients with IBD. Such markers have been well studied and demonstrate high specificity (90-100%) but rather low sensitivity (50-70%) for bowel disease\textsuperscript{37}. Thus, while these markers do not obviate the need for colonoscopy, they have shown potential in aiding in disease management. Furthermore, the ease of obtaining serological samples (draw peripheral blood from patient, allow sample to clot in the tube, remove serum for testing or storage), combined with the relative cost-effectiveness of their measurement, would make tests based on these markers widely appealing for both clinical and research applications.

1.2.1 A description of the anti-microbial antibodies identified in IBD

Many of the common and well-studied IBD associated serological markers, including anti-\textit{Saccharomyces cerevisiae} antibodies (ASCA), anti-CBir1, anti-outer membrane porin C (OmpC), and anti-I2 are antibodies which target specific microbial motifs. ASCAs are human IgA or IgG antibodies which specifically target the mannan motifs on the cell wall of \textit{S. cerevisiae}\textsuperscript{38} yet which are capable of also recognizing homologous sequences from other organisms more commonly found in the digestive tract. For example, ASCAs can also be generated against \textit{Candida albicans}\textsuperscript{39}. Anti-CBir1 targets a flagellin moiety and anti-OmpC a membrane porin of \textit{Escherichia coli}, while anti-I2 targets a membrane component of \textit{Pseudomonas fluorescens}\textsuperscript{40}. These microbes are part of the commensal microbiota, yet among healthy individuals rarely result in the production of antibodies.

Another commonly used serological marker with diagnostic use in studying IBD is perinuclear anti-neutrophil cytoplasmic antibody (pANCA). Rather than specifically recognizing microbial motifs, this marker is an autoantibody with specificity for an as yet unknown component of neutrophil granules\textsuperscript{41}. While ANCAs have been associated with several inflammatory diseases, detection of pANCA is a much more IBD-specific
phenomenon. Unlike ANCA, pANCA has nuclear localization of the staining pattern which can be detected by indirect immunofluorescence, and which disappears following treatment with DNase\textsuperscript{41}.

For each of these markers, antibody titres are assessed using enzyme linked immunosorbent assay (ELISA). Established positivity cutoff values, based on experimental results, are also commonly applied to enhance the diagnostic value and interpretability of data. Prometheus Therapeutics and Diagnostics (San Diego, CA) offers diagnostic testing for anti-microbial makers important in IBD, to provide more information to physicians attempting to make an accurate diagnosis of IBD or to distinguish between UC and CD. Table 1 describes the markers which are typically assessed by Prometheus and their described associations with IBD.
<table>
<thead>
<tr>
<th>Full Name</th>
<th>Antibody Marker</th>
<th>Target</th>
<th>IBD association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perinuclear antineutrophil cytoplasmic antibodies</td>
<td>pANCA</td>
<td>Autoantibodies directed against unknown component of neutrophil granules</td>
<td>UC</td>
</tr>
<tr>
<td>anti- <em>Saccharomyces cerevisiae</em> antibodies</td>
<td>ASCA (IgA/IgG)</td>
<td>Mannose sequences from phosphopeptidomannan in <em>S. cerevisiae</em></td>
<td>CD</td>
</tr>
<tr>
<td>Anti-outer membrane porin C</td>
<td>Anti-OmpC</td>
<td>Membrane porin, bacterial antigen originally isolated from <em>Escherichia coli</em></td>
<td>CD/UC</td>
</tr>
<tr>
<td>Anti-CBir1</td>
<td>Anti-CBir1</td>
<td>Motility associated bacterial antigen that is a component of flagellin</td>
<td>CD</td>
</tr>
<tr>
<td>Anti-I2</td>
<td>Anti-I2</td>
<td>Bacterial antigen isolated from <em>Pseudomonas fluorescens</em></td>
<td>CD/IC</td>
</tr>
</tbody>
</table>

Table 1: Serological markers included in the Prometheus antibody panel and associated with IBD.

More recent panels have incorporated genetic markers along with these serological ones, with the suggested benefit of increasing performance, according to some studies\(^42\). Additional markers, termed anti-glycan antibodies, target specific carbohydrate residues. Anti-laminaribioside carbohydrate antibodies (ALCA), anti-chitobioside carbohydrate antibodies (ACCA), anti-mannobioside carbohydrate antibodies (AMCA), respectively, recognize carbohydrate motifs from the outer membrane of bacterial cells. These three markers are detected at higher levels among individuals with CD compared to UC, and have been suggested to be useful in diagnosing IBD. Serological panels produced by Glycominds (Lod, Israel) include these markers, however they are less widely used than are the traditional markers and thus, beyond the scope of this thesis\(^43\).
1.2.2 Association of serological markers with CD and UC

The utility of serum associated markers in IBD diagnostics have been well documented. In each case, these markers are detected significantly more frequently among individuals with disease compared to healthy controls. These serological markers also have the potential to distinguish between CD and UC. Several studies have demonstrated a strong association between ASCA (IgA/IgG) and CD, and between pANCA and UC. ASCA IgA and IgG correlate well with one another, and overall ASCA positivity tends to correlate highly with small bowel CD and poorly with UC. Patients with IC have been shown to have ASCA levels which are intermediate between those found in individuals with CD and UC, an interesting observation given the intermediate nature of the IC phenotype. pANCA, on the other hand, is associated with UC and colonic disease location in CD. Both of these markers are also associated with more severe outcomes which commonly require surgical intervention. Moreover, individuals with CD and higher numbers of CD-associated markers are at increased risk of complicated disease behaviour and need for surgery. When measured alone, these markers are useful, however, the sensitivity and specificity of using them together is substantially increased.

Despite the diagnostic potential of ASCA and pANCA, approximately 30% of patients with IBD are not positive for either of these markers. Further, although fairly specific, there is some overlap in marker detection among phenotypic outcome groups, with 5-15% of UC patients positive for ASCA, and approximately 6-20% of CD patients positive for pANCA. This relationship becomes more complicated when different disease phenotypes are evaluated, as colonic CD often demonstrates a serological profile more similar to that of UC. In order to optimize diagnostic sensitivity and specificity, additional markers including anti-CBir1 and anti-OmpC, were identified and have demonstrated diagnostic potential.

Anti-OmpC is detected more frequently among individuals with CD and their unaffected relatives compared to those with UC and healthy controls. This marker has also been associated with internal penetrating and stricturing CD phenotypes. Yet the specificity of this marker in appropriately diagnosing IBD in the absence of the other markers has been called into question. In a small study examining its utility in a pediatric
cohort, seven patients were falsely found to have IBD based on anti-OmpC seroreactivity alone. Anti-CBir1 is also strongly associated with CD. This marker is detected more commonly among individuals with small bowel disease and requirement for resection, as well as among individuals with fibrostenotic and internal penetrating disease phenotypes. It is negatively associated with a more UC-like disease course. Furthermore, among CD patients who are positive for pANCA this marker was more commonly detected, suggesting that it may add specificity to the diagnostic capabilities of anti-microbial antibody panels. Interestingly, anti-CBir1 is detected more frequently among younger children (<8 years old) with IBD, and may therefore represent a better early indicator of IBD among this population. Anti-I2 is associated with CD, and a need for surgery. Interestingly, this marker was also associated with response following fecal diversion for medically resistant proctocolitis, with individuals positive for this marker more likely to experience clinical remission following this procedure. Seropositivity for all of these markers is also highly dependent on the length of time that an individual has had disease, as well as disease severity and tendency towards progression.

The majority of studies examining anti-microbial antibody markers have been conducted retrospectively. As such, the observation that serum levels of ASCA, pANCA and anti-OmpC have been shown to be fairly static over time, not varying greatly with disease activity or surgery, is important. Of the results from studies which do prospectively evaluate these markers, most findings are similar to those found in retrospective studies. One study demonstrated a substantial increase in the risk of developing IBD in first-degree relatives from multiply affected families, among individuals who were positive for at least one serological marker (measured prior to diagnosis). On the other hand, a small study evaluating pANCA positivity following colectomy showed a slight reduction in pANCA prevalence between individuals who had undergone surgery compared to those who had not. The few studies which have prospectively evaluated these markers have supported their predictive potential, and demonstrate that the described anti-microbial antibodies are detectable in many patients prior to symptom onset and IBD diagnosis, although often at lower titres.
1.2.3 IBD serology in pouch inflammatory complications

The beneficial potential of serological biomarkers in easily diagnosing and monitoring inflammatory pouch complications following IPAA is also apparent. As such, serological markers have also been examined in this context. Such studies have been limited by small sample sizes and unclear post-colectomy patient phenotypic classifications. Results to date have been mixed, with some demonstrating an association between pANCA, ASCA and anti-CBir1 and pouch outcome, with these markers detected at higher frequency among individuals with inflammatory complications compared to those with healthy pouches\(^{67,68-70}\). Other studies have shown no relationship\(^{71}\). An additional study demonstrated that ASCA was associated with the development of fistulas following surgery\(^{72}\). Whether these markers were present prior to IPAA, and thus could be used as predictors of post-surgical outcome is unknown. However the previously described stability of markers over time suggests that they may have a use in stratifying patient risk prior to colectomy. As well, a prospective study demonstrated that pre-colectomy levels of pANCA, and anti-CBir1 were associated with an increased risk of acute and chronic pouchitis development following IPAA\(^{67}\). The relationship between pouch outcome and serological markers is described in greater detail in Chapter 2.

1.2.4 The role of anti-microbial antibodies in inflammatory disease pathogenesis

Increased prevalence of several of the anti-microbial antibody markers (mainly ASCA and pANCA) have been detected in numerous inflammatory conditions, including IBD, ankylosing spondylitis\(^{73}\), chronic granulomatous disease\(^{74}\), and celiac disease\(^{75}\). In each case, the physiological processes which contribute to the specificity of antibodies to normally commensal organisms are unknown. While there is no evidence that these antibodies themselves have a pathogenic role, or are even detectable in the bowel mucosa\(^{76}\), their high prevalence among individuals with inflammatory diseases suggests that they may be byproducts of disease processes. Further, given that these markers are detected in numerous diseases, they may provide additional evidence that these diseases proceed via similar mechanisms.
The pathophysiological relevance of anti-microbial antibodies in IBD, and more generally in autoinflammatory diseases, is unclear. On the one hand it is possible that the antibodies detected in CD and UC, which target specific microbial surface antigens, are indicative of immunological recognition of a pathogenic organism which is of importance in diseases pathogenesis. In this case, marker specificity for flagellin (CBir1) and *E. coli* surface porin (OmpC) or the unknown protein detected from *Pseudomonas fluorescens* (I2) would suggest that these organisms may have a specific role in IBD. As will be discussed in section 1.4.8, specific strains of *E. coli* have indeed been detected more commonly in tissue samples from ileal CD, although have not been detected among all patients with disease. However, no specific correlations between anti-CBir1 and any organism, or anti-OmpC and *E. coli* have been documented. *P. fluorescens* has not been shown to associate with CD, although some evidence suggests that the *Pseudomonas* genus may be detected more frequently among children with ileal disease. Additionally, certain strains of *P. fluorescens* have been shown to increase epithelial permeability in cell culture. However, little evidence can be found supporting the hypothesis that a single organism is responsible for IBD, even among subsets of the population.

An alternative hypothesis is that the presence of anti-microbial antibodies in the serum of patients with IBD may be indicative of a more general breakdown of intestinal barrier function or of the innate immune response, resulting in an increase in the detection of microbes by the adaptive immune system. This in turn has been speculated to lead to increased production of antibodies targeting microbial motifs from normally commensal organisms. Observations that individuals with chronic granulomatous disease, a condition where failure to produce reactive oxygen species (an important innate immune process) leads to reduced ability to clear infections, also express these markers, would seem to suggest that innate immune function is important. Further, there is evidence that individuals with IBD have additional anti-microbial antibodies in their blood, with specificities for numerous different microbial antigens. In fact, enhanced diagnostic utility of antibodies against a mixture of bacterial antigens obtained from a preparation of the surface antigens of a single species (*Bacteroides vulgatus*), or complex community, compared to the individual markers currently tested, has been demonstrated. This suggests that the current evaluation of antibodies specific for certain epitopes may be missing important information. These observations are in keeping with a more
generalized breakdown in intestinal barrier function leading to increased antibody
detection. One study demonstrated a marginal correlation between increasing ASCA IgG
and intestinal permeability\textsuperscript{81}. However, few studies have definitively demonstrated that
increased gut permeability is associated with antibody titres or positivity.

It is also possible that, rather than a general breakdown in overall barrier function,
there is a loss of specific innate immune functions which typically prevent certain
antigens from eliciting adaptive immune processes. ASCA positive patients, for example,
have decreased expression of mannan-binding lectin (MBL)\textsuperscript{82}. This molecule acts as a
pattern-recognition receptor directed against oligomannan, (the epitope detected by
ASCA, which is also found on the cell surface of numerous other microorganisms). Upon
the binding of MBL to a microorganism, the complement system is activated, leading to
clearance of the targeted organism\textsuperscript{82}. This would suggest that specific host alterations are
responsible for the production of anti-microbial antibodies, and that these alterations may
also play a role in disease pathogenesis. A further discussion of the relationship between
host factors (genetics) and anti-microbial antibody positivity is included in section 1.3.5.

As with many aspects of IBD, determining whether a breakdown in barrier
function precedes inflammation, or is merely a consequence of macroscopic or
microscopic disease, is difficult. Inflammatory processes themselves result in a less
cohesive intestinal barrier and could, therefore, result in increased access of commensal
organisms to adaptive immune processes. Yet studies which have supported the
prognostic capabilities of antibody markers suggest that anti-microbial markers can be
detected in the serum of patients preceding the onset of inflammation\textsuperscript{66}. This suggests
that alterations in barrier function (should they occur), which lead to anti-microbial
antibody generation, precede the onset of IBD.

Despite the strong evidence regarding the diagnostic utility of these markers in
IBD, on their own they leave many unanswered questions regarding mechanisms of
disease pathogenesis. To gain a further understanding of disease mechanisms, and how
these serologic markers fit into a model of disease etiology, it is necessary to further
explore additional pathways, including host genetic and microbial factors.
1.3 Genetics

With the sequencing of the human genome ‘completed’ in 2001, it became clear that a great deal more work was required before a genetic roadmap of human disease was available. Early drafts of the human genome provided evidence of previously unrecognized variability in the genetic landscape, with some regions characterized by high gene density, and others more sparsely populated. Further, the number of coding genes was found to be lower than expected, with around 30-40K protein coding genes. On the other hand, these fewer genes are more complex than those described in other organisms, with greater potential for alternate splicing. While not providing a cure for all human genetic disease, the complete sequencing of the human genome provided many useful tools for better understanding the genetic variability between individuals and groups, and for evaluating the genetic contribution to many disorders. Subsequent projects such as the HapMap and 1000 Genomes Project, which both seek to discover, genotype and provide accurate haplotype information on human DNA polymorphisms in multiple populations, shed new light on human genetic diversity both within and between ethnic groups. These tools can be used to evaluate the complex architecture of genetically mediated diseases, and have identified millions of single nucleotide polymorphisms (SNP)s, allowing quantification of genetic variability across populations and phenotypes.

1.3.1 Analyzing the human genome

The role of genetic factors in health and disease is complex. Several diseases including sickle cell anemia and cystic fibrosis for example, result from disease-causing mutations in a single gene (monogenic). Yet other diseases, including diabetes, obesity and IBD demonstrate complex heritability, with numerous loci modulating disease risk. Indeed, while genetic variability likely impacts susceptibility to most human illness, with effects ranging from those seen in monogenic disorders, to the modulation of individual susceptibility towards infectious agents (ie. decreased susceptibility to HIV infection among CCL3L1 high copy number carriers), few of these interactions have been fully elucidated. In the case of complex diseases, incomplete penetrance (differential expression of a genetic trait despite genotype), polygenicity (multiple genes contributing...
to a phenotype) and differential epigenetic regulation (genomic modification not resulting from direct nucleotide changes) complicate genetic analysis as each may be responsible for alterations in phenotype. Genetic heterogeneity, where multiple genes may cause similar phenotypes, and copy number variation, where variation in the numbers of genes or gene segments may lead to alterations in phenotype, also complicate genetic analyses. Additionally, phenotype may be modified by a host of environmental factors which can be difficult to accurately measure.

Despite these challenges, several approaches have been used to evaluate the genetic contribution to specific diseases. Earlier genetic studies used existing knowledge of either a disease or basic cellular process, to identify candidate genes and then attempted to determine whether alterations to that gene were associated with disease pathogenesis. Such studies provided useful information on some disorders, but were small in scale and less effective for evaluating genetics in complex, multifactorial diseases where little is known regarding etiology. To address this limitation, positional cloning and linkage analysis were developed and have been used to identify chromosomal regions of interest which may contain loci involved in disease. Such studies allowed associations to be detected independently of functional knowledge.

Early linkage studies used microsatellite markers, composed of tandemly repeated nucleotides with different alleles containing altered numbers of repeats, regularly interspersed throughout the genome, to detect regions of interest. Recent studies more commonly make use of SNP markers, with substitutions (nucleotide switch ie. from A-G) the most common form of polymorphism measured in genotyping studies, and insertions (addition of an extra nucleotide into the genetic sequence) and deletions (removal of a nucleotide from the wildtype sequence) less common. These studies assume that there will be linkage between markers which are located near disease polymorphisms. Linkage between genetic loci results when the physical distance between markers is insufficient to allow regular crossing over to occur during meiosis. This leads to physically close loci being more commonly inherited together than are unlinked loci. In order to identify regions of interest, initial studies used transmission disequilibrium testing (TDT) to assess how often a specific marker allele is passed from a heterozygote parent to an affected offspring. Early studies made use of this technique in well characterized families
with both affected and unaffected members. This method is very powerful, and unlike the
case control studies described below, is reasonably robust to population stratification\textsuperscript{93}. Once a region of interest is found to associate with disease, increased resolution of that region can be obtained using more closely spaced polymorphisms\textsuperscript{91}. These approaches were useful in early genetic studies, but were both cost and labor intensive, and required a large number of family groups to provide adequate statistical power. Furthermore, changes with smaller effect sizes which are typical of complex diseases often went undetected. In order to ameliorate these issues, larger numbers of markers are required, as are increased sample sizes to improve power.

Current approaches to identifying polymorphisms or loci of interest in complex disease involve the use of case-control study designs with high-throughput technologies to specifically assess regions of interest, and genome-wide association (GWA) studies (some more rare diseases may also still use family based studies with transmission disequilibrium testing). Recent advances in genotyping technology has allowed the large scale testing of genetic variation among increased numbers of individuals with complex diseases. For such studies, allele frequencies of SNPs are measured in large cohorts (1000s) of both affected individuals and unrelated, healthy controls. When a marker is shown to have an altered allele frequency among cases compared to controls, that marker and the surrounding region are candidates for disease association\textsuperscript{94}. In order to achieve the most informative results, accurate phenotyping of populations is essential, and both disease and healthy control groups must be carefully selected to ensure that they are ethnically matched and to minimize population stratification. Healthy controls must be well characterized as well to ensure that they themselves are not at an increased risk of possessing risk factors for the trait in question (ie. having affected family members). With an appropriately large sample size, this approach allows identification of markers which confer smaller amounts of risk, and is thus useful for evaluating many complex traits.

1.3.2 SNP genotyping: platforms and analysis

Several genotyping platforms have been developed which allow the high throughput identification of loci of interest for genetic association studies. These can be
used for both candidate-SNP based analyses and larger-scale GWA studies. The application of large-scale genetic analyses to the study of complex diseases, and the subsequent requirement for increased accuracy has contributed to enhancing data quality, with the accuracy of SNP calls estimated to be greater than 98%\(^87\). Several different companies provide technologies for SNP-based genotyping (Table 2).
<table>
<thead>
<tr>
<th>Company</th>
<th>Technology</th>
<th>Availability of custom arrays</th>
<th>GWAS arrays available</th>
<th>Features</th>
</tr>
</thead>
</table>
| **Illumina**     | BeadArray  | YES, including Immunochip      | YES                   | - up to 4.5 million SNP coverage (OMNI quad beadchip)  
|                  |            |                                |                       | - can add additional 500K custom markers |
| **Affymetrix**   | GeneChip   | YES                            | YES                   | - up to 1.8 million markers  
|                  |            |                                |                       | 906 K SNPs, 946K CNVs (Genome wide human array 6.0) |
| **Sequenom**     | iPLEX      | YES                            | NO                    | - Custom SNP analysis of up to 40 SNPs per reaction |
| **Applied Biosystems** | TaqMan | YES                            | NO                    | - Custom SNP analysis with up to 4.5 million pre-designed SNP assays |

Table 2: Summary of the different technologies available for genotyping. This list is not intended to be exhaustive, but to provide a broad outline of the technologies available for SNP-based genotyping.

While scientists are able to select any number of candidate SNPs for analysis using custom platforms, GWA chips include set numbers of markers interspersed throughout the genome in an unbiased fashion. Some areas are poorly represented (ie. centromeres), however outside of these regions, markers are spaced regularly with varying levels of coverage depending on the technology. SNP markers are selected from either HapMap\(^{96}\) or the 1000 Genomes Project\(^{97}\) datasets and must be detectable in a sufficient proportion of the population (detection in \(\geq 1\%\) of the population is considered the standard definition of a polymorphism) to be measured in experiments. Current GWA
chips can provide information on up to 4.5 million SNP markers (Illumina), thus providing a huge amount of information on samples, relatively rapidly. There are, however, advantages to using a more targeted approach compared to GWA studies, as custom assays may reduce costs in the event where the association of candidate SNPs with disease is being measured or specific hypotheses tested. Custom assays may also allow scientists to explore regions of interest with increased resolution beyond what can typically be achieved through standard GWA chips.

Following genotyping, quality control filters are applied to ensure that only high quality results are included in analyses. Samples in which fewer than 95% of the SNPs analyzed produce results are typically excluded as inclusion of samples falling below this threshold results in higher error rates. SNPs with an experimental minor allele frequency (MAF) below 0.05 are also typically excluded. However, use of this threshold prevents rare variants from being analyzed98. Exclusion of such variants could have important ramifications for experimental results, as rare variants may, in some cases, play a role in pathogenesis. SNPs are also tested for agreement with the Hardy-Weinberg principle. SNPs showing a significant (p<0.001) deviation from equilibrium are excluded as this is indicative of population heterogeneity, non-random mating (unlikely in human studies) or genotyping problems93. Additional quality assurance measures can include performing a sex check to ensure that genotype information matches that obtained through patient phenotyping, and genotyping several replicate samples to determine sequencing error rates99.

Following quality trimming procedures, SNPs must be tested for associations with phenotypic outcome groups. The most basic association test involves comparing MAFs between phenotypes (ie. cases versus controls). Using a basic chi-square (parametric) or Fisher's exact (exact) test, association can be assessed and probability quantified to provide evidence as to whether a particular allele is associated with an increased or decreased risk of a phenotype of interest. The role of genetic factors in influencing quantitative traits (ie. height) can also be determined using a Wald test. However, these methods evaluate only differences in MAFs between outcome groups, which have little biological or functional relevance. More complex models which take into account the heritance pattern of alleles, and have the potential to adjust for additional covariates in the analysis, include logistic and linear regression with dominant, recessive, co-dominant or
additive inheritance models. A dominant model assumes that an allele will have an effect on phenotype regardless of whether an individual is homozygous or heterozygous for it. Recessive models imply just the opposite, with an allele's effect only observed when an individual is homozygous for that allele\textsuperscript{93}. Both of these models are especially useful for detecting associations when the MAF is very low. An additive model, on the other hand, assumes that the heterozygous state has an increased risk of a phenotype compared to wildtype, and that the homozygous minor allele has two times this risk\textsuperscript{93}. This closely approximates a co-dominant model, which assumes that the heterozygous state will be intermediate between the two homozygous forms, although without specifying that the relationship is directly additive\textsuperscript{93} (Table 3). These models all have the benefit of increasing power while also enhancing the biological relevance of results. In practice, while the co-dominant model provides decent power for detecting true associations and likely best represents the biological scenario of most complex disease genetics, slightly more power is obtained by concurrently testing the dominant, recessive and additive models and selecting the best p-value, followed by application of appropriate multiple testing correction\textsuperscript{100}.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Dominant</th>
<th>Recessive</th>
<th>Additive</th>
<th>Co-dominant</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AB</td>
<td>1</td>
<td>0</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>BB</td>
<td>1</td>
<td>1</td>
<td>2n</td>
<td>&gt;n</td>
</tr>
</tbody>
</table>

Table 3: Sample coding for genetic models typically used in genetic association analyses.

As with any analysis in which multiple, independent hypotheses are being tested, appropriate multiple testing correction must be carried out. This is especially critical in GWA studies because of the large number of SNPs being scrutinized, but is also important in cases where multiple candidate genes are being tested. Assuming a traditional p-value cutoff of 0.05 is used as an indicator of significance, with increasing numbers of tests, the probability of obtaining a p-value reaching this threshold purely by chance increases with the number of tests performed. To correct for this, one can apply a more stringent significance threshold, or can attempt to correct probabilities based on the number of tests. Multiplying all p-values by the total number of tests (Bonferroni
correction) provides a correction which removes most false positive results but likely at the expense of inflating type II errors. Additionally, given that some SNPs may be in linkage disequilibrium and therefore not truly independent, this correction may lead to the needless rejection of true associations\textsuperscript{93}. Despite this limitation, Bonferroni corrections have become standard of practice in many GWA studies. Application of this technique in conjunction with test reduction strategies such as selecting or prioritizing specific SNPs, conducting analyses on haplotype blocks rather than individual SNPs, or performing a two-stage analysis, is most effective.

Alternatives to Bonferroni include Sidak and Holmes corrections which follow many similar assumptions as Bonferroni, but are less stringent\textsuperscript{101}. These methods are rarely used in genetics research. The false-discovery rate (FDR) correction also assumes that tests are independent, and additionally requires all loci to conform to a similar distribution. FDR is less stringent than the methods previously described and attempts to identify important effects from the many 'significant' results detected in multiple experiments, by taking into account the number of hypotheses tested and the level of significance of each\textsuperscript{102}. Yet another approach is to use permutation testing to establish a p-value cutoff level which is appropriately stringent to achieve optimal levels of type I and II error\textsuperscript{101}.

Despite best efforts to reduce the likelihood of false positive results, mathematical correction is not always a sufficient guarantee of high quality results as these corrections do not take into account subtle population effects. Replication of findings in independently accrued populations is, therefore, considered the gold standard for genetic association analysis and has the ability to substantially reduce Type I error. Results which are replicated in multiple populations are considered true associations, while those which do not replicate are called into question, and may demand further analyses to determine whether they are true associations\textsuperscript{101}.

While these methods have been used to confirm that genetics do indeed play a role in many human diseases, the lack of functional information available which would provide valuable insight into disease mechanisms, especially in regards to disease pathogenesis, is lacking in many cases. For some SNPs, the marker itself may confer disease risk. In many others, genetic associations with specific loci have been found in chromosomal regions with genes of known or hypothesized function which could
reasonably be involved in disease. In others, however, marker SNPs are located in either
gene deserts, or regions with high gene density, producing none or numerous candidate
genes for disease association. This ambiguity highlights the need for further study beyond
genetic association studies.

1.3.3 - Genetic Associations with IBD

Several lines of evidence suggest that IBD is at least partially genetically
mediated: monozygotic twin concordance rates for IBD are significantly higher than are rates among dizygotic twins\textsuperscript{103}; individuals with a family history of IBD are at a significantly higher risk of developing the disease themselves\textsuperscript{7}; and the preponderance of disease among certain ethnic groups, most notably individuals of Ashkenazi Jewish heritage\textsuperscript{6} are all indirect evidence of a genetic role in disease pathogenesis. As such, a great deal of study has gone into evaluating different genetic markers in the context of IBD, first with small scale and candidate gene studies, and more recently with larger cohorts and GWA studies.

To date more than 163 SNPs have been associated with IBD, some specific to CD (18%) or UC (14%), with those remaining conferring risk or protection from both disorders\textsuperscript{104}. Of the loci associated exclusively with CD or UC, many show similar directionality between both disorders. Only SNPs in two genes (\textit{NOD2} and \textit{PTPN22}) have been associated with an increased risk of CD and protection against UC (although in the case of \textit{NOD2}, not reaching genome-wide significance levels in UC). Many of the SNPs associated with IBD are located in or around genes involved in innate or adaptive immune processes, suggesting the importance of host immune factors in disease pathogenesis. Indeed, among the most common IBD-associated SNPs are those near genes involved in biological processes including innate immune function, lymphocyte activation, maintenance of barrier function, and production of reactive oxygen species (ROS)\textsuperscript{105}. Furthermore, many of the SNPs associated with IBD are involved in other complex diseases or have been demonstrated to have a modulating effect on the pathogenesis of bacterially mediated diseases such as leprosy, or susceptibility to other mycobacterial diseases\textsuperscript{104}.
The most striking genetic association with IBD is that of the \textit{NOD2}\textsubscript{insC} variant (rs2066847) with an increased risk of ileal CD\textsuperscript{106, 107}. Despite its low MAF, this finding has been replicated in many experiments and in different patient populations. Furthermore, several additional SNPs in this gene have also been associated with CD\textsuperscript{108}. This variant is highly associated with risk of fibrostenotic and fistulizing phenotypes, ileal disease location, and requirement for surgery among both adult and pediatric cohorts, yet is protective against large bowel disease\textsuperscript{109, 110}. NOD2 signals downstream effectors after recognition of bacterial muramyl dipeptide (MDP), and specific viral motifs including single stranded RNA\textsuperscript{111, 112}. The role of variant forms of \textit{NOD2} in IBD pathogenesis has been well studied and suggests several possible mechanisms through which this loci may contribute to disease: variants of this protein may have a reduced ability to detect pathogens or to mount an appropriate innate immune response through NFkB induced production of pro-inflammatory molecules, resulting in adaptive immune processes becoming activated against normally non-pathogenic organisms\textsuperscript{113}. \textit{NOD2} has also been implicated in immune regulatory pathways, with an important role in establishing immune tolerance\textsuperscript{105}. Additionally, \textit{NOD2} variants may have decreased ability to activate autophagy pathways in response to infection. Along with \textit{NOD2}, \textit{ATG16L1} and \textit{IRGM} are components of the autophagy pathway which is important in bacterial recognition and processing, regulation of cell trafficking, and activation of both innate and adaptive immune pathways\textsuperscript{114}. Variants in these genes are also highly associated with IBD, \textit{ATG16L1} (rs12994997) exclusively with CD, and \textit{IRGM} (rs11741861) with both CD and UC, further supporting this pathway’s role in disease pathogenesis\textsuperscript{104}.

Several polymorphisms in \textit{IL23R}, have been associated with IBD both CD and UC, with different SNPs associated with an increased, and decreased risk of CD\textsuperscript{104, 115}. The protein encoded by this gene, as well as those from several other IBD-associated genes including \textit{TNFSF15}, \textit{STAT3}, \textit{IL12B}, \textit{CCR6} and \textit{JAK2} are important components of pathways involved in activation and differentiation of T-helper 17 (T\textsubscript{H17}) cells. This cell group is important in inflammatory processes, and inhibits development of regulatory T-cells (T\textsubscript{reg})\textsuperscript{116}. The minor allele at rs11209026 in \textit{IL23R}, which is protective against IBD development and results in a missense mutation in the protein, has been shown to decrease populations of IL23 responsive cells, and pro-inflammatory cytokine
Of further interest is the observation made in one study showing a moderate (though not significant after multiple testing correction) decrease in risk for the development of colorectal cancer among individuals who were carriers of \textit{IL23R} polymorphisms\textsuperscript{118}.

\textit{IL10} cytokine and receptor mutants are rare, however can lead to disease phenotypes in the offspring of consanguineously mating individuals. Children who are homozygous for these mutations experience a severe form of infantile onset IBD without seemingly requiring any environmental stimulus. Such mutations which result in a loss of function of \textit{IL10} or \textit{IL10} receptor, lead to a subsequent inability to produce inhibitors of proinflammatory cytokines. Successful treatment of individuals with this genotype has relied on bone marrow transplant\textsuperscript{113, 119, 120}.

Several polymorphisms have also been exclusively associated with UC. Among these are rs10797432 in \textit{TNFRSF14} and a variant at 7p22 (rs798502) believed to implicate \textit{GNA12}\textsuperscript{121 104}. Less is known about the pathological roles of the genes associated with these variants. However, \textit{TNFRSF14} encodes a member of the TNF superfamily with a role in innate immunity, and which has been shown to be an important inhibitor of inflammation in mouse models\textsuperscript{121}. \textit{GNA12} is involved in the assembly of tight junctions for intestinal epithelial cells, suggesting the importance of barrier function in UC pathogenesis\textsuperscript{122}. Additional SNPs which have been evaluated in IBD are included in Appendix Chapter 2, Supplementary Table 1.

Many of the variant SNPs associated with IBD are in non-coding regions, or have not been associated with a candidate gene. These SNPs are typically thought to be in linkage disequilibrium with other genetic variants which are the true contributors to pathogenesis. However, it is also possible that such apparently 'non-coding' SNPs may be in regions of transcriptional regulation. rs10065172, another variant near \textit{IRGM} for example, leads to altered binding of several regulatory miRNAs, with the risk allele leading to increased levels of IRGM protein detected in cells\textsuperscript{123}. Alternate splicing may also account for the function of some of these polymorphisms. A rare protective variant in an intron of \textit{CARD9}, for example, has been shown to result in a truncated mRNA transcript caused by a skipped exon. This variant protein is less functional than the wildtype\textsuperscript{108}. Furthermore, many of the IBD-associated SNPs have been associated with
altered expression levels of either nearby (cis) or genomically distant (trans) genes. These expression quantitative trait loci (eQTL)s, are described in a recent paper by Kabakchiev et al\textsuperscript{124}.

The majority of loci which have been associated with IBD have also been associated with additional comorbidities\textsuperscript{104}. A large number of IBD-associated loci are also involved in primary immunodeficiencies, ankylosing spondylitis and psoriasis among others\textsuperscript{125 104}. rs11209026 in \textit{IL23R}, for example, is protective against the development of both psoriasis and IBD\textsuperscript{126}. Several UC susceptibility loci including those near genes \textit{L2}, \textit{REL} and \textit{CARD9} are also associated with PSC, a liver condition which is much more common among individuals with UC\textsuperscript{127}. While most of the associations between polymorphisms and multiple disorders are similar in direction, a notable exception to this pattern is observed in rs2476601 in \textit{PTPN22}. While the minor allele of this variant is a risk factor for development of Type I diabetes, rhumatoid arthritis and vitiligo, the same allele is protective against the development of CD\textsuperscript{108}. As studies evaluating IBD genetics have been some of the largest, it is likely with the generation of more data through larger studies in other complex diseases that the number of disease overlapping loci will increase. Such data suggests shared pathways which are likely important in the etiology of both disorders.

Despite the work done in investigating the role of genetics in IBD, the 163 disease-associated loci explain only approximately 14\% of disease variance in CD, and 8\% in UC\textsuperscript{104}. Additional alterations in gene expression, epigenetic regulation, rare genetic variants or environmental factors may help to account for this ‘missing variability’. The SNPs described so far have been ‘common’ variants, all with minor allele frequencies above 1\%. However, a small, exploratory analysis examining patients with severe CD and UC, found that individuals possessing no known IBD polymorphisms, were often found to have rare, and in some cases previously undescribed variants in exomes of IBD-associated genes\textsuperscript{128}. Additionally, eQTL analysis has demonstrated that the common IBD-associated SNPs result in transcriptomic changes, with cis or trans gene products either up or down-regulated, suggesting that alterations in gene expression play a role in pathogenesis\textsuperscript{124}. Host environment also likely plays a strong role in pathogenesis, but challenges associated with investigating such risk factors have prevented significant advancement in this field. However, it has been suggested that the reason that genetic
studies to date have failed to explain greater amounts of heritability, is that IBD is not in fact a genetic disease. Instead, in a manner similar to peptic ulcer disease, which was also once thought to be genetically mediated\textsuperscript{129}, IBD is actually bacterial in nature, with a single organism (analogous to \textit{H. pylori} in the case of peptic ulcers\textsuperscript{130}) responsible for disease processes. This hypothesis remains controversial, and has not been supported by experimental evidence or clinical trials.

1.3.4 IBD genetic association studies in non-Caucasian populations

The majority of IBD-associated genetic polymorphisms have been studied primarily in Caucasian populations. Studies examining the role of these genetic factors in individuals of Ashkenazi Jewish heritage demonstrate that several of the strongest IBD-associated polymorphisms in Caucasians are also associated with IBD among this group\textsuperscript{131}. However, evidence for a role of these specific polymorphisms among individuals of other ethnic groups is sparse. Healthy populations of individuals classified as North American First Nations, for example, have many of the common IBD-associated polymorphisms detected at differential rates compared to those of Caucasian ancestry. For example, \textit{ATG16L1} and \textit{NOD2} polymorphisms are detected less frequently than they are among Caucasians, and risk variants in \textit{IL23R}, \textit{IL12B}, and \textit{TNFSF15} are detected more commonly\textsuperscript{132}. These differences may explain the reduced prevalence of IBD among individuals of this ethnicity, although raise puzzling questions when considered alongside evidence that other autoimmune conditions including diabetes and arthritis are more common among this group\textsuperscript{133, 134}.

Although IBD prevalence is lower in Asian countries than is typically observed in the West, a recent rise in the number of cases detected in these regions has been observed\textsuperscript{135}. Unlike Caucasian populations, the dominant IBD-associated polymorphisms in \textit{NOD2} are rare among those of Asian descent, and have not been associated with IBD\textsuperscript{136-138}. Furthermore, neither of the strongly IBD-associated SNPs in \textit{ATG16L1} or \textit{IRGM}, have shown disease associations among Asians, although several novel polymorphisms in these genes are associated with IBD among individuals of this ethnicity\textsuperscript{139, 140}. Alternative \textit{NOD2} variants can also be detected among these populations, and like the variants associated with disease in Caucasians, are typically
associated with severe, early onset CD. Furthermore, variants in this gene are also associated with an increased risk of infectious diseases including tuberculosis and leprosy\textsuperscript{138, 141}. The IL23R locus has shown mixed results, but the lack of replication of associations among Asian cohorts, compared to the strong associations seen among Caucasians suggests that this loci does not have the same role in IBD pathogenesis among these groups\textsuperscript{140, 142, 143}. The HLA region also shows similar associations with UC among Caucasian and Asian populations. TNFSF15, on the other hand is much more strongly associated with IBD among Asian groups than their Caucasian counterparts\textsuperscript{138}.

The cohorts used for many of the studies investigating IBD genetics among non-Caucasians are typically much smaller than those used for studies among Caucasians. However, findings in different ethnic groups suggest that many of the polymorphisms that have been well associated with disease pathogenesis among Caucasians, are rare among other ethnic groups in which IBD prevalence is on the rise. Given that too little time has passed for genetic drift to explain the recent spike in IBD among non-Western populations, one may speculate that recent environmental or dietary changes may be responsible for increased disease risk. Additional study of such factors among individuals of varied ethnicities will be of benefit in answering some of the fundamental questions regarding IBD pathogenesis.

1.3.5 Genetics in IPAA

As observed with IBD, individuals with inflammatory complications following IPAA (both pouchitis and CDL) have demonstrated increased prevalence of genetic risk factors. Unlike the large-scale IBD genetic association studies, those investigating genetic susceptibility to pouch inflammatory complications are typically characterized by low sample numbers and unclear phenotyping. However, several studies implicate genetic markers in pouch inflammatory pathologies, with specific markers discussed in greater detail in Chapter 3.

1.3.6 Evidence for genetic associations with anti-microbial antibody positivity
Observations demonstrating differences in the carriage rates of ASCA and pANCA between individuals of different ethnicities suggest that positivity for these markers is a complex phenomenon which may be partially genetically mediated\textsuperscript{144, 145}. First-degree relatives of patients with IBD are also more likely to be positive for anti-microbial markers than are unrelated healthy controls, further supporting this hypothesis. Several IBD-associated genetic polymorphisms, including those in \textit{NOD2}, \textit{ATG16L1}, and \textit{TLR5}, have been associated with antibody positivity\textsuperscript{146}. The \textit{NOD2} and \textit{ATG16L1} variants are associated with an increased number of antibody markers detected in serum, and have also both been associated with increased titres and positivity for ASCA, and in some studies, anti-OmpC\textsuperscript{147-150}. Additionally, anti-CBir1 is more prevalent in individuals with \textit{IRGM} polymorphisms, although this marker is detected at decreased levels among those with variant \textit{TLR5}\textsuperscript{146}. This marker is also detected at increased titres among individuals with at least one \textit{NOD2} polymorphism\textsuperscript{59}. The IBD-associated \textit{TNFSF15} polymorphism is associated with anti-OmpC positivity among individuals of Ashkenazi Jewish heritage \textsuperscript{151}. However, whether there is a causal relationship between these factors remains to be determined.

The clinical utility of combining such markers in managing and diagnosing patients with IBD has been a topic of much study. Recent work evaluating the diagnostic potential of these factors in IBD, or differentiating between CD and UC, suggest that combination panels, which include described serological markers and well-characterized gene polymorphisms, more accurately predict phenotype than do those measuring serological markers alone\textsuperscript{42, 152}. Panels which incorporate eight serological, four genetic and five inflammatory markers (ie. C-reactive protein) have been developed by Prometheus and are capable of diagnosing IBD with sensitivity of 63\% and specificity of around 90\%\textsuperscript{153}. The sub-classification (of patients with CD or UC) power of this panel has not yet been independently reviewed.

1.4 Bacteria

The reason for the dramatic recent rise in the prevalence of IBD and other autoimmune diseases through much of the developed world is unclear. For many years,
scientists postulated that an important factor is the increasingly 'clean' environments
typical of these areas\textsuperscript{154}. This so called 'hygiene hypothesis' suggests that exposure to
fewer bacteria, viruses or eukaryotic parasites, prevents the immune system from
developing properly during childhood and adolescence, leading to dysregulation and
disease. Thus, improvements in sanitation, increased use of antibiotics and vaccines, and
reduced exposure to pathogens, although exceedingly successful in improving public
health, may have had unpredictable consequences on the proper functioning of the
immune system. Evidence regarding the role of any single pathogenic organism in
preventing or causing autoimmune disease is lacking. It is possible that 'cleaner'
environments alter human microbial community composition as a whole, which in turn
prevents the development of host immune tolerance and homeostasis. However,
evaluating complex microbial communities in the context of the human host was, until
recently, too time consuming and expensive to be broadly undertaken. The application of
next-generation sequencing technologies, reduced sequencing cost, as well as an
increased availability of funding have greatly enhanced our knowledge of the structure of
human-associated microbial communities, and led to the development of better tools for
increasing the accuracy of information obtained from such experiments.

1.4.1 Phylogenetic organization and taxonomic classification of bacteria

Phylogenetics refers to the evaluation of the evolutionary relationship among
organisms, and is usually based on a combination of genetic and phenotypic information.
Taxonomy is informed by the phylogenetic relationships between groups, and refers to
the classification, identification and naming of organisms. Traditional measures of
organism relatedness and subsequent taxonomic assignments were based strictly on
phenotypic traits, however, the incorporation of genetic information into classification
schemes has resulted in a more accurate approximation of the relationships between
organisms\textsuperscript{155}. In some cases the increased availability of genetic data has led to confusion
surrounding specific organisms’ evolutionary relationships, with those once thought to be
closely related based on phenotype, shown to be less closely related based on DNA
sequence similarity – a problem specifically apparent in the analysis of microbes where
horizontal gene transfer (HGT) is quite common. Sequence-based approaches to
microbial analysis have become more widely used, leading to challenges in interpreting the results of sequencing experiments. Thus, where possible, both genetics and phenotype should be used to inform phylogenetic relationships and taxonomic assignment.

All living organisms are currently hierarchically classified into eight major levels, with domain being the most general, and species the most specific (Figure 1). Subgroupings are also often included, to provide additional information. There are 3 major domains (Archaea, Bacteria, Eubacteria), each of which have numerous sublevels, characterized by distinguishing features allowing organisms to be classified with their most similar cousins. To better illustrate the nature of these hierarchical systems one might consider our own species, *Homo sapiens*. A member of the genus *Homo*, the family Hominidae, the order Primate, class Mammalia, phylum Chordata, kingdom Anamalia, and domain Eukarya (Figure 1). Here it may be seen that analyses on any one of these levels could be conducted, each providing different amounts of information. While the most precise information is gained from the species-level assignment, there is also value in higher level analyses, especially in cases where, unlike our example of *Homo sapiens* above, species identification is not possible.
Accurate phylogenies and taxonomic assignments are critical to microbiome analysis and interpretation of results, especially for experiments which are conducted strictly based on DNA sequences. Our example of humans is a simplistic one, as the phylogenetic relationship between *Homo sapiens* and related species is extremely well defined. Bacteria, on the other hand, display more complex evolutionary relationships between one another, and new species and even phyla are still being discovered. Determining where novel species fit among known organisms and assigning them accurate taxonomic designations can be challenging. Further complicating bacterial phylogenetic analysis, is the fact that bacterial genomes are much more subject to variability than are those of eukaryotes, and the rapid doubling time characteristic of most prokaryotic cells allows populations to evolve quickly in response to stress. An example of this can be seen in the rapid development and spread of antibiotic resistance among bacteria exposed to a specific antibiotic\textsuperscript{156}.
Our inability to culture many recently identified organisms has led to increased emphasis being placed on genetic information. However, HGT between different species, and even more distantly related groups may confound many metagenomic analyses and present problems for accurately determining phylogenetic relationships. Transformation, transduction and conjugation all facilitate the transfer of DNA between different bacterial cells and can be used to share genes between distantly related organisms. Transferred genes conferring anti-microbial resistance or other functions which may confer benefit to the recipient, are more likely to be selected for and spread among a population. The major limiting factor to successful, stable and useful incorporation of a novel gene into the recipient cell is the ability of the recipient to utilize the donated gene, with more closely related organisms more likely to do so successfully. Locations with high bacterial population densities, such as that found in the human gut, are characterized by higher potential for HGT. Therefore, the plasticity in bacterial genomes caused by gene transfer must be taken into account when performing sequence based analyses of microbial community structure.

Although much of the bacterial genome is relatively plastic, certain genetic markers are more stable, and can therefore, be considered candidates for phylogenetic analysis. Most commonly used among these is the 16S rRNA gene, which is present in all Bacteria and Archaea although with different copy numbers in different species, ranging from one to as many as 15 per organism. The 16S rRNA gene encodes an essential component of the bacterial ribosome that is necessary for all protein synthesis, and as such, is comparatively well conserved. Further adding to this genes’ utility in phylogenetic analysis is the sequence itself, which is characterized by conserved regions suitable for use as annealing targets for PCR amplification of sequences from the majority of organisms present in a sample, flanked by regions of substantially higher variability which may be used in conjunction with curated 16S databases to identify organisms. While transfer of these units between organisms is believed to be relatively rare, it is possible for highly divergent species to utilize 16S genes transferred from remotely related species with only limited reduction in function observed among recipient cells. A summary of the advantages and disadvantages of 16S-based approaches are described in Table 4.
<table>
<thead>
<tr>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present in all bacteria and archaea</td>
<td>Present in multiple copy numbers through most organisms</td>
</tr>
<tr>
<td>Contains highly conserved regions suitable for universal primer design</td>
<td>Small number of organisms do not display as much conservation through these regions leading to primer bias</td>
</tr>
<tr>
<td>Contains regions of high variability suitable as unique identifiers</td>
<td>Regions of variability are often insufficient to provide species level resolution</td>
</tr>
<tr>
<td>Numerous well curated databases allowing sequence comparison and taxonomic assignment of organisms</td>
<td>Many databases contain sequences with errors</td>
</tr>
<tr>
<td>Well studied primer pairs available which are capable of amplifying most organisms with high specificity for bacteria</td>
<td>May lack specificity for certain bacterial groups and result in inaccurate estimations of community composition</td>
</tr>
</tbody>
</table>

Table 4: Advantages and disadvantages of using 16S rRNA gene-based analyses in construction of bacterial phylogenies and taxonomic assignment.

Alternative marker genes have been proposed including 23S rRNA, cpn60 and rpoB to address the limitations presented by 16S rRNA based analyses. 23S based sequencing tends to generate phylogenetic trees which are very similar to those produced by 16S based sequencing. These genes are also longer, and said to contain more phylogenetically informative stretches, while also containing conserved regions suitable for universal primer design\(^{160,161}\). *Cpn60* is more polymorphic than is the 16S rRNA gene, allowing increased resolution, in some cases to the species level\(^{162}\). *RpoB* is present in only a single copy number in bacteria, making it an attractive choice for more accurate microbial quantification\(^{163}\). The main limitation of using any of these markers is the lack of data available allowing them to be universally applicable for taxonomic analyses. While curated databases containing these sequences do exist, they are significantly less complete than are current 16S based ones such as Silva, Greengenes or the Ribosomal Database Project (RDP). As such, 16S rRNA-based sequencing remains the gold standard for sequence-based bacterial analyses.
1.4.2 Human microbiome analysis techniques

Perhaps surprisingly, bacteria exist only rarely in isolation, and are instead most commonly found in complex community assemblages or biofilms, in which numerous different organisms share a similar ecological niche\(^{164}\). In many cases, these organisms are co-dependent on one another, requiring metabolic support from additional members of the community for survival\(^ {165}\). Organisms residing in or upon higher order taxa display even more complex ecology, as the interaction between host and microbe requires a delicate balance to ensure that both symbionts are capable of optimal growth. The human microbiome is defined as the collection of organisms and their genomes inhabiting different anatomical locations both in and on the human host\(^ {166}\). The human gut alone is home to hundreds of trillions of microorganisms and contains more genetic information than that which exists in the human genome\(^ {167}\). In most cases, these organisms exist with the host in a symbiotic or mutualistic relationship: microbes help with the digestion of nutrients, prevent colonization of the host by pathogenic organisms, and aid in the proper development of both the intestinal epithelium and immune system\(^ {168}\), while the host provides nutrients and a suitable habitat for bacterial growth. However, recent studies have implicated a reduced diversity of microorganisms in the etiology of several chronic conditions including inflammatory bowel disease (IBD), diabetes and obesity\(^ {169-173}\).

Prior to the development and application of sequence-based molecular tools to microbiome analysis, culture-based methods were the norm. Traditional, culture based microbiology relies on the ability to cultivate viable organisms outside of their natural habitat, which can be difficult as many organisms which are well adapted to life in the human gut are not viable in different locations. Providing the necessary nutritional and atmospheric conditions to promote growth for all organisms is difficult, and in the past has resulted in underestimation of the complexity of the human gut microbial ecosystem. However, advantages of being able to directly culture organisms include the greater amount of information obtained regarding bacterial metabolism and growth requirements, and the potential future use of cultured strains in experimentally evaluating the interaction between microbes and host.

Early studies relying on traditional approaches to culture organisms significantly underestimated the complexity of the gut ecosystem in particular, in large part because of
the difficulties inherent in culturing many of the organisms occupying this niche\textsuperscript{174}. To solve this problem, culture-independent techniques have been applied to the analysis of complex microbial communities. Yet a surprising majority of human microbiome studies to date have focused on high level taxonomic assignments for analysis, providing information on whether specific phyla or families are associated with a particular habitat, phenotypic outcome, or environmental factor in the human host. For the most part, this is due to the limitations present in current technologies which have provided insufficient data to allow a more detailed understanding. However, applying a phylum level analysis to our human example above, humans would be classified together with 43,000 other species including tunicates, ascidiacea (aka sea squirts), and amphibians to name just a few. Thus, while some important information is gathered from this knowledge, it is clear that much ambiguity remains as well. As such, the advantage of increasing the resolution of microbiome studies is apparent, and the quest to obtain greater information from current technologies remains essential.

1.4.3 Culture-independent microbial community analysis

Recent sequence-based approaches to evaluating the microbiome have been developed to circumvent the limitations of culture and use bacterial DNA sequence as a proxy for estimation of organism identity, relative abundance and functional potential. However, the challenges and sources of error of sequence-based analyses are important to understand in order to accurately interpret results.

1.4.3.1 Sample procurement

The first step in analyzing the microbiome is to obtain samples from the site of interest. The type and number of samples to be obtained is dependent on the specific experimental question to be answered. In evaluating the gut microbiome, the most common samples used are stool, and biopsies obtained during patient endoscopy. While stool is commonly used, easy to get and provides a great deal of sample material, its microbial profile is substantially different from that obtained from tissue\textsuperscript{175}. Biopsies, on the other hand, are difficult to obtain and the microbiome may be altered by the
requirement that patients take laxatives prior to endoscopy\textsuperscript{176}. Furthermore, differences in micronutrient bioavailability or the presence of microscopic inflammation at the site of biopsy, may make it difficult to get an accurate picture of overall community structure. However, tissue samples have the benefit of providing a picture of the organisms present at sites immediately adjacent to the host and, therefore, most likely to interact with the immune system.

Due to the ubiquity of bacteria and bacterial DNA, contamination of all samples by environmental sequences is likely to occur regardless of measures taken to reduce these factors. Attempts should be made to minimize contamination, and equivalent sampling collection and handling procedures must be applied to all samples. Small changes in sample handling procedures profoundly influence observed community composition, with, for example, freezing stool samples altering the Bacteroidetes/Firmicutes ratio compared to samples which are never frozen\textsuperscript{177}.

1.4.3.2 Microbial DNA extraction from samples

Once samples have been obtained, total DNA is extracted directly, typically using a combination of mechanical and enzymatic disruption. Extraction method is highly dependent on the samples from which DNA is being extracted, with tissue samples requiring the use of slightly different protocols compared to samples from soil or stool. Furthermore, the relative efficiency of disrupting organisms is highly dependent on the structure of the bacterial cell\textsuperscript{178}. This means that obtaining an appropriate balance between an extraction techniques' harshness, which could result in shearing the DNA of easily lysed organisms (gram negatives for example), and mildness, which would prevent difficult to lyse organisms (gram positives, spores, mycobacteria) from having their DNA extracted, can have important implications for downstream community analysis, and may lead to alterations in apparent community composition\textsuperscript{179-181}.

While the advantages of optimizing extraction protocols to provide increasingly accurate views of bacterial community structure are obvious, no methods are currently capable of providing a truly unbiased DNA sample\textsuperscript{182, 183}. Extraction methods using a combination of bead beating for physical disruption and enzymatic digestion using
lysozyme, mutanolysin, lysostaphin, or combinations of these, appear to provide sequence information which most closely approximates community structure\textsuperscript{183}. Other studies have suggested that mechanical lysis steps, such as bead beating, actually decrease the overall amount of DNA obtained from samples\textsuperscript{184}. While no extraction protocol is yet capable of rendering perfect extraction efficiency across all cell types, ensuring methodological consistency between samples included in an experiment is critical for obtaining unbiased results. All extraction methods have demonstrated reproducible results when used to extract DNA from multiple identical samples (ie. samples obtained from the same stool sample)\textsuperscript{183}, suggesting that once an extraction method has been selected, it is important to extract all samples using the same protocol in order to ensure that different biases are not applied to samples included in the same experiment.

To provide even greater consistency between extraction batches, numerous commercial kits are available which can be modified through the addition of the above mentioned bead-beating steps and enzymatic lysis to generate reasonably accurate accounts of community structure. Such kits have the additional benefit of being subjected to stringent quality control procedures demanded of commercial manufacturers. Among these are kits provided by QIAGEN, MoBio, MP Biomedicals, and Sigma-Aldrich to name just a few. Most commonly used for human microbiome analyses are the QIAGEN DNA Mini, DNeasy and stool kits, and the MoBio PowerSoil isolation kit\textsuperscript{185}. While the former can be used to extract total DNA from tissues, the latter is primarily used in stool analyses. In each case, the presence of human and archaeal DNA can reduce the number of bacterial reads obtained during sequencing. However these can be easily detected and removed during downstream processing. Operator contamination is more difficult to manage and at present, can only be minimized by performing extractions in a clean environment, and by recording technician as well as kit lot numbers used for different extraction batches.

1.4.3.3 Overview of Sequence-based microbial analysis tools

Initial sequence-based microbial community analysis approaches involved construction of libraries, consisting of bacterial clones transformed with sequences of
interest from the community being studied\textsuperscript{186, 187}. Extracted DNA from each colony could then be sequenced using Sanger sequencing, and relative quantification of sequences could be inferred. While a significant advancement over culture-based approaches, this technique remains time consuming and laborious, and often does not provide the necessary coverage to accurately estimate the variability in the human microbiome.

DNA fingerprinting techniques, including terminal restriction fragment length polymorphism (T-RFLP), and single-strand conformation polymorphism (SSCP), have also been applied to microbial analysis in conjunction with normal or capillary electrophoresis, and while significantly faster than library construction, are incapable of providing adequate levels of resolution for characterization of complex communities\textsuperscript{174, 188, 189}. The utility of these techniques can be most easily observed in performing basic comparisons between organisms with well-characterized restriction patterns or 16S \textit{rRNA} secondary structure respectively\textsuperscript{190}. However, these methods dramatically under-represent community diversity and they do not provide quantitative data allowing estimation of the relative abundance of different organisms. In an interesting application of these techniques, Hiibel et al concurrently measured levels of the 16S \textit{rRNA} gene (DNA) and 16S \textit{rRNA} product to deduce a specific organisms’ cellular activity. In so doing, they were able to accurately predict the metabolic state of the organisms in question (ie. determine whether the organisms were alive and replicating, or inactive/dead)\textsuperscript{191}.

Sanger sequencing and quantitative (q) PCR, are accurate, able to provide long read lengths to achieve high degrees of sequence resolution, and relatively simple to use. In fact, qPCR is considered the gold standard for quantifying bacterial abundance\textsuperscript{192}. However, these methods are slow, and labour intensive and, in the case of qPCR, require the use of specifically designed and validated primers in order to provide taxonomically informative data. Thus, these methods have only limited potential for use in complex community analysis, especially in cases where little is known \textit{a priori}. Their widest application is in determining whether a specific organism is detected in a sample, and for confirming the results obtained through other, more high-throughput methodologies.

Additional indirect analysis techniques include fluorescence in situ hybridization (FISH) and the use of dot-blot or microarrays to assess whether particular sequences are present in a sample. Both of these techniques can be used for the detection, and relative
quantification of known sequences\textsuperscript{174}. In both cases, oligonucleotide probes targeting specific organisms are used to capture complementary sequences in a sample. The most commonly used microarray, PhyloChip (2nd generation), is capable of providing data on more than 8700 strains of bacteria and archaea\textsuperscript{193}. However, these methods are limited in that they require use of a specific probe in order to detect an organism, preventing novel microbes from being detected. Furthermore, cross-hybridization between sequences from non-target organisms may conflate results and, in the case of the PhyloChip, have been shown to lead to distortions in perceived community structure\textsuperscript{193}.

The use of next-generation sequencing circumvents many of these limitations. The two main technologies currently used for next-generation microbiome analysis are 454 pyrosequencing (Roche) and Illumina platforms. Several additional technologies from other companies are also available (Table 5), but are less widely used and as such will not be described in depth. Both 454 pyrosequencing and Illumina provide several platforms offering different levels of coverage (number of sequences in a sequencing run) and sequence lengths (base pair length of the sequence). While Illumina is able to provide more coverage at lower cost, 454 pyrosequencing is capable of generating longer sequences, often corresponding to an increase in taxonomic resolution (Table 5). Both Illumina and pyrosequencing-based technologies can be applied to either 16S \textit{rRNA} gene or metagenomic analyses. While the former focuses on determining which organisms are present in a sample, the latter can be used to provide functional information as well. However the added cost and (currently) limited resources available for downstream processing (ie genome assembly and functional interpretation) make metagenomic analyses significantly more resource intensive. As such, the focus of work discussed in this thesis relates to 16S-based approaches.
<table>
<thead>
<tr>
<th>Technology</th>
<th>Maximum Sequence Length</th>
<th>Sequences per run</th>
<th>Error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roche 454 pyrosequencing (GS FLX Titanium XL+)</td>
<td>&lt;1000bp</td>
<td>1,000,000</td>
<td>1.07-1.6% 194, 195</td>
</tr>
<tr>
<td>Illumina (MiSeq)</td>
<td>&lt;150bp paired end reads</td>
<td>1x10⁹</td>
<td>0.4-5% 194, 196</td>
</tr>
<tr>
<td>Applied Biosystems SOLiD</td>
<td>85</td>
<td>1.5x10⁹</td>
<td>0.01-0.5% 197, 198</td>
</tr>
<tr>
<td>Pacific BioSciences</td>
<td>&gt;2500 bp</td>
<td>3,000,000</td>
<td>13-15% 196, 199</td>
</tr>
<tr>
<td>Ion Torrent</td>
<td>&lt;250</td>
<td>11,000,000</td>
<td>0.5-1.8% 196, 200</td>
</tr>
</tbody>
</table>

Table 5: Summary of the next-generation sequencing technologies currently available 195. Listed in brackets are the newest technologies which are available, in the case of 454, the Titanium XL+ had not yet been adopted for use in 16S microbiome analyses at the time of writing this thesis 201.

1.4.3.4 A more detailed examination of next-generation sequencing and how it is applied to the analysis of the microbiome.

Current next-generation sequencing technologies are limited by the length of sequences they are able to provide, with ~400 bp being the upper limit of sequence lengths achievable through currently applied technologies. As such, it is not possible to utilize the entire length of the 16S *rRNA* gene (~1550 bp), and specific region(s) of the gene must be selected to target for analysis. The 16S *rRNA* gene is composed of nine regions of high variability, termed hypervariable regions, flanked by relatively conserved regions which can be used as binding sites for universal primers. Different regions have different biases, and levels of resolution, meaning that few comparisons can be made between experiments conducted by evaluating different sites (Figure 2). Despite the well-characterized universal nature of primers typically used in microbiome experiments, amplification biases remain, and are not consistent between conserved regions 202. Furthermore, specific groups such as the Bifidobacteria are consistently underrepresented.
regardless of the variable region investigated, usually because of universal primer mismatches\textsuperscript{203}. In considering which region to use, one must take into account the length of sequence possible through a given technology, and whether there are any specific organisms which are of interest which might be less well covered using a selected region. Commonly used segments include the V1-V3 and V4-V5 regions, both of which are able to provide genus level sequence resolution, with V1-V3 shown to be slightly more accurate\textsuperscript{204}.

![Regions of 16S rRNA molecule](image)

Figure 2: Regions of 16S rRNA molecule typically used for evaluating the human microbiome in 454 pyrosequencing or Illumina studies. \textsuperscript{1} indicates which phyla appear more abundant among communities in which this region is used for analysis\textsuperscript{194, 205}.

Regardless of the region selected for analysis, some human and archaeal sequences are also typically amplified along with bacterial sequences. Sequences of human origin tend to map primarily to the mitochondria. Although archaeal sequences are also produced in sequencing runs and may be of interest in certain experiments, since most primers used for human microbiome studies are not optimized for detection and analysis of this group of organisms, the quality of the data obtained is not sufficient for accurate analyses to be conducted\textsuperscript{206}.

Once a target region has been selected, well described primers\textsuperscript{194} are selected and used for amplification of the region of interest. In order to reduce the costs associated with sequencing, samples are multiplexed with the potential to include more than 96 samples together in the same run\textsuperscript{207}. Such multiplexing reduces the number of sequences obtained for each sample, however analysis of rarefaction curves can ensure that coverage remains adequate to appropriately characterize microbial communities. To prepare sequences for multiplexing and downstream analysis, 16S primers specific for the
region of interest and containing customized barcodes and adapter sequences for ligation to either 454 pyrosequencing beads or the Illumina flowcell are used. Barcode length is dependent on the degree of multiplexing desired, but are commonly approximately 10-base sequences, which are sufficiently unique compared to other barcodes to prevent sequencing errors leading to assignment of a sequence to an incorrect sample\textsuperscript{208}. These are sequenced along with the selected 16S region, and are subsequently used to accurately parse data.

Next-generation sequencing technologies each make use of different chemistries to provide sequence information. 454 pyrosequencing makes use of pyrophosphate chemistry for sequence resolution. A single query sequence is ligated to a bead and undergoes emulsion PCR to coat the bead with thousands of copies of that sequence. Beads are placed into a single well of a picotitre plate, and nucleotides are added sequentially (across all wells). Unincorporated nucleotides are removed prior to the addition of the next base. The pyrophosphate which is released when nucleotides are incorporated into the growing DNA strand, is measured using an ATP sulfurylase/luciferase reaction which generates light. The light is subsequently detected by a charge coupled device (ccd) camera and converted to a peak in a pyrogram. The height of the pyrogram peak is a function of the amount of light produced by the reaction\textsuperscript{209}. The sequential addition of bases allows determination of base calls, and the number of nucleotides incorporated is proportional to the intensity of the light output (up to 8bp).

Illumina uses a different method which involves deposition of sequences on a flowcell followed by isothermal bridge amplification to increase the sequence density at that specific location. Reverse strands are cleaved and washed away, leaving millions of copies of the original sequence for analysis. Sequencing is conducted base-by-base using fluorescently labelled nucleotides. Following incorporation and fluorescence reading, tags are removed and the next base may be added to the growing chain. Unlike pyrosequencing, paired end reads can be generated, in which sequencing is conducted from both ends of the DNA fragment\textsuperscript{201}. This can be used to create longer read lengths and to enhance the accuracy of measurements.
Sequencing error rates and common types are different between Illumina and 454 pyrosequencing, with Illumina sequencing more prone to mismatching, while 454 typically has higher rates of insertions and deletions, particularly around homopolymeric regions. Longer sequences are, on average, of higher quality than are shorter sequences, although may still contain erroneous calls. However, for both of these technologies, extended sequencing results in a dramatic rise in sequencing error rate with a higher number of errors occurring after 100 bases in Illumina and 400 bases in 454. Error rates for 454 tend to be spread across many sequences, making it difficult to remove all that are suspect without a corresponding reduction in sequence coverage. Given the run- and sequence-specific nature of error rates, in order to minimize propagation of errors through workflows, it is important to include standardized control sequences and error-corrected base callers in each run to allow estimation of the true error rate. Furthermore, downstream sequence processing which specifically targets and culls sequences with a higher probability of containing erroneous bases, is useful.

1.4.3.5 Sequence quality trimming, alignment and taxonomic assignment

In order to ensure the best quality sequencing information is included in analyses, and to prevent inaccurate sequence information from falsely inflating levels of diversity among samples, sequences are quality trimmed prior to taxonomic assignment and subsequent analysis. Typical methods of quality control are aimed at addressing the specific limitations of the sequencing technology being used. For example, given that shorter 454 sequences tend to have higher error rates, and extended homopolymers are a constant source of error, sequences which do not meet minimum length requirements (with the threshold used typically dependent on the technology used and choice of reagents), and which include extended homopolymeric stretches (typically 8 nucleotides) are removed from downstream processing. Sequences with homopolymers extending beyond 8 nucleotides have error rates above 0.01, as do sequences shorter than 200 nucleotides (error rates are defined as the total errors detected per base call). The presence of ambiguous bases, or primer mismatches are also indicative of lower quality sequences. More than two primer mismatches correspond to an error rate of 0.0056, while the presence of any ambiguous bases substantially increases error (above 0.01). The
expected alignment location of a sequence of interest can further be used to identify incorrect sequences. If an experiment is relying on primers specific to a given hypervariable region of 16S, yet preliminary database alignment suggests that a different region from that expected has a stronger alignment score, the quality of that sequence is called into question.

Rather than simply removing sequences with the features described above, alternative solutions include removing portions of sequences with overall low quality, thus reducing the amount of information lost at this stage\(^\text{211}\). Algorithms such as PyroNoise and DeNoiser have been developed which allow base calls to be corrected using the raw flowgram data\(^\text{212, 213}\). However, the increased computational intensity demanded by both of these approaches has prevented their widespread utility in large datasets.

An additional method to remove spurious sequences, proposed by Huse et al makes use of a pre-clustering step which identifies highly similar sequences and separates out those which are unique\(^\text{214}\). The logic behind this approach is that more abundant sequences, which cluster together based on a predetermined distance measure, are more likely to be real than those which are less abundant. Sequences which are rare and which fail to cluster into groups can then be removed. The simplest implementation of this algorithm is to perform pairwise comparisons between all sequences and determine the number of mismatches between them. Those with more than the defined cutoff (one mismatch is commonly used) will not be grouped together. This approach has been observed to decrease error rates to 0.0004, when applied in conjunction with the previously described steps\(^\text{202}\).

Providing more of a challenge for ensuring sequence quality, is the ubiquity of chimeric sequences throughout many sequencing projects. Chimeric sequences are an artefact of PCR amplification in which different parts of a sequence arise from different parent strands. Such sequences typically result from an incomplete sequence dissociating from its parent strand and acting as a primer for another, different sequence. This hybrid is then amplified leading to spurious groups. Because these sequences are not a reflection of poor sequencing quality, they are much more difficult to detect than are errors resulting from low quality reads. Chimeras are so ubiquitous in sequence-based analyses
that they can be found in many 16S databases, especially among organisms which have not yet been cultured. Chimeras are estimated to make up anywhere from 5-45% of sequences in a run and thus represent a huge source of error for next generation sequencing approaches\(^{215}\). Furthermore, 16S rRNA databases have been shown to contain a great many anomalies, with approximately 5% showing errors (RDP database release 9.22) prior to the widespread use of chimera detection algorithms\(^{216}\). However, through improvements in computer based algorithms for detecting chimeric sequences, and the application of standardized sequence processing pipelines which are capable of reducing the number of errors in data, improved accuracy in taxonomic assignment is possible.

Several algorithms have been developed which attempt to detect and remove chimeric sequences from datasets. Each can be used in conjunction with well-curated databases in which it is known that no chimeric sequences are found, or run independently of reference databases, with sequences in a run compared to one another. While database approaches have some advantages, especially in terms of decreased computer power and resources, they are forced to rely on often incomplete and inaccurate datasets. Because the highest quality sequences are available from previously cultured organisms, these databases would not be useful for samples in which many of the organisms have not previously been cultured. Further, in applying a more broad collection of database sequences, one may inadvertently incorporate lower quality reference reads which would not add to the accuracy of the analysis. As such, where possible it is advantageous to make use of a database free approach.

ChimeraSlayer detects a large proportion (87%) of chimeric sequences with greater than 4% pair divergence, meaning it is capable of detecting a large proportion of intra-genus chimeras (chimeras where both parent sequences are members of the same genus), and can be run with or without reference database dependence\(^{215}\). This algorithm represented a substantial improvement over those which had been previously developed, and was able to detect more erroneous sequences then earlier approaches such as Bellerephon or Pintail, especially for similar sequences\(^{217}\). Another approach, Perseus, identifies chimeras independently of a reference database. It assumes that parent strands of a chimera must have experienced at least one more round of replication than the chimera, therefore allowing restriction of searches to those sequences present in greater
abundance than the query. Unfortunately, this method assumes that chimeras result from two parents, not taking into account the possibility of increased numbers of parent strands\textsuperscript{218}. The most recent algorithm, UCHIME, shows significant improvements over ChimeraSlayer, especially in cases where the chimeric sequence are small or arose from more than two parents. A further advantage of UCHIME, is that it performs significantly faster than both Perseus and ChimeraSlayer\textsuperscript{218}.

Once raw sequences have been quality trimmed, they can be assigned to taxonomic outcome groups to generate more meaningful information for downstream analyses. The two main approaches include direct assignment of sequences to phylotypes and operational taxonomic unit (OTU)-based processing\textsuperscript{219}. These methods each have their own benefits and limitations. When sequences are assigned directly to a taxonomy based on comparisons of sequences with known databases, assignment is highly contingent on the accuracy of the sequencing platform and the reference database being used. The alternate approach, which is used quite commonly in microbiome research, seeks to group sequences based on sequence similarity, and independent of a reference database. These groups, termed OTUs, are composed of sequences clustered at a given similarity threshold (typically 97%). OTUs are artificial constructs which do not correspond well with biological units (ie. species)\textsuperscript{219}. Following clustering, representative sequences are selected for assignment of each OTU to a taxonomic group. OTU based methods often inflate estimates of sample diversity with direct phylotype assignment less subject to this bias\textsuperscript{202}. However the ability to group sequences independently of 16S databases for initial steps, especially in the context of early databases which were incomplete and plagued by erroneous reads, has made OTU-based analyses popular\textsuperscript{220}.

Clustering organisms together based on a given distance threshold can be time consuming and computationally intensive. Calculating the distance between two sequences involves either direct sequence comparisons, or may take into account the three dimensional structure of the rRNA molecule. Numerous algorithms have been developed to perform sequence alignments, all balancing a need for accuracy with limited computing power. Dynamic programming, including the Needleman-Wunsch, and Smith-Waterman algorithms were designed to align any types of sequences, and to guarantee that the optimal alignment is found. These methods involve construction of global and
local distance matrices respectively, which are then used to calculate optimal alignments
between paired sequences. These processes are incredibly computationally intensive
however, and therefore useful only on a small number of sequences\textsuperscript{221}. Secondary
structure based methods such as Infernal are particularly useful for more distantly related
organisms as, in these cases, sequences may be highly divergent, although protein folding
remains conserved. However, Infernal often overestimates distance calculations except in
cases where the distance between organisms is already large\textsuperscript{222}.

Word-based methods and heuristic algorithms have been designed to reduce the
amount of computer power required to generate high quality alignments. Such methods
use early results to influence and reduce the number of future calculations, and decrease
the number of overall calculations required and amount of computer resources. While not
guaranteeing optimal alignment, the speed of such approaches has made them popular for
aligning large numbers of query sequences. Included in this group are several of the most
commonly used database search algorithms. Both the Basic Local Alignment Search Tool
(BLAST)\textsuperscript{223} and FASTA\textsuperscript{224}, used by the National Centre for Biotechnology Information
(NCBI), make use of small 'words' extracted from query sequences and searched against
target sequences in a database. When word hits are found, dynamic programming is used
to extend those matches outward as long as a good alignment is possible. While FASTA
will provide only the best alignment, BLAST will display many 'top' alignments, with bit
scores allowing scientists to select options which make the most biological sense. Various
iterations of both of these algorithms can be used to align DNA, RNA, and proteins.

Calculating the distances or relative similarities between multiple sequences from
a run requires even more computer resources due to the sheer number of comparisons to
be made. Large amounts of data, such as that produced by next-generation sequencing,
makes the use of dynamic programming infeasible\textsuperscript{214}. To address this, multiple sequence
alignment (MSA) algorithms have been developed and applied to sequence processing
pipelines, and aim to reduce the number of comparisons conducted on a set of sequences.
Propagative or progressive MSA methods involve grouping together sequences which are
most alike first, and adding less similar sequences later. This dramatically reduces the
number of comparisons required, but has the disadvantage of propagating early errors
forward without correction. Iterative MSA methods attempt to address this problem by
continually realigning subgroups in order to classify sequences into an accurate phylogenetic tree. Studies have demonstrated that MSA-based algorithms including MUSCLE (progressive) and MAFFT (progressive and iterative functionality), falsely inflate the amount of genetic distance between 16S sequences and artificially increase the number of OTUs generated225.

Additional tools have been developed which provide accurate pairwise-distances and are widely used for OTU generation in microbiome research. To minimize resource use, the main algorithms used include the DNADIST program from the PHYLIP package and k-mer based distance calculations with various mer sizes. Additional procedures are offered through different analysis pipelines226. Using calculated distances, different heuristic algorithms are then used to cluster sequences at a given threshold, including but not limited to nearest, average and furthest neighbour. Additional algorithms include CD-HIT, UClust, ESPRIT, and BlastClust. The choice of method has been shown to have a strong impact on observed community richness, so selection of appropriate, well validated parameters is essential. One simulation and mock community study demonstrated that the average-neighbour algorithm is capable of producing the most robust OTU groupings219.

To further simplify sequence processing and analysis, several suites of tools have been developed and are available online, offering iterations of common or proprietary algorithms for sequence alignment and OTU generation, as well as providing methods for performing quality control checks and generating graphical representations of data. These include dotur, mothur, and Qiime among others. These utilities provide a fair amount of flexibility in sequence quality trimming and data clustering, allowing the user to specify which alignment and clustering algorithms are used. Further, within those algorithms users are given the flexibility of modifying a number of the default parameters to optimize their data. Qiime, for example, offers seven different clustering methods (including each of those listed above) for assignment of sequences to OTUs, allows the user to specify which 16S rRNA database to use for taxonomic assignment, and allows the user to view data either graphically or through alternate data formats for use in other programs (Qiime 1.6.0).
1.4.3.6 Statistical analysis of microbial data

As with any new technology, consensus on the most robust way of analyzing microbiome data has yet to be reached. Many studies make use of traditional ecological indicators of community composition including estimates of coverage, and alpha and beta diversity. Such tools are specifically useful as they are not impacted by the distribution of the data. Diversity is a concept which is used often in microbiome research, but can be difficult to interpret and quantify. In general, diversity estimates incorporate information regarding species richness (different organisms detected in a sample) and abundance (number of a specific type of organism present in a sample). Alpha diversity is a measure of the mean diversity within a specific sample or ecosystem, whereas beta diversity is typically considered to involve a comparison of diversities between ecosystems or samples\textsuperscript{227}. Numerous methods have been developed to estimate alpha diversity, with the Shannon, Simpson and Chao1 indices most commonly used. These metrics were initially created for macro-ecology, and have subsequently been applied to microbiome analyses\textsuperscript{228, 229}. However, of these measures, none control well for rare organisms, and their applicability for comparisons between samples must be carefully considered. As such, measures of diversity may be a useful tool which should only be used in conjunction with other, more detailed information.

Beta diversity, in the context of microbiome research has been used to perform comparisons between samples based strictly on their microbial composition (Qiime website). As such, each sample community is explicitly compared to others through calculation of the distance or variance between samples. Evaluation of these distances is complex and may make use of different calculation matrices such as Jaccard, Euclidian, weighted or unweighted Unifrac. Unifrac is the most popular technique used to calculate beta-diversity in microbiome analyses, as it takes into account the phylogenetic structure of a community in calculating distance matrices and probabilities. Furthermore, both frequency (weighted Unifrac) and dichotomous (unweighted Unifrac) data can be analyzed, which is useful for ensuring the accuracy of statistical inferences\textsuperscript{230}.

Analyses which aim to determine whether specific organisms or OTUs are associated with outcome may also be performed, and can be used to provide more detailed and specific information regarding associations between environments or
phenotypes and microbes. Microbiome data can be analyzed as a dichotomous variable, based on whether an organism is detected in an experiment, or as a semi-continuous variable based on relative abundance. For dichotomous results, presence of an organism may be a reflection of the coverage achieved in an experiment, making it difficult to make comparisons between sequencing runs or platforms. Abundance data can also be problematic as interpretation of count data obtained from next-generation sequencing, given the lack of standardized internal controls, makes data normalization between batches and experiments next to impossible. As such, most samples rely on frequency data, whereby a degree of normalization is achieved by converting raw count data to frequencies in samples. Neither of these methods address the issue of primer bias towards specific sequences or the differential sequencing efficiency of different samples. However, dichotomous and abundance data of individual organisms can be used together to assess whether there are common differences between phenotypic groups.

Differences in dichotomous or abundance data between groups must then be assessed using statistics. Major challenges in analyzing microbiome data include the lack of conformation of data to a normal distribution, discrete abundance data (organism or OTU counts) which is later converted to semi-continuous frequency data (continuous, but bound by 0 and 1), and the observation that many organisms in microbial communities are detected in only a few samples, resulting in zero-inflation of data. Common statistical tests rely on the assumption of normality to accurately estimate probability, and break down when this condition is not met. To address this, different data transformation methods have been applied to microbiome data, including various log-based methods (which require additional manipulation with zero-inflated data), arcsine square root, Dirichlet multinomial, and negative binomial models. The advantages of these approaches, is that once data has been transformed, traditional statistical tests including analysis of variance (ANOVA), linear regression and t-tests, may be applied to detect associations. These methods are typically more powerful than exact or non-parametric approaches and are therefore especially useful in cases in which smaller sample sizes negatively influence experimental power. On the other hand, exact and non-parametric statistics can be used on many different types of data, and do not make assumptions about data structure or rely on complex transformations for analysis. However, zero-inflated data remains problematic for all analyses as the exact meaning of a zero in next-
generation microbiome sequencing is unclear. This makes it difficult to obtain an accurate estimate of the variance among subjects for sparsely detected organisms\textsuperscript{233}. Obtaining similar results when multiple types of analyses are applied may be considered indicative of robust results.

Regardless of the chosen approach, proper correction for multiple testing must be performed in order to account for the number of tests being analyzed and avoid type I errors. Stringent correction approaches such as a Bonferroni, may inflate the type II error, resulting in the rejection of true results. As such, a balance must be struck between maintaining statistical stringency and power. Unlike genetics, the FDR method is most commonly used in microbiome analyses. In addition, some groups apply a higher significance threshold (Corrected P-value $<$ 0.3) combined with evidence of biological plausibility as a check for suspected associations, which can be effective when information is available regarding groups of interest, but is less so in analyzing novel organisms. However, biological plausibility in itself must not be used as the sole verification of results, as it does not necessarily imply that egregious analytical errors have not occurred. As such, this approach must be used with caution to prevent reporting inaccurate results.

Because of the comparative newness of microbiome research, gold-standard protocols for analysis have not yet been developed and widely accepted by the research community. Different groups make use of different methodologies from sampling to sequence processing and analysis, which have profound impacts on data quality and can prevent the accurate comparison of results between studies. Some of the methodological issues which can have an impact on perceived microbiome composition include: extraction method (whether physical, chemical or combination protocols are used), extraction kit (commercial variety or homemade), primer selection (differential primer bias)$^{234}$, sequencing method (Sanger vs 454 vs Illumina vs microarray), clustering and alignment algorithm, database used for taxonomic assignment, transformation method, and statistical approach. These issues may be minimized within a study by ensuring that standard protocols are used for all samples, however are difficult to account for in comparisons between studies, and the inherent biases of chosen methodologies must be considered in interpreting data. Despite these challenges, a great deal has been learned.
recently about the populations of microbes inhabiting different anatomical locations on
the human host.

1.4.4 The Human Microbiome – lessons from community level analysis

Before understanding which changes are necessary for the development of
disease, it is necessary to first attempt to identify a 'core' microbiome, or set of organisms
which are present at a given sampling location among many individuals. Depending on
the parameters used in identifying 'core' features, different results can be obtained. Using
the most universal definition of inclusion, that to be included as part of a 'core'
microbiome an organism must be detected in a site of interest among every individual
studied, several phyla are identified which fall into this category in the intestinal tract.
While most gut or stool microbiome studies typically identify up to 20 phyla in samples,
only Actinobacteria, Firmicutes, Bacteroidetes, Proteobacteria and Fusobacteria may be
considered to be universal in that they are detected across all samples. A study by Huse et
al attempted to identify a core set of organisms at the OTU level by compiling results
obtained from several studies of healthy subjects, and across many different biological
habitats. In this study, stool was found to have fewer 'core' OTUs than were oral cavity
sites, but more than were detected in skin and vaginal sites. When a lower threshold was
used for inclusion as a core organism such as detection of an OTU in 50% of individuals,
the number of OTUs reaching this threshold was approximately 15%. Notably, both
definitions of 'core' do not take into account organism abundance, so some of the
organisms were present in high numbers among certain samples, and were rare in others.
To date, no intra-individual ‘core’ organisms have been found which can be detected
across all body habitats. Additional work is now focusing on identifying factors which
are responsible for shaping an individuals’ microbiome, in hopes of explaining some of
this variability.

Microbial colonization of humans begins at birth, with mode of delivery being the
first factor known to influence the microbiome. Slightly later, breast feeding versus
formula feeding impact the development and composition of the gut associated
microbiome in early childhood, as does weaning, and helps to explain the large amount of
temporal variability undergone by microbial populations early in life. The speed of
change begins to level off as children approach adolescence, and large scale changes in microbiome structure are not typically observed again until much later in life. Typifying this change is the correlation between increasing age and decreasing *Bifidobacterium* levels. However, despite the relative stability of the microbiome through adulthood, in time-series experiments, few organisms can be detected across all sampling times (approximately 10% in stool, over 100 daily samples). Additionally, use of antibiotics leads to more drastic perturbations of the host microbiota. These agents are capable of significantly altering the microbiome, and while communities often revert to a state similar to that which existed prior to medication use, they frequently do not fully recover.

Evidence that the microbiome is at least partly shaped by dietary and environmental factors is compelling. Host organisms with diverse dietary patterns (herbivores, carnivores, omnivores) have differences in the structure of their microbiome which transcend phylogenetic relationships between host species. Furthermore, similar observations may be seen among human populations who consume broadly different diets. In a comparison between children from Africa and Italy where there are substantial differences in the amount of fibre, calories and protein consumed, broad differences in the gut microbiome are detected and include a significantly higher overall ratio of Firmicutes/Bacteroidetes among individuals from Europe compared to Africans. Furthermore, several genera are present exclusively among African children, including *Prevotella*, *Xylanibacter* (Bacteroidetes) and *Treponema* (Spirochaetes) which may all contribute to enhancing nutritional uptake from plant polysaccharides. Yet breast fed children in both locations displayed similarities which caused them to group independently of either larger group. This suggests that diet is an important driver of microbial composition among healthy, well matched individuals consuming different diets. Dietary patterns among Western populations have also shown evidence of association with microbiome profiles, with individuals consuming a protein/animal fat rich diet having increased *Bacteroides*, and those consuming more carbohydrates enriched for *Prevotella*. Consumption of large amounts of fruits and vegetables further alters the composition of the human microbiome.
Other factors which alter gut bacterial communities include smoking, medication use and ethnicity. Indeed, smoking has been associated with changes in the microbiome, as have various medications which were not previously recognized to have bactericidal effects, such as 5-aminosalicylic acid medications (5-ASA)s. Such differences could account for the differential risk of CD and UC among smokers and non-smokers, and the efficacy of different drugs in reducing inflammatory activity. However, these results need to be replicated in additional cohorts to determine their validity, as associations to date have been marginal. Additional studies have suggested that ethnicity may contribute to microbiome heterogeneity, which may make it difficult to evaluate the confounding effect of other environmental differences and the potential impact of genetic heterogeneity, on microbial composition.

Additional microbiome heterogeneity between subjects can be attributed to cooperation and competitive exclusion between the microbes themselves. Certain species or genera often co-associate with other distinct organisms, often independently of phylogenetic relationships. These co-associations likely result from synergistic relationships between groups. *Marvinbryantia formatexigens*, for example, grows more rapidly in the presence of formate, and has been observed to positively co-occur with *Butyribrio fibrisolvens*, a formate producer. Samples taken at birth, and following antibiotic treatment are more likely to consist of 'opportunistic' organisms, capable of fast growth, and without a stringent requirement for the presence of additional organisms to provide metabolic co-factors. As time following birth or removal of antibiotic application lengthens, fewer such opportunistic organisms are detected, as they are replaced by more symbiotic groups. This successional colonization likely has important implications following IBD-related surgery and may explain why the use of probiotics and antibiotics later in life often have transient effects on the microbiome.

The microbiome also plays a role in regulating its human niche, most notably through influencing and modulating the function of cells and the immune system. *Helicobacter pylori*, the causative agent in stomach and duodenal ulcers, for example, is known to modulate the macroenvironment of the stomach lumen, increasing or decreasing acid secretion depending on the location of infection. Other organisms directly impact the host immune system, such as segmented filamentous bacteria (SFB) which induces production of Th17 cells in rodents, and *Faecalibacterium prausnitzii*,...
which excretes an as yet unidentified compound with anti-inflammatory effects\textsuperscript{250, 251}. Additionally, temporal variation in microbial populations, including changes in population structures during a lifetime, suggest that the functional benefits conferred by certain organisms on immune development may have a finite opportunity to exert their effects. For example, while few studies have described detection of the immunomodulating organism SFB in humans, those which have suggest a near absence of SFB among adults compared to children\textsuperscript{252}. Organisms such as these with known function, represent only a tiny fraction of the intestinal microbiome, as such it is likely that with more study, a great many additional species with similar host modulating properties will be found.

In an attempt to better understand these relationships, the Human Microbiome Project (HMP) was conceived. In many ways analogous to the human genome project which first sought to obtain a roadmap of the human genome, the aim of the HMP is to catalogue the typical organisms present among healthy individuals. The HMP has done a great deal of work cataloguing the organisms present in numerous human sites, including the digestive tract, skin, oral and nasal cavities, and urogenital tract with the ultimate goal of determining the impact of microbial communities on human health. To answer this question, it is necessary to discover which organisms are present in specific sites, and what functions they are capable of performing. While both of these questions are important, to date, the majority of work has focused on the former, with little yet known about the functional implication of such findings. This work has lead to the generation of reference genome sequences for numerous previously unidentified or unstudied organisms, and has provided a basic catalogue of the bacteria present in different body habitats. Further work aimed at investigating specific disease phenotypes and evaluating how the human microbiota contributes to disease, either through altering immunity or other as yet unknown mechanistic pathways, has also been done.

1.4.5 Location-specific microbiome heterogeneity

The microbial composition of different environments is reflective of macro-ecological factors, including nutrient bioavailability, pH, salinity, and in the case of the human microbiome, immunological factors including components of the innate and
adaptive immune systems. These environmental differences shape the microbiome and lead to significant variation in microbial profiles between sampling sites. Along the digestive tract alone, it has been demonstrated that microbial profiles are significantly different between the mouth, stomach, small and large intestines and stool\textsuperscript{187, 253, 254}. Even physically close sites such as the dorsum of the tongue and hard palate can be substantially different, suggesting the importance of local selective pressure on community composition\textsuperscript{255}.

Numerous studies cataloguing site-specific bacterial composition have been conducted. The most diverse habitats in the human host are found in the oral cavity and lower large bowel\textsuperscript{248}. Interindividual variability is high as is temporal variability at a given site, with the oral cavity seemingly less susceptible to change over time than other sites. Few organisms are detected in all body habitats (including stool, skin sites, oral and nasal cavity)\textsuperscript{248, 256}. Interestingly, when finer mapping has been performed examining different skin micro-locations it appears that variation in environmental characteristics shape microbial populations, with sebum producing areas such as the forehead and face dominated by Propionibacterineae, and more dry locations such as the trunk and legs dominated by \textit{Staphylococcus} species\textsuperscript{256}. Other sites also appear to be dominated by different organisms or groups of organisms. Stool, for example, has a higher amount of \textit{Bacteroides}, and several genera from within the Firmicutes, vaginal sites tend to be dominated by \textit{Lactobacillus}, and the mouth by \textit{Streptococcus, Haemophilus, Prevotella} and \textit{Veillonella}\textsuperscript{248}. Differences between sampling locations within the digestive system are also apparent. As a general rule, diversity increases from the stomach to lower large bowel. Differences in the composition of the intestinal microbiome are striking between the ileum and the colon; however, subtle differences can also be observed between biopsies taken from left and right colonic tissue, including altered frequencies of \textit{Roseburia, Altistipes} and \textit{Fusobacterium}. Yet some of the largest differences in community structure are observed between stool and any intestinal biopsy location, suggesting caution must be used in interpreting the biological significance of stool-based studies\textsuperscript{246}.

Despite the recognized taxonomic diversity of the intestinal ecosystem, the functional capabilities of organisms in this site are more conserved than observed in
other, less diverse habitats. A recent study by Cantarel et al demonstrated that specific carbohydrate bioavailability in different locations is a primary driver of microbial population structure, and that although community membership composition may be different, the metabolic potential of regions with specific carbohydrate sources may be less variable. Furthermore, while stool and oral samples are taxonomically distinct, organisms from both sites demonstrate distinct functional potential from non-digestive sites as they more commonly possess genes involved in the breakdown of plant and fungal carbohydrates. It has been recognized, therefore, that despite phenotypic relatedness (at the family level), there is a high degree of variability in the number and type of metabolic processes encoded in bacterial genomes, suggesting that species-level resolution is necessary in order to accurately deduce functional potential. The importance of the relationship between taxonomy and sequence-deduced metabolic potential has not been determined. However, the incomplete nature of current metagenomic datasets, highlights the importance of adding to current reference sequence collections to allow more accurate functional inference.

1.4.6 Microbiome and Immunity

Germ free mice have altered immune structures compared to those raised with normal commensal bacteria. Alterations include smaller splenic lymphoid follicles, fewer splenic CD4+ T cells, smaller Peyer’s patches, fewer intestinal lymphocytes, and an immune profile skewed towards a more Th2 type response. These changes can be reversed following colonization with specific members of the commensal microbiota. Interestingly, whole microbes may not be required to shape this relationship, as specific microbial associated molecular patterns (MAMP)s may also be capable of inducing more ‘typical’ immune processes. *B. fragilis* polysaccharide A (PSA), for example, has the ability to increase the number of CD4+ T-cells to normal levels in germ free mice, independent of *B. fragilis* colonization. PSA alone enhances production of IL10, produced systemically by regulatory T-cells and by several other cell types in response to prolonged inflammation. Furthermore, *B. fragilis*-derived PSA appears to signal directly through TLR2 to promote immunogenic tolerance, a feature which is necessary for colonization of the host intestinal tract by commensals. Of note, is the observation
that several organisms falling into the *B. fragilis* group (which includes 10 species in total) are decreased among individuals with IBD compared to healthy controls\(^{260}\). Other organisms have also shown immunomodulatory potential, include SFB which induce production of Th17 cells in rodents through an as yet unidentified epitope\(^{261}\), and Clostridia groups IV and XIVa which induce the accumulation of T\(_{reg}\) cells in the lamina propria of the colon\(^{262}\).

Variation in the overall composition of the intestinal microbiota brought on by differences in hygiene, have also demonstrated influence on tissue gene expression patterns. Animals housed in a clean environment with microbial profiles dominated by Bacteroidetes and reduced Firmicutes, have increased expression of pathways involving immune response-IFN alpha/beta signaling, and antigen presentation by MHC I compared to animals housed in less sterile environments\(^{263}\). Of further interest is the observation that different microbes have the ability to alter host response in a location-specific manner; for example, interacting with the immune system in different ways depending on whether colonizing the intestinal lumen or blood\(^{168, 264}\). Additionally, stress-induced alterations in bacterial community structure\(^{265}\), and increased detection of bacterial lipopolysaccharide (LPS) in blood brought on by increased intestinal permeability, is at least partially responsible for the cytokine response (increased IL-1\(\beta\) and IL-18) typically observed in rats exposed to stressors\(^{266}\).

1.4.7 Genes, serology and microbiome

To date, no associations between the anti-microbial antibodies and different components of the microbiome have been observed. However, in a small number of pouch patients, several inflammatory markers including CRP, TNF-\(\alpha\) and erythrocyte sedimentation rate (ESR), showed evidence of correlation with specific microbes. In these patients, serum CRP was inversely correlated with Streptococcaceae, TNF-\(\alpha\) was correlated or inversely correlated with several different bacterial families, and ESR correlated with Staphilococcaceae and Bacillaceae but inversely with total Moraxellaceae\(^{267}\). This study was small, however, and the results have not been replicated in either another pouch cohort, or among individuals with colonic disease.
Gene-microbe associations are also beginning to be evaluated, in hopes of determining which endogenous host factors are responsible for a specific microbial 'phenotype'. Genetically distinct mouse strains harbour divergent microbial communities, regardless of housing conditions and diet, with different OTUs dominating the bacterial community in different strains. In total, 42 OTUs from the uncultured Firmicutes and Bacteroidetes were highly variable between ten different mouse strains tested\textsuperscript{268}. Furthermore, among humans, $NOD2$ genotype may be associated with microbial community composition, and differential frequencies of the groups Actinobacteria, Bacteroidetes, Clostridium GroupIV (Firmicutes) and Bacillus (Firmicutes), although this result requires validation in an additional cohort.

1.4.8 Microbiome in IBD

In the case of complex diseases, differences in bacterial growth patterns may have a profound impact on disease development. In IBD, the idea that bacteria occupying certain niches either alter gene expression patterns in those areas or interact with specific host features to alter immune response, leading to the development of location specific inflammation, has gained much traction. Tissue gene expression patterns have been observed to be affected by the presence of bacterial DNA, with several genes involved in inflammatory pathways either up or down regulated in response to this stimulus. In individuals with CD and UC, the number of gene transcripts displaying this phenomenon is more than doubled\textsuperscript{269}.

IBD and several other diseases including Type II diabetes and obesity, are characterized by a decreased level of microbial diversity, reduced numbers of typically ubiquitous organisms, and community structures which are more susceptible to drastic perturbation than are those of healthy individuals\textsuperscript{171, 172, 270, 271}. It has been suggested that decreased diversity may lead to communities which are less resistant to perturbation\textsuperscript{272}. However, it is also possible that even in the context of exposure to a wide variety of microbes, healthy guts will be dominated by a limited number of beneficial organisms\textsuperscript{263}. Numerous studies have implicated specific organisms and changes in community structure in the pathogenesis of IBD. Faecalibacterium prausnitzii and adherent-invasive $E.\ coli$ (AIEC) are perhaps the best studied organisms demonstrating association with
IBD. *F. prausnitzii* is decreased in the stool of many individuals with IBD, and AIEC shows evidence of being detected more frequently among individuals with ileal CD\cite{246,273}.

Numerous additional organisms have also been associated with inflammation, detected more or less frequently in stool or tissue biopsies from those with inflammation. Examples include *Roseburia*, the genus *Escherichia* and *Odoribacter* which have demonstrated associations with IBD, the *Roseburia* and *Escherichia* exclusively with ileal CD, and *Odoribacter* with ileal CD and pan-colitis\cite{246}. *Bacteroides* decrease in inflammation (both CD and UC), as do *Lactobacillus*, *Ruminococcus*, and *Bifidobacterium*. *Alistipes* are decreased in both ileal and colonic CD\cite{331}. Like *F. prausnitzii*, many of these organisms are known producers of immune-modulating compounds including butyrate\cite{246}, suggesting a potential important role in regulating the inflammatory response among individuals without IBD. Some of the organisms suggested to be increased in inflammation include members of the phylum Proteobacteria in ileal inflammation\cite{331}, as well as the genera *Methanobrevibacter*, *Campylobacter*, *Collinsella*, *Bifidobacterium* and sulfate reducing bacteria in both CD and UC compared to controls\cite{331,344}. Further microbial associations with IBD are discussed in Chapter 4.

Several murine studies have also demonstrated the importance of microbes in the etiology of IBD. Specific strains of *Proteus mirabilis* and *Klebsiella pneumoniae* derived from *T-bet*\textsuperscript{−/−} × *Rag2*\textsuperscript{−/−} mice are capable of transferring colitis to wild-type mice\cite{274}. NLRP6 or ASC deficiencies also lead to changes in the gut microbiota of mice, which are transferrable to wild type animals upon cohabitation. As components of multiprotein inflammasomes, these proteins are critical in innate immune response to MAMPs through triggering cleavage of pro-inflammatory cytokines. The changes to the microbiome engendered by knock-outs of these proteins include increases in *Prevotella*, an additional unknown member of the family *Prevotellaceae* and several members of the TM7 phylum\cite{275}.

Changes in host immunity resulting from the inflammatory process, also likely influence the function and composition of the gut microbiome. The lumen of the healthy colon is normally relatively anaerobic, and therefore supports the growth and expansion of obligately anaerobic organisms. During inflammation increased prevalence of facultative anaerobes (ie. *Enterobacteriaceae*) may be caused by increased production of
ROSs by the innate immune system as part of the inflammatory process\textsuperscript{276}. These organisms gain a competitive advantage in this scenario, as they are capable of using several ROS as terminal electron acceptors for anaerobic respiration\textsuperscript{276}. Some organisms (ie. Salmonella spp.), actually illicit inflammation in order to gain a competitive advantage over other commensal bacteria\textsuperscript{277}. Furthermore, chronic intestinal inflammation in mice has been demonstrated to contribute to the upregulation of bacterial stress response proteins, providing some measure of protection from oxidative stress\textsuperscript{278}. Such findings demonstrate the important role of the innate immune system in modulating the structure of the colonic microbiome. These changes may have an important role in propagating an inflammatory state.

A potential role for microbes in diagnosing or differentiating IBD subtypes has also been explored. The diagnostic potential of synthetic learning in microbial ecology (SLiME) analysis of 16S sequence data among stool samples from a pediatric cohort, observed that they were able to detect IBD with 80.3\% sensitivity and 69.7\% specificity. Differentiating between CD and UC was less accurate (specificity 49\%, sensitivity 95\%). Interestingly, this algorithm performed better at distinguishing CD from UC among individuals in remission than among those with active disease\textsuperscript{271}. This suggests that inflammation-specific microbial changes are likely similar between CD and UC, and that these may overwhelm signals which are characteristic of and specific to these phenotypes. While not as accurate as current serology and genetics based approaches, this method demonstrates a great deal of future potential, given the amount of study which is currently directed at this field.

Several studies have also begun to investigate the human virome and fungome (collection of viruses and micro-eukaryotes respectively, co-existing with humans) in the context of IBD. The human virome, for example, is highly individual-specific and appears to be dominated by bacteriophages\textsuperscript{279}. These particles are capable of shaping the composition of the observed microbiome through the transmission of genetic material to bacterial hosts, with high specificity for certain species\textsuperscript{280, 281}. While evidence for a viral role in IBD development is scant, mouse studies have indicated that in the context of specific IBD-associated genetic polymorphisms and a normal microbiota, viral infection with a murine novovirus may lead to the development of colitis\textsuperscript{282}. The diversity of
micro-eukaryotes, on the other hand, tends to be low and fungal communities tend to be more stable temporally than either the microbiome or virome\textsuperscript{283}. Contrary to the microbiome, fungal diversity in individuals with CD has been suggested to be higher than that observed in healthy controls\textsuperscript{284}. Although it is likely that these organisms play a role in influencing disease outcomes, additional studies are required to confirm their importance.

Given the apparent profound effect of microbes on health and disease, it seems logical that altering the microbiome through ingestion of specific microorganisms with known beneficial effects might result in positive outcomes for patients with IBD. The development of probiotic formulations, designed to treat a variety of conditions, have this aim in mind. In actual fact, the effect of probiotics on altering disease processes has met with limited success in the case of IBD\textsuperscript{285-289}, and has been only slightly more successful in the treatment of pouchitis\textsuperscript{287,290}. Several reasons may exist for this apparent contradiction: the organisms which are used in common probiotic formulas may not be those capable of modulating the host immune system, and rather their described benefits were determined based on studies which were associative in nature rather than causative. Alternatively, in the context of already-established inflammation, the use of probiotics may be insufficient to improve disease processes\textsuperscript{291}. It is also possible that the already-established microbiota prevents colonization of the host with non-native organisms. This final observation has been supported by studies demonstrating that, while transient changes to the host microbiota may be observed in response to probiotic organisms, these are most commonly rapidly reversed after probiotic treatment is stopped\textsuperscript{292}.

Fecal transplantation (FT) also has the potential to modulate the structure of the gut microbiome, and has been demonstrated to be an effective treatment for \textit{Clostridium difficile} infection in cases where antibiotics are unsuccessful. The use of this treatment for IBD is still being investigated, however, small reports and case studies have suggested that FT may result in decreased disease symptoms and medication use\textsuperscript{293}. However, one case study describes an individual with quiescent UC treated with fecal transplant for \textit{C. difficile}, developing a transient UC flare following transplantation\textsuperscript{294}. Such data highlights the need for additional study prior to the widespread use of this procedure in patients.
1.4.9 Microbiome and pouchitis

As with pouch genetic studies, those investigating the microbiome in pelvic pouches and pouch inflammation have been small, have not consistently made use of next generation sequencing approaches for analysis, and have had flaws in study design or sample collection and produced inconclusive results. A prospective study of ten patients found that pouches had increased levels of Enterobacteria and coagulase negative Staphylococci compared to pre-surgical ileum samples, and increased Bacteroides, Lactobacillus and Veillonella compared to pre-surgical rectal samples\(^{295}\). The majority of organisms detected in pouch samples belong to one of the usual four major phyla, although their prevalence is quite different from that typically seen in colonic tissue\(^{296}\). Little work has investigated whether differences in genus or species level groups are detectable between pouch and non-pouch samples, representing a large gap in current knowledge.

Some studies have shown that pouchitis episodes are characterized by decreased numbers of anaerobic bacteria, and increased numbers of aerobic organisms\(^{297}\) and a generalized alteration of the tissue associated microbiota, characterized by reduction in organism diversity\(^{296, 298, 299}\). Yet other studies have found no difference in bacterial diversity or ratio of aerobic to anaerobic organisms between individuals with pouchitis and those without, and that only rare organisms were differentially abundant between inflamed and uninflamed pouches\(^{267, 300}\). Recent work presented evidence that, at the family level, Enterococcaceae, Streptococcaceae, Lachnospiraceae and Alcaligenaceae were reduced in patients with pouchitis; however, these changes were not consistently statistically significant, and not replicated across all studies\(^{267, 296, 299}\). At the genus level, increased Roseburia, Clostridium, Lachnospira, Veillonella, Prevotella and Akkermansia were detected in pooled samples from individuals with pouchitis compared to FAP, while Bacteroides, Parabacteroides, Escherichia and Faecalibacterium were decreased\(^{301}\). Furthermore, treatment with ciprofloxacin and metronidazole, reduces the number of aerobic and anaerobic organisms, respectively, which are detected in inflamed pouches, with reductions closely mirroring improvements in symptoms among those treated with ciprofloxacin\(^{297}\).
1.5 Overarching theme - bringing it all together

One of the greatest difficulties in studying IBD is the difficulty in determining which factors are associated with disease development and which are the result of disease. To get an accurate picture of the environmental and bacterial triggers leading to disease development among a genetically susceptible host, it is necessary to either evaluate individuals prior to the onset of disease, or study the recurrence of inflammation in a well-characterized model, preferably human. The evidence which has been presented suggests that genetic and microbial factors both play a role in pouch inflammatory processes. The detection of anti-microbial antibodies in the serum of individuals with pouch inflammation, and the phenotypic similarities characteristic of both IBD and chronic pouch inflammation, suggest that these processes may proceed via common mechanisms. Thus, inflammation of the pouch following IPAA is an appropriate human model for studying de novo intestinal inflammation.

The precise changes which small bowel tissue undergoes in order to become susceptible to inflammation, among individuals in whom disease had been previously confined to the large bowel (UC), has important implications for pathology. A better understanding of the role of microbes in the context of a well defined genetic predisposition to bowel inflammation, will be useful for better understanding pouch inflammation, as well as IBD in general. Furthermore, additional insight relating to the subtle differences in phenotype between pouchitis and IBD, such as the antibiotic responsive nature of the majority of pouchitis cases, may be gained from a wider study examining these factors.

1.6 Hypothesis and Aims

It is hypothesized that pouch inflammatory complications arise from similar mechanisms to those involved in IBD pathogenesis (both CD and UC) and that susceptibility for both IBD and pouch inflammation result from a similar genetic predisposition and microbial profile, with anti-microbial antibody markers demonstrating patterns which are
characteristic of chronic inflammation. Furthermore, individuals with UC and pouches who demonstrate CD-like or chronic inflammatory phenotypes following IPAA will have a genetic, serological and microbial profile similar to individuals with CD. As such, the aim of this thesis was to evaluate the genetic, serological and microbial factors which are associated with pouch inflammatory complications in a large, well-characterized cohort of patients with IPAA.

The specific aims of this study were to evaluate:

1) Whether specific, well characterized IBD-associated serological markers are also associated with pouch inflammation.

2) Whether genetic factors which are known to be associated with IBD, or which are suggested to be involved in inflammatory or immune pathways, are associated with chronic pouch inflammation.

3) Whether specific alterations in the microbiota of the pelvic pouch were associated with pouch inflammation.
Chapter 2: Antimicrobial Antibodies Are Associated with a Crohn's Disease-Like Phenotype following Ileal Pouch-Anal Anastomosis

Reprinted from Clinical Gastroenterology and Hepatology, 10(5), Andrea D. Tyler, Raquel Milgrom, Wei Xu, Joanne M. Stempak, A. Hillary Steinhart, Robin S. McLeod, Gordon R. Greenberg, Zane Cohen, Mark S. Silverberg. Antimicrobial Antibodies Are Associated with a Crohn's Disease-Like Phenotype following Ileal Pouch-Anal Anastomosis, Pages 507-12, Copyright (2012), with permission from Elsevier
Title: Anti-Microbial Antibodies Are Associated with a Crohn's Disease-Like Phenotype following Ileal Pouch-Anal Anastomosis

Authors: Andrea D. Tyler1,2, Raquel Milgrom2, Wei Xu3, Joanne M. Stempak2, A. Hillary Steinhart2, Robin S. McLeod2, Gordon R. Greenberg2, Zane Cohen2, Mark S. Silverberg2

Specific Author Contributions:

Andrea D. Tyler: Conception and design of the study; generation, collection, assembly, analysis and interpretation of data; drafting and revision of the manuscript; approval of the final version of the manuscript

Raquel Milgrom: Collection and assembly of data; drafting of the manuscript

Wei Xu: Analysis and interpretation of data; drafting and version of manuscript; approval of the final version of the manuscript

Joanne M. Stempak: Design of the study; generation, collection, assembly and interpretation of data; drafting and revision of the manuscript; approval of the final version of the manuscript

A. Hillary Steinhart: Collection of data; Drafting and revision of the manuscript; approval of the final version of the manuscript

Robin S. McLeod: Collection of data; Drafting and revision of the manuscript; approval of the final version of the manuscript

Gordon R. Greenberg: Conception of the study; Drafting/review of manuscript; approval of the final version of the manuscript

Zane Cohen: Conception of the study; collection and interpretation of the data; Drafting/review of manuscript; approval of the final version of the manuscript

Mark S. Silverberg: Conception and design of the study; generation, collection, assembly, analysis and interpretation of data; drafting and revision of the manuscript; approval of the final version of the manuscript
Acknowledgements:

The authors would like to thank Diane Verbeeten and Lucy T. Zhang for their assistance with the study, and Harden Huang and Brenda O’Connor for maintaining and providing access to the Mount Sinai Hospital Pelvic Pouch Database. This study was supported by a grant from the Crohn’s and Colitis Foundation of Canada.
2.1 Abstract:

Background & Aims: Pouchitis and Crohn’s disease-like (CDL) complications of the pouch occur at rates near 50% and 20%, respectively, following colectomy with ileal-pouch anal anastomosis (IPAA) for ulcerative colitis (UC). We investigated whether anti-microbial antibodies are associated with pouch outcome following IPAA.

Methods: We studied clinical and endoscopic data from 399 individuals with UC who underwent colectomy with IPAA at Mount Sinai Hospital in Toronto, Canada; patients were classified as no pouchitis, chronic pouchitis (CP), or CDL. Serum samples were analyzed from 341 patients for antibodies against *Saccharomyces cerevisiae* (ASCA), OmpC, CBir1, and perinuclear neutrophil cytoplasm (pANCA).

Results: Of the subjects, 70.7% had no pouchitis, 16.8% developed CP, and 12.5% developed CDL. Smoking was associated with CDL ($P=.003$). Ashkenazi Jewish individuals more commonly had CP ($P<.008$). Of patients with CDL, 53.5% and 14.0% had positive test results for anti-CBir1 and ASCA (immunoglobulin [Ig] G), respectively, compared with 21.4% and 3.8% of those with no pouchitis and 28.3% and 5.0% of those with CP ($P<.0001$ and $P=.03$). Anti-CBir1 was associated with CDL, compared with no pouchitis ($P=2.8 \times 10^{-5}$; odds ratio [OR]=4.2; 95% confidence interval [CI], 2.2–8.3) or CP ($P=.011$; OR=2.9; 95% CI, 1.3–6.6). ASCA IgG was associated with CDL, compared to patients with no pouchitis ($P=.01$; OR=4.1; 95% CI, 1.4–12.3). In a model, levels of pANCA and the anti-microbial antibodies were associated with CP ($P=.029$) and CDL ($P=4.7 \times 10^{-4}$).

Conclusions: Anti-microbial antibodies and pANCA are associated with inflammatory complications of the pouch. The CDL phenotype is associated with factors that characterize Crohn’s disease, including smoking, anti-CBir1, and ASCA.

KEY WORDS: ileal pouch-anal anastomosis; serologic markers; inflammation; surgical complications, ulcerative colitis; inflammatory bowel disease; pouchitis
2.2 Introduction

Ulcerative colitis (UC) is a chronic inflammatory disease of unknown etiology affecting primarily the colon. Several studies have shown that greater than 20% of affected individuals will require surgical management at some point in their disease course. The surgical approach in the setting of fulminant or chronic, treatment refractory UC and colonic dysplasia is colectomy with ileal-pouch anal anastomosis (IPAA). Despite the surgical removal of the colon, de novo inflammation of the ileal reservoir, termed pouchitis, occurs in 12% to 50% of patients. Treatment of pouchitis is often sub-optimal. Symptoms may resolve after a defined treatment course with broad spectrum antibiotics, however, some cases require continuous therapy, immunomodulators or biologics in order to maintain symptom control. Up to 20% of patients with a pouch go on to develop a Crohn’s disease (CD)-like phenotype, characterized by inflammation in the afferent limb, the presence of proximal small bowel strictures unrelated to surgery, or perianal/abdominal fistulas or abscesses that occur at a remote time from the IPAA procedure.

Prior studies have identified factors which may be associated with inflammatory complications of the ileal pouch in IBD such as smoking, colonic disease extent, primary sclerosing cholangitis (PSC), and extra-intestinal manifestations (EIM). Family history of CD has been shown to increase the risk of a CD-like phenotype of the pouch. In addition, the utility of serological markers such as pANCA and ASCA, in the setting of pouchitis and CD-like complications, have been evaluated with variable results. The aim of this study was to determine which clinical and serological factors are associated with outcomes after IPAA in a large cohort of UC patients.

2.3 Materials and Methods

Study Population

More than 1600 patients having undergone IPAA prior to 2007 at Mount Sinai Hospital (MSH) in Toronto, Canada, and who were registered in a Pelvic Pouch database,
were contacted to participate in the study which was approved by the Research Ethics Board of MSH. All patients provided written informed consent.

Patients with a confirmed pre-colectomy diagnosis of UC who had a minimum of two years follow up after ileostomy closure were eligible for the study. Patients with confirmed or suspected pre-colectomy CD, inflammatory bowel disease unclassified (IBDU)/indeterminate colitis (IC), or familial adenomatous polyposis (FAP) were excluded. A pre-colectomy diagnosis of UC was based on clinical, endoscopic, histologic, and when available, radiologic evidence of disease as documented in patient clinical files. Colectomy pathology was also reviewed to verify the diagnosis. Patients with evidence of backwash ileitis (n = 11) pre-colectomy were included.

Clinical Data Collection

Study data was retrospectively collected by thorough clinical chart review and patient interview. Pre-colectomy disease extent was characterized using the Montreal Classification (E1, E2 and E3). Clinical data collected included gender, date and age of IBD diagnosis, time from diagnosis to colectomy, family history of IBD, ethnicity, EIMs including arthralgias, osteoporosis/osteopenia, erythema nodosum, pyoderma gangrenosum, PSC, ocular inflammation, previous appendectomy and post-surgical outcomes. Only 7 patients in the cohort had PSC so no analyses were performed for this group specifically. Smoking history was also documented: individuals smoking at least one cigarette per day were considered smokers and those having quit prior to recruitment ex-smokers.

Post-Surgical Classification

Review of clinical records and patient interview to document disease course following ileostomy closure was performed. Patients were grouped into one of three categories based on post-surgical outcomes (Table 6). The “no pouchitis” (NP) group included individuals who had no episodes of pouchitis or who had not experienced more than three acute episodes of pouchitis per year. An acute episode was defined as one where there was a complete clinical response to less than or equal to a two week course
of antibiotics (ciprofloxacin, metronidazole, or combination) which is the usual treatment protocol at MSH. The “chronic pouchitis” (CP) group included antibiotic-dependent and antibiotic-resistant patients who required either prolonged (>1 month) antibiotic therapy or required the use of second- or third-line medications (5-ASA, steroids, immunomodulators, biologics). The final group included individuals with a Crohn’s disease-like phenotype (CDL). These were defined by one or more of the following criteria: a) development of a perianal fistula greater than one year after ileostomy closure; b) a stricture proximal to the pouch which was not related to a surgical complication; c) evidence of endoscopic inflammation (ulceration, erythema, friability) in the afferent limb/pre-pouch ileum or more proximal small intestine. Additional adverse post-surgical outcomes such as outlet obstruction, anal strictures, or rectal cuff inflammation were documented, but did not lead to exclusion. Time to development of pouch outcome was defined as the time from ileostomy closure to the onset of symptoms and diagnosis.

Table 6: Outcome groups following IPAA

<table>
<thead>
<tr>
<th>No Pouchitis (NP)</th>
<th>Chronic Pouchitis (CP)</th>
<th>CD-like phenotype (CDL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- No episodes of pouchitis</td>
<td>- Patients requiring long-term (&gt;1month) use of antibiotics to maintain remission</td>
<td>- Abdominal or perianal fistula not related to surgical complication (&gt;1 year post-ileostomy closure)</td>
</tr>
<tr>
<td>-OR-</td>
<td>-OR-</td>
<td>-AND/OR-</td>
</tr>
<tr>
<td>- &lt;4 antibiotic responsive episodes of pouchitis per year</td>
<td>- Patients require second- or third-line medications (i.e. 5-ASA, steroids, immunomodulators, biologics) to maintain symptom control</td>
<td>- Inflammation in the afferent limb (pre-pouch ileum) or more proximal small bowel</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-AND/OR-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Proximal bowel stricture not related to surgical complication</td>
</tr>
</tbody>
</table>

Sample collection and serology testing

Blood samples were obtained by venipuncture and collected using standard serum or serum separation tubes at the time of study enrollment. Samples were spun and serum was aliquoted into sterile freezer vials and stored at -80°C for future testing. Serologic assays for a panel of anti-microbial antibodies were carried out at Prometheus.
Laboratories (San Diego, CA) using enzyme linked immunosorbent assay (ELISA) and immunofluorescence, as previously described. An antibody test was considered positive if the titres were greater than the cut off limits as provided by Prometheus: ASCA IgA >20 EU/mL; ASCA IgG >40 EU/mL; pANCA >12.1 EU/mL and pANCA IFA DNAse sensitive; anti-OmpC >16.4 EU/mL; anti-CBir1 >21 EU/mL.

Statistical Analysis

Descriptive statistics were provided with mean, standard deviation and range for continuous variables, and frequency and percentage for categorical variables. Chi-square test and ANOVA were used to detect the differences in clinical and serological variables between the NP, CP, and CDL subgroups. Logistic regression models were applied to estimate the odds ratio (OR) and corresponding 95% confidence interval (CI). We applied multivariate analysis on CP vs. NP and CDL vs. NP to identify important risk factors. The quartile sums of serum markers were explored in conjunction with other clinical factors including age, smoking, race, and pre-colectomy UC disease extent. For all analyses, two-sided tests were used. Results were considered significant if p-value was \( \leq 0.05 \). All statistical analyses were applied using SAS version 9.2 (SAS institute, Cary, NC).

2.4 Results

1427 patients from the MSH Pelvic Pouch Database were contacted and 420 patients with a pre-colectomy diagnosis of UC were included and provided consent to participate in the study. Twenty-one patients were excluded after medical records review: ten patients with FAP, ten with a pre-colectomy diagnosis of IC or IBDU and one due to a pre-colectomy diagnosis of microscopic colitis. The remaining 399 subjects were included in the study and had sufficient medical documentation for classification. Serum samples were provided by 341 of these individuals.

70.7% of the study population met criteria for the NP group, 16.8% for the CP group and 12.5% in the CDL group (Figure 3). Among individuals in the NP group, 60.3% had never experienced an episode of pouchitis, while 39.7% experienced at least
one but less than four episodes per year. The comparison of the clinical characteristics of
the study population by pouch outcome are shown in Table 7. The mean duration of
pouch follow up at the time of study entry was 9.2 ± 6.8 years with no significant
difference between the three groups. More than 90% of the population was Caucasian.
Ashkenazi Jewish ethnicity was significantly associated with pouch outcome (P=0.02)
with those affected by CP more likely to be of Ashkenazi Jewish heritage compared to
those with NP (P<0.008, OR 2.7). First degree family history of IBD was not associated
with outcome. Over 90% of the study population had extensive UC prior to surgery
(Montreal Classification E3). Only 1 of 399 subjects had distal UC (Montreal
Classification E1). There was no significant difference among pouch outcome in terms
of gender, duration of UC prior to colectomy, age at UC diagnosis or age at IPAA. Nine
patients had pre-colectomy evidence of skin tags or anal fissures but were still considered
to have a diagnosis of UC (6 NP, 2 CP, 1 CDL, P=0.91). Ten patients had evidence of
backwash ileitis prior to colectomy (6 NP, 2 CP, 2 CDL; P=0.71).
Figure 3: Proportions of patients (n=399) in each of the three outcome groups
Table 7: Relationship between demographic characteristics and outcome following IPAA. P-values are those resulting from the 3-way comparison.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>No Pouchitis n=282 (70.7%)</th>
<th>Chronic Pouchitis n=67 (16.8%)</th>
<th>Crohn’s Disease-Like n=50 (12.5%)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration from ileostomy closure to serum collection (years (SD))</td>
<td>8.9 (6.9)</td>
<td>9.6 (7.0)</td>
<td>10.1 (5.8)</td>
<td>0.32</td>
</tr>
<tr>
<td>Duration from UC diagnosis to surgery (years (SD))</td>
<td>6.7 (6.9)</td>
<td>7.2 (6.7)</td>
<td>6.2 (7.5)</td>
<td>0.77</td>
</tr>
<tr>
<td>Males, n (%)</td>
<td>136 (48.2%)</td>
<td>37 (55.2%)</td>
<td>22 (44.0%)</td>
<td>0.45</td>
</tr>
<tr>
<td>Mean Age, years (range) at UC diagnosis</td>
<td>30 (10-62)</td>
<td>29 (10-55)</td>
<td>29 (2-55)</td>
<td>0.72</td>
</tr>
<tr>
<td>Mean Age, years (range) at pouch surgery</td>
<td>37 (7-36)</td>
<td>37 (12-59)</td>
<td>36 (8-56)</td>
<td>0.67</td>
</tr>
<tr>
<td>Race/Ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>261 (92.6%)</td>
<td>62 (92.5%)</td>
<td>46 (92.0%)</td>
<td>0.89</td>
</tr>
<tr>
<td>Ashkenazi Jewish*</td>
<td>21 (8.0%)</td>
<td>12 (19.4%)</td>
<td>7 (15.2%)</td>
<td>0.02</td>
</tr>
<tr>
<td>UC extent, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>0</td>
<td>0</td>
<td>1(2.0%)</td>
<td>0.08</td>
</tr>
<tr>
<td>E2</td>
<td>25 (9.1%)</td>
<td>7 (10.4%)</td>
<td>2 (4.0%)</td>
<td>0.08</td>
</tr>
<tr>
<td>E3</td>
<td>251 (90.9%)</td>
<td>60 (89.6%)</td>
<td>47 (94.0%)</td>
<td>0.08</td>
</tr>
<tr>
<td>Smoking Status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoked</td>
<td>156 (55.3%)</td>
<td>31 (46.3%)</td>
<td>18 (36.0%)</td>
<td>0.02</td>
</tr>
<tr>
<td>Ex-smoker at study recruitment</td>
<td>111 (39.4%)</td>
<td>32 (47.7%)</td>
<td>23 (46.0%)</td>
<td></td>
</tr>
<tr>
<td>Smoke at study recruitment</td>
<td>11 (4.0%)</td>
<td>4 (5.9%)</td>
<td>7 (14.6%)</td>
<td></td>
</tr>
<tr>
<td>IBD Family History</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st degree CD</td>
<td>9(3.2%)</td>
<td>4(6.0%)</td>
<td>4(8.0%)</td>
<td>0.22</td>
</tr>
<tr>
<td>1st degree UC</td>
<td>37(13.1%)</td>
<td>10(14.9%)</td>
<td>5(10.0%)</td>
<td>0.73</td>
</tr>
<tr>
<td>2nd degree</td>
<td>38(13.5%)</td>
<td>17(25.4%)</td>
<td>6(12.0%)</td>
<td>0.04</td>
</tr>
<tr>
<td>Time from Ileostomy Closure to Pouch Outcome (years (SD))</td>
<td>NA</td>
<td>2.7 (4.1)</td>
<td>3.9 (5.0)</td>
<td>0.04</td>
</tr>
</tbody>
</table>

* Percentage of Caucasian population that is Ashkenazi Jewish. ◊ Data not available on six patients
In a three-way comparison, there was a statistically significant interaction between smoking history and pouch outcome ($P \leq 0.02$). Patients with a history of smoking (either current at recruitment or ex-smokers) were significantly more likely to have the CDL outcome (62.5%) compared to NP or CP (43.9%, 53.7% respectively; $P=0.003$).

The presence of either anti-CBir1 or ASCA IgG was significantly associated with a CDL phenotype (Figure 4). 53.5% and 14.0% of subjects in the CDL group were positive for anti-CBir1 and ASCA IgG respectively, compared to 21.4% and 3.8% in the NP group and 28.3% and 5.0% in the CP groups ($P<0.0001$ and $P=0.03$). Pairwise comparison between groups showed that anti-CBir1 was significantly associated with CDL when compared to both NP ($P=2.83 \times 10^{-5}$, OR=4.22, 95%CI 2.15-8.28) and CP ($P=0.011$, OR=2.91, 95%CI 1.28-6.61). Combined ASCA (IgA and/or IgG) was significantly associated with CDL outcome when compared to NP ($P=0.046$; OR=2.21, 95%CI 1.01-4.83). However, the relationship between ASCA IgG was much more strongly associated with CDL outcome when compared to NP ($P=0.01$, OR=4.13, 95% CI 1.39-12.27). We also examined the association between the number of positive antibodies and outcome (ASCA IgA and IgG combined, anti-CBir1, anti-OmpC, pANCA) (Table 8). As the number of antibody markers a patient was positive for increased, there was a greater likelihood of having a CDL outcome, both when pANCA was included ($P=4.69 \times 10^{-4}$) and excluded ($P=3.19 \times 10^{-5}$) (Table 8).
Figure 4: Overall levels of antibody positivity among three outcome groups
Table 8: Odds Ratios and Confidence intervals associated with increasing antibody positivity. Included in the analysis were combined ASCA IgA/IgG, anti-OmpC, anti-CBir1.

<table>
<thead>
<tr>
<th></th>
<th>Excluding pANCA</th>
<th></th>
<th>Including pANCA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (CI)</td>
<td>P-value</td>
<td>OR (CI)</td>
<td>P-value</td>
</tr>
<tr>
<td>No Pouchitis vs. Chronic Pouchitis</td>
<td>1.33 (0.91-1.92)</td>
<td>0.12</td>
<td>1.41 (1.03-1.92)</td>
<td>0.029</td>
</tr>
<tr>
<td>No Pouchitis vs. Crohn’s Disease-Like</td>
<td>2.50 (1.64-3.85)</td>
<td>&lt; 0.0001</td>
<td>1.89 (1.33-2.70)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Chronic Pouchitis vs. Crohn’s Disease-Like</td>
<td>1.75 (1.10-2.78)</td>
<td>0.018</td>
<td>1.30 (0.88-1.92)</td>
<td>0.187</td>
</tr>
</tbody>
</table>
Mean antibody titres for each of the markers in the three outcome groups are described as a supplementary figure (Appendix Chapter 2, Supplementary figure 1). In quartile sum analysis, only mean antibody titres for anti-CBir1 were significantly different between outcome groups (p=0.001). In a multivariate model, which included quartile sum of combined serum markers as well as age, smoking, race, and UC extent, it was found that for CP vs. NP, only the quartile sum of serum markers were marginally associated with disease (p=0.05). However, when comparing CDL with NP, both smoking (p=0.0023) and the quartile sum of serum markers (p= 0.0006) were associated with outcome.

The overall seroprevalence rate of pANCA was 44.6% in the entire cohort (n = 341) of patients with serum available. The rate was highest in CP with 55.0% of patients positive for this marker. However there were no significant differences in rates of pANCA positivity among the three outcome groups. Additionally, there were no differences in seroprevalence among any of the markers when comparing individuals with no pouchitis ever with those who had had episodes of acute antibiotic responsive pouchitis (data not shown). Finally, we compared both CP and CDL with the subset of NP individuals who had never experienced pouchitis (n=146) and found that these results were identical to those seen in the comparisons made using the entire NP cohort.

2.5 Discussion

The development of chronic pouchitis and a CD-like phenotype of the pouch following IPAA for UC are among the greatest risk factors for pouch failure following surgery. A model to help identify UC patients at risk for complications following IPAA would be very useful when counseling patients where colectomy is a consideration for management. A number of studies have examined clinical risk factors as predictors of outcome demonstrating that smoking, family history of CD and Ashkenazi Jewish ethnicity are associated with a greater risk of complications following IPAA. Several studies have also shown that the presence of PSC or other immune disorders are associated with poor pouch outcome. In this report, we confirm that Ashkenazi Jewish individuals and those who have a history of smoking (either current or ex-
smokers) were significantly more likely to develop inflammatory pouch complications. We did not observe any association between outcome and first-degree family history of IBD.

Anti-microbial antibodies have also been previously described as important markers associated with pouch outcomes including pouchitis and a CD-like phenotype. Typically, pANCA is associated with acute and chronic pouchitis, and ASCA with a CD-like phenotype after IPAA. Previously published work described a modest association between anti-CBir1 and either acute or chronic pouchitis however, this marker has not previously been associated with a CDL outcome. These observations are in keeping with our findings demonstrating the independent association of anti-CBir1 and ASCA IgG with the CDL outcome. Interestingly, the prevalence of anti-CBir1 among UC-IPAA patients with a CDL phenotype, is similar to rates previously reported in CD patients. Additionally, previous studies have shown that presence of both anti-CBir1 and pANCA together increased the risk of developing pouchitis. Extending these data, our results show that an increasing number of positive antibodies typically found in CD (ASCA, anti-CBir1 and anti-OmpC) are strongly associated with CDL outcome and also with chronic pouchitis when pANCA is included.

Several previous studies have suggested that pANCA is associated with chronic pouchitis and this marker has been associated with UC generally in numerous studies. In our study population 44.6% of subjects were pANCA positive and there were no significant differences in prevalence among the three outcome groups. This rate resembles that previously reported in other UC populations. The pANCA rate found in our general UC population (55.6%, data not published) is slightly higher than that described here, suggesting perhaps some diminution in antibody titre over time following colectomy. Conversely, the rates of anti-OmpC and anti-CBir1 reported in this population appear higher than the usual rates found in a UC population. The prevalence of ASCA in IPAA patients without CDL phenotype is similar to other UC cohorts. The divergent rates of anti-microbial antibody prevalence in this cohort compared to a general UC or CD cohort may reflect a change in response to alterations of the microbial flora in the ileal pouch compared to the flora found in the ileum of a UC patient prior to colectomy. It is
also conceivable that CD-related anti-microbial antibodies are found more frequently in severe UC requiring colectomy as was reported by Fleshner et al.\textsuperscript{67,311}. Further studies sampling serum both before and after colectomy will help to clarify this observation.

Genetic studies also suggest that considerable overlap may exist in early etiologic events leading to UC and CD as illustrated by reports of overlapping CD and UC risk variants\textsuperscript{317-319}. Many such polymorphisms are thought to result in increased intestinal permeability to microbial antigens triggering an adaptive immune response which may be characterized by the production of anti-microbial antibodies. \textit{NOD2} polymorphisms have previously been associated with increases in serum concentration of ASCA, anti-OmpC and anti-CBir1\textsuperscript{148}. The findings that anti-CBir1, ASCA, and a history of smoking, features typically associated with CD\textsuperscript{40,59}, were associated with the CDL phenotype following IPAA in a cohort of patients with a confirmed UC diagnosis prior to surgery, suggests that these individuals may be genetically susceptible to IBD via pathways which are important in both CD and UC phenotypes. Following surgery, such individuals may be more susceptible to a CD-like phenotype in response to changes in the microbiota of the pouch, thereby contributing to changes in anti-microbial antibody titres. We hypothesize that patients with CD-like outcomes after pelvic pouch surgery for UC are not a “missed” diagnosis of CD but rather an evolution of the IBD-phenotype to one that more closely resembles CD rather than UC. To date, few genetic factors have been implicated in the etiology of pouchitis. Polymorphisms in \textit{NOD2}, \textit{IL-1RA} and \textit{TNF} have shown only weak association with pouchitis, and no genetic variants have as yet been associated with the CDL outcome\textsuperscript{71,320,321}. However, serologic evidence such as that described in this study, continues to suggest a role for a breakdown in mucosal and innate immunity in the pathogenesis of both pouchitis and CD of the pouch.

The onset of ileal inflammation in the pelvic pouch model represents a unique opportunity to identify the mechanisms that contribute to intestinal inflammation more generally in IBD. To further characterize this process additional study is required to evaluate the genetic risk variants associated with pouch outcome and the composition of the pouch microbiome, ideally in a prospective fashion. In the interim, these data add to the accumulating literature suggesting that anti-microbial antibodies may be useful in
stratifying patients into groups that are at varying risk for inflammatory pouch complications.
Chapter 3: The NOD2insC polymorphism is associated with worse outcome following ileal pouch-anal anastomosis for ulcerative colitis

This article has been accepted for publication in Gut.

Title: The NOD2insC polymorphism is associated with worse outcome following ileal pouch-anal anastomosis for ulcerative colitis

Authors: Andrea D Tyler\textsuperscript{1,2}, Raquel Milgrom\textsuperscript{2}, Joanne M Stempak\textsuperscript{2}, Wei Xu\textsuperscript{3}, John H Brumell\textsuperscript{1,5,6}, Aleixo M Muise\textsuperscript{1,4,5}, Rishabh Sehgal\textsuperscript{7}, Zane Cohen\textsuperscript{1,2}, Walter Koltun\textsuperscript{7}, Bo Shen\textsuperscript{8}, Mark S Silverberg\textsuperscript{1,2}

Contributorship statement:

ADT implemented the trial in Toronto, data acquisition, analysis and interpretation, and drafting and revising the manuscript. RM data analysis and critical revision of the manuscript. JMS implemented the trial in Toronto and critical revision of the manuscript. WX statistical analysis and revision of the manuscript. JHB critical revision of the manuscript. AM critical revision of the manuscript. ZC critical revision of the manuscript and assisted with patient recruitment. RS implemented the trial for the Hershey site and monitored data collection for that site. WK implemented the trial in Hershey and critical revision of the manuscript. BS implemented the trial in Cleveland and critical revision of the manuscript. MSS initiated the collaborative project, designed data collection tools, implemented the trial in Toronto, and critical revision of the manuscript. ADT and MSS are guarantors.

Acknowledgements:

The authors would like to thank Diane Verbeeten and Lucy T. Zhang for their assistance with the study, and Harden Huang and Brenda O’Connor for maintaining and providing access to the Mount Sinai Hospital Pelvic Pouch Database. This study was supported by a grant from the Crohn’s and Colitis Foundation of Canada. Dr. Silverberg is partially supported by the Gale and Graham Wright Research Chair in Digestive Disease. Partial funding for genotyping was supported by a CIHR operating grant (MOP97756) to AMM and JHB.
3.1 Abstract:

Objective: Inflammatory complications after ileal pouch-anal anastomosis (IPAA) for ulcerative colitis (UC) are common. The aim of this study was to investigate whether genetic factors are associated with adverse pouch outcomes such as chronic pouchitis (CP) and a Crohn’s disease-like (CDL) phenotype. Design: 866 patients were recruited from three centres in North America: Mount Sinai Hospital (Toronto, ON), the Cleveland Clinic (Cleveland, OH) and Penn State Milton S. Hershey Medical Center (Hershey, PA). DNA as well as clinical and demographic information was collected. Subjects were classified into post-surgical outcome groups: no chronic pouchitis (NCP), CP, and CDL phenotype. Results: Clinical and genetic data was available on 714 individuals. 487 (68.2%) were classified as NCP, 118 (16.5%) CP, and 109 (15.3%) CDL. The presence of arthritis or arthropathy (p=0.02), primary sclerosing cholangitis (p=0.009) and duration of time from ileostomy closure to recruitment (p=0.001) were significantly associated with outcome. The NOD2insC (rs2066847) risk variant was the SNP most significantly associated with pouch outcome (p=7.4x10^{-5}). Specifically, it was associated with both CP and CDL when compared to NCP (odds ratio [OR]=3.2 and 4.3 respectively). Additionally, SNPs in NOX3 (rs6557421, rs12661812), DAGLB (rs836518), and NCF4 (rs8137602) were shown to be associated with pouch outcome with slightly weaker effects. A multi-variable risk model combining previously identified clinical (smoking status, family history of IBD), serologic (ASCA IgG, pANCA, and anti-CBir1) and genetic markers was constructed and resulted in an OR of 2.72 (p=8.89x10^{-7}) for NCP vs CP/CDL and 3.22 (p=4.11x10^{-8}) for NCP vs CDL respectively. Conclusion: Genetic polymorphisms, in particular, the NOD2insC risk allele, are associated with chronic inflammatory pouch outcomes among patients with UC and IPAA.
What is already known:
- Several genetic factors have previously been associated with inflammatory pouch outcomes in small cohorts.
- Patients with familial adenomatous polyposis and a pelvic pouch do not typically develop inflammatory pouch complications, suggesting genetic factors are important in pouchitis and a Crohn’s disease-like phenotype among individuals with ulcerative colitis.

What is new in this study:
- The NOD2 insC polymorphism, which is important in innate immune recognition of microbial muramyl dipeptide, is associated with inflammatory pouch outcomes.
- Additional genes which are mediators in the generation of reactive oxygen species may also have a role in pouch inflammation

Possible clinical implications:
- These polymorphisms may be useful for predicting which patients are at an increased risk of developing chronic pouch inflammatory complications following surgery for ulcerative colitis
- A combination of clinical, serologic and genetic factors may be a useful clinical tool for assisting in decision making for patients considering a colectomy and pelvic pouch for UC
3.2 Introduction:

Ulcerative colitis (UC) is a chronic inflammatory disorder of unknown etiology. Approximately 20% of affected individuals will require colectomy at some point in their disease course. Restorative proctocolectomy with ileal-pouch anal anastomosis (IPAA) is the surgical procedure of choice in the setting of fulminant or chronic, treatment-refractory UC and UC-associated colonic neoplasia. However, de novo inflammation of the ileal reservoir (pouchitis) is a common post-surgical complication with prevalence rates ranging from 12% to greater than 50%. An additional subset of patients (up to 17%) will develop a Crohn’s Disease (CD)-like phenotype characterized by abdominal or perianal fistulae or inflammation of the small bowel proximal to the pouch (afferent limb). The causes of ileal inflammation in the pouch in patients with a preoperative diagnosis of UC are unknown. Clearly the evolution of the microbiome of the pouch after closure of a diverting ileostomy plays a role, as patients rarely develop pouchitis prior to restoration of continuity of the fecal stream. The decreased risk of inflammation among those with familial adenomatous polyposis (FAP), and reports by some groups that a family history of CD may increase the risk for the CDL phenotype in the pouch, suggests that heritable susceptibility may be a critical factor in disease pathogenesis. However, previous genetic studies have been small in scale and have not yielded consistently reproducible results. Among others, variants in IL-1RN and TLR1 have been associated with pouch inflammation. Additional studies have tentatively linked NOD2 variants with chronic pouchitis. The aim of this study was to evaluate whether IBD susceptibility polymorphisms which have been previously implicated in CD, UC or pouchitis are associated with chronic pouch inflammatory outcomes following surgery for UC in a large IPAA cohort.

3.3 Methods:

Study Population
All study protocols were carried out in accordance with the research ethics boards at each center and informed consent was obtained from all patients prior to their enrolment. Patients having undergone colectomy with IPAA at Mount Sinai Hospital (MSH), Cleveland Clinic and the Penn State Milton S. Hershey Medical Center (HMC) were contacted by research staff between 2007 and 2010. The demographic and clinical features of subjects from MSH have been previously reported and described in detail\textsuperscript{329}, and the \textit{NOD2} genotype status has been previously reported on the HMC cohort\textsuperscript{321}. Included in the study were any patients with a confirmed pre-colectomy diagnosis of UC and who had had their ileostomy closed a minimum of one year prior to study enrolment. UC diagnosis was confirmed based on clinical, endoscopic and pathologic evidence – particularly from the colectomy specimen. Those having undergone their procedure outside of any of these institutions were included only in the event that medical documentation was available. Patients with CD, inflammatory bowel disease of the colon – type unclassified (IBD-U) or indeterminate colitis (IC) based on pre-colectomy medical chart review or colectomy surgical pathology, were excluded.

\textit{Clinical Data Collection}

Study data were obtained through detailed retrospective chart review and patient questionnaire with investigators blinded to patients’ genetic results. To ensure uniformity between sites, rigorous definitions were agreed upon and applied to classify patients into outcome groups. Data collected directly by patient interview included gender, age, family history of IBD, smoking status, and clinical symptoms of pouch function following surgery (daily number of bowel movements, incontinence, presence or absence of blood in stool). UC diagnosis date, surgical history and dates, presence of any extraintestinal manifestations of IBD such as arthritis or arthropathy, osteoporosis/osteopenia, erythema nodosum, pyoderma gangrenosum, primary sclerosing cholangitis (PSC) or ocular inflammation, as well as the post-surgical outcome were confirmed through medical chart review. Pre-colectomy disease extent was classified based on the Montreal classification\textsuperscript{5}. The presence of backwash ileitis (inflammation of the terminal ileum in individuals with pan-colitis) pre-colectomy was documented and patients were included only in the event that all other findings were consistent with UC.
Post-Surgical Classification

Patients were classified into one of three outcome groups based on clinical, endoscopic and histologic factors. The “no chronic pouchitis” (NCP) group included individuals without an episode of pouchitis, as well as those who experienced fewer than four clinical acute pouchitis episodes per year, each responding to two weeks or less of antibiotics (ciprofloxacin, metronidazole, or combination). This clinical definition of acute pouchitis has been previously described\(^2\). Typically, these subjects may not have had a full clinical assessment but clearly represent a distinct category as demonstrated by their immediate response to antibiotics. The “chronic pouchitis” (CP) group included antibiotic-dependent and antibiotic-refractory patients who required either prolonged (>1 month) antibiotic therapy, required medical intervention for pouchitis more than three times per year, or required the use of second- or third-line medications (5-ASA, steroids, immunomodulators, biologics). All such subjects had endoscopic assessment at some point during their disease course\(^2\). The final group was the “Crohn’s disease-like” phenotype (CDL). Individuals classified into this group met at least one of the following criteria: a) development of a perianal fistula more than one year after ileostomy closure documented through physical examination, examination under anesthesia or imaging; b) development of a stricture proximal to the pouch which was not related to a surgical complication and was confirmed by endoscopy or small bowel imaging; c) evidence of inflammation (ulceration, erythema, friability) extending above the pouch inlet and into the afferent limb/pre-pouch ileum or more proximal small intestine detected on pouchoscopy or upper endoscopy. Anastomotic ulceration or ulceration around the pouch inlet alone was not sufficient to classify patients into the CDL outcome group. Additional post-surgical outcomes including mechanical or surgical complications or cuffitis were also documented\(^2\). Time to diagnosis of pouchitis was defined as time from ileostomy closure to the time of onset of symptoms and diagnosis.

DNA Collection and Genotyping

3-6mL of whole blood was obtained by venipuncture using standard EDTA collection tubes at both MSH and HMC. DNA was extracted using the QIAGEN Gentra Puregene.
Blood Kit. DNA samples from the Cleveland Clinic were extracted from clotted blood using the Maxwell 16 Tissue DNA purification kit according to manufacturer’s protocol. DNA from all study sites was then stored in sealed Matrix screw top tubes at 4°C prior to spectrophotometric quality and quantity check using the NanoDrop 1000 (Thermo Scientific). The majority of DNA samples yielded concentrations ranging from 100-300 μg with all concentrations above 20ng/μL. Those falling below 18 ng/μL were whole genome amplified using the Illustra GenomiPhi HY DNA Amplification Kit (GE Healthcare). Single nucleotide polymorphism (SNP) genotyping was performed using the Illumina Goldengate custom SNP assay on the Illumina BeadStation500G (San Diego, CA) at The Center for Applied Genomics (TCAG, Toronto, Canada) and the Sequenome iPLEX platform (Génome Québec, Montreal, Canada). NOD2insC genotyping was performed using the TaqMan SNP Genotyping platform (TCAG).

Statistical Analysis and Quality Control

Descriptive statistics were reported as mean and range for continuous variables and frequencies and proportions for categorical variables. For phenotypic results, Fisher’s exact test and Pearson’s chi-square test were used to compare proportions; non-parametric tests were used to compare the continuous variables. PLINK version 1.06 was used to obtain descriptive statistics of the SNPs such as the allele frequency, genotype distribution and to test for Hardy-Weinberg equilibrium (HWE) for each marker based on Pearson’s chi-square test. All SNPs which were successfully genotyped in at least 95% of the study cohort, satisfied HWE criteria (p HWE>0.001) and had a minor allele frequency >0.01, were included in the analysis. To ensure that results were not due to population stratification, only Caucasian individuals were included in the genetic analysis. Logistic regression models were applied for the association analysis. Although an additive genetic model was used for the primary analysis we also explored dominant and recessive genetic models. Throughout this report the p-values are those obtained from the additive genetic model unless otherwise stated. Odds ratios (OR) and 95% confidence intervals (CI) were calculated. Two-sided statistical tests were applied and all analyses were performed with SAS 9.2 (SAS Institute, Cary, NC, USA).
A list of 646 SNPs was generated and chosen for the analysis in the initial cohort based on their previous known associations with CD, UC or pouchitis. To accommodate the Sequenom platform, the top 46 hits from the initial analysis as well as 20 previously described important CD/UC/pouchitis associated SNPs were selected for genotyping in the replication cohort (Appendix: Chapter 3, Supplementary table 1). Due to the large number of tests performed, nominal significance was defined in both the preliminary analysis and the combined cohort as \( p < 0.01 \). Stringent Bonferroni correction was applied to the final \( p \)-value to adjust for multiple comparisons with \( p < 7.5 \times 10^{-4} \) required to declare significance.

Finally, multivariate analysis was performed with factors previously shown by our group and others to be associated with pouch outcome. These included the serologic markers anti-saccharomyces cerevisiae antibody (ASCA) IgG, perinuclear anti-neutrophil cytoplasmic antibody (pANCA) and anti-CBir1\(^{329}\); clinical factors included smoking and family history of IBD; and the SNPs which were significantly associated with outcome in our Caucasian cohort. This analysis was performed in a subset of 341 patients in whom all of the preceding data was available. A step-wise procedure was performed to generate a risk score based on the significant risk factors from the multivariate analysis. The risk score was based on the weighted combination of the risk factors with the standardized logistic regression coefficient as the weight\(^{332}\). Receiver operating characteristic (ROC) curves were generated to calculate the area under the curve (AUC) and the sensitivity and specificity of the risk score\(^{332}\).

3.4 Results:

*Study Cohort*

The initial cohort consisted of 399 patients recruited from MSH and the second included 467 patients recruited from MSH, Cleveland Clinic and HMC, for a total cohort of 866. Within both the first and second cohorts, 339 and 401 respectively had both genotype and phenotype information available for a total genetic cohort of 740. Among these individuals 714 (96.5\%) were Caucasian and only these subjects were subsequently analyzed.
Clinical Variables Associated with Pouch Outcome

Clinical and phenotypic characteristics of the study population are shown in Table 9. A description of the pouch outcome groups among the combined Caucasian cohort is shown in Figure 5. There were 487 (68.2%) NCP, 118 (16.5%) CP, and 109 (15.3%) CDL. The mean age at UC diagnosis (29, range 2-59 years) and at IPAA (37, range 7-61 years) was the same between groups. Within the CDL group, 56.6% were diagnosed based on the presence of a fistula or abscess developing more than one year following surgery, 36.4% based on inflammation extending into the afferent limb, and 7.0% met both criteria.

Factors previously related to pouch outcome such as smoking and family history of IBD were not found to be significantly associated in this cohort. However, the presence of large-joint arthritis (p=0.02) and PSC (p=0.009) were significantly associated with both the CP and CDL outcomes. A longer duration of time from ileostomy closure to study enrolment was associated with a worse pouch outcome (p=0.001). The mean time to diagnosis of CP and CDL was 2.4 and 4.4 years respectively, well below the mean post-surgical follow-up time for any of the outcome groups.
Table 9: Clinical and phenotypic characteristics of the study population.

<table>
<thead>
<tr>
<th>Feature</th>
<th>No Chronic Pouchitis (n=487)</th>
<th>Chronic Pouchitis (n=118)</th>
<th>Crohn’s Disease-Like (n=109)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (% female)</td>
<td>47.0</td>
<td>43.2</td>
<td>50.5</td>
<td>0.55</td>
</tr>
<tr>
<td>Mean age at UC diagnosis (years, range)</td>
<td>30 (7-59)</td>
<td>30 (11-55)</td>
<td>28 (2-55)</td>
<td>0.85</td>
</tr>
<tr>
<td>Mean age at surgery (years, range)</td>
<td>37 (7-61)</td>
<td>37 (12-59)</td>
<td>36 (8-56)</td>
<td>0.84</td>
</tr>
<tr>
<td>Ashkenazi Jewish (% Jewish)</td>
<td>6.0</td>
<td>10.2</td>
<td>8.3</td>
<td>0.23</td>
</tr>
<tr>
<td>Smoking (n=704)</td>
<td></td>
<td></td>
<td></td>
<td>0.72</td>
</tr>
<tr>
<td>Never</td>
<td>59.5</td>
<td>58.5</td>
<td>54.1</td>
<td></td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>33.3</td>
<td>34.7</td>
<td>33.9</td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>5.7</td>
<td>6.8</td>
<td>9.2</td>
<td></td>
</tr>
<tr>
<td>Family history of IBD (%)</td>
<td>9.8</td>
<td>4.2</td>
<td>11.0</td>
<td>0.16</td>
</tr>
<tr>
<td>Duration of pouch (years) Mean (SD) (n=665)</td>
<td>8.6 (6.4)</td>
<td>9.1 (6.3)</td>
<td>11.0 (6.3)</td>
<td>0.001</td>
</tr>
<tr>
<td>Arthritis or arthropathy (%) (n=447)</td>
<td>10.6</td>
<td>19.4</td>
<td>22.2</td>
<td>0.02</td>
</tr>
<tr>
<td>Primary Sclerosing Cholangitis (%) (n=617)</td>
<td>1.4</td>
<td>6.7</td>
<td>4.2</td>
<td>0.009</td>
</tr>
</tbody>
</table>
Figure 5: Proportion of individuals in each specific outcome group.
Association of SNPs with Pouch Outcome

Of the 646 SNPs selected for analysis in the initial cohort of subjects (n = 369), 66 SNPs which met initial criteria for significance or were SNPs of importance in prior IBD genetic studies were genotyped in the additional cohort (n = 345). Of these, 12 had a low call rate and were not included in the analysis. Of the remaining 54 SNPs, five were significantly associated with outcome at a p-value threshold of p<0.01 in a three way comparison. However, only the \(NOD2\) insertion variant (\(NOD2\ insC\), rs2066847) remained significantly associated with pouch outcome after stringent Bonferroni correction (p=7.4x10^{-5}; \(p_{corr}=4.9x10^{-3}\)) (Table 10). Specifically, the \(NOD2insC\) variant was detected significantly more frequently among individuals with CP and CDL compared to those with NCP, with OR of 3.21 (CI=1.38-7.47) and 4.30 (CI=1.90-9.77), respectively (Table 11). This association remained significant even when the previously reported samples obtained from HMC were excluded.
Table 10: Single nucleotide polymorphisms (SNPs) associated with outcome at a nominal significance threshold of p<0.01. Bonferroni corrected p-values are also listed. The additive genetic model was applied except where indicated (*dominant genetic model).

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Cohort 1 (n=369)</th>
<th>Cohort 2 (n=345)</th>
<th>Combined Cohort (n=714)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>p-value</td>
<td>MAF NCP</td>
</tr>
<tr>
<td>NOD2insC</td>
<td>rs2066847</td>
<td>0.09</td>
<td>5.0x10^{-3}</td>
<td>0.01</td>
</tr>
<tr>
<td>NOX3</td>
<td>rs6557421</td>
<td>0.03</td>
<td>0.04</td>
<td>0.26</td>
</tr>
<tr>
<td>NOX3*</td>
<td>rs12661812</td>
<td>&gt;0.1</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>DAGLB</td>
<td>rs836518</td>
<td>0.02</td>
<td>3.3x10^{-3}</td>
<td>0.22</td>
</tr>
<tr>
<td>NCF4*</td>
<td>rs8137602</td>
<td>9.3 x10^{-4}</td>
<td>&gt;0.1</td>
<td>0.09</td>
</tr>
</tbody>
</table>

MAF= minor allele frequency. NCP = no chronic pouchitis, CP = chronic pouchitis, CDL = Crohn’s disease like.
Table 11: Odds ratios (OR) and confidence intervals (CI) for the five single nucleotide polymorphisms (SNPs) which are associated with outcome. (*dominant genetic model).

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>CP vs NCP</th>
<th>CDL vs NCP</th>
<th>CDL vs CP</th>
<th>CP/CDL vs NCP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>OR</td>
<td>CI</td>
<td>OR</td>
<td>CI</td>
</tr>
<tr>
<td>NOD2insC</td>
<td>rs2066847</td>
<td>3.21</td>
<td>1.38-7.47</td>
<td>4.30</td>
<td>1.90-9.77</td>
</tr>
<tr>
<td>NOX3</td>
<td>rs6557421</td>
<td>0.55</td>
<td>0.36-0.82</td>
<td>0.56</td>
<td>0.37-0.87</td>
</tr>
<tr>
<td>NOX3*</td>
<td>rs12661812</td>
<td>1.25</td>
<td>0.71-2.19</td>
<td>1.09</td>
<td>0.59-2.00</td>
</tr>
<tr>
<td>DAGLB</td>
<td>rs836518</td>
<td>0.93</td>
<td>0.63-1.38</td>
<td>0.45</td>
<td>0.27-0.74</td>
</tr>
<tr>
<td>NCF4*</td>
<td>rs8137602</td>
<td>1.01</td>
<td>0.58-1.74</td>
<td>1.10</td>
<td>0.63-1.93</td>
</tr>
</tbody>
</table>

NCP = no chronic pouchitis, CP = chronic pouchitis, CDL = Crohn’s disease like.
Given the low minor allele frequency of the \textit{NOD2}\textsubscript{insC} allele in our cohort, to fully assess the risk associated with variants in \textit{NOD2} and pouch related inflammatory complications, we examined the risk associated with compound heterozygosity or homozygosity for combinations of markers (rs2066847, rs2066845 and rs2066844). We found that increased numbers of variants at this loci were associated with increased risk of CDL when compared to NCP, (p=.002; OR=2.08; CI=1.31-3.50), however, no increased risk was observed when NCP was compared to CP.

Additional SNPs associated with outcome in the three-way analysis (p<0.01) were variants in \textit{NOX3} (rs6557421) and \textit{DAGLB} (rs836518), however, these did not remain significant after correction for multiple testing (Table 10). In the pairwise analysis, rs6557421 trended towards being protective for both of the inflammatory outcomes and rs836518 was protective against only the CDL phenotype (Table 11). Additional variants in \textit{NOX3} (rs12661812) and \textit{NCF4} (rs8137602) were found to be associated with outcome (p<0.01) using the dominant genetic model, with variants at both loci more common in individuals with inflammatory pouch phenotypes. However, these polymorphisms did not remain significantly associated with outcome after correction for multiple testing.

We next conducted analyses in which we compared the NCP group with the combined CP/CDL cohort (n\textsubscript{comb}=227). We found that both rs2066847 (\textit{NOD2}\textsubscript{insC}) and rs6557421 (\textit{NOX3}) were associated with outcome with the minor allele of rs2066847 (p=3.3x10\textsuperscript{-4}) increasing disease risk and that of rs6557421 (p=2.8x10\textsuperscript{-4}) protective against inflammatory phenotypes (Table 11).

Variants which had been previously associated with pouch complications, including those in \textit{TLR1} \textsuperscript{328}, \textit{CD14}, \textit{TLR9} \textsuperscript{326} as well as other \textit{NOD2} variants, were not found to be associated with outcome among our cohort (Table 12). \textit{IL1RN}\textsuperscript{527} trended towards significance in our initial analysis, however did not remain significant after correction for multiple testing. Additionally, many of the other well known IBD-associated genetic polymorphisms including those specific to CD (i.e. ATG16L1) and those associated with both CD and UC (i.e. IL23R), were not associated with any of the outcome groups.
Table 12: Single nucleotide polymorphisms (SNPs) previously associated with pouch outcome. IL1RN did not meet specified quality criteria for combined analysis (successfully genotyped in <95% of the cohort).

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>MAF</th>
<th>p-value</th>
<th>p-value corr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NCP</td>
<td>CP</td>
<td>CDL</td>
</tr>
<tr>
<td><em>TLR1</em>&lt;sup&gt;328&lt;/sup&gt;</td>
<td>rs4833103</td>
<td>0.47</td>
<td>0.43</td>
<td>0.51</td>
</tr>
<tr>
<td><em>CD14</em>&lt;sup&gt;326&lt;/sup&gt;</td>
<td>rs2569190</td>
<td>0.49</td>
<td>0.52</td>
<td>0.40</td>
</tr>
<tr>
<td><em>NOD2 (SNP8)</em>&lt;sup&gt;321&lt;/sup&gt;</td>
<td>rs2066844</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td><em>NOD2 (SNP12)</em>&lt;sup&gt;321&lt;/sup&gt;</td>
<td>rs2066845</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td><em>IL1RN</em>&lt;sup&gt;333*&lt;/sup&gt;</td>
<td>rs419598</td>
<td>0.25</td>
<td>0.20</td>
<td>0.35</td>
</tr>
<tr>
<td><em>TLR9</em>&lt;sup&gt;326*&lt;/sup&gt;</td>
<td>rs352140</td>
<td>0.457</td>
<td>0.473</td>
<td>0.513</td>
</tr>
</tbody>
</table>

* replication genotyping did not pass quality control measures. MAF=minor allele frequency. NCP = no chronic pouchitis, CP = chronic pouchitis, CDL = Crohn’s disease like
Multivariate analysis and risk score

In the preliminary multivariate analysis, five factors were found to independently associate with outcome and were included in the subsequent multivariate analysis, including rs2066847 (NOD2 insC), rs6557421 (NOX3), rs836518 (DAGLB), anti-CBir1 and smoking (Table 13). These factors were used to generate a weighted risk score, with the comparison of both NCP with CDL (p=4.11\times10^{-8}; OR=3.22, CI=2.12-4.89) and NCP with combined CP and CDL (p=8.89\times10^{-7}; OR=2.72; CI=1.89-3.91) resulting in highly significant associations. The sensitivity and specificity were assessed by generating ROC curves. The best model was that comparing CDL with NCP and containing all three genetic markers, as well as the serologic and clinical factors (sensitivity = 80.0%; specificity = 70.3%). This model performed significantly better than did those generated using smoking alone, or smoking and anti-CBir1 (Figure 6).
<table>
<thead>
<tr>
<th>Factor</th>
<th>CP vs. NCP OR (CI)</th>
<th>P-value</th>
<th>CDL vs. NCP OR (CI)</th>
<th>p-value</th>
<th>CDL vs. CP OR (CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOD2insC*</td>
<td>2.62 (0.39-17.64)</td>
<td>0.32</td>
<td>13.51 (2.00-90.9)</td>
<td>0.008</td>
<td>2.99 (0.45-20.0)</td>
<td>0.26</td>
</tr>
<tr>
<td>NOX3*</td>
<td>0.54 (0.31-0.94)</td>
<td>0.03</td>
<td>0.35 (0.15-0.78)</td>
<td>0.01</td>
<td>0.62 (0.24-1.59)</td>
<td>0.32</td>
</tr>
<tr>
<td>DAGLB*</td>
<td>1.10 (0.69-1.74)</td>
<td>0.70</td>
<td>0.29 (0.12-0.67)</td>
<td>0.004</td>
<td>0.25 (0.10-0.66)</td>
<td>0.005</td>
</tr>
<tr>
<td>anti-CBir1 (+/-)</td>
<td>2.30 (0.84-3.21)</td>
<td>0.15</td>
<td>6.38 (2.81-14.43)</td>
<td>&lt;0.0001</td>
<td>3.48 (1.36-8.90)</td>
<td>0.009</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current smoking vs. Never smoked</td>
<td>2.30 (0.55-9.71)</td>
<td>0.38</td>
<td>11.96 (3.07-46.53)</td>
<td>0.0002</td>
<td>5.68 (1.10-29.25)</td>
<td>0.04</td>
</tr>
<tr>
<td>Ex-smoker vs. Never smoked</td>
<td>1.52 (0.83-2.77)</td>
<td>0.99</td>
<td>2.20 (0.95-5.10)</td>
<td>0.30</td>
<td>1.16 (0.44-3.05)</td>
<td>0.18</td>
</tr>
</tbody>
</table>

*: For the genetic markers an additive genetic model to compare homozygote recessive vs. heterozygote vs. homozygote dominant genotypes was applied.

NCP = no chronic pouchitis, CP = chronic pouchitis, CDL = Crohn’s disease like.
OR=odds ratio; CI=confidence interval.
Figure 6: ROC curves generated for the risk score analysis.

\( NOD2 \text{insC} \) - rs2066847; \( NOX3 \) - rs6557421; \( DAGLB \) - rs836518
3.5 Discussion

Our results demonstrate that phenotypic characteristics, including PSC and arthritis, as well as several SNPs are associated with pouch outcome following IPAA in individuals with UC. PSC has previously been associated with chronic pouchitis as have general autoimmune disorders including arthritis. The co-occurrence of additional inflammatory disorders previously documented in CD and UC, such as those described above, with pouch complications suggest overlapping disease mechanisms may be important in the onset and propagation of inflammatory outcomes. Interestingly, we did not observe any association between family history of IBD or smoking and pouch outcome. The sample size of this study, which was approximately double that of previous studies, provided us with increased power to detect true associations. Our data describing no correlation between IBD family history or smoking and outcome, suggests that these previously described associations may have been the result of Type I error.

We have shown that several genetic polymorphisms are associated with pouch inflammation, confirming that indeed host genetic factors are critical in the etiology of both pouchitis and the Crohn’s disease-like phenotype. The well established \textit{NOD2}insC polymorphism, which has previously been implicated in ileal CD and which results in a truncation of the leucine rich repeat region of the NOD2 protein, is associated with chronic pouchitis and a Crohn’s disease-like outcome in our large cohort. While the precise effect of this polymorphism is unknown, gene knockout studies have demonstrated that loss of \textit{NOD2} is associated with reduced ability to detect microbial pathogens. Further, \textit{NOD2} variants have been implicated in intestinal allograft rejection and graft versus host disease following allogeneic stem cell transplantation. Our findings demonstrating the association between \textit{NOD2} and pouch complications, confirm results observed in several smaller cohorts, and suggests that studies failing to detect significant associations were likely underpowered. This growing body of evidence demonstrates that pathways which are important in ileal CD may also be critical in inflammation of the ileal pouch following surgery.

It is interesting to note that the \textit{NOD2}insC variant is typically only rarely found in UC. In the MSH UC patient population, for example, the \textit{NOD2}insC allele frequency is 0.014 which is comparable to the reported allele frequency in this study for those without
inflammatory pouch complications\textsuperscript{340}. However in the 227 subjects with CP or CDL, the \textit{NOD2}insC allele frequency is over 0.05. These patients were all carefully chart reviewed to exclude any clinical or pathological evidence of CD or IBDU prior to colectomy. Additionally, the colectomy pathology was all carefully reviewed and there were no features of CD seen in these subjects. These data therefore suggest that the utility of traditional clinical and histologic diagnostic classification and phenotyping is not sufficient to categorize subjects in terms of prognosticating outcome after pelvic pouch surgery. That is, a diagnosis of clinical UC may not imply that pouch complications are less likely to arise but rather, the subject’s genotype may be a more important determinant of outcome.

Other variants which demonstrated modest association with outcome include those which are located in the non-coding regions of \textit{NOX3} and \textit{DAGLB} or adjacent to \textit{NCF4}. \textit{NCF4}, encoding p40phox which is a component of the NADPH oxidase complex, was recently identified in a GWAS as an ileal CD susceptibility gene\textsuperscript{341}. Mechanistic studies have demonstrated that CD patients who are carriers of risk alleles at this locus have significantly decreased amounts of reactive oxygen species - important for host innate immunity and defense from microbial pathogens - generated from granulocyte-macrophage colony-stimulating factor-primed neutrophils compared to patients not carrying these mutations\textsuperscript{342}. \textit{NOX3}, another component of this complex in various tissue types, is also important in the production of reactive oxygen species \textsuperscript{343, 344}. Together, these findings suggest an important role for this pathway in the pathogenesis of not only pouch inflammatory outcomes but IBD in general. The function of \textit{DAGLB}, and a potential role for it in IBD pathogenesis remains unclear, however, the SNP in this gene is in a region of high linkage disequilibrium with RAC1; a gene involved in host immune defence and which has been previously associated with UC\textsuperscript{345}.

Other variants which have been previously associated with pouch outcome among smaller cohorts were not associated with outcome in our group. A possible explanation for the difference between our results and others is that this study has a much larger sample size, which would help to reduce Type I error. Additionally, population stratification may account for different results as some previous studies did not control for ethnicity \textsuperscript{328, 333}. To limit these effects, only Caucasian individuals were analyzed in
our study. We also applied very stringent statistical correction to our data in order to reduce the likelihood of reporting false positives, which may have been too stringent to allow detection of weaker associations. The four variants discussed in this paper which did not remain significant after multiple testing correction, may therefore warrant further investigation in an additional cohort to definitively assess their importance in pouchitis pathogenesis. This would ideally be attempted in the setting of the large International IBD Genetics Consortium Cohort but would require detailed phenotyping of pre-colectomy data as well as of pouch outcomes, both of which may not be readily available in existing databases.

We hypothesize that patients with NOD2\textsuperscript{insC} may be more likely to develop inflammatory pouch complications due to alteration in the microbial composition of the pouch mucosa. This hypothesis is supported by our data showing that UC patients with a CDL phenotype after pelvic pouch surgery are more likely to be positive for ASCA and anti-CBir1\textsuperscript{329}, serologic markers more typically associated with CD, and by other groups’ data demonstrating a strong association between the common UC associated serum marker, pANCA, and pouch outcome\textsuperscript{67, 68}. Additional evidence that NOD2 variants affect the composition of the ileal flora suggests that there may be a relationship between IBD associated genetic polymorphisms and the presence of serum antibodies\textsuperscript{337 148}. Our findings of the presence of both serologic markers of CD and genetic polymorphisms which have been previously associated with ileal CD showing association with pouch outcome, suggests that the mechanism of ileal inflammation which leads to a CP or CDL outcome in the pouch may be similar to those which lead to the development of inflammation within the ileum. Additionally, the inclusion of individuals with non-chronic pouchitis in the NCP group, and the lack of evidence for any associations between genotype and this outcome, suggests that the chronic inflammatory phenotypes proceed via diverging mechanistic pathways from those which lead to antibiotic responsive pouchitis. Furthermore, the overlap in genetic susceptibility between CP and CDL suggests that distinction between these outcomes may be of limited consequence compared to the more clinically relevant phenotype of chronic, medication refractory inflammation which is characteristic of both disorders and results from a common etiology.
While it is conceivable that these CDL patients were misclassified prior to colectomy, the stringent phenotypic classification which was used to determine which patients were included in this analysis would suggest that this is unlikely to be the case. Rather, our results combined with recent data published by Waterman et al. showing few differences in the prevalence of several IBD-associated SNPs between those with UC and those with CD\textsuperscript{317}, supports the concept of IBD as a mosaic of inflammatory disorders. The varying phenotypes associated with overlapping genetic susceptibility loci suggests that while genetic predisposition may be a key inflammatory mediator, other factors such as bacteria or as yet unknown environmental stimuli are necessary in order for immune dysregulation and inflammation to occur.

The etiology of pouch complications remains unknown, however, the data presented in this paper emphasizes the similarities between pouch inflammation and IBD in general, and suggests that the pelvic pouch model may be useful for evaluating factors which contribute to \textit{de novo} inflammation. It will be important to evaluate the impact of the pouch microbiome and the impact of host genotype on the composition of the microbiome to fully understand the mechanisms of pouch inflammation. Our data also demonstrate that a model for assessing risk of pouch complications which encompasses genetic and serologic factors may be a more useful tool than current clinical assessment.
Chapter 4: Characterization of the gut-associated microbiome in inflammatory pouch complications following ileal pouch-anal anastomosis.

Accepted May 13, 2013: PLOS One.

Authors: Andrea D. Tyler\textsuperscript{1,2}, Natalie Knox\textsuperscript{3}, Boyko Kabakchiev\textsuperscript{1,2}, Raquel Milgrom\textsuperscript{2}, Richard Kirsch\textsuperscript{5}, Zane Cohen\textsuperscript{1,2}, Robin S. McLeod\textsuperscript{2}, David S. Guttman\textsuperscript{6}, Denis O. Krause\textsuperscript{3,4}, Mark S. Silverberg\textsuperscript{1,2}

Author contributions:
ADT: Generation, collection, assembly, analysis and interpretation of data; statistical plan and analysis; drafting and revision of the manuscript. NK: assembly, analysis and interpretation of data; revision of the manuscript. BK: analysis and interpretation of data; statistical analysis; revision of the manuscript. RM: assembly of data and manuscript revision. RK: collection, assembly, analysis and interpretation of data. ZC: conception of the study; revision of the manuscript. RSM: conception of the study; revision of the manuscript. DSG: interpretation of data; revision of the manuscript. DOK: conception and design of the study; generation, assembly of data. MSS: conception and design of the study; drafting or revision of the manuscript; approval of the final version of the manuscript. ADT and MSS are guarantors.

Acknowledgements: The authors would like to thank Joanne Stempak, Lucy T. Zhang, Tsedey Legesse, Rachel Caplan, Fuqin Zhu and other members of the Silverberg and Krause labs for their assistance with the study. Additional thanks to Hillary A. Steinhart, Gordon R. Greenberg, Kenneth Croitoru, Geoff C. Nguyen, Alvin Newman, Helen M. MacRae, Robert Gryfe, Robert H. Riddell at the MSH Department of Gastroenterology and Pathology and research fellows Smita Halder and Matti Waterman at MSH for aiding in patient recruitment, and to Harden Huang and Brenda O’Connor for maintaining and
providing access to the Mount Sinai Hospital Pelvic Pouch Database. This study was supported by a grant from the Crohn’s and Colitis Foundation of Canada. Dr. Silverberg is partially supported by the Gale and Graham Wright Research Chair in Digestive Disease.
4.1 Abstract:

Introduction: Inflammatory complications following ileal pouch-anal anastomosis (IPAA) for ulcerative colitis (UC) are common and thought to arise through mechanisms similar to de novo onset inflammatory bowel disease. The aim of this study was to determine whether specific organisms in the tissue-associated microbiota are associated with inflammatory pouch complications.

Methods: Patients having previously undergone IPAA were recruited from Mount Sinai Hospital. Clinical and demographic information were collected and a pouchoscopy with biopsy of both the pouch and afferent limb was performed. Patients were classified based on post-surgical phenotype into four outcome groups: familial adenomatous polyposis (FAP) controls, no pouchitis, pouchitis, and Crohn’s disease-like (CDL). Pyrosequencing of the 16S rRNA V1-V3 hypervariable region, and quantitative PCR for bacteria of interest, were used to identify organisms present in the afferent limb and pouch. Associations with outcomes were evaluated using exact and non-parametric tests of significance.

Results: Analysis at the phylum level indicated that Bacteroidetes were detected significantly less frequently ($P<0.0001$) in the inflammatory outcome groups (pouchitis and CDL) compared to both FAP and no pouchitis. Conversely, Proteobacteria were detected more frequently in the inflammatory groups, although without reaching the same levels of significance. At the genus level, organisms associated with outcome were detected less frequently among the inflammatory groups compared to those without inflammation. Several of these organisms, including Bacteroides ($P_{\text{both sites}}<0.0001$), Parabacteroides ($P_{\text{pouch}}=1.2\times10^{-4}, P_{\text{afferent limb}}=2.2\times10^{-3}$), Blautia ($P_{\text{pouch}}=3.0\times10^{-3}, P_{\text{afferent limb}}<0.0001$) and Sutterella ($P_{\text{pouch}}=2.5\times10^{-3}, P_{\text{afferent limb}}=1.7\times10^{-3}$), were associated with outcome in both the pouch and afferent limb. These associations remained significant even following adjustment for antibiotic use, smoking, country of birth and gender. Individuals with quiescent disease receiving antibiotic therapy displayed similar reductions in these organisms as those with active pouch inflammation.

Conclusions: Specific genera are associated with inflammation of the ileal pouch, with a reduction of typically ubiquitous organisms characterizing the inflammatory phenotypes.
4.2 Introduction

Recent studies have implicated microorganisms in the etiology of several chronic conditions including inflammatory bowel disease (IBD)\textsuperscript{169}, diabetes\textsuperscript{171} and obesity\textsuperscript{173}. The evidence for a role of microorganisms in IBD, both Crohn’s disease (CD) and ulcerative colitis (UC), is compelling: numerous polymorphisms in genes associated with innate and adaptive immunity, as well as those associated with barrier function, have been associated with IBD, as have anti-microbial antibodies\textsuperscript{52, 121, 346}. Additionally, some studies have implicated adherent invasive \textit{Escherichia coli} (AIEC) and \textit{Mycobacterium avium} subspecies \textit{paratuberculosis} as potential contributors to pathogenesis\textsuperscript{347, 348}. On the other hand, decreased frequency of \textit{Faecalibacterium prausnitzii} among those with inflammation compared to healthy controls, suggests that this organism may have a protective effect\textsuperscript{250, 349}. Recent studies have shown that intestinal dysbiosis is associated with disease, and advances in culture-independent sequencing approaches have demonstrated a vast amount of heterogeneity within the microbiota of the gastrointestinal tract\textsuperscript{337, 350}. This highlights the need for further investigation in large and phenotypically diverse cohorts.

Post-surgical models of IBD are useful for studying the role of microbes, as recurrence can be viewed as a surrogate for \textit{de novo} onset of disease. Among UC patients, greater than 20% will require surgical management\textsuperscript{324}, for which the treatment of choice is a colectomy with ileal-pouch anal anastomosis (IPAA). Colectomy is often considered a definitive treatment for UC, however, \textit{de novo} inflammation of the ileal reservoir (pouchitis) is a common post-surgical complication with prevalence rates ranging from 12\% to greater than 50\%\textsuperscript{29, 306}. Additionally, 10-17\% of patients go on to develop a CD-like phenotype which is described as the development of abdominal or perianal fistulas or abscesses, or inflammation of the small bowel proximal to the pouch (afferent limb)\textsuperscript{307, 329}. IPAA is also the treatment of choice among individuals with familial adenomatous polyposis (FAP), however, inflammatory complications of the pouch among this group are very rare. Recent genetic studies have shown that among individuals with IPAA, those with polymorphisms in innate immune and bacterial sensing and recognition genes are at an increased risk for inflammatory complications\textsuperscript{351}. However, inflammation rarely develops in the absence of fecal flow, suggesting that this
genetic predisposition alone does not itself cause inflammation, and that microbial factors may have a critical role. The aim of this study was to characterize and evaluate the mucosal microbiome of individuals having undergone IPAA for treatment of UC or FAP.

4.3 Materials and Methods

Ethics Statement

This study was approved by and carried out in accordance with the Research Ethics Board of Mount Sinai Hospital (Toronto, Canada).

Subject Recruitment

Patients were recruited during regular pouch follow-up at Mount Sinai Hospital (MSH) in Toronto, Canada. Any patients with confirmed UC or FAP and who had undergone IPAA at least one year prior to recruitment were included in the study. Biopsies were taken from within the pouch itself (1 biopsy) and 5-10 cm into the afferent limb (1 biopsy), and were immediately placed into sterile, empty freezer vials and snap frozen in liquid nitrogen. Two additional biopsies from the same locations were sent to the MSH pathology lab for histological scoring. During the pouchoscopy, physicians documented the appearance of the pouch and afferent limb using previously described criteria for pouch inflammation. Peripheral blood was also collected for clinical evaluation of C-reactive protein (CRP) levels.

All subjects were classified into outcome groups based on a combination of long-term complications in conjunction with inflammatory activity at the time of the procedure. Those with FAP were classified as such, while the remaining groups were composed of individuals with UC prior to colectomy. To assess inflammation of both the pouch and afferent limb, endoscopic appearance (erythema, friability, ulceration) and histological (polymorphonuclear leukocyte infiltration, ulceration/erosions) scores at the time of the study endoscopy were considered (Chapter 4, Supplementary Table 1). These traits were utilized as they appear to be objective categories within the PDAI and PAS. Using these categories, a score greater than 3 was applied as the cutoff indicative of
inflammation. Individuals who had never been diagnosed with pouchitis and who had no evidence of inflammation of the pouch or afferent limb at the time of sample collection were classified into the “no pouchitis” group; individuals with evidence of inflammation of only the pouch at the time of endoscopy, based on the criteria described above, were classified as “pouchitis”; individuals with documented evidence of inflammation of the afferent limb or proximal small bowel, historically, or at the time of their procedure, or who had developed a stricture or fistula no sooner than one year following the closure of their ileostomy were classified into the “CD-like” group (CDL). This group was made up of individuals who had both active inflammation as well as those without. Individuals with a previous diagnosis of CDL or chronic pouchitis, based on documented evidence of complications and a need for long-term medical therapy, but who had an inflammatory score below our cutoff for both the pouch and afferent limb, and no fistulae, were considered a separate group (“quiescent”) and were not included in the primary analysis.

Microbial DNA Extraction and Analysis

Biopsies were thawed and processed in two batches at the Department of Animal Science, University of Manitoba (Winnipeg, Canada). The QIAGEN DNeasy Blood and Tissue Kit (QIAGEN, California, USA) with an additional bead beating step was used to extract microbial DNA. Barcoded primers specific for the V1-V3 hypervariable region of the 16S rRNA molecule were used for 454 pyrosequencing using the GS FLX Titanium assay (Roche, USA)\textsuperscript{353}. Sequences were processed and assigned to a taxonomy using mothur (Chapter 4, Supplementary Methods).

Statistical analysis

Taxonomic groups which were detected in less than 5% of the samples were excluded from further analysis. Analyses were performed separately on pouch and afferent limb samples in order to avoid falsely inflating significance. Genus level results were analyzed using presence-absence and frequency data with exact and non-parametric methodologies (Chapter 4, Supplementary Methods). Confirmation of results and adjustment for confounding variables was performed using exact logistic regression. FDR corrected p-values below 0.05 were considered significant. All statistical analyses were
conducted using STATA version 11.1 (StataCorp Texas, USA) and R version 2.13.1 (R Foundation for Statistical Computing, Vienna, Austria). Additional analyses were carried out using the Linear Discriminant Analysis (LDA) Effect Size (LEfSe)\textsuperscript{354} tool (version 1.1.0).

Quantitative PCR (qPCR)

Organisms previously implicated in bowel inflammation including \textit{F. prausnitzii}, \textit{Clostridial} cluster IV, AIEC, and \textit{Roseburia} were evaluated using qPCR with previously described primer pairs (Chapter 4, Supplementary Table 2). Relative abundance was estimated as has been previously described, using the 16S rRNA \textit{Eubacteria} gene as a reference (Chapter 4, Supplementary Methods)\textsuperscript{355}. Results from individual samples were log\textsubscript{2} transformed and analyzed using the Kruskal-Wallis test. Additionally, primers specific for the \textit{Bacteroides} genus were used to confirm pyrosequencing results by evaluating the correlation between dichotomized (Matthew’s) and frequency (Spearman) pyrosequencing and qPCR data.

4.4 Results

Patient phenotypic and clinical characteristics

274 patients were recruited to this study. To ensure a balanced study design, 78 were selected for analysis based on phenotype, with 71 classified into the four primary outcome groups. Seven patients were excluded from the initial analysis having been diagnosed with chronic pouchitis or CDL which was effectively treated with antibiotics or other medications (quiescent). In total, there were 18 patients in the FAP group, 19 no pouchitis, 15 pouchitis, and 19 CDL. Among those classified as CDL, two had only perianal fistulae and an inflammation score of zero in both the pouch and afferent limb (no evidence of inflammation at the time of endoscopy). One additional patient included in this group had evidence of only mild inflammation in both the pouch and afferent limb (inflammatory score less than or equal to 3), with additional features of the CDL phenotype. Inflammatory activity, as measured by our objective inflammatory score, was
low among individuals in the FAP and no pouchitis groups, and higher in both the pouchitis and CDL groups in the pouch (P<0.01). In the afferent limb, only individuals in the CDL group displayed evidence of inflammation (P<0.01)(Chapter 4, Supplementary figure 1).

Interestingly, C-reactive protein (CRP) levels were similarly elevated among each of the UC groups, compared to individuals with FAP (P=0.002)(Table 14). There was an approximately equal split across genders through each of the outcome groups and the mean age at UC or FAP diagnosis and time from ileostomy closure to sampling were approximately the same (Table 14). Smoking was significantly more common (P=0.007) among individuals who were in the FAP group. The majority of patients (54%) had taken antibiotics, (ciprofloxacin, metronidazole, or combination) at some point following their pouch procedure (either for pouch complications, or for reasons not related to bowel inflammation). Antibiotic use at the time of pouchoscopy, or during the month immediately prior is documented in Table 14. Individuals in the FAP and no pouchitis groups who were on antibiotics (amoxicillin, cefixime) reported sinus or urinary tract infections as the reason for medication use. One individual in the FAP group took medications because of possible pouchitis-like symptoms. However, endoscopic and histological examinations at several time-points suggest that an inflammatory outcome was not the appropriate diagnosis. Another individual in this group was characterized as having pouch inflammation based on mild pouch erythema and histology showing discrete patchy polymorphonuclear leukocyte infiltration in both the pouch and afferent limb. This individual had no symptoms consistent with pouchitis and was taking Sulindac for periampullary polyps.
Table 14: Phenotypic characteristics of individuals among the four outcome groups.

<table>
<thead>
<tr>
<th></th>
<th>FAP (n=18)</th>
<th>No Pouchitis (n=19)</th>
<th>Pouchitis (n=15)</th>
<th>Crohn’s disease-like (n=19)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (% female)</td>
<td>55.6</td>
<td>26.3</td>
<td>26.7</td>
<td>52.6</td>
<td>0.14</td>
</tr>
<tr>
<td>Mean age at recruitment (years(SE))</td>
<td>38.1 (2.1)</td>
<td>52.8 (3.3)</td>
<td>41.3 (3.8)</td>
<td>46.3 (3.1)</td>
<td>0.009</td>
</tr>
<tr>
<td>Mean age at UC/FAP diagnosis (years (SE))</td>
<td>23.1 (3.0)</td>
<td>31.9 (3.3)</td>
<td>27.8 (3.0)</td>
<td>30.5 (2.9)</td>
<td>0.20</td>
</tr>
<tr>
<td>Time from surgery to sample collection (years (SE))</td>
<td>11.1 (1.5)</td>
<td>10.6 (1.7)</td>
<td>9.3 (1.6)</td>
<td>9.9 (1.3)</td>
<td>0.88</td>
</tr>
<tr>
<td>C-reactive protein (mean (SE))</td>
<td>2.0 (0.6)</td>
<td>7.1 (3.1)</td>
<td>9.8 (3.4)</td>
<td>7.7 (2.3)</td>
<td>0.002</td>
</tr>
<tr>
<td>Antibiotic previous month (% using, SE)</td>
<td>11.1 (7.4)</td>
<td>10.5 (7.0)</td>
<td>13.3 (8.7)</td>
<td>42.1 (11.3)</td>
<td>0.07</td>
</tr>
<tr>
<td>Antibiotics ever (% used, SE)</td>
<td>11.1 (7.4)</td>
<td>42.1 (11.3)</td>
<td>73.3 (44.2)</td>
<td>89.5 (30.6)</td>
<td>0.03</td>
</tr>
<tr>
<td>Biologics (% using, SE)</td>
<td>0</td>
<td>0</td>
<td>6.7 (6.4)</td>
<td>10.5 (7.0)</td>
<td>0.44</td>
</tr>
<tr>
<td>Smoking at recruitment (% using, SE)</td>
<td>41.2 (11.6)</td>
<td>5.3 (16.2)</td>
<td>6.7 (6.4)</td>
<td>5.3 (5.1)</td>
<td>0.007</td>
</tr>
</tbody>
</table>

FAP = familial adenomatous polyposis; SE=standard error. Individuals who were classified as smoking at recruitment reported at least one cigarette per day within the last month.
Bacterial community diversity and phylum composition of the ileal pouch

The pyrosequencing experiment provided 557,215 (mean of 7143 sequences per sample) raw sequences with a median read length of 453 basepairs (range 250-531). However, following quality trimming and chimera checking, 345,466 (mean of 4429 sequences per sample) high quality reads remained. To ensure that our stringent inclusion criteria did not adversely affect our ability to detect organisms in our samples, Good’s coverage was calculated for each sample, and was on average greater than 92% for all outcome groups. A single sample from the pouchitis group had coverage which was below 80%. Diversity, measured using the Shannon and the inverse Simpson diversity indices, was lower among samples obtained from the pouchitis group (Chapter 4, Supplementary table 3) in the pouch. Results in the afferent limb were similar, although no significant differences between outcome groups were observed at this site.

The Firmicutes, Proteobacteria, Bacteroidetes, and Actinobacteria were the most diverse phyla present with the majority of genera detected belonging to one of these groups (Figure 7A; Chapter 4, Supplementary Figure 2). Firmicutes, Proteobacteria, Bacteroidetes and Fusobacteria were numerically the most abundant organisms (Figure 7B; Chapter 4, Supplementary Figure 2). Additionally, Actinobacteria were detected among 72% of the samples, although at very low abundance. There was no significant difference in the number of genera detected per phylum between the four outcome groups. In both the pouch and afferent limb, a significantly greater proportion of sequences in the FAP and no pouchitis groups were Bacteroidetes with, on average, 37.4% and 26.7% of the community composition made up of organisms from these groups compared to the 1.9% and 6.6% in the pouchitis and CDL groups respectively in the pouch (P=0.0001) (Figure 7B, Chapter 4, Supplementary Table 4). Results in the afferent limb were similar. While there was no significant association observed in the four-way analysis, Proteobacteria were detected more commonly among individuals with the CDL phenotype compared to FAP (P=0.004). Similar trends were observed between the pouchitis and FAP group, although not reaching statistical significance. .
Figure 7: Phylum level comparisons between four outcome groups.

A) Proportion of the total number of genera detected in each major phylum. B) Proportion of sequences detected belonging to each phylum. Depicted results are averaged between afferent limb and pouch samples, with significance assessed separately for each site. Individual results for the pouch and afferent limb are included in Supplementary figure 2. FAP=familial adenomatous polyposis, CDL=Crohn’s disease-like.
Microbial composition of the ileal pouch at the genus level

In total, 287 genera were detected. Of these 83 were present in more than 5% of samples. There was no statistical difference in genera positivity or abundance between sampling locations (pouch vs. afferent limb). However, to avoid bias from including multiple samples from the same individual, analysis was conducted separately on the different biopsy sites.

Fisher’s exact test was used to conduct a four-way comparison evaluating whether different proportions of individuals in each outcome group had a specific genus detected at pouchoscopy. Within the pouch, 15 genera were associated with outcome at a nominal significance threshold of $P \leq 0.05$, with five (\textit{Bacteroides}, \textit{Parabacteroides}, \textit{Moryella}, \textit{Sutterella} and \textit{Blautia}) remaining significant after FDR correction for multiple testing ($P_{\text{corr}} \leq 0.05$) (Figure 8A, Chapter 4, Supplementary Table 5). In the afferent limb, 16 genera were significantly associated with outcome, with seven (\textit{Blautia}, \textit{Bacteroides}, \textit{Dorea}, \textit{Anaerococcus}, \textit{Sutterella}, \textit{Parabacteroides}, and \textit{Icertae Sedis \textit{Erysipelotrichaceae}}) remaining significant after correction (Figure 8B; Chapter 4, Supplementary Table 5). These genera were typically detected less frequently among individuals with pouchitis and CDL compared to those with the non-inflammatory outcomes.
Figure 8: Proportion of patients positive for genera which were significantly associated with outcome (P_{corr}<0.05) after correction for multiple testing. * represent pairwise comparisons which were significant after correction. A) Pouch samples. B) Afferent limb samples. I.S.=Incertae sedis.
As the CDL group was somewhat heterogeneous, containing individuals both with and without active inflammation at the time of their pouchoscopy, we performed an additional analysis, including only CDL individuals with inflammation at the time of sampling (n=16). The results were similar to those in the previous analysis, although *Moryella* in the pouch and *Anaerococcus* in the afferent limb did not remain significant after multiple testing correction (Chapter 4, Supplementary Figure 3). When we examined the effect of inflammation on bacterial composition we found that genera associated with inflammation were similar to those described in our four group comparison, with *Bacteroides* \( (P=2.5\times10^{-5}) \) and *Parabacteroides* \( (P=2.5\times10^{-4}) \) less commonly detected among inflamed samples (Figure 9).
Figure 9: Proportion of patients positive for genera in individuals with inflamed vs not-inflamed pouches. Depicted results are those associated with inflammation at a nominal p-value threshold of P<0.01. * demonstrates associations which remained significant after correction for multiple testing (P_{corr}<0.05).
We next examined the relationship between quantitative genera abundance and outcome. Results in both locations were similar to those observed for the dichotomized results. In both sites, *Bacteroides*, *Blautia*, and *Sutterella* were positively associated with outcome, as were, *Parabacteroides* and *Moryella* in only the pouch samples (*P* corr < 0.05) (Chapter 4, Supplementary Figure 4). In each case, genera were detected at lower levels in the pouchitis and CDL groups compared to those with non-inflamed tissue.

Across all groups, use of one or more antibiotics at the time of the pouchoscopy or during the preceding month was positively associated with the presence of *Lactobacillus* in both the pouch and afferent limb (pouch *P* = 4.6x10^-4; afferent limb *P* = 8.9x10^-4). Also associated with antibiotic use in the afferent limb were *Dorea* (*P* = 4.0x10^-4) and *Ralstonia* (*P* = 1.4x10^-4), with *Ralstonia* detected more commonly among individuals who were on antibiotics, and *Dorea* less common among individuals on these medications. Several additional genera, including those which were associated with outcome were shown to be marginally effected by antibiotic use at a nominal significance threshold (*P*<0.05), however did not remain significant after correction for multiple testing. Genera abundance data showed similar results, with *Lactobacillus* and *Haemophilus* comprising an increased proportion of the microbiome among individuals on antibiotics in both the pouch and afferent limb, and *Ralstonia*, detected at higher frequency in the afferent limb. *Dorea* was detected at a lower frequency in the afferent limb among individuals on antibiotics. No other medication use appeared to significantly influence microbial composition, although few patients were available for this analysis.

53.5% of our population had taken antibiotics at some point in the past (pre- or post-colectomy) (Table 14). To verify that our observed associations were between inflammatory phenotype and specific genera, and to ensure that our observations were not due to antibiotic use, we performed our analysis again evaluating only individuals who were not on antibiotics during the month (n=57), or 6 months (n=55) preceding their pouchoscopy. Results in the dichotomous analysis for both locations and time points demonstrated that genera previously found to be associated with outcome after correction were all associated in this smaller cohort (*P*<0.05) with the exception of *Moryella* in the
pouch and *Parabacteroides* in the afferent limb (Chapter 4, Supplementary Figure 5). All previously described organisms remained significant in the frequency analysis.

Exact logistic regression was performed with smoking, gender and Canadian vs. other country of birth included as co-variates. In both anatomical locations, *Bacteroides*, *Sutterella* and *Blautia*, remained significantly associated with outcome ($P<0.01$). *Moryella*, and *Parabacteroides* were associated with outcome only in the pouch and *Dorea* and *Anaerococcus* in the afferent limb (Table 15). Adjusted odds ratios demonstrated a significant decrease in likelihood of pouch inflammatory complications associated with the above genera. No significant results in this pair-wise analysis were detected for the FAP vs. no pouchitis or pouchitis vs. CDL comparisons.
Table 15: Association between outcome and bacterial positivity in multivariate analysis with smoking, birth country and gender included as co-variates.

<table>
<thead>
<tr>
<th>Pouch</th>
<th>FAP vs. Pouchitis</th>
<th>FAP vs. CDL</th>
<th>No Pouchitis vs. Pouchitis</th>
<th>No Pouchitis vs. CDL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P-value</td>
<td>OR</td>
<td>P-value</td>
<td>OR</td>
</tr>
<tr>
<td>Bacteroides</td>
<td>.002*</td>
<td>.06</td>
<td>.002*</td>
<td>.04</td>
</tr>
<tr>
<td>Parabacteroides</td>
<td>.006*</td>
<td>.05</td>
<td>.004*</td>
<td>.05</td>
</tr>
<tr>
<td>Moryella</td>
<td>.04</td>
<td>.08</td>
<td>.009*</td>
<td>.05</td>
</tr>
<tr>
<td>Sutterella</td>
<td>.01</td>
<td>.05</td>
<td>.001*</td>
<td>.02</td>
</tr>
<tr>
<td>Blautia</td>
<td>.009*</td>
<td>.07</td>
<td>.31</td>
<td>.25</td>
</tr>
</tbody>
</table>

Afferent Limb

<table>
<thead>
<tr>
<th>I.S. (Erysipelotrichaceae)</th>
<th>P-value</th>
<th>OR</th>
<th>P-value</th>
<th>OR</th>
<th>P-value</th>
<th>OR</th>
<th>P-value</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sutterella</td>
<td>.06</td>
<td>.11</td>
<td>.006*</td>
<td>.04</td>
<td>.14</td>
<td>.20</td>
<td>.02</td>
<td>.07</td>
</tr>
<tr>
<td>Anaerococcus</td>
<td>.34</td>
<td>.28</td>
<td>.21</td>
<td>.22</td>
<td>.06</td>
<td>.13</td>
<td>.008*</td>
<td>.06</td>
</tr>
<tr>
<td>Parabacteroides</td>
<td>.13</td>
<td>.21</td>
<td>.02</td>
<td>.10</td>
<td>.14</td>
<td>.23</td>
<td>.03</td>
<td>.13</td>
</tr>
</tbody>
</table>

Described nominal p-values marked with a * are those which remain significant after correction for multiple testing ($P_{corr}$<0.05). Genera tested are those which were significant in the preliminary analysis. No significant associations were seen for the FAP vs No Pouchitis or the Pouchitis vs CDL comparisons. FAP = familial adenomatous polyposis, CDL = Crohn’s disease like. OR = odds ratio I.S.= Incertae sedis.
Finally, analysis conducted using LEfSe confirmed many of our previous results, and highlighted additional organisms with potential roles in the inflammatory phenotypes (Chapter 4, Supplementary Figures 6). These included the class Bacilli (phylum Firmicutes) which was detected at higher frequency among inflammatory groups, and Proteobacteria detected more often among individuals with CDL (Chapter 4, Supplementary Figures 6).

qPCR validation and targeted organism analysis

qPCR was performed on 86 samples from 43 individuals (FAP=12; No Pouchitis=10; Pouchitis=11; CDL=8; quiescent=2). *Bacteroides* qPCR results correlated with those obtained from pyrosequencing (Matthew’s correlation coefficient = 0.78; Spearman correlation coefficient = 0.74). Furthermore, detection rates among the outcome groups were similar with *Bacteroides* detected less frequently in the pouchitis and CDL groups than in either FAP or no pouchitis groups ($P<0.01$). Additional organisms associated with outcome included the *Roseburia* spp. / *E. rectale* group (pouch) and *Clostridium* cluster IV (afferent limb). These organisms were present in increased numbers in the FAP group compared to both pouchitis and CDL groups (Chapter 4, Supplementary Figure 7).

Pyrosequencing results for treated quiescent, medication dependent samples

Seven patients not included in our analysis were classified as quiescent. These individuals were on long term therapy to maintain remission, and displayed no features of inflammation at the time of their study visit. Nominally significant differences ($P<0.01$) were observed between the quiescent and FAP groups in several genera in both the pouch (*Bacteroides* and *Blautia*) and the afferent limb (*Blautia*, *Bacteroides*, and Incertae Sedis Erysipelotrichaceae) in both the dichotomized and frequency analyses (Chapter 4, Supplementary Figure 8). In both the pouch and afferent limb, there were no significant differences between the pouchitis or CDL groups and the quiescent individuals. Despite the lack of evidence of inflammation, samples taken from individuals on long-term
therapy because of a diagnosis of chronic, medication dependent disease, seemed to have a similar microbial profile among pouchitis-associated genera to the inflammatory outcome groups.

4.5 Discussion

These data demonstrate decreased microbial diversity among individuals with pouchitis compared to those with non-inflamed pouches, with broad changes at the phylum and genus level observed. Similar reductions in diversity have also been demonstrated in non-surgical bowel inflammation\textsuperscript{169, 350}. Interestingly, both the non-parametric Shannon and inverse Simpson indices demonstrated that all three pre-colectomy UC groups had decreased diversity compared to the FAP group. Previous histological evidence suggesting that subclinical inflammation may be present in individuals with pre-colectomy UC, although not in those with FAP may offer an explanation for this finding\textsuperscript{356}. Thus, the observation that individuals classified into the no pouchitis group had microbial characteristics intermediate between those seen in individuals with pouchitis and FAP, may be related to subclinical inflammation present in this group which was not captured by our reporting, but which was reflected in increased CRP levels detected in this group. Also of note is the observation that the CDL group did not demonstrate decreased diversity compared to individuals without pouch inflammation. This may be related to the fact that several individuals with CDL were not experiencing active pouch inflammation at sampling, and that the inflammatory activity detected in this group at the time of pouch endoscopy was more heterogeneous than was that observed in the pouchitis group. Indeed, this group had, on average, slightly lower inflammatory scores than did the pouchitis group.

At the phylum level, significantly decreased levels of Bacteroidetes and marginal increases in the Proteobacteria were associated with pouch outcome, among individuals with inflamed pouches compared to those without inflammation. This confirms results from other pouch studies, although results have been mixed in examining these organisms in the context of non-surgical IBD\textsuperscript{296, 301, 350, 357, 358}. However, our study demonstrates that while Actinobacteria is detected in a majority of samples (72%), it is present at low
abundance (mean 0.8%). Fusobacteria, on the other hand, were detected in only 28% of samples, yet among those individuals in whom these organisms were detected they were dominant members of the bacterial community. This contrasts reports from others \(^{296}\), and can most likely be attributed to the increased sequence coverage, and unique phenotypes explored in our experiment.

Phylum level comparisons, while interesting, are of limited use in predicting biological function. We subsequently demonstrated that specific genera are associated with pouch outcome among our patient cohort. Among the Clostridia group XIVa, *Blautia*, *Moryella* and *Dorea* were convincingly associated with outcome, detected more frequently and in greater abundance among individuals without inflammatory complications. Interestingly, a recent study of healthy subjects showed that *Blautia* was one of the only genera detected in samples from all individuals, suggesting that it is a common component of a healthy gut microbiota \(^{359}\). Furthermore, previous studies have suggested that individuals with CD, including those in remission, have decreased levels of Clostridia-related organisms detected \(^{360,361}\). Both *Bacteroides* and *Parabacteroides* were also strongly associated with outcome. These closely related genera have both been previously implicated in pouch inflammation \(^{296,301}\). Decreased levels of *Bacteroides* have also been previously observed in some studies in inflamed UC and CD tissue compared to healthy controls, while other studies have shown increases in this organism to be associated with disease \(^{362-365}\). This further suggests the importance of this organism in bowel inflammation, regardless of host or specific anatomic site. Furthermore, *Bacteroides* species, specifically *B. fragilis*, have been shown to modulate immune function, stimulating immune tolerance through Toll-like receptor (TLR)\(^{259}\). *Sutterella* was the only genus from the Proteobacteria phylum associated with outcome, and in contrast with previous studies suggesting a potential pathogenic role for this organism in IBD or no association at all \(^{366}\), was detected less frequently among inflammatory groups. We noted an increase in the prevalence of Proteobacteria and the class Bacilli among the inflamed groups, however this association was less dramatic. Further study investigating the relevance of these groups of organisms is necessary to determine the significance of this finding.
We specifically evaluated whether organisms previously associated with bowel inflammation, including CD, were also involved in pouch inflammation to determine whether similar biological processes might be of importance in both disorders. Contrary to previous studies, we did not observe an association between *F. prausnitzii* or AIEC and inflammatory outcome in our cohort. However, our results do support previous observations that additional members of the Clostridium group\textsuperscript{358}, and *Roseburia* spp./*E. rectale* may be associated with pouch inflammation.

The use of subjects who had an IPAA for FAP, allowed us to evaluate changes occurring in the pouch microbiome which were relevant specifically to individuals with an inherited predisposition to intestinal inflammation. Individuals with FAP rarely develop pouch inflammation, despite the fact that their digestive tract anatomy is identical to those with UC. Additionally, subclinical inflammation may be present in individuals with pre-colectomy UC, without a clinical diagnosis of pouchitis, as our finding of elevated levels of CRP among this group may suggest. This subclinical inflammation has important clinical and experimental implications, and is reflected in our observations illustrating substantial changes between the microbiome of individuals with inflammatory pouch complications compared to those with FAP. More moderate, yet similar alterations occur in the microbiome of individuals with UC when comparing those with and without inflammation. This likely reflects underlying differences in host genetic predisposition in UC patients resulting in specific alterations in the pouch microbiome that characterize the inflammatory phenotypes.

While there were no statistical differences in medication use between outcome groups at the time of or immediately prior to pouchoscopy, it is conceivable that medications, particularly antibiotics, may have impacted the tissue associated microbiota. Yet the described genera were detected significantly less frequently among individuals in the inflammatory outcome groups even when individuals who had taken antibiotics in the preceding month or six months were excluded from the analysis. This would suggest that the effect of inflammation on the microbiome supersedes the impact of antibiotic use. Alternatively, if shifts in the microbiome of the pouch precede and possibly induce pouch inflammation, then the use of antibiotics may not result in long term beneficial effects on bacterial composition or lead to a reversion to a “healthy” pouch microbiome as would be
suspected by the typically antibiotic responsive nature of pouchitis. This may in turn be responsible for the often recurrent nature of pouchitis following treatment. It is also possible that past antibiotic use (beyond one month prior to procedure) leads to permanent changes to the microbiome which are detectable long after cessation of medications. Studies evaluating the long-term effect of antibiotics on intestinal microbial composition have shown mixed results with some studies showing a relatively quick reversion to a profile similar to the pre-medication state\textsuperscript{367}, while others suggest long-term changes, especially with antibiotic use in childhood\textsuperscript{241,368}. Interestingly, antibiotic use in childhood has been associated with an increased risk of CD, suggesting that changes to the microbiota which occur early in life, whether they persist long-term or not, may have a drastic impact on subsequent disease susceptibility\textsuperscript{11}. However, as the levels of these genera had recovered to detectable levels in FAP and individuals without pouchitis, despite historic antibiotic use (data not shown), yet remained undetectable in the most pouchitis and CDL individuals, we conclude that this decrease was associated with the phenotype rather than medication use.

Interestingly, we did not observe differences in the microbial composition of the pouch compared to the afferent limb among patients regardless of the inflammatory status of either site. The anatomical alterations resulting from the IPAA procedure, namely the surgically induced continuity between the pouch and afferent limb, likely promotes similar bacterial community structure between these two locations. Previous studies have documented the occurrence of variability in the microbiome between inflamed and non-inflamed neighbouring sites, although none were consistent across many individuals\textsuperscript{357}. This suggests that individual-specific differences in the microbiome may be important in pathology. Further, the absence of evidence for differences between inflamed and uninflamed locations in the same individual suggests that variable host genetics or location-specific gene expression patterns may be of importance in establishing disease location and phenotype. It is also possible that microbial variation results from the inflammatory process itself. In this case, loss of specific organisms may contribute towards propagating disease but not initiate disease processes. Prospective studies are required to evaluate this hypothesis.
Of particular interest, is the observation that the seven individuals who were taking medication for chronically inflamed pouch phenotypes, and who had successfully achieved resolution of their symptoms, did not demonstrate a change in their microbial composition to resemble that of individuals with healthy pouches. While it appears that the samples from individuals with quiescent disease were most similar to those from individuals with pouchitis, it is conceivable that with a larger sample size differences between these groups may have been observed. Additionally, changes at lower taxonomic levels (ie. species), which were undetectable in this experiment could also have been associated with different phenotypes. It is interesting to speculate, however, that perhaps the reason for patients’ continual requirement for medical therapy is their inability to be colonized by the ‘beneficial organisms’ described by our group.

The majority of associations described in this report suggest that a loss of organism diversity is associated with inflammation. It is possible that individuals with pouch inflammation may undergo a loss of specific anti-inflammatory organisms that have important roles in modulating the immune response. Many of the organisms detected less frequently in the inflammatory groups have been previously shown to have anti-inflammatory potential: members of the Clostridia XIVa group (Blautia, Dorea and Moryella) produce butyrate, a known anti-inflammatory molecule\(^{369,370}\), and Bacteroides produce vitamin K, which is associated with decreased levels of serum inflammatory markers (ie TNFa, CD40 etc)\(^{371}\). Alternatively, the absence of these typically ubiquitous organisms may allow increased colonization of tissue by atypical species. While we did not observe increased prevalence of specific genera in our cohort, it is conceivable that small changes among several groups of organisms may not have been detected. In any event, endogenous gut organisms have the capacity to exert profound effects on the function of the immune system, and play a pivotal role in maintaining gut immune homeostasis\(^{259,264,372}\). Such changes in immune function may also be reflected by the presence of anti-microbial antibodies in the serum which are also known to be associated with pouch outcomes\(^{67,329}\). In the future, functional metabolites from these organisms may represent novel diagnostic targets or have therapeutic potential for patients with active disease.
Our experiments have demonstrated that a reduction in specific organisms is associated with inflammatory outcomes. These changes are characteristic of both pouchitis and the Crohn’s disease-like phenotype, and include several organisms which have been previously associated with non-surgical IBD. These observations demonstrate that the role of these microbes in bowel inflammation may be common and independent of surgical history or anatomic location of disease. Our previous findings showing that IBD-associated genetic polymorphisms are also important in pouch pathology, in conjunction with this work suggests that like non-surgical IBD, pouch inflammation arises in individuals with inherited genetic predisposition and changes in the tissue associated microbiome. However, the event triggering disease onset remains elusive, and will require large, prospective studies to evaluate.
4.6 Chapter 4, Supplementary Methods

Microbial DNA Extraction and Analysis

Sequences were parsed based on barcodes using a customized Perl script. Reads with an average quality score below 35 were removed prior to sequence processing. Data was then quality trimmed and aligned to a SILVA reference database (version SSUref_102) using mothur (version 1.12.3)\textsuperscript{373}. This resulted in removal of sequences which were less than 250 base pairs long (excluding primers), and/or which contained ambiguous bases or homopolymers extending eight or more nucleotides. The ChimeraSlayer algorithm\textsuperscript{215} was used to remove potentially chimeric sequences from the dataset. Additionally, a pre-cluster step was performed, as has been previously recommended by Huse et al., allowing detection and removal of rare spurious sequences\textsuperscript{214}. Individual sequences were then assigned to taxonomic outcome groups, at the genus level where possible, using the Silva taxonomy with mothur\textsuperscript{374}. A Bayesian classification approach with 100 iterations, and an applied bootstrap cutoff of 60 was used to assign sequences to a taxonomic outcome groups.

Statistical analysis

Taxonomic groups which were detected in less than 5% of the samples were excluded from further analysis. Analyses were performed separately on pouch and afferent limb samples in order to avoid falsely inflating significance. Results were first dichotomized based on whether or not a given genera was detected within a sample. Fisher’s exact test was applied to determine whether there were statistically significant differences between groups. Results at the genus level were confirmed using exact logistic regression with smoking, country of birth and gender included as covariates to further validate results. Continuous abundance was calculated by normalizing the sequences for each taxon against the total number of sequences detected in a sample\textsuperscript{375}. Genera abundance was assessed for significance using the non-parametric Kruskal-Wallis and Wilcoxon signed rank test. FDR corrected p-values below .05 were considered significant in both the dichotomized and continuous abundance analyses. All statistical analyses were conducted
using STATA version 11.1 (StataCorp Texas, USA) and R version 2.13.1 (R Foundation for Statistical Computing, Vienna, Austria). Additional analyses were carried out using the Linear Discriminant Analysis (LDA) Effect Size (LEfSe)\textsuperscript{354} tool (version 1.1.0) to get a more robust understanding of differences in community structure across phenotypic outcome groups (Significance threshold $P<.05$).

Quantitative PCR (qPCR)

Organisms previously implicated in inflammation including $F. \text{ prausnitzii}$, Clostridial cluster IV, AIEC, and $Roseburia$ were evaluated using qPCR with previously described primer pairs (Supplementary Table 2). Relative abundance was estimated as has been previously described\textsuperscript{355}, with Eubacterial primers used to construct a standard curve. Reactions were carried out using the Power SYBR green PCR master mix (Applied Biosystems, Foster City, CA) with the AB 7300 system and sequence detection software (version 1.3; Applied Biosystems, Foster City, CA). Results from individual samples were log$_2$ transformed and analyzed using Kruskal-Wallis test. Additionally, primers specific for the Bacteroides genus were used to confirm pyrosequencing results by evaluating the correlation between dichotomized (Matthew’s) and frequency (Spearman) pyrosequencing and qPCR data. Dichotomous results were generated based on a pyrosequencing detection limit of $10^{-4}$ sequences per sample, estimated from the theoretical number of sequences possible per sample in the multiplexed pyrosequencing reaction.
5.1 Conclusions and general discussion

This project involved accrual of the largest IBD-IPAA patient population yet assembled, and studied a variety of important factors with the potential to contribute towards, or arise in response to, ileal inflammation. As such, well powered experiments allowed these analyses to be conducted with a high degree of statistical stringency. In each aspect of this study, steps were taken to ensure that the effects of potentially confounding factors were minimized, including restricting the sample size to include only Caucasians in the genetic analysis, and controlling for antibiotic use, country of birth, gender and smoking status when studying the microbiome. Furthermore, the strict patient phenotyping and classification guidelines, allowed us to ensure that observed associations were with the target outcomes of interest. As such, the findings presented in this thesis, of alterations in the SNP frequency of the \textit{NOD2}insC polymorphism, the frequency and abundance of organisms in the pouch microbiome and seropositivity of antimicrobial antibodies, suggests a model of how inflammatory conditions of the ileum may arise.

The size of the patient cohort, as well as the strict phenotyping criteria used in these studies, ensured that patients with acute, antibiotic responsive pouchitis as well as chronic pouchitis and a Crohn’s disease-like phenotype, were included in analyses. The subclassification of patients into either the antibiotic responsive or refractory outcome groups allowed determination of factors which were associated with disease chronicity, and which with a more transient disease course. Observations of genetic and serological profiles among individuals with chronic inflammatory phenotypes which differed from those seen in individuals with acute pouchitis (which were included in the no pouchitis group, although subgroup analyses were performed), suggest that these inflammatory outcomes proceed via different mechanisms. However, findings of similar alterations to the microbiota among individuals with active pouchitis (chronic and acute) and inactive, but medication dependent chronic pouchitis and CDL, suggest that similar bacterial alterations are involved in both pathologies. Additional exploration of this phenomenon in another large cohort will be useful in confirming these findings.

Decreased microbial diversity among individuals with pouchitis compared to those with non-inflamed pouches, with broad changes at the phylum and genus level were
observed. One of the major findings was that *Bacteroides* were detected significantly less frequently in inflamed tissue compared to that which is uninflamed. Previous reports have demonstrated that certain species of *Bacteroides*, specifically *B. fragilis*, modulate immune function, and stimulate immune tolerance through TLR2\(^{259}\). This, combined with observations by other groups that different members of this genus are able to decrease circulating cytokine levels\(^{259}\), strongly suggests that *Bacteroides* spp. play an important role in downregulating the immune response. Interestingly, *B. fragilis*, which has been shown to have important immuno-regulatory effects via the production of PSA\(^{376}\), binds to mucins in the mucus layer, rather than exclusively within the intestinal lumen where organisms typically reside\(^{377}\). This intimate association between host and microbe, demonstrates the ability of this organism to adapt to a specific host niche which is resistant to colonization by other organisms. Other species within this genus, and the closely related *Parabacteroides*, also produce PSA and related compounds, although they have not been studied to the same degree as has this organism\(^{378}\).

The other genera which were associated with inflammatory outcomes have not been as well studied. *Blautia* and *Moryella*, newly described members of the Clostridia Group XIVa\(^{379, 380}\), were detected less frequently among individuals with inflammation. *Blautia* is detected often in microbiome studies of the lower digestive tract of healthy individuals\(^{359}\), but its precise function has not been elucidated. Both *Blautia* and *Moryella* are members of the larger Clostridia group, however, and are therefore hypothetical producers of butyrate, a molecule with anti-inflammatory potential\(^{369}\). The role of *Sutterella* in modulating or propagating inflammation is more difficult to determine, as some studies suggest that members of this group promote inflammation, while others have shown just the opposite effect\(^{381, 366}\). Additional work will be required in order to determine the exact role of this group in IBD.

qPCR was also used to more closely evaluate whether particular *Clostridial* groups and strains of organisms, which have been previously associated with bowel inflammation, were also involved in pouch inflammation. Members of the *Clostridium* cluster IV and XIV groups may be protective against the development of IBD, and play a role in maintaining mucosal immune homeostasis\(^{187}\). Both *F. prausnitzi* (*Clostridium* cluster IV) and the *Eubacterium rectale/Roseburia* (*Clostridium* cluster XIV) groups
produce butyrate, via varied pathways\textsuperscript{382}. This short-chain fatty acid is thought to contribute towards decreasing levels of IL-12 and IFN-gamma production and leads to higher secretion of IL10\textsuperscript{250}. Additionally, organisms from the \textit{Clostridium} cluster IV group have the capability of increasing the abundance of T\textsubscript{reg} cells in the colon, but not in small bowel tissue, via activation of intestinal epithelial cells\textsuperscript{262}. Mice colonized by these organisms, and displaying a resultant increase in T\textsubscript{reg} cells throughout the colon, are resistant to the development of experimental colitis\textsuperscript{262}. Contrary to previous studies\textsuperscript{245, 343}, no statistically significant association between \textit{F. prausnitzii} and inflammatory outcome was observed in this cohort. Yet there did appear to be a moderate reduction in the frequency of this organism among inflamed samples. qPCR experiments did, however, demonstrate a reduced prevalence of the \textit{Eubacterium rectale/Roseburia} group among the inflammatory groups compared to uninflamed samples. This further emphasizes the critical role of bacterially-produced immune-modulating compounds in pouch inflammation.

AIEC has also been previously associated with IBD, detected more frequently in ileal tissue of individuals with CD\textsuperscript{347}. AIEC strains have the ability to adhere to and to invade intestinal epithelial cells via FimH, are capable of surviving and replicating in macrophages without inducing cell death, and induce increased production of the pro-inflammatory cytokine TNF-\textalpha\textsuperscript{383}. While there was no statistically significant association between this organism and outcome in our cohort, increased abundance of AIEC among individuals with the CDL phenotype was observed (not statistically significant), mirroring results seen in previous IBD studies\textsuperscript{347}. However, the high relative frequency of this organism in individuals with FAP, compared to UC patients without pouchitis, was a novel and surprising finding. Previous studies have observed higher levels of certain mucosa-adherent strains of \textit{E. coli} among individuals with colorectal cancer\textsuperscript{384}, suggesting that this organism may have a pathogenic role in bowel cancer development. This might also explain the higher prevalence of this organism in individuals with FAP, despite the lack of inflammation in this group. Additional evaluation in a larger cohort is required to confirm this finding.

Together, these results suggest that a reduction in bacterially produced immune-modulating compounds is important in IBD pathogenesis. These compounds likely play a
beneficial role in helping to maintain immune homeostasis among individuals without bowel disease, and may have therapeutic potential among individuals with already established inflammation. The observed associations between decreased organism diversity, and reductions in the prevalence of specific organisms and disease status, may be the result of two possible mechanisms: disease development may be initiated in response to a reduction in beneficial organisms, which is caused by an unknown factor. Alternatively, it is possible that microbial variation results from the inflammatory process itself. In this case, the increased prevalence of anti-microbial antibodies among individuals with inflammation in conjunction with observations of alterations in the microbiome may suggest that an aberrant host response to intestinal microbes contributes to a reduction in organism diversity, including loss of species with beneficial effects on the host. In this case, genetic or immune processes may drive the development of inflammation, and also contribute towards altering the composition of the gut microbiota. The loss of specific organisms may then contribute towards propagating disease but would not, themselves, initiate disease processes. In either scenario, the observation of specific reductions in microorganisms with immune-modulating properties, and the lack of strong evidence for an association with organisms which are more likely to be pathogenic, highlights the importance of 'good' bacteria in regulating gut immunity and homeostasis. However, as many of the changes to the microbiome which have been observed involve reduced detection of organisms which are capable of producing anti-inflammatory compounds, it is difficult to determine which comes first: decreased immune tolerance towards commensals, or reduction in organisms with anti-inflammatory properties. These observations do, however, emphasize the importance of cross-talk between commensal organisms occupying the pelvic pouch and the immune system in maintaining immune homeostasis.

The observed differences in the prevalence of several organisms (Blautia most notably, but also observed in Bacteroides and Dorea (afferent limb)) between the active pouchitis and the CD-like group may be a result of altered inflammatory activity, or may suggest that divergent mechanisms of disease onset and propagation occur in each of these phenotypic outcome groups. The CDL group was more heterogeneous than was the pouchitis group, made up of individuals with inflammatory scores ranging from zero to greater than ten. Further, there was a slight, but not significant, reduction in the mean
inflammatory score in the CD-like group compared to the pouchitis group. When only the individuals with inflammation at the time of their pouchoscopy were included in the CD-like group, the prevalence of all organisms but *Blautia* was similar in the active pouchitis and CDL groups. Furthermore, no significant difference in organism prevalence between these groups was observed in the pairwise analysis. It is possible that removal of these individuals resulted in reduced power, and an inability to detect significant associations with outcome. In this case, the increased prevalence of these organisms in individuals with CD-like phenotypes (intermediate levels between individuals with no pouchitis and pouchitis), may be indicative of important changes in the microbiome contributing to, or resulting from different disease mechanisms.

These results also highlight the need for more robust mechanisms for measuring inflammatory activity at intestinal sites. While several organisms were associated with our four phenotypic outcome groups, fewer associations were seen when we compared individuals with inflammation to those without. CRP levels among the three pre-surgery UC groups were also similarly elevated compared to individuals without UC, suggesting ongoing inflammatory activity was present despite the absence of endoscopic or histological evidence. Further, this activity appeared to be capable of influencing the composition of the pouch microbiome. Such observations suggest that a clinical diagnosis of inflammation may be non-specific, when compared to molecular indicators of disease activity. The relevance of such low-level inflammation in the context of patient outcomes is unclear.

Additionally complicating the study of the human microbiome is the observation that bacterial populations vary over time, with changes in bacterial communities resulting from alterations in diet, smoking, country of residence, or medication use. Yet alterations in the microbiome alone appear insufficient to cause chronic inflammation, as demonstrated by the observation that individuals with antibiotic responsive pouchitis have microbial profiles which resemble those seen in individuals with chronic inflammation. The pouch inflammation in patients with antibiotic responsive pouchitis is uniquely responsive to these medications, and can be rapidly resolved. However, among individuals with a genetic predisposition for IBD, it is possible that normal or abnormal variation in the microbiome tips the balance of immunity, and lead to
disease. When this happens, increased immune recognition of commensal organisms leads to the generation of anti-microbial antibodies, and contributes towards propagating inflammation. Enhanced immune response likely causes further reductions in the numbers of beneficial organisms present in the mucosa. As such, the specific combination of altered host genotype and intestinal bacterial dysbiosis may be sufficient to result in chronic inflammation.

Several genes are associated with increased susceptibility towards CD, IBD (both CD and UC) and UC alone\textsuperscript{104}. As a predominantly UC patient cohort, therefore, IPAA patients are more likely to have higher rates of UC and IBD-associated polymorphisms than are individuals without a previous diagnosis of IBD (ie. individuals with FAP prior to colectomy). This observation likely explains the high prevalence of pouch inflammatory complications among those who undergo IPAA for UC compared to FAP. Several of the UC-associated polymorphisms tag genes which are involved in intestinal barrier function such as \textit{HNF4A} and E-cadherin (\textit{CDH1})\textsuperscript{385}. Additionally, several CD-associated genetic polymorphisms are also associated with chronic pouch inflammation, confirming that additional host genetic factors are important determinants of phenotypic outcome following surgery, and may suggest mechanistic differences between acute (antibiotic responsive) and chronic pouch phenotypes. The well established \textit{NOD2}insC polymorphism results in a truncation of the leucine rich repeat region of the NOD2 protein which has demonstrated evidence of an inhibited ability to detect microbes\textsuperscript{336, 337}. Other genetic variants which demonstrated modest association with chronic inflammation of the pouch include those which are located in the non-coding regions of \textit{NOX3} and \textit{DAGLB} or adjacent to \textit{NCF4}. Individuals who are carriers of additional risk alleles at these loci have significantly decreased production of reactive oxygen species (which are important for defense from microbial pathogens) compared to patients not carrying these mutations\textsuperscript{342, 343, 344}. \textit{NOD2} and \textit{NCF4} polymorphisms have also been demonstrated to associate strongly with ileal CD\textsuperscript{106, 107, 341, 386}. These findings, in conjunction with results showing that microbes are also important in pathogenesis, suggest an important role for innate immune activity in response to commensal microbes in the development of chronic pouch inflammatory outcomes. It is likely that additional polymorphisms are also associated with pouch complications; however, these may only be detectable in larger sample sizes. Many of the SNPs associated with IBD have extremely low effect sizes,
requiring the use of extremely large cohorts (sample size greater than 10,000) to measure\textsuperscript{104}. However, these results definitively establish that host genotype, specifically loci previously associated with ileal inflammation, is critical to the pathogenesis of chronic pouch inflammatory complications.

The changes in immune function which are brought on by a reduced number of anti-inflammatory organisms, and increased host susceptibility due to decreased innate immune activity, is also reflected by the presence of anti-microbial antibodies in the serum of individuals with chronic pouch inflammatory phenotypes. This hypothesis is supported by data showing that UC patients with a CD-like phenotype after pelvic pouch surgery are more likely to be positive for ASCA and anti-CBir1, and by that of other groups’ demonstrating a strong association between the common UC associated serum marker, pANCA, and pouch outcome\textsuperscript{67, 68}. The presence of these anti-microbial antibodies may be representative of a broader breakdown in mucosal function. While individuals with chronic pouch inflammatory phenotypes were commonly positive for markers typically associated with CD (ASCA and anti-CBir1), no association was seen among individuals with acute, antibiotic responsive pouchitis. The genotype of these individuals (acute pouchitis) was also similar to that seen in individuals who had never experienced inflammatory pouch complications. This suggests that microbial alterations brought on by the pouch surgery alone are not sufficient to promote the development of anti-microbial antibodies. Instead, as with IBD, anti-microbial antibody positivity results from a complex interaction between genetically-mediated alterations to the gut barrier or innate immune functions of the mucosa, and the tissue-associated microbiome.

These results suggest that patients with \textit{NOD2}insC (and possibly with variants in the additional genes which show modest associations with outcome) are more likely to develop chronic inflammatory pouch complications in response to alteration in the microbial composition of the pouch mucosa. Decreased prevalence of organisms with immune-modulating functions such as \textit{Clostridium} group IV, XIV, and \textit{Bacteroides}-like organisms, resulting from either random variation or via host-mediated alterations in diet or gene expression (or numerous other types of environmental instability), among patients who are also carriers of these genetic susceptibility loci, results in an inability to downregulate the inflammatory response directed towards commensal organisms at the
intestinal epithelium. The increased adaptive immune activity resulting from this perturbation and imbalance leads to the production of anti-microbial antibodies, which can be detected in the serum of patients with IBD (Figure 10). Since \textit{NOD2} polymorphisms are associated with increases in serum concentration of ASCA, anti-OmpC and anti-CBir1, it is likely that alterations in this gene are a critical mediator in gut barrier permeability\textsuperscript{148}. Furthermore, these findings of the presence of both serologic markers of CD and genetic polymorphisms, which have been previously associated with ileal CD, showing association with pouch outcome, suggests that the mechanism of ileal inflammation leading to chronic pouch inflammation is similar to that which leads to the development of inflammation in the ileum.

Figure 10: Hypothetical model of the pathogenesis of ileal inflammation
Additional proposed mechanisms of pouchitis development including ischemia, mechanical failure, NSAID use and deposition of collagen on the pouch mucosa may effect a subset of patients with pouch complications. However, this may be ruled out in most cases based on the length of time between pouch surgery and inflammation development, and the absence of medication use prior to symptom onset.

The majority of individuals in whom pouchitis develops, respond rapidly to the administration of broad spectrum antibiotics. Yet the diagnosis of these individuals with UC prior to surgery suggests that they are genetically predisposed towards the development of chronic, colonic inflammation. Furthermore, the bacterial composition of the pouch brought on by inflammation, is the same as that which is observed in individuals with chronic pouch inflammation. Despite these similarities, they do not develop a chronic inflammatory phenotype following surgery. It is possible that given additional time, acute, medication-responsive pouchitis may evolve into a chronic inflammatory phenotype. However, the long follow-up time between pouch surgery and phenotyping in this cohort suggests that this is unlikely to be the case. Alternatively, differences in the gene expression of key mediators of inflammation, between colonic and ileal tissue could be responsible for the variable phenotypes. Such differences would contribute to chronic disease in large bowel tissue, and acute disease in ileal pouches. It may also be possible that specific differences in the mucosa-associated microbiome between colonic inflammation and ileal pouch inflammation lead to differences in phenotype and outcome. However, the most likely explanation for this phenotypic difference is that the absence of key polymorphisms among the acute pouchitis group, which are important mediators of ileal inflammation (ie. NOD2insC, NCF4) prevents the development of chronic inflammation in the ileal pouch among individuals with antibiotic responsive pouchitis. These individuals may, therefore, experience inflammation of ileal tissue, but do not experience long-term inflammatory complications when exposed to the same microbial stimuli.

Observations of recurrence of chronic inflammation among individuals who are clinically diagnosed with UC prior to colectomy, suggests to some that among these patients a missed diagnosis of CD is the reason for inflammation recurrence. Alternatively, patients with chronic inflammation and CD-like outcomes after pelvic
pouch surgery for UC may not be a “missed” diagnosis of CD, but rather evolution of the IBD-phenotype to one that more closely resembles CD rather than UC. These results in conjunction with data published by Waterman et al. which showed that there are few differences in the prevalence of several IBD-associated SNPs between those with UC and those with CD\textsuperscript{317}, support the concept of IBD as a mosaic of inflammatory disorders. The varying phenotypes associated with IBD arise via a specific combination of genetic (including alterations in gene expression, and epigenetic regulation) and microbial predisposing factors. Antibody profiles in serum reflect the underlying pathogenic forces contributing to disease development, rather than specificity for the clinical subtype into which patients are classified. Furthermore, the genetic susceptibility overlap between individuals with chronic pouchitis and those with the CD-like phenotype suggests that distinguishing between these outcomes is of limited importance compared to the more clinically relevant phenotype of chronic, medication refractory inflammation characteristic of both disorders and potentially arising from a common etiology. This concept of IBD has interesting implications for disease management, both prior to and following surgical intervention. Therefore, future disease prognostication and medical treatments must take into account clinical phenotype as only one aspect of disease management, with additional pathway-specific factors also taken into account in determining which patients are candidates for a specific procedure or medication.

The findings of this study also shed light on different avenues for investigation in the development of novel molecular tests and more targeted therapies for managing IBD. Molecules which stimulate the production of innate immune compounds or which enhance the function of the epithelial barrier may prevent commensal organisms from stimulating a severe and damaging adaptive immune response. Development and application of more effective probiotics, such as members of the genera identified by our group, might also have a beneficial effect on modulating disease course. Among patients in whom surgery is a consideration, knowledge of the specific genotype and antimicrobial antibody profile will help stratify patients based on their risk of post-surgical chronic inflammatory outcomes. This approach will ensure that patients who are more likely to develop complications are discouraged from undergoing IPAA, or are closely monitored following surgery to ensure that they are rapidly diagnosed and treated in the event that they develop post-surgical complications. Individuals in whom pouchitis does develop,
who have a profile suggestive of chronic inflammatory phenotypes, may be candidates for more rapid escalation of therapy. Furthermore, identification of individuals who are more susceptible to the development of either pouchitis or IBD, based on the diagnosis with IBD of a family member, and the detection of risk polymorphisms or anti-microbial serological markers, may be candidates for preventive therapy before disease onset. Such individuals may be good candidates for probiotic or prebiotic therapies to enhance the numbers of beneficial bacteria present in healthy pouch tissue. By ensuring that these organisms are present in individuals at increased risk of disease, the effect of their immunomodulating compounds may prevent other, more pro-inflammatory organisms from becoming dominant members of the microbiome, and therefore enhance immune homeostasis.

This hypothesis has broader applications as well, given that numerous other diseases including PSC, arthritis, and diabetes have shown similar genetic and antimicrobial antibody profiles to those observed in IBD. These conditions are also characterized by intestinal bacterial dysbiosis. Furthermore, additional autoimmune conditions have shown evidence of subclinical inflammation in the intestinal tract prior to diagnosis, as well as increased intestinal permeability. In a rat model of type I diabetes, for example, jejunal enteropathy and increased numbers of intraepithelial lymphocytes were detected among animals predisposed to the development of diabetes prior to symptom onset\(^{387}\). Further, even prior to diagnosis, specific pathogen treated mice may have the natural history of their disease course altered via the introduction of specific microorganisms, with several bacterial groups providing protection against the development of diabetes\(^{388}\). Such findings demonstrate the importance of gaining a better understanding of the complex relationship between gut immunity and the tissue associated microbiome of the digestive tract in order to better understand processes contributing to the development of immune-mediated conditions.

These results further demonstrate the utility of the pelvic pouch model for studying \textit{de novo} IBD and suggest that it may have important implications for understanding disease processes in both IBD and other immune-mediated conditions. Findings of similarities in genetic susceptibility towards chronic pouch inflammatory outcomes and CD and altered populations of microbes in the pouches of individuals with
inflammation, which suggest that chronic inflammation arises from similar mechanisms, provide a model for evaluating the instigating processes in de novo disease development. Thus, the application of well-designed prospective studies which capture different time points following ileostomy closure, will likely shed further light on initiating factors in de novo ileal disease onset.

5.2 Future directions

The work described in this thesis highlights important factors in the pathogenesis of inflammatory pouch complications. The similar genetic predisposition observed in pouch complications and IBD (notably ileal CD) highlighted by these studies, and specifically the role of innate and adaptive immune processes in modulating disease risk, have been suggested. Additional host genes and pathways of importance which may be specific to pouch inflammation or previously undiscovered IBD-associated polymorphisms, may be identified using GWA studies. These studies will be conducted in this cohort combined with cohorts from other centres in the future.

The cross-sectional nature of these studies makes it difficult to determine causality, and the utility of the identified serological and microbial markers for prognostication remains unknown. While the predictive value of SNP markers is high considering that host genetics are stable and do not change over time in response to external stimuli, the same cannot be said regarding either anti-microbial antibody positivity or intestinal microbiome composition. Some studies have shown serological markers to be stable longitudinally, yet to date no well-powered study has shown that pre-surgical anti-microbial antibody profiles are similar following surgery. Additionally, while the microbial profiles identified are definitively associated with pouch inflammation, their role in disease initiation is unclear. As many studies have pointed out, it is possible that the described changes are the result of inflammatory processes, and not contributory towards disease onset or pathogenesis. However, although observed changes may be caused by the inflammatory status of the tissue, the alterations to the microbiome which are characteristic of inflammation may have a role in propagating inflammatory processes and immune dysregulation. Future analyses evaluating both serology and the
microbiome prospectively and involving serial sampling (from multiple sample sites in the case of the intestinal microbiota) will be useful for determining which factors are important in the development of inflammation and which may be suitable for disease diagnostics. Prospective microbiome studies will therefore be useful in determining which organisms may be responsible for disease onset, and how those organisms contribute towards propagating inflammation.

While this study has documented alterations in the microbiome of pouches with inflammation, the functional effect of the loss of specific organisms is unknown. In future, metabolomic analysis should be conducted to evaluate whether reductions in the organisms we have described results in a cumulative reduction in the quantity of butyrate, or other short chain fatty acids available to the host. Further, to determine which additional microbial functional pathways are important in pouch health, it will be important to perform a metagenomic analysis in pouch samples. This will allow us to confirm our hypothesis that dysregulation of the inflammatory response is propagated via a reduction in the production of immune modulating compounds. This approach may also help to highlight additional microbial pathways and metabolites with pathogenic or therapeutic potential.

It would also be useful for patients to have a measure for objectively evaluating risk of disease recurrence following surgery. This score would likely encompass genetic, demographic (smoking history) and phenotypic (presence of extra-intestinal manifestations of disease, medication use) characteristics, pre-surgery serological results, and incorporate early changes to the pouch microbiota prior to, and immediately following ileostomy closure. Since it is unclear whether the ileal or colonic microbiome, or both, contribute to the colonization of the pouch, the utility in measuring the microbiota of these sites prior to surgery for inclusion as part of a risk score is unknown. However, should evidence of predictive capabilities arise, they could be of great benefit.

The field of microbiome analysis is so new, that many different analyses and studies are necessary to get a more comprehensive understanding of the changes which are responsible for disease initiation, propagation and recurrence. Immediate evaluation of the differences between anatomical sites, and between pouches and non-surgical IBD may be useful for determining why pouchitis is responsive to antibiotics in more than
50% of cases, whereas non-surgical IBD is not. Another challenge going forward will be to determine which genetic and microbial factors together contribute to establishing a particular serological profile, and which host factors, be they genetic, dietary, or lifestyle, are important in shaping the composition of the host microbiome. A greater evaluation of these relationships will be critical in obtaining a more comprehensive understanding of the relationship between colonizer and host and how they may impact immune function.

The relationship between the organisms inhabiting the digestive tract, and the host are complex, and are only beginning to be understood. This work shows clear evidence for an etiological role of genetic and microbial factors in pouch inflammatory complications. While the mechanisms which are responsible for disease initiation and propagation remain a mystery, this work has identified similarities and differences between pouch inflammation and non-surgical IBD, with potential implications for a more broad understanding of IBD in general. Future evaluation of the pathways and organisms identified by this work may lead to the generation of novel therapeutics or diagnostic strategies which are capable of enhancing patient care.
Appendix

Chapter 3, Supplementary Table 1: List of SNPs included in this analysis

<table>
<thead>
<tr>
<th>CHR</th>
<th>SNP</th>
<th>Included in Replication</th>
<th>CHR</th>
<th>SNP</th>
<th>Included in Replication</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs1044879</td>
<td></td>
<td>1</td>
<td>rs3845461</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs10489628</td>
<td>Yes</td>
<td>1</td>
<td>rs3845466</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs1061622</td>
<td></td>
<td>1</td>
<td>rs396991</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs1061624</td>
<td></td>
<td>1</td>
<td>rs4266924</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs10789230</td>
<td></td>
<td>1</td>
<td>rs4353135</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs10797887</td>
<td></td>
<td>1</td>
<td>rs447536</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs1089676</td>
<td></td>
<td>1</td>
<td>rs448740</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs1089677</td>
<td></td>
<td>1</td>
<td>rs4645983</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs10911364</td>
<td></td>
<td>1</td>
<td>rs526201</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs10911365</td>
<td></td>
<td>1</td>
<td>rs6426833</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs11209003</td>
<td></td>
<td>1</td>
<td>rs6669582</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs11209008</td>
<td></td>
<td>1</td>
<td>rs6672995</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs11265519</td>
<td></td>
<td>1</td>
<td>rs7517847</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs11465802</td>
<td></td>
<td>1</td>
<td>rs763110</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs11465804</td>
<td></td>
<td>1</td>
<td>rs789180</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs11578964</td>
<td>Yes</td>
<td>1</td>
<td>rs789181</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs11584383</td>
<td></td>
<td>1</td>
<td>rs789185</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs11588654</td>
<td></td>
<td>1</td>
<td>rs789191</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs11811630</td>
<td></td>
<td>1</td>
<td>rs9286879</td>
<td>Yes</td>
</tr>
<tr>
<td>1</td>
<td>rs12035082</td>
<td></td>
<td>1</td>
<td>rs9988642</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs12094228</td>
<td></td>
<td>2</td>
<td>rs10210302</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs12567232</td>
<td></td>
<td>2</td>
<td>rs1048610</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs12753665</td>
<td></td>
<td>2</td>
<td>rs10495565</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs13294</td>
<td>Yes</td>
<td>2</td>
<td>rs1056204</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs1336480</td>
<td></td>
<td>2</td>
<td>rs10929587</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs17129659</td>
<td></td>
<td>2</td>
<td>rs10929590</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs17419032</td>
<td></td>
<td>2</td>
<td>rs1143634</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs1800896</td>
<td></td>
<td>2</td>
<td>rs11684747</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs1801274</td>
<td>Yes</td>
<td>2</td>
<td>rs12469362</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs2064689</td>
<td></td>
<td>2</td>
<td>rs12471449</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs2274910</td>
<td></td>
<td>2</td>
<td>rs13003464</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs2297909</td>
<td></td>
<td>2</td>
<td>rs13412102</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs2343331</td>
<td></td>
<td>2</td>
<td>rs1441090</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs2476601</td>
<td></td>
<td>2</td>
<td>rs16944</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs2490271</td>
<td>Yes</td>
<td>2</td>
<td>rs1839669</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs2819130</td>
<td></td>
<td>2</td>
<td>rs1880439</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs2902440</td>
<td></td>
<td>2</td>
<td>rs2001658</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>rs2241879</td>
<td></td>
<td>4</td>
<td>rs4833103</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>rs2241880</td>
<td></td>
<td>4</td>
<td>rs9999226</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>rs2276338</td>
<td></td>
<td>5</td>
<td>rs1002922</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>rs2289472</td>
<td></td>
<td>5</td>
<td>rs10045431</td>
<td>Yes</td>
</tr>
<tr>
<td>rs2925757</td>
<td>5</td>
<td>rs10075090</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3792106</td>
<td>5</td>
<td>rs10077785</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3828309</td>
<td>Yes</td>
<td>5</td>
<td>rs10277</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs419598</td>
<td>Yes</td>
<td>5</td>
<td>rs1050152</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4464248</td>
<td>5</td>
<td>rs10512734</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4676406</td>
<td>5</td>
<td>rs1065154</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4676410</td>
<td>5</td>
<td>rs11739135</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6431660</td>
<td>5</td>
<td>rs11747270</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6733000</td>
<td>5</td>
<td>rs12521868</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6743984</td>
<td>5</td>
<td>rs1295685</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7603516</td>
<td>5</td>
<td>rs1295686</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs780094</td>
<td>5</td>
<td>rs13361189</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs883399</td>
<td>5</td>
<td>rs1373692</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs917997</td>
<td>5</td>
<td>rs1535375</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1128535</td>
<td>5</td>
<td>rs155788</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11707680</td>
<td>5</td>
<td>rs17234657</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11919639</td>
<td>Yes</td>
<td>5</td>
<td>rs17622208</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12695416</td>
<td>5</td>
<td>rs1800925</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs13067120</td>
<td>5</td>
<td>rs1992660</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs151618</td>
<td>5</td>
<td>rs1992662</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2030582</td>
<td>5</td>
<td>rs20541</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2230590</td>
<td>5</td>
<td>rs2066960</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2276707</td>
<td>5</td>
<td>rs2188962</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3197999</td>
<td>5</td>
<td>rs2241349</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs352140</td>
<td>5</td>
<td>rs2450484</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3732383</td>
<td>5</td>
<td>rs2548993</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3732807</td>
<td>5</td>
<td>rs255969</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3814055</td>
<td>5</td>
<td>rs2569190</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3924462</td>
<td>5</td>
<td>rs2631372</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4625</td>
<td>5</td>
<td>rs272867</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6439924</td>
<td>5</td>
<td>rs272888</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6784598</td>
<td>5</td>
<td>rs2853694</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs864623</td>
<td>5</td>
<td>rs348601</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9858542</td>
<td>Yes</td>
<td>5</td>
<td>rs3734007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10488959</td>
<td>5</td>
<td>rs3792876</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs16853571</td>
<td>5</td>
<td>rs3844312</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs230530</td>
<td>5</td>
<td>rs4495224</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4613763</td>
<td>Yes</td>
<td>6</td>
<td>rs2235673</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4797</td>
<td>6</td>
<td>rs2335674</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4958847</td>
<td>Yes</td>
<td>6</td>
<td>rs2235675</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs502729</td>
<td>6</td>
<td>rs2244502</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs513165</td>
<td>6</td>
<td>rs2267657</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs515110</td>
<td>6</td>
<td>rs2294676</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs565280</td>
<td>6</td>
<td>rs2294677</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6887695</td>
<td>6</td>
<td>rs2294679</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7711505</td>
<td>6</td>
<td>rs2294681</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7720838</td>
<td>6</td>
<td>rs2294683</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9292777</td>
<td>6</td>
<td>rs2295119</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10080901</td>
<td>6</td>
<td>rs2301436</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs1016259</td>
<td></td>
<td>rs231943</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>---</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs10223752</td>
<td>6</td>
<td>rs231944</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs1041981</td>
<td>6</td>
<td>rs231946</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs10484545</td>
<td>6</td>
<td>rs231947</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs10484603</td>
<td>6</td>
<td>rs231948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs10484604</td>
<td>6</td>
<td>rs231952</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs11758155</td>
<td>6</td>
<td>rs231954</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs11758386</td>
<td>6</td>
<td>rs231956</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs11961314</td>
<td>6</td>
<td>rs231959</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs12195525</td>
<td>6</td>
<td>rs231961</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs12196156</td>
<td>6</td>
<td>rs2395182</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs12202693</td>
<td>6</td>
<td>rs2395185</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs12526198</td>
<td>6</td>
<td>rs2395269</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs12661812</td>
<td>6</td>
<td>rs3093661</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs13201732</td>
<td>6</td>
<td>rs3093662</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs14362</td>
<td>6</td>
<td>rs3093664</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs1546894</td>
<td>6</td>
<td>rs3093665</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs162990</td>
<td>6</td>
<td>rs3094204</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs163007</td>
<td>6</td>
<td>rs3117223</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs17309827</td>
<td>6</td>
<td>rs3129882</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs1799724</td>
<td>6</td>
<td>rs3130453</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs1799964</td>
<td>6</td>
<td>rs3130501</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs1800629</td>
<td>6</td>
<td>rs3132946</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs1800750</td>
<td>6</td>
<td>rs3134603</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs2229094</td>
<td>6</td>
<td>rs3134954</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs2235668</td>
<td>6</td>
<td>rs361525</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs2235670</td>
<td>6</td>
<td>rs3749930</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs2235671</td>
<td>6</td>
<td>rs3749946</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs2235672</td>
<td>6</td>
<td>rs3763313</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs4248160</td>
<td>6</td>
<td>rs9480140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs4248161</td>
<td>6</td>
<td>rs995482</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs4334841</td>
<td>7</td>
<td>rs1016189</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs438239</td>
<td>7</td>
<td>rs10238136</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs443198</td>
<td>7</td>
<td>rs10247855</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs4516954</td>
<td>7</td>
<td>rs10278217</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs4645839</td>
<td>7</td>
<td>rs1045642</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs6239</td>
<td>7</td>
<td>rs10499343</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs6557420</td>
<td>7</td>
<td>rs10951982</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs6557421</td>
<td>7</td>
<td>rs1406161</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs660895</td>
<td>7</td>
<td>rs1456893</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs6900372</td>
<td>7</td>
<td>rs17224513</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs6908425</td>
<td>7</td>
<td>rs1860695</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs6919626</td>
<td>7</td>
<td>rs2075820</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs6932514</td>
<td>7</td>
<td>rs2240478</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs6935359</td>
<td>7</td>
<td>rs2302004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs746868</td>
<td>7</td>
<td>rs2302006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs7739889</td>
<td>7</td>
<td>rs2303364</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs7746082</td>
<td>7</td>
<td>rs3789243</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs7754577</td>
<td>7</td>
<td>rs3800752</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs7758080</td>
<td>Yes</td>
<td>7</td>
<td>rs3813517</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------------</td>
<td>-------</td>
<td>------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs7769639</td>
<td>7</td>
<td>rs4720672</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs7774274</td>
<td>7</td>
<td>rs6947579</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs793697</td>
<td>7</td>
<td>rs6951997</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs849884</td>
<td>7</td>
<td>rs6960867</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs849886</td>
<td>7</td>
<td>rs6962966</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs909253</td>
<td>7</td>
<td>rs6964587</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs9268480</td>
<td>7</td>
<td>rs6976629</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs9271568</td>
<td>7</td>
<td>rs702484</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs9275184</td>
<td>7</td>
<td>rs7783685</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs9277554</td>
<td>7</td>
<td>rs836488</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs9322523</td>
<td>7</td>
<td>rs836518</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs9322524</td>
<td>7</td>
<td>rs836547</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs9371890</td>
<td>7</td>
<td>rs836554</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs9384323</td>
<td>7</td>
<td>rs836559</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs9384326</td>
<td>8</td>
<td>rs10505007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs9397819</td>
<td>8</td>
<td>rs11362</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs946964</td>
<td>8</td>
<td>rs11775611</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs9478652</td>
<td>8</td>
<td>rs2044999</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs9480136</td>
<td>8</td>
<td>rs2386922</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs9480137</td>
<td>8</td>
<td>rs2456449</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>rs2979040</td>
<td>10</td>
<td>rs2241833</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>rs4871857</td>
<td>10</td>
<td>rs2289310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>rs6557634</td>
<td>10</td>
<td>rs3218621</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>rs830772</td>
<td>10</td>
<td>rs3936503</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>rs916552</td>
<td>Yes</td>
<td>10</td>
<td>rs4379776</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>rs921720</td>
<td>11</td>
<td>rs10501701</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs10821091</td>
<td>11</td>
<td>rs10765205</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs11534415</td>
<td>11</td>
<td>rs10830273</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs1322054</td>
<td>11</td>
<td>rs10830277</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs1885383</td>
<td>11</td>
<td>rs11018622</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs2295800</td>
<td>11</td>
<td>rs1176713</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs34155071</td>
<td>Yes</td>
<td>11</td>
<td>rs11821838</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>rs3789879</td>
<td>11</td>
<td>rs11826748</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs3789882</td>
<td>11</td>
<td>rs12574346</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs4263839</td>
<td>11</td>
<td>rs12576495</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs4378065</td>
<td>11</td>
<td>rs12799930</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs4574921</td>
<td>11</td>
<td>rs17193643</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs4743484</td>
<td>11</td>
<td>rs17195409</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs4878061</td>
<td>11</td>
<td>rs1793004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs4986790</td>
<td>Yes</td>
<td>11</td>
<td>rs1827428</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs4986791</td>
<td>11</td>
<td>rs2036418</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs7849191</td>
<td>11</td>
<td>rs2164521</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>rs7868736</td>
<td>11</td>
<td>rs2164522</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs10509093</td>
<td>11</td>
<td>rs2712800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs1058198</td>
<td>11</td>
<td>rs317147</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs10761659</td>
<td>Yes</td>
<td>11</td>
<td>rs317155</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs10763976</td>
<td>Yes</td>
<td>11</td>
<td>rs317185</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs10826240</td>
<td>11</td>
<td>rs317194</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs10883365</td>
<td>11</td>
<td>rs3793973</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------------</td>
<td>-----</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs10886462</td>
<td>Yes</td>
<td>11</td>
<td>rs3897708</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs10995239</td>
<td>11</td>
<td>rs4127488</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs10995271</td>
<td>11</td>
<td>rs492840</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs11187736</td>
<td>11</td>
<td>rs495344</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs11190140</td>
<td>11</td>
<td>rs497279</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs12360212</td>
<td>11</td>
<td>rs521765</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs1248696</td>
<td>11</td>
<td>rs564699</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs17582416</td>
<td>11</td>
<td>rs595518</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs1800682</td>
<td>11</td>
<td>rs6483779</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs2165046</td>
<td>11</td>
<td>rs671492</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs2234978</td>
<td>11</td>
<td>rs7927894</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs224136</td>
<td>11</td>
<td>rs7929532</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>rs7933900</td>
<td>15</td>
<td>rs1365242</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>rs7947257</td>
<td>15</td>
<td>rs1617207</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>rs8176785</td>
<td>15</td>
<td>rs1648281</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>rs8176786</td>
<td>15</td>
<td>rs1648282</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>rs10735810</td>
<td>15</td>
<td>rs1648305</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>rs11175593</td>
<td>15</td>
<td>rs166575</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>rs11181674</td>
<td>Yes</td>
<td>15</td>
<td>rs16939752</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>rs11568820</td>
<td>15</td>
<td>rs16939755</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>rs1544410</td>
<td>15</td>
<td>rs16977681</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>rs1558744</td>
<td>15</td>
<td>rs16977686</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 12  | rs2228224  | Yes | 15  | rs16977750| Yes
| 12  | rs2234649  | 15  | rs16977754|
| 12  | rs2292657  | 15  | rs1706765 |
| 12  | rs4149621  | 15  | rs1706803 |
| 12  | rs731236   | 15  | rs1706810 |
| 12  | rs767455   | 15  | rs1706813 | Yes
| 12  | rs7975232  | 15  | rs175088  |
| 12  | rs886898   | 15  | rs17594910|
| 13  | rs11617463 | 15  | rs17595239|
| 13  | rs17837226 | 15  | rs1863469 |
| 13  | rs2185150  | 15  | rs1898298 |
| 13  | rs3764147  | 15  | rs2001616 |
| 13  | rs9545740  | 15  | rs2020216 |
| 14  | rs11832    | 15  | rs2068982 |
| 14  | rs1188157  | 15  | rs2252479 |
| 14  | rs1243160  | 15  | rs2292464 |
| 14  | rs1303     | 15  | rs2458236 | Yes
| 14  | rs17580    | Yes | 15  | rs2470666 |
| 14  | rs17751769 | 15  | rs2470667 |
| 14  | rs8647     | 15  | rs2554452 |
| 14  | rs709932   | Yes | 15  | rs2554459 |
| 14  | rs8004738  | 15  | rs2576090 |
| 15  | rs11072060 | 15  | rs2576092 |
| 15  | rs11072064 | 15  | rs269856  |
| 15  | rs11637215 | 15  | rs269863  |
| 15  | rs11856030 | 15  | rs269866  | Yes
<table>
<thead>
<tr>
<th>Chromosome</th>
<th>SNP ID</th>
<th>Value</th>
<th>Chromosome</th>
<th>SNP ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>rs11858195</td>
<td>Yes</td>
<td>15</td>
<td>rs311893</td>
</tr>
<tr>
<td>15</td>
<td>rs12442417</td>
<td>15</td>
<td>rs3784576</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs12908110</td>
<td>15</td>
<td>rs3784577</td>
<td>Yes</td>
</tr>
<tr>
<td>15</td>
<td>rs12910807</td>
<td>15</td>
<td>rs438866</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs13380419</td>
<td>15</td>
<td>rs4774518</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs4774519</td>
<td>17</td>
<td>rs1053023</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs4777102</td>
<td>17</td>
<td>rs11080344</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs6493135</td>
<td>17</td>
<td>rs1137933</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs6493136</td>
<td>17</td>
<td>rs11653716</td>
<td>Yes</td>
</tr>
<tr>
<td>15</td>
<td>rs6495116</td>
<td>17</td>
<td>rs16967637</td>
<td>Yes</td>
</tr>
<tr>
<td>15</td>
<td>rs687128</td>
<td>Yes</td>
<td>17</td>
<td>rs17722851</td>
</tr>
<tr>
<td>15</td>
<td>rs7175817</td>
<td>17</td>
<td>rs2071214</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs7177539</td>
<td>17</td>
<td>rs2293152</td>
<td>Yes</td>
</tr>
<tr>
<td>15</td>
<td>rs7183046</td>
<td>17</td>
<td>rs2297515</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs751151</td>
<td>17</td>
<td>rs2297516</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs8025912</td>
<td>17</td>
<td>rs2297518</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs8038236</td>
<td>17</td>
<td>rs2306580</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs8041911</td>
<td>17</td>
<td>rs2314809</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs905614</td>
<td>17</td>
<td>rs2314810</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>rs916977</td>
<td>17</td>
<td>rs2872507</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs10431923</td>
<td>Yes</td>
<td>17</td>
<td>rs3729508</td>
</tr>
<tr>
<td>16</td>
<td>rs1049255</td>
<td>17</td>
<td>rs3730013</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs10521209</td>
<td>17</td>
<td>rs3730017</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs12597188</td>
<td>17</td>
<td>rs3744508</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs12933505</td>
<td>17</td>
<td>rs3794756</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs13333449</td>
<td>17</td>
<td>rs3794764</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs16960173</td>
<td>17</td>
<td>rs4794986</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs1968752</td>
<td>17</td>
<td>rs4795067</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs2066843</td>
<td>Yes</td>
<td>17</td>
<td>rs6503695</td>
</tr>
<tr>
<td>16</td>
<td>rs2066844</td>
<td>Yes</td>
<td>17</td>
<td>rs6503698</td>
</tr>
<tr>
<td>16</td>
<td>rs2066845</td>
<td>Yes</td>
<td>17</td>
<td>rs7211777</td>
</tr>
<tr>
<td>16</td>
<td>rs2066847</td>
<td>Yes</td>
<td>17</td>
<td>rs744166</td>
</tr>
<tr>
<td>16</td>
<td>rs2076756</td>
<td>Yes</td>
<td>17</td>
<td>rs8072199</td>
</tr>
<tr>
<td>16</td>
<td>rs3180279</td>
<td>17</td>
<td>rs8077981</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs3785273</td>
<td>17</td>
<td>rs944725</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs3794624</td>
<td>17</td>
<td>rs9906835</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs4673</td>
<td>Yes</td>
<td>17</td>
<td>rs9910698</td>
</tr>
<tr>
<td>16</td>
<td>rs5743289</td>
<td>17</td>
<td>rs991804</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs6564261</td>
<td>18</td>
<td>rs2542151</td>
<td>Yes</td>
</tr>
<tr>
<td>16</td>
<td>rs8050910</td>
<td>18</td>
<td>rs7228236</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs8460</td>
<td>18</td>
<td>rs8098673</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs9932581</td>
<td>18</td>
<td>rs937815</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs9935563</td>
<td>18</td>
<td>rs9949882</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>rs9940427</td>
<td>19</td>
<td>rs1020075</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>rs10459953</td>
<td>19</td>
<td>rs11880212</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>rs10512488</td>
<td>19</td>
<td>rs11881691</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>rs1457092</td>
<td>22</td>
<td>rs13058338</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>rs1545620</td>
<td>22</td>
<td>rs1476002</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>rs17130</td>
<td>22</td>
<td>rs1883112</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4, Supplementary Table 1: Endoscopic and histological characteristics included in the objective pouchitis score used to define inflammation.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Severity</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erythema</td>
<td>None, mild, severe</td>
<td>0, 2, 3</td>
</tr>
<tr>
<td>Friability</td>
<td>None, mild, severe</td>
<td>0, 1, 2</td>
</tr>
<tr>
<td>Ulceration</td>
<td>None, mild, severe</td>
<td>0, 2, 3</td>
</tr>
<tr>
<td>Polymorphonuclear leukocyte infiltration</td>
<td>None, discrete and patchy (largely confined to surface)</td>
<td>0, 1, 2, 3</td>
</tr>
<tr>
<td>Ulcerations/Erosions</td>
<td>Total possible score</td>
<td>Total score for inflammation</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>epithelium), moderate (±) crypt abscesses or cryptitis, extensive (±) crypt abscesses or cryptitis</td>
<td>14</td>
<td>&gt;3</td>
</tr>
</tbody>
</table>
Chapter 4, Supplementary Table 2: Primers used for pyrosequencing and qPCR experiments. Primers listed 5’-3’.

<table>
<thead>
<tr>
<th>Primer Target</th>
<th>Primer Names</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>VI-V3 specific universal primer set</td>
<td>Gray 28F / Gray 519R&lt;sup&gt;353&lt;/sup&gt;</td>
<td>GAGTTTGATCNGCCTGAG</td>
<td>GTNTTACNGCNGGGCTGAG</td>
</tr>
<tr>
<td><em>Eubacteria</em></td>
<td>341-357F / 518-534R&lt;sup&gt;355&lt;/sup&gt;</td>
<td>CCTACGGAGGCAGCAG</td>
<td>ATTACCGGCTGCTGG</td>
</tr>
<tr>
<td><em>Bacteroides spp.</em></td>
<td>Bac303F / Bfr-Fmrev&lt;sup&gt;389&lt;/sup&gt;</td>
<td>GAAGGTCCCCACATTG</td>
<td>CGCKACTTGGCTGGTCAG</td>
</tr>
<tr>
<td><em>Roseburia spp. and E. rectale</em></td>
<td>RrecF / Rrec630mR&lt;sup&gt;389&lt;/sup&gt;</td>
<td>GCCGTRCGGCAAGTCTGA</td>
<td>CCTCCGACGCTCTAGTCGAC</td>
</tr>
<tr>
<td>Clostridial cluster IV</td>
<td>Clep866mF / Clept1240m R&lt;sup&gt;389&lt;/sup&gt;</td>
<td>TTAACAAACTAGTATCCACTGG</td>
<td>ACCTTCCTCCGTTTGTCAAC</td>
</tr>
<tr>
<td><em>Faecalibacterium prausnitzii</em></td>
<td>FPR-2F / Fprau645R&lt;sup&gt;389&lt;/sup&gt;</td>
<td>GGAGGAAGAGGTCTCGG</td>
<td>AATCCGCTACCTCTGCACT</td>
</tr>
<tr>
<td><em>Adherent invasive E. coli (AIEC)</em></td>
<td>EcoliFimH2F / EcoliFimH2 R&lt;sup&gt;355&lt;/sup&gt;</td>
<td>GCCGTTGGGCCTTTATGTG</td>
<td>TCATCGGCTGTTATAGTTGTTGTCT</td>
</tr>
</tbody>
</table>
Chapter 4, Supplementary Table 3: Mean diversity measures across all four phenotypic outcome groups. Values expressed are means and standard error of the mean for pouch and afferent limb samples.

<table>
<thead>
<tr>
<th></th>
<th>Pouch</th>
<th>Afferent Limb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Parametric Shannon Diversity Index</td>
<td>Inverse Simpson Diversity Index</td>
</tr>
<tr>
<td>FAP</td>
<td>4.15 (0.13)</td>
<td>25.70 (4.82)</td>
</tr>
<tr>
<td>No pouchitis</td>
<td>3.76 (0.15)</td>
<td>19.94 (3.06)</td>
</tr>
<tr>
<td>Pouchitis</td>
<td>3.29 (0.17)</td>
<td>12.14 (1.89)</td>
</tr>
<tr>
<td>CDL</td>
<td>3.76 (0.11)</td>
<td>20.16 (2.42)</td>
</tr>
<tr>
<td>P-value (4 group comparison)</td>
<td>0.006</td>
<td>0.04</td>
</tr>
</tbody>
</table>
Chapter 4, Supplementary Table 4: Description of the association between bacterial abundance of the four dominant phyla among pouch and afferent limb samples. Nominal p-values are depicted for the overall analysis and for pairwise comparisons. Highlighted rows are those which were significant using the Kruskal-Wallis test and FDR correction for multiple testing. No significant associations were observed in the FAP vs No pouchitis or pouchitis vs. CDL comparisons.

<table>
<thead>
<tr>
<th>Location</th>
<th>Phylum</th>
<th>P-value</th>
<th>FAP vs Pouchitis</th>
<th>FAP vs CDL</th>
<th>No pouchitis vs Pouchitis</th>
<th>No pouchitis vs CDL</th>
<th>4-way analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pouch</td>
<td>Firmicutes</td>
<td>0.64</td>
<td>0.76</td>
<td>0.74</td>
<td>0.51</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>0.02</td>
<td>4.0x10^{-3}</td>
<td>0.19</td>
<td>0.03</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bacteroidetes</td>
<td>1.0x10^{-4}</td>
<td>1.0x10^{-4}</td>
<td>5.0x10^{-4}</td>
<td>1.0x10^{-3}</td>
<td>1.0x10^{-4}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fusobacteria</td>
<td>0.27</td>
<td>0.14</td>
<td>0.79</td>
<td>0.83</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>Afferent Limb</td>
<td>Firmicutes</td>
<td>0.66</td>
<td>1.0</td>
<td>0.66</td>
<td>0.90</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>0.06</td>
<td>0.01</td>
<td>0.19</td>
<td>0.07</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bacteroidetes</td>
<td>1.0x10^{-4}</td>
<td>2.0x10^{-4}</td>
<td>1.0x10^{-3}</td>
<td>0.01</td>
<td>1.0x10^{-4}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fusobacteria</td>
<td>0.91</td>
<td>0.91</td>
<td>0.58</td>
<td>0.89</td>
<td>0.96</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4, Supplementary Table 5: Genera achieving a nominal significance level of $P<0.05$ in four-way comparison. FET=Fisher’s exact test (dichotomous results); KWT=Kruskal-Wallis Test (continuous results). Highlighted rows are those which remained significant in either analysis after FDR correction for multiple testing.

<table>
<thead>
<tr>
<th>Location</th>
<th>Phylum</th>
<th>Genus</th>
<th>FET P-value</th>
<th>FET P-value (corr)</th>
<th>KWT P-value</th>
<th>KWT P-value (corr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afferent Limb</td>
<td>Actinobacteria</td>
<td>Uncultured Coriobacteriaceae</td>
<td>0.03</td>
<td>0.17</td>
<td>0.03</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Bacteroidetes</td>
<td>Bacteroides</td>
<td>6.9x10^{-6}</td>
<td>2.9x10^{-4}</td>
<td>4.2x10^{-6}</td>
<td>3.4x10^{-4}</td>
</tr>
<tr>
<td></td>
<td>Firmicutes</td>
<td>Anaerococcus</td>
<td>2.8x10^{-3}</td>
<td>0.05</td>
<td>8.4x10^{-3}</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IS Erysipelotrichaceae</td>
<td>4.5x10^{-3}</td>
<td>0.05</td>
<td>4.0x10^{-3}</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Roseburia</td>
<td>0.02</td>
<td>0.15</td>
<td>0.08</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Uncultured Family XIII IS</td>
<td>0.01</td>
<td>0.09</td>
<td>0.03</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Sutterella IS</td>
<td>0.01</td>
<td>0.09</td>
<td>0.03</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Haemophilus</td>
<td>0.13</td>
<td>0.39</td>
<td>5.9x10^{-3}</td>
<td>0.12</td>
</tr>
<tr>
<td>Pouch</td>
<td>Actinobacteria</td>
<td>Corynebacterium</td>
<td>0.02</td>
<td>0.15</td>
<td>0.03</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Bacteroidetes</td>
<td>Bacteroides</td>
<td>3.6x10^{-6}</td>
<td>3.0x10^{-4}</td>
<td>5.7x10^{-7}</td>
<td>4.7x10^{-3}</td>
</tr>
<tr>
<td></td>
<td>Firmicutes</td>
<td>Parabacteroides</td>
<td>1.2x10^{-4}</td>
<td>4.9x10^{-3}</td>
<td>7.8x10^{-4}</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Morayella</td>
<td>2.5x10^{-3}</td>
<td>0.05</td>
<td>1.4x10^{-3}</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Blautia</td>
<td>3.0x10^{-3}</td>
<td>0.05</td>
<td>2.3x10^{-4}</td>
<td>9.5x10^{-3}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dorea</td>
<td>0.02</td>
<td>0.13</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Parabacteroides</td>
<td>9.6x10^{-3}</td>
<td>0.10</td>
<td>0.02</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IS Erysipelotrichaceae</td>
<td>0.02</td>
<td>0.14</td>
<td>0.03</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dialis</td>
<td>0.02</td>
<td>0.13</td>
<td>0.02</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Uncultured Lachnospiraceae</td>
<td>0.04</td>
<td>0.20</td>
<td>6.9x10^{-3}</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Uncultured Ruminococaceae</td>
<td>0.04</td>
<td>0.20</td>
<td>7.1x10^{-3}</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Subdoligranulum</td>
<td>0.01</td>
<td>0.13</td>
<td>0.03</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Uncultured Family XIII IS</td>
<td>8.5x10^{-3}</td>
<td>0.10</td>
<td>7.7x10^{-3}</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Roseburia</td>
<td>0.07</td>
<td>0.31</td>
<td>0.04</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sutterella IS</td>
<td>2.5x10^{-3}</td>
<td>0.05</td>
<td>2.4x10^{-3}</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rhodobacter</td>
<td>0.01</td>
<td>0.09</td>
<td>0.03</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Haemophilus</td>
<td>0.13</td>
<td>0.39</td>
<td>5.9x10^{-3}</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>Sutterella</td>
<td>1.7x10^{-3}</td>
<td>0.03</td>
<td>8.5x10^{-4}</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rhodobacter</td>
<td>0.03</td>
<td>0.17</td>
<td>0.05</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Haemophilus</td>
<td>0.12</td>
<td>0.37</td>
<td>0.05</td>
<td>0.21</td>
</tr>
</tbody>
</table>
Chapter 2, Supplementary figure 1: Mean antibody titres expressed in EU/mL. Error bars represent standard errors. Titres of anti-CBir1 were significantly associated with outcome (P=0.001)
Chapter 4, Supplementary figure 1: Distribution of inflammatory activity scores through each of the phenotypic outcome groups. Mean values for each group are indicated with a red line. * indicates results which were significantly increased compared to FAP and no pouchitis in the pouch, and to FAP, no pouchitis and pouchitis in the afferent limb.
Chapter 4, Supplementary figure 2: Phylum level comparisons between four outcome groups for the pouch and afferent limb. FAP=familial adenomatous polyposis, CDL=Crohn’s disease-like.
Chapter 4, Supplementary figure 3: Proportion of patients positive for genera which were significantly associated with outcome ($P_{corr}<0.05$) following removal of individuals without evidence of inflammation at study pouchoscopy from the CDL group. * represent pairwise comparisons which were significant ($P_{corr} < 0.05$) after correction. A) Pouch samples. B) Afferent limb samples. I.S. = Incertae sedis.
Chapter 4, Supplementary figure 4: Mean and standard error of the frequency of genera significantly associated with outcome ($P_{corr} < 0.05$) in A) pouch and B) afferent limb. FAP=familial adenomatous polyposis, CDL=Crohn’s disease-like. Panel i) Bacteroides, ii) other significant organisms.
Chapter 4, Supplementary figure 5: Proportion of patients positive for genera which were significantly associated with outcome ($P_{corr} < 0.05$) in the preliminary analysis, among the cohort of patients on no antibiotic therapy (n=57). * represent pairwise comparisons which were significant ($P_{corr} < 0.05$) after correction. A) Pouch samples. B) Afferent limb samples. I.S. = Incertae sedis.
Chapter 4, Supplementary Figure 6: A) Cladograms demonstrating the results obtained through LDA Effect Size (LEfSe) analysis. Highlighted results are those which were increased in the corresponding group. Significant associations are highlighted with outcome group(s) with increased proportions indicated. B) Differential abundance of organisms detected at significantly different frequencies via LEfSe.
Cladogram – Pouch samples, four group comparison

Significant associations as highlighted, with outcome group(s) where increased proportion was detectable indicated.
Cladogram – Pouch samples, FAP vs Pouchitis

Significant associations as highlighted, with outcome group(s) where increased proportion was detectable indicated.
Cladogram – Pouch samples, FAP vs Crohn’s disease-like

Significant associations as highlighted, with outcome group(s) where increased proportion was detectable indicated.
Cladogram – Pouch samples, No Pouchitis vs Crohn’s disease-like

Significant associations as highlighted, with outcome group(s) where increased proportion was detectable indicated.
Cladogram – Pouch samples, No Pouchitis vs Pouchitis

Significant associations as highlighted, with outcome group(s) where increased proportion was detectable indicated.
Cladogram – Afferent Limb samples, four group comparison

Significant associations as highlighted, with outcome group(s) where increased proportion was detectable indicated
Cladogram – Afferent Limb samples, FAP vs Pouchitis

Significant associations as highlighted, with outcome group(s) where increased proportion was detectable indicated.
Cladogram – Afferent Limb samples, FAP vs Crohn’s disease-like

Significant associations as highlighted, with outcome group(s) where increased proportion was detectable indicated
Significant associations as highlighted, with outcome group(s) where increased proportion was detectable indicated.
Significant associations as highlighted, with outcome group(s) where increased proportion was detectable indicated.
Chapter 4, Supplementary Figure 7: log$_2$ transformed relative abundance of organisms of interest from real-time quantitative PCR in the A) pouch and B) afferent limb. Significant results are marked with an astrix ($P_{corr}<0.05$). FAP=familial adenomatous polyposis, CDL=Crohn’s disease-like.
Chapter 4, Supplementary figure 8: Proportion of individuals in each outcome group positive for genera which were previously associated with pouch inflammatory outcomes. FAP=familial adenomatous polyposis, Quiescent=seven individuals requiring long-term medical therapy to maintain remission. Lines represent associations between the quiescent group and others which reached nominal significance (P<0.05). A) Pouch samples. B) Afferent limb samples.
A

Proportion of Patients Positive

Genus

B

Proportion of Patients Positive

Genus
References


45. Buckland MS, Mylonaki M, Rampton D, et al. Serological markers (anti-
Saccharomyces cerevisiae mannan antibodies and antineutrophil cytoplasmic
antibodies) in inflammatory bowel disease: diagnostic utility and phenotypic

inflammatory bowel disease: clinical role and review of the literature. Dis Colon

47. Adams RJ, Heazlewood SP, Gilshenan KS, et al. IgG antibodies against common
gut bacteria are more diagnostic for Crohn's disease than IgG against mannan or

Crohn's disease into immunologically homogeneous subgroups with distinct

cytoplasmic antibodies in patients with Crohn's disease define a clinical subgroup.

50. Zholudev A, Zurakowski D, Young W, et al. Serologic testing with ANCA,
ASCA, and anti-OmpC in children and young adults with Crohn's disease and
ulcerative colitis: diagnostic value and correlation with disease phenotype. Am J

51. Amre DK, Lu SE, Costea F, et al. Utility of serological markers in predicting the
early occurrence of complications and surgery in pediatric Crohn's disease

52. Ferrante M, Henckaerts L, Joossens M, et al. New serological markers in
inflammatory bowel disease are associated with complicated disease behaviour.

bowel disease are associated with earlier age at onset, complicated disease
behavior, risk for surgery, and NOD2/CARD15 genotype in a Hungarian IBD

antibodies combined with antineutrophil cytoplasmic autoantibodies in

55. Peyrin-Biroulet L, Standaert-Vitse A, Branche J, et al. IBD serological panels:

outer membrane porin C in relatives of patients with Crohn's disease.
Gastroenterology 2006;130:1078-85.

57. Dubinsky MC, Lin YC, Dutridge D, et al. Serum immune responses predict rapid
disease progression among children with Crohn's disease: immune responses


179. Salonen A, Nikkila J, Jalanka-Tuovinen J, et al. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of


266. Maslanik T, Tannura K, Mahaffey L, et al. Commensal bacteria and MAMPs are necessary for stress-induced increases in IL-1beta and IL-18 but not IL-6, IL-10 or MCP-1. PLoS ONE 2012;7:e50636.


317. Waterman MS, JM.; Kabakchiev, B.; Xu, W.; Griffiths, AM.; Greenberg, GR.; Steinhart, AH.; Bernstein, CN.; Silverberg, MS. Genotypic Overlap Between
Crohn's disease (CD) and ulcerative colitis (UC) in a Large Cohort of Patients with Inflammatory Bowel Disease (IBD). Gut 2010;59:A60.


